

PhD

3rd
CYCLE

FCUP
2025

U.PORTO

MetaDef - The Evolution of the Chemical
DEFensome in METAzoa

André Miguel Moreira Machado

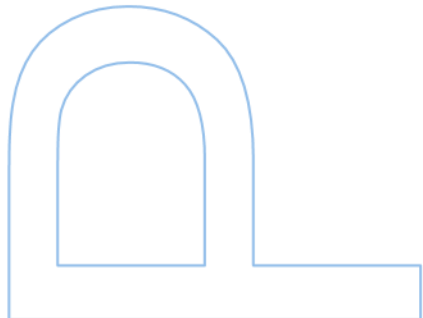
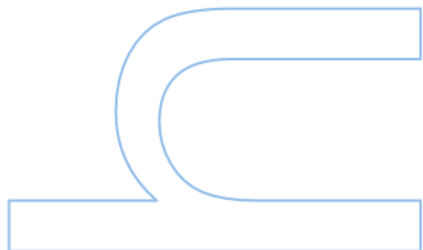
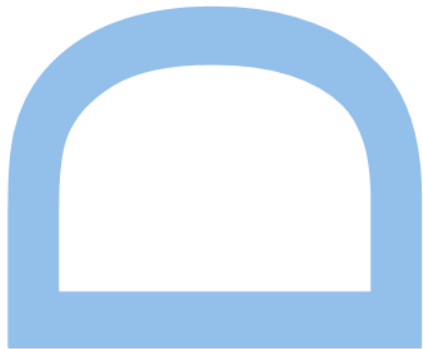
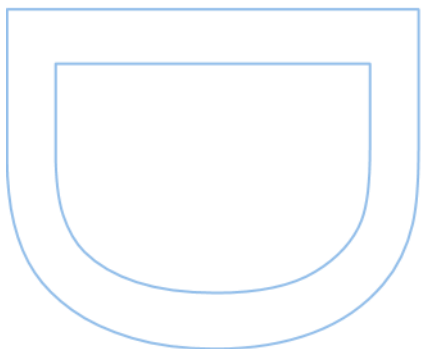
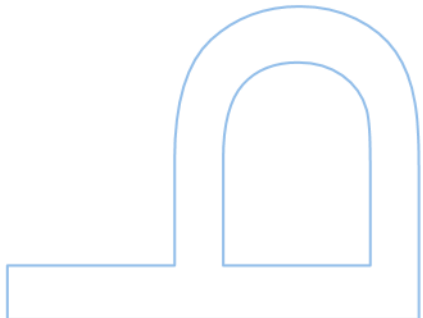
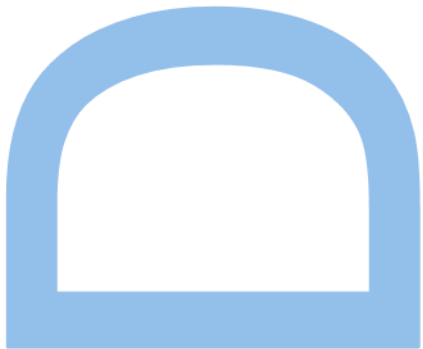
FC



MetaDef

The Evolution of the Chemical DEFensome in METAzoa

André Miguel Moreira Machado
Biology Doctoral Program
Biology Department
Faculty of Sciences of the University of Porto
2025



MetaDef

The Evolution

of the Chemical

DEFensome

in METAzoa

André Miguel Moreira Machado

Biology Doctoral Program
Biology Department
Faculty of Sciences of the University of Porto
2025

Supervisor

Luís Filipe Costa de Castro , Assistant Professor at
FCUP and Researcher at CIIMAR

Co- Supervisor

Miguel Alberto Machado e Santos, Professor at
FCUP and Researcher at CIIMAR

Raquel Frade Ruivo, Researcher at CIIMAR



Ao meu núcleo, Ana e Martim.

Agradecimentos

A dissertação aqui apresentada resulta de um esforço conjunto de mais de cinco anos de trabalho, meu e de muitas outras pessoas que se cruzaram comigo nesta jornada, tanto a nível pessoal como profissional. Toda e qualquer pessoa que se tenha cruzado, e que por lapso não seja explicitamente aqui mencionada, deixo o meu obrigado.

Primeiramente, quero agradecer à Fundação para a Ciência e a Tecnologias (FCT), à Faculdade de Ciências da Universidade do Porto (FCUP) e ao Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR) pelo financiamento e disponibilização de infraestruturas necessárias para a realização deste trabalho. Além disso, o meu muito obrigado a toda a comunidade do CIIMAR, que desempenhou um papel importante em todo o processo de realização desta dissertação.

Como não poderia deixar de ser, quero também agradecer a toda a equipa de orientação que me guiou e apoiou durante os últimos quatro anos – Dr. Filipe Castro, Dra. Raquel Ruivo e Dr. Miguel Santos. Cada um de vós foi essencial num ou múltiplos níveis da realização desta dissertação, e por isso vos agradeço.

Ao meu orientador e mentor Filipe, quero agradecer por tudo o que me proporcionou, tanto na diversidade como na variedade de tópicos, pelo apoio e por me ter feito crescer enquanto pessoa e profissional, não só durante esta dissertação, mas durante os últimos 10 anos. A sua mentoria e visão curiosa e inovadora guiaram e moldaram a forma como vejo e faço ciência, levando-me a ser melhor cientista e a fazer um balanço cuidado entre pragmatismo e detalhe. Obrigado ainda pela exigência e forma crítico-constructiva com que abordas cada assunto, tanto a nível profissional como pessoal. Foi por todos estes motivos que fiz questão de te incluir na equipa de orientação e, sem dúvida, foi a decisão correta.

À minha coorientadora e colega Raquel, agradeço todo o apoio incondicional que me deste durante toda esta jornada. Para além de seres uma cientista fenomenal e uma inspiração, contribuíste muito para a forma meticulosa como abordo cada tópico e estudo que fazemos. Não poderia deixar de agradecer também todo o esforço que fizeste e o papel crucial que tiveste durante todo o processo, em específico neste último ano 2025, onde sempre acreditaste de forma otimista que o processo ia ser concluído, independentemente das razões para que o mesmo estivesse mais moroso.

Ao meu coorientador Miguel Santos, agradeço a postura compreensiva e otimista, bem como a confiança depositada em mim durante todo o processo. Agradeço ainda

todo o apoio logístico e burocrático que me forneceu, bem como a forma expedita com que o mesmo me foi dado.

De seguida, quero agradecer a todos os colegas dos grupos AGE, AEE e EDEC, por terem feito ou fazerem parte deste percurso e por me terem ajudado a ser melhor pessoa e cientista, entre os quais Raul Fonseca, Elza Fonseca, Mónica Lopes Marques, Miguel Cordeiro, Diogo Oliveira, David Barros, Duarte Gonçalves, Teresa Neuparth, e colegas que já não estão no CIIMAR, mas foram essenciais para este percurso como Giulia Secci e Joana Soares. Estes agradecimentos estendem-se também a todos os estudantes com os quais tive oportunidade de conviver e apoiar de alguma forma durante o seu percurso académico, entre eles o Luís Alves, Marcos Domingues, Rodrigo Samico, Filipe Caramelo e mais recentemente o meu “pupilo” João Moutinho. Vocês ensinaram-me que ensinar é uma das melhores e mais eficientes formas de aprender.

Adicionalmente, quero agradecer a todos os excelentes investigadores do CIIMAR e de outras instituições com quem tive oportunidade de colaborar, entre eles Rute da Fonseca, Sérgio Fernández Boo, Ana Veríssimo, Óscar Moroing e Jonathan Wilson.

Aos meus amigos e núcleo no CIIMAR, Rui Pinto, André Gomes-dos-Santos, Manuel Nande e Elsa Froufe, quero agradecer a oportunidade e honra que me deram de trabalharmos juntos, aprendermos, crescermos e confienciarmos durante todo este percurso. Para além de serem cientistas excecionais, são humanos incríveis e foram absolutamente essenciais para que esta dissertação visse a luz do dia. Entre conversas, sobre política, imobiliário, astros, agricultura, entre muitos outros tópicos, o vosso apoio e força constante foram cruciais para completar todas as tarefas e desafios que foram surgindo. Ainda, o vosso otimismo e boa disposição constante foram e continuam a ser basilares para todo o meu crescimento enquanto pessoa e profissional. Sem dúvida, vocês são uma inspiração dentro e fora do CIIMAR. Por tudo isto, o meu obrigado.

Ao meu núcleo familiar,

À minha companheira e futura esposa Ana Ferreira, esta dissertação é tanto minha como tua. Entre todos os momentos, de alegria, lágrimas e suor, tu foste sempre o meu porto de abrigo. Obrigado por todo o amor, acolhimento, cumplicidade, amizade e força que me deste durante todo o processo. Para mim, é um privilégio, poder conviver diariamente e partilhar vida contigo, enquanto lutamos pelos nossos objetivos. Ainda, é absolutamente inspirador ver a forma lutadora como abraças todos os desafios e, à tua maneira, me ajudas a fazer o mesmo. Sem dúvida, és essencial na minha vida. Ao meu filho Martim Machado, que nasceu poucos meses antes deste desafio ter início, quero agradecer todos os momentos de felicidade, desafios e a forma como moldaste a minha

visão do mundo. Apesar de seres uma pequena pessoa, fizeste-me crescer exponencialmente em pouco tempo e foste um dos principais motores para que tudo isto fosse alcançado e finalizado. Obrigado, meu filho.

Ao meu pai João Machado, quero agradecer por ter sempre acreditado em mim. Ao meu irmão Ricardo Machado e cunhada Susana, quero agradecer todo o apoio que me deram. Ricardo, obrigado pelo suporte incondicional, amor e forma como me ajudaste a ultrapassar esta fase. Tu contribuístes de forma massiva para que eu conseguisse terminar esta dissertação e para que consiga sempre chegar mais longe, em todos os meus objetivos. À minha mãe, quero agradecer por ser uma das pedras basilares da minha vida, não só em termos profissionais, mas principalmente em termos pessoais. O teu suporte faz-me acreditar que vou chegar sempre onde quero, e que tudo o que parece difícil se torna acessível se mudarmos a perspetiva. Obrigado por me acompanhares, acolheres e dares-me força para vencer todo e qualquer obstáculo na vida. À minha madrinha Alexandra, ao Paulo e ao meu primo Gonçalo, obrigado pelos bons momentos partilhados, que, embora poucos, foram significativos. Aos meus sogros Vitória Oliveira e João Ferreira, aos tios Paulo, Rosário e Eduardo, quero agradecer todo o carinho, consideração e acima de tudo a forma como acreditaram que conseguiria completar esta fase com sucesso.



Resumo

Durante milhões de anos, os sistemas de defesa contra químicos em metazoários evoluíram de forma a adaptar-se, resistir e integrar diversos ambientes, e em alguns casos reutilizar esses químicos como parte do seu próprio sistema. A arquitetura destes sistemas contempla um conjunto de genes e vias metabólicos conservados evolutivamente, que permitem aos organismos metabolizar e reagir a compostos químicos independentemente da sua origem, seja ela ambiental ou proveniente de organismos. Este conjunto de genes e vias metabólicos é comumente conhecido como defensoma químico (CD). Este sistema molecular compreende uma rede de genes evolutivamente conservados, fatores de transcrição e enzimas que permitem aos organismos sobreviver, adaptarem-se e proliferarem em diversos ecossistemas e ambientes.

Apesar da amplitude e diversidade do CD, desde o aparecimento do “Antropoceno”, o sistema do CD tem sido desafiado numa escala sem precedentes devido à libertação de compostos químicos de origem antropogénica e de outras atividades humanas. Naturalmente, todo este cenário tem despertado o interesse científico em compreender como o CD está organizado, como interage com outros sistemas biológicos e de que forma impacta os diferentes níveis organizacionais biológicos. Não obstante, e apesar da sua importância e dos esforços científicos mais recentes, o conhecimento sobre estes componentes, vias e mecanismos integrados do CD continua limitado e pouco caracterizado em muitas espécies não-modelo

Esta tese aborda precisamente a lacuna anteriormente mencionada através da expansão e identificação de variações na composição génica global do CD. Neste contexto, abordei o tema através da expansão de recursos genómicos e transcriptómicos em pontos chave da filogenia dos metazoários. Por forma a validar as diferentes estratégias utilizadas, explorei e cataloguei alguns elementos de duas famílias génicas centrais no CD – os recetores nucleares (NRs) e as monooxigenases dependentes de flavina (FMOs) – em quatro espécies de deuterostómios vertebrados e invertebrados, regularmente classificados como não modelos. Estas análises geraram novo conhecimento, contribuindo para a compreensão da história evolutiva, relevância funcional e integração de vários elementos do CD com outros sistemas biológicos, incluindo o imunológico e lipídico.

As análises genómicas permitiram expandir o reportório génico do CD em vertebrados e invertebrados, através da identificação e catalogação de vários NRs e

FMOs nas espécies *Scomber colias* e *Hydrolagus colliei*, bem como clarificar e identificar duplicações génicas e padrões de presença-ausência, bem como elementos conservados que refletem a sua trajetória evolutiva. As abordagens de transcriptómica nas dos invertebrados, *Hediste diversicolor* e *Paracentrotus lividus*, clarificaram ainda a sincronia espaço-temporal entre os NRs, FMOs, e genes relacionados, bem como o seu potencial papel na coordenação da homeostase lipídica, reprodução e metabolismo de xenobióticos. Os diferentes padrões de coexpressão reforçaram ainda esta potencial sincronia entre NRs heterodimerizantes e coativadores, fornecendo assim alvos moleculares para futuras análises, enquanto sugerem uma ligação entre a sinalização lipídica e atividade do CD. Adicionalmente, as análises transcriptómicas em células do sistema imunitário de *P. lividus*, quando expostas a um agressor biológico, demonstraram a presença e envolvimento de NRs, respetivos coativadores e FMOs numa rede alargada de defesa químico-imunológica. Estes resultados sugerem que os componentes do CD interagem com diversas vias do sistema imunitário inato, podendo contribuir para a atenuação do stress oxidativo, regulação da inflamação e estabilização da homeostase quando o organismo está sujeito a uma ameaça biológica. A identificação de membros da família génica dos FMOs nestas condições forneceu ainda novos contributos para a compreensão da especialização de elementos do CD no sistema imunológico de equinóides, reforçando assim a complementaridade entre diferentes sistemas biológicos ao longo dos metazoários.

De uma perspetiva bioinformática, estabelecemos e implementámos um conjunto de ferramentas que pode ser virtualmente aplicado a qualquer espécie de metazoário. As abordagens genómicas incluíram um conjunto de ferramentas que permitiram a construção de novas referências genómicas, desde a recolha de indivíduos até à construção do seu genoma a nível cromossómico. Estas incluíram os processos de sequenciação, limpezas primárias de dados, ocultação de elementos genómicos repetitivos, montagem do próprio genoma, bem como processos de anotação funcional de elementos genómicos (elementos repetitivos, genes, transcritos, proteínas, entre outros). Adicionalmente, desenvolvemos ficheiros de integração de programas de processamento transcriptómico, para auxiliar a anotação funcional de genomas, utilizando métodos de construção *de novo* ou baseados numa referência. Ainda, implementámos diferentes programas de análise computacional, de forma a reconstruir cenários evolutivos, avaliar relevância funcional, e integrar dados de diferentes sistemas biológicos, indivíduos ou espécies.

No global, esta tese permitiu estabelecer uma ferramenta de dupla face genómico-transcriptómica para identificação, análise e avaliação dos diferentes módulos do

defensoma químico em espécies não modelo. Adicionalmente, também expande o atual conhecimento sobre a evolução do defensoma, demonstrando importantes correlações entre elementos do CD e outros sistemas biológico, enquanto enfatiza a importância da caracterização global do CD em espécies pouco investigadas. Além disso, esta tese realça a importância das novas técnicas de sequenciação na elucidação dos mecanismos moleculares subjacentes ao defensoma químico.

Palavras-chave: defensoma químico, recetores nucleares, monooxigenases dependentes de flavina, metazoários

Abstract

Over millions of years, the internal chemical defences of metazoans have evolved to adapt, resist and integrate environmental cues, and in some cases reuse them as part of their own biological systems. The architecture of this system involves a diverse and conserved set of genes and pathways that enable organism to identify, metabolise, and counteract chemically or biologically derived compounds. This set of genes and pathways is collectively referred to as the ‘chemical defensome’ (CD). This molecular system comprises a network of evolutionarily conserved specialised genes, transcription factors and enzymes that allow organisms to survive, adapt and thrive in diverse ecosystems and environments.

Despite the diversity and breadth of the CD, the onset of Anthropocene era has challenged this system at an unprecedented rate due to the rapid release of man-made chemicals and other human activities. This has naturally driven scientific interest in understanding how the CD is organised and how it interacts with other systems at different levels of biological organisation. However, despite its importance and the efforts of recent research, knowledge of CD components, pathways, and integrative mechanisms remains limited or poorly characterised in many non-model species.

This thesis addresses the aforementioned gap by exploring and identifying variations in the overall gene composition. In this context, I approached this topic by expanding genome and transcriptome resources at key nodes of Metazoa phylogeny: invertebrate protostomes (*Hediste diversicolor*) and deuterostomes (*Paracentrotus lividus*), early-branching Chondrichthyes (*Hydrolagus colliei*), a non-model teleost (*Scomber colias*). To validate this strategy, I next explored these novel generated resources to decipher the catalogue of several elements of two key CD gene families—nuclear receptor (NRs) and/or flavin-containing monooxygenases (FMOs)—in four non-model vertebrate and invertebrate species. These analyses provide new insights into the evolutionary history, functional relevance and integration of these elements with other biological systems, including immune and lipid pathways.

Genome-wide analyses expanded the CD repertoire in vertebrates by cataloguing multiple NR and FMO components in *Scomber colias* and *Hydrolagus colliei*, enabled us to identify duplications and presence-absence patterns, as well as conserved elements that reflect their evolutionary trajectories. Further transcriptomic-wide approaches in the gonadal tissue of the invertebrate species *Paracentrotus lividus* further highlighted the spatiotemporal synchronicity between NRs, FMOs, and related genes, and their putative role in coordinating lipid homeostasis, reproduction, and xenobiotic metabolism. Co-

expression patterns reinforce such synchronicity between potential heterodimerizing NRs and coactivators providing molecular targets for future analysis and suggests a link between lipid signalling and defensome activity. Furthermore, transcriptomic analyses of the immune cells transcriptomes of *P. lividus* exposed to biological stressors revealed the involvement of NRs, their co-activators, and FMOs in a broader chemical-immune defence network. These findings suggest that CD components interact with pathways of the innate immune system, possibly contributing to the mitigation of oxidative stress, the regulation inflammation, and the stabilisation of homeostasis under biological threat. The identification of FMO members in these conditions provided new insights into the specialisation of CD elements in echinoid immunity, further emphasising the complementarity of different biological systems across metazoans.

From a bioinformatics perspective, we have established and implemented a universal framework that can be applied virtually to any metazoan species. Genomic approaches encompass an array of universal pipelines that enable the construction of new genomic references from scratch, up to chromosome-scale assemblies. These include sequencing, quality control, repeat masking, genome assembly processes and the functional annotation of genomic features (repetitive elements, genes, transcripts, proteins, among others). Additionally, we developed transcriptomic pipelines to support genome annotation, establish relative and differential expression profiles, and generate novel transcripts from sequencing through to functional annotation, using *de novo* or reference-based methods. Additionally, multiple analytical pipelines were implemented to reconstruct evolutionary scenarios, evaluate functional relevance, and integrate data across biological systems, individuals and species.

Overall, this thesis establishes a two-tier genomic-transcriptomic framework for the identification, analysis and assessment of chemical defensome modules in non-model species. It also expands current knowledge of defensome evolution, demonstrating links between CD elements and other biological systems and emphasising the importance of broadening defensome characterisation to include unexplored species. Furthermore, the thesis highlights the role of next-generation sequencing in elucidating the molecular mechanisms underpinning the chemical defensome.

Keywords: chemical defensome, nuclear receptors, flavin-containing monooxygenases, metazoans

Table of Contents

Author Statement.....	XIV
List of Tables	XVI
List of Figures	XVIII
List of Abbreviations	XXII
1. Chapter 1 – Introduction	1
1.1. The Evolution of the Metazoa.....	1
1.1.1. Taxonomic and Palaeontologic early stages of Metazoa evolution.....	1
1.1.2. Phylogeny and the Metazoa.....	2
1.1.3. The early-branches of the Metazoa.....	3
1.1.4. Protostomes.....	5
1.1.5. Deuterostomes	6
1.2. Timelapse overview: From the DNA to the Omics Era.....	7
1.2.1. Early stages of sequencing	7
1.2.2. Omics and Next-generation sequencing	9
1.2.3. Databases and Bioinformatics	11
1.2.4. Bioinformatics in Genomics and Transcriptomics.....	12
1.2.5. Genomics and Transcriptomics in large consortia.....	14
1.3. The Origins of the Chemical Defensome	15
1.3.1. Origin and impact of xenobiotics in ecosystems.....	15
1.3.2. From chemical defence to the chemical defensome.....	17
1.3.3. Molecular overview of chemical defensome – The three Phases	19
1.3.4. Chemical defensome in Metazoa	27
1.4. Nuclear Receptors and Flavin-Containing Monooxygenases.....	31
1.4.1. Nuclear Receptors in the Chemical Defensome.....	31
1.4.1. Flavin-Containing Monooxygenases in the Chemical Defensome	33
1.5. Model Species.....	36

1.5.1. <i>Scomber colias</i>	36
1.5.2. <i>Hydrolagus colliei</i>	37
1.5.3. <i>Paracentrotus lividus</i>	40
1.5.4. <i>Hediste Diversicolor</i>	41
1.6 – Objectives	44
1.6.1 – Global objectives	44
1.6.2 – Specific objectives	44
1.7. Dissertation outline	46
2. Chapter 2 – A genome assembly of the Atlantic chub mackerel (<i>Scomber colias</i>): a valuable teleost fishing resource.....	48
2.1 – Abstract.....	49
2.2 – Background and context.....	49
2.3 – Methods	51
2.4 – Data Validation.....	57
2.5 – Reuse Potential	63
2.6 – Data Availability.....	64
2.7 – References.....	65
3. Chapter 3 – The Complete Genome Sequence of <i>Hydrolagus colliei</i> (Lay & Bennett, 1839), the spotted ratfish	70
3.1. Abstract.....	71
3.2. Background & Summary.....	72
3.3. Methods	74
3.4. Data Records	82
3.5. Technical Validation	82
3.6. References.....	88
4. Chapter 4 – Chemical Defensome in <i>Paracentrotus lividus</i>	97
4.1. The male and female gonad transcriptome of the edible sea urchin, <i>Paracentrotus lividus</i> : Identification of sex-related and lipid biosynthesis genes	97

4.1.1. Abstract	98
4.1.2. Introduction	98
4.1.3. Material and methods	99
4.1.4. Results and Discussion.....	101
4.1.5. Conclusion.....	107
4.1.6. References	107
4.2. Unravelling the main immune repertoire of <i>Paracentrotus lividus</i> following <i>Vibrio anguillarum</i> bath challenge	110
4.2.1. Abstract	111
4.2.2. Introduction	111
4.2.3. Material and methods	112
4.2.4. Results.....	114
4.2.5. Discussion	115
4.2.6. Conclusion.....	118
4.2.7. References	119
5. Chapter 5 – A whole-body transcriptome assembly of the annelid worm <i>Hediste diversicolor</i>	121
5.1. Abstract.....	122
5.2. Introduction	122
5.3. Data description	123
5.4. Conclusion	125
5.5. Data Availability.....	125
5.6. References.....	125
6. Chapter 6 – General Discussion and Future Perspectives	127
6.1. The Chemical defensome: From its origin through concept to current frameworks in Metazoans	127
6.2. The NGS as game changer in Chemical Defensome	130
6.3. Evolution of chemical defensome in metazoans	134
6.3.1. From Phenotype to chemical defensome adaptations in Metazoans.....	134

6.3.2. Transcription factors and biotransformation enzymes in metazoa species	135
6.4. Future perspectives.....	143
7. Conclusion.....	145
References	146
Attachments.....	197
Attachment 1: Authorization of use of PDF paper	197
Attachment 2: Authorization of use of PDF paper	198
Attachment 3: Authorization of use of PDF paper	200
Attachment 4: Authorization of use of PDF paper	202

Author Statement

In this thesis, the candidate worked closely with supervisors and other researches as co-authors to complete all or part of the work presented in this thesis. In specifics, the candidate is totally responsible for the conception and complete drafts of Chapter 1, 6 and 7 with supervisory team making any necessary adjustments. In remain Chapters, the candidate made substantial contributions.

Briefly, this thesis encompasses four publications that have been published in international peer-reviewed journals and one article in final preparation for submission. All articles were integrated within the thesis as chapters (first author or join first authors are underlined).

Chapter 2

Machado, A. M., Gomes-dos-Santos, A., Fonseca, M. M., Fonseca, R. R. da, Veríssimo, A., Felício, M., Capela, R., Alves, N., Santos, M., Salvador-Caramelo, F., Domingues, M., Ruivo, R., Froufe, E., Castro, L. F. C., Gomes-dos-Santos, A., Fonseca, M. M., Fonseca, R. R. da, Veríssimo, A., Felício, M., Castro, L. F. C. (2022b). A genome assembly of the Atlantic chub mackerel (*Scomber colias*): a valuable teleost fishing resource. Gigabyte, 2022, 1–21.

Chapter 3

Machado, A. M., Gomes-dos-Santos, A., Almeida, T., Neves, F., Moutinho, J., Froufe, E., Wilson, J., Ruivo, R., Ohta, Y., Flajnik, M., Kawaguchi, A., Kuraku, S., Veríssimo, A., Castro, L. F. C. The Complete Genome Sequence of *Hydrolagus coliei* (Lay & Bennett, 1839), the spotted ratfish. IN PREPARATION

Chapter 4.1

Machado, A. M., Fernández-Boo, S., Nande, M., Pinto, R., Costas, B., & Castro, L. F. C. (2022). The male and female gonad transcriptome of the edible sea urchin, *Paracentrotus lividus*: Identification of sex-related and lipid biosynthesis genes. Aquaculture Reports, 22, 100936.

Chapter 4.2

Fernández-Boo, S., Machado, A., Castro, L. F. C., Azeredo, R., & Costas, B. (2024). Unravelling the main immune repertoire of *Paracentrotus lividus* following *Vibrio anguillarum* bath challenge. *Fish & Shellfish Immunology*, 147, 109431.

Chapter 5

Machado, A. M., Samico, R., Domingues, M., Hagemann, A., Valente, L. M. P., Malzahn, A. M., Gomes-dos-Santos, A., Ruivo, R., Navarro, J. C., Monroig, Ó., & Castro, L. F. C. (2024). A whole-body transcriptome assembly of the annelid worm *Hediste diversicolor*. *Marine Genomics*, 74, 101084.

List of Tables

Chapter 2 – A genome assembly of the Atlantic chub mackerel (*Scomber colias*): a valuable teleost fishing resource.

Table 2.1 – General statistics of read datasets used to perform the <i>Scomber colias</i> genome assembly.....	58
Table 2.2 – Statistics of the <i>Scomber colias</i> genome assembly.	60
Table 2.3 – Report of RepeatMasker software. This report contains statistics of repetitive elements in the <i>Scomber colias</i> genome assembly.	61
Table 2.4 – Structural annotation report of the <i>Scomber colias</i> genome assembly.	61
Table 2.5 – Functional annotation report of <i>S. colias</i> genome assembly.	63

Chapter 3 – The Complete Genome Sequence of *Hydrolagus colliei* (Lay & Bennett, 1839), the spotted ratfish.

Table 3.1 – MixS descriptors and accession number of tissue sample and raw data used in the initial assembly of <i>H. colliei</i> genome.....	75
Table 3.2 – General characteristics and applications of the sequencing datasets.	76
Table 3.3 – Statistics of <i>H. colliei</i> genome assembly and annotation.	83
Table 3.4 – Percentage of the contig genome assembly anchored to 35 chromosomes.	84
Table 3.5 – Flavin-containing monooxygenases in Chondrichthyes.	87

Chapter 4 – Chemical defensome in *Paracentrotus lividus*

Subchapter 4.1 – The male and female gonad transcriptome of the edible sea urchin, *Paracentrotus lividus*: Identification of sex-related and lipid biosynthesis genes

Table 4.1.1 – General statistics of transcriptome assembly and annotation process. a) Stats of raw data quality control. b) Stats of transcriptome assembly and annotation.	102
--	-----

Subchapter 4.2 – Unravelling the main immune repertoire of *Paracentrotus lividus* following *Vibrio anguillarum* bath challenge

Table 4.2.1 – Total and differential cell counts in control and exposed groups. Data is presented in 106 cells/mL in total cell concentration while differential counting is presented as percentage of total population. Phagocytic activity is presented as % of cells with engulfed <i>Candida albicans</i> cells. All data is presented as mean ± _standard deviation.	114
Table 4.2.2 – Humoral innate immune parameters in perivisceral fluid and cell lysate samples of control and exposed groups. Data is presented as mean ± _standard deviation. *denotes significant differences between groups (p < 0.05).	115
Table 4.2.3 – Differential expressed proteins in males and females of <i>P. lividus</i> after gene ontology searching in Panther DB.	115
Table 4.2.4 – List of genes presented in Biological process and Molecular function categories according to Panther DB analysis.	117
Table 4.2.5 – Enriched KEGG pathways on challenged group according to gProfiler analysis.....	117

Chapter 5 – A whole-body transcriptome assembly of the annelid worm *Hediste diversicolor*

Table 5.1 – MixS descriptors of the <i>Hediste diversicolor</i> sample.....	123
Table 5.2 – Statistics of the <i>Hediste diversicolor</i> transcriptome assembly versions, raw (V0), decontaminated (V1), final (V2).....	124

List of Figures

Chapter 1 – Introduction

Figure 1.1 – Phylogenetic tree of Metazoa with Choanoflagellata as outgroup. Division of metazoan into major clades.	4
Figure 1.2 – Schematic overview of omics subfields of molecular biology.	10
Figure 1.3 – Graphic showing the number of genomes, species with genome public available and genomes with annotation in NCBI.	15
Figure 1.4 – Chemical defensome overview.	20
Figure 1.5 – Illustration showing how drugs, chemicals and xenobiotics activate receptor-mediated genes and form heterodimers, and their involvement in each phase of drug metabolism.	26
Figure 1.6 – Photograph of Atlantic chub mackerel, <i>Scomber colias</i>	36
Figure 1.7 – Photograph of spotted ratfish, <i>Hidrolagus colliei</i>	38
Figure 1.8 – Photograph of common purple sea urchin, <i>Paracentrotus lividus</i>	40
Figure 1.9 – Photograph of common ragworm, <i>Hediste diversicolor</i>	42

Chapter 2 – A genome assembly of the Atlantic chub mackerel (*Scomber colias*): a valuable teleost fishing resource.

Figure 2.1 – Photograph of Atlantic chub mackerel, <i>Scomber colias</i> . The specimen was caught in 2020 and used for Pacbio HiFi genome assembly.	50
Figure 2.2 – Bioinformatics workflow used to perform the genome assembly of <i>Scomber colias</i> species.	53
Figure 2.3 – Genomescope2 plot with k-mer spectra content and fitted models of the <i>Scomber colias</i> Illumina PE dataset.	57
Figure 2.4 – Circular mitochondrial genome assembly of <i>Scomber colias</i> , obtained from the Illumina PE dataset (equal to that obtained from the PacBio HiFi long reads assembly). From the centre to the outmost features: GC content distribution; sequencing depth distribution of aligned Paired-End reads; gene elements (i.e., PCGs, rRNA genes, tRNA genes)	58

Figure 2.5 – Validation of the genome assembly and annotation process. (a) K-mer analyses of the *Scomber colias* genome assembly (Merqury). (b) Maximum Likelihood phylogenetic tree based on the concatenated alignments of amino acid sequences of 392 single-copy orthologs retrieved by OrthoFinder. Bootstrap values are shown next to the nodes. (c) BUSCO scores were obtained from searching the proteomes of the 3 Scombriformes species with genome annotation available, against the actinopterygii_odb10 (n:3640) lineage. 61

Figure 2.6 – Pairwise sequentially Markovian coalescent (PSMC) estimates from the *Scomber colias* genome assembly. Estimations were obtained using a generation time of 2 years and genome-wide mutation rate of 7.3×10^{-9} mutations/site/generation. Effective population size (N_e) is presented in the left vertical axis, and changes estimated up to the present, over the last 3 myr, on the horizontal axis. 64

Chapter 3 – The Complete Genome Sequence of *Hydrolagus coliei* (Lay & Bennett, 1839), the spotted ratfish.

Figure 3.1 – A) Photograph of spotted ratfish, *Hidrolagus colliei* (Provided by Martin Flajnik). B) Potential distribution of *H. colliei* in North America (Retrieved from FishBase Database - https://www.fishbase.se/summary/Hydrolagus_colliei.html). 74

Figure 3.2 – Bioinformatic workflow used in the genome assembly and annotation of *H. colliei*. 79

Figure 3.3 – Quality-control analyses of the *H. colliei* genome. A) Stacked histogram showing the read k-mer frequency spectrum of the *H. collie* genome assembly at the contig level. B) BlobTools read coverage plot showing the taxonomic affiliation of each contig to a taxonomic family within the RefSeq database. C) Hi-C interactions among 35 chromosomes. The intensity of the interactions increases from white to dark red. Each central square represents a pair of *H. colliei* chromosomes. 85

Chapter 4 – Chemical defensome in *Paracentrotus lividus*

Subchapter 4.1 – The male and female gonad transcriptome of the edible sea urchin, *Paracentrotus lividus*: Identification of sex-related and lipid biosynthesis genes

Figure 4.1.1 – Sampling and bioinformatic workflow of the study. (a) Sampling and experimental setup used to perform RNA extraction and sequencing. (b) Bioinformatics

workflow used to process the raw datasets, perform the differential gene expression and obtain biological insights. 100

Figure 4.1.2 – Histological sections of gonads used for transcriptome sequencing. Females (a–c) were in stage III (a, b) and IV (c) of maturation. All males (d–f) were in stage III of maturation. Scale bar (200 μ m). NP: nutritive phagocyte; O: oocyte; S: spermatid. 101

Figure 4.1.3 – Functional annotation and clean-up analyses of the V0 and V1 versions of transcriptome assembly of *P. lividus*. (a) Venn diagram of the transcripts (V0 transcriptome) codifying to an open reading frame or matching with NCBI, UNIPROT, and PFAM databases. (b) Length distribution of the transcripts in versions V0 and V1 of the transcriptome assembly. (c) Blast-x analysis of the V0 and V1 versions of the *P. lividus* transcriptome assembly. 103

Figure 4.1.4 – Heatmap of up and down-regulated genes. (a) Sexual development and determination genes. (b) Genes related to the biosynthesis and storage of lipids. 105

Sub-Chapter 4.2 – Unravelling the main immune repertoire of *Paracentrotus lividus* following *Vibrio anguillarum* bath challenge

Figure 4.2.1 – Representation of the trial and sampling from Experiment 2. 113

Figure 4.2.2 – Data of the reference transcriptome from males and females of the sea urchin *Paracentrotus lividus*. A. Principal component analysis (PCA) of the samples. B. Heatmap from the DEGs obtained after analysis of reads by Degust. C. Percentage of variance by MDS dimension of the PCA. D. Volcano plot of the DEGs from males and females. 116

Figure 4.2.3 – Transcriptome data from control and challenged *P. lividus*, sampled after the infection trial. A. Principal component analysis (PCA) of the samples. B. Heatmap from the DEGs (>2 FC and p-value <0.05) obtained after analysis of reads by Degust. C. Percentage of variance by MDS dimension of the PCA. D. Volcano plot of the DEGs (>2 FC and p-value <0.05) from control (C) and challenged (I) groups. 116

Figure 4.2.4 – Protein processing in endoplasmic reticulum pathway as showed in KEGG database. Red stars indicate modulated genes after 4 h challenge with *V. anguillarum*. 118

Chapter 5 – A whole-body transcriptome assembly of the annelid worm *Hediste diversicolor*

Figure 5.1 – a) Photograph of common ragworm, *Hediste diversicolor*. b) Bioinformatic workflow used in the transcriptomic approach. The workflow was created with BioRender.com. 123

List of Abbreviations

AHE	ANCHORED HYBRID ENRICHMENT
AA	AMINOACID
ABC	ATP-BINDING CASSETE
ADH	ALDEHYDE DEHYDROGENASES
AGAT	ANOTHER GTF/GFF ANALYSIS TOOLKIT
AhR	ARYL HYDROCARBON RECEPTOR
ARE	ANTIOXIDANT RESPONSE ELEMENTS
ARNT	ARYL HYDROCARBON RECEPTOR NUCLEAR TRANSLOCATOR
ATP	ADENOSINE TRIPHOSPHATE
BLAST	BASIC LOCAL ALIGNMENT SEARCH TOOL
BUSCO	BENCHMARKING UNIVERSAL SINGLE-COPY ORTHOLOGS
BV	BAEYER-VILLIGER
BVMOs	BAYER-VILLIGER MONOOXYGENASES
BWA	BURROWS-WHEELER ALIGNER
CAR	CONSTITUTIVE ANDROSTANE RECEPTOR
CCS	CIRCULAR CONSENSUS SEQUENCES
CD	CHEMICAL DEFENSOME
cDNA	COMPLEMENTARY DNA
CDS	CODING REGION
cirRNA	CIRCULAR RNA
CLR	CONTINUOUS LONG READ
CNC-bzip	CAP 'N' COLLAR -BASIC LEUCINE ZIPPER
CPM	COUNTS PER MILLION
CYP	CYTOCHROME P450
DAO	DIAMINE OXIDASE
DBD	DNA-BINDING DOMAIN
DDBJ	DNA DATA BANK OF JAPAN
DEG	DIFFERENTIAL EXPRESSED GENES
DGE	DIFFERENTIAL GENE EXPRESSION
DME	DRUG-METABOLIZING ENZYMES
DNA	DEOXYRIBONUCLEIC ACID
DToL	DARWIN TER OF LIFE
EDCs	ENDOCRINE DISRUPTORS
EH	EPOXIDE HYDROLASE
EMBL	EUROPEAN MOLECULAR BIOLOGY LABORATORY

EPRE	ELECTROPHILE RESPONSE ELEMENTS
ER	ESTROGEN RECEPTOR
ESTs	EXPRESSED SEQUENCE TAGS
FAD	FLAVIN ADENINE DINONUCLEIC
FC	FOLD-CHANGE
FCUP	FACULTY OF SCIENCES OF THE UNIVERSITY OF PORTO
FMOs	FLAVIN-CONTAINING MONOOXYGENASES
FXR	FARNESOID-X-RECEPTOR
gDNA	GENOMIC DNA
GE	GENE EXPRESSION
GO	GENE ONTOLOGY
GR	GLUCOCORTICOID RECEPTOR
GST	GLUTATHIONE S-TRANSFERASES
HGP	HUMAN GENOME PROJECT
Hi-C	CHROMATIN CONFORMATION CAPTURE (3C) METHODS
HiFi	HIGH FIDELITY
HMMER	HIDDEN MARKOV MODEL-BASED SEQUENCE ANALYSIS
HMW	HIGH MOLECULAR WEIGHT
HRE	HORMONE RESPONSE ELEMENTS
HTS	HIGH-THROUGHPUT SEQUENCING
IMTA	INTEGRATED MULTITROPHIC AQUACULTURE
IPMA	INSTITUTO DO MAR E DA ATMOSFERA
KAT	THE K-MER ANALYSIS TOOLKIT
KEGG	KYOTO ENCYCLOPEDIA OF GENES AND GENOMES
LBD	C-TERMINAL LIGAND-BINDING DOMAIN
LINES	LONG INTERSPERSED NUCLEAR ELEMENTS
lncRNA	LONG NON-CODING RNA
LTR	LONG TERMINAL REPEATS
LXR	LIVER-X-RECEPTOR
MAO	MONOAMINE OXIDASES
MAREs	MAF RECOGNITION ELEMENTS
MATES	MULTIDRUG AND TOXIN EXTRUSION
MBP	MEGA BASE PAIRS
MHC	MAJOR HISTOCOMPATIBILITY COMPLEX
MMseqs2	MANY-AGAINST-MANY SEQUENCE SEARCHING
mRNA	MESSANGER RNA
MS2	MALE-SPECIFIC BACTERIOPHAGE, STRAIN 2

MT	METHYLTRANSFERASES
mtDNA	MITOCHONDRIAL DNA
MyA	MILLION YEARS AGO
NADP	NICOTINAMIDE ADENINE DINUCLEOTIDE PHOSPHATE
NATs	N-ACETYLTRANSFERASES
NCBI	NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION
ncRNA	NON-CODING RNA
NGS	NEXT GENERATION SEQUENCING
NHI	NATIONAL INSTITUTE OF HEALTH
NMOS	N-HYDROXYLATING MONOOXYGENASES
NO	NITRIC OXIDE
NRF2	NUCLEAR FACTOR ERYTHROID 2-RELATED FACTOR 2
NRs	NUCLEAR RECEPTORS
NT	NUCLEOTIDE
OAT	ORGANIC ANION TRANSPORTERS
OCT	ORGANIC ANION TRANSPORTERS
ONT	OXFORD NANOPORE TECHNOLOGIES
ORF	OPEN READING FRAME
PacBio	PACIFIC BIOSCIENCES
PAHs	POLYCYCLIC AROMATIC HYDROCARBONS
PAS	PER-ARNT-SIM
PCR	POLYMERASE CHAIN REACTION
PE	PAIRED-END
Phrap	PHIL'S REVISED ASSEMBLY PROGRAM
Phred	PHIL'S READ EDITOR
PPAR	PEROXISOME PROLIFERATOR-ACTIVATED RECEPTORS
PSMC	PAIRWISE SEQUENTIALLY MARKOVIAN COALESCENT
PXR	PREGNANE-X-RECEPTOR
QC	QUALITY-CONTROL
QUAST	QUALITY ASSESSMENT TOOL FOR GENOME ASSEMBLIES
QV	MEAN QUALITY VALUE
RAD-SEQ	RESTRICTION SITE ASSOCIATED DNA SEQUENCING
Rad-seq	RESTRICTION SITE ASSOCIATED DNA SEQUENCING
RAR	RETINOIC ACID RECEPTOR
RAS	RECIRCULATION AQUACULTURE SYSTEM
RBM	READ-BACK MAPPING
RefSeq	NCBI REFERENCE SEQUENCE DATABASE

RNA	RIBONUCLEIC ACID
RNA	RIBONUCLEIC ACID
RNA-Seq	RNA SEQUENCING
ROS	REACTIVE OXYGEN SPECIES
rRNA	RIBOSOMAL RNA GENES
rRNA	RIBOSOMAL RIBONUCLEIC ACID
RXR	RETINOID-X-RECEPTOR
SBL	SEQUENCING BY LIGATION
SBS	SEQUENCING BY SYNTHESIS
SE	SINGLE-END
SINES	SHORT INTERSPERSED NUCLEAR ELEMENTS
siRNA	SMALL INTERFERING RNA
SLC	SOLUTE CARRIERS
SLR	SYNTHETIC LONG-READS
SMLS	SINGLE-MOLECULE LONG-READ SEQUENCING
SMRT	SINGLE MOLECULE REAL-TIME
SMS	SINGLE-MOLECULE SEQUENCING
SNPs	SINGLE NUCLEOTIDE POLYMORPHISMS
snRNA	SMALL NUCLEAR RNA
SULTs	SULFOTRANSFERASES
TE	TRANSPOSABLE ELEMENTS
TLR	TOLL-LIKE RECEPTORS
TM	TRANSMEMBRANE DOMAINS
TMA	TRIMETHYLAMINE
ToL	TREE OF LIFE
TPM	TRANSCRIPTS PER MILLION
TPMT	THIOPURINE S-MTHYLTRASFERASE
tRNA	TRANSFER RNA
UGTs	UPD GLUCURONOSYLTRANSFERASES
UP	UNIVERSITY OF PORTO
USA	UNITED STATES OF AMERICA
VDR	VITAMIN D RECEPTOR
WGD	WHOLE GENOME DUPLICATION
WGS	WHOLE GENOME SEQUENCING
XREs	XENOBIOTIC RESPONSE ELEMENTS

1. Chapter 1 – Introduction

1.1. The Evolution of the Metazoa

Humans have sought to understand the origin and formation of Earth, from the mineral, chemical and physical makeup to the foundation and evolution of living organisms. One of the primary fields involved in this process is evolutionary biology, where scientists use multidisciplinary approaches, such as phylogeny, palaeontology, taxonomy, molecular evolution, comparative genomics among others, to scrutinize the intricates of life and its evolution.

1.1.1. Taxonomic and Palaeontologic early stages of Metazoa evolution

Taxonomy and phylogeny are pivotal and complementary subfields within evolutionary biology. Taxonomy describes, classifies and assigns species names based on shared characteristics, including morphological, genetic and ecological, while phylogeny reveals the evolutionary road, history and relationships between organisms (Reif, 2002; Thompson & Walsh, 2014). Functionally, taxonomy provides a systematic framework in which hierarchical categories are organized in domains, kingdoms, phylum, class, order, family, genus and species. Currently, this system is composed of three domains, Bacteria, Archaea, Eukarya, encompassing all living beings (Cavalier-Smith, 1981, 1998; Ruggiero et al., 2015).

Chronologically, the earliest evidence of life dates from the Precambrian era (4.6 billion - 541 million years ago (Mya)), specifically in Early Proterozoic (2.5 billion - 1 billion years ago), marked by the emergence of prokaryotic organisms of the Archaea and Bacteria domains (Knoll et al., 2016). The biota changes during this period were significant and crucial to the emergence of the first multicellular organisms in Late Proterozoic Era (1 billion - 541 Mya). The Late Proterozoic Era, encompasses two periods: Cryogenian Period (720 - 635 Mya), marked by two significant global glaciations (known as the Sturtian and Marinoan) that dramatically affected the biogeochemical cycles and ecosystems; and, the Ediacaran Period (635 - 541 Mya), characterized by a significant increase of biological diversity, including both eukaryotic-rich and eukaryote-lean microbial communities, as well as the geochemical impact of the these communities in the increase of atmospheric oxygen levels (Dahl et al., 2010; Hoffman et al., 1998, 2017; Knoll et al., 2006; Pehr et al., 2018). Some of the most relevant fossils of the Ediacaran period were discovered in the Doushantuo Formation in China (630 - 564 Ma)

and Deep Spring Formation at Mount Dunfee, Nevada (USA). These records pointed to the appearance of early-multicellular specimens, including algae, acritarchs, and animal embryos and adults with soft-bodied and various body plans (Budd, 2008; Erwin, 2015; Smith et al., 2016; Xiao et al., 2014).

Together, the emergence of multicellularity and the drastic climatic changes of the Late Proterozoic Era paved the way to the Cambrian explosion (541 - 485 Mya). The Cambrian period was notable by the rapid diversification and adaptation of organisms belonging to the Animalia kingdom to various ecological niches alongside the development of feeding, locomotion and reproductive strategies (Marshall, 2006; Na & Kiessling, 2015). During this time, multiple morphological innovations arose, including different body plans (Na & Kiessling, 2015; Vannier et al., 2012). Regarding the feeding system, the first rudimentary trophic chains appeared, laying the foundation for the complex food web observed today (Dunne et al. 2008). The fossils from this period—some of them found in the Burgess Shale formation (508 Mya) in Canadian Rockies of British Columbia—reported several exceptionally well-preserved organisms (Gaines et al., 2012). These included soft-bodies specimens as well as the anatomical innovations mentioned above, primarily associated with the marine environment (Mángano & Buatois, 2014; Nanglu et al., 2020).

1.1.2. Phylogeny and the Metazoa

Complementary to palaeontology, the phylogeny includes time and assesses the relationships between organisms, by integrating molecular data. Notwithstanding, the palaeontology, in specific the fossil reports, plays a crucial role in the phylogenetic analyses once they are often used as reference to calibrate the molecular clocks and help determine the timing of divergences among major lineages (Parfrey et al., 2011). Before scrutinizing the relationship among the Animalia organisms, it is essential to define some phylogenetic concepts such as monophyletic, paraphyletic and polyphyletic groups. Briefly, in the monophyletic group all members share a common ancestor. On the other hand, in paraphyletic group includes a common ancestor and some of its descendants, but not all. Lastly, the polyphyletic group can contain several organisms without sharing a common ancestral. In this case other characteristics or similarities are used to group organisms (example: flying) (Ashlock, 1971). These concepts as well as some phylogenetic terms are often used in phylogeny to define the relationships among organisms and/or genes (Koonin, 2005; Maddison et al., 1984), such as: clade (single branch on the tree of life with descendants and common ancestral), sister clade (two lineages that emerge from the same node on a phylogenetic tree), outgroup (organism

or group of organisms that are closely related to, but not part of the group in study), orthologues (homologous genes by speciation), and paralogues (homologous genes by gene duplication).

1.1.3. The early-branches of the Metazoa

The early-metazoan organism emerged between the Cryogenian and the Ediacaran and almost all lineages and phyla of Metazoa emerged during these periods (Dos Reis et al., 2015; Erwin et al., 2011; Hedges et al., 2004). Phylogenetically, the Metazoa are a monophyletic group, and a sister clade of the choanoflagellates (King et al., 2008). The choanoflagellates have been used to study the transition of unicellular to multicellular in Metazoa species. Morphologically, these organisms (e.g. *Monosiga brevicollis*) are unicellular and contain a single flagellum, used to create a water current, surrounded by a collar of microvilli used for the filtering and feeding of the organism (Brunet & King, 2017; Holland, 2011; King et al., 2008). These characteristics are similar to the ones found in early-branching species of the Metazoa (e.g. the flagellum structure of *Monosiga brevicollis* seems to be functionally similar to the choanocytes, specialized filter feeding cells of sponges).

The Metazoa are subdivided into two major groups: the non-bilaterian and bilaterian. The non-bilaterian typically exhibit a radial symmetry or asymmetry, decentralized nervous system and are diploblastic (development of two germ layers during the embryonic development, ectoderm and endoderm) (Martindale, 2005). The non-bilaterians include four phyla, Porifera (sponges), Cnidaria (jellyfish, corals, and sea anemones), Ctenophora (comb jellies), and Placozoa (Figure 1.1). The organisms of these phyla can reproduce sexually through gametes or asexually, via budding or binary fission, and display diverse morphologies and body constitution (e.g. Porifera (porous body structure); Cnidaria (polyp and medusa); Ctenophora (gelatinous bodies and rows of cilia used for locomotion); Placozoa (simple, flat, and multicellular body plan) (Borchiellini et al., 2001; Brusca et al., 2022; Kayal et al., 2013; Schierwater et al., 2009; van Soest et al., 2012). In contrast, Bilateria are usually characterized by the bilateral symmetry (three axes, front (anterior) to back (posterior), top (dorsal) to bottom (ventral), and left to right, in which just the left and the right plans are symmetric), centralized nervous system, and the development of three germ layers during the embryonic development (triploblastic; ectoderm, mesoderm, and endoderm) (Hejnol & Martindale, 2008; L. Z. Holland et al., 2013; Minelli, 2007).

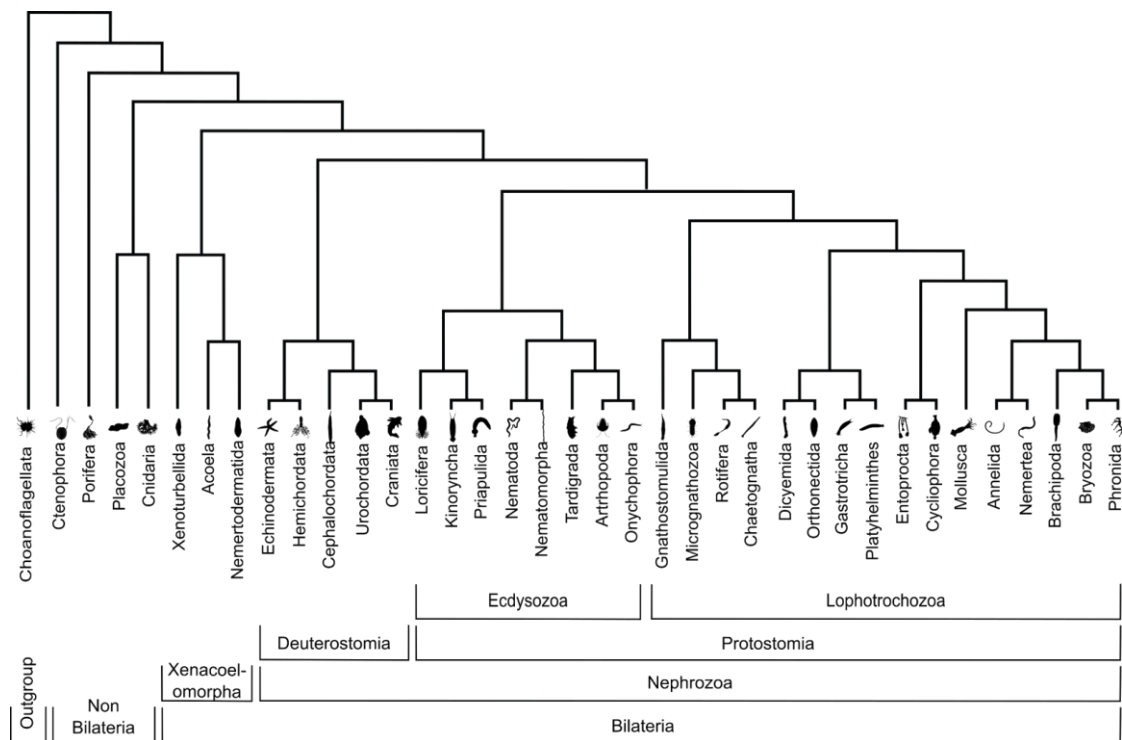


Figure 1.1 – Phylogenetic tree of Metazoa with Choanoflagellata as outgroup. Division of metazoan into major clades (adapted from (Fernández & Gabaldón, 2020)).

Bilateria group contains more than the 33 phyla encompassing 1.5 million species (Laumer et al., 2019; Ruggiero et al., 2015) (Figure 1.1). Most bilaterian organisms can be classified based on the presence of coelom, pseudocoelom or partial coelom (acoelomates). Coelom is defined as a fluid-filled body cavity that is completely enclosed by mesoderm in coelomates (true coelom). In pseudocoelomates, the body cavity is only partially surrounded by mesoderm, while acoelomates lack a body cavity altogether (Adoutte et al., 1999, 2000; Giribet et al., 2000; Wallace et al., 1996). The emergence of coelom was a major advancement in the evolution of Bilateria body plan, allowing the development of organ systems, including circulatory, respiratory, and excretory systems. The presence of excretory systems and the type of coelom led to the sub-classification of Bilaterians in two major clades, Xenacoelomorpha and Nephrozoa (Hejnol & Pang, 2016) (Figure 1.1). The organisms of Xenacoelomorpha clade lack a true coelom and specialized excretory system, showed a bilateral symmetry and a flattened morphology (Cannon et al., 2016). Phylogenetically this clade is considered a sister group of Nephrozoa and it is used as an early branch of bilaterian (Cannon et al., 2016). On the other hand, Nephrozoa are characterized by the presence of true coelom and sophisticated excretory systems and nerve cords. The coelomic cavity in Nephrozoa

plays an important role, once allowing the development of internal organs and facilitating the movement of these organs in relation to the body wall (Hejnol et al., 2009).

In Nephrozoa, the coelom formation is also used to subdivide the Nephrozoa into two other groups of animals, the Deuterostomes and Protostomes (Figure 1.1). While in the protostomes the coelomic cavities are usually developed from mesoderm embryonic tissue (Schizocoely), in the Deuterostomes typically the coelomic sacs became from an endodermal layer during embryo development and after formation of the primitive gut—the archenteron (Enterocoely) (Adoutte et al., 1999; Nielsen et al., 1996; Ruiz-Trillo, Baguña, et al., 2002; Ruiz-Trillo, Paps, et al., 2002). In addition to coelom formation Protostomia and Deuterostomia species diverge in other key features such as: (1) the blastopore fate, that is the order of mouth and anus formation during the embryonic development — in protostomes the mouth is formed primarily from blastopore and anus is developed secondary while in deuterostomes the opposite occurs, with the anus arising from blastopore and mouth developing posteriorly (Nielsen, 2019); (2) the cleavage pattern, that in protostome embryonic cells tends to be spiral and determined in the early stages of development, while in deuterostomes, embryonic cells typically exhibit radial configuration and indeterminate cleavage during the early development, allowing for the potential development of several types of cells (Martín-Durán et al., 2012; Nielsen, 2017) ; and, (3) the global developmental pattern, with both protostomes and deuterostomes able to pursue indirect and direct development, however in deuterostomes the larval and adult stage often share more similarities than those in protostomes (Cameron et al., 2000; Martín-Durán et al., 2016; Nielsen, 2017).

1.1.4. Protostomes

In protostomes, the developmental patterns and peculiar morphologies are some of the most important features to sub-classified the members of this clade into Ecdysozoa or Lophotrochozoa (Figure 1.1). Most members of the Lophotrochozoa superphylum, display spiralian development and formation of a ciliated trochophore larva (e.g. Mollusca, Annelida, and Platyhelminthes, Nemertea, Brachiopoda) (Hejnol, 2010; Nielsen, 2005, 2012; Paps et al., 2015). In terms of body structure, Lophotrochozoa shows a wide variety of body plans, being the most common forms: segmented (e.g. annelids) and unsegmented (e.g. molluscs) (Seaver, 2003; Steinmetz et al., 2011). Throughout Lophotrochozoa, Annelida and Mollusca phyla comprise some of the most iconic and well-studied species. The Annelida phylum encompasses marine, freshwater and terrestrial species, various forms (polychaetes, oligochaetes, and leeches) of segmented bodies and are essential to the trophic Interactions and nutrient cycling

(Thorp et al., 2016). Notwithstanding, molluscs present soft bodies, subdivided in three main parts (head-foot; visceral mass and mantle), protected by a shell (the shell can be internal or external) and commonly are associated with marine and freshwater environments (Ponder et al., 2019). On the other hand, Ecdysozoa are characterized by moulting their exoskeletons in a process called ecdysis. This superphylum includes some of the most successful phyla in terms of diversity and adaptability to various environments. Among these phyla, we can find: Arthropoda, which encompasses insects, arachnids, crustaceans, and myriapods. Members of this phylum display a segmented form, with head, thorax, and abdomen, revested by an chitinous exoskeleton and can be found in several ecological niches (Budd & Telford, 2009); Nematoda phylum comprises species with cylindrical shape, unsegmented form, tough cuticle and usually associated to the parasitism of several species (Blaxter, 2009) ; Tardigrada species with an exceptional resilience and ability to survive harsh environments through a process called Cryptobiosis (Møbjerg et al., 2011); or Priapulida a basal lineage of Ecdysozoa, with cylindrical body divided in trunk and an retractable introvert (Rota-Stabelli et al., 2010; Webster et al., 2006).

1.1.5. Deuterostomes

The deuterostomes are primarily subdivided in two groups, encompassing three phyla. While the Chordata phylum *per se*, is a single group, the Ambulacraria group comprises other two phyla (Hemichordata and Echinodermata) (Figure 1.1)

The species of Ambulacraria group are characterized by the presence of pharyngeal gill slits during some stage of development, a decentralized nervous system and a global indirect development, involving a free-swimming larval stage (Holland & Anderson, 2015; Lin et al., 2024). Despite the similarities across the group, each phyla possesses specific features. Hemichordata species (mainly composed of worms and pterobranchs) usually display a tripartite morphology contemplating proboscis (for feeding), collar (nervous system), and trunk (digestive system), a primitive body plan and nerve net, with a notochord-like structure and some centralized features (Gerhart, 2006; Lowe et al., 2003; Simakov et al., 2015) The Echinodermata members (starfish, sea urchins, and sea cucumbers) exhibit a pentameral symmetry with the body organized around a central axis, a unique water vascular system (acting like a network of fluid-filled canals that assist locomotion, feeding, and respiration), and a nervous system characterized by a nerve ring and radial nerves extending into the arms (Burke et al., 2014; Holland & Anderson, 2015)

The Chordata phylum comprises a broad array of species and characteristics. This phylum includes invertebrate and vertebrate species and is subdivided in three sub-phyla: Urochordata (or Tunicata), Cephalochordata, and Vertebrata (Figure 1.1) (Cameron et al., 2000). Along Chordata evolution, each sub-phyla displays the heredity traits and innovations from the common ancestor. For instance, the notochord, one of the main innovations in Chordata, has been evolving in different ways. In Urochordata the notochord is only present in the larval stages (the adults usually are sessile), whereas Cephalochordata species retained notochord throughout life, providing structural support. In vertebrates, the notochord appears during the early stages of embryo development but is later replaced by vertebral column (backbone) during the later stages of embryogenesis (Satoh et al., 2012). In terms of body structure, adult urochordates have a sac-like shape with siphons and are covered by a layer made of cellulose-like substance. Cephalochordates on the other hand, possess a fish-like body with distinct head section and segmented muscles. Vertebrates display a distinct anatomy, usually subdivided in three main components, head, trunk and tail. Notably, vertebrates also feature a skull encasing the brain and highly developed and compartmentalized organ systems. From a phylogenetic perspective, the urochordates (usually represented by *Ciona intestinalis* in molecular studies), vertebrates (usually is represented by human or mouse) and Cephalochordata (usually represented by lancelets) are considered a monophyletic group and most probably shared a common ancestor (Cameron et al., 2000; Laumer et al., 2019; Satoh et al., 2014). Although the phylogenetic relationships among the chordate subphyla were historically debated, most recent phylogenomic studies support a closer relationship between vertebrates and urochordates, rather than cephalochordates (Cameron et al., 2000; Laumer et al., 2019; Satoh et al., 2014).

1.2. Timelapse overview: From the DNA to the Omics Era

In the last 70 years several discoveries and innovations have revolutionized biological research. From the discovery of double helix of DNA and evolution of sequencing technologies till the emergence of new sub-fields in molecular and computational biology, all contributions have enabled the comprehensive understanding of complex biological systems.

1.2.1. Early stages of sequencing

In early 1950s, Watson, Crick, Franklin and Wilkins unfolded the three-dimensional structure of deoxyribonucleic acid (DNA) (Watson & Crick, 1953; Zallen, 2003). This breakthrough in understanding of genetic information storage and transmission led to

development of new approaches to visualise and translate this information in a human-readable format. These challenges were addressed using sequencing technologies and “read” methods. Initially these methods were difficult to apply, as they were slow and labour-intensive. They were also hindered by the inherent properties of DNA (e.g. the large chain lengths) as well as the lack of knowledge and information about the enzyme restrictions properties (Giani et al., 2020; Heather & Chain, 2016; Hutchison, 2007). Almost in parallel, and spurred by Watson and Crick’s discovery, Robert Holley and co-workers sequenced in 1965 the first molecule of nucleic acid, a 76-base pair of transfer ribonucleic acid (tRNA) (from *Saccharomyces cerevisiae* (Holley et al., 1965)). Three years after, in 1968, Wu and colleagues purified and sequenced the first DNA molecule (Wu & Kaiser, 1968; R. Wu & Taylor, 1971). Following these milestones, the scientific community focused on developing high throughput sequencing methods with greater accuracy and low error rate. In the early stages of the transcriptomic Era several authors explored the properties of RNA, mainly at ribosomal (rRNA) and tRNA level. Thus, in 1972 and 1976, respectively, the first complete protein-coding gene (coat protein of male-specific bacteriophage, strain two (MS2)) and the first complete RNA genome (RNA bacteriophage MS2) were sequenced (Fiers et al., 1976; Jou et al., 1972). In 1975, Sanger and Coulson proposed the Sanger sequencing approach, one of most predominant and used technology to sequence DNA (Sanger & Coulson, 1975). This technology paved the way to the genomic era with the sequencing of the first complete DNA genome, the bacteriophage ϕ X174 (Sanger et al., 1977). Despite its low error rate, and the ability to produce relatively short DNA fragments, this technique was limited in terms of throughput and scalability (Bruijns et al., 2018; Lu et al., 2016). The 1980s-2000s were marked by technological advances in the Genomic and transcriptomic fields, with the creation of complementary DNA (cDNA) libraries, expressed sequence tags (ESTs) and microarray technologies. These repeated innovations allowed the exploration of the entire transcriptome of model organisms and gene expression quantification of thousands of known genes simultaneously (Lowe et al., 2017). On the other hand, during this period in genomics one of the most important landmarks of this period emerged, Human Genome Project (HGP), launched in 1990, which was completed in 2003 (Gibbs, 2020). Yet, during this period several other organisms had their genome sequenced, from Prokaryotes to Eukaryotes; e.g. *Haemophilus influenzae* (Fleischmann et al., 1995); *Saccharomyces cerevisiae* (Goffeau et al., 1996); *Escherichia coli* and *Bacillus subtilis* (Blattner et al., 1997; Kunst et al., 1997); the worm *Caenorhabditis elegans* (Consortium*, 1998); *Arabidopsis thaliana* (Kaul et al., 2000); and the *Drosophila melanogaster* (Adams et al., 2000; Myers et al., 2000). Together, all these projects established the genomic basis to evolutionary biology, and provided

unprecedented insights into the genetic underpinnings of metazoan diversity, adaptation, and diversification.

1.2.2. Omics and Next-generation sequencing

In the early 2000s, the advancements in sequencing and omic technologies designed a set of approaches commonly known as next-generation sequencing (NGS) technologies. The omic subfields (e.g. genomics, transcriptomics, metabolomics, lipidomics, epigenomics) emerged from molecular biology with the aim of collectively characterizing and quantifying biological molecules, including DNA, RNA, proteins, and metabolites. These approaches are frequently used to analyse organisms, species, and populations at molecular level (Figure 1.2). In particular, transcriptomics and genomics are two close subfields usually used in parallel. By default, genomics is a study of the complete set of DNA within an organism, encompassing all genome compartments and organizational structures. This includes the primary structural units—chromosomes—the fundamental units of heredity, the genes (comprising exons and introns), the non-coding DNA (ncRNA and regulatory sequences), the repetitive elements (including transposable elements and satellite DNA), mitochondrial and plastid genomes, the genome architecture and evolutionary patterns, among other aspects (Cho et al., 2009; Jackman et al., 2020; Rinn & Chang, 2012; Seehausen et al., 2014; Shapiro & Von Sternberg, 2005). Complementary, transcriptomics provided a dynamic perspective on how the genetic content is expressed and regulated in several biological contexts. In other words, the transcriptomics investigates the complete set of RNA transcripts produced from a genome at any time. This sub-field explore, the messenger RNA (mRNA), tRNA, rRNA, small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), small Interfering RNA (siRNA), microRNA (miRNA), long non-Coding RNA (lncRNA) and circular RNA (circRNA) (Figure 1.2) (Carthew & Sontheimer, 2009; Dunin-Horkawicz et al., 2006; Guo et al., 2014; Lowe et al., 2017; Warner et al., 2018).

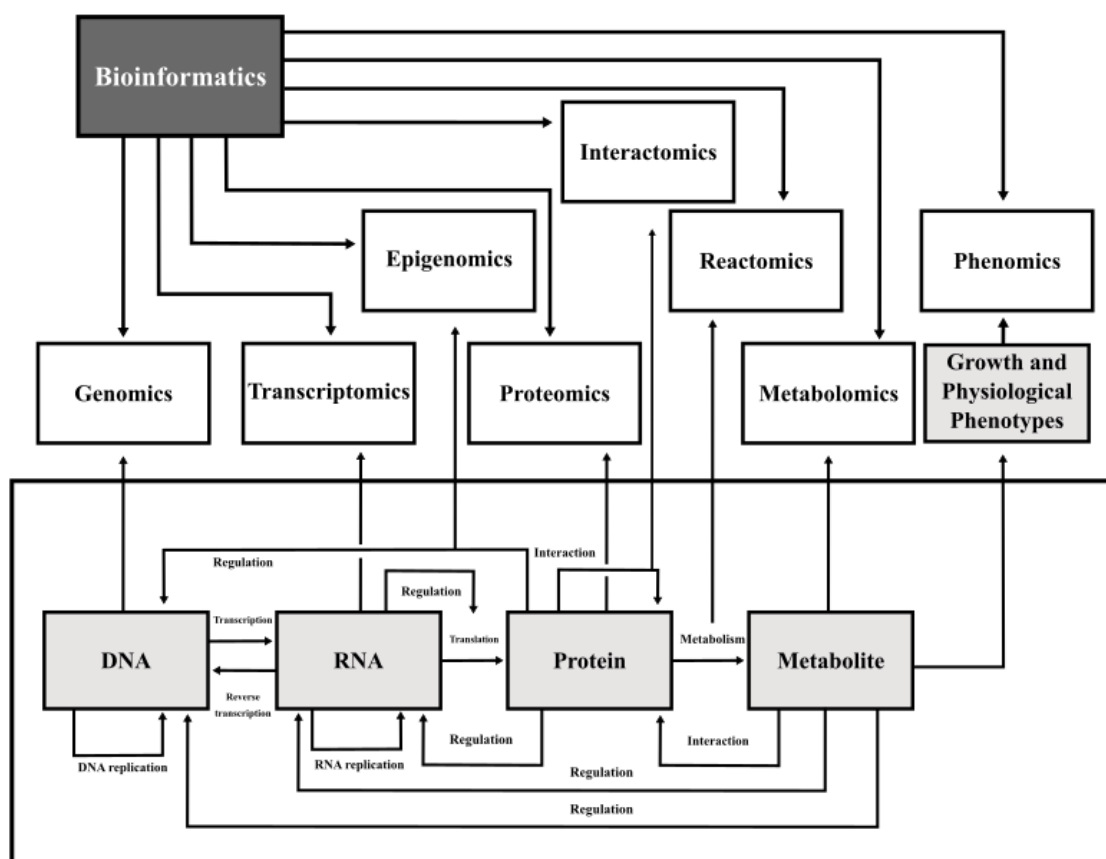


Figure 1.2 – Schematic overview of omics subfields of molecular biology (adapted from (Li et al., 2013)).

At the same time, during early stages of the NGS era, novel sequencing methods, such as Illumina and Ion Torrent, revolutionised the sequencing by drastically reducing the costs and enabling the high-throughput sequencing of entire genomes and transcriptomes. The major advantage of these methods lies in their ability to sequence millions of reads simultaneously. Despite this advancement, almost all approaches shared a common rationale: library preparation, DNA amplification, and sequencing (Giani et al., 2020; Goodwin et al., 2016; Heather & Chain, 2016; Hutchison, 2007; Shendure et al., 2017). Initially mostly methodologies produced short reads. However, over the last decade, both subfields have seen an accelerated adoption of long reads technologies. Short and long read strategies are typically applied depending on the biological question and sequencing goals. To generate short reads, one or multiple sequencing strategies can be used, sequencing by synthesis (SBS), sequencing by ligation (SBL) and synthetic long-reads (SLR). Conversely, long reads are generally produced using single-molecule sequencing (SMS), also referred to as single-molecule long-read sequencing (SMLS) (Goodwin et al., 2016; Sedlazeck et al., 2018). Among the above-mentioned strategies, the SBS strategy has been the most widely used, due to its ability to generate large volumes of data, high accuracy and low cost. Nowadays, this

methodology is predominantly employed by Illumina in several platforms, ranging from Solexa Genome Analyzer till the cutting-edge machine NovaSeq 6000. The NovaSeq 6000 offers high accuracy, multiple read-length options, competitive pricing, and a diverse array of library preparation methods (Goodwin et al., 2016). Despite its accuracy, short-read sequencing strategies have some limitations, particularly when used in the sequencing of large genomic datasets of organisms with large genome sizes, high repetitive content, and structural variability (Goodwin et al., 2016; Sedlazeck et al., 2018). To overcome these limitations, two SMS strategies have been developed: Single Molecule Real-Time (SMRT) sequencing, provided by Pacific Biosciences (PacBio), and the Nanopore sequencing commercialized by Oxford Nanopore Technologies (ONT) (Giani et al., 2020; Goodwin et al., 2016; Heather & Chain, 2016; Levene et al., 2003; Shendure et al., 2017; van Dijk et al., 2018). Initially, both PacBio and Nanopore sequencing companies invested in reads with an error rate around 5-15%. Pacbio with continuous long reads (CLR) achieved an average read length of 20 kbp, while nanopore read length ranged from 10 to 100 Kbp. Subsequently, Pacbio prioritised reducing sequencing error rates through Circular Consensus Sequences (CCS) technologies, by producing High fidelity reads (HiFi), while Nanopore focused on enhancing contiguity with Ultra-Long Read Sequencing Kits capable of produce read size 50kbp-4Mbp (Goodwin et al., 2016; Jain et al., 2018; Wenger et al., 2019). Importantly, both CLR and Nanopore sequencing methods often require third-party sequencing strategies (e.g. Illumina short reads) to error-correct and polish the long reads (Espinosa et al., 2024).

1.2.3. Databases and Bioinformatics

Not surprisingly, the advances in sequencing technologies led to an exponential number of genomic and transcriptomic projects, which produced thousands of Gigabases of datasets. In parallel, the public funding repositories such as GenBank, the US National Institute of Health (NIH) sequence database (<https://www.ncbi.nlm.nih.gov/genbank/statistics/>), European Molecular biology Laboratory (EMBL) Nucleotide Sequence Data Library (<https://www.ebi.ac.uk/history>) and the DNA Data Bank of Japan (DDBJ) (<https://www.ddbj.nig.ac.jp/about/index-e.html>) continued to be developed in order to accommodate such a quantities of data. The free availability of these repositories, combined with advances in sequencing technologies, has been instrumental in the democratization of genomics and transcriptomics. While NGS has broad application in field of transcriptomics, metagenomics or other genomic partitioning strategies like Anchored Hybrid Enrichment (AHE) or Restriction Site

Associated DNA Sequencing (Rad-seq), whole genome sequencing (WGS) remains the cornerstone and ultimate goal of genomics (Lewin et al., 2018).

Along with databases, also computational biology has undergone a major overhaul in the last 50 years. Computational biology provided the mathematical models, algorithms, and computational techniques to understand biological systems. In the NGS context, computational biology was essential to generate novel and efficient algorithms to process and analyse massive datasets generated by sequencing technologies. In parallel emerged the bioinformatics—a specialized subfield of computational biology—specifically focused on the analysis and interpretation of biological data. In NGS, this subfield is involved in the production of software tools, pipelines and databases, that facilitate the storage, processing, interpretation and analyses of biological information (Gauthier et al., 2019). One of the main goals of Bioinformatics is to standardize and integrate automation in NGS analyses, increasing efficiency and reproducibility while reducing the time, manual intervention and global effort needed in biological research (Spjuth et al., 2015). Nowadays, bioinformatics continues to evolve and the volume of data available led the researchers to develop workflow-based software platforms with user-friendly interfaces, that allowed the biologists to focus on biological interpretation rather than the intricacies of data handling. Usually, these platforms are fully automated and can be locally hosted, when local computational infrastructure is available, or cloud-based if no infrastructure is available (Krampis et al., 2012).

1.2.4. Bioinformatics in Genomics and Transcriptomics

In genomics and transcriptomics, the development of algorithms and tools progressed alongside the evolution of sequencing technologies. During the early stages of sequencing, bioinformatic tools were designed to handle Sanger. However, with the advent of NGS, new algorithms and software were developed to manage short-reads and, later, the long-read technologies. Simultaneously, processes for handling datasets became increasingly sophisticated, particularly with advancements in raw sequence search, alignment, mapping, assembly, and annotation.

In NGS, one of the most fundamental processes involves the nucleotide and protein sequence searching by similarity. To address this, BLAST was developed by Altschul and colleagues in 1997 (Altschul et al., 1997). Posteriorly, with the growing volume and diversity of datasets, new tools such as DIAMOND; Many-against-Many sequence searching (MMseqs2); Hidden Markov Model-based sequence analysis (HMMER) have been developed (Buchfink et al., 2014; Finn et al., 2011; Steinegger &

Söding, 2017). These tools are used as alternatives to BLAST in complementary ways to address specific research questions, improving both the speed and accuracy of sequence searches.

During the Sanger era, several software tools, such as Phil's Read Editor (Phred) and Phil's revised assembly program (Phrap), were developed to ensure quality control of sequencing raw data and initial genome assemblies (Ewing et al., 1998; Ewing & Green, 1998). These tools paved the way for more advanced software, with similar purposes, such as Trimmomatic, in the NGS era. In parallel, new tools and algorithms, including Celera, SOAPdenovo, SPAdes, ABySS, Trinity, Oases or IDBA-Trans were created to assemble genomic and transcriptomic datasets generated by NGS technologies (Sohn & Nam, 2018). These tools were instrumental to establish the first transcriptomic species references for various species based on short-read approaches. Additionally, in several cases, they served as foundation for the development of second and third generation of tools designed to process long reads provided by PacBio and Oxford Nanopore. Tools such as FALCON, Canu and Flye, Wtdbg2, NextDenovo and more recently, tools like Hifiasm and HiCanu have been and continue to be essential to assembly and processing of long read data (Cheng et al., 2021; Hu et al., 2024; Nurk et al., 2020; Ruan & Li, 2019; Sedlazeck et al., 2018; T. Zhang et al., 2022). The co-development of tools for structural and functional annotation of genomes and transcriptomes has also been a priority since the foundation of NGS. These tools rely on several fundamentals of comparative genomics, employing sequence searching methods to query publicly available databases and comparing genomic structures and features across species. Typically, for structure prediction, these methods utilise ab-initio software and models, such as Augustus or GeneMark, to predict genome or transcriptome structures, alongside with the integration of complementary omic technologies (Brûna et al., 2020; Stanke et al., 2004, 2006). Most pipelines incorporate several software tools, each providing different levels of evidence. Together, these evidences are then orchestrated and organized to deliver the structural and functional annotation of a given reference. Some of the most widely used pipelines, including Maker, Braker, EnTAP, Trinotate are continually evolving into new versions, adding new technologies, new genomic or transcriptomics features and new methods that enhance accuracy and speed of analyses (Brûna et al., 2021; Campbell et al., 2014; Hoff et al., 2015, 2019).

From a different perspective, the constant improvement of the assembly and annotation process, along with the creation of new references, has driven efforts to develop faster methods for aligning and mapping short and long reads to these

references. These methods are essential not only for the annotation and assembly processes themselves but also for addressing other biological questions at various levels, including intra- and interspecies comparisons, populations studies, and even ecosystems-level analyses. Some of the most renowned software for aligning and mapping such as BWA, Bowtie, STAR and HISAT were created, others such as BLASR and Minimap2 for long reads (Dobin et al., 2013; Kim et al., 2015, 2019; Li, 2018; H Li & Durbin, 2009).

1.2.5. Genomics and Transcriptomics in large consortia

To address the large-scale production of WGS datasets, significant research groups and consortia have been established to advance ambitious projects such as the Earth Biogenome Project (<https://www.earthbiogenome.org/affiliated-project-networks>), Darwin tree of life (DToL) Project, European Reference Genome Atlas (ERGA), DNAzoo Project and other related sub-projects like Vertebrate genome project or Squalomix. The primary aim of these projects is to sequence, assemble and annotate a vast number of genomes across the Metazoa kingdom, and setting new structures and features across across the Metazoa kingdom, and setting new benchmarks in genomics (Blaxter, 2022; Dudchenko et al., 2017; Kuraku et al., 2022; Lewin et al., 2022; Paez et al., 2022). In addition to this, these projects are addressing another significant challenge: the disparity of data available across the tree of life (Figure 1.3). In the early years of the NGS era, sequencing projects focused predominantly on economically relevant groups or model organisms, such *Homo sapiens* and other mammals, and only a couple years ago they began to delve deeply into non-model organisms. As result, groups that were underrepresented or entirely absent in ToL are now beginning to gain their first genomic datasets available (Stephan et al., 2022). These datasets are applied in multiple genomic-derived subfields such phylogenomics, population genomics, adaptive genomics, conservation genomics, epigenomics, metagenomics and mitochondrial genomics (Ashurst et al., 2005; Barghi et al., 2020; DeSalle et al., 2017; Formenti et al., 2022; Hohenlohe et al., 2021; Kapli et al., 2020; Mehrmohamadi et al., 2021; Quince et al., 2017).

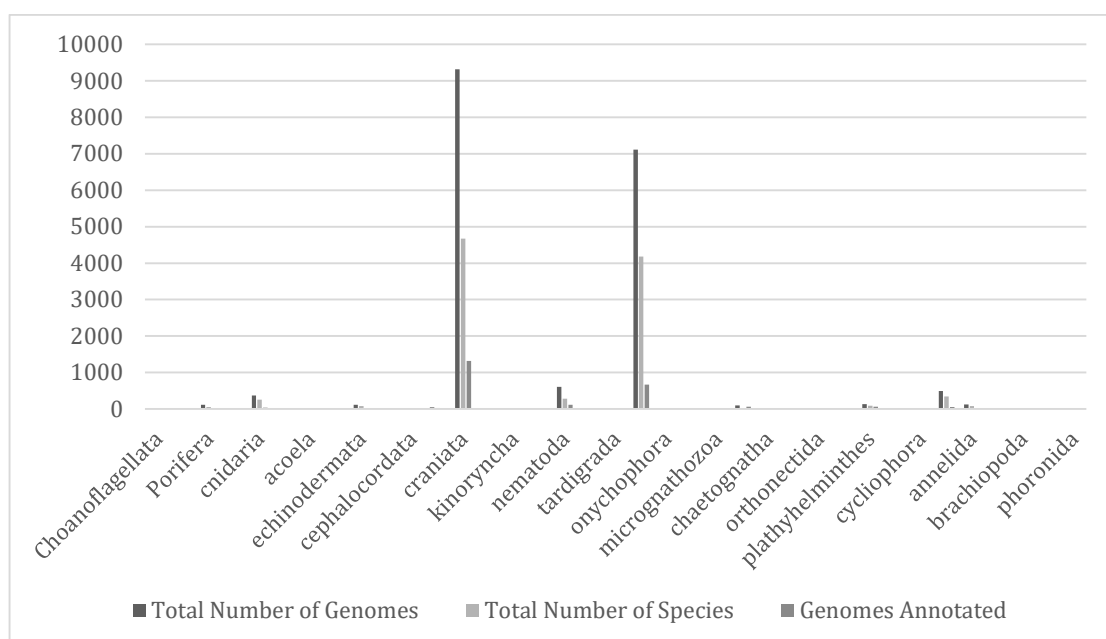


Figure 1.3 – Graphic showing the number of genomes, species with genome public available and genomes with annotation in NCBI (Retrieved from NCBI in 24/12/2024)

1.3. The Origins of the Chemical Defensome

The evolution of metazoan organisms and their capacity to resist, process and detoxify chemical compounds—whether at endogenous or exogenous level—requires a delicate balance between genetic adaptations and the environmental pressures. In nature, this balance has developed over millions of years, shaped by interactions between multiple taxa organisms, environments, as well as physical, chemical and geological phenomena. In last decades, humans have significantly accelerated this process, disrupting the natural equilibrium, by creating, changing and manipulating new and old materials and chemicals. The defence mechanisms of metazoan species have evolved in parallel with the emergence of natural chemicals, however the emergence of man-made/synthetic chemicals has introduced new challenges and consequences to these organisms.

1.3.1. Origin and impact of xenobiotics in ecosystems

To understand the metazoan mechanism of defence against a new set of chemicals, it is crucial to comprehend and review the trajectory of human beings and their impact on the Anthropocene epoch. The Anthropocene epoch, a period characterised by significant human impact on the Earth's geology and ecosystems, has been marked by continuous development and release of large amounts of synthetic chemical compounds (Brown et al., 2015; Sullivan et al., 2017). These chemical

compounds have been primarily introduced into the environment through industrial activities, such as agriculture, textile industry, urbanization and food processing, leading to profound ecological and health consequences. Presently, these substances are widespread in nearly all environments, with the aquatic systems being the most frequent final destinations and, consequently, among the most dramatically impacted (Sumpter, 2005).

Many of these chemical compounds are reported to pose significant risks to both human health and the environment (Noguera-Oviedo & Aga, 2016). Typically, xenobiotics are synthetic or natural, organic or inorganic, and are defined as exogenous chemical substances naturally absent in a given species or present at abnormal or unexpected concentrations (Park et al., 2014; Patterson et al., 2010; Rieger et al., 2002). Environmentally, xenobiotic concentrations can range from nanopollutant concentrations to high concentrations (using ng/L to µg/L units) depending on the specific compound and its source. Yet, among these contaminants, there is a significant class of compounds, classified as endocrine disruptors (EDCs), known to alter certain hormone levels of organisms—either increasing or decreasing the levels—by directly mimicking or blocking transcriptional activation of the receptors (Diamanti-Kandarakis et al., 2009; Tyler et al., 1998).

The origin and impact of these compounds have been extensively studied in both humans and wildlife. The trophic chain is the main pathway through which synthetic chemical compounds and EDCs reaching human beings and wildlife, through ingestion, inhalation and absorption (Bergman et al., 2013; Gore et al., 2015). Along this route they impact nearly all levels and Metazoa taxa as well as the ecosystem properties and interactions. In addition, many of these compounds exhibit bioaccumulation and biomagnification potential. Exemplifying, these contaminants can be transported through the terrestrial or aquatic food webs, where they may be ingested, inhaled or absorbed by animals and/or plants. These organisms can then be predated or captured by other animals, leading to the direct transmissions of these compounds. Furthermore, the continuous predation can lead to the mixing of several compounds and biomagnification of EDCs within the trophic chain (Previšić et al., 2021). Additionally, continuous exposure to these compounds can result in multigenerational and transgenerational effects on populations. These effects are particularly impacting when populations and/or species are permanently exposed during the vulnerable stages, such as foetus or neonate development. During these periods, the prolonged exposure can alter and disrupt the epigenetic germline markers, leading to adverse phenotypes across multiple generations (Anway & Skinner, 2006; Robaire et al., 2022; Walker & Gore, 2011; Xin et al., 2015).

In humans, the link between EDC compounds and several disorders and diseases is well-established. Initially, a direct correlation between some EDCs and reproductive and/or developmental conditions was identified, as well as some types of cancers (Heindel et al., 2015). More recently, numerous studies in both laboratory and animal research have demonstrated that several environmental EDCs also influence adipogenesis and obesity (Heindel et al., 2015). Naturally, these conditions further improve the risk of metabolic diseases such as type 2 diabetes and metabolic syndrome, thereby raising the probability of cardiovascular, pulmonary and liver issues (Casals-Casas & Desvergne, 2011; Gore et al., 2015; Papalou et al., 2019). In wildlife, several authors have reported EDCs substances as chronic and persistent, in multiple groups: for example birds, amphibians, fishes, insects, mollusks, bivalves, crustaceans, cnidarians (Barros et al., 2017; Blazer et al., 2011; Han et al., 2011; Pachura-Bouchet et al., 2006; Regnault et al., 2018; Sebastiano et al., 2016; Xia et al., 2010). In the case of aquatic species, for instance *Danio rerio*, the EDCs effects impact various dimensions, from the growth and reproduction parameters till stress response and social interactions (Han et al., 2011; Xia et al., 2010). In 2018, da Silva and co-workers reviewed and assessed several studies on the main effects of EDCs in various aquatic species. In this study it was found that, in amphibians and fishes, the reproduction and embryonic development are the most affected, leading to cases of gonadal abnormalities, hermaphroditism, and, consequently to an increase in mortality and reproductive unsuccess (da Silva et al., 2018). These and similar patterns (for instance imposex) also occur in invertebrate species. The imposex (irreversible development of male secondary sexual characteristics in females) occurs in marine gastropods when the specimens are exposed to the endocrine disruptor tributyltin (Gibbs & Bryan, 1986; Matthiessen & Gibbs, 1998; Santos et al., 2002).

1.3.2. From chemical defence to the chemical defensome

The mechanisms of chemical defence in Metazoa species are crucial for survival, protection, and the thriving of organisms in diverse ecosystems and environments. Commonly, the term “chemical defence” is applied at organismal level and associated with the mechanism of production and utilization of endogenous (e.g. production of secondary metabolites) and/or exogenous chemical compounds to protect or deter organisms from external threats such as parasites, predators or competitors (Berenbaum, 1995; Blum, 2012; Mithöfer & Boland, 2012). Nevertheless, the chemical defence definition can be applied at different organizational levels (e.g. molecular, cellular, tissue/organs, population, ecosystem), depending on the mechanisms of action,

the type and origin of chemical compounds, and the environmental context (Crespi et al., 2022). From a cellular perspective, chemical defence encompasses a set of components and biochemical pathways employed for self-protection against harmful chemicals. Naturally, all these elements are interconnected, operating like an army that provides resistance to several xenobiotics, transforms and eliminates external and internal toxic compounds, maintains cellular homeostasis and protect organisms from harmful effects of xenobiotics and/or endogenous compounds such as reactive oxygen species (ROS) (Bard, 2000; Iyanagi, 2007; Li et al., 2007). At the molecular level, these defence mechanisms are constituted by thousands of transcription factors, genes and enzymes, organized and integrated, in a complex and interactive network of pathways, commonly referred to as the “chemical defensome” (CD). The term “chemical defensome” was first introduced in the early 2000s by Goldstone and co-workers (Goldstone, 2008; Goldstone et al., 2006; Stegeman et al., 2010). Both the term and its definition are associated to the early stages of the NGS era, during which entire genome and transcriptome references were massively sequenced, assembled and annotated. In the molecular context, whole genome/transcriptome sequencing projects facilitated the identification of gene portfolios and a better understanding of the interactions and relationships between chemical-genes, chemical-proteins, and, gene-gene in several species. Complementary, the transcriptomic projects have also been essential in identifying the transcription factors that regulate chemical defensome genes enhancing our understanding of regulatory networks (Aranguren-Abadía et al., 2020; De Marco et al., 2017; Eide et al., 2021; Goldstone, 2008; Goldstone et al., 2006; Kim et al., 2019; Roncalli et al., 2025; Shinzato et al., 2012; Strobel et al., 2022; Yadetie et al., 2012, 2022). While most of these projects assess and monitor the impact of harmful compounds on organisms, species, populations and/or environments, other research focuses almost exclusively on the economic potential of exploiting these genes, in agriculture, disease control, pharmaceutical and drug development potential (Aurade et al., 2010; Bariami et al., 2012; De Marco et al., 2017; Encerrado-Manriquez et al., 2025; Shan et al., 2021). From an ecological and conservation perspective, to define the CD catalogue of genes has also been proven to be essential to understand the colonisation and, consequently, the adaptation and resilience of certain species in various environments, particularly in cases of severe pollution and habitat degradation. For example, invasive species such as *Neogobius melanostomus* possesses a natural advantage over native and endogenous species with a more limited catalogue of CD genes (Adrian-Kalchhauser et al., 2020).

1.3.3. Molecular overview of chemical defensome – The three Phases

To understand the chemical defensome elements and their function, it is crucial to comprehend the diversity, type and origin of chemical compounds. The origin of such compounds is diverse. In the context of CD, the exogenous man-made xenobiotics derived from drugs and pharmaceuticals (e.g. paracetamol and ibuprofen), pesticides and herbicides (e.g. organophosphates and glyphosate), industrial and environment pollutants (e.g. polycyclic aromatic hydrocarbons (PAHs), dioxins, heavy metals) and food-derived (e.g. acrylamide and benzoates) are the most common (Chen et al., 2015; Gibson & Harwood, 2002; Žur et al., 2018). Regarding endogenous compounds processed by CD, these can be categorized in five major groups: reactive metabolites, bile Acids and heme products, steroids hormones, neurotransmitters and endogenous toxins (Goldstone, 2008; Goldstone et al., 2006; Goodwin et al., 2000; Marrone et al., 2012). Several of these compounds are essential for cellular function; however, in high concentrations they can become detrimental and toxic. The equilibrium of these compounds within each category is essential for the cell. For instance, reactive metabolites such as superoxide (O_2^-) and nitric oxide (NO) are essential as signalling molecules for cell homeostasis; however, in excess, they can damage DNA, proteins, and lipids (He et al., 2017). Bile acids play a crucial role in lipid absorption and cholesterol homeostasis, while heme products are pivotal for oxygen transport and mitochondrial reactions; however, in unbalanced concentrations, may become cytotoxic or promote oxidative stress, respectively (Khan & Quigley, 2011; Lefebvre et al., 2009). Steroid hormones such as cortisol have a significant impact on glucose metabolism, nevertheless, dysregulations in cortisol metabolism can contribute to metabolic disorders (Kivimäki et al., 2022). Regarding endogenous toxins, urea is essential for nitrogen homeostasis. Nonetheless, impaired excretion or imbalance may lead to uremic toxicity, and consequently increase oxidative stress and cellular dysfunction (Weiner et al., 2014).

Naturally, these chemical compounds enter or are transported into and between cells using different mechanisms, depending on their size, polarity and chemical features. These mechanisms include passive or facilitated diffusion, for lipophilic and hydrophilic molecules, respectively, active transport via adenosine triphosphate (ATP) dependent transporters to move molecules against concentration gradients, paracellular transport for small and water-soluble molecules, and complete systems such as endocytosis for large molecules (Blais et al., 2007). These processes are commonly used to import and export xenobiotic chemicals to and from cells. Once inside the cells, chemicals are subject to the metabolism of xenobiotics, which involves several chemical defensome related transcription factors, genes and pathways. These elements are

organized within the cell according to their function and properties. Among these are elements responsible for sensing and identification, structures involved in processing and metabolism, and specialized pathways for the inactivation and elimination of compounds (Stegeman et al., 2010). In this context, the xenobiotics frequently bind with specific receptors, notably transcription factors, triggering a cascade of signalling events that lead to transcriptional activation of chemical defensome genes (Xu et al., 2005).

In detail, the metabolism of xenobiotics is subdivided into three main phases, phase I (oxidizing biotransformation), phase II (reducing and conjugation enzymes) and phase III (transporters and toxicants eliminations) (Figure 1.4) (Xu et al., 2005).

1.3.3.1. Phase I

Typically, the phase I serves as the entry point of the detoxification process and initial point of biotransformation, during which several enzymes, by adding or removing functional groups from a substrate, increase their reactivity and solubility (Testa & Krämer, 2007a, 2007b). Phase I encompasses three main reactions, oxidation, reduction and hydrolysis. Among these, oxidation reactions are the most frequent, introducing or exposing polar functional groups on xenobiotic molecules, thereby increasing the solubility of lipophilic compounds and facilitating their subsequent conjugation and excretion (Testa & Krämer, 2007a, 2007b).

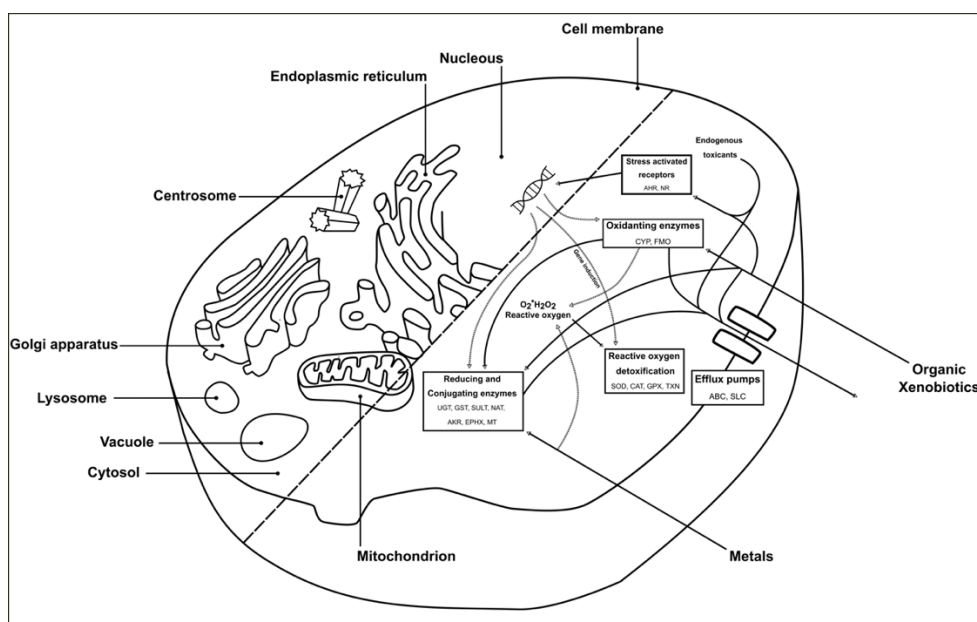


Figure 1.4 – Chemical defensome overview (Adapted from (Goldstone et al. 2006))

In contrast, the reduction reactions generally involve the gain of electrons or hydrogen atoms and are mediated by reductase enzymes. These reactions have been described as playing a crucial role in the processing of compounds encompassing reducible

functional groups such as azo ($-N=N-$), nitro ($-NO_2$), and carbonyl groups (Testa & Krämer, 2007a). The third reaction, hydrolysis, involve enzymes such as amilases (amide), esterases (ester), epoxide hydrolases (EH) (epoxide), among others, to cleave the respective compounds. Usually, this cleavage is performed by adding a water molecule to the xenobiotics, breaking their chemical bonds and converting them into a simpler and reactive fragment (Testa & Krämer, 2007b)

The phase I of xenobiotics encompasses a set of gene families such as Alcohol and Aldehyde dehydrogenases (ADH and ALDH), Monoamine Oxidases (MAO) or other Amide Oxidases like Diamine oxidase (DAO) or Semicarbanize-sensitive amine oxidase (SSAO), esterases and amidases, flavin-containing monooxygenases (FMO) and one superfamily of enzymes that catalyzed the majority of reactions, the cytochrome P450 (CYP) (Cashman, 1995, 2018; Fretland & Omiecinski, 2000; Gonzalez & Yu, 2006; Guengerich, 1999; Meyer, 1996; Nebert & Russell, 2002; Nebert & Vasiliou, 2004; Poulsen & Ziegler, 1995; Rendic & Guengerich, 2015; Seidegård & Ekström, 1997; Shih et al., 1999; Vasiliou et al., 2000, 2004; Zanger & Schwab, 2013). The cytochrome P450 enzymes are a diverse group of heme-containing monooxygenase that play a critical role in the activation of a large group of compounds. With more than 18 families and 42 subfamilies of genes in *Homo sapiens*, it is expected that members of this superfamily catalyze more than 90% of enzymatic reactions in chemical metabolism (Nebert & Russell, 2002; Rendic & Guengerich, 2015; Zanger & Schwab, 2013). Moreover, some studies postulate that 75% of all drug metabolism reactions are catalysed by only five CYPs, CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4 (Rendic & Guengerich, 2015). These findings are in line with various studies where CYP1A and CYP1B subfamilies have been identified as crucial to the processing of polycyclic aromatic hydrocarbons, while regulated by aryl hydrocarbon receptor (AHR) (Denison & Nagy, 2003; Shimada & Fujii-Kuriyama, 2004). Additionally, also CYP2 and CYP3 families have been largely associated with the metabolism of xenobiotics, being responsible for the metabolism of the major part of drugs (e.g. CYP2C19 – Omeprazol; CYP2D6 – Codeine and several antidepressants; and CYP3A4 with broad range of substrates and reported as metabolize more than 50% of all drugs) (Guengerich, 1999; Zanger & Schwab, 2013). In parallel, also FMO enzymes play a vital role in drug and xenobiotic metabolism. Flavin-containing monooxygenases are responsible for the N- and S- oxygenation of soft nucleophilic heteroatom small molecules, usually containing sulfur, selenium, nitrogen, among others (Cashman & Zhang, 2006; Poulsen & Ziegler, 1979; Ziegler et al., 1971). Within the FMO family, several genes have been studied, including FMO3 and FMO5. While FMO3 is clear involved in the metabolism of Trimethylamine (TMA), the FMO5 has

been documented as a non-standard monooxygenase with the ability to perform Baeyer-Villiger oxidation and the potential to be engaged in the metabolism of aldehydes and ketones (Cashman, 2005; Cashman & Zhang, 2006; Fiorentini et al., 2016; Krueger & Williams, 2005). The involvement of other families of enzymes such MAO in chemical metabolism, specifically in drug metabolism, has been shown to be less frequent. Rendic and Guengerich estimate that only 3% of drug metabolism reaction are catalyzed by member of these gene families (Rendic & Guengerich, 2015). In the metabolism of xenobiotics, these gene families participate in oxidation, activation and deactivation of various compounds. While ADH and ALDH are involved in metabolism of of alcohols, MAO participates in deamination of xenobiotic amines. Other enzymes, such as esterases and amidases, are commonly associated with activation and deactivation of ester and amide -containing compounds, facilitating their clearance from cells (Cashman, 2005; Shih et al., 1999; Testa & Krämer, 2007b; Vasiliou et al., 2000; Yu et al., 2003; Zanger & Schwab, 2013).

1.3.3.2. Phase II

In phase II, the functionalized xenobiotics originating in phase I are usually conjugated with endogenous hydrophilic groups in order to increase their solubility and facilitate their detoxification. In this phase, specific sets of enzymes are responsible for these conjugations, adding specific groups to either activated or deactivated xenobiotics. Overall, this phase encompasses five major families of enzymes; the UDP glucuronosyltransferases (UGTs), sulfotransferases (SULTs), Glutathione S-transferases (GSTs), N-acetyltransferases (NATs) and Methyltransferases (MT). Usually, UGT enzymes catalyze the transfer of glucuronic acid to substrates in a biochemical process known as glucuronidation (Rowland et al., 2013; Tukey & Strassburg, 2000) process accounts for approximately 35% phase II conjugations in drug metabolism (Evans & Relling, 1999; Yang et al., 2017). This superfamily encompasses more than 22 proteins, distributed across five families and six subfamilies (Rowland et al., 2013). Among these, UGT1A1 and UGT2B7 are frequently regarded as the most clinically relevant, as they have been proven to play roles in inherited diseases of bilirubin metabolism (at endogenous level) and in biotransformation of drug xenobiotics (exogenous level), such as Irinotecan - commonly used in chemotherapy of several cancers - and morphine (Rowland et al., 2013; Tukey & Strassburg, 2000)

Similarly, sulfotransferases also play a vital role in the mechanism of drug detoxification. These enzymes are part of a process called sulfonation, in which a sulphate group is transferred to a hydroxyl or amine group of a xenobiotic and/or

endogenous compound (Gamage et al., 2006; Glatt, 2000). Importantly, this superfamily comprises more than 47 enzymes in mammals, including two of the most studied SULTs enzymes: SUL1A1 and SULT2A1, involved in metabolism of phenolic drugs, (e.g. paracetamol), and steroids (e.g. dehydroepiandrosterone, DHEA), respectively (Jancova et al., 2010; Nagata & Yamazoe, 2000; Yamamoto et al., 2015).

The glutathione S-transferases are essential for catalysing reactions involving hydrophobic and electrophilic substrates. Typically, these enzymes act as catalysts in a process called glutathionylation, in which an electrophilic substrate is conjugated with the glutathione (GSH) tripeptide in order to perform a nucleophilic substitution or addition and neutralizing the reactivity of the compound (Hayes et al., 2005; Mannervik et al., 1988). This superfamily encompasses at least 21 GST enzymes in *Homo sapiens*, distributed in three main categories: cytosolic, microsomal and mitochondrial. Interestingly, cytosolic GST are further subdivided into at least six classes, Alpha (GSTA), Mu (GSTM), PI (GSTP), Theta (GSTT), Omega (GSTO), Zeta (GSYZ), Sigma (GSTS) (Hayes et al., 2005; Marrs, 1996; Nebert & Vasiliou, 2004; Tan & Low, 2018). Despite their primary function in conjugation of oxidative stress products, several studies have highlighted their additional roles in the metabolism of PAHs (GSTM1) and various xenobiotics, including anticancer agents (GSTP1) (Hayes et al., 2005; Townsend & Tew, 2003). Another important process within the phase II is the acetylation, which involves N-acetyltransferases catalysing the addition of an acetyl group ($-\text{COCH}_3$) to aromatic or heterocyclic amine, or hydrazine compounds, thereby increasing their solubility or reducing their reactivity (Butcher et al., 2002). Despite their significance, only a few genes are associated with this family, including NAT1 and NAT2, which are responsible for the acetylation of hydroxylamines, isoniazid, sulphonamides and other pharmacological drugs (Fullam et al., 2008; Hein, 2002; Sim et al., 2008). Finally, the methyl conjugation enzymes are involved in the detoxification of several drugs and xenobiotics. The mechanism of methyltransferases action is similar to that of other transferases. In methylation, a methyl group is transferred to nucleophilic functional groups such as hydroxy, amine or thiol groups (Abdelraheem et al., 2022; Weinshilboum et al., 1999; Weinshilboum & Wang, 2004). This superfamily includes thousands of members, though the precise number of genes, phylogenetic families and subfamilies remain unclear, primary due the extensive nucleotide and amino acid variation among the methyltransferases. Among these, thiopurine S-Methyltransferase (TPMT) is perhaps the most extensively studied, owing to its catalytic role action in metabolizing thiopurine substrates such as azathioprine and 6-mercaptopurine (6-MP) - agents

commonly used in immunosuppressive therapy and chemotherapy of leukemia, and organ transplantation (Weinshilboum et al., 1999; Weinshilboum & Wang, 2004).

1.3.3.3. Phase III

The phase III of drug metabolism involves the transport and/or expulsion of drugs and their metabolites across cellular membranes. This transport is primarily mediated by the members of two superfamilies of proteins: ATP-binding cassette (ABC) and solute carriers (SLC) transporters. Both ABC and SLC genes possess unique mechanisms of action. While the ABC proteins are capable of actively effluxing compounds across membranes, against their concentration gradients, by using ATP as an energy source, SLC transporters facilitate the diffusion of solutes or act as secondary active transporters. Interestingly, one of the main distinctions between these two types of transporters lies in their functional directionality: while ABC transporters primarily expel solutes, SLC transporters often function as gatekeeper, enabling the bidirectional movement of solutes, in accordance with electrical gradients (Giacomini et al., 2010; Nigam, 2014). This bidirectional nature of SLCs has led several authors to propose the existence of a potential fourth phase in drug metabolism, phase 0, a pre-metabolic drug stage, in which SLCs transporters act as uptake systems, enabling the active import into the cell (Döring & Petzinger, 2014; Y. Li et al., 2019).

The human ABC superfamily of genes comprise 48 members distributed in seven subfamilies, ABC(A-G), and is cross all eukaryotes. These membrane proteins transport a wide range of substrates and are structurally formed by two ATP-binding domains (also known as nucleotide-binding folds, NBFs) and two transmembrane domains (TM). This structure, 2NBFs + 2 TM represents the conventional vision of a complete ABC transporter. However, some ABC transporters function as dimers (both homo- and hetero) formed by two half transporters. The NBFs are commonly used to identify the ABC transporters based in sequence similarity and to classify them into one of the seven pre-established subfamilies. Structurally, the subfamilies ABC (A, C) consist only of full transporters; ABC (D, G, H) include only half transporters; ABC (B) contains both full and half transporters; and finally, ABC (E, F) proteins include that comprise NBFs without TM. The members of ABC (E, F) are not considered functional transporters, although they are phylogenetic related to the remaining ABC members. Among the ABC transporters, members of B, C and G subfamily are most strongly associated with drug metabolism (Alam & Locher, 2023; Dean et al., 2001; Dean & Annilo, 2005; Hediger et al., 2004, 2013). The ABCB1 gene encodes a canonical efflux pump and has been widely associated with the transport of drugs. Numerous studies have reported its essential role

in Ivermectin sensitivity. The members ABCC subfamily (also known as multidrug resistance proteins) also play key roles in drug metabolism, with several members being involved with the export of anticancer drugs, as well as sulphate, glutathione and glucuronide-conjugated xenobiotics (e.g., ABCC1, ABCC2) (Borst et al., 2000; Giacomini et al., 2010). The ABCG2 has also been reported as model gene for studying drug metabolism. This broad-spectrum transporter is a major player in transporting a wide range of xenobiotics, including several anticancer drugs such as topotecan (Jonker et al., 2002).

Complementarily to ABC transporters, SLC transporters are essential for the movement of solutes across membranes. SLC superfamily is one of the largest and most diverse families of transporters in metazoans. Comprising approximately 350-450 members across 52-66 families in humans, SLC members mediate the facilitated or active transport of several xenobiotic and endogenous solutes across membranes in all eukaryotes (Ferrada & Superti-Furga, 2022; Hediger et al., 2004, 2013; Höglund et al., 2011; Lin et al., 2015). In recent years, multiple SLC families have been associated with the uptake or excretion of drugs. Among them, OATP1B1 and OATP1B3, from SLCO family, are frequently linked with hepatic uptake of statins and the transport of chemotherapeutic compounds. Members of SLC22 (including organic anion and cation transporters, OATs/OCTs) and the SCL10 family have also been consistently associated with excretion of sulphate-, glutathione-, and glucuronide-conjugated compounds. Furthermore, members of SLC47 (multidrug and toxin extrusion, MATEs) play a vital role in the excretion of organic cations into urine (Hediger et al., 2013; Lin et al., 2015; Liu, 2019; Pelis & Wright, 2014).

Overall, ABC and SLC transporters play an essential role completing the three-phase stages of the drug metabolism by eliminating drug compounds from cells and facilitating their final excretion, thereby reducing the intracellular drug accumulation and maintaining cellular homeostasis.

1.3.3.4. Main regulators of Phase I, II, III drug-metabolizing enzymes

The three phases of drug metabolism are regulated by members of three principal families or superfamilies of chemical sensors: Per-ARNT-Sim (PAS) family, classical nuclear receptors (NRs), and elements of Cap 'n' Collar -Basic Leucine Zipper (CNC-bZIP) transcription factor family.

The PAS family comprise genes containing a PAS domain and typically regulates gene expression either through protein-protein interactions or by acting as transcription factors. The Aryl hydrocarbon receptor (Ahr) is the most well-studied xenobiotic sensor

of this family. In its inactive state, AhR is located in the cytoplasm, bound to chaperones (e.g. HSP90). When activated by ligands such as dioxin, it undergoes a configurational change, dissociates from chaperone and is translocated into the nucleus. Within the nucleus, it dimerizes with Aryl hydrocarbon receptor nuclear translocator (ARNT), forming an AhR-ARNT heterodimer with the ability to bind canonical xenobiotic response elements (XREs). After binding to XREs, coactivators are recruited, which leads to the regulation of target gene expression (McIntosh et al., 2009; Möglich et al., 2009; Xu et al., 2005). In the context of CD, AhR has been associated with regulation of CYP1 DMEs and several GSTs and UGTs (Figure 1.5) (Xu et al., 2005).

The nuclear receptors also play a crucial role in the regulation of drug-metabolizing enzymes (DMEs). This superfamily includes genes such as the pregnane-X-receptor regulator (PXR), constitutive androstane receptor (CAR), farnesoid-X-receptor (FXR), liver-X-receptor (LXR), peroxisome proliferator-activated receptors (PPARs), among others, which are engaged in the three phases of drug metabolism (Francis et al., 2003; Kliewer et al., 1998, 2002; Xu et al., 2005). Mechanistically, some NRs are able to dimerize with a partner nuclear receptor (equivalent to ARNT in the AhR pathway), the retinoid-X-receptor (RXR). Despite the functional similarities between ARNT and RXR, there are at least two fundamental differences: the ligand-binding ability (RXR can bind to ligands, whereas ARNT is constitutively nuclear and ligand-independent), and the DNA-binding specificity of their complexes (AhR-ARNT binds to XREs, whereas NRs-RXR bind to hormone response elements (HRE) (Evans & Mangelsdorf, 2014; Mangelsdorf & Evans, 1995). Regarding CD, the nuclear receptors regulate the majority of drug metabolism reactions, acting directly and indirectly on the CYP2, CYP3, CYP4 and CYP7 families in Phase I; on several UGTs, SULTs, GSTs, NATs in Phase II; and, on numerous ABCs and SLCs in Phase III (Figure 1.5) (Goldstone et al., 2006; Xu et al., 2005).

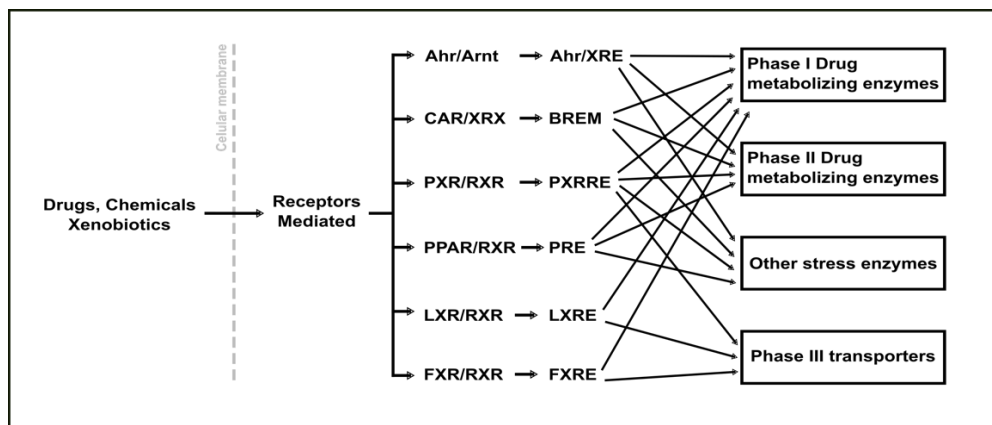


Figure 1.5 – Illustration showing how drugs, chemicals and xenobiotics activate receptor-mediated genes and form heterodimers, and their involvement in each phase of drug metabolism (adapted from (Xu et al., 2005)).

This illustration shows how drugs, chemicals, and xenobiotics activate specific receptor-mediated genes, potentially form heterodimers, and are involved in each phase of drug metabolism. Lastly, the CNC-bZIP transcription factor family is mainly responsible for the regulation of oxidative stress responses in Phase II and III of drug metabolism. While the bZIP domain is common to several transcription factors, the Cap 'n' Collar domain is specific to this family (Motohashi et al., 2002). Among CNC-bZIP family members, NRF2 (Nuclear factor erythroid 2-related factor 2) is the best known and most extensively studied. Similar to several members of the PAS or NRs families, NRF2 is basally located in the cytoplasm under control of inhibitory proteins such as KEAP1. When exposed to ROS, electrophiles or xenobiotic aggressors, the NRF2 is released and translocated to the nucleus, where it uses small Maf (Musculoaponeurotic fibrosarcoma oncogene) proteins to form a heterodimer complex (NRF2-MafG). Subsequently, this complex binds to antioxidant or electrophile response elements (ARE / EpRE) or Maf recognition elements (MAREs), thereby activating or repressing the transcription of target genes (Niture et al., 2014; Tonelli et al., 2018). Importantly, transcription factors of this family have been associated to the regulation of critical stages of drug metabolism, such as the glutathione conjugation or glucuronidation of several substrates in Phase II (E.g. By mediating genes such as GSTA1, GSTP1, UGT1A1) and the efflux of conjugated metabolites in Phase III (e.g. several members of the ABCC and ABCG families) (Nguyen et al., 2003; Xu et al., 2005)

1.3.4. Chemical defensome in Metazoa

The study of chemical defensome as a concept emerged in the last 20 years. The primary questions concerning the CD involved the identification of the global portfolio of CD genes, variations in the typology and number of CD elements, and the categorization and relationships between CD genes among species. Before the CD emerged as a network concept, numerous authors explored CD at gene or gene family level in a wide range of metazoan species. Notwithstanding, in this subchapter (1.3.4), I will summarize some key studies that highlight the variation and composition of CD portfolio across Metazoa.

As mentioned in Section 1.3.3 of this thesis, the CD encompass a network of transcription factors, genes, and enzymes that provide organisms with defences against chemical stressors, by sensing, transforming, and eliminating chemical toxins and xenobiotics. This portfolio of genes is highly dependent on the habitat, fitness adaptation, environmental pressure, and developmental stage of organisms (Roncalli et al., 2025) (Roncalli et al. 2025). Since the identification of the “first” catalogue of chemical

defensome genes by Goldstone and colleagues in *Strongylocentrotus purpuratus* (*S. purpuratus*), several other species of vertebrates and invertebrates have been scrutinized (Goldstone et al., 2006). In the sea urchin study, the authors used *in silico* analyses to establish the composition of CD gene families and their expression at both the adult and embryonic stages. Furthermore, they performed a highly descriptive and comparative study involving several other chordates—human, zebrafish and tunicates (Goldstone et al., 2006). One of the main goals of this study was to evaluate CD components in the deuterostome group. The authors found that most families or superfamilies of CD genes are conserved among Deuterostomia groups. Notwithstanding, they also identified several families of genes, such as, ABC, UGT and CYP, that were expanded in the sea urchin when compared with the remaining species. After this breakthrough in 2006, several studies adopted the same principles to identify and understand the complete or partial CD elements and/or their functions in chordates. Within the chordates, mammalian subclass and teleosts infraclass were the most explored – *Scomber colias*, *Salmo trutta*, *Squalius cephalus*, *Lepomis gibbosus*, *Barbus barbus*, *Cottus gobio*, *Gasterosteus aculeatus*, *Oryzias latipes*, *Fundulus heteroclitus*, *Basilichthys microlepidotus*, *Pterygoplichthys anisitsi*, *Chionodraco hamatus* and *Trematomus loennbergii* (Cortés-Miranda et al., 2024; De Marco et al., 2017; Eide et al., 2018, 2021; Franco, Araújo, et al., 2025; Franco, Hollender, et al., 2025; Lu et al., 2025; Machado, Gomes-dos-Santos, et al., 2022; Parente et al., 2017; Stegeman et al., 2010; Strobel et al., 2022; Yadetie et al., 2012, 2022). Despite the consistent findings across several species of deuterostomes, in 2012, Yadetie et al. found considerable discrepancies in the CD of the tunicate, *Oikopleura dioica* (Yadetie et al., 2012). In this species, the authors reported a reduced repertoire of CD genes, with some key genes and gene families missing: subfamily CYP1, and the master regulators AhR and PPAR (Yadetie et al., 2012). Interestingly, several of these CD genes had already been identified years earlier in other tunicates, *Ciona intestinalis* and *Ciona savignyi* (Goldstone et al., 2007; Nelson et al., 2013). Notwithstanding, variations in CD gene repertoire, as result of duplications, expansions or loss in chemical defensome is relatively common in metazoan, but hampered by missing genome/transcriptome data. The CYP superfamily provides as model example of such family diversifications and expansion. In 2013, Goldstone and co-workers performed a comprehensive analysis of the CYP superfamily across Metazoa and found highly variable gene numbers among species (e.g., the choanoflagellate, *Monosiga brevicollis*, contains only seven CYPs, while *Brachiostomata floridae* and *S. purpuratus* possess more than 216 and 120 CYPs, respectively) (Nelson et al., 2013). In vertebrates, variations in CYP genes are even more pronounced, particularly in CYP 1, 2 and 3 families, which have undergone considerable

expansions and diversification during vertebrate evolution (Nebert & Dalton, 2006). Additionally, other gene families such as, UGT, SULTs and GST families have also undergone multiple duplications events, allowing either a narrowed or broadened specificity for multiple substrates (Burchell et al. 1995; Hayes et al. 2005). In the evolution of ABC and SLC transporters, several studies have shown a broad conservation across vertebrates, alongside recent duplications and gene losses in certain lineages (Dean & Annilo, 2005; Ferrada & Superti-Furga, 2022; Höglund et al., 2011). In addition to the expansions and diversification, gene loss has also shaped the evolution of vertebrates chemical defensome. An example of this phenomenon is the case of the master regulator PXR in teleosts. Eide and colleagues (2018) inspected the genomes of 77 teleosts and found the PXR gene absent in 34 of the 77 species analysed. In this study, authors suggest that AhR may have acquired an additional regulatory role, controlling part of the original PXR target gene set (Eide et al., 2018). Similarly, in Chondrichthyes, PXR was also suggested to be lost in Elasmobranchii (sharks and rays) yet retained in Holocephali (chimaeras) (Fonseca et al., 2019, 2020).

Looking beyond the Deuterostomes, in the 2006 study by Goldstone and colleagues, the authors proposed that the majority part of CD gene families were also present in protostomes, suggesting a pre-Bilateria origin of the orchestrated CD (Goldstone et al., 2006). This suggestion was further confirmed by subsequent analyses of three cnidarians—*Nematostella vectensis*, *Acropora digitifera* and *Hydra magnipapillata* — in which authors found a CD gene family composition similar to that in deuterostomes (Goldstone, 2008; Reitzel et al., 2008; Shinzato et al., 2012). In these studies, authors performed extensive genomic and transcriptomic surveys, identifying CYPs, GSTs and ABCs as members of the three phases of drug metabolism (Goldstone, 2008; Shinzato et al., 2012). Interestingly, in *N. vectensis* only CYPs and ABC superfamilies display a significant family diversity, while the remaining superfamilies appear to have a limited number of gene expansions or novel genes (Goldstone, 2008). On the other hand, in *A. digitifera* only EH and NAT genes of phase I and II of drug metabolism, respectively, are expanded. Curiously, in the above studies, authors argue that the major part of CD elements acts at the embryonic stage, establishing that CD is mainly active in early development stages of these organisms (Shinzato et al., 2012).

Among protostomes, several Ecdysozoan and lophotrochozoan species have been studied. Usually, the arthropod Ecdysozoans, are the primary targets owing to their impact in agriculture, importance for the chemical industry and economic relevance (e.g., insecticide resistance). For instance, in 2017 De Marco and co-workers identified in the transcriptome of *Anopheles stephensi* larvae, a portfolio of 272 CD genes. About 177 of

the 272 were differentially expressed when subjected to a 48-h experimental assay with permethrin, with 67% of these genes belong to phase I, 68,5% to phase II and 53% to phase 0/III of drug metabolism (De Marco et al., 2017). In copepods the CD portfolio has also been examined in several species. Different authors have reported a lower number of genes for most CD families (with the exception of ABC, GST and AKR families) and a higher number in oxidative enzymes, compared with *S. purpuratus* and *N. vectensis* (Choi et al., 2021; Roncalli et al., 2025; Yadetie et al., 2022; Zhou et al., 2018). An identical pattern has been found in crustaceans. In 2021, the genome sequencing project of *Homarus americanus* species provided a complete overview of CD in lobsters (Polinski et al., 2021). In this study, Polinsky and co-authors found several CD gene families expanded in lobster, including CYPs, GTXs, SODs and SLCs. Among these, the authors highlight the expansion of CYP6 family of genes, specific to Arthropods, and commonly associated with pesticide resistance in *Drosophila* and mosquitoes (Edi et al., 2014; Polinski et al., 2021).

The Lophotrochozoans have a diverse repertoire of CD elements. The CD has been described in rotifers, molluscs, annelids and other phyla; however very few studies have documented the CD as integrated networks. The first global overview of CD in this group was performed by Zhang and co-workers in 2012 on the genome of oyster (*Crassostrea gigas*) (Wang et al., 2012). In that study, authors categorized and identified few CD families, including CYPs, MAO, GPX, among other. Interestingly, they found expansions in some specific CYP families. Similarly, in rotifers several authors have catalogued CD genes and evaluated the impact of xenobiotic compounds, such as benzo[a]pyrene and other compounds, on specific gene families such as CYPs, SULTs or GSTs. Although analyses in rotifers have been exhaustive, the authors did not report or analyse CD as a network or explore the interplay between CD components (Kim et al., 2013; Won et al., 2017). Finally, and in fragmentary manner, other species of lophotrochozoans, such as, *Scrobicularia plana*, *Corbicula fluminea*, *Cappitella teleta*, *Platynereis dumerilli*, *Hediste diversicolor*, or species of *Mytilus*, have been assessed for CD components, mainly in response to xenobiotic exposure (Bonnafé et al., 2015; Catalano et al., 2012; Cruzeiro et al., 2016; Curpan et al., 2022; Dejong & Wilson, 2014; García-Alonso et al., 2011).

1.4. Nuclear Receptors and Flavin-Containing Monooxygenases

1.4.1. Nuclear Receptors in the Chemical Defensome

NRs are the most abundant family of ligand-activated transcription factors in the Metazoa kingdom and have been described as essential for the regulation of several endogenous processes such as homeostasis, development, inflammation and control of circadian rhythms (Chawta et al., 2001; Mangelsdorf & Evans, 1995). Typically, the NR activity is controlled by the binding of several ligands, including hormones, metabolites, and both xeno- and endobiotics. Regarding xenobiotics, NRs act as sensors, activating response mechanism if a xenobiotic is detected (McDonnell and Norris 2002). After transcription factor binding, the NR-mediated transcription is initiated and the regulation of gene expression begins (Chawta et al., 2001). In terms of structure, NRs usually possess a modular architecture, characterized by three regions (N-terminal; Hinge and C-terminal) and two domains (the C-terminal ligand-binding domain (LBD) and the DNA-binding domain (DBD)). Among these structures, the DBD sequence has been described as the most conserved across species, while other structures, such as the LBD, are more variable. Several authors have associated the level of conservation of each component with its function. While the LBD is usually responsible for ligand binding and recognition, the DBD binds the conserved HRE of the target gene. Mechanistically, the NRs are typically inactivated in cytosol under the control of heat shock protein (e.g. HSP90) or in the nucleus under the influence of corepressors such as nuclear receptor corepressor (N-CoR) and silencing mediator of retinoid and thyroid receptors (Lavinsky et al., 1998). When the ligands are detected, corepressors are displaced and ligands attach to the LBD domain of NRs, leading to a conformational change and exposure of the nuclear localization signals (in the case of cytoplasm receptors). Additionally, the conformational changes promote attachment to other receptors to form a dimer. During dimerization, NR can adopt and act in three different configurations; monomers (no addition of other receptors; recognizing half-sites of HREs), homodimer (attachment to and identical receptor), heterodimer (pairing with a different receptor). Subsequently, the NR is translocated into the nucleus (if it is not already located there), binds with the HREs, and recruit co-regulatory proteins, thereby activating or repressing gene expression of the target gene (Francis et al., 2003; Germain et al., 2003; Gronemeyer et al., 2004; Mangelsdorf & Evans, 1995).

In *Homo sapiens*, there are 49 NRs described and subdivided into eight subfamilies, NR1-8, based on their conserved domains, ligand activity and evolutionary analyses (Bookout et al., 2006; Francis et al., 2003; Germain et al., 2003; Robinson-

Rechavi et al., 2001). Interestingly, most NR families display identical modular structures, three regions and two domains, DBD and LBD. Curiously, NRs members of subfamily 0 and 7 have different configurations. While the members of the NR0 subfamily (non-canonical NRs) display only one domain (DBD or LBD), members of the NR7 subfamily, can possess two DBDs (Germain et al., 2003; Gronemeyer et al., 2004; Huang et al., 2015; Kaur et al., 2015; Wu et al., 2007).

Within the NRs repertoire we find different levels of substrate specificity. While some NRs display broad action and recognize a wide range of chemical structures (e.g. PXR, the so-called master regulator of xenobiotic metabolism) others act locally in recognition of specific ligands. Among NRs gene families, NR1 is the most specialized subfamily for regulating ligand-dependent and xenobiotic-sensing DMEs (Francis et al., 2003; Xu et al., 2005). This sub-family comprises several master regulators such as PXR (NR1I2) and CAR (NR1I3) that regulate gene expression of several CYPs, UGTs, SULTs and ABC transporters. The CAR receptor has been strongly associated with the induction of CYP2B subfamily members by phenobarbital, via a phenobarbital-response element module (PBREM) (L. Chen et al., 2018; Francis et al., 2003; Xu et al., 2005). PXR can also induce several members of CYP2B subfamily; however, it acts mainly on CYP3A members. PXR is a promiscuous receptor that combines characteristics of both steroidal and non-steroidal nuclear receptors and encompasses several mechanisms of action. Usually PXR acts in conjugation (as a heterodimer) with another nuclear receptor from subfamily 2, the retinoid X receptor (RXRs – NR2B1,2,3) (Kliwer et al., 1998, 2002; Wang et al., 2003). The RXR receptor is a master co-regulator essential in the formation of heterodimers with several others NR such as CAR or PXR, PPAR, LXR, FXR. Importantly, RXR is usually silent, being the heterodimer activated by ligands of its partner NR (Figure 1.5) (Evans & Mangelsdorf, 2014; Mangelsdorf & Evans, 1995).

Other NR1 receptors, such as the retinoic acid receptor (RARs – NR1B1,2,3) the vitamin D receptor (VDR – NR1I1), FXR (NR1H4), LXRs (NR1H2,3), and PPARs (NR1C1,2,3), are also vital in regulation of DMEs, however, they tend to be more selective with respect to ligands or primarily act in endogenous processes (Burris et al., 2013; Carlberg et al., 1993; Haussler et al., 1998; Peet et al., 1998; H. Wang et al., 1999; Xu et al., 2005). Within the NR3 and NR4 subfamilies, some members are directly and indirectly engaged in the regulation of DMEs. In the NR3 subfamily, the glucocorticoid receptor (GR – NR3C1) responds to corticosteroids, the estrogen receptor (ER α – NR3A1) regulates CYPs and conjugating enzymes in estrogen-sensitive tissues, and the progesterone receptor (PR – NR3C3) is involved in the regulation of efflux transporters (Kadmiel & Cidlowski, 2013; McDonnell & Norris, 2002; Mohammed et al., 2015). By

contrast, the NR4 members associated with DMEs are less frequent and appear to have an indirect impact, such as Nur77 – NR4A1 or Nor1 – NR4A3 (L. Zhang et al., 2018; Zhao & Bruemmer, 2010).

1.4.1. Flavin-Containing Monooxygenases in the Chemical Defensome

Flavin-containing monooxygenases belong to the Group B of flavin/NADPH dependent enzymes in the Baeyer—Villiger monooxygenases (BV) class and are present in all domains (Phillips & Shephard, 2008). FMOs have the ability to catalyse NADPH-dependent oxidation reactions of a wide range of xenobiotics, especially those containing nitrogen-, sulfur- or phosphorus heteroatoms (Cashman, 1995; Cashman & Zhang, 2006; Poulsen & Ziegler, 1995; Ziegler, 1988, 2002). These enzymes are monomeric proteins composed by single polypeptide chains containing a flavin adenine dinucleic (FAD)-binding domain, and act mainly in phase I of drug metabolism in complement to CYP enzymes. Despite the ability of CYP to process heteroatom-containing compounds, the process frequently produces products with increased toxicity. In contrast, processing via FMOs reduce the potential toxicity of products by producing polar, oxygenated metabolites (Cashman, 1995). Similar to CYPs, FMOs are NADPH- and oxygen-dependent. Importantly, and unlike CYPs, FMOs do not require a cytochrome P454 co-protein to be in a catalytically active state. Instead, the FAD domain is automatically reduced to FADH₂ if NADPH is available. In the presence of oxygen, FADH₂ reacts immediately with O₂ forming a C4a-hydroperoxyflavin intermediate (FAD-OOH). This represents the ready-state form of FMO enzyme and is usually the predominant state inside the cell. When FAD-OOH comes into contact with the nucleophile, it releases one oxygen atom to the compound, producing a product ready for excretion. At the end of the process, this reaction releases water and regenerates the oxidase FAD, which, in presence of NADPH and O₂ will automatically restart the cycle. This simple process, combined with the selective and narrow substrate range, makes this enzyme stable and specialist in producing consistent products with lower level of toxicity (Krueger & Williams, 2005). Importantly, the majority of FMOs are not xenobiotic-inducible via NRs like CYP members. Rather, FMOs are constitutively expressed in cells, suggesting a regulation by the developmental and physiological state (Cashman, 1995; Phillips & Shephard, 2008).

The FMO portfolio in *Homo sapiens* comprises five functional genes (FMO1-FMO5) and six non-functional pseudogenes (FMO6; FMO7-11P). These genes usually share over 40% sequence identity and display numerous endogenous and exogenous roles (Cashman, 1995; Phillips & Shephard, 2019). Among the FMO members, FMO1

and FMO3 are major players in drug metabolism of xenobiotics. While FMO1 has a broad substrate range, FMO3 tends to be more selective. Frequently both enzymes share the same substrates (e.g. benzydamine, itopride or tamoxifen); however, in some cases they act as substrate-specific enzymes (e.g. chlorpromazine for FMO1; procainamide for FMO3) (Cashman, 2005; Phillips & Shephard, 2017). The dietary-derived compound trimethylamine (TMA) is an example of a substrate specificity in FMO3. Usually, the detoxification of TMA leads to the formation of trimethylamine *N*-oxide (TMAO) (Cashman et al. 2005). When the FMO3 is mutated, the organism is unable to process TMA, leading to the genetic disorder trimethylaminuria, also known as fish odor syndrome (Dolphin et al., 1997; Lang et al., 1998). In general, human FMO3 has the ability to oxygenate primary, secondary and tertiary amines (Cashman et al. 2005). The FMO1 share same capacity but appears to be more efficient in tertiary amines (Phillips & Shephard, 2017). Interestingly, FMO1 can also metabolize TMA, although with significantly lower efficiency compared with FMO3 (Veeravalli et al., 2018). As mentioned above, FMO regulation is highly tissue-specific and development-dependent. In human fetal stage, FMO1 is the predominant liver FMO, however after birth, FMO1 ceases expression in liver and becomes predominantly expressed in kidney. Conversely, before birth, FMO3 is silent in liver, but after birth, FMO3 is switched on, becoming one of the most expressed FMOs in the human liver (Cashman, 2002, 2005; Koukouritaki et al., 2002; Phillips & Shephard, 2017). Similarly, to FMO3, also FMO5 plays an important role in liver and often exhibits equal or greater expression levels in relation to FMO3. Additionally, FMO5 shows high expression in other tissues such stomach, pancreas and small intestine (Hernandez et al., 2004; Phillips & Shephard, 2017). In small intestine, FMO5 is predominating over other FMOs. Another important characteristic of human FMO5 is the ability to be induced by drugs such rifampicin, hyperforin and synthetic progestin — a unique behaviour in FMOs of humans. Regarding substrates, FMO5 enzymes are selective, showing capacity to oxygenate compounds such as nomifensine, ranitidine or phospho-sulindac amines (Phillips & Shephard, 2017). Recent studies have also identified FMO5 as capable of performing Baeyer-Villiger (BV) reactions, with unique capabilities within FMO family to catalyse carbonyl compounds (Fiorentini et al., 2016; George Lai et al., 2011). These findings have prompted several authors to explore additional functions of FMO5. As a result, in 2023, Phillips and co-workers identified FMO5 involvement in several other pathways such NRF2, PPARA and PPARG (Phillips et al., 2023).

The FMO2 is one of the least-studied gene of FMO family. In most cases, this gene encodes a truncated, non-functional protein in humans, which led the scientific

community to disregard this gene in comparison with other FMOS (Cashman, 2005; Dolphin et al., 1998). Nevertheless, several studies have found FMO2 to be important, mainly in the development of population-specific drugs (Krueger et al., 2005). In terms of gene expression and substrates, FMO2 is primarily expressed in lung, and can metabolize alkylamines. So far, few substrates have been identified for FMO2, including ethionamide, methimazole and thioacetazone (Krueger et al., 2005; Phillips & Shephard, 2017). Additionally, it has also been shown that FMO2 can produce toxic sulfenic acids from thiourea compounds, which may be related to the natural inactivation of the gene in most human populations (Henderson et al., 2004). Like FMO2, FMO4 has been poorly studied. This enzyme is frequently associated to non-xenobiotic mechanisms. General FMO studies have repeatedly reported low FMO4 expression in several tissues (Bracquemond et al., 2025; Cashman & Zhang, 2006; Phillips & Shephard, 2017; Veeravalli et al., 2018). Despite this, perceptions of FMO4 appear to be changing: in 2025, Bracquemond and co-workers identified FMO4 as an important player in lung cancer, with vital functions in the controlling ferroptosis during cancer development (Bracquemond et al., 2025).

Evolutionarily, FMOs is an ancient group of enzymes, present in all domains of life from bacteria to humans (Schlenk, 1998). In vertebrates and mammals FMOS are relatively well characterized in terms of both the number the FMO members and their main functions (Bailleul et al., 2021, 2023; Krueger & Williams, 2005; Nicoll et al., 2019). Regarding metazoan species, several studies have attempted to established the evolutionary pattern of FMOs; however, the large number of gene duplications as well gene losses has made this process complex (Mascotti et al., 2015; Nicoll & Mascotti, 2023). In an effort to understand the representation and presence of FMO members across eukaryotes, Nicoll and Mascotti in 2023 created a phylogenetic framework to cluster and categories FMO into four groups, initially created by Paul and co-workers in 2021, across eukaryotes (Nicoll & Mascotti, 2023; Paul et al., 2021). In this study authors, found the absence of FMO in several metazoan phyla and postulated that presence or absence of these enzymes are directly correlated with their importance in metabolism (Nicoll & Mascotti, 2023). Moreover, they also reconfirmed the paraphyletic origin of Metazoa FMOs with YUCCA-type flavin monooxygenases (YUCCAS – not present in Metazoa) group first established in 2015 and 2016 by Mascotti and colleagues (Mascotti et al., 2015, 2016).

1.5. Model Species

The presented thesis encompassed a set of datasets and research developed across several species of Metazoans. These include an annelid worm, *Hediste diversicolor* (ragworm), the Echinodermata *Paracentrotus lividus* (purple sea urchin), and vertebrates such as the teleost *Scomber colias* (atlantic chub mackerel) and various species of Chimaeriformes, with special focus on *Hydrolagus colliei* (spotted ratfish). All these species were selected due to their economic and scientific relevance and with the purpose of filling phylogenetic and evolutionary gaps. Although, the majority of chemical defensome elements and genes were sourced from public available databases such as NCBI and/or Ensemble, the datasets produced and resulting gene catalogues were crucial to explore an evolutionary perspective of CD.

This sub-chapter details the species analysed in this thesis, along with their significance in this context.

1.5.1. *Scomber colias*

The Atlantic chub mackerel, *Scomber colias* (Gmelin, 1789), is a small pelagic fish widely distributed across the Atlantic Ocean and Mediterranean Sea (Figure 1.6) (Velasco et al., 2011). This species, together with other species of Scombridae family, including bonitos and tunas, has great importance in coastal pelagic, oceanic commercial and sports fisheries (Collette). For these reasons, several researchers have sought to understand multiple aspects of the biology and fisheries of this species (Correia et al., 2021; Gamito et al., 2016; Martins et al., 2013).



Figure 1.6 – Photograph of Atlantic chub mackerel, *Scomber colias* (retrieved from (Machado et al., 2022).)

From an ecological perspective, *S. colias* is essential to the trophic chain. It feeds of zooplankton and small pelagic fishes, while also serving as food source for several large predators, including larger fish and marine mammals (Velasco et al., 2011). In this context, *S. colias* acts as key intermediary between primary producers and top predators.

In terms of applications and importance, *S. colias* is particularly relevant in fisheries. Its widespread availability and nutritional profile combined fishing restrictions imposed to other species (e.g. *Sardina pilchardus*), has leveraged the *S. colias* captures and landings over the last years – in some cases doubling annually (Correia et al., 2021; Ferreira et al., 2020; Karakoltsidis et al., 1995; Martins et al., 2013).

Moreover, due to increasing interest *S. colias* and its natural abundance, several groups have investigated its potential to bioaccumulation and biomagnification of pollutants, including heavy metals and trace elements. These processes occur through direct contact or ingestion. Such elements are often released into the aquatic systems as a result of human activities such as industry, agriculture, mining, and waste disposal (Lozano-Bilbao et al., 2024). Another significant source of contamination is the ingestion of microplastics. These particles frequently accumulate contaminants, that are ingested by marine species. The contaminants are then biomagnified through the trophic chain, ultimately reaching the top marine predators or humans, in case of fish consumption (Herrera et al., 2019).

In this thesis, *S. colias* was selected as a representative species to scrutinize several elements of the chemical defensome in teleosts. Its economic relevance in Portugal, along with its abundance, which facilitated further *in situ* assays, were the primary reasons for its inclusion in this dissertation. Additionally, several studies regarding biology, physiology and populations dynamics of *S. colias* have been conducted in recent years, making it a strong candidate for exploring the impact of man-made chemicals on the CD gene repertoire of fishes. Furthermore, I had previously undertaken molecular work involving this species, prior to the beginning of this thesis, which facilitated the identification of several chemical defensome genes in *S. colias*. Nevertheless, despite the availability of some molecular datasets (e.g., liver transcriptome dataset (Machado et al., 2018)), several elements of CD remained unidentified. This gap led me to develop several bioinformatic pipelines, sequence and assemble the first genome of the species, and identify the nuclear receptor repertoire of genes.

1.5.2. *Hydrolagus coliei*

Chondrichthyes (sharks, skates, rays and chimaeras) represent one of the most iconic classes of organisms in aquatic environment. This class of top predators encompasses more than 1088 species, catalogued in 16 orders, 61 families and 199 genera (Nelson, 2016; Weigmann, 2016). Highly diverse and exhibiting peculiar phenotypes, Chondrichthyes are distributed globally and inhabit in almost all oceans and

seas (Nelson, 2016; Weigmann, 2016). One of the most intriguing subclasses within Chondrichthyes, Holocephali, comprises at least 37 species distributed in three families (*Callorhinchidae*, *Rhinochimaeridae*, and *Chimaeridae*). These organisms are considered “living fossils” having evolved more than 300 million years ago, and are characterised by unique morphological, anatomical, physiological, behavioural characters (Barnett et al., 2009; Didier et al., 2012; Smith et al., 2004).

The *Hydrolagus colliei*, commonly known as spotted ratfish (Lay and Bennett, 1839), is a member of chimaeridae family, widely distributed from the southeast Alaska to the tip of Baja and in the northern Gulf of California (Figure 1.7) (Barnett et al., 2009; Smith et al., 2004). Unlike other Holocephalans, which are typically found in continental shelves and slopes and/or Deepsea cold waters, only *Hydrolagus colliei* and organisms of Callorhinchidae family has been observed from shallow water to depths of up to 1029 meters (0–1029 m) (Didier 2004; Inoue et al. 2010; Lisney 2010; Weigmann 2016; Berio et al. 2024).



Figure 1.7 – Photograph of spotted ratfish, *Hidrolagus colliei* (provided by Martin Flajnik).

The spotted ratfish predominantly feeds of hard preys (durophagy) such as gastropod molluscs, typically found over rocky, mud, and gravel bottoms (Huber et al., 2008; Johnson & Horton, 1972; Mathews, 1975). Notwithstanding, it is also considered an opportunistic feeder, complementing its diet with food sources like echinoderms, annelids, and small fishes, or by chance intracohort cannibalism (Allen & Horn, 2006; Johnson & Horton, 1972). Ecologically, this species is considered essential not only due its high relative abundance, ranking among the ten most abundant ground fishes in U.S. West Coast, but also because serve as both predator and prey within the trophic chain (Allen & Horn, 2006; Barnett et al., 2012; Keller et al., 2006). In addition to ecological

relevance, this cartilaginous fish holds indirect relevance for fisheries, once is an important food source for commercial valuable species such Pacific halibut (*Hippoglossus stenolepis*) or Spiny dogfish (*Squalus acanthias*) (Johnson & Horton, 1972) Moreover, due its wide distribution, *H. colliei* has been employed as bioindicator and biomonitor for highly persistent and available chemicals contaminants, with strong potential for bioaccumulation and biomagnification, including heavy metals such as copper, zinc, lead, mercury, as well as metalloids, like selenium and arsenic (Bergés-Tiznado et al. 2021; Tiznado-Salazar et al. 2023) .

During this dissertation, *H. colliei* species was incorporated as representative model of Holocephalans to explore CD in this lineage. The selection of *H. colliei* was relevant due its phylogenetic position as one of the most ancient jawed vertebrates. Thus, this species was essential as an intermediate model to highlight chemical defensome evolution across vertebrates, while clarified the phylogenetic relationships in the early branches of the vertebrate tree of life. Moreover, this integration followed previously studies wherein, I and other collaborators, identified and catalogued several genes related with chemical defensome and immune system, namely nuclear receptors and nuclear receptors and major histocompatibility complex (MHC) genes, in Chimaeriformes species (Almeida et al., 2021; Fonseca et al., 2020). Yet, and despite, the abundance of literature and genomic datasets available for other Chondrichthyans, particularly sharks and rays, they genomic datasets are notably complex. Usually, these organisms exhibit high genome size variations (2,86 and 17,05 pg), considerable levels of heterozygosity and large genomic variations among the populations. Moreover, the atypical gene evolution of these taxa, remains a challenge in terms of gene mining and large-scale evolutionary analyses (Hara et al. 2018; Torralba Sáez et al. 2024).

In contrast, the genomes of chimaeras are relatively small and homogenous, with a uniform number and categories of genes across species, thereby being more suitable to genomic analyses (Fonseca et al. 2020; Almeida et al. 2021). Although several datasets are available for Chimaeriformes, we found limited taxonomic representation and structural genome completeness in the available genomic resources, leading to a large number of CD elements unidentified. To address this gap, I produced a comprehensive new dataset, encompassing the entire process from the sequencing till the assembly and annotation, using cutting edge genomic technologies.

1.5.3. *Paracentrotus lividus*

Paracentrotus lividus (Figure 1.8), commonly known as purple sea urchin, is a widely distributed echinoid in Europe, occurring from the Mediterranean Sea to the eastern Atlantic coast, and from Scotland to Southern Morocco, Canary and Madeira Islands (Boudouresque and Verlaque, 2001). This species inhabits in coastal ecosystems and plays a crucial role in benthic communities, once it is one of the main responsible for maintaining of equilibrium of the algae and seagrass densities. On the other hand, also is an important source of food for other species of the tropic chain (Boudouresque and Verlaque, 2001).



Figure 1.8 – Photograph of common purple sea urchin, *Paracentrotus lividus* (provided by Sergio Fernández-Boo).

Despite its ecological importance, *P. lividus* populations has suffered a continual decline in last years, partly due to uncontrolled and intense harvesting (Bertocci et al. 2012; Bertocci et al. 2018). One of the main reasons of this exploitation lies in the economic importance of this species. Widely used in gastronomy, this species is considered as a delicacy in several cultures. Moreover, the Anthropocene has further contributed to this populational declining throughout increased pollution and/or habitat destruction (Yeruham et al. 2015; Ohgaki et al. 2019; Steneck and Pauly 2019; Milito et al. 2020). In an effort to mitigate these impacts, several research groups have sought to established the entire life cycle of *P. lividus* in aquaculture. However, despite these efforts, limited knowledge regarding optimal diet and maintenance schedules in aquaculture systems remain a bottleneck (Brown & Eddy, 2015; Liu & Chang, 2015).

From a different perspective, the *P. lividus* has been increasingly utilised in the fields of ecotoxicology and embryology. Owing to its wide geographical distribution, easy of sampling, well-known biology, as well as, capacity to bioaccumulate high levels of pollutants, *P. lividus*, along with other sea urchins, has been used as bioindicator (Milito et al. 2020). Indeed, the highly sensitivity of this species made it suitable to analyse, assess and monitor the accumulation and presence of several endocrine disruptors, pesticides, PHAs and emerging contaminants in environment (Gharred et al. 2022).

In this dissertation, *P. lividus* was among the first model organisms to be considered. This echinoderm was selected not only to its economic relevance in Portugal, but also because of the abundance of public available literature and previously tested experimental conditions, particularly within the field of ecotoxicology. Moreover, *P. lividus* is phylogenetically related with *Strongylocentrotus purpuratus* species, the first organism in which the complete set of chemical defensome genes was scrutinized by Goldstone and colleagues in 2006 (Goldstone et al. 2006). Subsequently, in 2012, Marrone and co-workers identified a partial set of chemical defensome genes in *P. lividus*. Collectively, these factors supported to the inclusion of this species in the present thesis. Despite the scientific interest on this species, I found that molecular resources were limited. Therefore, through a collaborative effort, I produced and analysed the first transcriptomic datasets of gonads and immune cells in this organism, with the aim of complement the public data already available.

1.5.4. *Hediste diversicolor*

The ragworm *Hediste diversicolor* (O.F. Müller, 1776), is a polychaete annelid belonging to the Nereididae family and genus *Hediste* (old *Nereis* genus) (Figure 1.9) (Einfeldt et al., 2014; Gomes-dos-Santos, Hagemann, et al., 2021). It is widely distributed along the Atlantic coast of Europe and North America, as well as Mediterranean and Pacific Ocean, primary inhabiting in estuarine and lagoon environments (Miloslavich et al., 2010; Silva et al., 2020). Due to its broad distribution and high ecological, environmental, and commercial value, this species has become a priority for scientific research (Pombo et al. 2020).



Figure 1.9 – Photograph of common ragworm, *Hediste diversicolor* (retrieved from (Machado et al., 2024)).

Ecologically, this benthic species plays a crucial role in enhancing nutrient recycling, sediment oxygenation and micro-environment creation, through a process known as bioturbation. During this process, and while feeding, *H. diversicolor* construct burrows, turn over sediments, and improves the sediment-water interface (Gillet et al. 2018). Moreover, *H. diversicolor* presents an outstanding tolerance to wide fluctuation of salinities, temperatures, and oxygen levels (Buikema and Cairns 1980; Kristensen 1983; Ozoh, 1992). These physiological proprieties render it an excellent candidate for ecotoxicological studies of several pollutants, including pharmaceuticals, pesticides, polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), and emerging contaminants (Pires et al., 2016; Silva et al., 2020). Additionally, *H. diversicolor* has been extensively employed as bioindicator to toxicity assessment and environmental biomonitoring of pollution (Nunes et al. 2016; Pires et al. 2016). Economically, this species is primarily used as bait in recreational fishing and aquaculture. In recreational fishing, a major challenge with harvesting *H. diversicolor* is the overlap between fishing season and the breeding season (Scaps, 2002). In aquaculture, the species is particularly valuable when integrated into established systems such as recirculation aquaculture system (RAS) or in integrated multitrophic aquaculture (IMTA) (Pombo et al. 2020). Within these systems, *H. diversicolor* shows a great potential to integration in multiple stages. On one hand, it can be intensively produced using waste streams from land-based aquaculture while assist the process of sludge treatment (Wang et al. 2019). On the other hand, also can be used as food, due its well-balanced nutritional profile, for several commercial species integrated on these systems.

In the context of this thesis, this model organism emerged naturally due to its practical applications and economic relevance in Portugal. Moreover, the abundance of public available literature and bioassays, particularly within the field of ecotoxicology, made it an excellent candidate for cataloguing, and highlighting the chemical defensome in invertebrates. Additionally, as previously noted, the physiological proprieties make it well-suited for exploring the relationship between number of genes, evolutionary events, such as gene duplication or gene loss, and physiological phenotype.

Nevertheless, I found several limitations in terms of molecular data resources. As mentioned by Gomes-dos-Santos et. al. in 2021, despite extensive morphological and bioassays data available, there remains a scarcity of mitogenomic, genomic, transcriptomic and other molecular resources (Gomes-dos-Santos et al., 2021). Consequently, during this thesis focused in the production of the first large-scale datasets for this species.

1.6 – Objectives

1.6.1 – Global objective

The global aim of this thesis is to explore the repertoire of chemical defensome components and their variability across metazoan species. To achieve this, I employed a combination of transcriptomic and genomic next-generation sequencing techniques, integrated with bioinformatic approaches, to: (1) fill phylogenetic gaps using species with economic and evolutionary relevance; (2) generate high-quality molecular datasets to provide extensive resources with practical applications in emerging fields such as population genomics, comparative genomics and transcriptomics, phylo-transcriptomics, and phylogenomic; and, (3) develop a framework and test case using the chemical defensome in the context of novel omic approaches.

1.6.2 – Specific objectives

1. To identify lineages with gaps in molecular resources across each phylum within Metazoa kingdom.
2. To pinpoint species with high commercial or phylogenetic relevance that currently lack molecular resources, but are relevant for assessing the chemical defensome in Metazoa species.
3. To develop, provide and optimize bioinformatic pipelines for the analysis, exploration, and production of high-quality molecular datasets in selected species. To this end, I have produced bioinformatic pipelines to the fields of genomics, transcriptomics and comparative genomics/transcriptomics.
4. To produce the first high-quality reference genome for the teleost Atlantic chub mackerel, *Scomber colias*. Beyond its economic importance in Portugal, and relevance in fisheries management, this dataset was essential to explore the chemical defensome elements in Scombriformes.
5. To generate the first whole-genome dataset for a neglected Chondrichthyan species, *Hydrolagus colliei* (*H. colliei*). Species from Chimaeriformes order are phylogenetically early-branching and are key for the understanding of the evolutionary journey of chemical defensome components, such as NRs and FMOs, in early vertebrate ancestry.
6. To generate and provide the first gonadal whole-transcriptome dataset for the echinoderm *Paracentrotus lividus* (*P. lividus*), with the aim of identify differential expressed genes between males and females, and correlating gene expression profiles with the chemical defensome genes expressed in gonadal tissue. This

dataset was instrumental in identifying genes with potential application in aquaculture.

7. To generate and provide the first immune cell whole-transcriptome dataset for the purple sea urchin, *P. lividus*, aiming to identify immune gene portfolio genes in both males and females. This study provided a broad overview of the immune genes repertoire in the species and enabled the correlation of the immune-related genes with the chemical defensome portfolio expressed in immune cells. The datasets were essential to characterizing the immune gene portfolio in purple sea urchin, leveraging its aquaculture potential.
8. To generate the first whole-transcriptome dataset for the annelid species *Hediste diversicolor*, for further characterizing of the chemical defensome.

1.7. Dissertation outline

This dissertation is presented as a coherent and well-defined structure collection of scientific articles, organised into chapters, which have been accepted for publication or are in preparation.

Chapter 1 provides the conceptual framework that contextualises the subsequent chapters.

Chapter 2 describes the application of a set of general genomic and transcriptomic tools, which led to the first high-quality genome assembly for the teleost *Scomber colias*. This chapter was instrumental in establishing the genomic and transcriptomic pipelines that support the discoveries presented in some of the remaining chapters. Additionally, it provided the first overview of the chemical defensome in *S. colias*. This analysis enabled the identification and exploration of the FMO and NR gene families, as well as the evaluation of the presence-absence of key CD genes, such as the xenobiotic sensor PXR (Machado, et al., 2022).

In **Chapter 3**, we took the implementation of global genomic-transcriptomic tools further by adding an extra step to the genome strategies implemented in Chapter 2. This involved applying chromatin conformation capture (3C) methods (Hi-C). This enabled the creation of the first chromosome-level genomic resource for the chimaeriform species *Hydrolagus collie*. Further analyses expanded the characterisation of CD in chondrichthyans by identifying the repertoire of FMOs in this species and in several others. Additionally, the presence of several NRs was confirmed in *H. colliei*.

Chapter 4.1 describe our use of core transcriptomic approaches to produce and characterise the male and female gonadal transcriptomes of *P. lividus*. This study adopted a species-specific approach involving the construction and implementation of *de novo* transcriptome references alongside gene expression and enrichment analyses. This chapter identified several sex-related genes and uncovered a portfolio of genes involved in lipid biosynthesis. In the context of the chemical defensome, it revealed and confirmed the presence of several members, particularly NRs and FMOs, which are involved in coordinating lipid homeostasis and xenobiotic metabolism (Machado, et al., 2022).

In **Chapter 4.2**, we conducted transcriptomic analyses of *P. lividus* immune cells after they were exposed to biological stressors. This study characterised the immune cell transcriptomes of both males and female individuals, identifying immune signatures associated with biological challenge. We applied a range of transcriptomic tools here,

including reference construction and annotation, gene expression analysis, enrichment testing and functional interpretation. Several CD elements were identified within the immune signatures, including NRs and FMOs. This indicates that *P. lividus* orchestrates a chemical-immune defence in response to biological stress. Furthermore, we identified CD elements that potentially contribute to the mitigating oxidative stress, regulating inflammation, and maintaining homeostasis. Overall, this chapter demonstrated the complementarity of physiological systems under stress biological stress (Fernández-Boo et al., 2024).

In **Chapter 5**, we extended the transcriptomic frameworks previously implemented to another metazoan species, the annelid *Hediste diversicolor*, to further analyse the identification of FMOs and NRS. We produced the first whole-transcriptome assembly for this species here, performed functional annotation, and prepared all datasets for future exploration (Machado et al., 2024).

2. Chapter 2 – A genome assembly of the Atlantic chub mackerel (*Scomber colias*): a valuable teleost fishing resource.

André M. Machado, André Gomes-dos-Santos, Miguel Fonseca, Rute R. da Fonseca, Ana Veríssimo, Mónica Felício, Ricardo Capela, Nélson Alves, Miguel Santos, Filipe Salvador-Caramelo, Marcos Domingues, Raquel Ruivo, Elsa Froufe, L. Filipe C. Castro

Gigabyte

(DOI: [10.46471/gigabyte.40](https://doi.org/10.46471/gigabyte.40))

A genome assembly of the Atlantic chub mackerel (*Scomber colias*): a valuable teleost fishing resource

André M. Machado^{1,2}, André Gomes-dos-Santos^{1,2}, Miguel M. Fonseca¹, Rute R. da Fonseca^{3,4}, Ana Veríssimo^{5,6}, Mónica Felício⁷, Ricardo Capela^{1,2}, Nelson Alves^{1,2}, Miguel Santos^{1,2}, Filipe Salvador-Caramelo^{1,2}, Marcos Domingues¹, Raquel Ruivo¹, Elsa Froufe¹ and L. Filipe C. Castro^{1,2,*}

- 1 CIIMAR – Interdisciplinary Centre of Marine and Environmental Research, U. Porto – University of Porto, Porto, Portugal
- 2 Department of Biology, Faculty of Sciences, U. Porto - University of Porto, Portugal
- 3 Center for Global Mountain Biodiversity, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark
- 4 Center for Macroecology, Evolution, and Climate, GLOBE Institute, University of Copenhagen, Denmark
- 5 CIBIO - Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO - Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal
- 6 BIOPOLIS - Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal
- 7 Portuguese Institute for the Sea and Atmosphere, I.P. (IPMA), Portugal

ABSTRACT

The Atlantic chub mackerel, *Scomber colias* (Gmelin, 1789), is a medium-sized pelagic fish with substantial importance in the fisheries of the Atlantic Ocean and the Mediterranean Sea. Over the past decade, this species has gained special relevance, being one of the main targets of pelagic fisheries in the NE Atlantic. Here, we sequenced and annotated the first high-quality draft genome assembly of *S. colias*, produced with PacBio HiFi long reads and Illumina paired-end short reads. The estimated genome size is 814 Mbp, distributed into 2,028 scaffolds and 2,093 contigs with an N50 length of 4.19 and 3.34 Mbp, respectively. We annotated 27,675 protein-coding genes and the BUSCO analyses indicated high completeness, with 97.3% of the single-copy orthologs in the Actinopterygii library profile. The present genome assembly represents a valuable resource to address the biology and management of this relevant fishery. Finally, this genome assembly ranks fourth in high-quality genome assemblies within the order Scombriformes and first in the genus *Scomber*.

Submitted: 22 November 2021
Accepted: 09 February 2022
Published: 14 February 2022

* Corresponding author. E-mail:
lfilipecastro@gmail.com;
filipe.castro@ciimar.up.pt

Published by GigaScience Press.

Preprint submitted at <https://doi.org/10.1101/2021.11.19.468211>

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Gigabyte, 2022, 1–21

Subjects Genetics and Genomics, Evolutionary Biology, Marine Biology

DATA DESCRIPTION

Background and context

The family Scombridae is divided into 2 subfamilies (Gasterochismatinae and Scombrinae), with 15 genera and around 49 described species, comprising mackerels, bonitos, and tunas [1]. The representative genus of the Scombridae, i.e., *Scomber*, includes 4 species: *S. scombrus*, *S. japonicus*, *S. australasicus*, and *S. colias*. The Atlantic chub mackerel, *Scomber colias* (Gmelin, 1789) (NCBI:txid338315, FishBase ID:54736), is a small coastal pelagic fish that is distributed widely, being found in the Atlantic Ocean from the Bay of



Figure 1. Photograph of Atlantic chub mackerel, *Scomber colias*. The specimen was caught in 2020 and used for PacBio HiFi genome assembly.

Biscay to South Africa (including the Canary, Madeira, Azores, and Saint Helena Islands), and in the Mediterranean Sea (Figure 1) [2]. *Scomber colias* is usually found at depths of up to 300 m and occupies a key position in the trophic web. This species acts as a link between primary producers and top predators, since it feeds mainly on zooplankton and small pelagic fish, and is an essential element of the diet of larger pelagic fish (e.g., tuna, swordfish, and sharks) and marine mammals (e.g., dolphins and seals) [3]. Besides its ecological importance, *S. colias* also supports important commercial fisheries for several countries across its distribution range, being an important component in the diet of local populations [1, 4]. This is probably related to its nutritional value, as this mackerel is a valued source of important fatty acids for human nutrition, particularly docosahexaenoic acid (DHA), an omega-3 fatty acid [5, 6]. Additionally, *S. colias* is used as bait for the tuna longline and handline fisheries, and is caught in purse seine and pelagic trawl fisheries which target sardines and anchovies [7].

The availability of *S. colias* makes it a sustainable marine resource [6] and a viable alternative to the European sardine (*Sardina pilchardus*), which is under fishing restrictions due to a population decline. Curiously, fluctuations in abundance and a northwards shift in the distribution of *S. colias*, with a likely inverse relationship with sardine abundance, have been recently demonstrated [8]. Due to its ecological and economic importance, *S. colias* has been the focus of several recent studies on different aspects of its fisheries and biology [3, 8, 9]. Yet, genomic resources for the species are still limited. Presently a (liver) transcriptome [10], a mitogenome [11], and single-nucleotide polymorphism (SNP) data obtained through restriction site-associated DNA sequencing [12], have been described for the species. With the vast majority of the world's fish stocks already in collapse, and with climate change as additional pressure, information on fish genomes is becoming a pressing tool to address conservation efforts [13, 14]. Here, we report the first high-quality draft genome of *S. colias*, assembled with Illumina and Pacific Biosciences (PacBio) Single Molecule High-Fidelity (HiFi) reads. This resource provides a critical platform to uncover the species' adaptive physiological potential in a changing environment. Specifically, it will help understand the current observed populational northward shift, postulated to be part of a more general expansion of species from warmer areas [8]. Moreover, being one of the genomes with higher quality within the family Scombridae and the first within the *Scomber*

genus, this information will help to improve the conservation, management, and sustainable exploitation of this valuable fish resource as well as that of its highly valued congeners.

METHODS

Sampling and DNA extraction

Two specimens of *S. colias* were collected at 2 sampling time points. The first specimen was collected in 2017, during the “Programa Nacional de Amostragem Biológica” managed by the Instituto Português do Mar e da Atmosfera (IPMA), in North Atlantic waters (41.501944 N 8.851667 W). From this individual, 2 tissue types were collected, and were stored in 100% ethanol (muscle) or RNA *later* (liver). Liver tissue was used to produce and describe the first liver transcriptome of *S. colias* [10]. Muscle tissue was used in the present study, for genomic DNA (gDNA) extraction using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. The gDNA was then used for Illumina paired-end (PE) sequencing (described below). The second specimen was caught in 2020, near Mira, Portugal (40.5588270 N 9.4529720 W). Immediately upon harvesting, the muscle was snap frozen in liquid nitrogen. The frozen tissue was shipped to Brigham Young University DNA Sequencing Center (BYU), where gDNA with high molecular weight was extracted from 1.1 g of muscle using the QIAGEN Genomic-tip 20/G kit. The quality and concentration of the gDNA were assessed with Qubit Fluorometric system (ThermoFisher), and the fragment size was determined with a fragment analyser (Agilent Technologies, RRID:SCR_013575) before loading on the PacBio Sequel II system (PacBio Sequel II System, RRID:SCR_017990).

DNA sequencing libraries construction and sequencing

For the first DNA sample, Illumina PE library preparation and sequencing were carried out by Macrogen, Inc. (Seoul, Korea), using Illumina HiSeq X Ten platform (Illumina HiSeq X Ten, RRID:SCR_016385), with 250 bp PE configuration. For the second specimen, PacBio HiFi library preparation and sequencing were performed at BYU, following the manufacturer's recommendations [15]. The size-selected fraction had a mean read length of 15.3 kbp and was selected on the SageELF system (Sage Science, RRID:SCR_014808). The sequencing was conducted on two single-molecule, real-time (SMRT) cells using Sequel II system v.9.0, with a run time of 30 h, and 2.9 h pre-extension. The circular consensus analysis was performed in SMRT® Link v9.0 [16] under default settings (the statistics of raw data generated from each PacBio SMRT cell can be viewed in Additional File 1 [17, 18]).

Raw data quality control, clean-up, and genome size estimation

Both short- and long-read datasets were assessed by FastQC v.0.11.8 software (FastQC, RRID:SCR_014583). Trimmomatic v.0.38 software (Trimmomatic, RRID:SCR_011848) [19] was used to filter and remove low quality reads as well as the adaptors of the Illumina dataset (LEADING:5 TRAILING:5 SLIDINGWINDOW:4:20 MINLEN:50). Next, trimmed datasets were used to check the overall characteristics of the *S. colias* genome (i.e., genome size, heterozygosity, unique content), through GenomeScope 2.0 [20]. Briefly, Jellyfish v.2.2.10 software (Jellyfish, RRID:SCR_005491) [21] was used to build *k*-mer frequency distributions, and the final *k*-mer counts (*k*-mer 21, 25, 31) were submitted to the GenomeScope 2.0 online platform. On the other hand, HiFi reads were filtered in two ways (Figure 2). First, mitochondrial reads were removed by BLAST searches (BLASTN, RRID:SCR_001598) using a

prebuilt database of mitochondrial sequences (database build protocol: (1) select all complete mitogenomes present in the nucleotide database of the National Center for Biotechnology Information (NT-NCBI); (2) select by taxon (Actinopterygii; [txid:7898](#)); (3) sequence length filter 15,000–50,000 bp; (4) build a database with the makeblastdb application of NCBI-BLAST+ v.2.9.0). Second, to filter out possible sources of contamination (artefactual or biological), HiFi reads were checked by BLAST (BLASTN) against NT-NCBI. Only HiFi reads with match hits over 90% identity and query coverage of 50% in the Actinopterygii taxon (NCBI:[txid:7898](#)), or without match hits at all, were considered for further analysis (Figure 2).

Mitochondrial genome assembly

Given that 2 specimens were used for the distinct sequencing approaches, i.e., PacBio HiFi and Illumina PE, the whole mitochondrial genome (mtDNA) was assembled and characterised for both specimens. For specimen 1, trimmed Illumina PE reads were used to assemble mtDNA in GetOrganelle v.1.7.1 [22] with optimised parameters (-F animal_mt -w 121 -R 10 -k 85,95,105,115,125) (Figure 2). For specimen 2, a new pipeline was designed to produce a mtDNA assembly from the PacBio HiFi long reads (Figure 2). The PacBio HiFi mtDNA reads, previously filtered (see above), were corrected using Hifiiasm v.0.13-r308 (Hifiiasm, [RRID:SCR_021069](#)) [23] with optimised parameters (-write-ec). Since Hifiiasm is not optimised to assemble circular molecules (which are expected for mtDNA), the corrected PacBio HiFi mtDNA reads were assembled using Unicycler v.0.4.8 [24], a software package designed to assemble bacterial genomes and so optimised for circular assemblies, with default parameters. Annotation and visual representation of both mtDNA assemblies were produced using MitoZ v.2.3 [25] with optimised parameters (-genetic_code 2; -clade Chordata; -topology circular), using the PE reads for coverage plotting. Furthermore, annotations were manually validated by comparison with other mitochondrial genomes of the genus *Scomber*, available at NCBI (see Data Availability).

Nuclear genome assembly and assessment

For whole-genome assembly a combined approach, using short- and long-read assemblies, was applied (Figure 2). While long-read assemblies were mainly used to produce the primary assembly, short-read assemblies were used to scaffold and improve the contiguity of the basal assembly. In summary, short-read assemblies were performed with the W2RAP pipeline v.0.1 [26], following the authors' protocol. First the *k*-mer analyses toolkit (KAT) v.2.4.1 software (KAT, [RRID:SCR_016741](#)) [27] hist module was applied to determine the ideal *k*-mer cut-off, before W2RAP with optimised parameters (-t 30; -m 500; -min_freq 14; -d 32; -dump_all 1; -k: 144, 180, 200, 224) was used to produce 4 assemblies (Figure 2). To generate the long-read assembly, multiple software and parameters were initially tested. PacBio HiFi reads were assembled in Hifiiasm v.0.13-r308 [23] with a range of parameters (*k* = 21, 25, 31, 41, 45, 51; *l* = 0, 2) and in HiCanu v.2.1.1 [28] with optimised parameters (default). While Hifiiasm generated 2 pseudo-haplotypes per assembly, HiCanu generated 1 merged assembly. To choose the "best" assembly we applied a series of analyses, including Bandage (a bioinformatics application for navigating *de novo* assembly graphs easily) v.0.8.1 [29] and manual inspection; Benchmarking Universal Single-Copy Orthologs (BUSCO) v.5.2.2 (BUSCO, [RRID:SCR_015008](#)) [30] with Eukaryota and Actinopterygii databases was used to assess the gene completeness of the assemblies, and Quality Assessment Tool for Genome Assemblies



Figure 2. Bioinformatics workflow used to perform the genome assembly of *Scomber colias* species.

(QUAST) v.5.0.2 (QUAST, RRID:SCR_001228) [31], to check general metrics of the assemblies (Figure 2). Due to discrepancies in the length of the Hifiasm primary and alternative pseudo-haplotypes, we chose to concatenate them in a single assembly. At this point, the assembly with the highest complete BUSCO scores, highest contiguity (N50), and longest contig, was selected for further analysis. The pseudo-haplotypes were separated by *purge_dups* v.1.2.5 (*purge_dups*, RRID:SCR_021173) [32]. After the first round of purging and inspection by *k*-mer plot, produced by the KAT tool, cutoffs were manually adjusted.

To assess the influence of purge_dups in the genome, BUSCO (rate of deduplicates) and QUAST (N50 and genomic length per pseudo-haplotype) were used. Next, to improve the contiguity and quality of the assembly, short-read assemblies were used to structurally scaffold the assembly without the introduction of any new bases in the assembly, similar to the literature [33, 34] (Figure 2). The 4 short-read assemblies were inputted to the Long Interval Nucleotide K-mer Scaffoldolder (LINKS) v.1.8.7 [35], being used as long reads; using several distance values, i.e., -d 0.5, 1.5, 3, 9, 27 kb, the primary assembly was rescaffolded interactively for 5 rounds (additional parameters: -k 21 -e 0.5). Furthermore, the scaffolded genome and the long-read assemblies, initially produced by Hifiasm and HiCanu and discarded based on contiguity and completeness, were inputted to Cobbler v.0.6.1 [36] and RAILS v.1.5.1 [36] pipeline, with default parameters. This allowed gap filling of ambiguity regions (produced by short-read scaffolding), and further rescaffolding using long-read information. To evaluate the final assembly, several metrics and software were used. In addition to BUSCO and QUAST metrics, read back mapping of paired-end (PE) reads with Burrows-Wheeler Aligner (BWA) v.0.7.17-r1198 (BWA, RRID:SCR_010910) [37], long reads with Minimap2 v.2.17 (Minimap2, RRID:SCR_018550) [38] and RNA sequencing (RNA-Seq) with Hisat2 v.2.2.0 (HISAT2, RRID:SCR_015530) [39, 40], were also applied. To check consensus quality (QV) and k-mer completeness we used Merqury v.1.1 [41] (Figure 2).

Repeat masking, gene prediction, and annotation

The repetitive elements of the genome were predicted and masked by RepeatMasker v.4.0.7 (RepeatMasker, RRID:SCR_012954) [42] using homologous comparisons and *ab initio* predictions. First, the *de novo* library of repetitive elements was created with the RepeatModeler v.2.0.1 (RepeatModeler, RRID:SCR_015027) [43]. Next, the *ab initio* library, as well as the Dfam_consensus-20170127 (Dfam, RRID:SCR_021168) [44] and RepBase-20181026 (Repbase, RRID:SCR_021169) [45], were used in RepeatMasker to softmask the *S. colias* genome assembly. The genome annotation was performed with the BRAKER2 pipeline v.2.1.6 (BRAKER, RRID:SCR_018964) [46–48]. Initially, the liver RNA-Seq reads (accession number: SRR6367407 [10]) were downloaded, mapped against the *S. colias* genome assembly using Hisat2 v.2.2.0 [39, 40] with default parameters, and converted to BAM and sorted files using Samtools v.1.9 (SAMTOOLS, RRID:SCR_002105) [49]. Additionally, we collected 89 proteomes from NCBI RefSeq (RefSeq, RRID:SCR_003496) [50] and Ensembl (Ensembl, RRID:SCR_002344) [51] databases. The species and accession numbers of the proteomes used in the genome annotation of *S. colias* can be consulted in Additional File 2 [17, 18]. Of these, 82 species belong to the class Actinopterygii (32 taxonomic orders): 81 with genome assembly at chromosome level, and 1 at scaffold level. As of the date of this genome annotation, only 1 Scombriforme genome, *Thunnus orientalis*, was annotated at scaffold level. The remaining 7 proteomes were selected from other vertebrate non-teleost animal models: *Callorhynchus milii*, *Amblyraja radiata*, *Scyliorhinus canicula*, *Lepisosteus oculatus*, *Petromyzon marinus*, *Mus musculus*, and *Homo sapiens*. Next, the RNA-Seq alignment, as well as all the above-mentioned proteomes, were inputted to the BRAKER2 pipeline with optimised parameters (–etpmode; –softmasking; –UTR = of; –crf; –cores = 30). The final file of predictions (braker.gtf) was further filtered by evidence, keeping only gene predictions with RNA-Seq or protein evidence (using BRAKER2 auxiliary scripts; selectSupportedSubsets.py), then converted to .gff3 format (using the Augustus auxiliary scripts; gtf2gff.pl) and post-processed with Another Gtf/Gff Analysis Toolkit (AGAT)

v.0.6.0 [52]. The post-processing stage involved the correction of overlapping gene prediction coordinates and the removal of small or incomplete protein-coding genes (i.e., coding for <100 amino acids (aa); lacking start or stop codons). Furthermore, the proteins were extracted with AGAT and subject to functional annotation using InterProScan v.5.44.80 (InterProScan, RRID:SCR_005829) [53] and BLASTP (BLASTP, RRID:SCR_001010) searches against RefSeq [50] and UniProtKB/SwissProt (UniProtKB, RRID:SCR_004426) [54] databases. The homology searches were performed with DIAMOND v.2.0.11.149 (DIAMOND, RRID:SCR_016071) [55] with optimised parameters (-k 1, -b 10, -e 1e-5, -ultra-sensitive, -outfmt 6). Finally, the genome and the annotation datasets were integrated using JBrowse2 (JBrowse, RRID:SCR_001004) [56], a dynamic web platform for genome visualisation and analysis that allows easy and interactive exploration of provided data (<http://portugalfishomics.ciimar.up.pt/app/scombercolias/>). The FASTA file containing the genome was indexed with Samtools faidx v.1.9 [49] and added to the JBrowse component, along with the annotation file sorted with GenomeTools v.1.6.1 (GenomeTools, RRID:SCR_016120) [57], and indexed with Samtools tabix v.1.9 [58]. In addition to the JBrowse component, NCBI-BLAST+ v.2.12.0 [59] was integrated into the webpage, allowing BLAST results from the genome, mRNA, protein-coding sequences (CDS), and proteins, directly from the website.

Phylogenomics

To generate a phylogenomic analysis, the proteomes of 15 selected Actinopterygii species, including the Scombriformes species *Thunnus maccoyii* and *T. orientalis*, were downloaded from public databases. The species and accession numbers used in the phylogenomic analyses can be consulted in Additional File 2 [17, 18]. Single-copy orthologs between these 15 species and *S. colias* were retrieved from the protein datasets by constructing protein family clusters using OrthoFinder v.2.4.0 (OrthoFinder, RRID:SCR_017118) [60] with optimised parameters (-M). This resulted in a total of 392 single-copy orthologous sequences that were individually aligned using MUSCLE v.3.8.31 (MUSCLE, RRID:SCR_011812) [61] with default parameters. Each alignment was trimmed using TrimAl v.1.2 (trimAl, RRID:SCR_017334) [62] with a gap threshold of 0.5 with optimised parameters (-gt 0.5), and afterwards concatenated using FASconCAT-G [63]. Phylogenetic inferences were conducted in IQ-Tree v.1.6.12 (IQ-Tree, RRID:SCR_017254) [64] with optimised parameters (-bb 10000 -nt AUTO -st AA). The best-fit molecular evolutionary model used in the phylogenetic analyses was JTT+R4, which was selected by ModelFinder [65] implemented within IQ-Tree.

Assessing the nuclear receptor and the “chemical defensome” repertoire in *Scomber colias*

To demonstrate the value of the present genome resource, we collected the repertoire of nuclear receptors (NRs) in *S. colias* via TBLASTN (TBLASTN, RRID:SCR_011822) searches in the primary genome assembly with default parameters. Protein sequences of DNA-binding domains and ligand-binding domains in *H. sapiens* NRs were collected from the RefSeq [50] database and used in a query (NP_000466.2, NP_068804.1, NP_003241.2, XP_005257609.1, NP_001349802.1, NP_068370.1, NP_599022.1, NP_009052.4, NP_001351014.1, XP_005260464.1, NP_002948.1, NP_001257330.1, NP_003288.2, XP_016862607.1, NP_001273031.1, NP_005645.1, NP_001278159.1, NP_004442.3, NP_000167.1, XP_005268879.1, NP_004950.2, NP_201591.2).

Next, regions aligning with *H. sapiens* sequences were collected, translated to protein using the Bio.Seq module of Biopython v.1.75 (Biopython, RRID:SCR_007173) [66], and blasted (BLASTP) against a local database containing the NR proteins of *Danio rerio* (*D. rerio* NRs database protocol: (1) NRs sequences and classifications were retrieved from [67]; (2) an NRs database was built using the makeblastdb application of NCBI-BLAST+ v.2.12.0). For each NRs sequence in *S. colias*, the best blast hit in the *D. rerio* database was collected. In some cases, several NRs of *S. colias* matched the same receptor in *D. rerio*. In these cases, the nucleotide sequences of *S. colias* were again validated against the NT-NCBI database, and all sequences matching different GenIDs in the same organism were kept in the final table of NRs. In parallel, and to assess the genome annotation performed by BRAKER2, the genomic coordinates of regions aligning with *H. sapiens* were searched and identified in the annotation files.

To identify the genes related to the *chemical defensome*, target genes were selected based on a previous report profiling the “chemical defensome” of teleost species [68]. Next, gene names were used as queries to search the deduced *S. colias* genome annotation, a simple but successful approach for well-annotated genomes such as *D. rerio* [68]. When gene names were not retrieved from *S. colias* genome annotation (i.e., *ftlh*, *gstp*, *hsph*, *maf*, *nme8*, *slc21*), further TBLASTN searches were performed in the primary genome assembly with optimised parameters (-max_hsps 1 to keep the best query-subject pair), using *D. rerio* sequences as a query.

Demography with pairwise sequentially Markovian coalescent (PSMC)

To explore the variation in the demographic history of the species, a pairwise sequentially Markovian coalescent (PSMC) (PSMC, RRID:SCR_017229) strategy was applied [69], following the authors' instructions. Briefly, PE short reads were aligned to the repeated masked genome assembly using BWA v.0.7.17-r1198 (BWA, RRID:SCR_010910) [37] with optimised parameters (BWA-MEM), and the output converted to BAM format and sorted using Samtools v.1.9 [49] (function: sort; parameters: default). Next, Picard Tools v.2.19.2 (Picard, RRID:SCR_006525) was used to remove duplicate reads (function: MarkDuplicates; parameters: default), and SAMtools for mapping quality filtering and SNP calling (function: mpileup; parameters: -Q 30 -q 30 -C 50). BCFtools v.1.9 (SAMtools/BCFtools, RRID:SCR_005227) was applied to extract consensus sequences (function: call; parameters: -c), and the subscript vcutils (from SAMtools) was used for filtering the output for a minimum depth of 25, a maximum depth of 150, and a min RMSmapQ of 20 (function: vcf2fq; parameters: -d 25 -D 150 -Q 20). The resulting fastq file was converted to a PSMC-compatible input format using fq2psmcfa with a minimum quality threshold of 20 (parameters: -q 20). Inferences of population history were performed by running PSMC for 25 iterations with optimised parameters (-N 15, -r 5, -p 4*4 + 13*2 + 4*4 + 6) following recent PSMC estimations on Scombriformes [70]. Furthermore, to account for uncertainties in the PSMC estimates, bootstrapping of 100 replicates was performed using the split face script provided by PSMC authors. Finally, to scale the demographic estimations, a mutation rate (μ) of 7.3×10^{-9} substitutions/site/generation was used, based on a recent estimation for the Scombriformes species *Thunnus albacares* [70], and a generation time for *S. colias* of 2 years [7, 71].

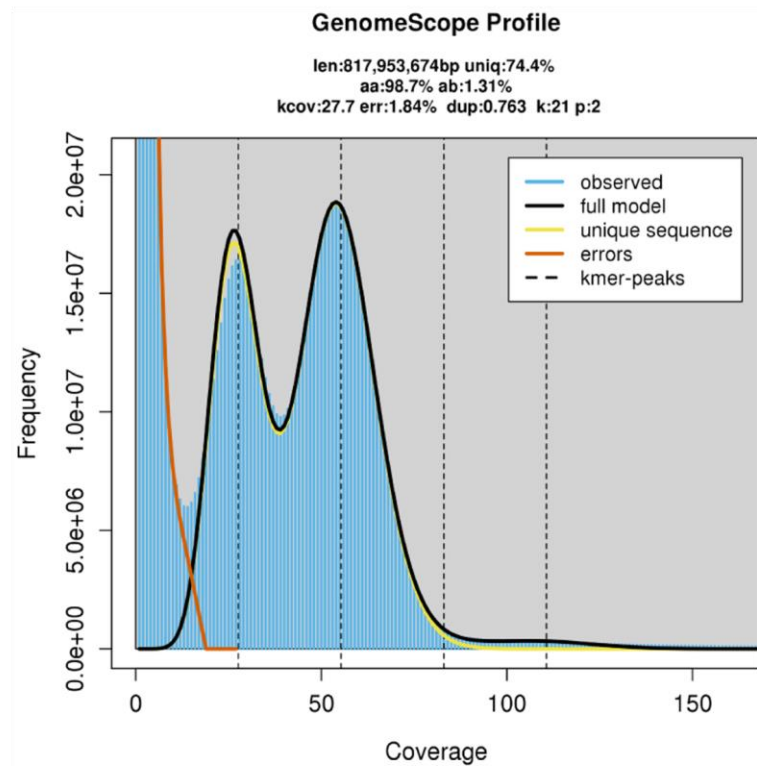


Figure 3. Genomescope2 plot with *k*-mer spectra content and fitted models of the *Scomber colias* Illumina PE dataset.

DATA VALIDATION

To produce the *S. colias* genome assembly, 2 sequencing strategies were used: Illumina PE short reads and PacBio HiFi long reads. The PE dataset was used to assess the genomic proprieties of the *S. colias* species and scaffold the long-read assembly, while HiFi reads were used to perform the primary genome assembly and gap closing (Figure 2).

The Illumina sequencing yielded 149 M of PE reads and the PacBio sequencing generated 1.7 M of HiFi reads (Table 1). Trimmed short reads were used to estimate the genome size (817 Mbp), heterozygosity rate (1.31%), and genome repeat content (approximately 26%), using GenomeScope2 (Figure 3). The complete statistics of GenomeScope2 can be consulted in Additional File 3 [17, 18]. In parallel, the HiFi dataset was inspected, and mitochondrial reads, as well as possible sources of contamination, were removed (amounting to 0.31% of the initial dataset) (Table 1).

For the mtDNA assemblies, a total of 38,868 mtDNA PE reads were filtered by GetOrganelle and a total of 792 mtDNA PacBio HiFi reads were filtered by BLASTN search. The 2 assemblies had the same length, 16,570 bp, and differed from each other by 0.29%

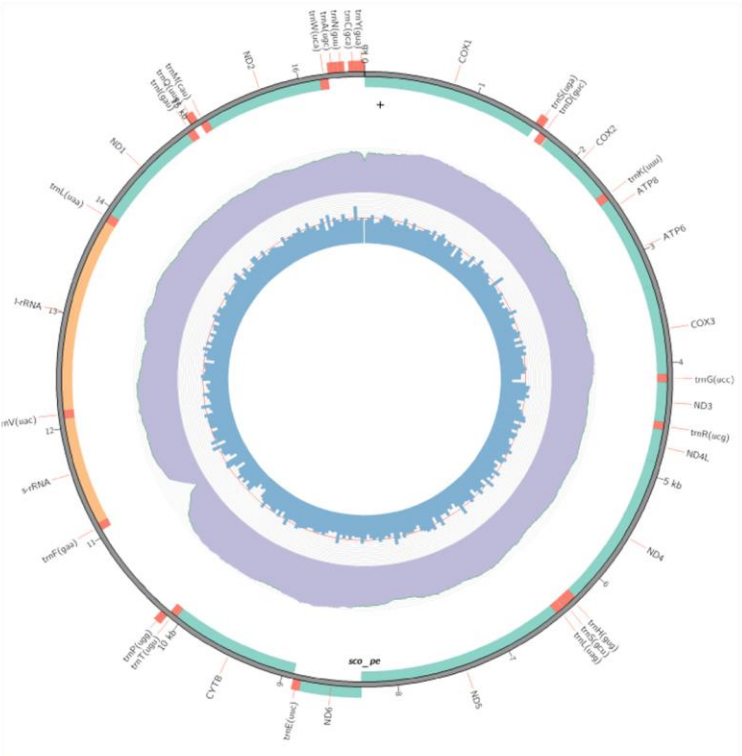


Figure 4. Circular mitochondrial genome assembly of *Scomber colias*, obtained from the Illumina PE dataset (equal to that obtained from the PacBio HiFi long reads assembly). From the centre to the outmost features: GC content distribution; sequencing depth distribution of aligned Paired-End reads; gene elements (i.e., PCGs, rRNA genes, tRNA genes).

Table 1. General statistics of read datasets used to perform the *Scomber colias* genome assembly.

Sample	Sequencing type	Library type	Platform	Insert size (bp)	Number of reads (before clean-up)	Number of reads (after clean-up)	Application
Sco_PH	WGS	Long reads	PacBio Sequel II System	15,500	1,792,104	1,786,541	Genome Assembly, Gap Closing, Assessment
Sco_PE	WGS	Short reads	HiSeq X Ten	478	149,564,893	84,738,393	Scaffolding, Assessment

(uncorrected *p*-distances). Furthermore, the PE and PacBio HiFi mtDNA assemblies differed from the *S. colias* mtDNA assembly available on NCBI (accession number [AB488406.1](#) [11]), by 0.35% and 0.40% respectively (uncorrected *p*-distances). The mtDNA gene content and arrangement is as expected for most fishes and is standard for vertebrates [72], consisting of 13 protein-coding genes, 22 transfer RNA (trn), and 2 ribosomal RNA (rrn) (Figure 4).

The primary genome assembly was produced using filtered PacBio HiFi reads and the below software packages and settings. Following the above-mentioned criteria (Material and Methods: Nuclear genome assembly and assessment) the Sco_k21 assembly was selected, with both pseudo-haplotypes merged and subjected to purge_dups. Detailed statistics of Hifiiasm and HiCanu genome assemblies can be consulted in Additional File 4 [17, 18]. Although the purge_dups generated a primary and an alternative assembly, only the primary assembly was used in subsequent steps. At the same time, 4 short-read genome assemblies were performed with W2RAP software, and contigs with over 500 bp were used as “long reads” to scaffold the primary assembly. Additional File 5 shows QUAST and BUSCO statistics for the PE genome assemblies [17, 18]. Importantly, during the scaffolding process, only structural information of short-read assemblies was used, without the inclusion of bases. Lastly, the remaining non-basal long read assemblies were used to fill gaps inserted during the scaffolding stage. The final assembly (primary assembly) of *S. Colias* yielded a genome size of 814 Mbp, distributed in 2,028 scaffolds and 2,093 contigs with an N50 length of 4.19 and 3.34 Mbp, respectively. On the other hand, the alternative assembly had 807 Mbp and 5908 contigs with an N50 length of 0.47 Mbp (Table 2). The BUSCO analyses, at the nucleotide level, in Eukaryota and Actinopterygii datasets, showed high levels of completeness for both primary (97.3% and 97.9% of single-copy orthologs) and alternative (93.3% and 96% single-copy orthologs) assemblies (Table 2). Consistently, Merqury determined high QV (primary, 56.53%; alternative, 54.99%) and *k*-mer completeness (primary, 86.11%; alternative, 84.60%) values for both assemblies (Table 2). In the primary assembly, the *k*-mer analyses (via Merqury) showed a low level of *k*-mer duplication in the genome (colour blue, green, purple, and orange in Figure 5a), indicating a high level of haplotype uniqueness (red colouring in Figure 5a), and a similar *k*-mer distribution pattern to GenomeScope2 (performed with Illumina PE reads). Additionally, we found a high mapping rate in the Illumina, PacBio, and RNA-Seq reads, against the primary assembly of 95%, 99.8%, and 90.02%, respectively. Overall, these results provide evidence of the high quality of the *S. colias* genome assembly (Table 2). Our *S. colias* genome assembly ranks fourth in high-quality genome assemblies within the order Scombriformes and first in the genus *Scomber* (Additional File 6) [17, 18].

The RepeatMasker software masked 29.62% of bases in the primary genome assembly. The masked regions were predominantly linked to DNA elements (11.66%), long interspersed nuclear elements (4.11%), long terminal repeats (2.58%), and simple repeats (2.88%). Furthermore, 8.62% of the genome was masked and annotated as “unclassified”, and only a small percentage were classified as short interspersed nuclear elements, small RNA, or satellite repeats (Table 3). The genome annotation process generated about 27,675 protein-coding genes and 30,999 protein-coding sequences. On average, we found 9.5 exons and 1,656 bp lengths per CDS (Table 4). Of the CDS, 30,355 had at least 1 BLASTP hit in SwissProt or RefSeq databases, 27,101 were identified in the InterPro database, and 21,664 of these were classified as belonging to a specific homolog superfamily (Table 5).

To validate the protein-coding sequences we performed phylogenetic analysis (via OrthoFinder) and BUSCO analysis (using the Actinopterygii library profile) (Figure 5b, c). Of the 16 Actinopterygii proteins datasets inputted to OrthoFinder, 98.3% were assigned to 29,066 orthogroups, with 12,334 orthogroups present in all species. All OrthoFinder statistics can be consulted in Additional File 7 [17, 18]. Furthermore, a total of 392 single-copy orthologues were retrieved by OrthoFinder and used for the phylogenomic

Table 2. Statistics of the *Scomber colias* genome assembly.

Table 2. Statistics of the <i>Scomber colias</i> genome assembly.			
Assembly	Alternative	Primary	
	Contigs (Sco_k21_a_c)	Contigs (Sco_k21_p_c)	Scaf olds (Sco_k21_p_s)
Number of contigs (≥10,000 bp)	5,908	2,093	2,028
Number of contigs (≥50,000 bp)	2,417	1,123	1,078
Number of contigs (≥100,000 bp)	1,593	704	662
Number of contigs (≥200,000 bp)	1,025	456	417
Number of contigs (≥500,000 bp)	421	235	209
Number of contigs (≥1,000,000 bp)	123	155	138
Total length (≥10,000 bp)	807,928,680	813,976,802	814,072,661
Total length (≥50,000 bp)	721,244,010	781,696,683	782,480,923
Total length (≥100,000 bp)	662,374,873	751,893,146	752,912,084
Total length (≥200,000 bp)	580,469,606	716,806,065	718,068,371
Total length (≥500,000 bp)	385,329,197	648,055,626	653,890,381
Total length (≥1,000,000 bp)	180,689,595	591,655,104	603,146,189
Largest contig (Mbp)	3,248	22,804,600	22,804,600
Total length (Mbp)	807,936	813,977	814,072
GC (%)	39.94	40.09	40.09
N50 (Mbp)	0,466	3,342	4,190
K-mer completeness (%)	84.602		86.1077
Consensus quality	56.5369		54.9969
Read back mapping PE (%)	-		95.0
Read back mapping PH (%)	-		99.8
Read back mapping RNA-Seq (%)	-		90.2
BUSCO statistics (databases)			
Eukariota**	T: 93.3, C: 90.2 [S: 88.6, D: 1.6], F: 3.1, M: 6.7, n: 255		
Actinopterygii**	T: 96.0, C: 94.8 [S: 91.9, D: 2.9], F: 1.2, M: 4.0, n: 3640		

* Statistics are based on contigs and scaf olds of size ≥1000 bp. ** (T, total BUSCOs found (completed + fragmented), %; C, complete BUSCOs [S, complete and single copy, %; D, complete and duplicated, %]; F, fragmented, %; M, missing, %; n, number of sequences in database).

* Statistics are based on contigs and scaffolds of size ≥1000 bp. ** (T, total BUSCOs found (completed + fragmented), %; C, complete BUSCOs [S, complete and single copy, %; D, complete and duplicated, %]; F, fragmented, %; M, missing, %; n, number of sequences in database).

analysis. Alignment, trimming and concatenation of all single-copy orthologues, resulted in a final 120,886 aa-long supermatrix alignment that was used for phylogenomic inference in IQ-Tree. The resulting Maximum Likelihood phylogenetic tree has maximum support for almost all nodes (Figure 5b). The phylogeny recovered the reciprocal monophyletic Acanthopterygii groups Pelagiaria, Eupercaria, Anabantaria, Carangaria, and Ovalentaria, with Pelagiaria being the basal clade and represented by the 3 Scombriformes, including *S. colias* (Figure 5b). These results are in accordance with the most recent phylogenomic study of ray-finned fishes [73], as well as the Ensembl Compara Species Tree of Ensembl database [51]. BUSCO analysis showed the *S. colias* proteome with 93.6% of the groups complete, 2% fragmented, and 4.4% missing (Figure 5c). In comparison, *T. maccoyii* had 99.8% BUSCO groups complete, while *T. orientalis* had but 82.8%. These results are expected, since the *T. maccoyii* genome assembly, part of the Vertebrate Genome Project [74], was built at chromosome level, with multiple technologies (including 46x PacBio data, 46x 10X Genomics Chromium data, BioNano data, and Arima Hi-C data) and several manual curation steps [75]. In contrast, both *T. orientalis* [76] and *S. colias* were built at scaffold level using only short- and long-read information.

We further explored the quality of the annotation by investigating the repertoire of the NRs superfamily in the *S. colias* assembly. NRs are critical molecular physiology components, with vital roles in animal physiology and disruption [77]. Moreover, their

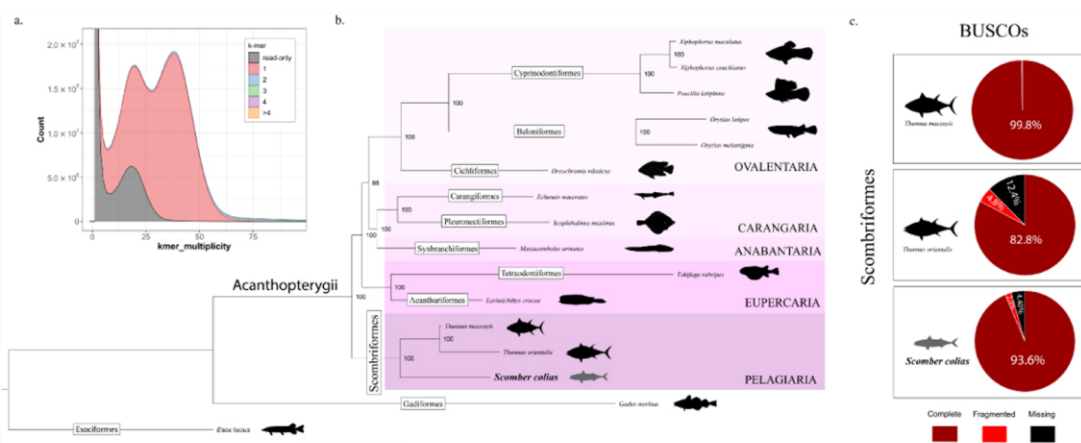


Figure 5. Validation of the genome assembly and annotation process. (a) K-mer analyses of the *Scomber colias* genome assembly (Merqury). (b) Maximum Likelihood phylogenetic tree based on the concatenated alignments of amino acid sequences of 392 single-copy orthologs retrieved by OrthoFinder. Bootstrap values are shown next to the nodes. (c) BUSCO scores were obtained from searching the proteomes of the 3 Scombriformes species with genome annotation available, against the actinopterygii_odb10 (n:3640) lineage.

Table 3. Report of RepeatMasker software. This report contains statistics of repetitive elements in the *Scomber colias* genome assembly.

Total number of sequences		2,028	
Total length (bp)		814,072,661 bp	
GC level (%)		40.09	
Number of bases masked		241,071,029 bp (29.62%)	
Type	Number of elements	Length in Genome	Percentage of Genome
SINEs	16,132	2,679,916	0.33
ALUs	0	0	0.00
MIRs	7,082	1,280,739	0.16
LINEs	113,089	33,426,533	4.11
LINE1	8,048	4,651,362	0.57
LINE2	57,670	14,551,177	1.79
L3/CR1	697	123,438	0.02
LTR elements	82,410	20,969,171	2.58
ERV1	10	279	0.00
ERV1-MaLRs	0	0	0.00
ERV1-class1	22,786	4,702,084	0.58
ERV1-class1 11490		576,448	0.07
DNA elements	623,126	94,930,706	11.66
hAT-Charlie	27,952	5,526,534	0.68
TcMar-Tigger	169	46,619	0.01
Unclassified	278,199	70,161,089	8.62
Total interspersed repeats	-	222,167,415	27.29
Small RNA	11,380	1,807,250	0.22
Satellites	18,552	3,093,792	0.38
Simple repeats	84,014	23,465,814	2.88
Low complexity	959	200,769	0.02

exact NR gene complement in vertebrate lineages has been shown to vary [67]. We were able to deduce the existence of 76 NRs in the *S. colias* genome, detailed in Additional File 8,

Table 4. Structural annotation report of the *Scomber colias* genome assembly.

Structural Annotation	Result
Number of genes	27,675
Number of mRNAs	30,999
Number of CDSs	30,999
Number of exons	295,102
Number of introns	264,103
Number of exon in CDS	295,102
Number of intron in CDS	264,103
Number of introns in exon	264,103
Number of introns in intron	235,209
Number gene overlapping	71
Number of single exon genes	2,036
Number of single exon mRNA	2,105
Mean mRNAs per gene	1.1
Mean CDSs per mRNA	1.0
Mean exons per mRNA	9.5
Mean introns per mRNA	8.5
Mean exons per CDS	9.5
Mean introns in CDSs per mRNA	8.5
Mean introns in exons per mRNA	8.5
Mean introns in introns per mRNA	7.6
Total gene length	269,856,447
Total mRNA length	310,580,471
Total CDS length	51,346,678
Total exon length	51,346,678
Total intron length	259,233,793
Total intron length per CDS	259,233,793
Total intron length per exon	259,233,793
Total intron length per intron	35,919,947
Mean gene length	9,750
Mean mRNA length	10,019
Mean CDS length	1,656
Mean exon length	173
Mean intron length	981
Mean intron in exon length	981
Mean intron in intron length	152
Longest gene	242,447
Longest mRNA	242,447
Longest CDS	98,436
Longest exon	14,939
Longest intron	76,003
Longest CDS piece	14,939
Shortest gene	303
Shortest mRNA	144
Shortest CDS	18
Shortest exon	3
Shortest intron	30

in line with the repertoire described for other teleost species [78]. Among the retrieved NRs we found those that are key components of the “chemical defensome”—an ensemble of related and unrelated genes that protect organisms against chemical stressors, and are thus critical under anthropogenic chemical build-up and climate change scenarios—such as the xenobiotic-inducible pregnane X receptor (*pxr*, *nr1i2*) [68, 79]. Subsequent analysis, using gene names, further suggested the presence of gene annotations for the vast majority of the reported members of the teleost “chemical defensome” in *S. colias*, similarly to that

Table 5. Functional annotation report of *S. colias* genome assembly.

Functional Annotation	Number
Swiss-Prot/ RefSeq	30,355
InterPro	27,101
CDD	12,832
Coils	7,705
GO	18,643
Gene3D	22,209
HAMAP	463
KEGG	1,402
MetaCyc	1,140
MobiDBlite	16,765
PIRSF	1,755
PRINTS	7,143
Pfam	25,708
PROSITE patterns	8,082
PROSITE profiles	16,229
Reactome	7,376
SFDL	114
SMART	14,906
SUPERFAMILY	21,664
TIGRFAMs	1,427

described for *D. rerio* [68]. Additional BLAST searches were performed for a reduced set of genes (*fthl*, *gstp*, *hsph*, *maf*, *nme8*, and *slc21*), uncovering possible homologs for this gene subset, except for a single member of the GST family (*gstp*). The chemical defensome repertoire identified in *S. colias* species is detailed in Additional File 9 in the associated data entries [17, 18].

We additionally validated our dataset by examining the present population structure of the species, since the genome may also provide clues regarding its past demographic history [69]. One popular method to produce these inferences is the pairwise sequentially Markovian coalescent (PSMC) model, here applied to the *S. colias* final genome assembly. Since PSMC requires an estimation of the genome-wide mutation rate, and since this has never previously been produced for *S. colias*, we used the recently estimated genome-wide mutation rate of the yellowfin tuna, *T. albacares*, of 7.3×10^{-9} mutations/site/generation [70]. The results suggest a past population expansion between 160,000–115,000 years ago, with maximum effective population size (N_e) of 36,000 during the end of the Mid-Pleistocene Transition, corresponding to the Eemian (i.e., the last interglacial period) and the transition between Marine Isotope Stages (MIS) 5 and 6 (Figure 6). This population expansion is followed by an apparent decrease in the N_e to around 25,000 at the beginning of the Late Pleistocene, corresponding to the beginning of the Last Glacial Period. These results, suggesting the influence of climatic changes from the Pleistocene glaciation cycles on the N_e , are following other recent studies on Scombriformes, such as the Pacific Sierra mackerel, *Scomberomorus sierra* [80], and the Indo-Pacific Yellowfin tuna *T. albacares* [70], as well as in other pelagic marine species such as the killer whale [81].

REUSE POTENTIAL

This study reports the first genome assembly of Atlantic chub mackerel. *Scomber colias* is a valuable marine resource, with a high impact on the fisheries of several countries along the west coast of the Atlantic Ocean and the Mediterranean Sea. Ecologically, this species

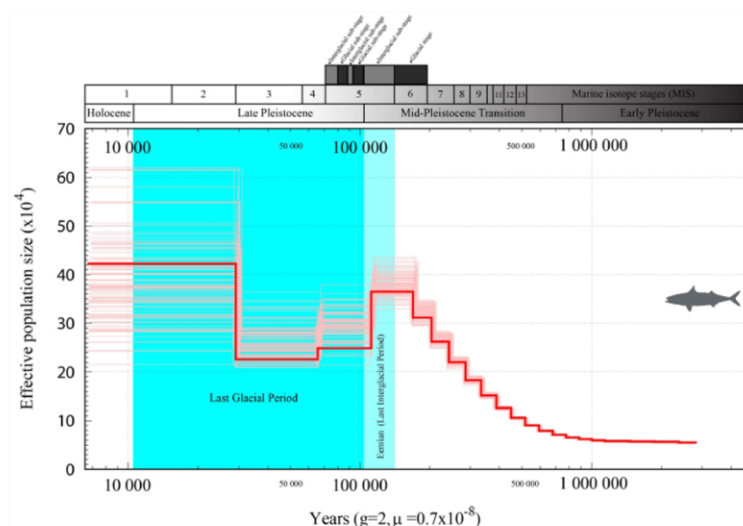


Figure 6. Pairwise sequentially Markovian coalescent (PSMC) estimates from the *Scomber colias* genome assembly. Estimations were obtained using a generation time of 2 years and genome-wide mutation rate of 7.3×10^{-8} mutations/site/generation. Effective population size (N_e) is presented in the left vertical axis, and changes estimated up to the present, over the last 3 myr, on the horizontal axis.

establishes an important link between primary producers and top predators of the marine trophic web. Despite the ecological and economic importance of *S. colias*, few genomic resources are available for this species. Thus, this genome is timely and is expected to contribute to the effective conservation, management, and sustainable exploitation of *S. colias* species in the Anthropocene. Additionally, this genome will be a key tool to decipher biological features of *S. colias*, such as population dynamics, ecology, and physiology.

DATA AVAILABILITY

Raw datasets of PacBio HiFi and Illumina sequencing were deposited in the NCBI Sequences Read Archive under Bioproject [PRJNA769550](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA769550). Additionally, both primary and alternative pseudo-haplotype assemblies were submitted to NCBI GenBank (accession numbers [JA.JDFG000000000](https://www.ncbi.nlm.nih.gov/nuccore/JA.JDFG000000000) and [JA.JDFH000000000](https://www.ncbi.nlm.nih.gov/nuccore/JA.JDFH000000000)). Mitochondrial genome assemblies and annotations were submitted to GenBank (accession numbers [OK501306](https://www.ncbi.nlm.nih.gov/nuccore/OK501306) and [OK501307](https://www.ncbi.nlm.nih.gov/nuccore/OK501307)). The four W2RAP assemblies, as well as genome annotation files and supplementary tables, were uploaded to Figshare online repository [17]. Additional data is available at the GigaScience GigaDB repository [18]. Genome and annotation datasets also can be interactively explored at <http://portugalfishomics.ciimar.up.pt/app/scombercolias/>.

DECLARATIONS

LIST OF ABBREVIATIONS

aa: amino acid; AGAT: Another Gtf/Gf Analysis Toolkit; bp: base pair(s); BUSCO: Benchmarking Universal Single-Copy Ortholog; BWA: Burrows-Wheeler Aligner;

CDS: protein-coding sequences; DHA: docosahexaenoic acid; gDNA: genomic DNA; HiFi: High-Fidelity; KAT: *k*-mer analyses toolkit; LINKS: long interval nucleotide *k*-mer scaffold; Mbp: megabase pair(s); mtDNA: mitochondrial genome; NCBI: National Center for Biotechnology Information; NRs: nuclear receptors; NT-NCBI: nucleotide database of NCBI; PacBio: Pacific Biosciences; PE: paired-end; PSMC: pairwise sequentially Markovian coalescent; QUASt: Quality Assessment Tool for Genome Assemblies; QV: consensus quality value; RNA-Seq: RNA sequencing; SMRT: single molecule, real-time; SNP: single-nucleotide polymorphism.

ETHICS APPROVAL

This work has been approved by the CIIMAR ethics committee and by CIIMAR Managing Animal Welfare Body (ORBEA) according to the European Union Directive 2010/63/EU.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

LFCC designed and conceived this work; MF, RC, and NA collected the samples; AMM, AGS, EF, LFCC wrote the manuscript; AMM, AGS, MMF, FC, MS, MD, RdF, RR, AV, and LFCC coordinated and carried out the bioinformatics analyses. All authors read, revised, and approved the final manuscript.

FUNDING

This research was funded by COMPETE 2020, Portugal 2020, and the European Union through the ERDF (grant number 031342), and by FCT through national funds (PTDC/CTA-AMB/31342/2017), and is part of the CIIMAR-lead initiative Portugal-*Fishomics*. The Foundation for Science and Technology (FCT) Portugal supported AMM (DFA/BD/8069/2020), AGS (SFRH/BD/137935/2018), AV (DL57/2016), NA (DFA/BD/6218/2020). RRdF thanks the Villum Foundation for its funding of the Center for Macroecology, Evolution, and Climate (DNRF96).

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- Collette BB, Nauer CE. Scombrids of the world. An Annotated and Illustrated Catalogue of Tunas, Mackerels, Bonitos and Related Species Known to Date. *FAO Species Cat.*, 1983; 2: 2–137. <http://www.fao.org/3/ac478e/ac478e00.htm>. Accessed 20 Jan 2020.
- Hernández JC, Ortega ATS, Castro Hernandez JJ et al. Synopsis of biological data on the chub mackerel (*Scomber japonicus* Houttuyn, 1782). *FAO Fish Synop.*, 2000; 157: 1–77. <https://agris.fao.org/agris-search/search.do?recordID=XF2000393177>. Accessed 20 Feb 2020.
- Velasco EM, del Arbol J, Baro J et al. Age and growth of the Spanish chub mackerel *Scomber colias* of southern Spain: a comparison between samples from the NE Atlantic and the SW Mediterranean. *Rev. Biol. Mar. Oceanogr.*, 2011; 46: 27–34. doi:10.4067/S0718-19572011000100004.
- Gamito R, Pita C, Teixeira C et al. Trends in landings and vulnerability to climate change in different fleet components in the Portuguese coast. *Fish Res.*, 2016; 181: 93–101. doi:10.1016/j.fishres.2016.04.008.
- Karakoltsidis PA, Zotos A, Constantinides SM. Composition of the commercially important mediterranean finfish, crustaceans, and molluscs. *J Food Compos. Anal.*, 1995; 8: 258–273. doi:10.1006/jfca.1995.1019.

- 6 Ferreira I, Gomes-Bispo A, Lourenço H et al. The chemical composition and lipid profile of the chub mackerel (*Scomber colias*) show a strong seasonal dependence: Contribution to a nutritional evaluation. *Biochimie*, 2020; **178**: 181–189. doi:10.1016/j.biochi.2020.09.022.
- 7 Carvalho N, Perrotta RG, Isidro E, Age, growth and maturity in the chub mackerel (*Scomber japonicus* Houttuyn, 1782) from the Azores. *Arquipél. Ciênc. Biol. Mar.*, 2002; **19**: 93–99. https://repositorio.uac.pt/bitstream/10400.3/169/1/pp93_99_Carvalho_et_al.pdf.
- 8 Martins MM, Skagen D, Marques V et al. Changes in the abundance and spatial distribution of the Atlantic chub mackerel (*Scomber colias*) in the pelagic ecosystem and fisheries of Portugal. *Sci. Mar.*, 2013; **77**: 551–563. doi:10.3989/scimar.03861.07B.
- 9 Vasconcelos J, Afonso-Dias M, Faria G, Atlantic chub mackerel (*Scomber colias*) spawning season, size and age at first maturity in Madeira waters. *Arquipelago Life Mar. Sci.*, 2012; **29**: 43–51. <https://www.researchgate.net/publication/273319854>.
- 10 Machado AM, Felício M, Fonseca E et al. A resource for sustainable management: De novo assembly and annotation of the liver transcriptome of the Atlantic chub mackerel, *Scomber colias*. *Data Br.*, 2018; **18**: 276–284. doi:10.1016/j.dib.2018.03.013.
- 11 Catanese G, Machado M, Infante C, Evolutionary relatedness of mackerels of the genus *Scomber* based on complete mitochondrial genomes: Strong support to the recognition of Atlantic *Scomber colias* and Pacific *Scomber japonicus* as distinct species. *Gene*, 2010; **452**: 35–43. doi:10.1016/j.gene.2009.12.004.
- 12 Rodríguez-Espeleta N, Bradbury IR, Mendibil I et al. Population structure of Atlantic mackerel inferred from RAD-seq-derived SNP markers: effects of sequence clustering parameters and hierarchical SNP selection. *Mol. Ecol. Resour.*, 2016; **16**: 991–1001. doi:10.1111/1755-0998.12518.
- 13 Ravi V, Venkatesh B, The divergent genomes of teleosts. *Annu. Rev. Anim. Biosci.*, 2018; **6**: 47–68. doi:10.1146/annurev-animal-030117-014821.
- 14 Formenti G, Theissinger K, Fernandes C et al. The era of reference genomes in conservation genomics. *Trends Ecol. Evol.*, 2022; **37**(3): 197–202. doi:10.1016/j.tree.2021.11.008.
- 15 PACBIO. Procedure & Checklist – Preparing HiFi SMRTbell® Libraries using the SMRTbell Express Template Prep Kit 2.0. Pacific Biosciences, CA; 2021; <https://www.pacb.com/wp-content/uploads/Procedure-Checklist-Preparing-HiFi-SMRTbell-Libraries-using-SMRTbell-Express-Template-Prep-Kit-2.0.pdf>.
- 16 PACBIO. SMRT link software installation (v9.0). Pacific Biosciences, CA; 2020; https://www.pacb.com/wp-content/uploads/SMRT_Link_Installation_v90.pdf.
- 17 Machado A, A draft genome assembly of the Atlantic chub mackerel (*Scomber colias*) using Illumina and Pacbio HiFi reads. Figshare Dataset. 2021; <https://doi.org/10.6084/m9.Figshare.17025506.v4>.
- 18 Machado AM, Gomes-dos-Santos A, Fonseca MM et al. Supporting data for "A genome assembly of the Atlantic chub mackerel (*Scomber colias*): a valuable teleost fishing resource". *Giga Science Database*. 2022; <http://dx.doi.org/10.5524/100978>.
- 19 Bolger AM, Lohse M, Usadel B, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 2014; **30**: 2114–2120. doi:10.1093/bioinformatics/btu170.
- 20 Ranallo-Benavidez TR, Jaron KS, Schatz MC, GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.*, 2020; **11**: 1–10. doi:10.1038/s41467-020-14998-3.
- 21 Marçais G, Kingsford C, A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics*, 2011; **27**: 764–770. doi:10.1093/bioinformatics/btr011.
- 22 Jin JJ, Yu W Bin, Yang JB et al. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.*, 2020; **21**: 241. doi:10.1186/s13059-020-02154-5.
- 23 Cheng H, Concepcion GT, Feng X et al. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods*, 2021; **18**: 170–175. doi:10.1038/s41592-020-01056-5.
- 24 Wick RR, Judd LM, Gorrie CL et al. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.*, 2017; **13**: e1005595. doi:10.1371/journal.pcbi.1005595.
- 25 Meng G, Li Y, Yang C et al. MitoZ: A toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res.*, 2019; **47**: 63. doi:10.1093/nar/gkz173.
- 26 Clavijo BJ, Garcia Accinelli G, Wright J et al. W2RAP: A pipeline for high quality, robust assemblies of large complex genomes from short read data. bioRxiv. 2017; 110999. <https://doi.org/10.1101/110999>.

- 27 Mapleson D, Accinelli GG, Kettleborough G et al. KAT: A k-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics*, 2017; **33**: 574–576. doi:10.1093/bioinformatics/btw663.
- 28 Nurk S, Walenz BP, Rhie A et al. HiCanu: Accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.*, 2020; **30**: 1291–305. doi:10.1101/gr.263566.120.
- 29 Wick RR, Schultz MB, Zobel J et al. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*, 2015; **31**: 3350–3352. doi:10.1093/bioinformatics/btv383.
- 30 Manni M, Berkeley MR, Seppey M et al. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.*, 2021; **38**: 4647–4654. doi:10.1093/molbev/msab199.
- 31 Gurevich A, Saveliev V, Vyahhi N et al. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, 2013; **29**: 1072–1075. doi:10.1093/bioinformatics/btt086.
- 32 Guan D, McCarthy SA, Wood J et al. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, 2020; **36**: 2896–2898. doi:10.1093/bioinformatics/btaa025.
- 33 Jones S, Taylor G, Chan S et al. The Genome of the Beluga Whale (*Delphinapterus leucas*). *Genes (Basel)*, 2017; **8**: 378. doi:10.3390/genes8120378.
- 34 Taylor GA, Kirk H, Coombe L et al. The Genome of the North American Brown Bear or Grizzly: *Ursus arctos* ssp. *horribilis*. *Genes (Basel)*, 2018; **9**: 598. doi:10.3390/genes9120598.
- 35 Warren RL, Yang C, Vandervalk BP et al. LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *Gigascience*, 2015; **4**: 35. doi:10.1186/s13742-015-0076-3.
- 36 Warren RL, RAILS and Cobbler: Scaffolding and automated finishing of draft genomes using long DNA sequences. *J Open Source Softw.*, 2016; **1**: 116. doi:10.21105/joss.00116.
- 37 Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 2010; **26**: 589–595. doi:10.1093/bioinformatics/btp698.
- 38 Li H, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 2018; **34**: 3094–3100. doi:10.1093/bioinformatics/bty191.
- 39 Kim D, Paggi JM, Park C et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.*, 2019; **37**: 907–915. doi:10.1038/s41587-019-0201-4.
- 40 Kim D, Langmead B, Salzberg SL. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods*, 2015; **12**: 357–360. doi:10.1038/nmeth.3317.
- 41 Rhie A, Walenz BP, Koren S et al. Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.*, 2020; **21**: 245. doi:10.1186/s13059-020-02134-9.
- 42 Chen N. Using Repeat Masker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics*, 2004; **5**: 4–10. doi:10.1002/0471250953.bi0410s25.
- 43 Smit AFA, Hubley R. RepeatMasker Open-1.0. <http://www.repeatmasker.org>.
- 44 Hubley R, Finn RD, Clements J et al. The Dfam database of repetitive DNA families. *Nucleic Acids Res.*, 2016; **44**: D81–D89. doi:10.1093/nar/gkv1272.
- 45 Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA*, 2015; **6**: 1–6. doi:10.1186/s13100-015-0041-9.
- 46 Hof KJ, Lange S, Lomsadze A et al. BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, 2015; **32**: 767–769. doi:10.1093/bioinformatics/btv661.
- 47 Hof KJ, Lomsadze A, Borodovsky M et al. Whole-genome annotation with BRAKER. *Methods Mol. Biol.*, 2019; **1962**: 65–95. doi:10.1007/978-1-4939-9173-0_5.
- 48 Brůna T, Hof KJ, Lomsadze A et al. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.*, 2021; **3**: lqaa108. doi:10.1093/nargab/lqaa108.
- 49 Danecek P, Bonfield JK, Liddle J et al. Twelve years of SAMtools and BCFtools. *Gigascience*, 2021; **10**(2): giab008. doi:10.1093/gigascience/giab008.
- 50 O'Leary NA, Wright MW, Brister JR et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, 2016; **44**: D733–D745. doi:10.1093/nar/gkv1189. Accessed 15 May 2021.

- 51 Yates AD, Achuthan P, Akanni W et al. Ensembl 2020. *Nucleic Acids Res.*, 2020; **48**: D682–D688. doi:10.1093/nar/gkz966.
- 52 Dainat J. AGAT : Another Gf Analysis Toolkit to handle annotations in any GTF/GFF format. Zenodo. (Version v0.6.0). 2021; <https://www.doi.org/10.5281/zenodo.4637977>.
- 53 Jones P, Binns D, Chang HY et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 2014; **30**: 1236–1240. doi:10.1093/bioinformatics/btu031.
- 54 Bateman A, Martin MJ, Orchard S et al. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.*, 2021; **49**: D480–D489. doi:10.1093/nar/gkaa1100. Accessed 15 May 2021.
- 55 Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*, 2014; **12**: 59–60. doi:10.1038/nmeth.3176.
- 56 Buels R, Yao E, Diesh CM et al. JBrowse: A dynamic web platform for genome visualization and analysis. *Genome Biol.*, 2016; **17**: 66. doi:10.1186/s13059-016-0924-1.
- 57 Gremme G, Steinbiss S, Kurtz S. Genome tools: A comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 2013; **10**: 645–656. doi:10.1109/TCBB.2013.68.
- 58 Li H. Tabix: Fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics*, 2011; **27**: 718–719. doi:10.1093/bioinformatics/btq671.
- 59 Zhang Z, Schwartz S, Wagner L et al. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.*, 2000; **7**: 203–214. doi:10.1089/10665270050081478.
- 60 Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.*, 2015; **16**: 157. doi:10.1186/s13059-015-0721-2.
- 61 Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, 2004; **32**: 1792–1797. doi:10.1093/nar/gkh340.
- 62 Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T et al. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 2009; **25**: 1972–1973. doi:10.1093/bioinformatics/btp348.
- 63 Kück P, Longo GC, FASconCAT-G: Extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front. Zool.*, 2014; **11**: 81. doi:10.1186/s12983-014-0081-x.
- 64 Nguyen L-T, Schmidt HA, von Haeseler A et al. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.*, 2015; **32**: 268–274. doi:10.1093/molbev/msu300.
- 65 Kalyaanamoorthy S, Minh BQ, Wong TKF et al. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods*, 2017; **14**: 587–589. doi:10.1038/nmeth.4285.
- 66 Chapman B, Chang J. Biopython: Python tools for computational biology. *ACM SIGBIO Newsl.*, 2000; **20**: 15–19. doi:10.1145/360262.360268.
- 67 Fonseca E, Machado AM, Vilas-Arrondo N et al. Cartilaginous fishes offer unique insights into the evolution of the nuclear receptor gene repertoire in gnathostomes. *Gen. Comp. Endocrinol.*, 2020; **295**: 113527. doi:10.1016/j.ygcen.2020.113527.
- 68 Eide M, Zhang X, Karlén OA et al. The chemical defensome of five model teleost fish. *Sci. Rep.*, 2021; **11**: 1–13. doi:10.1038/s41598-021-89948-0.
- 69 Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature*, 2011; **475**: 493–496. doi:10.1038/nature10231.
- 70 Barth JMI, Damerau M, Matschiner M et al. Genomic differentiation and demographic histories of atlantic and indo-pacific yellowfin tuna (*Thunnus albacares*) populations. *Genome Biol. Evol.*, 2017; **9**: 1084–1098. doi:10.1093/gbe/evx067.
- 71 Martins MM. Growth variability in Atlantic mackerel (*Scomber scombrus*) and Spanish mackerel (*Scomber japonicus*) of Portugal. *ICES J. Mar. Sci.*, 2007; **64**: 1785–1790. doi:10.1093/icesjms/fsm163.
- 72 Satoh TP, Miya M, Mabuchi K et al. Structure and variation of the mitochondrial genome of fishes. *BMC Genom.*, 2016; **17**: 719. doi:10.1186/s12864-016-3054-y.
- 73 Hughes LC, Orti G, Huang Y et al. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc. Natl. Acad. Sci. USA*, 2018; **115**: 6249–6254. doi:10.1073/pnas.1719358115.

- 74 Rhie A, McCarthy SA, Fedrigo O et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, 2021; **592**: 737–746. doi:10.1038/s41586-021-03451-0.
- 75 Howe K, Chow W, Collins J et al. Significantly improving the quality of genome assemblies through curation. *Giga Science*, 2021; **10**: 1–9. doi:10.1093/gigascience/giaa153.
- 76 Suda A, Nishiki I, Iwasaki Y et al. Improvement of the Pacific bluefin tuna (*Thunnus orientalis*) reference genome and development of male-specific DNA markers. *Sci. Rep.*, 2019; **9**: 1–12. doi:10.1038/s41598-019-50978-4.
- 77 Santos MM, Ruivo R, Capitão A et al. Identifying the gaps: Resources and perspectives on the use of nuclear receptor based-assays to improve hazard assessment of emerging contaminants. *J Hazard. Mater.*, 2018; **358**: 508–511. doi:10.1016/j.jhazmat.2018.04.076.
- 78 Bertrand S, Thisse B, Tavares R et al. Unexpected novel relational links uncovered by extensive developmental profiling of nuclear receptor expression. *PLoS Genet.*, 2007; **3**: 2085–2100. doi:10.1371/journal.pgen.0030188.
- 79 Eide M, Rydbeck H, Tørresen OK et al. Independent losses of a xenobiotic receptor across teleost evolution. *Sci. Rep.*, 2018; **8**: 1–13. doi:10.1038/s41598-018-28498-4.
- 80 López MD, Alcocer MU, Jaimes PD. Phylogeography and historical demography of the Pacific Sierra mackerel (*Scomberomorus sierra*) in the Eastern Pacific. *BMC Genet.*, 2010; **11**: 34. doi:10.1186/1471-2156-11-34.
- 81 Moura AE, Van Rensburg CJ, Pilot M et al. Killer whale nuclear genome and mtDNA reveal widespread population bottleneck during the last glacial maximum. *Mol. Biol. Evol.*, 2014; **31**: 1121–1131. doi:10.1093/molbev/msu058.

3. Chapter 3 – The Complete Genome Sequence of *Hydrolagus colliei* (Lay & Bennett, 1839), the spotted ratfish

André M. Machado, André Gome-dos-Santos, Tereza Almeida, Fabiana Neves, João Moutinho, Elsa Froufe, Jonathan Wilson, Raquel Ruivo, Yuko Ohta, Martin Flajnik, Akane Kawaguchi, Shigehiru Kuraku, Ana Veríssimo, L. Filipe C. Castro

(In preparation)

The Complete Genome Sequence of *Hydrolagus colliei* (Lay & Bennett, 1839), the spotted ratfish.

André M. Machado^{1,2}, André Gome-dos-Santos¹, Tereza Almeida^{3,4}, Fabiana Neves^{3,4}, João Moutinho^{1,2}, Elsa Froufe¹, Jonathan Wilson⁵, Raquel Ruivo¹, Yuko Ohta^{6,7}, Martin Flajnik⁸, Akane Kawaguchi^{6,7}, Shigehiru Kuraku^{6,7}, Ana Veríssimo^{3,4*}, L. Filipe C. Castro^{1,2*}

¹CIIMAR/CIMAR — Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Avenida General Norton de Matos, S/N, P 4450-208 Matosinhos, Portugal

²FCUP — Department of Biology, Faculty of Sciences, University of Porto, Rua do Campo Alegre 1021/1055, 4169-007 Porto, Portugal

³CIBIO — Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661, Vairão, Portugal

⁴BIOPOLIS — Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661, Vairão, Portugal

⁵Department of Biology, Wilfrid Laurier University, Waterloo, ON, N2L 3C5, Canada

⁶Molecular Life History Laboratory, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

⁷Department of Genetics, Sokendai (Graduate University for Advanced Studies), Mishima, Shizuoka 411-8540, Japan

⁸Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD, USA

*Correspondence to: Shigehiru Kuraku (skuraku@nig.ac.jp), Ana Veríssimo (verissimoac@gmail.com), Luís Filipe Costa de Castro (filipe.castro@ciimar.up.pt)

Keywords: Chondrichthyes, Chimaeras, Genome assembly, flavin-containing monooxygenases

3.1. Abstract

Chimaeriformes represents one of the most intriguing orders within the class Chondrichthyes, comprising more than 37 species distributed across three families (*Callorhynchidae*, *Rhinochimaeridae*, and *Chimaeridae*). These organisms are globally distributed and exhibit a unique combination of morphological, anatomical, physiological and behavioural traits. *Hydrolagus colliei*, also known as spotted ratfish (Lay and Bennett, 1839), was one of the first species described within this order and is frequently used as representative of the *Chimaeridae* family. It has played an essential role in elucidating the emergence of several anatomical features and systems in gnathostomes.

Despite, its evolutionary significance among vertebrates, there is a paucity of omic and molecular resources available for this species. In this study we present a chromosome-scale reference genome assembly for *H. colliei* using a combination of PacBio HiFi and Hi-C technologies. The final PacBio genome assembly spans 1,14 Gb, comprising 2,196 contigs, with an N50 length of 2,1 Mb and a total Vertebrate BUSCO score of 93.4%. Hi-C chromosomal mapping enabled the reconstruction of 35 chromosome sequences, representing 85,3% of the total contig content. Additionally, RNA-Seq datasets for twelve tissues were generated and employed for genome annotation. This allowed to identify several repetitive sequences, including, SINEs, LINEs, LTR elements as well as 22,621 coding genes encompassing 27,671 proteins. This high-quality chromosome-scale assembly along with the transcriptomic datasets, provides a valuable omic resources for further studies investigating the evolutionary history of gnathostomes, as well as the biodiversity and biology of *H. colliei* and other Chondrichthyes.

3.2. Background & Summary

The origin of cartilaginous fish, the earliest-branching lineage of the gnathostomes, dates back to approximately 450 million years (Andreev et al. 2017). Extant species fall into two lineages: the (sharks, skates, and rays) and Holocephali (chimaeroids), comprising more than 1088 species, catalogued in 16 orders, 61 families and 199 genera (Nelson et al., 2016; Weigmann, 2016). Holocephali, representing only 4% of Chondrichthyes diversity, possess distinctive and intriguing biology - including an holostylic jaw suspension, non-replaceable hyper-mineralized tooth-plates and complete, non-suspensory hyoid arch, an unusual adaptive immune system, or the lack of a gastric stomach – and offer key insights to decipher the evolution of gnathostomes (e.g., Barnett et al., 2009; Castro et al., 2014; Didier et al., 2012; Lisney, 2010; Smith et al., 2004).

The spotted ratfish *Hydrolagus colliei* (Lay and Bennett, 1839), a member of the Chimaeridae family, occurs from the southeast Alaska to the tip of Baja and in the northern Gulf of California and is among the few Chimaeriformes reported from shallow water to depths of 1029 meters (0–1029 m) (Didier 2004; Inoue et al. 2010; Lisney 2010; Weigmann 2016; Berio et al. 2024) (Figure 3.1 A and B). This species predominantly feeds of hard-shelled preys (durophagy), but also opportunistically feeds on echinoderms, annelids, and small fishes, and occasionally engages in intracohort cannibalism (Johnson and Horton 1972; Mathews 1975; Allen and Horn 2006; Huber et al. 2008). Ecologically important as both predator and prey, it ranks among the ten most abundant ground fishes in U.S. West Coast and indirectly fisheries by serving as food

source for commercial valuable species such Pacific halibut (*Hippoglossus stenolepis*) or Spiny dogfish (*Squalus acanthias*) (Johnson and Horton 1972; Allen and Horn 2006; Keller et al. 2006; Barnett et al. 2012). Interestingly, recent studies have utilized this species as a bioindicator of bioaccumulation and biomagnification of highly persistent chemical compounds such as copper, zinc, lead, mercury, as well as metalloids, like selenium and arsenic, thus demonstrating the ecotoxicological potential of this species (Bergés-Tiznado et al. 2021; Tiznado-Salazar et al. 2023).

Despite its ecological, fisheries and evolutionary and recent ecotoxicological significance, *omic* resources for this and other holocephalans remain scarce (Nishimura et al. 2022). Among limited molecular resources available, most are mitochondrial genomes, primarily used for phylogenetic placement (e.g., (Gomes-dos-Santos et al., 2020; Gomes-dos-Santos, Vilas-Arrondo, et al., 2021; Matos et al., 2024; Vilas-Arrondo et al., 2022). In contrast, the nuclear genome sequencing efforts for species such elephant shark (*Callorhinchus mili*), small-eyed rabbitfish (*Hydrolagus affinis*) and more recently the silver chimaera (*Chimaera phantasma*) have elevated Holocephali as valuable comparative in vertebrate genomics (Fonseca et al., 2020; Nakatani et al., 2021; Teramura et al., 2024; Venkatesh et al., 2014).

Here, we report a chromosome-level genome assembly of the spotted ratfish, generated using highly accurate long reads from PacBio HiFi sequencing and high-throughput chromosome conformation capture (Hi-C) technologies. Additionally, transcriptomic datasets from twelve tissues were sequenced to support the functional annotation of the genome. This annotation encompasses repetitive regions, non-coding RNAs, mRNAs, proteins and gene predictions. Finally, to demonstrate the intrinsic value of this dataset, we evaluated the status of a family of enzymes essential for the chemical defence against xenobiotics, the flavin-containing monooxygenases (FMOs) in the spotted ratfish and other Chondrichthyes. Together, these resources are essential for understanding the biology, molecular features, and evolutionary history of Chimaeriformes and gnathostomes.

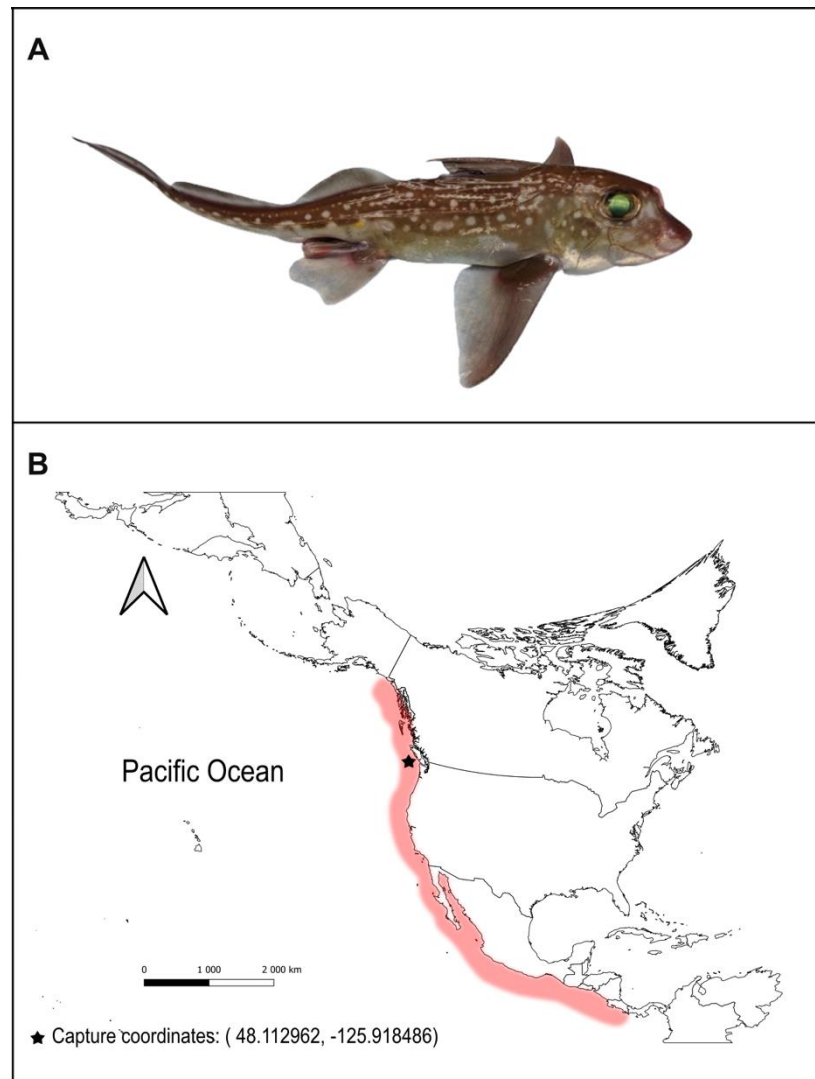


Figure 3.1 – A) Photograph of spotted ratfish, *Hydrolagus colliei* (Provided by Martin Flajnik). B) Potential distribution of *H. colliei* in North America (Retrieved from FishBase Database - https://www.fishbase.se/summary/Hydrolagus_colliei.html).

3.3. Methods

Animal sampling

In this study were used various samples of *Hydrolagus colliei*. The sample employed for PacBio approach was obtained from an adult male specimen (18.2 cm snout-vent length) captured using hook and line on continental shelf waters off Vancouver (Eastern North Pacific). This specimen was transported to laboratory, dissected, and samples from different tissues were collected and stored in RNA later at -80 °C (Table 3.1 and 3.2). For Hi-C analyses, a female *H. colliei* larva that had ceased at Aquamarine Fukushima was used. The specimen had been frozen at -80 °C and was later dissected to obtain muscle tissue samples (Table 3.1 and 3.2).

Item	Accession Numbers of NCBI
Investigation_type	Eukaryote
Project_name	<i>Hydrolagus colliei</i> (spotted ratfish)
Specimen	A
Number of Samples	1
Lat_lon	48.112962, -125.918486
Geo_loc_name	Eastern North Pacific, off Vancouver
Tissues	Kidney
Collection_date	2018/08/09
Biome	Sea water (ENVO:00002149)
Feature	Ocean (ENVO:00000015)
Material	Sea water (ENVO:00002149)
seq_meth	WGS - Pacbio HiFi
Collector	Living Elements, Ltd.
Maturity	Mature
Datasets Generated	Accession Numbers of NCBI
Bioproject	PRJNA1026076
Raw datasets	SRR31870140
Genome Assembly	JBMHJC000000000

Table 3.1 – MixS descriptors and accession number of tissue sample and raw data used in the initial assembly of *H. colliei* genome.

PacBio HiFi - DNA extraction, library construction, and sequencing

For PacBio HiFi sequencing, the two samples of *H. colliei* brain and kidney were shipped to Brigham Young University (BYU) (Table 3.1 and 3.2). At BYU, DNA extraction, PacBio HiFi library preparation, and sequencing were performed using the manufacture's protocol guidance (<https://www.pacb.com/wp-content/uploads/Procedure-Checklist-Preparing-HiFi-SMRTbell-Libraries-using-SMRTbell-Express-Template-Prep-Kit-2.0.pdf>). The size-selection was carried-out with SageELF, and the sequencing was performed on two single-molecule real-time (SMRT) cells using the Sequel II system v.9.0 (run time 30h, pre-extension 2,9 h). The in-silico analyses of circular consensus were conducted using the SMRT® Link v9.0 with the default parameters (https://www.pacb.com/wp-content/uploads/SMRT_Link_Installation_v90.pdf).

Specimen	Sequencing Strategy	Sequencing type	Tissue	Library type	Platform	Number of reads	Application
----------	---------------------	-----------------	--------	--------------	----------	-----------------	-------------

A	PacBio HiFi - Cell1	WGS	Kidney	Long reads	PacBio Sequel II system	1,602,179	Genome assembly
A	PacBio HiFi - Cell2	WGS	Kidney	Long reads	PacBio Sequel II system	1,691,573	Genome assembly
B	Hi-C - Illumina	WGS	Muscle	Short reads	Novaseq 6000	431,000,000	Genome size estimation, Genome Anchoring
C	RNA-Seq	RNA-Seq	Gill	Short reads	Novaseq 6000	23,395,451	Genome annotation
C	RNA-Seq	RNA-Seq	Heart	Short reads	Novaseq 6000	23,540,175	Genome annotation
C	RNA-Seq	RNA-Seq	Liver	Short reads	Novaseq 6000	22,221,844	Genome annotation
C	RNA-Seq	RNA-Seq	Muscle	Short reads	Novaseq 6000	21,327,138	Genome annotation
C	RNA-Seq	RNA-Seq	Thymus	Short reads	Novaseq 6000	22,102,960	Genome annotation
C	RNA-Seq	RNA-Seq	Skin	Short reads	Novaseq 6000	23,706,805	Genome annotation
C	RNA-Seq	RNA-Seq	Spiral valve	Short reads	Novaseq 6000	21,278,108	Genome annotation
C	RNA-Seq	RNA-Seq	RBC	Short reads	Novaseq 6000	23,501,394	Genome annotation
D	RNA-Seq	RNA-Seq	Spleen	Short reads	Novaseq 6000	23,232,299	Genome annotation
D	RNA-Seq	RNA-Seq	Kidney	Short reads	Novaseq 6000	24,997,489	Genome annotation
D	RNA-Seq	RNA-Seq	Testis	Short reads	Novaseq 6000	22,337,017	Genome annotation
D	RNA-Seq	RNA-Seq	Liver	Short reads	Novaseq 6000	23,536,807	Genome annotation

Table 3.2 – General characteristics and applications of the sequencing datasets.

RNA-Seq - RNA extraction, library construction, and sequencing

The total RNA was extracted from eleven tissues. At the end of the extraction process, the eleven samples were shipped to Novogene (China), where eleven strand-specific poly(A) libraries - one per sample – were constructed and sequenced using 150 bp paired-end reads on the NovaSeq PE150 platform (Table 3.2).

Hi-C - In situ Hi-C library preparation and sequencing

The protocols described in (Schloissnig et al. 2021; Kawaguchi and Tanaka 2023), with modifications, were used to prepare the in situ Hi-C library for genome assembly. The library preparation was started from 200 mg of the frozen muscle tissue. The frozen muscle tissue was ground to a powder with a mortar and pestle under liquid nitrogen. The ground tissue was collected directly into 1 % formaldehyde with 10 % FCS/PBS in 50 ml conical tubes, and gently rotated for 10 min at room temperature. Fixation was

stopped by the addition of ice-cold glycine (125 mM final concentration) and tubes were kept on ice for 15 min with occasional inversion for mixing. Cells were collected by centrifugation (350 x g for 10 min at 4 °C), and washed with ice-cold PBS twice (350 x g for 10 min at 4 °C). Cell pellets were then resuspended in ice-cold lysis buffer (10 mM Tris-HCl pH7.5, 10 mM NaCl, 0.2 % NP-40, x50 Protease inhibitor EDTA free (SIGMA; P8340)) and incubated on ice for 30 min with occasional agitation. The nuclei were then collected by the centrifugation (300 x g for 10 min at 4 °C) and washed with 500 µl of 1.25 x NEBuffer3 at the same speed. The nuclei were then resuspended in 1 ml of 1.25 x NEBuffer3 and transferred to DNA low-binding tubes (500 µl each). SDS (0.5 % final concentration) was added and the tubes were incubated for 2 hours at 37°C with agitation (950 rpm). Triton X-100 (3.3 % final concentration) was then added to quench SDS, and the nuclei were incubated with agitation (950 rpm) for 2 hours at 37°C. The nuclei were pelleted with centrifugation (100 x g for 1 min at room temperature) and the pellets were resuspended in 450 µl of 1x DpnII digestion buffer (NEB; R0543M). DpnII digestion was performed with agitation (950 rpm) for overnight at 37 °C. After gDNA digestion, the enzyme was inactivated by heating the tubes to 65 °C for 20 min. Overhung filling was carried out using biotin-14-dATP, dCTP, dGTP, and dTTP, along with 50 U of Klenow (NEB) for 1 hours at 37°C with repeated agitation (700 rpm, 10 sec and rest 30 sec repeat program for 1 hour in a thermal cycler). Ligation was performed overnight at 18 °C with 2000 U of T4 DNA ligase. After de-crosslinking step with Protease K in SDS buffer for overnight at 65 °C, RNase A treatment was conducted for 2 hours at 37 °C. Subsequently, DNA extraction was performed with phenol/chloroform and EtOH precipitation steps, and the ligated gDNA products were eluted in 20 µl of TE buffer. The yield was determined using Qubit3 device, and 2.5 µg of Biotinylated DNA was used for the library preparation. To remove excess biotin, un-ligated fragment ends were then treated with 3U of T4 DNA polymerase for 30 min at 37°C, and the reaction was stopped by adding EDTA (10 mM final concentration). DNA was sonicated using Covaris S220 system to generate DNA fragments with the size peak around 400 bp (Covaris settings: duty factor: 10 %; peak incident power: 5; cycles per burst: 200; time: 70sec). After end repair with T4 DNA polymerase, Klenow Large fragment and T4 DNA polynucleotide kinase (all from NEB) in the presence of dNTPs in T4 DNA ligation buffer (for 30 min at room temperature), the DNA was purified with Monarch DNA purification kit (NEB). A double size selection using Ampure beads (Beckman) was performed (see (Schloissnig et al. 2021; Kawaguchi and Tanaka 2023)). From the DNA pools, Hi-C ligation products were isolated using pre-washed MyOne Streptavidin C1 Dynabeads (Life Technologies, Cat. No. 65001) on a magnetic stand in binding buffer (5 mM Tris pH8.0, 0.5 mM EDTA, 1 M NaCl) for 30 min at room temperature. dA-tailing was performed on the beads with

Klenow exo- (for 1 hours at 37 °C). After two washes in binding buffer and one wash in T4 DNA ligation buffer, Illumina adapters were ligated to Hi-C ligation products bound to streptavidin beads in T4 DNA ligase (for 2 hours at room temperature). After washing twice with wash buffer (5 mM Tris pH8.0, 0.5 mM EDTA, 1 M NaCl, 0.05 % Tween-20) and then once with binding buffer, the Hi-C products bound to the beads were resuspended in a final volume of 20 µl 1x NEBuffer2. The captured Hi-C DNA was amplified by PCR amplification cycles with NEBNext® High-Fidelity 2X PCR Master Mix (NEB, Cat.# M0541S) for 7 cycles. The final Hi-C libraries were purified with x1 volume of AMPure XP beads. The concentration of the Hi-C library was determined using Tapestation (Agilent), and the Hi-C libraries were sequenced with PE150 setting on NovaSeq 6000.

Raw data clean-up and contig genome assembly

The Pacbio HiFi datasets obtained from BYU were inspected and polished with several bioinformatic tools. Initially, the Deepconsensus tool was applied to the raw subreads from two SMRT cells to improve the mean quality value (QV) of the consensus sequences (Figure 3.2).

To obtain the contig-level genome assembly of *H. colliei*, several software tools and parameter configurations were applied. During this process we tested Hifiasm v.0.19.5 with multiple parameter sets ($l=2$, $l=3$, and $--hg-size=1,1g$), Flye v.2.8.3 ($--pacbio-hifi$; $-g 1.1g$), wtdbg2 v.2.5 with the following options ($-x ccs$; $--rescue-low-cov-edges$) and Peregrine (pghifid) (Figure 3.2) (Kolmogorov et al. 2019; Ruan and Li 2019; Cheng et al. 2021). The contig-level assemblies were evaluated using multiple tools. Gene content completeness was assessed with the Benchmarking Universal Single-Copy Orthologs (BUSCO) v.5.2.2 with Eukaryota and Vertebrata databases (v.odb10) (Simão et al. 2015; Manni et al. 2021). Structural features of the assemblies were inferred with Quality Assessment Tool for Genome Assemblies (QUAST) v.5.0.2 (Gurevich et al. 2013). Read back mappings of PacBio HiFi and RNA-Seq datasets were performed with Minimap2 v.2.24 and Hisat2 v.2.2.0, respectively (Kim et al., 2015, 2019; H. Li, 2018) (Figure 3.2). Kmer distributions were evaluated with Kmer Analysis toolkit v.2.4.1 (KAT) (Mapleson et al. 2017). Based in the evaluation of these metrics we selected the version “Hifiasm_ $l=2$ ” of the contig-level assembly: This version presented the highest BUSCO score, contiguity and N50, and the most equilibrated genome size between pseudo-haplotypes, in terms of total size, while also exhibiting the lowest proportion of missing BUSCO genes (Supplementary Table S1).

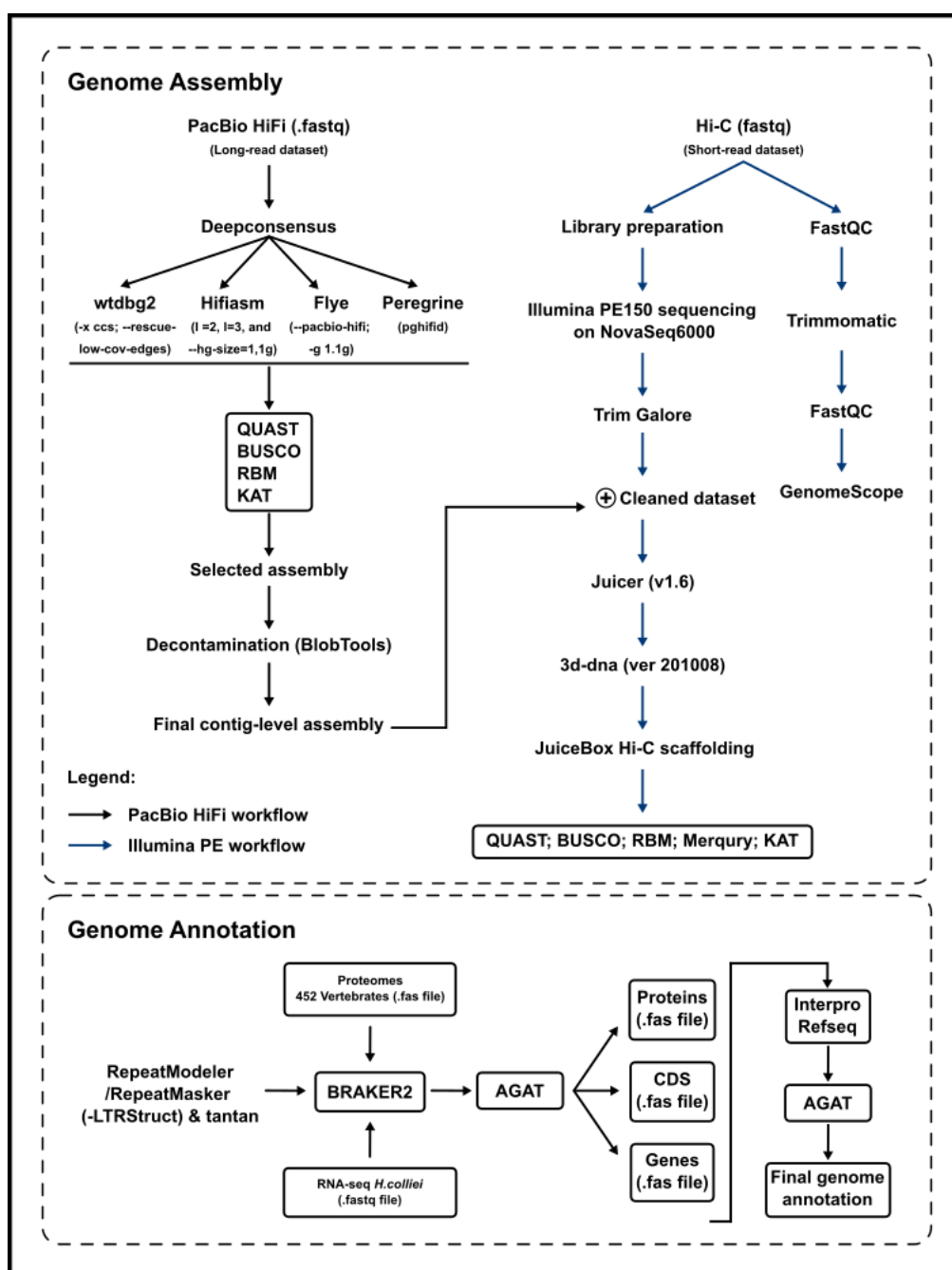


Figure 3.2 – Bioinformatic workflow used in the genome assembly and annotation of *H. coliei*.

Before proceeding with the Hi-C anchoring analyses we conducted the final evaluations of potential exogenous contaminations with BlobTools v.1.1.1. To this end, the RefSeq database was queried using BLASTX function of DIAMOND v.2.0.13.151, with the default setting (Buchfink et al. 2014; O’Leary et al. 2016). The resulting. outfmt6 and .bam file (output from BLASTX analyses and read back mapping (RBM), respectively) were used as input for BlobTools (Figure 3.2). The taxonomic classification was performed at order level, ensuring that the final assembly was free of contamination (Laetsch and Blaxter, 2017).

Hi-C scaffolding and final genome assessment

The obtained Hi-C read pairs were evaluated using FastQC to assess general sequences features, including the total number of reads, Phred quality scores and GC content (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The datasets were then used to estimate genomic features of the species, such as genome size, heterozygosity, repetitive content. These features were inferred using Jellyfish v.2.2 and GenomeScope v2-0 with a K-mer length of 21 (Marçais and Kingsford 2011; Ranallo-Benavidez et al. 2020) (Figure 3.2). Subsequently, these reads were processed with Trim_Galore v0.6.8 (<https://github.com/FelixKrueger/TrimGalore>) specifying the options ‘--phred33 --stringency 2 --quality 30 --length 25’ and aligned to the HiFi sequence contigs with the program Juicer v1.6 (Durand et al. 2016), and using its results, HiFi sequence contigs were scaffolded with 3d-dna (version 201008; Dudchenko et al., 2017) specifying the options ‘-m haploid -i 1000 --editor-repeat-coverage 15 -r 2’ to be consistent with the chromatin contact profiles (Figure 3.2).

Repeat masking and genome annotation

The genome annotation process was conducted through three main steps; repeat masking to identify and mask repetitive elements, structural annotation using BRAKER2 to detect genomic features such as genes, proteins and mRNAs, and functional annotation through integration of several databases to characterise and identify these features. Briefly, repeat masking was carried out in three phases. Firstly, a *de novo* library was created based on the *H. colliei* genome assembly using the RepeatModeler v.2.0.133 (Flynn et al., 2020). This *ab initio* library was then merged with Dfam_consensus-20170127 and RepBase-20181026, to create a comprehensive master library. Finally, both the genome assembly and the master library were input into RepeatMasker, which soft-masked the genome (Figure 3.2) (Tarailo-Graovac and Chen 2009; Bao et al. 2015; Hubley et al. 2016).

Structural annotation of *H. colliei* genome was performed using the BRAKER2 v.2.1.6 pipeline (Brůna et al., 2021). This pipeline integrates *ab initio* gene prediction tools and allows the inclusion of supplementary data such as RNA-seq and protein sequences from the target organism and or closely related species. Prior to gene prediction, twelve RNA-seq datasets generated in this study were mapped to the *H. colliei* genome assembly with Hisa2 v.2.2.0 (Table. 3.2) (Kim et al., 2019). These alignments were subsequently converted to BAM and sorted using Samtools v.1.9 (Li et al., 2009). In parallel, 452 proteomes were collected from RefSeq NCBI database, including all Chondrichthyes and Chimaeriformes available, as well as model organism

such as *Homo sapiens*, *Mus musculus*, *Gallus gallus*, *Danio rerio* among others vertebrates (Supplementary Table 2) (Jones et al. 2014; O’Leary et al. 2016). The RNA-seq alignments and protein datasets were input into BRAKER2 pipeline with the following parameters (-etpmode; - softmasking; -UTR = off; -crf; - cores =30). BRAKER2 employed GeneMark-EP+ and AUGUSTUS for gene prediction and training based on transcriptomic evidence, while the ProHint pipeline map the proteins to the predicted genes (Stanke et al. 2006; Brúna et al. 2020). The final output of this pipeline is a GTF annotation file, that includes gene models with several levels of evidence (in silico, transcriptomic, proteomic, or a combination of thereof). For this study we retained only those gene models supported by RNA-seq and/or protein evidences. This filtration as well the format conversion (from GTF to GFF3) were performed using BRAKER2 auxiliary scripts (selectSupportedSubsets.py and gtf2gff.pl).

The functional annotation was performed using a combination of datasets and software tools. Initially, the annotation file was filtered, uniformized and corrected using Another Gtf/Gff Analysis Toolkit (AGAT) v.1.0.0 (Dainat et al. 2022). The following AGAT scripts were executed in sequence with default parameters: 1: agat_sp_manage_functional_annotation.pl; 2: agat_sp_fix_overlapping_genes.pl; 3: agat_sp_add_start_and_stop.pl; 4: agat_sp_fix_cds_phases.pl; 5: agat_sp_statistics.pl; 5: agat_sp_extract_sequences.pl (with -p protein flag enabled). In the end of this process the protein sequences were generated, and, after a quick validation with BUSCO software, were submitted to InterProScan v.5.44.80 tools and queried against protein databases, such as RefSeq) and UniProtKB/SwissProt (O’Leary et al. 2016; Bateman et al. 2017). The sequences homology searches were conducted using Diamond v.0.11.149 tool in BLASTP mode with the following parameters (-k 1 -b 20 -e 1e-5 --sensitive --outfmt 6). The resulting homology and domain data were incorporated into the final GFF3 file with AGAT script (agat_sp_manage_functional_annotation.pl).

Identification of flavin-containing monooxygenases in Chondrichthyes

To identify the repertoire of flavin-containing monooxygenases in Chondrichthyes we employed exploratory homology-based approaches. Initially, we performed BLASTP searches against the RefSeq database, using the FMO protein sequences of *Homo sapiens* species (retrieved from NCBI) as references (FMO1 - NP_001269621.1; FMO2 - NP_001451.2; FMO3 - NP_001002294.1; FMO4 - NP_002013.1; FMO5 - NP_001452.2). Next, the homology searches were conducted against an internal RefSeq database, constructed with DIAMOND v.0.11.149 software, using the following paremeters “(-k 1 -b 20 -e 1e-5 --sensitive --outfmt 6)”. In parallel, we also created a protein database based on *H. colliciei* braker2 ORF predictions. Overall, these searches

enabled the identification of potential FMO candidates in *H. colliei* and in all species present in the RefSeq database, with particular focus on Chondrichthyes.

3.4. Data Records

The Pacbio HiFi, Illumina and RNA-seq raw sequencing outputs were deposited at the NCBI within the BioProject: PRJNA1026076 (Table 3.1). The chromosome-level genome assembly was deposit at GenBank under accession number JBMHJC000000000 (Table 3.1).

3.5. Technical Validation

Genome assembly

The assembly of the *H. colliei* genome, at both contig and chromosome level, was evaluated using several approaches. Structural metrics were analysed using QUAST tool, which indicate that the *H. colliei* genome comprised a total size of 1,114 Gb, initially organised into 2,078 contigs and lastly resolved into 35 chromosomes (85.3% of the contig assembly content) (Table 3.3 and 3.4). The N50 lengths of the contig and chromosome assemblies were 2,20 Mb and 31,54 Mb, respectively (Table 3.3). Gene content completeness was assessed using BUSCO tool. BUSCO results showed similar levels of completeness between both genome versions (contig and chromosome) against Vertebrata and Metazoa databases. At the chromosome level, BUSCO reported 96.1% and 93.3%, for Metazoa and Vertebrata databases, respectively. More specifically, 90.06% and 89.8% were identified as complete single-copy genes, 2.4% and 1.8% as complete duplicated genes, 3.1% and 1.7% as fragmented genes and only 3.9% and 6.7% of genes were missing in each respective database (Table 3.3). Additionally, read back mapping was employed as an extra evaluation metric. The high-ranking mapping rates to the initially assembly, evidence a strong integration of the initial sequencing dataset in both genome versions (Table 3.3).

Genome Assemblies				
Structural level	Contig		Chromosome	
Version	Primary Haplotype	Alternate Haplotype	Consensus Haplotype	Consensus Haplotype
Genome size (Gb)	1,11	1,04	1,14	1,14
Contig / Scaffold N50 (Mb)	1,44	1,46	2,19	31,54
Largest sequence (Mb)	12,97	11,26	13,24	158,17
Number of sequences	2,695	2,053	2078	1,842
Number of chromosomes	-	-	-	35
N's per 100 kbp	0	0	0	80.91
GC Content (%)	43.56	43.23	43.65	43.65
Average Illumina read-mapping rate (%)	99,47	99,17	99,67	99,67
Average Pacbio HiFi read-mapping rate (%)	99,78	99,55	99,89	99,89
Genome BUSCO (Eukariota)* (%)	T:94%,C:90.9% [S:88.2%,D:2.7%] F:3.1%,M:6.0%,n:255	T:93%,C:91.0% [S:89.0%,D:2.0%] F:2.0%,M:7.0%,n:255	T:96%,C:93.3% [S:90.2%,D:3.1%] F:2.7%,M:4.0%,n:55	T:96.1%,C:93.0% [S:90.6%,D:2.4%] F:3.1%,M:3.9%,n:255
Genome BUSCO (Vertebrata)* (%)	T:91.8%,C:89.4% [S:87.4%,D:2.0%] F:2.4%,M:8.2%,n:3354	T:89.6%,C:87.8% [S:86.3%,D:1.5%] F:1.8%,M:10.4%,n:3354	T:93.4%,C:91.6% [S:89.3%,D:2.3%] F:1.8%,M:6.6%,n:354	T:93.3%,C:91.6% [S:89.8%,D:1.8%] F:1.7%,M:6.7%,n:3354
Genome Annotation				
Number of genes	-	-	-	22,621
Number of mRNA sequences	-	-	-	27,671
Number of protein sequences	-	-	-	27,671
Average gene length (bp)	-	-	-	20,845
Average intron length (bp)	-	-	-	2,442
Average exon length (bp)	-	-	-	156
Genes annotated by Interproscan (=>1 Database)	-	-	-	20,364
Genes with blast hit in NCBI database (=>1)	-	-	-	19,928
Proteins BUSCO (Vertebrata)* (%)	-	-	-	T:95.4%,C:91.4% [S:75.5%,D:15.9%] F:3.0%,M:5.6%,n:3354

*Statistics are based on contigs and scaffolds of size ≥ 10000 bp. * (T, total BUSCOs found (completed + fragmented), %; C, complete BUSCOs [S, complete and single copy, %; D, complete and duplicated, %]; F, fragmented, %; M, missing, %; n, number of sequences in database)

Table 3.3 – Statistics of *H. collei* genome assembly and annotation.

Number scaffolds	Number of bp (%)	Percentage of total genome (%)
Chromossome 1	158170875	13,86%
Chromossome 2	139723355	12,24%
Chromossome 3	112144134	9,83%
Chromossome 4	91120134	7,98%
Chromossome 5	43616500	3,82%
Chromossome 6	31540993	2,76%
Chromossome 7	30054760	2,63%
Chromossome 8	25637785	2,25%
Chromossome 9	24673444	2,16%
Chromossome 10	22210733	1,95%
Chromossome 11	20092888	1,76%
Chromossome 12	19843245	1,74%
Chromossome 13	19647255	1,72%
Chromossome 14	19239510	1,69%
Chromossome 15	19060887	1,67%
Chromossome 16	18525232	1,62%
Chromossome 17	17910000	1,57%
Chromossome 18	15488590	1,36%
Chromossome 19	14704491	1,29%
Chromossome 20	13872139	1,22%
Chromossome 21	13511009	1,18%
Chromossome 22	13494516	1,18%
Chromossome 23	12377364	1,08%
Chromossome 24	12247237	1,07%
Chromossome 25	11732962	1,03%
Chromossome 26	11195446	0,98%
Chromossome 27	10796962	0,95%
Chromossome 28	8439861	0,74%
Chromossome 29	5105496	0,45%
Chromossome 30	4132788	0,36%
Chromossome 31	3751137	0,33%
Chromossome 32	3445235	0,30%
Chromossome 33	2638111	0,23%
Chromossome 34	1832500	0,16%
Chromossome 35	1695280	0,15%
Unplaced Scaffolds	167684110	14,69%

Table 3.4 – Percentage of the contig genome assembly anchored to 35 chromosomes.

Prior to Hi-C anchoring, we inspected the contig-level genome assembly for potential contaminations with BlobTools. As shown in Figure 3.3B, the BlobTools read coverage plot clearly demonstrate a predominance of hits in Callorhinchidae family followed by hits assigned to Rajidae family (Figure 3.3B). Yet, Figure 3.3B shows a high proportion of undefined or non-Chimaeriformes matches – yet phylogenetically related groups such Elasmobranch – underscores the limited availability of molecular resources for Chimaeriformes species. Additionally, the Kmer spectra of the contig-level assembly shows a low-level duplication, prior to chromosome anchoring (Figure 3.3A). In the chromosome-level assembly, Hi-C heatmaps reveal well-defined clusters along diagonal position indicating strong interactions between sequences. The lack of prominent non-adjacent/non-diagonal sequences demonstrated successfully anchoring of most sequences (Figure 3.3C).

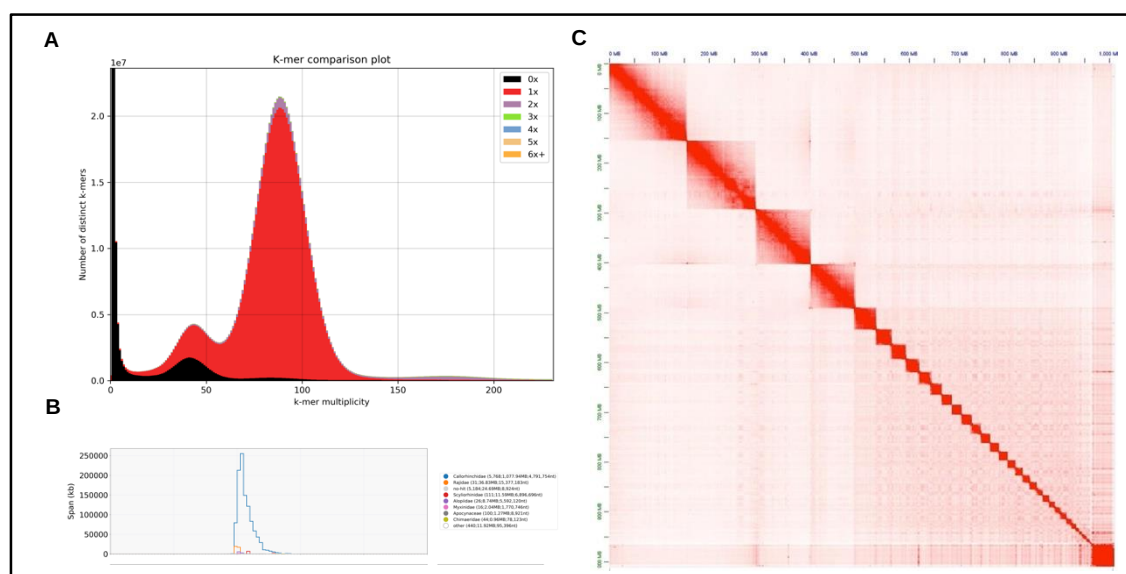


Figure 3.3 – Quality-control analyses of the *H. coliei* genome. A) Stacked histogram showing the read k-mer frequency spectrum of the *H. coliei* genome assembly at the contig level. B) BlobTools read coverage plot showing the taxonomic affiliation of each contig to a taxonomic family within the RefSeq database. C) Hi-C interactions among 35 chromosomes. The intensity of the interactions increases from white to dark red. Each central square represents a pair of *H. coliei* chromosomes.

Genome annotation

The functional annotation was conducted using a comprehensive pipeline. Initially, the chromosome-level genome assembly was subject to repeat masking, structural annotation via BRAKER2, and functional annotation using databases such InterProscan and NCBI-RefSeq databases. The structural annotation, enabled the identification of 22,621 genes codifying 27,671 protein coding sequences (Table 3.3 and Supplementary Table S4). These sequences were also evaluated with BUSCO, which

revealed a completeness of 95.4% in Vertebrata Database (Table 3.3). Globally, the genome assembly and annotation of *H. colliei* demonstrate low redundancy, high completeness, contiguity and overall quality, substantiating the robustness of the dataset presented in this study.

Short case study of Flavin-containing monooxygenases

To demonstrate the potential applicability of these datasets, we conducted a brief case study. We began by exploring FMOs in *H. colliei* and then expanded our analyses to include other Chondrichthyes species. FMOs are part of a network of genes involved in the defending organisms against chemical aggressors, known as the 'chemical defensome' (Goldstone et al. 2006; Goldstone 2008; Stegeman et al. 2010). These genes are essential to all forms of life and have been studied in several species across the metazoan kingdom, primarily in Tetrapoda. However, their identification and characterization in the earliest gnathostome lineages, particularly Chondrichthyes and Chimaeriformes, remains unclear (Krueger and Williams 2005; Nicoll et al. 2019; Bailleul et al. 2023; Nicoll and Mascotti 2023). Table 3.5 reports the identification of approximately 46 FMO candidates in twenty Chondrichthyes species. As with the elephant shark, two FMO candidates were identified in *H. colliei*. Interestingly, Elasmobranch species exhibited a heterogeneous FMO composition, particularly in *Hemitrygon akajei*, which seem to display a potential gene expansion.

Subclass	Species	Taxon ID	GeneID	Gene Symbol
Holocephali	<i>Hydrolagus colliei</i>	7873	HCOG00000011761	Best hit - LOC103180026
	<i>Hydrolagus colliei</i>	7873	HCOG00000011763	Best hit - LOC103180025
	<i>Callorhinchus milii</i>	7868	103180025	LOC103180025
	<i>Callorhinchus milii</i>	7868	103180026	LOC103180026
Elasmobranchii	<i>Carcharodon carcharias</i>	13397	121289233	LOC121289233
	<i>Carcharodon carcharias</i>	13397	121274925	LOC121274925
	<i>Chiloscyllium plagiosum</i>	36176	122554592	LOC122554592
	<i>Chiloscyllium punctatum</i>	137246	140479648	LOC140479648
	<i>Hemiscyllium ocellatum</i>	170820	132812464	LOC132812464
	<i>Hemiscyllium ocellatum</i>	170820	132818968	LOC132818968
	<i>Hemitrygon akajei</i>	2704970	140736636	LOC140736636
	<i>Hemitrygon akajei</i>	2704970	140737396	LOC140737396
	<i>Hemitrygon akajei</i>	2704970	140737248	LOC140737248
	<i>Hemitrygon akajei</i>	2704970	140737244	LOC140737244
	<i>Hemitrygon akajei</i>	2704970	140736601	LOC140736601
	<i>Hemitrygon akajei</i>	2704970	140737397	LOC140737397
	<i>Hemitrygon akajei</i>	2704970	140737395	LOC140737395
	<i>Hemitrygon akajei</i>	2704970	140737393	LOC140737393
	<i>Hemitrygon akajei</i>	2704970	140737392	LOC140737392
	<i>Hemitrygon akajei</i>	2704970	140737249	LOC140737249
	<i>Hemitrygon akajei</i>	2704970	140737242	LOC140737242
	<i>Hemitrygon akajei</i>	2704970	140715820	LOC140715820
	<i>Heptranchias perlo</i>	212740	137325512	LOC137325512
	<i>Heterodontus francisci</i>	7792	137356371	LOC137356371
	<i>Heterodontus francisci</i>	7792	137356302	LOC137356302
	<i>Heterodontus francisci</i>	7792	137373229	LOC137373229
	<i>Hypanus sabinus</i>	79690	132402145	fmo5
	<i>Hypanus sabinus</i>	79690	132402146	LOC132402146
	<i>Leucoraja erinaceus</i>	7782	129701192	LOC129701192
	<i>Leucoraja erinaceus</i>	7782	129694814	LOC129694814
	<i>Leucoraja erinaceus</i>	7782	129716176	LOC129716176
	<i>Mobula birostris</i>	1983395	140206098	LOC140206098
	<i>Mobula birostris</i>	1983395	140206097	LOC140206097
	<i>Mobula hypostoma</i>	723540	134354682	LOC134354682
	<i>Mobula hypostoma</i>	723540	134354683	LOC134354683
	<i>Narcine bancroftii</i>	1343680	138763499	LOC138763499
	<i>Pristiophorus japonicus</i>	55135	139268732	LOC139268732
	<i>Pristiophorus japonicus</i>	55135	139255435	LOC139255435
	<i>Pristis pectinata</i>	685728	127566445	LOC127566445
	<i>Pristis pectinata</i>	685728	127567849	LOC127567849
	<i>Rhincodon typus</i>	259920	109917356	LOC109917356
	<i>Scyliorhinus canicula</i>	7830	119964875	LOC119964875
	<i>Scyliorhinus canicula</i>	7830	119960797	LOC119960797
	<i>Scyliorhinus torazame</i>	75743	140399616	LOC140399616
	<i>Scyliorhinus torazame</i>	75743	140426589	LOC140426589
	<i>Stegostoma tigrinum</i>	3053191	125456437	LOC125456437

Table 3.5 – Flavin-containing monooxygenases in Chondrichthyes.

Code Availability

Most of the software used in this study were run using the default parameters, following the authors recommendations. Details regarding any non-standard parameters, along with software versions, are provided in the Material and Methods section of this manuscript.

Acknowledgements

AMM was supported by the Portuguese Foundation for Science and Technology (FCT) under the grants DFA/BD/8069/2020. We also extend our thanks to Shinya Yamauchi at Aquamarine Fukushima for animal sampling.

Author contributions

L.F.C.C., A.V., S.K. designed and conceived this work. T.A., F.N., M.F. collected the samples. A.M.M., J.M., A.G.S., Y.O., J.W., A.K., carry on all the analysis. A.M.M., L.F.C.C., A.V., E.F., R.R., M.F. wrote the first version of the manuscript.

All authors read, revised, and approved the final manuscript.

Competing interests

The authors have no conflict of interest to declare.

3.6. References

- Allen, L. G., & Horn, M. H. (2006). *The ecology of marine fishes: California and adjacent waters*. Univ of California Press.
<https://doi.org/10.1525/california/9780520246539.001.0001>
- Andreev, P. S., Coates, M. I., Karatajūtė-Talimaa, V., Shelton, R. M., Cooper, P. R., & Sansom, I. J. (2017). Egestolepis and its kin, the earliest monodontode chondrichthyans. *Journal of Vertebrate Paleontology*, 37(1), e1245664.
<https://doi.org/10.1080/02724634.2017.1245664>
- Bailleul, G., Yang, G., Nicoll, C. R., Mattevi, A., Fraaije, M. W., & Mascotti, M. L. (2023). Evolution of enzyme functionality in the flavin-containing monooxygenases. *Nature Communications* 2023 14:1, 14(1), 1–10. <https://doi.org/10.1038/s41467-023-36756-x>
- Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, 6(1), 1–6. <https://doi.org/10.1186/s13100-015-0041-9>

- Barnett, L. A. K., Earley, R. L., Ebert, D. A., & Cailliet, G. M. (2009). Maturity, fecundity, and reproductive cycle of the spotted ratfish, *Hydrolagus coliei*. *Marine Biology*, 156(3), 301–316. <https://doi.org/10.1007/s00227-008-1084-y>
- Barnett, L. A. K., Ebert, D. A., & Cailliet, G. M. (2012). Evidence of stability in a chondrichthyan population: case study of the spotted ratfish *Hydrolagus coliei* (Chondrichthyes: Chimaeridae). *Journal of Fish Biology*, 80(5), 1765–1788. <https://doi.org/10.1111/J.1095-8649.2011.03216.X>
- Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-AJee, H., Cowley, A., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., Castro, L. G., Figueira, L., ... Zhang, J. (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, 45(D1), D158–D169. <https://doi.org/10.1093/nar/gkw1099>
- Bergés-Tiznado, M. E., Véliz-Hernández, I. J., Bojórquez-Sánchez, C., Zamora-García, O. G., Márquez-Farías, J. F., & Páez-Osuna, F. (2021). The spotted ratfish *Hydrolagus coliei* as a potential biomonitor of mercury and selenium from deep-waters of the northern Gulf of California. *Marine Pollution Bulletin*, 164, 112102. <https://doi.org/10.1016/J.MARPOLBUL.2021.112102>
- Berio, F., Charron, R., Dagouret, J. M., De Gasperis, F., Éon, A., Meunier, E., Simonet, M., Verschraegen, N., & Hirel, N. (2024). Husbandry conditions of spotted ratfish (*Hydrolagus coliei*, Chimaeriformes) in aquaria for successful embryonic development and long-term survival of juveniles. *Zoo Biology*, 43(2), 188–198. <https://doi.org/10.1002/ZOO.21813>
- Brûna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021a). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, 3(1), lqaa108. <https://doi.org/10.1093/nargab/lqaa108>
- Brûna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021b). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, 3(1), lqaa108. <https://doi.org/10.1093/NARGAB/LQAA108>
- Brûna, T., Lomsadze, A., & Borodovsky, M. (2020). GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics and Bioinformatics*, 2(2). <https://doi.org/10.1093/NARGAB/LQAA026>

- Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 2014 12:1, 12(1), 59–60. <https://doi.org/10.1038/nmeth.3176>
- Castro, L. F. C., Gonçalves, O., Mazan, S., Tay, B. H., Venkatesh, B., & Wilson, J. M. (2014). Recurrent gene loss correlates with the evolution of stomach phenotypes in gnathostome history. *Proceedings of the Royal Society B: Biological Sciences*, 281(1775). <https://doi.org/10.1098/RSPB.2013.2669>
- Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* 2021 18:2, 18(2), 170–175. <https://doi.org/10.1038/s41592-020-01056-5>
- Dainat, J., Hereñú, D., Davis, E., Crouch, K., LucileSol, Agostinho, N., pascal-git, & tayyrov. (2022). *AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format*. (1.0.0). <https://doi.org/10.5281/ZENODO.7255559>
- Didier, D. A. (2004). Phylogeny and classification of extant Holocephali. *Biology of Sharks and Their Relatives*. CRC Press, Boca Raton, 4, 115–138. <https://doi.org/10.1201/b11867>
- Didier, D. A., Kemper, J. M., & Ebert, D. A. (2012). Phylogeny, Biology, and Classification of Extant Holocephalans. *Biology of Sharks and Their Relatives: Second Edition*, 97–122. <https://doi.org/10.1201/b11867>
- Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., & Aiden, E. L. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* (New York, N.Y.), 356(6333), 92–95. <https://doi.org/10.1126/SCIENCE.AAL3327>
- Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems*, 3(1), 95–98. <https://doi.org/10.1016/j.cels.2016.07.002>
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, 117(17), 9451–9457. <https://doi.org/10.1073/PNAS.1921046117>
- Fonseca, E., Machado, A. M., Vilas-Arrondo, N., Gomes-dos-Santos, A., Veríssimo, A., Esteves, P., Almeida, T., Themudo, G., Ruivo, R., Pérez, M., da Fonseca, R., Santos, M. M., Froufe, E., Román-Marcote, E., Venkatesh, B., & Castro, L. F. C. (2020).

Cartilaginous fishes offer unique insights into the evolution of the nuclear receptor gene repertoire in gnathostomes. *General and Comparative Endocrinology*, 295, 113527. <https://doi.org/10.1016/j.ygcen.2020.113527>

Goldstone, J. V. (2008). Environmental sensing and response genes in cnidaria: The chemical defensome in the sea anemone *Nematostella vectensis*. *Cell Biology and Toxicology*, 24(6), 483–502. <https://doi.org/10.1007/s10565-008-9107-5>

Goldstone, J. V., Hamdoun, A., Cole, B. J., Howard-Ashby, M., Nebert, D. W., Scally, M., Dean, M., Epel, D., Hahn, M. E., & Stegeman, J. J. (2006). The chemical defensome: Environmental sensing and response genes in the *Strongylocentrotus purpuratus* genome. *Developmental Biology*, 300(1), 366–384. <https://doi.org/10.1016/j.ydbio.2006.08.066>

Gomes-dos-Santos, A., Arrondo, N. V., Machado, A. M., Veríssimo, A., Pérez, M., Román, E., Castro, L. F. C., & Froufe, E. (2020). The complete mitochondrial genome of the deep-water cartilaginous fish *Hydrolagus affinis* (de Brito Capello, 1868) (Holocephali: Chimaeridae). *Mitochondrial DNA Part B*, 5(2), 1810–1812. <https://doi.org/10.1080/23802359.2020.1749154>

Gomes-dos-Santos, A., Vilas-Arrondo, N., Machado, A. M., Veríssimo, A., Pérez, M., Baldó, F., Castro, L. F. C., & Froufe, E. (2021). Shedding light on the Chimaeridae taxonomy: the complete mitochondrial genome of the cartilaginous fish *Hydrolagus mirabilis* (Collett, 1904) (Holocephali: Chimaeridae). *Mitochondrial DNA Part B*, 6(2), 420–422. <https://doi.org/10.1080/23802359.2020.1870887>

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>

Huber, D. R., Dean, M. N., & Summers, A. P. (2008). Hard prey, soft jaws and the ontogeny of feeding mechanics in the spotted ratfish *Hydrolagus coliei*. *Journal of The Royal Society Interface*, 5(25), 941–953. <https://doi.org/10.1098/RSIF.2007.1325>

Hubley, R., Finn, R. D., Clements, J., Eddy, S. R., Jones, T. A., Bao, W., Smit, A. F. A., & Wheeler, T. J. (2016). The Dfam database of repetitive DNA families. *Nucleic Acids Research*, 44(D1), D81–D89. <https://doi.org/10.1093/nar/gkv1272>

Inoue, J. G., Miya, M., Lam, K., Tay, B. H., Danks, J. A., Bell, J., Walker, T. I., & Venkatesh, B. (2010). Evolutionary Origin and Phylogeny of the Modern Holocephalans (Chondrichthyes: Chimaeriformes): A Mitogenomic Perspective. *Molecular Biology and Evolution*, 27(11), 2576–2586. <https://doi.org/10.1093/MOLBEV/MSQ147>

Johnson, A. G., & Horton, H. F. (1972). *Length-weight relationship, food habits, parasites, and sex and age determination of the ratfish, Hydrolagus coliei (Lay and Bennett)*. 421–429.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30(9), 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>

Kawaguchi, A., & Tanaka, E. M. (2023). Chromosome Conformation Capture for Large Genomes. *Methods in Molecular Biology (Clifton, N.J.)*, 2562, 291–318. https://doi.org/10.1007/978-1-0716-2659-7_20

Keller, A. A., Horness, B. H., Tuttle, V. J., Wallace, J. R., Simon, V. H., Fruh, E. L., Bosley, K. L., & Kamikawa, D. J. (2006). *The 2002 US West Coast upper continental slope trawl survey of groundfish resources off Washington, Oregon, and California: Estimates of distribution, abundance, and length composition*. <https://doi.org/10.13140/RG.2.2.24116.17284>

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357–360. <https://doi.org/10.1038/nmeth.3317>

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8), 907–915. <https://doi.org/10.1038/s41587-019-0201-4>

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* 2019 37:5, 37(5), 540–546. <https://doi.org/10.1038/s41587-019-0072-8>

Krueger, S. K., & Williams, D. E. (2005). Mammalian flavin-containing monooxygenases: Structure/function, genetic polymorphisms and role in drug metabolism. In *Pharmacology and Therapeutics* (Vol. 106, Issue 3, pp. 357–387). Pergamon. <https://doi.org/10.1016/j.pharmthera.2005.01.001>

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, 6, 1287. <https://doi.org/10.12688/f1000research.12232.1>

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>

- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/https://doi.org/10.1093/bioinformatics/btp352>
- Lisney, T. J. (2010). A review of the sensory biology of chimaeroid fishes (Chondrichthyes; Holocephali). *Reviews in Fish Biology and Fisheries* 2010 20:4, 20(4), 571–590. <https://doi.org/10.1007/S11160-010-9162-X>
- Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, 38(10), 4647–4654.
- Mapleson, D., Accinelli, G. G., Kettleborough, G., Wright, J., & Clavijo, B. J. (2017). KAT: A K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics*, 33(4), 574–576. <https://doi.org/10.1093/bioinformatics/btw663>
- Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27(6), 764–770.
- Mathews, C. P. (1975). Some observations on the ecology of *Galeus piperatus* Springer and Wagner, a little known shark endemic to the Northern Gulf of California*. *Journal of Fish Biology*, 7(1), 77–82. <https://doi.org/10.1111/J.1095-8649.1975.TB04577.X>
- Matos, A., Vilas-Arrondo, N., Gomes-dos-Santos, A., Veríssimo, A., Román-Marcote, E., Baldó, F., Moreno-Aguilar, J., Pérez, M., Lopes-Lima, M., Froufe, E., & Castro, L. F. C. (2024). The complete mitogenome of the Atlantic longnose chimaera *Rhinochimaera atlantica* (Holt & Byrne, 1909). *Mitochondrial DNA Part B*, 9(7), 886–891. <https://doi.org/10.1080/23802359.2024.2378127>
- Nakatani, Y., Shingate, P., Ravi, V., Pillai, N. E., Prasad, A., McLysaght, A., & Venkatesh, B. (2021). Reconstruction of proto-vertebrate, proto-cyclostome and proto-gnathostome genomes provides new insights into early vertebrate evolution. *Nature Communications*, 12(1), 1–14. <https://doi.org/10.1038/s41467-021-24573-z>
- Nelson, J. (2016). *Fishes of the World* (Fifth edition). John Wiley & Sons. <https://doi.org/10.1002/9781119174844>
- Nicoll, C. R., Bailleul, G., Fiorentini, F., Mascotti, M. L., Fraaije, M. W., & Mattevi, A. (2019). Ancestral-sequence reconstruction unveils the structural basis of function in mammalian FMOs. *Nature Structural & Molecular Biology* 2019 27:1, 27(1), 14–24. <https://doi.org/10.1038/s41594-019-0347-2>

- Nicoll, C. R., & Mascotti, M. L. (2023). Investigating the biochemical signatures and physiological roles of the FMO family using molecular phylogeny. *BBA Advances*, 4, 100108. <https://doi.org/10.1016/J.BBADVA.2023.100108>
- Nishimura, O., Rozewicki, J., Yamaguchi, K., Tatsumi, K., Ohishi, Y., Ohta, T., Yagura, M., Niwa, T., Tanegashima, C., & Teramura, A. (2022). Squalomix: shark and ray genome analysis consortium and its data sharing platform. *F1000Research*, 11, 1077. <https://doi.org/10.12688/f1000research.123591.1>
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufu, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., ... Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44(D1), D733–D745. <https://doi.org/10.1093/nar/gkv1189>
- Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, 11(1), 1–10. <https://doi.org/10.1038/s41467-020-14998-3>
- Ruan, J., & Li, H. (2019). Fast and accurate long-read assembly with wtdbg2. *Nature Methods* 2019 17:2, 17(2), 155–158. <https://doi.org/10.1038/s41592-019-0669-3>
- Schloissnig, S., Kawaguchi, A., Nowoshilow, S., Falcon, F., Otsuki, L., Tardivo, P., Timoshevskaya, N., Keinath, M. C., Smith, J. J., Voss, S. R., & Tanaka, E. M. (2021). The giant axolotl genome uncovers the evolution, scaling, and transcriptional control of complex gene loci. *Proceedings of the National Academy of Sciences of the United States of America*, 118(15), e2017176118. <https://doi.org/10.1073/pnas.2017176118>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Smith, M., Warmolts, D., Thoney, D., & Hueter, R. (2004). The elasmobranch husbandry manual: captive care of sharks, rays and their relatives. *Special Publication of the Ohio Biological Survey*, 589(145), 4.
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research*, 34(suppl_2), W435–W439. <https://doi.org/10.1093/NAR/GKL200>

Stegeman, J. J., Goldstone, J. V., & Hahn, M. E. (2010). Perspectives on zebrafish as a model in environmental toxicology. *Fish Physiology*, 29(C), 367–439. [https://doi.org/10.1016/S1546-5098\(10\)02910-9](https://doi.org/10.1016/S1546-5098(10)02910-9)

Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. In *Current Protocols in Bioinformatics: Vol. Chapter 4* (Issue SUPPL. 25, p. Unit 4.10). <https://doi.org/10.1002/0471250953.bi0410s25>

Teramura, A., Kadota, M., Hirase, S., Kuraku, S., & Kikuchi, K. (2024). Silver chimaera genome assembly and identification of the holocephalan sex chromosome sequence. *BioRxiv*, 2024.08.20.608884. <https://doi.org/10.1101/2024.08.20.608884>

Tiznado-Salazar, V. M., Bojórquez-Sánchez, C., Zamora-García, O. G., Márquez-Farías, J. F., Páez-Osuna, F., & Bergés-Tiznado, M. E. (2023). Bioaccumulation of Essential and Potentially Toxic Elements in the Muscle and Liver of the Spotted Ratfish (*Hydrolagus coliei*) From Deep-Sea Waters off the Northern Gulf of California. *Biological Trace Element Research*, 201(5), 2536–2545. <https://doi.org/10.1007/s12011-022-03330-3>

Venkatesh, B., Lee, A. P., Ravi, V., Maurya, A. K., Lian, M. M., Swann, J. B., Ohta, Y., Flajnik, M. F., Sutoh, Y., Kasahara, M., Hoon, S., Gangu, V., Roy, S. W., Irimia, M., Korzh, V., Kondrychyn, I., Lim, Z. W., Tay, B. H., Tohari, S., ... Warren, W. C. (2014). Elephant shark genome provides unique insights into gnathostome evolution. *Nature* 2014 505:7482, 505(7482), 174–179. <https://doi.org/10.1038/nature12826>

Veríssimo, A., Castro, L. F. C., Muñoz-Merida, A., Almeida, T., Gaigher, A., Neves, F., Flajnik, M. F., & Ohta, Y. (2023). An Ancestral Major Histocompatibility Complex Organization in Cartilaginous Fish: Reconstructing MHC Origin and Evolution. *Molecular Biology and Evolution*, 40(12). <https://doi.org/10.1093/MOLBEV/MSAD262>

Vilas-Arrondo, N., Gomes-dos-Santos, A., Pérez, M., Baldó, F., Veríssimo, A., Catarino, D., Machado, A. M., Román-Marcote, E., Bañón, R., Froufe, E., & Castro, L. F. C. (2022). A mitochondrial genome assembly of the opal chimaera, *Chimaera opalescens* Luchetti, Iglésias et Sello 2011, using PacBio HiFi long reads. *Mitochondrial DNA Part B*, 7(3), 434–437. <https://doi.org/10.1080/23802359.2022.2044403>

Weigmann, S. (2016). Annotated checklist of the living sharks, batoids and chimaeras (Chondrichthyes) of the world, with a focus on biogeographical diversity. *Journal of Fish Biology*, 88(3), 837–1037. <https://doi.org/10.1111/JFB.12874>

Supplementary Material

Below is the link to the electronic supplementary material.

[https://www.dropbox.com/scl/fo/3xxegggmyeg5iiejp1gni/APM2_Mpjax-qdyGTElsp9Y?rlkey=pcbl132z623l38kk3i5my5ebc&st=44y66pec&dl=0](https://www.dropbox.com/scl/fo/3xxegggmyeg5iiejp1gni/APM2_Mpjax_qdyGTElsp9Y?rlkey=pcbl132z623l38kk3i5my5ebc&st=44y66pec&dl=0)

Supplementary Table S1. Statistics of *H. colliei* genome assemblies at contig level. The assemblies were performed with several softwares and different configurations.

Supplementary Table S2. Species and accession number of the proteomes used in genome annotation of *H. colliei* genome assembly.

Supplementary Table S3. Full annotation report of *H. colliei* genome assembly provided by Agat tool.

4. Chapter 4 – Chemical Defensome in *Paracentrotus lividus*

4.1. The male and female gonad transcriptome of the edible sea urchin, *Paracentrotus lividus*: Identification of sex-related and lipid biosynthesis genes

André M. Machado*, Sergio Fernández-Boo*, Manuel Nande, Rui Pinto, Benjamin Costas, L. Filipe C. Castro

(*Joint First Authors)

Aquaculture Reports

(DOI: [/10.1016/j.aqrep.2021.100936](https://doi.org/10.1016/j.aqrep.2021.100936))

Aquaculture Reports 22 (2022) 100936



Contents lists available at ScienceDirect

Aquaculture Reports

journal homepage: www.elsevier.com/locate/aqrep

The male and female gonad transcriptome of the edible sea urchin, *Paracentrotus lividus*. Identification of sex-related and lipid biosynthesis genes

André M. Machado^{a,1}, Sergio Fernández-Boo^{a,1}, Manuel Nande^a, Rui Pinto^a, Benjamin Costas^{a,b}, L. Filipe C. Castro^{a,c,*}

^a CIIMAR - Interdisciplinary Centre of Marine and Environmental Research, U. Porto - University of Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos s/n, 4450-208 Matosinhos, Portugal

^b Instituto de Ciências Biomédicas Abel Salazar (ICBAS-UP), U. Porto University of Porto, Rua de Jorge Viterbo Ferreira 228, 4050-313 Porto, Portugal

^c Department of Biology, Faculty of Sciences, U. Porto - University of Porto, 4169-007 Porto, Portugal

ARTICLE INFO

Keywords:

Paracentrotus lividus

Gonad

Sexual development

Lipid biosynthesis genes

ABSTRACT

Paracentrotus lividus is the most abundant, geographically distributed, and desirable echinoid species in Europe. Although economically important, this species has scarce genomic resources available. Here, we produced and comprehensively characterized the male and female gonad transcriptome of *P. lividus*. The *P. lividus* transcriptome assembly has 53,865 transcripts, an N50 transcript length of 1842 bp, and an estimated gene completeness of 97.4% and 95.6% in Eukaryota and Metazoa BUSCO databases, respectively. Differential gene expression analyses yielded a total of 3371 and 3351 up-regulated genes in *P. lividus* male and female gonad tissues, respectively. Additionally, we analyzed and validated a catalog of transcripts pivotal in sexual development and determination (206 transcripts) as well as in biosynthesis and storage of lipids (119 transcripts) in male and female specimens. This study provides a valuable resource and will contribute to the future conservation of the *P. lividus* species as well as its exploitation in an aquaculture setting.

1. Introduction

Paracentrotus lividus, commonly known as the purple sea urchin, is the most abundant echinoid species in Europe, with an overall distribution from the Mediterranean Sea to the eastern Atlantic coast, from Scotland to Southern Morocco, Canary and Madeira Islands (Ghisaura et al., 2016). In the last few years, *P. lividus* populations have continuously decreased due to intense harvesting, habitat destruction, and climate/environmental changes (i.e. temperature, red tides, and ocean acidification) (Bertocci et al., 2012, 2018; Yeruham et al., 2015; Mos et al., 2016; Ohgaki et al., 2019). The main pressure in this species, the intense harvesting, emerges from the high price market of the gonads, which has led to populational collapses in some geographic areas (Fernández-Boo et al., 2012; Ouréns et al., 2015). In an economically context, the purple sea urchin is highly desirable in gourmet markets, not only because their gonads are considered a delicacy but also due to

their high protein and polyunsaturated fatty acid (PUFA) content, especially in omega-3 and omega-6 (Prato et al., 2018; Baião et al., 2019; Rocha et al., 2019). Additionally, the market price of sea urchin gonads is dependent on taste, color, and firmness. For example, while the testis usually has a sweet and milky flavor, the ovaries taste bitter and sour (Phillips et al., 2009, 2010). These characteristics together with the gonadosomatic index of sea urchins establish the market price of the captures. The reproductive period of *P. lividus* is variable, and several factors such as diet, availability of food resources, photoperiod, and water temperature can condition their spawn (Ghisaura et al., 2016). Usually, in the Atlantic coast *P. lividus* spawn is annual (Garmendia et al., 2010; Fernández-Boo et al., 2018), while in the Mediterranean Sea is bi-annual (Selleme and Guillou, 2007). The reproductive apparatus of sea urchin is composed of five gonads with different color patterns between sexes, while males present a yellow-orange pattern, female gonads are red-orange. The gonads are composed of two main types of

* Corresponding author at: CIIMAR - Interdisciplinary Centre of Marine and Environmental Research, U. Porto - University of Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos s/n, 4450-208 Matosinhos, Portugal.

E-mail address: filipe.castro@ciimar.up.pt (L.F.C. Castro).

¹ These Authors contributed equally.

<https://doi.org/10.1016/j.aqrep.2021.100936>

Received 17 September 2021; Received in revised form 29 October 2021; Accepted 1 November 2021

Available online 13 November 2021

2352-5134/© 2021 The Author(s).

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

A.M. Machado et al.

Aquaculture Reports 22 (2022) 100936

cells: germinal cells where the gametes are produced and stored, and somatic cells defined as nutritive phagocytes which are the main storage of nutrients and energetic reserves of the animal (Garmendia et al., 2010; Ghisaura et al., 2016).

Sea urchin aquaculture is based on the enhancement of gonad yield, also known as bulking. This approach consists of the capture of wild animals to improve their gonadosomatic index, gonad sensory attributes, or manipulation of the reproductive cycle to commercialize when wild animals are not available (Walker et al., 2015). Despite multiple attempts to develop intensive aquaculture practices, the poor knowledge of dietary composition for juveniles and the period necessary to reach market size are the main bottlenecks in *P. lividus* production (James et al., 2015; Liu and Chang, 2015). Moreover, sea urchin aquaculture has been hampered by the lack of genomic resources, with some minor examples involving the production of sea urchin triploids to increase the roe size by suppression of gametogenesis and the enhancement of nutritive phagocyte growth (Böttger et al., 2011; Walker et al., 2015). Importantly, to select parental individuals with valuable genetic traits (e.g. gonad size, color, or lipid content), for production purposes, it is crucial to develop *omic* resources for this species.

Decisively, in the last few years, several strategies have been applied to study and comprehend multiple traits with value in the production context (Liu et al., 2005; Wang, Ding et al., 2020). Next-generation sequencing approaches, both at the genomic and transcriptomic level, are now established to explore general biological species features, whether from the commercial or fundamental point of view (Sea Urchin Genome Sequencing, 2006). For instance, transcriptome studies in different sea urchin species have been conducted recently to investigate the genes involved in the synthesis of polyunsaturated fatty acids in *Strongylocentrotus nudus* (Jia et al., 2017; Wei et al., 2019), the sex-related genes in *Mesocentrotus nudus* (Sun et al., 2019) or the response to ocean acidification in *S. purpuratus* (Evans et al., 2017). Other species with less economic value, but highly important to local communities, also have been studied and information regarding genes involved in the development, fertilization, toxin effects, or immune system response against pathogens are now available (Gaitan-Espitia et al., 2016; Laruson et al., 2018). In line with transcriptomic studies, the number of whole-genome sequencing projects in sea urchin species has drastically increased after the massive effort in the genome sequencing of *S. purpuratus* (Sea Urchin Genome Sequencing, 2006; Janies et al., 2016; Kinjo et al., 2018; Davidson et al., 2020). In *P. lividus*, genomic and transcriptomic resources are scarce in public databases. The few studies available were focused on sea urchin development, mainly in the use of embryos as an experimental model for evolutionary and ecotoxicological fields (Gildor et al., 2016; Ruocco et al., 2016; Chassé et al., 2018; Tato et al., 2018; Galasso et al., 2019; Morroni et al., 2019); also, the proteins involved in the attachment by their tube feet were studied for their use in industrial and medical purposes (Pjeta et al., 2020). Regarding the gonadal tissue of adult specimens, only two articles were released, one studying the protein patterns of gonads in males and females at different stages of development (Ghisaura et al., 2016), and another measuring the gene expression level of different pollution biomarkers after metal exposure (Di Natale et al., 2019).

Here, we reported an in-depth transcriptome analysis of the gonads from three adult males and females. These analyses establish the basis for new studies on gonad development, sex differentiation, growth, and also lipid and color traits. Through the functional annotation and phylogenetic analyses, multiple genes pivotal involved in sex determination and differentiation, gametogenesis, lipid metabolism, and PUFA biosynthesis were scrutinized in this species. This study will be highly valuable not only to improve the information of this keystone species in the Atlantic and the Mediterranean Sea, but also to improve and boost the sea urchin aquaculture in Europe.

2. Material and methods

2.1. Experimental setup, RNA extraction, and sequencing

Ten adult sea urchins were collected in July 2018 in Vila Chã beach (41.295160 N, 08.737073 W) by scuba diving (68.6 ± 4.9 mm diameter; 123.10 ± 22.82 g weight; 13.51 ± 2.29% gonadal somatic index) (Fig. 1a). Initially, four females and six males were identified after gonad extraction by visual identification in a light microscope. A small piece of the gonad was immediately stored in an Eppendorf with 1 ml of RNA later (Sigma) at 4 °C for 24 h and then, frozen at -80 °C until RNA extraction. Another piece of the gonad was stored for histological analysis according to Rocha et al. (2019). Sea urchin histological sections were evaluated with light microscopy and all females were in gonadal stage IV and all males in gonadal stage III according to Machado et al. (2019). For RNA extraction, 50–100 mg of gonad tissue was homogenized in 1 ml of trizol (NZY Tech, PT) using a Precellys homogenizer (Bertin Inst., France). RNA was extracted according to the manufacturer's instructions. RNA concentration was measured in a DeNovix DS-11 spectrophotometer (DeNovix Inc, USA) and 4 µg of RNA were subjected to DNase treatment (RQ1 – Promega) according to the manufacturer's instructions. After DNase treatment, RNA was isolated and cleaned using the Total RNA isolation kit (NZY Tech, PT) according to the manufacturer's protocol and diluted in a final volume of 40 µl in mQ water. Finally, the concentration was measured again in a DeNovix DA-11 spectrophotometer. RNA integrity was observed in a 2% agarose gel and the best 3 males and 3 females were selected for RNA-Sequencing. RNA integrity and quantity were evaluated using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The RNA integrity number was > 6.1 in all samples. In the end, six samples (3 females and 3 males) were shipped to Novogene (Honk Kong) Company Limited, and sequenced using Illumina HiSeq-4000 platform (150 × 2bp, paired-end, 30 million sequencing reads).

2.2. Raw data clean up

Initially, the RNA-Seq quality profile of each sample was assessed with FastQC (v.0.11.8) (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). After that, Trimmomatic (v.0.38) (Bolger et al., 2014) was used to trim and drop reads with quality scores below 5, at leading and trailing ends, with an average quality score below 20 in a 4 bp sliding window and with less than 36 bases length. Next, the error correction method, Rcorrector (v.1.0.3) (Song and Florea, 2015), was applied to correct random sequencing errors, with the default settings. In the final stage of the clean-up process, Centrifuge (v.1.0.3-beta) (Kim et al., 2016) software was employed (Fig. 1b). On this approach, all corrected raw reads were screened against the NCBI-nt database (<ftp://ftp.ncbi.nlm.nih.gov/pub/infphilo/centrifuge/data/>) (v.nt_2018_3_3) to obtain a taxonomic classification of each contig, with a minimum hit length of 50. Reads labeled by Centrifuge as non-Echinodermata (NCBI: taxid7586) were considered to be contaminants and excluded from the dataset.

2.3. De novo assembly transcriptome, decontamination, and quality assessment

After quality filtering control, clean reads were first concatenated and then assembled using Trinity (v. 2.8.4) (Grabherr et al., 2011; Haas et al., 2013) with the specific parameter (SS_lib_type RF). To remove possible sources of contamination, all transcripts were blasted against NCBI-nt (Download; 20/11/2019) and Univec (Download; 02/04/2019) databases using Blast-n tool (v. 2.9.0) (Fig. 1b). Transcripts with a match to Echinodermata phylum (NCBI: taxid7586), with an e-value cut-off of 1e-5, identity score of 95%, and minimum alignment length of 100 bp, or without matches at all in NCBI-nt database, were retained. Transcripts matching other taxa than Echinodermata in NCBI-nt and Univec

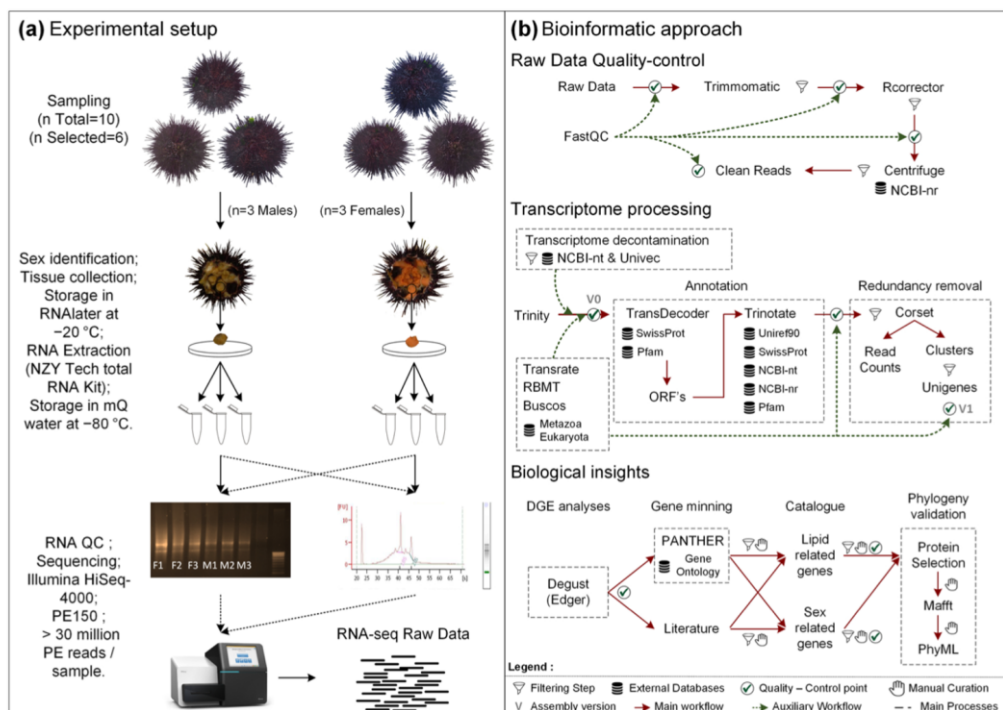


Fig. 1. Sampling and bioinformatic workflow of the study. (a) Sampling and experimental setup used to perform RNA extraction and sequencing. (b) Bioinformatics workflow used to process the raw datasets, perform the differential gene expression and obtain biological insights.

database were considered exogenous to the *P. lividus* transcriptome (Assembly V0) and excluded from this dataset.

2.4. Functional annotation

The Trinotate pipeline (v. 3.0.2) (<https://github.com/Trinotate/Trinotate>) was used to perform the functional annotation of the transcriptome (Bryant et al., 2017). In the first step, the TransDecoder software (v.5.3.0) (<https://transdecoder.github.io/>) was applied to predict the open reading frames (ORFs) with at least 100 amino acids. Next, both amino acid and nucleotide sequences of Assembly V0 were blasted with a blast-n tool (v. 2.9.0) (Altschul et al., 1997) and blast-x/p tools of DIAMOND software (v 0.9.24) (Buchfink et al., 2015), against the non-redundant database of NCBI (NCBI-nr) (Download; 20/01/2020), NCBI-nt (Download; 30/03/2019), Swiss-Prot (Download; 18/02/2020) (The UniProt Consortium, 2016), Uniref90 (Download; 04/09/2019) (Suzek et al., 2007) databases, and searched in Pfam (Download; 18/02/2020) (Punta et al., 2012), eggNOG (Powell et al., 2011), Kyoto Encyclopedia of Genes and Genomes pathways (Kanehisa and Goto, 2000). The final report generated by the Trinotate pipeline was filtered by an e-value cut-off of $1e-5$ (Fig. 1b).

2.5. Transcriptome filtering and redundancy removal

The transcriptome filtering and redundancy removal were performed based on several criteria. First, the functional annotation and ORF prediction of the Assembly V0 version were scrutinized. Thus, all transcripts with at least one blast hit (blast -n, x or p in Uniref90, NCBI-nt or NCBI-nr Databases) in an Echinodermata species or coding to a protein were

collected (Annotated Assembly V0) (Fig. 1b). Second, the raw reads were mapped onto the Annotated Assembly V0 with Bowtie2 (v. 2.3.4.2) (Settings; -no-mixed -no-discordant -end-to-end -all -score-min L,- 0.1,- 0.1) and filtered with Corset (v.1.0.9) (Davidson and Oshlack, 2014) software. While all the transcripts containing at least 10 reads mapping were kept, the transcripts with few reads mapping were considered spurious and discarded. In addition to the filtering, by read coverage, the Corset software clustered transcripts, based on the ratio of shared reads and expression patterns. Finally, and to remove the redundancy, only the longest transcript per cluster (Unigene) was collected (Assembly V1) (Fig. 1b). To evaluate the quality of both transcriptome versions (V0 and V1), several strategies were applied. Completeness and gene content of the transcriptomes were assessed using Metazoa and Eukaryota lineage-specific profile libraries of Benchmarking Universal Single-Copy Orthologs tool (BUSCO v. 3.0.2) (Simão et al., 2015). Accuracy was assessed by the percentage of original clean sequence reads mapped against the Assembly V0 (RMBT) using Bowtie2 (v. 2.3.4.2) (Langmead and Salzberg, 2012) with default settings. Structural integrity and general stats of the transcriptome were examined using TransRate (v. 1.0.3) (Lang-Unnash, 1992) with default settings (Fig. 1b).

2.6. Differential gene expression analyses

Differential gene expression (DGE) analyses were performed using the Unigenes matrix of raw counts, generated by Corset software during the filtering steps, and the Degust (v.4.1.1) platform (Powell, 2019) (<http://degust.erc.monash.edu/>). All Unigenes containing less than 1 CPM (count per million mapped reads) in at least three samples were

A.M. Machado et al.

Aquaculture Reports 22 (2022) 100936

removed from the dataset. In addition, the multidimensional scaling plot was applied to check the variance of the samples. Thereafter, the edgeR (v.3.26.8) package (Robinson et al., 2009) of R (v.3.6.1) was used to perform the DGE analyses between male and female samples of *P. lividus*. During the DGE analyses, the trimmed mean of M-values (TMM) method (Robinson and Oshlack, 2010) was applied to perform the normalization of the values across the samples. Finally, the Unigenes were considered differentially expressed if the values of False Discovery Rate – corrected (FDR) p -value < 0.05 and $\log_2$ [fold change] ≥ 2 . The heatmaps were performed in Heatmapper Expression tool (<http://heatmapper.ca/expression/>), within the Average Linkage clustering method and the Euclidean Distance measurement method (Babicki, Arndt et al., 2016).

2.7. Gene ontology analyses and gene cataloging

To perform the gene ontology (GO) analyses we used the blast-x annotations of the differentially expressed genes (DEGs) against the Uniref90 database. To do that, two datasets were used; firstly, the total number of DEGs; second only the upregulated genes in males and females. Technically, the Uniref90 Id's were subjected to Panther v.15.0 for gene function classification using the sea urchin *Strongylocentrotus purpuratus* as a model organism (Mi et al., 2019). Identification of sex and lipid-related genes was done by searching specific GO keywords (Sex-related terms: reproductive process GO:0022414; reproduction GO:0000003; developmental process GO:0032502; signaling GO:0023052 and Lipid Related terms: biogenesis GO:0071840; metabolic process GO:0008152; catalytic activity GO:0003824) in panther output and trinotate report, as well as, different genes already described in the bibliography as being involved in the sex differentiation and lipid related pathways in the Echinodermata phylum.

2.8. Phylogenetic analysis

Full-length amino acid (aa) sequences were used for phylogenetic analysis, to reduce the impact of nucleotide composition bias. The lipid

and sex-related genes full-length aa sequences were selected accordingly to their relevance within the main pathways previously identified by Panther and after an exhaustive analysis of the related literature (Suppl. Table 1). After collecting the target sequences, orthologous proteins were gathered from the NCBI database, using the blast-p algorithm with an e-value cut-off of $1e-5$ and at least 75% identity in echinoderms (*S. purpuratus*, *Apostichopus japonicus*, *Asterias rubens*, and *Crassostrea gigas*) and vertebrates model species (*Danio rerio*, *Xenopus tropicalis*, *Gallus gallus*, *Mus musculus*, and *Homo sapiens*) (Suppl. Table 1). Multiple alignments were performed with all orthologs sequences using MAFFT v7.402 free software (Katoh et al., 2019) with the L-INS-I algorithm (Katoh and Standley, 2013). Subsequently, the alignments were manually reviewed, and short sequences and gaps were removed. To remove gaps we used the GapStrip/Squeeze v2.1.0 (<http://www.hiv.lanl.gov>) software and all columns of the multi-alignment with 95% nucleotides full filled were kept for the phylogenetic analyses. Phylogenetic trees were constructed using the PhyML 3.0 server (Guindon et al., 2010), using the maximum likelihood method with the default parameters (specific substitution model details in Suppl. Table 2). Later the phylogenetic trees were visualized by Dendroscope (Huson et al., 2007) and rooted with the sequences of the mollusk, *C. gigas*.

3. Results and discussion

3.1. Histological analysis

In the case of females, the gonads presented an advanced stage of development being two of them at stage III (F1, F2) and one at stage IV (F3) according to the scale of Byrne (1990). Stage III is considered a premature stage. The nutritive phagocytes are displaced from the central position to the periphery by the large oocytes (Fig. 2a–c). Stage IV is a mature stage with ovaries closely-packed, most of them with a diameter higher than 90 μ m. All males were in stage III of maturation. This stage is characterized by mature testes packed with spermatozoa (S) and nutritive phagocytes (NP), clearly limited to the periphery (Fig. 2d–f). The

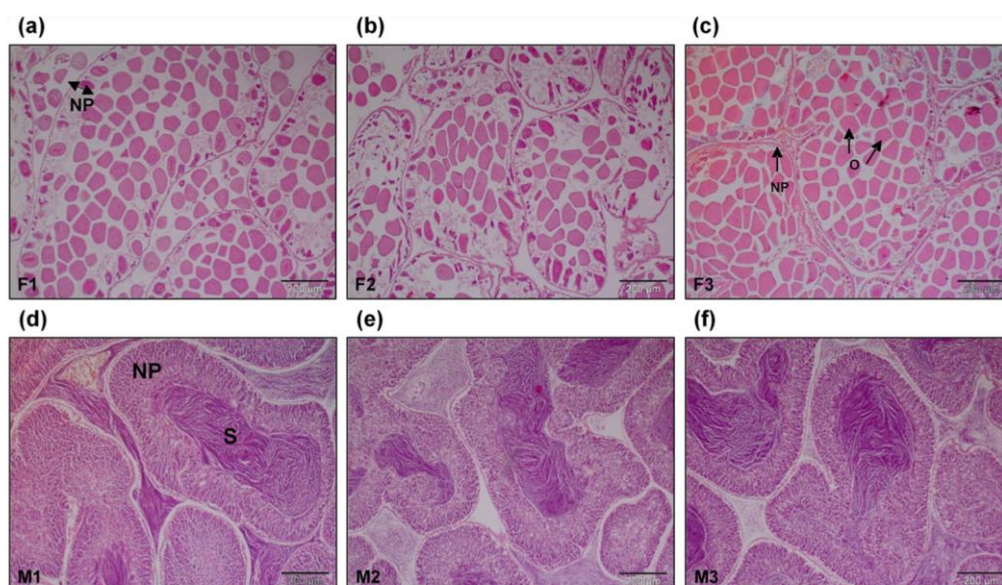


Fig. 2. Histological sections of gonads used for transcriptome sequencing. Females (a–c) were in stage III (a, b) and IV (c) of maturation. All males (d–f) were in stage III of maturation. Scale bar (200 μ m). NP: nutritive phagocyte; O: oocyte; S: spermatozoa.

collection of individuals, in similar stages of maturation, allowed us to obtain a comparative dataset between males and females.

3.2. Sequencing data

From the ten *P. lividus* specimens collected, six (three males (M1, M2, M3) and three females (F1, F2, F3)) were selected to perform gonadal RNA-Seq sequencing. The RNA-Seq approach generated a total of 204,865,648 PE reads, which after validation with several quality-control and filtering software, were reduced to 203,895,472. Despite the conservative approach used, to perform the clean-up, a high percentage of raw reads (99.53% of the initial datasets) showed Phred score $\geq$ Q20 which indicates the high quality of the initial dataset. At the end of this process, all the clean raw reads were deposited to the NCBI database and can be consulted under the BioProject accession: PRJNA625933 (Table 1a).

3.3. The de novo assembly and filtering

To build the transcriptomic reference of *P. lividus*, we applied *de novo* assembler Trinity. Briefly, the six samples of *P. lividus* were pooled together and inputted in the Trinity assembler as a unique dataset. Next, the Trinity software selected about 44,665,675 (21.91%) reads during

the normalization stage and produced the first draft assembly of the transcriptome. Importantly, this version of the transcriptome assembly was submitted to an extensive quality control against the NCBI-nt and UniVec databases, which allowed to remove sources of contaminations and to build the first version of the gonadal transcriptome of *P. lividus* (Assembly V0) (Table 1b). Unexpectedly, this transcriptome version showed a huge number of transcripts (more than 1 M), an N50 transcript length of 720 bp, and a total length size of 756,496,790 bp. Biologically, these values can be explained by several factors. For instance, certain features, such as interspecific variation and heterozygosity can impact the building of *de novo* assembly references. On the other hand, high repeat content can confuse the *de novo* assemblers leading to high rates of fragmentation, spliced transcripts in many sub-transcripts, and production chimeric or artefactual transcripts (Lima et al., 2017). These characteristics were underlined in the genome projects of the sea urchins, *S. purpuratus*, *Hemicentrotus pulcherrimus*, and *Lytechinus variegatus*, in which the authors have found high levels of heterozygosity as well as repeat content and several difficulties in the building of the genome references. To remove the redundancy and mitigate the impact of these values in further analyses, we next applied several filtration steps. These filters have used three main approaches, codifying status, functional annotation, and read coverage of the transcripts.

The codifying status of the transcripts was determined with the

Table 1

General statistics of transcriptome assembly and annotation process. a) Stats of raw data quality control. b) Stats of transcriptome assembly and annotation.

a.				
Samples	Raw sequencing reads	Excluded reads (Trimmomatic Filter)	Excluded reads (Centrifuge Filter)	Final number of reads
F1	37,221,549	11,248 (0.03%)	191,701 (0.52%)	37,018,600 (99.45%)
F2	33,469,876	3293 (0.01%)	164,797 (0.49%)	33,3017,86 (99.50%)
F3	37,324,320	4086 (0.01%)	235,280 (0.63%)	37,084,954 (99.36%)
M1	31,156,243	5221 (0.02%)	89,786 (0.29%)	31,061,236 (99.70%)
M2	30,472,104	2811 (0.01%)	91,064 (0.30%)	30,378,229 (99.69%)
M3	35,221,556	5023 (0.01%)	165,866 (0.47%)	35,050,667 (99.51%)
Totals	204,865,648	31,682 (0.02%)	938,494 (0.46%)	203,895,472 (99.53%)
b.				
Input Data	ALL Datasets			
Selected reads during normalization stage (Trinity)	44,665,675 (21.91%)			
Transcriptome Assembly Versions	Assembly V0	Assembly V1		
Total number of Transcripts	1,310,553	53,865		
n50 transcript length (bp)	720	1842		
n90 transcript length (bp)	268	665		
Largest transcript (bp)	13,405	13,405		
Number of transcripts over 1k nn	161,146	26,773		
Total Assembled bases (Mb)	756,496,790	73,102,927		
GC %	0.37	0.41		
Back Mapping Reads %	95.40	85.09		
BUSCO (Euk;Met) ^a	-			
Complete (%)	93.8/ 93.0	92.4/ 92.4		
Single (%)	8.3/ 9.4	54.1/ 57.4		
Multi (%)	85.5/ 83.6	38.3/ 35.0		
Fragment (%)	5.6/ 5.1	5.0/ 3.2		
Missing (%)	0.6/ 1.9	2.6/ 4.4		
Total Busco Groups (%)	99.4/ 98.1	97.4/ 95.6		
Functional Annotation	-			
No. of transcripts with ORF	138,992	35,723		
Blast-x hits in Ncbi-nt	122,159	31,257		
Blast-x/ p hits in Ncbi-nr	149,473	34,584		
Blast-x/ p hits in Uniref90	154,859	35,692		
Blast-x/ p hits in SwissProt	93,187	22,939		
N° of transcripts with GO terms (blast/ PFAM)	95,252	23,478		
N° of transcripts with Kegg Pathways	78,288	18,959		
N° of transcripts with eggNOG/ COG	73,527	17,865		
No. of transcripts with PFAM	80,094	21,334		
No. Transcripts in > 1 database	206,543	49,048		

Euk: Dataset with 303 genes of Eukaryota library profile.

Met: Dataset with 978 genes of Metazoa library profile.

^a (Euk/ Met).

A.M. Machado et al.

Aquaculture Reports 22 (2022) 100936

TransDecoder software. From 1,310,553 initial transcripts, only 138,992 transcripts codify to an ORF with 100 or more amino acids. The amino acid and nucleotide sequences were then functionally annotated using a set of different databases (Fig. 3a). After searches in several databases, the Uniref90, NCBI-nt and NCBI-nr databases were selected to filter the transcriptome, and about 185,644 transcripts were found to have at least one blast hit in an echinoderm species. Thus, the Annotated version of the transcriptome (Annotated Assembly V0), was obtained through the sum of coding transcripts and transcripts with blast hit but without an ORF (225,700 transcripts - 138,992 protein-coding transcripts; 86,708 additional transcripts with blast hit). Subsequently, the 225,700 transcripts were inputted in the Corset software to remove all transcripts with scarce evidence of read mapping and to cluster the remaining transcripts based on multi-mappings. In the end, Corset grouped 141,138 transcripts (> 10 reads mapping) in the 53,865 clusters/Unigenes (Assembly V1). Both versions of the transcriptome assembly as well the annotation reports and open read frames can be consulted in Figshare digital repository (<https://doi.org/10.6084/m9.figshare.16480542.v1>).

Overall, this strategy has significantly reduced the size and the redundancy of the transcriptome assembly and is similar to previous analyses with sea urchins transcriptomes (Chen et al., 2015; Gaitán-Espitia et al., 2016; Gaitán-Espitia and Hofmann, 2017; Wong et al., 2019; J. Zhang et al., 2019; Zhang et al., 2019a, 2019b; Shi et al., 2020). The final version (Assembly V1) of *P. lividus* transcriptome had only 4.11% of the initial number of transcripts (53,865), an N50 transcript length of 1842, more than double the first version, and 1/10 of the initial transcriptome length (Table 1b). Importantly, both transcriptome versions (V0 and V1) were carefully inspected, assessed, and compared with three methods.

First, the gene content was analyzed using the BUSCO tool. Searching the two libraries profiles, Eukaryota and Metazoa, we observed high completeness of the transcriptome assemblies. While in V0, 99.4% and 98.1% of the total gene groups were found in both databases, in the V1

we identified 97.4% and 95.6% respectively. Importantly, all BUSCOs analyses showed a high percentage (> 92%) of genes classified as complete, and almost no gene content was lost during the filtration process. On the other hand, the number of missing genes in both versions per database is remarkably low (V0, Eukaryota - 0.6%, Metazoa - 1.9%; V1, Eukaryota - 2.6%, Metazoa - 4.4%). Comparing these values with others in the literature, we can conclude that *P. lividus* gonadal transcriptome is one of the most complete among sea urchins reported to date (Table 1b). For instance, in the sea urchins *Strongylocentrotus intermedius* (Zhan et al., 2019), in *Echinometra* sp. embryos (Uthicke, Deshpande et al., 2019) and in giant red sea urchin *Mesocentrotus franciscanus* (Gaitán-Espitia and Hofmann, 2017), *de novo* assemblies achieved 82.5%, 91%, 88% of complete genes in the eukaryotic database, respectively.

Secondly, we performed a structural comparison between V1 and V0 assemblies. Fig. 3b showed a clear decrease in the percentage of short contigs (< 600 bp) from the V0 (43.7%) to the V1 (22.4%) of the transcriptome. Consistently, in the remaining classes, the V1 version always showed higher percentages of transcripts than the V0 version, indicating that small contigs were largely affected by the filtration steps, while the percentages of the large contigs remained almost unmodified.

In the final test, the rate of back mapping reads to V0 and V1 transcriptomes were analyzed. As expected, the RBMT decrease 10,31% from the V0 to the V1 (Table 1b). Using the structural and RBMT analyses it is possible to conclude that a part of the reads not mapped in the V1 version belongs to small and fragmented contigs removed during the filtration step. This result is coherent with the biological features of sea urchins, with the presence of high repetitive regions leading to high percentages of fragmented and spliced transcripts in several sub-transcripts. Lastly, the top 10 *Echinodermata* species with the highest number of transcripts matching the NCBI-nr database in both V0 and V1 Assembly versions are shown in Fig. 3c. As can be consulted, both versions displayed a very similar distribution.

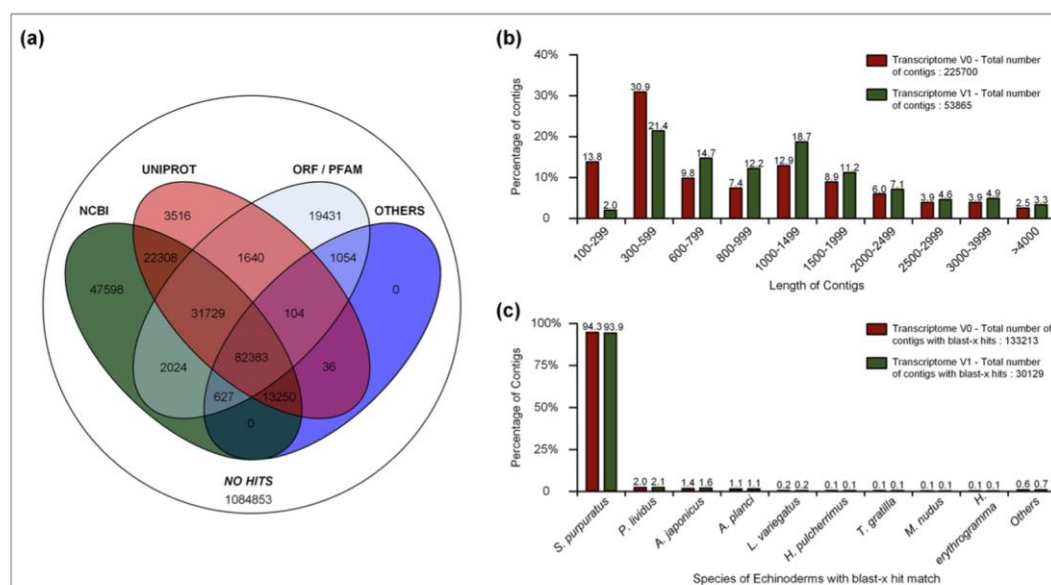


Fig. 3. Functional annotation and clean-up analyses of the V0 and V1 versions of transcriptome assembly of *P. lividus*. (a) Venn diagram of the transcripts (V0 transcriptome) codifying to an open reading frame or matching with NCBI, UNIPROT, and PFAM databases. (b) Length distribution of the transcripts in versions V0 and V1 of the transcriptome assembly. (c) Blast-x analysis of the V0 and V1 versions of the *P. lividus* transcriptome assembly.

3.4. Functional annotation

The functional annotation was performed to the V0 version of the gonadal transcriptome and thereafter gathered to the V1 version. Overall, the V0 version showed only 15.75% (206,543) of the total transcripts annotated against at least one database, and 6.29% annotated in all databases (Fig. 3a; Table 1b). On the other hand, the V1 dataset of Unigenes presented a ratio 91.05% (49,048) of annotated transcripts (Table 1b, Suppl. Table. 3). Once the remaining analyses explored the V1 assembly the following results of functional annotation will be focused on this version. In KEGG, eggNOG, and Pfam databases we found between 33.17% and 39.61% of the Unigenes. On the other hand, in the general databases of NCBI-nt, NCBI-nr, UNIREF90, and Swiss-Prot, 31,257; 34,584; 35,692, and 22,939 matches were found. The blast-x results against the NCBI-nr database showed 34,537 match hits in 615 species. Of these, 30,129 (87.24%) Unigenes matched against species of Echinodermata, 597 against Cnidaria, 189 against Mollusca, 184 against Arthropoda, and 3061 against others Phyla. Among the top 10 Echinodermata species with the highest number of unigenes, the *S. purpuratus* species stand out from the remaining species, with 93.9% of the hits (Fig. 3c). Although several genomes and transcriptomes of sea urchins are available (e.g. *Sea Urchin Genome Sequencing, 2006; Cameron et al., 2015; Cary et al., 2018; Kinjo et al., 2018; Davidson et al., 2020*), to the date of these analyses only *S. purpuratus* species had a uniformized genome annotation, CDS and proteins available in NCBI databases. In addition, the *P. lividus* and *S. purpuratus* are closely related species in terms of phylogeny, which makes this result expected (Montgiardino Koch et al., 2018). The remaining eight species of Fig. 3c have between 2% and 0.1% of the hits and comprise one sea star, *Acanthaster planci* (Class: Asteroidea), one sea cucumber, *A. japonicus* (Class: Holothuroidea), and seven other sea urchins (Class: Echinoidea).

3.5. Differential expression analyses

The differential gene expression analyses between the male and female samples were performed in the Degust platform. After submitting the matrix raw of counts to the platform, all Unigenes were filtered by the CPM (1 CPM in at least 3 samples). This filter allowed gathering only Unigenes with strong read support (17,581). Next, the variance was evaluated across the samples using the multidimensional plot. Importantly, both female and male samples are grouped in two independent clusters. Moreover, the female samples presented higher variance than male samples (Suppl. Fig. 1). The DGE analyses yielded a total of 6722 Unigenes with $p\text{-value} < 0.05$ and $\log_2[\text{fold change}] \geq 2$. Overall, were found 3371 Unigenes up-regulated in *P. lividus* males and 3351 up-regulated in *P. lividus* females (Suppl. Fig. 2, Suppl. Table. 4). Moreover, 163 and 162 Unigenes were found to be female and male-specific, respectively (Suppl. Table. 5, 6). Regarding the annotation status of the DEG Unigenes (DEGs), about 5278 DEGs codifying to an ORF, and 6267 have a match hit annotation in UniProt or NCBI databases.

3.6. Gene ontology analyses

The ontology analyses included all DEG and were mainly focused on the Biological Process category. These analyses showed most of the genes related to the cellular process (35%), metabolic process (23.8%), and biological regulation (14.9%), while only a few were directly involved in reproduction (0.8%) or reproductive process (0.8%) (Suppl. Fig. 3). When the gene ontology annotation was determined per sex, once again most of the genes were related with the cellular process (32.6% F vs 37.3% M), metabolic process (24% F vs 23.7% M), and biological regulation (15.8% F vs 14% M), while reproduction has low values (0.3% F vs 1.2% M) (Suppl. Fig. 4). Notwithstanding, the reproduction gene ontology term has 4 times more expressed genes in males than females. In terms of the developmental process the opposite occurs, with more DEGs present in females (4% F vs 1.8% M) (Suppl.

Table 7; Suppl. Fig. 4).

Next, to identify candidate genes involved in sex differentiation, determination, and gonad development of *P. lividus*, the DEG genes were filtered using key GO terms and literature in the Echinodermata phylum. According to Panther classification, only 32 entries were associated with reproduction (GO:0000003) and reproductive process (GO:0022414) gene ontology terms. When the Trinotate report was scrutinized, a higher number of genes linked to gonad development and sex differentiation (GO:0008584, GO:0007548) were found. The Spermatogenesis category (GO:0007283) was the most abundant category with 115 entries all of them found up-regulated in males, corresponding to 57.21% of the total sex-related genes found in males (Suppl. Table 8). The second category with more entries was Spermatid development (GO:0007286) with 2 up-regulated genes in females and 49 in males. Then, oogenesis (GO:0048477) appeared with 7 entries in females and 10 in males. After that, males gonad development (GO:0008584) with 13 entries 6 in females and 7 in males; fertilization (GO:0009566) with 11 entries, all of them in males; oocyte maturation (GO:0001556), 9 entries, 1 in females and 8 in males; males meiosis (GO:0007140), 5 entries, 1 in females and 4 in males; gonad development (GO:0008406), 4 entries, 2 in females and 2 in males; females gonad development (GO:0008585), 3 entries, all in males, and finally oocyte development (GO:0048599) with 2 entries, one in females and another in males. The highest number of entries were related to lipid metabolic process (GO:0006629) with 30 entries, 15 in females and 15 in males; lipid transport (GO:0006869) with 21 hits, 14 in females and 7 in males was the second most abundant group; cholesterol metabolic process (GO:0008203) with 13 hits, 7 in females and 6 in males; lipid glycosylation (GO:0030259) with 12 hits all in females and lipid storage (GO:0019915) with 10 hits, 5 in females and 5 in males were the most abundant groups. Additionally, some genes involved in fatty acid elongation (GO:0030497) and the fatty acid elongase process (GO:0009922) were found (Suppl. Table 9).

Initially, the total number of genes found in female and male gonads was similar. However, after filtering a higher number of genes related to sex and reproduction was detected in males (Fig. 4a, Suppl. Table 8). Moreover, the DGE present in males was 3.28 times higher than in females (154 vs 47). In contrast, in lipid-related genes a total of 119 genes were obtained, 48 hits of genes up-regulated in males and 68 in females. In this case, as expected, the highest number of genes involved in fatty acid production was found in sea urchin females (Fig. 4b, Suppl. Table 9).

3.7. Catalog sex and lipid-related unigenes

3.7.1. Unigenes involved in sexual development/determination

As previously described, sea urchins are diploid species with heteromorphic chromosome sex mechanisms of the XY genes (Lipani et al., 1996). A characteristic of *P. lividus* is the lower number of chromosomes ($2n = 36$) in comparison with other echinoid species ($2n = 42-44$) (Lipani et al., 1996). Usually, sex determination in gonochoric species such as *P. lividus* involves a robust transcriptional regulation (Nef et al., 2005). Interestingly, in the majority of these species, the male specimens have a higher number of genes being expressed in the gonads (Nef et al., 2005; Tao et al., 2013; Teaniniuraitemoana et al., 2014; Shen et al., 2020). Notwithstanding, there are some species where the opposite also occurs (González-Castellano et al., 2019; Piprek et al., 2019).

Sex differentiation includes several processes regulated by a high variety of genes and transcription factors. Among these genes, Sox genes are some of the well described in the literature as key factors for sex determination. In *P. lividus* we found five Sox genes differentially expressed (Suppl. Table 8). While Sox30 (DN1680_c0_g1_i5) was only found up-regulated in males, other Sox genes such as SoxB1, Sox21 and Sox4 (DN7136_c0_g1_i2; DN7136_c0_g2_i4; DN18728_c0_g1_i1) were uniquely found up-regulated in female gonads. Sox30, also known as Sex determination factor Y, has been described as a key regulator of

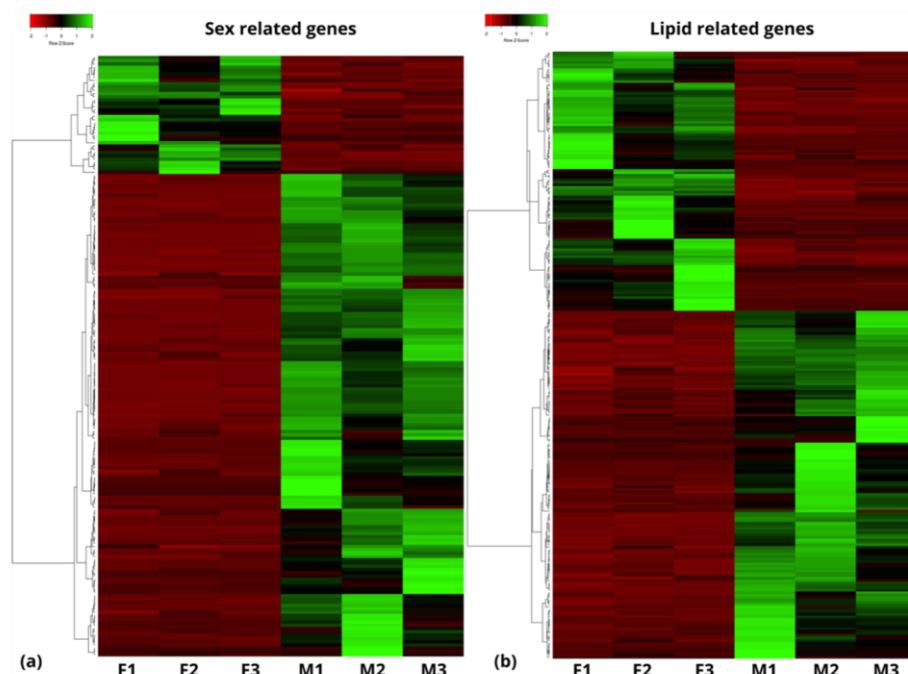


Fig. 4. Heatmap of up and down-regulated genes. (a) Sexual development and determination genes. (b) Genes related to the biosynthesis and storage of lipids.

transcription in mouse spermatogenesis (Zhang et al., 2018), suggesting an important role in *P. lividus* testis development. On the other hand, the transcription factor SoxB1, one of the most expressed genes in *P. lividus* female gonads, is described as predominantly expressed in females and can have a function of oocyte protection and regulating the expression of ZP proteins (Yue et al., 2015). Despite, the finding of several Sox genes, other important genes such as Sox9, a major transcription factor for testis development (Zhang et al., 2018), were not found in *P. lividus*. Although unexpected, this result is coherent with the literature, where this and other important genes did not show differential gene expression in testis or ovaries of sea urchins' species (Sun et al., 2019). Moreover, in Echinodea class seems to exist different Sox genes being differentially expressed in gonads. Interestingly, in *M. nudus* species four different Sox genes such as Sox1, 3, 6, and 9 were found (Sun et al., 2019) and none of them were found in *P. lividus* transcriptome. These differences can be explained by the differential developmental stages of the gonads in the different studies (Yue et al., 2015).

The Dmrt1 gene (DN11344_c0.g1.i5), usually involved in testis development, also was found up-regulated in *P. lividus* testis. As in many other species, including sea urchins, Dmrt1 is described as playing a major role in sex determination and differentiation in several vertebrate and invertebrate species (Zhang and Zarkower, 2017; Nagasawa et al., 2019; Sun et al., 2019). Consistently with the report of (Sun et al., 2019) in *M. nudus* species, also the Nanos (DN9673_c0.g1.i2) gene in *P. lividus* is up-regulated in female gonads. Although Nanos has a key role in embryonic development in sea urchin (Fujii et al., 2006), in adult ovaries has a pivotal role in proliferation and survival of germline stem cells and cyst development (Forbes and Lehmann, 1998; De Kauckelaere et al., 2018). Additionally, several genes related with the Notch and Delta pathway (e.g. Neurogenic locus notch homolog protein 1 and 2 (DN2510_c0.g1.i1; DN17254_c0.g1.i4); notch gene (DN9493_c1.g2.i3), notch ligand (DN44109_c0.g2.i3) and Delta gene (DN44109_c0.g2.i3)

were found in female gonads. While in adults this pathway is an essential regulator of cell proliferation during development and oogenesis (Feng et al., 2014; Irlles et al., 2016; Sun et al., 2019; J. Zhang et al., 2019; Zhang et al., 2019a, 2019b), in sea urchin embryos is related to the development process regulation (Materna and Davidson, 2012). Another gene involved in the developing gonads and granulosa cells in adult mice is the GATA-type Zinc finger protein 1 (GLP-1) (Li et al., 2007; Pangas and Rajkovic, 2015). This gene, not characterized in *S. purpuratus*, also was expressed in *P. lividus* females (DN31305_c0.g1.i2) and previously described in *M. nudus* (Sun et al., 2019). Importantly, this gene seems to have a crucial role in the regulation of oocyte meiosis and formation of primordial follicles, once the GLP-1 knock-out in both males and females of mice results in infertility (Li et al., 2007).

Regarding the spermatogenesis regulation, several genes such as DMC1 (DN19900_c0.g1.i5), Spo11 (DN82908_c0.g1.i1), Hsd3 (DN1732_c1.g1.i2) and two SPATA genes, 24 and 45 (DN632_c1.g1.i2; DN847_c0.g1.i3), were found up-regulated in male gonads. DMC1 is a recombinase essential for meiosis in the male testis, regulating the homologous recombination of chromosomes. Depletion of DMC1 leads to malformation of sperm or aneuploidy spermatocytes in several species (Chen et al., 2016). DMC1/LIM15 (DN19900_c0.g1.i5) from *S. purpuratus* was found in the testis of *P. lividus* indicating that as in other eukaryotes, the expression of DMC1 is essential for the development of the spermatocytes in sea urchin. Also, Spo11 is essential in meiotic recombination by generating double-strand breaks and it is required for meiotic synapsis (Romanienko and Camerini-Otero, 2000). Two genes of the SPATA family were found in the transcriptome, SPATA24, and SPATA45. The spermatogenesis-associated (SPATA) family consists of several genes with critical roles in spermatogenesis. For instance, the knock-out of these genes in humans led to several problems in sperm motility, sperm production, and germ cell development (Sujit et al., 2020). Finally, several testis-specific serine/threonine

A.M. Machado et al.

Aquaculture Reports 22 (2022) 100936

protein kinases (Tssk) (e.g. *tssk1*, 4 and 5; DN1880_c1.g1.i12; DN20093_c0.g1.i3; DN2835_c2.g1.i6) were detected in the gonadal transcriptome, mainly associated to the male gonad. In general, these genes are associated with spermatogenesis regulation and they are widespread across the Eukaryota domain, including sea urchins (Wang et al., 2016; Sun et al., 2019).

3.7.2. Lipid Unigenes determination

The main components during gametogenesis in *P. lividus* female gonads are proteins, but lipids also play a remarkable role in the viability of gametes (Sanna et al., 2017). Lipids, such as phospholipids (PL) and cholesterol, are structural components of cell membranes and the key to somatic growth (Liu et al., 2007). It increases the synthesis and storage of lipids during gonadal development, reaching maximum values in stage IV of females and stage III of males (final stage), mainly for energy reservoirs (Sanna et al., 2017).

Several genes involved in the biosynthesis of long-chain polyunsaturated fatty acid (LC-PUFA) during gametogenesis, such as putative fatty acid elongation protein 3 (up-regulated in males) (DN12877_c0.g1.i2), elongation of very long-chain fatty acids protein 6 (DN6887_c0.g2.i2), or fatty acyl desaturases A (Fads A) (DN230613_c0.g1.i7), were identified in the *P. lividus* transcriptome. Fads A mediates the introduction of unsaturation (double bond) into a fatty acyl chain (Guillou et al., 2010), showing $\Delta 5$ -desaturated activity in which it uses 20: 3n-6 and 20: 4n-3 to transform into ARA and EPA (Kabeya et al., 2017). LC-PUFA biosynthesis activity was also reported in other sea urchin species such as *S. intermedius* (Han et al., 2019). The gene expression of Fads A in the gonad is congruent with ARA and the EPA as the most abundant LC-PUFAs in the *P. lividus* gonad (Sanna et al., 2017). Different acyl-CoA ligase family members involved in long-chain-fatty-acid-CoA ligase (LC-PUFA-acid CoA ligase) such as LC-PUFA CoA ligase 1 and 6 (DN12809_c0.g1.i8 and DN79126_c1.g1.i1 respectively) were up-regulated in gonads of females while the LC-PUFA CoA ligase 3, 4 and 5 (DN3246_c0.g1.i1, DN3764_c0.g1.i13, and DN12660_c0.g1.i4) were found up-regulated in gonads of males (Suppl. Table 9).

Another gene involved in the LC-PUFA incorporation into the mitochondria, mainly through the hydrolysis of triglycerides (García-Rincón et al., 2016), was carnitine O-palmitoyltransferase 2 (DN25985_c0.g2.i1). This Unigene showed higher expression in female gonads than in males of *P. lividus*. Consistently, this gene showed crucial activity in transcriptome analysis for fatty acid metabolism of other sea urchins such as *S. intermedius* (Wang et al., 2019).

Furthermore, PUFAs play a fundamental role as the main components of complex molecules such as phospholipids or triglycerides. An example of intermediate formation products is found in the sn1-specific diacylglycerol lipase beta that catalyzes the hydrolysis of arachidonic acid (AA) esterified diacylglycerol (DAG) to produce the main endocannabinoid, 2-arachidonoylglycerol (2-AG), which can be further cleaved by downstream enzymes to release arachidonic acid (AA) for cyclooxygenase (COX) mediated eicosanoid production. Also, DAG is an intermediate in the glycerolipid and glycerophospholipids metabolism for the biosynthesis of TAG and PL. In male gonads, high expression of sn1-specific diacylglycerol lipase beta (DN17156_c0.g2.i1) was detected.

Phospholipids are part of the cell membrane and essential for cell growth (Byrd, 1975). Most of the phospholipids are incorporated through the diet, but *P. lividus* has the capacity for its endogenous biosynthesis (Byrd, 1975). Thus, phosphatidylcholine (PC) is one of the major phospholipids in the sea urchin (Mita et al., 1994). Importantly, the Kennedy pathway route is used for phospholipids (Gibellini and Smith, 2010). Cholinephosphotransferase 1 (chpt1) is the last enzyme in charge of obtaining PC from 1,2-diacyl-sn-glycerol and CDP-choline (EC 2.7.8.2). Differential expression analyses showed cholinephosphotransferase 1 (DN72675_c0.g2.i1) up-regulated in female gonads. In contrast, a significantly lower chpt1 activity was identified in oocytes of

Arbacia punctulata compared to sperm (Ewing, 1973). Furthermore, this study results showed a higher expression of lysophosphatidylcholine acyltransferase 2 (DN14424_c0.g1.i10) in male gonads, as an alternative route for the biosynthesis of PC as a final product. The enzyme encoded by lysophosphatidylcholine acyltransferase 2 plays a role in phospholipid metabolism, specifically in the conversion of lysophosphatidylcholine to phosphatidylcholine in the presence of acyl-CoA (Law et al., 2019). The alternative pathways to obtaining PC between males and females may be related to the composition of PUFA and LC-PUFA in positions sn-2 and sn-1,3. Thus, in females of *P. lividus* gonads, the sn-1,3 position of LC-PUFA such as ARA or EPA determines the use of TAGs in response to acclimatization or involvement in reproduction, respectively (Sanna et al., 2017). Also, the male gametes of the sea urchin use PC and TAG as energy (Mita et al., 1994), while in the gonad, PC is incorporated into the cell structure (Ewing, 1973).

Several genes are involved in TAG biosynthesis, such as diacylglycerol kinase zeta (DN39457_c0.g1.i2), nuclear envelope phosphatase-regulatory subunit 1 (DN10024_c0.g1.i6), microsomal triglyceride transfer protein large subunit (DN3378_c1.g1.i6), and were identified in the gonadal transcriptome. Most of the genes related to triglycerides' metabolism were up-regulated in females gonadal. Thus, TAGs are the highest lipid class in gonads in *P. lividus*, varying seasonally due to lipid turnover after each period (Sanna et al., 2017). Similarly, during the egg development of the sea urchin, *Anthodidaris crassispinus*, TAGs are stored and later used in the early stages of development (Yasumasu, Hino et al., 1984). In contrast, in spermatozoa, the TAGs proportion is low or maybe absent (Mita et al., 1994). This is consistent with the high gene expression of triglycerides genes, related to gonadal development, found in *P. lividus* samples.

Furthermore, several genes related to cholesterol metabolism, transport, and binding such as scavenger receptor class B member 1 (DN1029_c0.g1.i3), NPC intracellular cholesterol transporter 1 (DN7961_c0.g1.i6), low-density lipoprotein (LDL) receptor-related protein 6 (DN74960_c0.g2.i3), and Apolipoprotein D (DN8923_c0.g2) were identified (Suppl. Table 9). These genes were mainly expressed in female gonads and have functions related to membrane transport, and binding of high-density lipoprotein cholesterol (HDL), low-density lipoprotein (LDL), and cholesterol. Moreover, genes such as sterol O-acyltransferase 1 have increased activity in male gonads. This enzyme plays a role in lipoprotein assembly and dietary cholesterol absorption (Yang et al., 1997). In female ovary and *P. lividus* spermatozoa, cholesterol was identified as a relevant lipid component (Mita et al., 1994; López-Hernández et al., 1999).

3.8. Phylogenetic validation

One of the biggest challenges in gene ortholog identification is related to variations of gene repertoire between species, as a consequence of loss, duplication, or appearance of paralogs genes (Altenhoff and Dessimoz, 2009). Here, we phylogenetically validated 12 relevant target genes, annotated in our transcriptome, with their orthologs in other species of echinoderms sea urchin (*S. intermedius*), starfish (*A. rubens*) and sea cucumbers (*A. japonicus*), vertebrate model species of teleosts (*D. rerio*), amphibians (*X. tropicalis*), birds (*G. gallus*), mammals (*M. musculus* and *H. sapiens*) and mollusks (*C. gigas*). This is a powerful method to detect annotation and assembly errors caused during the transcriptome assembly for non-model species (Guang et al., 2021). Genes with functional characterization, validation, and description in the literature for echinoderms were included in the analysis (Fads, PPARs and Gonadotropin-Releasing Hormone-Type) (Kabeya et al., 2017; Tian et al., 2017; Capitão et al., 2020). Our results showed that, all target lipid and sex-related full-length aa sequences clustered into a strongly supported clade of echinoderms with the orthologs of sea urchin (*S. intermedius*), starfish (*A. rubens*), and sea cucumbers (*A. japonicus*) (Suppl. Fig. 5). A similar topology was identified in almost all trees (Suppl. Fig. 5), one group of vertebrates, another of

A.M. Machado et al.

Aquaculture Reports 22 (2022) 100936

echinoderms, and one of mollusks (*C. gigas*). As expected, genes involved in lipid metabolism such as *Fads A* and *chpt1* were clustered in the echinoderm clade with 99% posterior probability support and 100% with the other sea urchin (*S. intermedius*) (Suppl. Table 5). The study of Kabeya et al. (2017) performed the functional characterization of *P. lividus* $\Delta 5$ desaturase (*Fads A*) and validated it by phylogenetic analysis using mollusks, other sea urchins (*S. purpuratus* and *L. variegatus*), and other species of marine echinoderms such as lilies (*Oxycoranthus japonicus*), sea cucumbers (*Sclerodactyla briareus*) and starfish (*Patiria miniata*). Supported by the results of Kabeya et al. (2017), we were able to obtain a similar result in our phylogenetic analysis that validates the functional annotation of key target genes in our transcriptome.

4. Conclusion

To conclude, the gonad transcriptome from both sexes of sea urchins led to the identification of a larger number of genes involved in sex determination, sex differentiation, gonad development, spermatogenesis and oogenesis, and also in several pathways of lipid metabolism, fatty acid elongase process, lipid storage, and cholesterol metabolic process. The results obtained in this work will provide useful data on *P. lividus* sea urchin gonad tissue and it will contribute to the future of the conservation of the species and the exploitation for aquaculture purposes.

CRedit authorship contribution statement

André M. Machado: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft. **Sergio Fernández-Boo:** Data curation, Formal analysis, Investigation, Methodology, Visualization, Resources, Writing – original draft. **Manuel Nande:** Formal analysis, Investigation, Methodology, Visualization, Writing – review & editing. **Rui Pinto:** Investigation, Methodology, Visualization, Writing – review & editing. **Benjamin Costas:** Methodology, Validation, Writing – review & editing. **L. Filipe C. Castro:** Conceptualization, Methodology, Validation, Supervision, Project administration, Resources, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the project CRAGIAMP – PTDC/BIA-BQM/30232/2017, co-financed by COMPETE 2020, Portugal 2020, and the European Union through the ERDF. This research was also supported by national funds through FCT – Foundation for Science and Technology within the scope of UIDB/0443/2020 and UIDP/04423/2021. AMM is funded by Foundation for Science and Technology (DFA/BD/8069/2020).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.aqrep.2021.100936.

References

- Altenhoff, A.M., Dessimoz, C., 2009. Phylogenetic and functional assessment of orthologs inference projects and methods. *PLoS Comput. Biol.* 5 (1), e1000262.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25 (17), 3389–3402.
- Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J.R., Maciejewski, A., Wishart, D.S., 2016. Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res.* 44 (W1), W147–W153.
- Baião, L.F., Rocha, F., Costa, M., Sá, T., Oliveira, A., Maia, M.R.G., Fonseca, A.J.M., Pintado, M., Valente, L.M.P., 2019. Effect of protein and lipid levels in diets for adult sea urchin *Paracentrotus lividus* (Lamarck, 1816). *Aquaculture* 506, 127–138.
- Bertocci, I., Domínguez, R., Freitas, C., Sousa-Pinto, I., 2012. Patterns of variation of intertidal species of commercial interest in the Parque Litoral Norte (north Portugal) MPA: comparison with three reference shores. *Mar. Environ. Res.* 77, 60–70.
- Bertocci, I., Blanco, A., Franco, J.N., Fernández-Boo, S., Arenas, F., 2018. Short-term variation of abundance of the purple sea urchin, *Paracentrotus lividus* (Lamarck, 1816), subject to harvesting in northern Portugal. *Mar. Environ. Res.* 141, 247–254.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120.
- Böttger, S.A., Eno, C.C., Walker, C.W., 2011. Methods for generating triploid green sea urchin embryos: an initial step in producing triploid adults for land-based and near-shore aquaculture. *Aquaculture* 318 (1), 199–206.
- Bryant, D.M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M.B., Payzin-Dogru, D., Lee, T.J., Leigh, N.D., Kuo, T.-H., Davis, F.G., Bateman, J., Bryant, S., Guzikowski, A. R., Tsai, S.L., Coyne, S., Ye, W.W., Freeman Jr., R.M., Peshkin, L., Tabin, C.J., Regev, A., Haas, B.J., Whitel, J.L., 2017. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* 18 (3), 762–776.
- Buchnik, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12 (1), 59–60.
- Byrd, E.W., 1975. Phospholipid metabolism following fertilization in sea urchin eggs and embryos. *Dev. Biol.* 46 (2), 309–316.
- Byrne, M., 1990. Annual reproductive cycles of the commercial sea urchin *Paracentrotus lividus* from an exposed intertidal and a sheltered subtidal habitat on the west coast of Ireland. *Mar. Biol.* 104 (2), 275–289.
- Cameron, R.A., Kudrarkar, P., Gordon, S.M., Worley, K.C., Gibbs, R.A., 2015. Do echinoderm genomes measure up? *Mar. Genom.* 22, 1–9.
- Capitão, A., Lopes-Marques, M., Páscoa, I., Ruivo, R., Mendiratta, N., Fonseca, E., Castro, L., Santos, M.M.J.E., 2020. The Echinodermata PPAR: functional characterization and exploitation by the model lipid homeostasis regulator tributyltin. *Environ. Pollut.* 263 Pt B, 114467.
- Cary, G.A., Cameron, R.A., Hinman, V.F., 2018. EchinoBase: tools for echinoderm genome analyses. *Methods Mol. Biol.* 1757, 349–369.
- Chassé, H., Aubert, J., Boulben, S., Le Corguillè, G., Corre, E., Cormier, P., Morales, J., 2018. Translational analysis at the egg-to-embryo transition in sea urchin. *Nucleic Acids Res.* 46 (9), 4607–4621.
- Chen, J., Cui, X., Jia, S., Luo, D., Cao, M., Zhang, Y., Hu, H., Huang, K., Zhu, Z., Hu, W., 2016. Disruption of *dmc1* produces abnormal sperm in medaka (*Oryzias latipes*). *Sci. Rep.* 6, 30912.
- Chen, Y., Chang, Y., Wang, X., Qiu, X., Liu, Y., 2015. De novo assembly and analysis of tissue-specific transcriptomes revealed the tissue-specific genes and profile of immunity from *Strongylocentrotus intermedius*. *Fish. Shelf. Immunol.* 46 (2), 723–736.
- Davidson, N.M., Oshlack, A., 2014. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biol.* 15 (7), 410.
- Davidson, P.L., Guo, H., Wang, L., Berrio, A., Zhang, H., Chang, Y., Soborowski, A.L., McClay, D.R., Fan, G., Wray, G.A., 2020. Chromosomal-level genome assembly of the sea urchin *Lytechinus variegatus* substantially improves functional genomic analyses. *Genome Biol. Evol.* 12 (7), 1080–1086.
- De Keuckelaere, E., Hulpiau, P., Saey, Y., Bex, G., van Roy, F., 2018. Nanos genes and their role in development and beyond. *Cell. Mol. Life Sci.* 75 (11), 1929–1946.
- Di Natale, M., Bennici, C., Blondo, G., Masullo, T., Monastero, C., Tagliavia, M., Torri, M., Costa, S., Ragusa, M.A., Cuttitta, A., Nicosia, A., 2019. Aberrant gene expression profiles in Mediterranean sea urchin reproductive tissues after metal exposures. *Chemosphere* 216, 48–58.
- Evans, T.G., Pespeni, M.H., Hofmann, G.E., Palumbi, S.R., Sanford, E., 2017. Transcriptomic responses to seawater acidification among sea urchin populations inhabiting a natural pH mosaic. *Mol. Ecol.* 26 (8), 2257–2275.
- Ewing, R.D., 1973. Cholinephosphotransferase activity during early development of the sea urchin, *Arbacia punctulata*. *Dev. Biol.* 31 (2), 234–241.
- Feng, Y.-M., Liang, G.-J., Pan, B., Qin, X., Zhang, X.-F., Chen, C.-L., Li, L., Cheng, S.-F., De Felici, M., Shen, W., 2014. Notch pathway regulates female germ cell meiosis progression and early oogenesis events in fetal mouse. *Cell Cycle* 13 (5), 782–791.
- Fernández-Boo, M., Fernández, L., Freire, J., 2012. History and management strategies of the sea urchin *Paracentrotus lividus* fishery in Galicia (NW Spain). *Ocean Coast. Manag.* 69, 265–272.
- Fernández-Boo, S., Pedrosa-Oliveira, M.H., Afonso, A., Arenas, F., Rocha, F., Valente, L. M.P., Costas, B., 2018. Annual assessment of the sea urchin (*Paracentrotus lividus*) humoral innate immune status: tales from the north Portuguese coast. *Mar. Environ. Res.* 141, 128–137.
- Forbes, A., Lehmann, R., 1998. Nanos and Pumilio have critical roles in the development and function of *Drosophila* germline stem cells. *Development* 125 (4), 679–690.
- Fujii, T., Mitsunaga-Nakatsubo, K., Saito, I., Iida, H., Sakamoto, N., Akasaka, K., Yamamoto, T., 2006. Developmental expression of Hpnanos, the *Hemicentrotus pulcherrimus* homologue of nanos. *Gene Expr. Patterns* 6 (5), 572–577.
- Gaitán-Espitia, J.D., Sanchez, R., Bruning, P., Cardenas, L., 2016. Functional insights into the testis transcriptome of the edible sea urchin *Loxechinus albus*. *Sci. Rep.* 6, 6.
- Gaitán-Espitia, J.D., Hofmann, G.E., 2017. Gene expression profiling during the embryo-to-larva transition in the giant red sea urchin *Mesocentrotus franciscanus*. *Ecol. Evol.* 7 (8), 2798–2811.

- Galasso, C., D'Aniello, S., Sansone, C., Ianora, A., Romano, G., 2019. Identification of cell death genes in sea urchin *Paracentrotus lividus* and their expression patterns during embryonic development. *Genome Biol. Evol.* 11 (2), 586–596.
- García-Rincón, J., Darzon, A., Beltrán, C., 2016. Speract, a sea urchin egg peptide that regulates sperm motility, also stimulates sperm mitochondrial metabolism. *Biochim. Biophys. Acta (BBA) Bioenerg.* 1857 (4), 415–426.
- Garmendia, J.M., Menchaca, I., Belzunce, M.J., Franco, J., Revilla, M., 2010. Seasonal variability in gonad development in the sea urchin (*Paracentrotus lividus*) on the Basque coast (Southeastern Bay of Biscay). *Mar. Pollut. Bull.* 61 (4–6), 259–266.
- Ghisaura, S., Loi, B., Biosa, G., Baroli, M., Pagnozzi, D., Roggio, T., Uzzau, S., Anedda, R., Addis, M.F., 2016. Proteomic changes occurring along gonad maturation in the edible sea urchin *Paracentrotus lividus*. *J. Proteom.* 144, 63–72.
- Gibellini, F., Smith, T.K., 2010. The Kennedy pathway-de novo synthesis of phosphatidylethanolamine and phosphatidylcholine. *IUBMB Life* 62 (6), 414–428.
- Gildor, T., Malik, A., Sher, N., Avraham, L., Ben-Tabou de-Leon, S., 2016. Quantitative developmental transcriptomes of the Mediterranean sea urchin *Paracentrotus lividus*. *Mar. Genom.* 25, 89–94.
- González-Castellano, I., Manfrin, C., Pallavicini, A., Martínez-Lage, A., 2019. De novo gonad transcriptome analysis of the common littoral shrimp *Palaeomon serratus*: novel insights into sex-related genes. *BMC Genom.* 20 (1), 757.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Maucci, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29 (7), 644–652.
- Guang, A., Howison, M., Zapata, F., Lawrence, C., Dunn, C.W., 2021. Revising transcriptome assemblies with phylogenetic information. *PLoS One* 16 (1), e0244202.
- Guilou, H., Zdravcov, D., Martin, P.G., Jacobsson, A., 2010. The key roles of elongases and desaturases in mammalian fatty acid metabolism: Insights from transgenic mice. *Prog. Lipid Res.* 49 (2), 186–199.
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59 (3), 307–321.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., Williams, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8 (8), 1494–1512.
- Han, L., Ding, J., Wang, H., Zuo, R., Quan, Z., Fan, Z., Liu, Q., Chang, Y., 2019. Molecular characterization and expression of SfAd1 in the sea urchin (*Strongylocentrotus intermedius*). *Gene* 705, 133–141.
- Huson, D.H., Richter, D.C., Rausch, C., DeZulian, T., Franz, M., Rupp, R., 2007. Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinform.* 8 (1), 460.
- Iries, P., Eshaer, N., Pülsch, M.-D., 2016. The Notch pathway regulates both the proliferation and differentiation of follicular cells in the panoistic ovary of *Blattella germanica*. *Open Biol.* 6 (1), 150197.
- James, P., Srikavupio, S.I., Mortensen, A., 2015. Sea urchin aquaculture in Norway. *Echinoderm. Aquaculture* 147–173.
- Janies, D.A., Witter, Z., Linchango, G.V., Foltz, D.W., Miller, A.K., Kerr, A.M., Jay, J., Reid, R.W., Wray, G.A., 2016. Echinodb, an application for comparative transcriptomics of deeply-sampled clades of echinoderms. *BMC Bioinform.* 17, 48.
- Jia, Z., Wang, Q., Wu, K., Wei, Z., Zhou, Z., Liu, X., 2017. De novo transcriptome sequencing and comparative analysis to discover genes involved in ovarian maturity in *Strongylocentrotus nudus*. *Comp. Biochem. Physiol. Part D: Genom. Proteom.* 23, 27–38.
- Kabeya, N., Sanz-Jorquera, A., Carboni, S., Davie, A., Oboh, A., Monroig, O., 2017. Biosynthesis of polyunsaturated fatty acids in sea urchins: molecular and functional characterisation of three fatty acyl desaturases from *Paracentrotus lividus* (Lamarck 1816). *PLoS One* 12 (1), e0169374.
- Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30 (4), 772–780.
- Katoh, K., Rozewicki, J., Yamada, K.D., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20 (4), 1160–1166.
- Kim, D., Song, L., Breitwieser, F.P., Salzberg, S.L., 2016. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res.* 26, 1721–1729.
- Kinjo, S., Kiyomoto, M., Yamamoto, T., Ikee, K., Yaguchi, S., 2018. HpbBase: a genome database of a sea urchin, *Hemicentrotus pulcherrimus*. *Dev., Growth Differ.* 60 (3), 174–182.
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9 (4), 357–359.
- Lang-Unnath, N., 1992. Purification and properties of *Plasmodium falciparum* malate dehydrogenase. *Mol. Biochem. Parasitol.* 50, 17–26.
- Larson, A.J., Coppard, S.E., Pespeni, M.H., Reed, F.A., 2018. Gene expression across tissues, sex, and life stages in the sea urchin *Triptenaes gratilla* [Toxopneustidae, Odontophora, Camarodonta]. *Mar. Genom.* 41, 12–18.
- Law, S.-H., Chan, M.-L., Marathe, G.K., Parveen, F., Chen, C.-H., Ke, L.-Y., 2019. An updated review of lysophosphatidylcholine metabolism in human diseases. *Int. J. Mol. Sci.* 20 (5), 1149.
- Li, S., Lu, M.M., Zhou, D., Hammes, S.R., Morrissey, E.E., 2007. GLP-1: a novel zinc finger protein required in somatic cells of the gonad for germ cell development. *Dev. Biol.* 301 (1), 106–116.
- Lima, L., Snaimeri, B., Sacomoto, G., Lopez-Maestre, H., Marchet, C., Miele, V., Sagot, M.-F., Lacroix, V., 2017. Playing hide and seek with repeats in local and global de novo transcriptome assembly of short RNA-seq reads. *Algorithms Mol. Biol.* 12, 2–22.
- Lipani, C., Vitturi, R., Sconzo, G., Barbata, G., 1996. Karyotype analysis of the sea urchin *Paracentrotus lividus* (Echinodermata): evidence for a heteromorphic chromosome sex mechanism. *Mar. Biol.* 127 (1), 67–72.
- Liu, H., Chang, Y.-q., 2015. Sea urchin aquaculture in China. *Echinoderm Aquac.* 127–146.
- Liu, H., Kelly, M.S., Cook, E.J., Black, K., Orr, H., Zhu, J.X., Dong, S.L., 2007. The effect of diet type on growth and fatty-acid composition of sea urchin larvae, I. *Paracentrotus lividus* (Lamarck, 1816) (Echinodermata). *Aquaculture* 264 (1), 247–262.
- Liu, X.-I., Chang, Y.-q., Xiang, J.-h., Cao, X.-b., 2005. Estimates of genetic parameters for growth traits of the sea urchin, *Strongylocentrotus intermedius*. *Aquaculture* 243 (1), 27–32.
- López-Hernández, J., González-Castro, M.J., Piñero-Sotelo, M., 1999. Determination of sterols in sea urchin gonads by high-performance liquid chromatography with ultraviolet detection. *J. Chromatogr. Sci.* 37 (7), 237–239.
- Machado, I., Moura, P., Pereira, F., Vasconcelos, P., Gaspar, M.B., 2019. Reproductive cycle of the commercially harvested sea urchin (*Paracentrotus lividus*) along the western coast of Portugal. *Invertebr. Biol.* 138 (1), 40–54.
- Materna, S.C., Davidson, E.H., 2012. A comprehensive analysis of Delta signaling in pre-gastrular sea urchin embryos. *Dev. Biol.* 364 (1), 77–87.
- Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., Thomas, P.D., 2019. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.* 14 (3), 703–721.
- Mita, M., Oguchi, A., Kikuyama, S., Yasumasu, I., De Santis, R., Nakamura, M., 1994. Endogenous substrates for energy metabolism in spermatozoa of the sea urchins *Arbacia lixula* and *Paracentrotus lividus*. *Biol. Bull.* 186 (3), 285–290.
- Mongiardino Koch, N., Coppard, S.E., Lessios, H.A., Briggs, D.E.G., Mool, R., Rouse, G. W., 2018. A phylogenomic resolution of the sea urchin tree of life. *BMC Evol. Biol.* 18 (1), 189.
- Morroni, L., Sartori, D., Costantini, M., Genovesi, L., Magliocco, T., Ruocco, N., Buttino, I., 2019. First molecular evidence of the toxicogenetic effects of copper on sea urchin *Paracentrotus lividus* embryo development. *Water Res.* 160, 415–423.
- Mos, B., Byrne, M., Dworjanyan, S.A., 2016. Biogenic acidification reduces sea urchin gonad growth and increases susceptibility of aquaculture to ocean acidification. *Mar. Environ. Res.* 113, 39–48.
- Nagasawa, K., Thitipheera, T., Osada, M., 2019. Phenotypic Stability of Sex and Expression of Sex Identification Markers in the Adult Yesso Scallop *Mizuhopecten yessoensis* throughout the Reproductive Cycle (vol 9, 277, 2019). *Animals* 9 (9).
- Nef, S., Schaad, O., Stallings, N.R., Cederroth, C.R., Pletti, J.L., Schaefer, G., Malki, S., Dubois-Dauphin, M., Boizet-Bonhoure, B., Descombes, P., Parker, K.L., Vassalli, J.D., 2005. Gene expression during sex determination reveals a robust female genetic program at the onset of ovarian development. *Dev. Biol.* 287 (2), 361–377.
- Ohgaki, S.-I., Kato, T., Kobayashi, N., Tanase, H., Kumagai, N.H., Ishida, S., Nakano, T., Wada, Y., Yusa, Y., 2019. Effects of temperature and red tides on sea urchin abundance and species richness over 45 years in southern Japan. *Ecol. Inl.* 96, 684–693.
- Ouriens, R., Naya, I., Freire, J., 2015. Mismatch between biological, exploitation, and governance scales and ineffective management of sea urchin (*Paracentrotus lividus*) fisheries in Galicia. *Mar. Policy* 51, 13–20.
- Pangas, S.A., Rajkovic, A., 2015. Chapter 21 - follicular development: mouse, sheep, and human models. In: Plant, T.M., Zeleznik, A.J. (Eds.), *Knobil and Neill's Physiology of Reproduction*, fourth ed. Academic Press, San Diego, pp. 947–995.
- Phillips, K., Bremer, P., Slocock, P., Hamid, N., Delahunty, C., Barker, M., Kissick, J., 2009. Effect of gender, diet and storage time on the physical properties and sensory quality of sea urchin (*Evechinus chloroticus*) gonads. *Aquaculture* 288 (3), 205–215.
- Phillips, K., Niimi, J., Hamid, N., Slocock, P., Delahunty, C., Barker, M., Sewell, M., Bremer, P., 2010. Sensory and volatile analysis of sea urchin roe from different geographical regions in New Zealand. *LWT - Food Sci. Technol.* 43 (2), 202–213.
- Piprek, R.P., Damulewicz, M., Tassan, J.-P., Kloc, M., Kubiak, J.Z., 2019. Transcriptome profiling reveals male- and female-specific gene expression pattern and novel gene candidates for the control of sex determination and gonad development in *Xenopus laevis*. *Dev. Genes Evol.* 229 (2), 53–72.
- Pjeta, R., Lindner, H., Kremsler, L., Salvenmoser, W., Sobral, D., Ladurner, P., Santos, R., 2020. Integrative transcriptome and proteome analysis of the tube foot and adhesive secretions of the sea urchin *Paracentrotus lividus*. *Int. J. Mol. Sci.* 21 (3), 946.
- Powell, D., 2019. *drpowell/degust 4.1.1. Zenodo*.
- Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Müller, J., Arnold, R., Rattei, T., Letunic, I., Doerks, T., Jensen, L.J., von Mering, C., Bork, P., 2011. eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.* 40 (D1), D284–D289.
- Prato, E., Fanelli, G., Angioni, A., Biondolino, F., Parlapiano, I., Papa, L., Denti, G., Seci, M., Chiantore, M., Kelly, M.S., Ferranti, M.P., Addis, P., 2018. Influence of a prepared diet and a macroalgae (*Ulva* sp.) on the growth, nutritional and sensory qualities of gonads of the sea urchin *Paracentrotus lividus*. *Aquaculture* 493, 240–250.
- Punta, M., Coghill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L., Eddy, S.R., Bateman, A., Finn, R.D., 2012. The Pfam protein families database. *Nucleic Acids Res.* 40, D290–D301.

- Robinson, M.D., Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11 (3), R25.
- Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2009. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1), 139–140.
- Rocha, F., Rocha, A.C., Baiao, L.F., Gadelha, J., Camacho, C., Carvalho, M.L., Arenas, F., Oliveira, A., Maia, M.R.G., Cabrita, A.R., Pintado, M., Nunes, M.L., Almeida, C.M.R., Valente, L.M.P., 2019. Seasonal effect in nutritional quality and safety of the wild sea urchin *Paracentrotus lividus* harvested in the European Atlantic shores. *Food Chem.* 282, 84–94.
- Romanienko, P.J., Camerini-Otero, R.D., 2000. The mouse Spo11 gene is required for meiotic chromosome synapsis. *Mol. Cell* 6 (5), 975–987.
- Ruocco, N., Costantini, M., Santella, L., 2016. New insights into negative effects of lithium on sea urchin *Paracentrotus lividus* embryos. *Sci. Rep.* 6, 32157.
- Sanna, R., Siliari, S., Melis, R., Loi, B., Baroli, M., Roggio, T., Uzzau, S., Anedda, R., 2017. The role of fatty acids and triglycerides in the gonads of *Paracentrotus lividus* from Sardinia: Growth, reproduction and cold acclimatization. *Mar. Environ. Res.* 130, 113–121.
- Sea Urchin Genome Sequencing, C., 2006. The Genome of the Sea Urchin *Strongylocentrotus purpuratus*. *Science* 314 (5801), 941–952.
- Sellem, F., Guillou, M., 2007. Reproductive biology of *Paracentrotus lividus* (Echinodermata: Echinoidea) in two contrasting habitats of northern Tunisia (southeast Mediterranean). *J. Mar. Biol. Assoc. U.K.* 87 (3), 763–767.
- Shen, F., Long, Y., Li, F., Ge, G., Song, G., Li, Q., Qiao, Z., Cui, Z., 2020. *De novo* transcriptome assembly and sex-biased gene expression in the gonads of Amur catfish (*Silurus asotus*). *Genomics* 112 (3), 2603–2614.
- Shi, D., Zhao, C., Chen, Y., Ding, J., Zhang, L., Chang, Y., 2020. Transcriptomes shed light on transgenerational and developmental effects of ocean warming on embryos of the sea urchin *Strongylocentrotus intermedius*. *Sci. Rep.* 10 (1), 7931.
- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31 (19), 3210–3212.
- Song, L., Flores, L., 2015. Recorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience* 4 (1), 48.
- Sujit, K.M., Singh, V., Trivedi, S., Singh, K., Gupta, G., Rajender, S., 2020. Increased DNA methylation in the spermatogenesis-associated (SPATA) genes correlates with infertility. *Andrology* 8 (3), 602–609.
- Sun, Z.H., Zhang, J., Zhang, W.J., Chang, Y.Q., 2019. Gonadal transcriptomic analysis and identification of candidate sex-related genes in *Mesocentrotus nudus*. *Gene* 698, 72–81.
- Suzek, B.E., Huang, H., McGarvey, P., Mazumder, R., Wu, C.H., 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 23 (10), 1282–1288.
- Tao, W., Yuan, J., Zhou, L., Sun, L., Sun, Y., Yang, S., Li, M., Zeng, S., Huang, B., Wang, D., 2013. Characterization of gonadal transcriptomes from Nile Tilapia (*Oreochromis niloticus*) reveals differentially expressed genes. *PLoS One* 8 (5), e63604.
- Tato, T., Saigüero-González, N., León, V.M., González, S., Beiras, R., 2018. Ecotoxicological evaluation of the risk posed by bisphenol A, triclosan, and 4-nonylphenol in coastal waters using early life stages of marine organisms (*Ischrysis galbana*, *Mytilus galloprovincialis*, *Paracentrotus lividus*, and *Acartia clausi*). *Environ. Pollut.* 232, 173–182.
- Teaniniuraitemoana, V., Huvet, A., Levy, P., Klopp, C., Lhuillier, E., Gaertner-Mazouni, N., Gueguen, Y., Le Moullac, G., 2014. Gonad transcriptome analysis of pearl oyster *Pinctada margaritifera*: identification of potential sex differentiation and sex determining genes. *BMC Genom.* 15 (1), 491.
- The UniProt Consortium, 2016. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45 (D1), D158–D169.
- Tian, S., Egertová, M., Bphick, M.R., 2017. Functional characterization of paralogous gonadotropin-releasing hormone-type and corazonin-type neuropeptides in an echinoderm. *Front. Endocrinol.* 8, 259, 259–259.
- Uthicke, S., Deshpande, N.P., Liddy, M., Patel, F., Lamare, M., Wilkins, M.R., 2019. Little evidence of adaptation potential to ocean acidification in sea urchins living in Future Ocean conditions at a CO₂ vent. *Ecol. Evol.* 9 (17), 10004–10016.
- Walker, C.W., Böttger, S.A., Unuma, T., Watts, S.A., Harris, L.G., Lawrence, A.L., Eddy, S.D., 2015. Enhancing the commercial quality of edible sea urchin gonads — technologies emphasizing nutritive phagocytes. *Echinoderm Aquac.* 263–286.
- Wang, H., Ding, J., Ding, S., Chang, Y., 2019. Transcriptome analysis to characterize the genes related to gonad growth and fatty acid metabolism in the sea urchin *Strongylocentrotus intermedius*. *Genes Genom.* 41 (12), 1397–1415.
- Wang, H., Ding, J., Ding, S., Chang, Y., 2020. Integrated metabolomic and transcriptomic analyses identify critical genes in eicosapentaenoic acid biosynthesis and metabolism in the sea urchin *Strongylocentrotus intermedius*. *Sci. Rep.* 10 (1), 1697.
- Wang, X., Li, H., Fu, G., Wang, Y., Du, S., Yu, L., Wei, Y., Chen, S., 2016. Testis-specific serine/threonine protein kinase 4 (Tsk4) phosphorylates Odf2 at Ser-76. *Sci. Rep.* 6 (1), 22861.
- Wei, Z., Liu, X., Zhou, Z., Xu, J., 2019. *De novo* transcriptomic analysis of gonad of *Strongylocentrotus nudus* and gene discovery for biosynthesis of polyunsaturated fatty acids. *Genes Genom.* 41 (5), 583–597.
- Wong, J.M., Gaitán-Espitia, J.D., Hofmann, G.E., 2019. Transcriptional profiles of early stage red sea urchins (*Mesocentrotus franciscanus*) reveal differential regulation of gene expression across development. *Mar. Genom.* 48, 100692.
- Yang, H., Cromley, D., Wang, H., Billheimer, J.T., Sturley, S.L., 1997. Functional expression of a cDNA to human acyl-coenzyme A: cholesterol acyltransferase in Yeast: species-dependent substrate specificity and inhibitor sensitivity. *J. Biol. Chem.* 272 (7), 3980–3985.
- Yasumasu, I., Hino, A., Suzuki, A., Mita, M., 1984. Change in the triglyceride level in sea urchin eggs and embryos during early development. *Dev. Growth Differ.* 26 (6), 525–532.
- Yerham, E., Rilov, G., Shpigiel, M., Abelson, A., 2015. Collapse of the echinoid *Paracentrotus lividus* populations in the Eastern Mediterranean—result of climate change? *Sci. Rep.* 5, 13479.
- Yue, H., Li, C., Du, H., Zhang, S., Wei, Q., 2015. Sequencing and *De Novo* Assembly of the Gonadal Transcriptome of the Endangered Chinese Sturgeon (*Acipenser sinensis*). *PLoS One* 10 (6), e0127332.
- Zhan, Y., Li, J., Sun, J., Zhang, W., Li, Y., Cui, D., Hu, W., Chang, Y., 2019. The impact of chronic heat stress on the growth, survival, feeding, and differential gene expression in the sea urchin *Strongylocentrotus intermedius*. *Front. Genet.* 10, 301.
- Zhang, D., Xie, D., Lin, X., Ma, L., Chen, J., Zhang, D., Wang, Y., Duo, S., Feng, Y., Zheng, C., Jiang, B., Ning, Y., Han, C., 2018. The transcription factor SOX30 is a key regulator of mouse spermiogenesis. *Development* 145 (11), dev164723.
- Zhang, J., Han, X., Wang, J., Liu, B.-Z., Wei, J.-L., Zhang, W.-J., Sun, Z.-H., Chang, Y.-Q., 2019a. Molecular cloning and sexually dimorphic expression analysis of nanos2 in the Sea Urchin, *Mesocentrotus nudus*. *Int. J. Mol. Sci.* 20 (11), 2705.
- Zhang, T., Zarkower, D., 2017. DMRT proteins and coordination of mammalian spermatogenesis. *Stem Cell Res.* 24, 195–202.
- Zhang, W., Lv, Z., Li, C., Sun, Y., Jiang, H., Zhao, M., Zhao, X., Shao, Y., Chang, Y., 2019a. Transcriptome profiling reveals key roles of phagosome and NOD-like receptor pathway in spotting diseased *Strongylocentrotus intermedius*. *Fish Shellfish Immunol.* 84, 521–531.
- Zhang, W., Wang, Z., Leng, X., Jiang, H., Liu, L., Li, C., Chang, Y., 2019b. Transcriptome sequencing reveals phagocytosis as the main immune response in the pathogen-challenged sea urchin *Strongylocentrotus intermedius*. *Fish Shellfish Immunol.* 94, 780–791.

4.2. Unravelling the main immune repertoire of *Paracentrotus lividus* following *Vibrio anguillarum* bath challenge

Sergio Fernández-Boo*, **André M. Machado***, L. Filipe C. Castro, Rita Azeredo, Benjamin Costas

(*Joint First Authors)

Fish and Shellfish Immunology

(DOI: /10.1016/j.fsi.2024.109431)

Fish and Shellfish Immunology 147 (2024) 109431



Contents lists available at ScienceDirect

Fish and Shellfish Immunology

journal homepage: www.elsevier.com/locate/fsi

Full length article

Unravelling the main immune repertoire of *Paracentrotus lividus* following *Vibrio anguillarum* bath challenge

Sergio Fernández-Boo^{a,*}, André Machado^{a,1}, L. Filipe C. Castro^{a,b}, Rita Azeredo^{a,c}, Benjamin Costas^{a,c}

^a Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR). Terminal de Cruzeiros Do Porto de Leixões, Av. General Norton de Matos S/n, 4450-208, Matosinhos, Portugal

^b Departamento de Biologia, Universidade Do Porto, Rua Do Campo Alegre, S/n, Edifício FC4, 4169-007, Porto, Portugal

^c Instituto de Ciências Biomédicas Abel Salazar (ICBAS-UP), Universidade Do Porto, Rua de Jorge Viterbo Ferreira 228, 4050-313, Porto, Portugal



ARTICLE INFO

Keywords:

Innate immunity
Transcriptomics
Sea urchin
Lectin
Vibrio anguillarum

ABSTRACT

Paracentrotus lividus is the most abundant echinoid species in the North East Atlantic Ocean and Mediterranean Sea. Although there is abundant genomic information of the species, there is no deep characterisation of the genes involved in the immune response. Here, a reference transcriptome of male and female coelomocytes was produced. The generated *P. lividus* transcriptome assembly has 203,511 transcripts, N50 transcript length of 1079 bp, and more than 90% estimated gene completeness in Eukaryota and Metazoa BUSCO databases, respectively. Differential gene expression analyses showed 54 and 55 up-regulated genes in *P. lividus* female and male coelomocyte tissues, respectively. These results suggest a similar immune gene repertoire between sexes. To examine the immune response, *P. lividus* was challenged with *Vibrio anguillarum*, one of the candidate pathogens for bald disease. Immune parameters were evaluated at cell and humoral levels, as well as the expression analysis of immune related genes at an early response stage. No differences were found at cellular and humoral levels with the exception of the increase of nitric oxide in perivisceral fluid of challenged animals. At the gene expression level, a total of 2721 genes were upregulated in challenged animals, 13.6 times higher expression than control group. Our analysis revealed that four major KEGG pathways were enriched in challenged animals: Autophagy (KEGG:04140), Endocytosis (KEGG:04144), Phagosome (KEGG:04145) and Protein processing in endoplasmic reticulum (KEGG:04141). Several toll-like receptors (TLR), scavenger receptors cysteine-rich (SRCR) or nucleotide-binding oligomerisation domain like receptors (NLR) were identified as major family genes for pathogen recognition and immune defence. This study provides a valuable transcriptomic resource and unfolds the molecular basis of immune response to *V. anguillarum* exposure. Overall, our findings contribute to the conservation effort of the *P. lividus* populations, as well as its sustainable exploitation in an aquaculture context.

1. Introduction

Sea urchins are key model organisms in biology, widely used to study reproduction, embryology or ecotoxicology, and with high potential for aquaculture purposes [1,2]. Among sea urchins, *Paracentrotus lividus* is the most common species of the North East Atlantic Ocean [3]. This species is affected by several biotic and abiotic factors specially related with the rise of temperatures, that lead to mass mortality events and ultimately limit distribution range of the species [4]. Rise of temperature combined with multiple anthropogenic stressors can be related with an

impairment of the immune response increasing the probability to be infected with opportunistic bacteria [5,6]. Although some pathogenic agents were described affecting sea urchins, such as the amoeba *Neoparamoeba branchiophila* [7] or some bacterial agents such as *Vibrio alginolyticus* [8,9], *Vibrio anguillarum* and *Aeromonas salmonicida* [5,10] and more recently *Vibrio splendidus* [11], the reality is more complex and diseases such as bald sea urchin disease or the spotting disease are considered multifactorial [12]. Several reports of mass mortality events in the last years related with increase of sea surface temperature [8,13,14] and posterior infection with bacteria [5] suggested that bald disease

* Corresponding author.

E-mail address: sboo@ciimar.up.pt (S. Fernández-Boo).

¹ Both authors contribute equally to this manuscript.

<https://doi.org/10.1016/j.fsi.2024.109431>

Received 29 November 2023; Received in revised form 31 January 2024; Accepted 5 February 2024

Available online 10 February 2024

1050-4648/© 2024 Published by Elsevier Ltd.

is not a unique pathogen or stressful condition underscoring these events and that isolated bacteria from affected sea urchins can be opportunistic and not specific [15].

Sea urchins possess a wide variety of mechanisms to counteract infections and avoid possible pathogens. Their innate immune system is composed by immune cells called coelomocytes and humoral factors released to the perivisceral coelom space [16]. Coelomocytes comprehend a heterogeneous cellular population, which varies among species. In the case of *P. lividus*, five different types of cells were previously described: white and red amoebocytes, petaloid and philipodial phagocytes and vibratile cells [17]. Red and white amoebocytes seem to be involved in the rapid recognition, clotting and encapsulation of external agents. Moreover, red amoebocytes possess a known antibacterial activity with the release of the red pigment echinochrome [18–21]. Phagocytic cells, the most abundant cell type of sea urchins that represents around 80% of the total population, are involved in removal of bacteria and cellular debris, but also in encapsulation and clotting [18,19]. Finally, the role of vibratile cells is not completely understood, but it was suggested that these cells are involved in the immobilization and inhibition of bacterial growth [20]. Also, immune cells together with other epithelial cells release humoral factors to the coelom space recognizing and inhibiting the growth of pathogens or neutralizing and destroying pathogens by a huge variety of tools such as hemolysis, proteases, lysozyme or reactive oxygen and nitrogen species among others [16,22,23].

Transcriptomic studies are one of the best approaches to address the complexity of the immune system and the effective immune response of a species. In the case of sea urchins, a high quantity of information is available in some species, especially in *Strogilocentrotus purpuratus* after the initial sequencing of the genome [24], with a dedicated database of the immune related genes available for this species [25]. This includes a large gene repertoire of pathogen-associated molecular patterns (PAMPs) including, 222 Toll-like receptors, 203 NATCH domain-LRR genes or 218 scavenger receptor cysteine-rich proteins. Recently, the whole nuclear genome of *P. lividus* was also sequenced [2], which will improve our capacity to decipher the immune function in echinoderms. Yet, although there is some available information on the immune response of *P. lividus* sea urchin after being challenged by certain stress factors or bacterial agents, the majority of the studies were focused on cellular and humoral responses [19,23,26,27]. Moreover, very few experimental designs utilized a transcriptomic approach to inform on the molecular landscape of modified pathways [28].

In this study, a comprehensive dataset of the transcriptome of *P. lividus* coelomocytes is presented. A de novo coelomocyte reference transcriptome was built and used to analyse gene expression differences between male and female coelomocytes. Simultaneously, we conducted a challenge with *Vibrio anguillarum* and analysed innate immune parameters. Lastly, we performed a gene expression analysis using RNA-Seq and identified some key genes responsible for the early immune response in this species.

2. Material and methods

2.1. Specimen collection and exposition trial

Experiment 1 – Coelomocyte gene expression of males and females: A coelomocyte reference transcriptome was constructed using the same 10 *P. lividus* previously collected at the Vila Chã beach by Ref. [29]. Animals were collected and immediately transported to CIMAR facilities for coelomic fluid (CF) extraction through the peristomal membrane with 10 mL syringes assembled with 21 gauge needles. At least 8 mL of CF were collected and maintained in 15 mL tubes in ice until centrifugation at 400×g for 10 min at 4 °C. The pellet (coelomocytes) was stored immediately at –80 °C until use. After a careful evaluation of sex through tissue imprint by light microscopy, three males and three females were selected for RNA extraction and

sequencing.

Experiment 2 – Challenge of sea urchins with *Vibrio anguillarum*: Sixty commercial size (7.07 ± 0.57 cm diameter; 123.64 ± 21.05 g; 10.77 ± 2.68 gonadosomatic index) sea urchins *P. lividus* were collected at Vila Chã beach (41.295160 N, 08.737073 W) by scuba diving. Animals were kept in acclimation during five days in six tanks (30 L) with 10 randomly distributed individuals per tank. The system was an open system, with gentle aeration and a water temperature of 20 ± 0.2 °C and 35‰ salinity. During acclimation animals were fed with 50 g of *Ulva lactuca* per tank per day.

After five days of acclimation, a challenge with *Vibrio anguillarum* was performed. Bacteria were grown in 500 mL Tryptic Soy Broth medium for 18 h, at 25 °C in 1 L Erlenmeyer with gentle agitation. For that, the volume of all tanks was decreased to 10 L and animals from three tanks were subjected to a bath bacterial challenge (3×10^{12} cfu/L) during 4 h (with 30 min' interval between tanks to allow sampling after the same time of bacterial challenge). The other three tanks remained undisturbed and were regarded as control. Then, exposed and control animals were sampled at the end of the bacterial challenge for evaluation of immune parameters and RNA-Seq analysis (Fig. 1).

Coelomocytes of three sea urchins per tank were collected according to Ref. [16] for RNA sequencing. Briefly, 0.5 mL of CF was used for differential cell counting in a Malassez chamber and the remaining amount (around 7.5 mL) was then centrifuged for 10 min at 400×g and 4 °C. The supernatant (perivisceral fluid (PF)) was stored at –80 °C for humoral immune parameters evaluation and pellet (coelomocytes) was stored in 1.5 mL of Trizol (NZY Tech, PT) at –80 °C until RNA extraction.

Finally, the remaining seven animals were used for evaluation of the immune parameters. For that, CF was extracted and 0.5 mL were used for cell counting. In 4 individuals the remaining CF was centrifuged at 400×g, 10 min and 4 °C and the PF was stored in aliquots of 1 mL and stored at –80 °C, while coelomocytes were resuspended in FSW at a proportion of 5×10^5 cells per mL. Cells were then sonicated 3×5 min in cold water and centrifuged at 5400×g for 10 min at 4 °C for cellular debris removal. The supernatant was stored at –80 °C until use. The CF from the remaining 3 sea urchins was used directly for phagocytosis assay (see section 2.2).

2.2. Innate immune parameters evaluation

For total and differential cell counting, an aliquot of 0.5 mL of CF was stored in a 1.5 mL tube containing 0.5 mL of ISO-EDTA anticoagulant (20 mM Tris, 0.5 M NaCl, 70 mM EDTA pH 7.5). Then, cells were counted in a Malassez chamber differentiating between red and white amoebocytes, total phagocytes (philipodial and petaloid) and vibratile cells.

Phagocytosis assay was performed using 5 mL of CF, collected from three sea urchins per tank, against 5 mL of *Candida albicans* at 10^7 cells/mL in filtered sea water for 1 h with a ratio 1:5 (10^5 coelomocytes for 5×10^5 *C. albicans* cells). After that, a cytospin was done using 200 µL of cell suspension and the percentage of phagocytosis was determined by light microscopy after staining with hematoxylin-eosin staining according to Ref. [30].

Bactericidal activity against *V. anguillarum*, protease and lysozyme activity, total protein and nitric oxide content were done according to Ref. [16] in perivisceral fluid and cell extracts respectively. Haemolytic activity was done using 20 µL of sample diluted in Mg-EGTA-GVB buffer (1:1) (Isotonic Veronal buffered saline solution, 0.1 % gelatin, 20 mM EDTA and 10 mM Mg^{+2} , pH 7.3) in 96 well-plates. 20 µL of rabbit red blood cells (Probiologica Lda., Portugal) were added at a final concentration of 2.5×10^8 cells/mL in saline solution (distilled water with 0.9 % NaCl). Samples were incubated for 100 min at room temperature with continuous shaking, the reaction was stopped by adding 150 µL of cold EDTA_GVB (without Magnesium). Samples were then centrifuged for 2.5 min at 120×g and the extent of haemolysis was estimated by

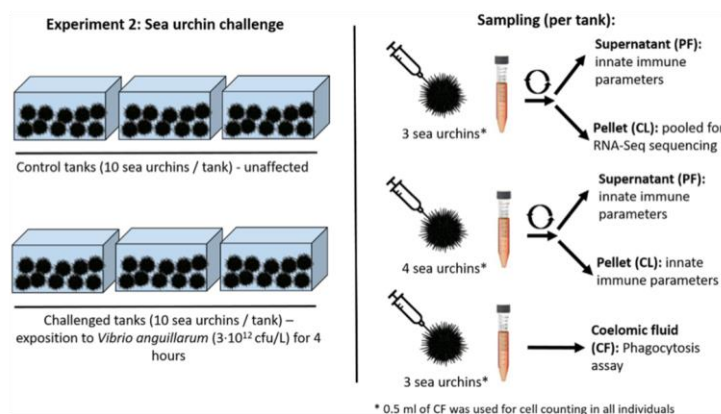


Fig. 1. Representation of the trial and sampling from Experiment 2.

measuring the optical density of the supernatant at 414 nm (BioTek, USA). A positive control (100 % haemolysis) was included using 40 μ L of distilled water instead of sample dilution, as well as a negative control (0 % haemolysis), using saline solution instead of sample. The haemolytic activity was calculated as the percentage of lysed cells.

Phagocytic activity was evaluated by examining at least 200 cells in each slide by light microscope according to Ref. [30]. The number of cells engulfing *C. albicans* cells was calculated in percentage.

2.3. RNA extraction and sequencing

RNA from the six specimens (comprising 3 males and 3 females) used to build the reference transcriptome assembly was extracted according to protocol described in Ref. [29] (Supplementary Table I).

The second group of animals was used to perform, exclusively, differential expression analyses. In these analyses, nine individuals per condition (control and challenged) were used, the number of males and females were similar in both conditions. Briefly, samples were kept in 15 mL falcon tubes with 1.5 mL of Trizol (NZY Tech, PT) and maintained at -80°C until processing. The extraction was done using the Total RNA Isolation Kit (NZY Tech, PT) according to manufacturer's instructions. RNA concentration was measured in a DeNovix DS-11 spectrophotometer (DeNovix Inc, USA) and 4 μ g of RNA were subjected to DNase treatment (RQ1 – Promega) according to manufacturer's instructions. After DNase treatment, RNA was isolated once again and cleaned using the Total RNA isolation kit (NZY Tech, PT) according to manufacturer's instruction and eluted in a final volume of 40 μ L in mQ water. Finally, the concentration was measured again in a DeNovix DA-11 spectrophotometer. RNA integrity was observed in a 2 % agarose gel stained with GelGreen nucleic acid stain (NZYTech, PT). Before sequencing, we chose to utilize pooled samples rather than individual ones. These pools were created for each condition, by combining the RNA of three individuals per tank (2 μ g per individual), resulting in three control pools and three exposed pools. The integrity and quantity of RNA for each pool was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The RNA integrity number was >6.1 for all samples (Supplementary Tables I and II (Mixes descriptors/ Samples QC)).

At the end, six samples (3 control pools and 3 exposed pools) were shipped to Novogene Company Limited (Hong Kong), and sequenced using the Illumina HiSeq-4000 platform (100 x 2 bp, paired-end, 20 million sequencing reads per sample).

2.4. Transcriptomic analyses

In the transcriptomic analyses, the process was initiated by conducting quality control checks on all sequenced samples, which included references (comprising 3 male and 3 female samples) and differential gene expression samples (consisting of 3 control pools and 3 exposed pools). FastQC software (version 0.11.8) (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), Trimmomatic (version 0.38) (configured with parameters: LEADING:5, TRAILING:5, SLIDINGWINDOW:4:5, MINLEN:36) [31], and Roorrector (version 1.0.3) software [32] (Supplementary Tables II Samples QC) were used.

To assemble the reference transcriptome, we concatenated the six samples, conducted a de novo assembly, and applied part of the filtration process outlined in Ref. [29]. This process guarantees an accurate quantification of gene expression and downstream analyses. Despite the recent release of genome assembly of *P. lividus*, the annotation files are not yet useable for general transcriptomic analyses [2]. In summary, Trinity (v. 2.8.4) [33,34] software was employed to generate the initial version of the transcriptome assembly. The nucleotide database of NCBI (NCBI-nt) [35] and the UniVec databases were used to remove contaminants from the transcriptome (Transcriptome V1). Additionally, Corset (v.1.0.9) [36] software was utilized for redundancy removal (Transcriptome V2). Quality control statistics for the two transcriptome versions were assessed using the Benchmarking Universal Single-Copy Orthologs tool (BUSCO v. 3.0.2) [37], TransRate (v. 1.0.3) [38], and Bowtie 2 (v. 2.3.4.2) [39].

For transcriptome annotation, a pipeline described by Ref. [40] was followed. Structural annotation, which included the prediction of open reading frames (ORFs), was carried out using Transdecoder. Subsequently, the resulting structural annotation files were processed using the Gtf/Gff Analysis Toolkit (AGAT) (version 0.8.0) [41]. On the other hand, functional annotation was conducted with the InterProScan tool (version 5.44.80) [42], along with Blast-n, Blast-p, and Blast-x searches using the NCBI Blast-n tool [43] and the DIAMOND software (version 2.0.13) [44] against the NCBI-nt database and the protein databases of NCBI (NCBI-RefSeq – Reference Sequence Database [45]; NCBI-nr – non-redundant database of NCBI [35]). Following this step, and preceding the execution of differential gene expression analyses, the annotation information was integrated with the counts that had been pre-computed using the Corset software.

Data of gene counts from both analysis, males vs female coelomocytes and challenged vs control, were uploaded into Degust server tool for analysis of differential expression genes (DEGs) [46] using a minimal of 1 CPM in at least 3 samples. For differential expression analysis, Edge

R method with a false discovery rate cut-off of 0.05 and a fold change of 2 was applied for Male vs Female samples as well as for Control vs Challenged condition. Only annotated sequences were present in result section. Before proceeding with differential gene expression analysis all sequences without functional annotation were removed.

2.5. Functional enrichment analysis

Significant enriched KEGG pathways were identified using the g:GOS tool of gprofiler server [47]. Gene ID of DEGs obtained from Degust analysis were uploaded and run as multiquery using the *Strogyloncentrotus purpuratus* Ensembl genome. The significant threshold used was g:SCS threshold with a *p* value of 0.05.

2.6. Statistical analysis

All statistical analyses from evaluation of immune parameters were performed using SPSS statistical analysis software (IBM, v.26, USA). Results were expressed as mean $\pm$ standard deviation. After verifying normality and variance homogeneity, a one-way analysis of variance (ANOVA) was carried out. Tukey HSD post hoc test was performed to identify significant differences among groups with a level of significance of *p* < 0.05.

3. Results

3.1. Differential cell counts and phagocytic activity

No differences were found between control and bacterial challenged animals regarding the total and cell category counts or phagocytic activity. The cell concentration average was 1.2 and 1.1 million cells per mL in control and exposed groups, respectively (Table 1). In all samples the most common group of cells were the phagocytic cells with more than 60 % of the population (Table 1), followed by white amoebocytes (16 % in control and 20 % in challenged group), vibratile cells with 10 % of the population in both groups and finally red amoebocytes which represented 7 % in control group and 9 % in exposed group. Regarding the phagocytic activity, although the contact with bacteria provoked an increase in the percentage of phagocytosis it was not enough to be significant (Tables I).

3.2. Humoral immune parameters analyses

After the 4 h challenge with bacteria, animals were sampled and the coelomic fluid (CF) was extracted. Several parameters were evaluated in the perivisceral fluid (PF) and in a homogenate of the coelomocytes (cell lysate, CL). None of the parameters evaluated presented differences between groups with the exception of the nitric oxide content in the PF (Table II). Nitric oxide had a concentration of 13.78 ± 6.56 μ M/ml in the exposed group while a concentration of 7.78 ± 1.58 μ M/ml was measured in control group. Results showed a high haemolytic activity in CL samples (more than 80 %) while PF showed around 40 % of

haemolytic activity. Regarding bactericidal activity, a higher activity was measured in CL samples (around 45 %) compared to PF (around 28 %) (Table II).

3.3. Transcriptomic analysis

3.3.1. Transcriptome assembly and annotation

During this study, over 257 million reads across a total of 12 samples were generated (see Supplementary Table II for QC). These included six samples used for constructing the reference transcriptome and performing differential gene expression analyses between male and female coelomocytes, as well as six samples used to analyse the gene expression profiles of the coelomocytes of *P. lividus* under *V. anguillarum* challenge.

Regarding the results pertaining to the reference transcriptome of sea urchin coelomocytes from both males and females, a total of 203,511 transcripts were identified, with an N50 length of 1079bp. The BUSCO analysis demonstrated a high level of completeness, with more than 90 % gene completeness in the Eukaryota and Metazoa databases. Furthermore, the read-back mapping analyses revealed that over 93 % of the raw reads were retained in the final transcriptome assembly. At the end of the annotation procedure, approximately 36,877 transcripts encoding open reading frames were identified, and 54,380 transcripts exhibited a blast hit with at least one of the searched databases (Supplementary Table III Assemblies).

3.3.2. Differential gene expression between wild-type male and female specimens

The analysis of differential gene expression between male and female individuals encompassed a total of 44,368 transcripts following normalization tests in Degust. Out of these, 338 exhibited differential expression but only 109 of them were annotated. Among those, 54 were up-regulated in females, while 55 were up-regulated in males (Supplementary Table IV). After a gene ontology analysis against the Panther database, we found just 20 genes, 10 per sex, associated with a protein class using the genome of *S. purpuratus* as reference (Table 3, Supplementary Table V). Principal component analysis suggests a clear difference among sexes (Fig. 2A), with more than 70 % of the differences being explained in dimension 1 (Fig. 2B). Finally, heatmap and volcano plot showed the homogeneous differentiation between sexes (Fig. 2C&D). Functional enrichment analysis on gProfiler showed no significant enrichment functions between sexes.

3.3.3. Transcriptome analysis of control and bacterial challenged pools

The analysis of differential gene expression between control and bacterial challenged pools involved a total of 38,108 transcripts after normalization tests in Degust. Among these, 4052 transcripts exhibited differential expression of which 2921 were annotated (Supplementary Table VI). In this case, and in contrast to transcriptome of male and female coelomocytes, the challenged group presents a 13.6 times higher expression than the control group. A total of 2721 DEGs were up-regulated in the bacterial challenged group compared to 200 DEGs of the control group (Fig. 3).

A functional analysis on Panther Database using the genome of *S. purpuratus* as reference showed a great difference between groups in terms of gene expression. Regarding Biological process categories (Supplementary Table VII), cellular process (GO:0009987) was the most expressed one in challenged group (693 genes), followed by metabolic process (GO:0008152) (420 genes), biological regulation (GO:0065007) (270 genes), localization (GO:0051179) (181 genes), response to stimulus (GO:0050896) (146 genes) among others (Table 4), while in the control group the most represented category was cellular process (GO:0009987) with only 5 genes.

Regarding molecular function classification (Supplementary Table VIII), binding (GO:0005488) and catalytic activity (GO:0003824) were the most expressed ones in the exposed group with 391 and 386 genes respectively, while in the control group the most represented

Table 1

Total and differential cell counts in control and exposed groups. Data is presented in 10^6 cells/mL in total cell concentration while differential counting is presented as percentage of total population. Phagocytic activity is presented as % of cells with engulfed *Candida albicans* cells. All data is presented as mean $\pm$ standard deviation.

	Control		Challenged	
Total cell count (10^6 cells/mL)	1.27	$\pm$ 0.51	1.10	$\pm$ 0.43
Vibratile cells (%)	10.39	$\pm$ 7.12	9.44	$\pm$ 4.89
White amoebocytes (%)	16.09	$\pm$ 7.74	20.38	$\pm$ 9.77
Red amoebocytes (%)	7.04	$\pm$ 3.98	9.21	$\pm$ 6.46
Phagocytic cells (%)	66.33	$\pm$ 11.66	60.96	$\pm$ 14.53
Phagocytic activity (%)	23.44	$\pm$ 8.10	27.56	$\pm$ 10.50

Table 2

Humoral innate immune parameters in perivisceral fluid and cell lysate samples of control and exposed groups. Data is presented as mean $\pm$ standard deviation. *denotes significant differences between groups ($p < 0.05$).

	Perivisceral fluid					Cell lysate				
	Control		Challenged			Control		Challenged		
Bactericidal activity (%)	28.59	$\pm$	9.14	27.88	$\pm$	7.92	45.57	$\pm$	12.17	45.38
Nitric oxide (μ M)	7.77	$\pm$	1.58	13.78*	$\pm$	6.55	67.06	$\pm$	7.73	73.44
Protein content (μ g/ml)	0.36	$\pm$	0.14	0.29	$\pm$	0.09	0.83	$\pm$	0.27	0.92
Lysozyme (μ g/ml)	0.27	$\pm$	0.11	0.23	$\pm$	0.18	0.41	$\pm$	0.10	0.49
Haemolytic activity (%)	43.59	$\pm$	17.73	40.33	$\pm$	20.96	84.85	$\pm$	6.19	83.53
Protease activity (%)	0.99	$\pm$	0.36	0.88	$\pm$	0.06	0.86	$\pm$	0.10	0.86

Table 3

Differential expressed proteins in males and females of *P. lividus* after gene ontology searching in Panther DB.

Differential expression	Protein class	Mapped Ids	Gene name
Females	chromatin/ c chromatin-binding, or -regulatory protein (PC00077)	XP_011665124	HIT-type domain- containing protein
	intercellular signal molecule (PC00207)	XP_011672459	Fibrinogen C-terminal domain-containing protein
	metabolite interconversion enzyme (PC00262)	XP_782723	Uncharacterized protein
	translational prot ein (PC00263)	XP_795427	40S ribosomal protein S12
	translational prot ein (PC00263)	XP_785096	Uncharacterized protein
	transmembrane signal receptor (PC00197)	XP_030847655	G.PROTEIN_RECEP_F1_2 domain-containing protein
	transmembrane signal receptor (PC00197)	XP_030853863	Deleted in malignant brain tumors 1 protein
	transmembrane signal receptor (PC00197)	XP_030835081	Uncharacterized protein
	transporter (PC00227)	XP_030837755	Uncharacterized protein
	viral or transposable element protein (PC00237)	XP_030832208	Reverse transcriptase domain-containing protein
	cytoskeletal protein (PC00085)	XP_030838296	TCTP domain-containing protein
	intercellular signal molecule (PC00207)	XP_011672459	Fibrinogen C-terminal domain-containing protein
	metabolite interconversion enzyme (PC00262)	XP_030850344	Uncharacterized protein
Males	metabolite interconversion enzyme (PC00262)	XP_030834458	F5/8 type C domain- containing protein
	protein modifying enzyme (PC00260)	XP_030853781	Uncharacterized protein
	scaffold/ adaptor protein (PC00226)	XP_030851019	Tetraspanin
	translational prot ein (PC00263)	XP_795427	40S ribosomal protein S12
	translational prot ein (PC00263)	XP_011680849	40S ribosomal protein S9
	transmembrane signal receptor (PC00197)	XP_030845725	G.PROTEIN_RECEP_F1_2 domain-containing protein
	viral or transposable element protein (PC00237)	XP_030828392	Reverse transcriptase domain-containing protein

category was catalytic activity (GO:0003824) with just 5 genes (Table 4).

After the analysis of classification by Gene Ontology terms on Panther, an enrichment analysis was performed on gProfiler (Supplementary Table XI). This analysis revealed that four major KEGG pathways were enriched in exposed animals: Autophagy (KEGG:04140) with 8 entries, Endocytosis (KEGG:04144) with 23 entries, Phagosome (KEGG:04145) with 13 entries and Protein processing in endoplasmic reticulum (KEGG:04141) with 14 entries (Table 5, Fig. 4).

Also, regarding pathogen recognition proteins, four different lectins were identified (Supplementary Table VI) all of them with a higher expression in challenged animals. Six copies of alpha-N-acetylgalactosamine, two C-type lectin Thr-1, one Melectin and one endoplasmic reticulum lectin isoform X1. The complement system was represented by 28 genes all of them up-regulated in challenged animals.

4. Discussion

A complete overview of the main immune genes from *P. lividus* were investigated. In order to define a reference transcriptome database, an analysis of the gene expression of males and females' coelomocytes was performed. This analysis revealed that a low differential gene expression was detected between both sexes. Also, a similar pattern of expression was observed in *Arbacia lixula*, where just 119 genes were differentially expressed between males and females with 73 being up-regulated in males and 47 in females [48]. In this study, a total of 109 annotated DEGs with 55 up-regulated in males and 54 in females were identified. No enriched functions were found in either sex, indicating a similar immune response regardless of sex. Although some reports showed differential activity in immune parameters pre- and post-spawning between sexes, females typically exhibit stronger immune responses than males, which could be a strategy to guarantee the success of the next generation [49]. Our findings did not follow this hypothesis, as a consequence of transcriptomic analysis. Nevertheless, it is important to note that all animals involved in this study were at stages III and IV of maturation, in accordance with [50]. This corresponds to a stage that falls midway in the gonad maturation process, as demonstrated in Ref. [29]. This intermediate stage may have a comparatively lower impact on gene expression in coelomocytes compared to individuals at stages VI (full maturation, pre-spawning) or stage I (ripeness, post-spawning). It could be interesting to know if different maturation stages could have or not influence in the gene expression of coelomocytes between sexes although this comparison was not the goal of the present study.

Regarding the bacterial challenge, no differences were observed neither in the total neither in the differential cell counts between challenged and control group. It was previously documented that coelomocyte populations can change during time after infection, but that data was obtained in injected individuals and at 24 h post-injection [51]. The present study did not show alteration in coelomocyte populations after 4 h of contact with *V. anguillarum*, suggesting that these changes might occur later during immune response. Also, immune related parameters on the perivisceral fluid or the cell lysate showed no differences between

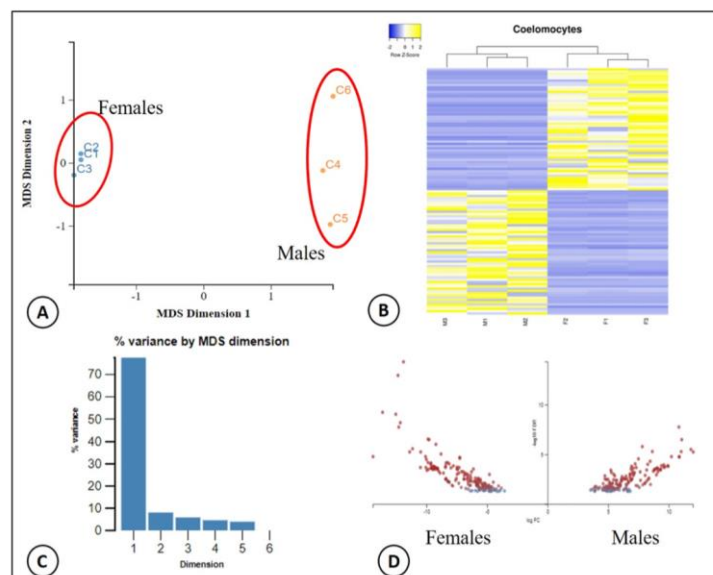


Fig. 2. Data of the reference transcriptome from males and females of the sea urchin *Paracentrotus lividus*. A. Principal component analysis (PCA) of the samples. B. Heatmap from the DEGs obtained after analysis of reads by Degust. C. Percentage of variance by MDS dimension of the PCA. D. Volcano plot of the DEGs from males and females.

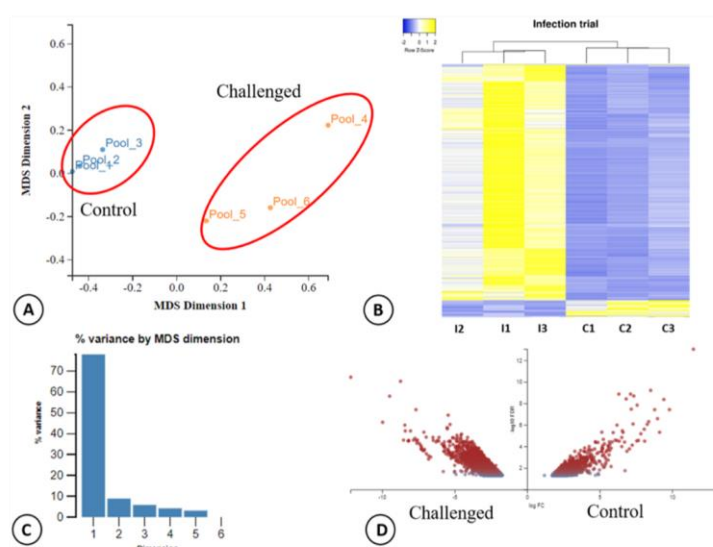


Fig. 3. Transcriptome data from control and challenged *P. lividus*, sampled after the infection trial. A. Principal component analysis (PCA) of the samples. B. Heatmap from the DEGs (> 2 FC and p-value < 0.05) obtained after analysis of reads by Degust. C. Percentage of variance by MDS dimension of the PCA. D. Volcano plot of the DEGs (> 2 FC and p-value < 0.05) from control (C) and challenged (I) groups.

control or challenged groups with the exception of an increase in nitric oxide (NO) concentration in PF of challenged sea urchins. NO is a molecule that could act as a modulator of the immune system that can increase with the presence of cytokines [16,52]. The presence of

antigens from bacteria could lead to recognition by coelomocytes and trigger the production of NO and also the expression of genes related with recognition and early immune response. Previous data on *P. lividus* showed an increase of NO concentration in summer months which are

Table 4

List of genes presented in Biological process and Molecular function categories according to Panther DB analysis.

Biological Process	Number of genes	
	Challenged	Control
Cellular process (GO:0009987)	693	5
Reproductive process (GO:0022414)	1	0
Localization (GO:0051179)	181	1
Biological process involved in interspecies interaction between organisms (GO:0044419)	3	0
Reproduction (GO:0000003)	1	0
Biological regulation (GO:0065007)	270	0
Response to stimulus (GO:0050896)	146	0
Pigmentation (GO:0043473)	1	0
Signalling (GO:0023052)	106	0
Developmental process (GO:0032502)	31	0
Rhythmic process (GO:0048511)	1	0
Multicellular organismal process (GO:0032501)	9	0
Locomotion (GO:0040011)	13	0
Biological adhesion (GO:0022610)	16	0
Metabolic process (GO:0008152)	420	3
Growth (GO:0040007)	1	0
Immune system process (GO:0002376)	4	0
Biological regulation (GO:0065007)	0	1
Molecular function		
Transporter activity (GO:0005215)	41	1
Translation regulator activity (GO:0045182)	16	1
Transcription regulator activity (GO:0140110)	45	0
Catalytic activity (GO:0003824)	386	5
Cytoskeletal motor activity (GO:0003774)	3	0
Molecular function regulator (GO:0098772)	36	0
ATP-dependent activity (GO:0140657)	29	0
Molecular transducer activity (GO:0060089)	25	0
Molecular adaptor activity (GO:0060090)	17	0
Structural molecule activity (GO:0005198)	5	3
Binding (GO:0005488)	391	3
Molecular function regulator (GO:0098772)	0	1

also the months with a higher presence of bacteria and possible pathogens in the water [16]. These results suggest that at short-term exposure to *V. anguillarum* was not able to provoke a clear effect on the coelomocyte protein expression to measure differences in the perivisceral fluid or cell lysates. Indeed, the bacterial challenge was able to activate gene expression in coelomocytes indicating that *P. lividus* recognizes bacteria as a threat. Among these genes, 4 different lectins such as alpha-N-acetylgalactosamine lectin and C-type lectin Thr-1 which binds oligosaccharides present in bacterial cell wall [53] and two lectins related with the endoplasmic reticulum were found, the malectin and endoplasmic reticulum lectin isoform X1. These lectins are related to folding, sorting and targeting of glycoproteins through the secretory pathway [54], which function was significantly enriched in challenged animals. These proteins, in association with vesicular integral membrane protein (VIP36) that was also up-regulated in exposed animals, are involved in the sorting and transport of proteins from ER to Golgi apparatus to maturation [54,55].

As it could be observed, the gene expression showed some functions that were significantly up-regulated in *P. lividus* challenged against *V. anguillarum*. In fact, all the up-regulated genes encode functions related to autophagy, endocytosis, phagosome and protein processing in endoplasmic reticulum, which are closely immune-related.

Autophagy is a key process for recycling cytoplasmic components during stress. It is a widely conserved mechanism identified from yeast to mammals [56]. The occurrence of autophagy is regulated by numerous proteins and pathways but serine/threonine protein kinases play a major role on it [56,57]. Two serine/threonine kinases were found up-regulated in autophagy functions as well as several Rab related proteins which are small GTPases that are involved in the formation of autophagosomes in the macroautophagy process [58]. Although autophagy was mainly studied in the response of sea urchin embryos to several contaminants [59], their role in immune response also appears

Table 5

Enriched KEGG pathways on challenged group according to gProfiler analysis.

KEGG Pathway	Gene ID	Description
Autophagy KEGG:04140	LOC373418	GTP-binding protein Rab 1, transcript variant X3
	LOC373525	MAP kinase
	LOC576066	dual specificity mitogen-activated protein kinase 1, transcript variant X1
	LOC580852	serine/threonine-protein phosphatase 2A catalytic subunit beta isoform
	LOC588017	ubiquitin-like modifier-activating enzyme ATG7
	LOC752803	endophilin-B1, transcript variant X3
	rab7	rab7 GTPase homolog SURab7
	rab8	GTPase homolog
	LOC373178	kinesin heavy chain
	LOC574910	charged multivesicular body protein 2a
Endocytosis KEGG:04144	LOC575201	sorting nexin-3
	LOC575255	ras-related protein Rab-35, transcript variant X1
	LOC575871	ADP-ribosylation factor 4
	LOC576276	heat shock cognate 71 kDa protein
	LOC576486	E3 ubiquitin-protein ligase NRDP1, transcript variant X1
	LOC577320	charged multivesicular body protein 6
	LOC577465	epsin-2, transcript variant X1
	LOC577633	ADP-ribosylation factor 1-like 2
	LOC578627	ras-related protein Rab-5B, transcript variant X1
	LOC581581	charged multivesicular body protein 5
Phagosome KEGG:04145	LOC581807	vacuolar protein sorting-associated protein 28 homolog
	LOC582450	ADP-ribosylation factor
	LOC584477	cell division cycle 42 (GTP binding protein, 25 kDa)-like
	LOC585665	EH domain-containing protein 1
	LOC586454	vacuolar protein sorting-associated protein VTA1 homolog
	LOC591745	vacuolar protein sorting-associated protein 29
	LOC752803	endophilin-B1, transcript variant X3
	rab11	rab11 GTPase homolog SURab11p
	rab7	rab7 GTPase homolog SURab7
	rab8	GTPase homolog
Protein processing in endoplasmic reticulum KEGG:04141	Sp-Smad2/3	Sp-Smad2/3
	c3	complement component C3
	calret	calreticulin
	LOC373364	integrin beta-C subunit
	LOC575822	V-type proton ATPase subunit F
	LOC578627	ras-related protein Rab-5B, transcript variant X1
	LOC581537	ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac 1)-like
	LOC583066	ras-related C3 botulinum toxin substrate 1
	LOC584331	tubulin alpha-1 chain
	LOC590766	vesicle trafficking protein SEC22b
Protein processing in endoplasmic reticulum KEGG:04141	LOC593221	V-type proton ATPase 16 kDa proteolipid subunit
	LOC593690	ras-related C3 botulinum toxin substrate 1
	LOC753734	tubulin alpha chain
	rab7	rab7 GTPase homolog SURab7
	calret	calreticulin
	ECaSI/ PDI	ER chaperone
	Gp96	heat shock protein gp96
	hsp 110	egg receptor for sperm
	LOC576276	heat shock cognate 71 kDa protein
	LOC578220	dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit dad 1, transcript variant X2
Protein processing in endoplasmic reticulum KEGG:04141	LOC579272	ubiquitin recognition factor in ER-associated degradation protein 1
	LOC579681	UBX domain-containing protein 1
Protein processing in endoplasmic reticulum KEGG:04141	LOC580188	ubiquitin thioesterase OTU1

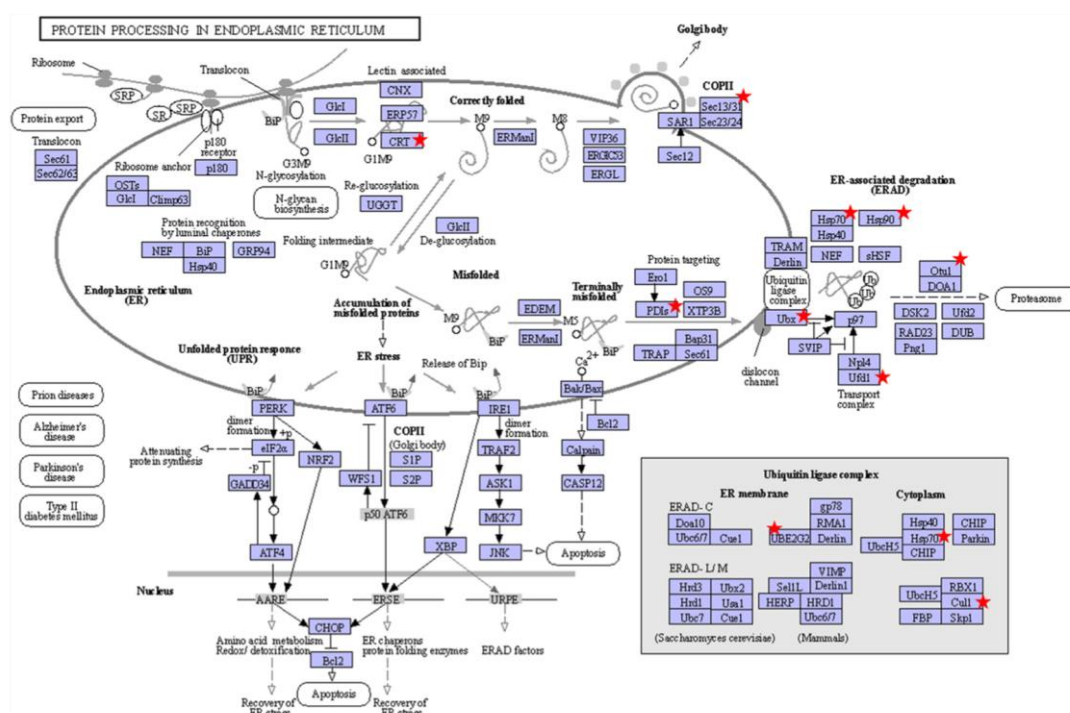
(continued on next page)

KEGG Pathway	Gene ID	Description
	LOC583776	protein SEC13 homolog
	LOC585800	ubiquitin-conjugating enzyme E2 G2
	LOC586525	protein disulf de-isomerase A4
	LOC588440	dnaJ homolog subfamily B member 11
	LOC589630	cullin-1

Pathways related with endocytosis and phagosome have already been described in *P. lividus* injected with LPS. A proteomic study of coelomocytes showed these functions up-regulated at 1, 3 and 6 h post injection, while the phagocytosis was the only pathway still up-regulated at 24 h post injection [60]. Endocytosis is a general term describing a process in which cells internalize external material through the membrane. It is usually divided in pinocytosis and phagocytosis, depending on the size of the engulfed molecule. In this case, it seems that coelomocytes are being activated by molecules released from bacteria. As the phagocytosis pathway was not clearly up-regulated, it seems that at this point coelomocytes are engulfing material secreted by *V. anguillarum* in order to destroy it and to present antigens to other cells via toll-like receptors (TLR), scavenger receptors cysteine-rich (SRCR) or nucleotide-binding oligomerisation domain like receptors (NLR) [61]. In this work, 3 TLRs, 5 SRCR and 17 NLR were also up-regulated in animals challenged with *V. anguillarum* (Supplementary Table VI) suggesting that endocytosis was the main way of presenting antigens to initiate the response against the pathogen. Also, a high quantity of small GTPases that act as membrane organizers and mediate directionality and specificity of vesicles [62] was identified. This point shows the high immune

In parallel to autophagy, endocytosis and phagosome pathways, the protein processing in endoplasmic reticulum (PPER) showed alterations. Proteins could follow three major pathways on the ER. Firstly, the production of proteins to be secreted to extracellular space or new proteins to fight against the pathogen via maturation in Golgi body, in which previously described lectins seem to be involved. The second one could be related to the unfolded protein response (UPR). Upon ER stress, cells activate mechanisms to cope with protein folding alterations. Once ER reaches an irreversible stress, this pathway eliminates damaged cells by apoptosis or autophagy [64,65]. And the third one is endoplasmic-reticulum-associated protein degradation (ERAD), in which misfolded proteins are ubiquitinated and the subsequent degradation is done via proteasome [66]. In this case, the high quantity of proteins related to the ubiquitin ligase complex indicates that up-regulation of PPER is related to ERAD pathway and in consequence to the degradation of proteins through proteasome complex.

This work presents an overview of the main immune genes involved in the pathogen recognition and immune response against the opportunistic bacteria *V. anguillarum* by the sea urchin *P. lividus* utilising a novel reference transcriptome of the coelomocytes constructed de novo. Also, a comparison between sexes identified some genes differentially expressed between males and females but without any function enriched. On the other hand, several functions were found significantly enriched in challenged animals to bacteria. Autophagy, endocytosis



8

S. Fernández-Boo et al.

Fish and Shellfish Immunology 147 (2024) 109431

phagosome and protein processing in endoplasmic reticulum pathways were up-regulated in challenged animals which indicated that a higher activation of the immune response occurred just at 4 h post exposure. Our work significantly increases the knowledge on echinoderm immunity and also expand the *omic* resources available for this key species in ecosystem functioning.

CRediT authorship contribution statement

Sergio Fernández-Boo: Data curation, Formal analysis, Investigation, Methodology, Visualization, Resources, Project administration, Writing – original draft. **André Machado:** Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft. **L. Filipe C. Castro:** Methodology, Validation, Writing – review & editing. **Rita Azeredo:** Formal analysis, Investigation, Methodology, Visualization, Writing – review & editing. **Benjamin Costas:** Conceptualization, Methodology, Validation, Supervision, Resources, Writing – review & editing.

Data availability

All data included in this research is available on NCBI and the supplementary data file of the manuscript

Acknowledgements

This work was supported by the project CRAGIAMP – PTDC/BIA-BQM/30232/2017, co-financed by COMPETE 2020, Portugal 2020, and the European Union through the ERDF. This research was also supported by national funds through FCT – Foundation for Science and Technology within the scope of UIDB/0443/2020 and UIDP/04423/2020. SFB, RA, BC and LFC were financed by the FCT Scientific Employment Stimulus program with reference CEECINST/00027/2021/CP2789/CT0002, 2022.03248.CEECIND, 2020.00290.CEECIND and 10.54499/CEECINST/00133/2018/CP1510/CT0004 respectively.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.fsi.2024.109431>.

References

- [1] E. Paredes, Biobanking of a marine invertebrate model organism: the sea urchin, *J. Mar. Sci. Eng.* 4 (7) (2016), 10.3390.
- [2] F. Mariétaz, et al., Analysis of the *P. lividus* sea urchin genome highlights contrasting trends of genomic and regulatory evolution in deuterostomes, *Cell Genomics* 3 (4) (2023) 100295.
- [3] C.F. Boudouresque, M. Verlaque, Ecology of *Paracentrotus lividus*, in: J. M. Lawrence (Ed.), *Developments in Aquaculture and Fisheries Science*, Elsevier, 2001, pp. 177–216.
- [4] E. Yeruham, et al., Collapse of the echinoid *Paracentrotus lividus* populations in the Eastern Mediterranean—result of climate change? *Sci. Rep.* 5 (2015) 13479.
- [5] M. Sweet, M. Bulling, J.E. Williamson, New disease outbreak affects two dominant sea urchin species associated with Australian temperate reefs, *Mar. Ecol. Prog. Ser.* 551 (2016) 171–183.
- [6] D.A. Leite Figueiredo, et al., Ocean acidification affects parameters of immune response and extracellular pH in tropical sea urchins *Lytechinus variegatus* and *Echinometra lucunter*, *Aquat. Toxicol.* 180 (2016) 84–94.
- [7] I. Dyková, et al., Neoparamoeba branchiphila infections in moribund sea urchins *Diadema aff. antillarum* in Tenerife, Canary Islands, Spain, *Dis. Aquat. Org.* 95 (3) (2011) 225–231.
- [8] S. Clemente, et al., Sea urchin *Diadema africanum* mass mortality in the subtropical eastern Atlantic: role of waterborne bacteria in a warming ocean, *Mar. Ecol. Prog. Ser.* 506 (2014) 1–14.
- [9] J.C. Bauer, C.M. Young, Epidermal lesions and mortality caused by vibriosis in deep-sea Bahamian echinoids: a laboratory study, *Dis. Aquat. Org.* 39 (3) (2000) 193–199.
- [10] K.W.P. Gilles, S. J., Disease in sea urchin *Strongylocentrotus purpuratus*: experimental infection and bacterial virulence, *Dis. Aquat. Org.* 1 (1986) 105–114.
- [11] D. Grech, et al., *Vibrio splendidus* clade associated with a disease affecting *Paracentrotus lividus* (Lamarck, 1816) in Sardinia (Western Mediterranean), *J. Invertebr. Pathol.* 192 (2022) 107783.
- [12] M. Sweet, Chapter 12 - sea urchin diseases: effects from individuals to ecosystems, in: J.M. Lawrence (Ed.), *Developments in Aquaculture and Fisheries Science*, Elsevier, 2020, pp. 219–226.
- [13] F. Gizzi, et al., Before and after a disease outbreak: tracking a keystone species recovery from a mass mortality event, *Mar. Environ. Res.* 156 (2020) 104905.
- [14] D. Girard, et al., A mass mortality of subtidal intertidal populations of the sea urchin *Paracentrotus lividus*: analysis of potential links with environmental conditions, *Mar. Ecol. Prog. Ser.* 33 (3) (2012) 377–385.
- [15] P.T. Becker, E. Egea, I. Eeckhaut, Characterization of the bacterial communities associated with the bald sea urchin disease of the echinoid *Paracentrotus lividus*, *J. Invertebr. Pathol.* 98 (2) (2008) 136–147.
- [16] S. Fernández-Boo, et al., Annual Assessment of the Sea Urchin (*Paracentrotus lividus*) Humoral Innate Immune Status: Tales from the North Portuguese Coast, *Marine Environmental Research*, 2018.
- [17] V. Matranga, et al., Monitoring chemical and physical stress using sea urchin immune cells, *Prog. Mol. Subcell. Biol.* 39 (2005) 85–110.
- [18] A. Pansino, V. Matranga, Sea urchin immune cells as sentinels of environmental stress, *Dev. Comp. Immunol.* 49 (1) (2015) 198–205.
- [19] A. Romero, B. Novoa, A. Figueras, Cell mediated immune response of the Mediterranean sea urchin *Paracentrotus lividus* after PAMPs stimulation, *Dev. Comp. Immunol.* 62 (2016) 29–38.
- [20] V. Queiroz, et al., Comparative study of coelomocytes from *Arbacia lixula* and *Lytechinus variegatus*: cell characterization and in vivo evidence of the physiological function of vibratile cells, *Fish Shellfish Immunol.* 110 (2021) 1–9.
- [21] C.J. Coates, et al., Echinochrome A release by red spherule cells is an iron-withholding strategy of sea urchin innate immunity, *J. Innate Immun.* 10 (2) (2018) 119–130.
- [22] F. Ramírez-Gómez, J.E. García-Arriaga, Echinoderm immunity, *Invertebr. Surviv. J.* 7 (2) (2010) 211–220.
- [23] M. Chiaramonte, et al., Stress and immune response to bacterial LPS in the sea urchin *Paracentrotus lividus* (Lamarck, 1816), *Fish Shellfish Immunol.* 92 (2019) 384–394.
- [24] C. Sea Urchin Genome Sequencing, The genome of the sea urchin *Strongylocentrotus purpuratus*, *Science (New York, N.Y.)* 314 (5801) (2006) 941–952.
- [25] J.P. Rast, et al., Genomic insights into the immune system of the sea urchin, *Science* 314 (5801) (2006) 952–956.
- [26] P. Pagliara, L. Stabili, Zinc effect on the sea urchin *Paracentrotus lividus* immunological competence, *Chemosphere* 89 (5) (2012) 563–568.
- [27] L. Stabili, P. Pagliara, The sea urchin *Paracentrotus lividus* immunological response to chemical pollution exposure: the case of lindane, *Chemosphere* 134 (2015) 60–66.
- [28] M. Di Natale, et al., Aberrant gene expression profiles in Mediterranean sea urchin reproductive tissues after metal exposures, *Chemosphere* 216 (2019) 48–58.
- [29] A.M. Machado, et al., The male and female gonad transcriptome of the edible sea urchin, *Paracentrotus lividus*: identification of sex-related and lipid biosynthesis genes, *Aquaculture Reports* 22 (2022) 100936.
- [30] S. Fernández-Boo, et al., Is palial mucus involved in *Ostrea edulis* defenses against the parasite *Bonamia ostreae*? *J. Invertebr. Pathol.* 169 (2020) 107259.
- [31] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics* 30 (15) (2014) 2114–2120.
- [32] L. Song, L. Florea, Rcorrecor: efficient and accurate error correction for Illumina RNA-seq reads, *GigaScience* 4 (1) (2015) 48.
- [33] M.G. Grabherr, et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome, *Nat. Biotechnol.* 29 (7) (2011) 644–652.
- [34] B.J. Haas, et al., De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis, *Nat. Protoc.* 8 (8) (2013) 1494–1512.
- [35] N.R. Coordinators, Database resources of the national center for biotechnology information, *Nucleic Acids Res.* 44 (D1) (2016) D7–D19.
- [36] N.M. Davidson, A. Oshlack, Corset: enabling differential gene expression analysis for de novo assembled transcriptomes, *Genome Biol.* 15 (7) (2014) 410.
- [37] F.A. Simão, et al., BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, *Bioinformatics* 31 (19) (2015) 3210–3212.
- [38] R. Smith-Unna, et al., TransRate: reference-free quality assessment of de novo transcriptome assemblies, *Genome Res.* 26 (8) (2016) 1134–1144.
- [39] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, *Nat. Methods* 9 (4) (2012) 357–359.
- [40] A. Gomes-dos-Santos, et al., The gill transcriptome of threatened European freshwater mussels, *Sci. Data* 9 (1) (2022) 494.
- [41] D.H. Jacques Dainat, Pucholt Pascal, NBISweden/AGAT: AGAT-v0.6.0 (v0.6.0), Zenodo, 2021.
- [42] P. Jones, et al., InterProScan 5: genome-scale protein function classification, *Bioinformatics* 30 (9) (2014) 1236–1240.
- [43] C. Camacho, et al., BLAST+: architecture and applications, *BMC Bioinf.* 10 (1) (2009) 421.
- [44] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, *Nat. Methods* 12 (1) (2015) 59–60.
- [45] K.D. Pruitt, T. Tatusova, D.R. Maglott, NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, *Nucleic Acids Res.* 35 (Database issue) (2007) D61–D65.
- [46] D. Powell, Drpowl/Degust 4.1.1, Zenodo, 2019.
- [47] U. Raudvere, et al., g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update), *Nucleic Acids Res.* 47 (W1) (2019) W191–W198.

S. Fernández-Boo et al.

Fish and Shellfish Immunology 147 (2024) 109431

- [48] R. Pérez-Portela, C. Leiva, Sex-specific transcriptomic differences in the immune cells of a key atlantic-mediterranean sea urchin, *Front. Mar. Sci.* 9 (2022) 908387.
- [49] V. Arizza, et al., Gender differences in the immune system activities of sea urchin *Paracentrotus lividus*, *Comp. Biochem. Physiol. Mol. Integr. Physiol.* 164 (3) (2013) 447–455.
- [50] M. Byrne, Annual reproductive cycles of the commercial sea urchin *Paracentrotus lividus* from an exposed intertidal and a sheltered subtidal habitat on the west coast of Ireland, *Mar. Biol.* 104 (2) (1990) 275–289.
- [51] M.A. Barela Huddell, L. Grayfer, L.C. Smith, Coelomocyte populations in the sea urchin, *Strongylocentrotus purpuratus*, undergo dynamic changes in response to immune challenge, *Front. Immunol.* 13 (2022).
- [52] Q. Jiang, et al., Transcriptional activation and translocation of ancient NOS during immune response, *FASEB J.* 30 (10) (2016) 3527–3540.
- [53] M.E. Mnich, R. van Dalen, N.M. van Sorge, C-type lectin receptors in host defense against bacterial pathogens, *Front. Cell. Infect. Microbiol.* 10 (2020).
- [54] J.D. Schrag, et al., Lectin control of protein folding and sorting in the secretory pathway, *Trends Biochem. Sci.* 28 (1) (2003) 49–57.
- [55] T. Schallus, et al., Malectin: a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein N-glycosylation, *Mol. Biol. Cell* 19 (8) (2008) 3404–3414.
- [56] K. Chen, Y. Shao, C. Li, ULK induces autophagy by targeting Beclin-1 in *Vibrio splendidus* challenged *Apostichopus japonicus*, *Aquaculture* 532 (2021) 736022.
- [57] Y. Zhan, et al., A novel p38 MAPK gene in the sea cucumber *Apostichopus japonicus* (Ajp38) is associated with the immune response to pathogenic challenge, *Fish Shellfish Immunol.* 81 (2018) 250–259.
- [58] C.F. Bento, et al., The role of membrane-trafficking small GTPases in the regulation of autophagy, *J. Cell Sci.* 126 (5) (2013) 1059–1069.
- [59] R. Chiarelli, et al., Autophagy as a defense strategy against stress: focus on *Paracentrotus lividus* sea urchin embryos exposed to cadmium, *Cell Stress & Chaperones* 21 (1) (2016) 19–27.
- [60] L. Inguglia, et al., Changes in the proteome of sea urchin *Paracentrotus lividus* coelomocytes in response to LPS injection into the body cavity, *PLoS One* 15 (2) (2020) e0228893.
- [61] M. Chiaramonte, R. Russo, The echinoderm innate humoral immune response, *Ital. J. Zool.* 82 (3) (2015) 300–308.
- [62] M. González-Gaitán, H. Stenmark, Endocytosis and signaling: a relationship under development, *Cell* 115 (5) (2003) 513–521.
- [63] T. Hibino, et al., The immune gene repertoire encoded in the purple sea urchin genome, *Dev. Biol.* 300 (1) (2006) 349–365.
- [64] C. Hetz, The unfolded protein response: controlling cell fate decisions under ER stress and beyond, *Nat. Rev. Mol. Cell Biol.* 13 (2) (2012) 89–102.
- [65] T. Yoritoma, D.J. Klionsky, Endoplasmic reticulum stress: a new pathway to induce autophagy, *Autophagy* 3 (2) (2007) 160–162.
- [66] L. Lemus, V. Goder, Regulation of endoplasmic reticulum-associated protein degradation (ERAD) by ubiquitin, *Cells* 3 (3) (2014) 824–847.

5. Chapter 5 – A whole-body transcriptome assembly of the annelid worm *Hediste diversicolor*

André M. Machado*, Rodrigo Samico*, Marcos Domingues, Andreas Hagemann, Luísa M.P. Valente, Arne M. Malzahn, André Gomes-dos-Santos, Raquel Ruivo, Juan Carlos Navarro, Óscar Monroig, L. Filipe C. Castro

(*Joint First Authors)

Marine Genomics

(DOI: [/10.1016/j.j.margen.2024.101084](https://doi.org/10.1016/j.j.margen.2024.101084))



Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Data



A whole-body transcriptome assembly of the annelid worm *Hediste diversicolor*

André M. Machado^{a,b,1}, Rodrigo Samico^{a,b,1}, Marcos Domingues^a, Andreas Hagemann^c, Luísa M.P. Valente^{a,d}, Arne M. Malzahn^c, André Gomes-dos-Santos^a, Raquel Ruivo^a, Juan Carlos Navarro^e, Óscar Monroig^{e,**}, L. Filipe C. Castro^{a,b,*}

^a CIIMAR/ CIMAR — Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Avenida General Norton de Matos, S/N, 4450-208 Matosinhos, Portugal

^b Department of Biology, Faculty of Sciences, University of Porto, Rua do Campo Alegre 1021/1055, 4169-007 Porto, Portugal

^c Department of Fisheries and New Biomedicine Industry, SINTEF Ocean, Trondheim, Norway

^d CBAS, Instituto de Ciências Biomédicas de Abel Salazar, Universidade do Porto, 4050-313 Porto, Portugal

^e Instituto de Acuicultura Torre de la Sal (IATS), CSIC, 12595, Ribera de Cabanes, Castellón, Spain

ARTICLE INFO

Keywords:
Ragworm
Annelida
Nereididae
Hediste
Aquaculture

ABSTRACT

The Annelida phylum is composed of a myriad of species exhibiting key phenotypic adaptations. They occupy key ecological niches in a variety of marine, freshwater and terrestrial ecosystems. Importantly, the increment of omic resources is rapidly modifying the taxonomic landscape and knowledge of species belonging to this phylum. Here, we comprehensively characterised and annotated a transcriptome of the common ragworm, *Hediste diversicolor* (OF Müller). This species belongs to the family Nereididae and inhabits estuarine and lagoon areas on the Atlantic coasts of Europe and North America. Ecologically, *H. diversicolor* plays an important role in benthic food webs. Given its commercial value, *H. diversicolor* is a promising candidate for aquaculture development and production in farming facilities, under a circular economy framework. We used Illumina next-generation sequencing technology, to produce a total of 105 million (M) paired-end (PE) raw reads and generate the first whole-body transcriptome assembly of *H. diversicolor* species. This high-quality transcriptome contains 69,335 transcripts with an N50 transcript length of 2313 bp and achieved a BUSCO gene completeness of 97.7% and 96% in Eukaryota and Metazoa lineage-specific profile libraries. Our findings offer a valuable resource for multiple biological applications using this species.

1. Introduction

The emergence of Next-Generation Sequencing (NGS) has provided the ability to sequence DNA at great speed and accuracy, allowing multiple approaches and innovative biological applications (Zhang et al., 2011). NGS emerged in the aftermath of the Human Genome Project (Gibbs, 2020) and revolutionized the sequencing of DNA and RNA, making it more automated and cost-effective. This advancement led to exponential growth in our understanding and capacity to tackle biological questions (Behjati and Tarpey, 2013). Furthermore, the continuous expansion of transcriptomic datasets has introduced

innovative strategies to enhance the conservation and management of important species (Connon et al., 2018).

The common ragworm, scientifically known as *Hediste diversicolor* (O.F. Müller, 1776), inhabits estuarine and lagoon areas of the Atlantic coasts of Europe and North America (Virgilio and Abbiati, 2004) (Fig. 1a). This aquatic annelid species belongs to the Nereididae family, *Hediste* genus (old *Nereis* genus), and holds significant ecological importance and commercial value (Einfeldt et al., 2014; Pombo et al., 2020). Its ecological contributions encompass the construction of burrows, through sediment reworking by bioturbation activity, which enhances the sediment-water interface and create new micro-

* Corresponding author at: CIIMAR/ CIMAR — Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Avenida General Norton de Matos, S/N, 4450-208 Matosinhos, Portugal.

** Corresponding author.

E-mail addresses: oscar.monroig@csic.es (Ó. Monroig), filipe.castro@ciimar.up.pt (L.F.C. Castro).

¹ Equal contribution

<https://doi.org/10.1016/j.margen.2024.101084>

Received 20 October 2023; Received in revised form 5 January 2024; Accepted 6 January 2024

Available online 12 January 2024

1874-7787/© 2024 Elsevier B.V. All rights reserved.

environments (Gillet et al., 2018). Due their remarkable ability to tolerate wide fluctuations in salinity (Neuhoff, 1979; Wolff, 1973) and temperature (Newell and Branch, 1980), *H. diversicolor* specimens are commonly found across a broad geographic range. Consequently, they are frequently employed for toxicity assessment and biomonitoring of pollution levels (Buikema, 1980; Ozoh, 1992). However, this species faces increasing threats due to its exploitation as fishing bait and the absence of proper capture management (Scaps, 2002). This difficulty is exacerbated by the overlap between their reproductive peak (spring and summer), and sport fishing season. Compounding the issue is that *H. diversicolor* can reproduce only once during their lifetime (Olive and Garwood, 1981). In this context, the development of intensive production protocols is valuable. While a standard and successful practice is not yet available as occurs in other nereid counterparts such as *Alitta virens*, their ability to thrive on a variety of food sources including sidestrems (Malzahn et al., 2023) has made *H. diversicolor* a promising candidate to be commercially exploited in farming facilities with recirculation aquaculture system (RAS) or in integrated multitrophic aquaculture (IMTA) set-ups (Pombo et al., 2020).

In this study, we employed a combination of NGS approaches and a comprehensive bioinformatics workflow to create a transcriptome for *H. diversicolor*. Despite the numerous pertinent studies on this species, which have primarily centred on its life cycle, distribution, and toxicology, there is a scarcity of omic resources datasets (Gomes-dos-Santos et al., 2021; Moutinho Cabral et al., 2022). We anticipate that our transcriptome assembly provides opportunities for future research with this organism, aimed at enhancing its conservation and promoting a sustainable exploration and exploitation.

2. Data description

2.1. Sampling, RNA extraction, and Illumina sequencing

An adult *H. diversicolor* individual was collected at Trondheim Fjord, in Leangbukta, Norway at 63.439151 N, 10.474605 E (Fig. 1a, Table 1). Initially, the whole organism was stored in RNA later and further deposited at the Interdisciplinary Center of Marine and Environmental Research – CIIMAR (Filipe Castro, filipe.castro@ciimar.up.pt) under the voucher number 4HDIIV3.

Total RNA extraction from the head and the middle body (reciprocal distance from the head and the posterior end - pygidium) was performed by NZY Total RNA Isolation kit (NZYTech, Lda. - Genes and Enzymes)

Table 1
MixS descriptors of the *Hediste diversicolor* sample.

	Value
Investigation_type	Eukaryote
Project_name	<i>Hediste diversicolor</i> transcriptome
Lat_lon	63.439151 N, 10.474605 E
Geo_loc_name	Norway: Trondheim Fjord, Leangbukta
Collection_date	10/07/2020
Biome	Coastal sea water (ENVO:00002150)
Feature	Coastal water body (ENVO:02000049)
Material	Sea water (ENVO:00002150)
Env_package	RNA later
Seq_meth	Illumina HiSeq 4000
Assembly_method	Trinity
Collector	Andreas Hagemann
Tissues	Head, Middle body
Maturity	Adult
Tissue sample size	1 cm
Assembly	Trinity v.2.13.2
BioProject accession	PRJNA860539
TSA	GKPK0100000

following the manufacturer's recommendations. A DS-11 Series Spectrophotometer was used to evaluate RNA concentration (ng/l) and quality (OD260/280 ratio values). Strand-specific libraries were built by Macrogen, Inc. These libraries included inserts between 250 and 300 bp, and were sequenced on the Illumina HiSeq4000 platform using 150 bp paired-end reads.

2.2. Processing of sequencing datasets and transcriptome assembly

Initially, we inspected the raw datasets (head and middle body) using the FastQC software (version 0.11.8) (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Next, the Trimmomatic (version 0.38) (Bolger et al., 2014) software was used to quality-filter the raw reads and remove any Illumina adapters (Parameters: LEADING:5 TRAILING:5 SLIDINGWINDOW:4:5 MINLEN:36). In the end, the Rcorrector (version 1.0.3) (Song and Florea, 2015), a kmer-based error correction method was executed with the default settings to correct any random sequencing errors (Fig. 1b). Importantly, this software corrected about 30 M of the 101 M reads in first sample and 28 M of the 108 M reads in the second one. Afterwards, the FastQC was applied to verify the final quality of the dataset. Despite the conservatory approach used, few reads were discarded during the pre-processing stage and all final reads of both

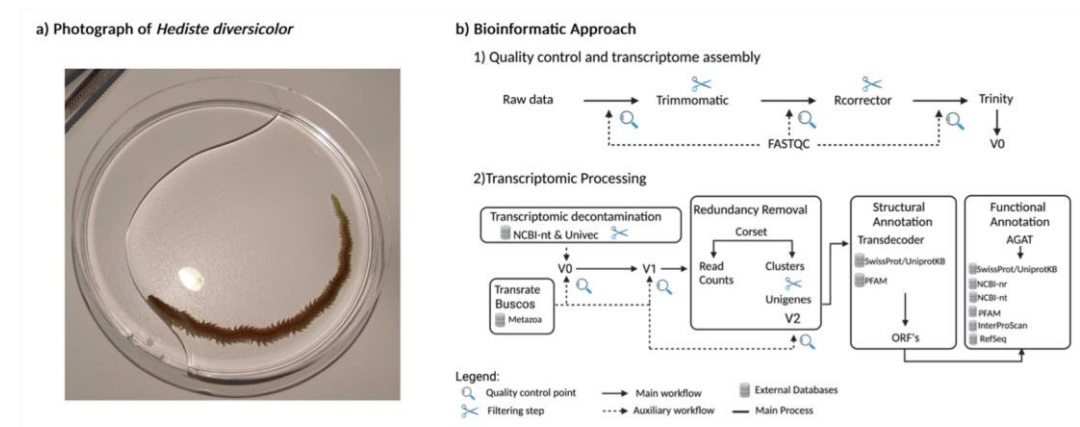


Fig. 1. a) Photograph of common ragworm, *Hediste diversicolor*. b) Bioinformatic workflow used in the transcriptomic approach. The workflow was created with BioRender.com. Figure should be coloured.

samples presented Phred scores higher than 30.

To build the *H. diversicolor* transcriptome assembly, we used the Trinity (version 2.13.2) (Grabherr et al., 2011; Haas et al., 2013) software with the default settings and a single k-mer approach. Before run the Trinity software, both datasets were merged. Next, the merged file was submitted to Trinity, with the default parameters, and the initial transcriptome assembly (Assembly V0) produced. To assess the V0 transcriptome assembly, we used Transrate (version 1.0.1) (Smith-Unna et al., 2016) software to assess general metrics of the assembly, the Benchmarking Universal Single-Copy Orthologs (BUSCO version 3.0.2) (Simão et al., 2015) – with two lineage-specific cprof libraries (Metazoa and Eukaryota) – to determine the gene completeness of the assembly and a read back mapping (RBM) approach to determine the percentage of clean reads present in the transcriptome assembly. To align the reads against the transcriptome assembly we used the Bowtie2 (version 2.3.5) (Langmead and Salzberg, 2012) software and to quantify the percentage of reads mapping we used the samtools stats (version 1.9) (Li et al., 2009). In this version of the transcriptome assembly, we found 280,456 transcripts, with an n50 transcript length of 2512 bp, an n90 transcript length of 361 bp, and a mean transcript length of 1081.46 bp (Table 2). The BUSCO analysis presented a high level of gene completeness in both databases, 100% (S: 13.2%; D: 86.8%) and 99.8% (S: 14.3%; D: 85.5) in Eukaryota and Metazoa databases, respectively (Table 2). The RBM analyses showed about 98.10% of the initial clean reads present in the transcriptome (Table 2).

2.3. Post-assembly processing stage

After the transcriptome assembly and initial validations, we ran the decontamination analysis. In summary, we used the nucleotide database of NCBI (nt-NCBI; downloaded on 17/02/2022) and UniVec (downloaded on 17/02/2022) database to identify and remove possible sources of contamination (e.g., vectors, adapters, biological contamination). These analyses were done using blast-n (version 2.12.0) versus the nt-NCBI dataset with the settings (-num_threads 15 -max_target_seqs 1 -max_hsps 1) and versus the Univec dataset with the parameters (-reward 1 -penalty -5 -gapopen 3 -gapextend 3 -dust yes -soft_masking true -evalue 700 -searchsp 1,750,000,000,000 -outfmt 6 -num_threads

15). To the nt-NCBI searches, all the transcripts with the best match hits out of the Lophotrochozoa taxon (Taxonomy ID: 1206795) were considered contaminations and were excluded from the V0 transcriptome assembly. The remaining contigs were preserved in the transcriptome and used in subsequent analyses, whether they had or had not match hits in the Lophotrochozoa taxon. All transcripts having a match hit in the Univec database were deemed exogenous to the *H. diversicolor* transcriptome and removed from the dataset. In the end of this process, we obtained the decontaminated assembly (Assembly V1) (Table 2). To remove redundancy of the datasets we executed the Corset tool (Davidson and Oshlack, 2014). The program removes transcripts with less than 10 read counts (spurious transcripts) and hierarchically cluster the remaining transcripts based on the fraction of common reads multi-mapped to the transcriptome. Next, the fetchClusterSeqs.py script of Corset-tools (version 1) (<https://github.com/Adamtaranto/Corset-tools/blob/master/fetchClusterSeqs.py>) with the parameter (-longest) was used to select the longest transcripts from each cluster (Assembly V2).

To perform the assessment of V1 and V2 versions of the transcriptome assembly we used the same strategy used to V0. While the TransRate analyses showed an n50 transcript length of 2487 (V1) and 2313 bp (V2), and n90 transcript length of 358 (V1) and 556 bp (V2) the BUSCO analysis showed 99.6% (V1) and 97.7% (V2) of gene completeness in Eukaryota database and 99.1 (V1) and 96% (V2) against Metazoa database. Despite the large reducing in the number of transcripts from V1 (272,156) to V2 (69,335), the BUSCO analyses showed that almost no gene content was lost. The redundancy removal step affected mainly transcripts with small length (0-800 bp) and consequently the mean length of the transcripts of the transcriptome, 1071.22 (V1) and 1335.81 bp (V2) (Table 2). In Supplementary Fig. 1 and Supplementary Table S1, we can examine the distribution of sequence lengths in the V2 version of the *H. diversicolor* transcriptome assembly (Supplementary Fig. 1 and Table S1). Furthermore, Supplementary Fig. 1 and Supplementary Table S2 reveal that the number of isoforms was also modestly influenced by the redundancy removal process (Supplementary Fig. 1 and Table S2).

2.4. Transcriptome annotation

The annotation of the *H. diversicolor* transcriptome was subdivided in two steps, namely structural and functional annotation. First, to perform the structural annotation we executed the TransDecoder (version 5.3.0) (<https://github.com/TransDecoder/TransDecoder/>). TransDecoder allowed us to determine possible coding regions within transcript sequences and produce the open reading frames (ORFs). To do that, we used both homology and protein domain searches - homologies searches with Blast-p (Camacho et al., 2009) tool in UniProtKB/SwissProt database (Bateman et al., 2021), and domain searches with hmmscan software of hmmer package (version 3.3.2) in PFAM database (Finn et al., 2011).

To perform the functional annotation, we used a set of tools. Initially Gtf/Gff Analysis Toolkit (AGAT) (version 0.8.0) (Dainat, 2021) allowed us to uniformize and produce the initial structural annotation file (in gff3 format) using the transcriptome assembly file (.fasta) and the Transdecoder output file (.gff). In the end, the protein and transcript fasta files were obtained with names properly uniformized and formatted. In the last phase of the annotation, the functional annotation, we applied the InterProScan tool (version 5.44.80) (Jones et al., 2014) and Blast-n/p/x searches in several databases. While the InterPro and RefSeq databases were queried for the proteins per species using the Blast-p/x tool of DIAMOND software (version 0.9.24) (Buchfink et al., 2014; O'Leary et al., 2016), the NCBI-nt and NCBI-nr databases were searched for transcripts using the Blast-n/x tool of DIAMOND software. In the end, the AGAT program combined all blast (.outfmt6 files) and InterProScan (.tsv file) outputs into the gff3 annotation file. After the functional annotation, we conducted a detailed analysis of the BLAST searches

Table 2
Statistics of the *Hediste diversicolor* transcriptome assembly versions, raw (V0), decontaminated (V1), final (V2).

Raw Sequences	<i>Hediste diversicolor</i>		
Reads used in assembly	105,011,718		
Quality Metrics of Assembly	Assembly V0	Assembly V1	Assembly V2
Number of transcripts	280,456	272,156	69,335
N50 transcript length (bp)	2512	2487	2313
N90 transcript length (bp)	361	358	556
Transcriptome length (mb)	330.6	298.9	96.8
Mean transcript length (bp)	1081.46	1071.22	1335.81
Longest transcript	22,036	22,036	22,036
Number of transcripts over 1 K nt	76,694	73,343	25,074
Number of transcripts over 10 K nt	48.97	366	181
GC %	0.40	0.40	0.39
Rate Mapping %	98.10	96.47	91.69
BUSCO Complete (euk/ met) %	100	99.6	97.7
Single	13.2/ 14.3	19.1/ 17.4	89.8/ 88.2
Duplicated	86.8/ 85.5	80.5/ 81.7	7.9/ 7.8
Fragmented	0.0/ 0.1	0.3/ 0.5	2.0/ 2.7
Missing	0.0/ 0.1	0.1/ 0.4	0.3/ 1.3
Annotation	–	–	–
Number of transcripts with ORF	–	–	32,006
Transcripts with Blast hit in at least one database	–	–	25,453
Transcripts with KEGG Pathways	–	–	1168
Transcripts with Interpro hit	–	–	21,745
Transcripts with PFAM	–	–	19,088
Transcripts with GO terms	–	–	13,402

(Blast-x) against the nr-NCBI database, using an E-value threshold of 1e-5. These analyses enabled us to identify 25,189 blast-x matches. In the supplementary Fig. 1 and Table S3 depict the top 10 species with the most significant blast-x matches against the nr-NCBI database. This outcome suggests a shortage of publicly available molecular resources for the Annelida phylum. In the Polychaeta class, the lack of genomic and transcriptomic resources is even more visible, with most species without available gene annotation resources. Upon closer examination, it becomes evident that only 32.75% of these hits corresponded to Annelida species (Supplementary Fig. 1 and Table S3). In the top five species of Supplementary Fig. 1c, we found that *Capitella telata* is the annelid with the highest number of sequences manually annotated in the non-redundant database of NCBI. Additionally, *Lingula anatina* is one of the closest phylogenetic species (non-annelid) with a sequenced genome, as well as structural and functional annotation in the non-redundant nucleotide database of NCBI (Luo et al., 2015).

3. Conclusion

In this study, we have generated the whole-body transcriptome assembly of the common ragworm *H. diversicolor*. The results and resources presented here represent a significant advancement in the study of this group of organisms and will contribute to the management of this species.

Author contributions

L.F.C.C. and O.M. designed and conceived this work.
A.H., A.M.M., collected the samples.
A.M.M., R.S., A.H., A.M.M., L.M.P.V., M.D., A.G.S., J.C.N., O.M. carry on all the analysis.
A.M.M., R.S., O.M. and L.F.C.C. wrote the first version of the manuscript.
All authors read, revised, and approved the final manuscript.

CRediT authorship contribution statement

André M. Machado: Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. **Rodrigo Samico:** Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. **Marcos Domingues:** Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – review & editing. **Andreas Hagemann:** Data curation, Formal analysis, Methodology, Resources, Writing – review & editing. **Luisa M.P. Valente:** Data curation, Funding acquisition, Investigation, Resources, Writing – review & editing. **Arne M. Malzahn:** Data curation, Formal analysis, Funding acquisition, Resources, Writing – review & editing. **André Gomes-dos-Santos:** Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – review & editing. **Raquel Ruivo:** Formal analysis, Investigation, Validation, Writing – review & editing. **Juan Carlos Navarro:** Data curation, Funding acquisition, Resources, Validation, Writing – review & editing. **Oscar Monroig:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Supervision, Validation, Writing – original draft, Writing – review & editing. **L. Filipe C. Castro:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare no competing interests (financial, professional, or personal relationships) that could influence the work here reported.

Data availability

The raw reads for each sample were deposited at the NCBI Sequence Read Archive under BioProject PRJNA860539. The transcriptome shotgun assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GPKK00000000. The version described in this paper is the first version, GPKK01000000. The remaining information, was uploaded to fshare (link: <https://doi.org/10.6084/m9.fshare.24236311.v1>). In detail, the files uploaded to fshare include the version V1 and V2 of the transcriptome assemblies, transcript files (Hediste_diversicolor_genes.fa.gz), messenger RNA file (Hediste_diversicolor_mrna.fa.gz), open reading frames predictions (Hediste_diversicolor_ods.fa.gz), open reading frames proteins predictions (Hediste_diversicolor_proteins.fa.gz) as well as the annotation files (Hediste_diversicolor_annotation_sorted.gff3.gz). All software, versions and parameters can be consulted in the Data Description section. Software programs with no parameters associated were used with the default settings.

Acknowledgements

This research was partially supported by the ERA-NET BlueBio COFUND Project SDESTREAM [Grant ID 68], co-funded through national funds provided by FCT [BLUEBIO/0005/2019]; Agência Estatal de Investigación [PCI2020-111960], NRC – Norwegian Research Council [#311701]. A.M.M. was funded by the Portuguese Foundation for Science and Technology (FCT) under the Grant [DFA/BD/8069/2020], and L.F.C.C. by 10.54499/CEECINST/00133/2018/CP1510/C0004. Additional strategic funding was provided by FCT [UIDP/04423/2020].

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.margen.2024.101084>.

References

- Bateman, A., Martin, M.J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Da Silva, A., Denny, P., Dogan, T., Ebenezer, T.G., Fan, J., Castro, L.G., Garmiri, P., Georgiou, G., Gonzales, L., Halton-Elis, E., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Jokinen, P., Joshi, V., Jyothi, D., Lock, A., Lopez, R., Luciani, A., Luo, J., Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Menchi, M., Mishra, A., Moulang, K., Nightingale, A., Oliveira, C.S., Pundir, S., Qi, G., Raj, S., Rice, D., Lopez, M.R., Saidi, R., Sampson, J., Sawford, T., Speretta, E., Turner, E., Tyagi, N., Vasudev, P., Volynkin, V., Warner, K., Watkins, X., Zaru, R., Zellner, H., Bridge, A., Poux, S., Redaschi, N., Almo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.C., Bolleman, J., Boutet, E., Breuza, L., Casals-Casas, C., de Castro, E., Echioukh, K.C., Coudert, E., Cuhe, B., Doche, M., Dornevil, D., Estreicher, A., Famiglietti, M.L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Keller, G., Kerhornou, A., Lara, V., Le Mercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T.B., Paesano, S., Pedruzzi, I., Pilboud, S., Pourcel, L., Pozzato, M., Pruess, M., Rivoire, C., Sigrist, C., Sonesson, K., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Ross, K., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.S., Zhang, J., 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 49, D480–D489. <https://doi.org/10.1093/nar/gkaa1100>.
- Behjati, S., Tarpey, P.S., 2013. What is next generation sequencing? Arch. Dis. Childhood - Educ. Practice 98, 236–238. <https://doi.org/10.1136/ARCHDISCHILD-2013-304340>.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Buchfink, B., Xie, C., Huson, D.H., 2014. Fast and sensitive protein alignment using DIAMOND. Nat. Methods <https://doi.org/10.1038/nmeth.3176>.
- Buikema, J., 1980. Aquatic Invertebrate Bioassays. ASTM International.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421. <https://doi.org/10.1186/1471-2105-10-421>.

- Connon, R.E., Jeffries, K.M., Komoroske, L.M., Todgham, A.E., Fanguy, N.A., 2018. The utility of transcriptomics in fish conservation. *J. Exp. Biol.* 221 <https://doi.org/10.1242/JEB.148833>/20435.
- Dainat, J., 2021. AGAT: Another GFF Analysis Toolkit to handle annotations in any GTF/GFF format. <https://doi.org/10.5281/zenodo.3552717>.
- Davidson, N.M., Oshlack, A., 2014. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biol.* 15, 410. <https://doi.org/10.1186/s13059-014-0410-6>.
- Einfeldt, A.L., Doucet, J.R., Addison, J.A., 2014. Phylogeography and cryptic introduction of the ragworm *Hediste diversicolor* (Annelida, Nereididae) in the Northwest Atlantic. *Invertebr. Biol.* 133, 232–241. <https://doi.org/10.1111/IVB.12060>.
- Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. <https://doi.org/10.1093/nar/gkr367>.
- Gibbs, R.A., 2020. The human genome project changed everything. *Nat. Rev. Genet.* 21 <https://doi.org/10.1038/s41576-020-0275-3>, 2020. 10 21, 575–576.
- Gillet, P.G., Mouloud, M., Mouneyrac, C., 2018. Chapter 6. The key role of the species *Hediste diversicolor* (Polychaeta, Nereididae) in estuarine ecosystems. In: Kovács, A., Nagy, P. (Eds.), *Advances in Marine Biology*, pp. 1–226.
- Gomes-dos-Santos, A., Hagemann, A., Valente, L., Malzahn, A.M., Monroig, Ö., Froute, E., Castro, L.F.C., 2021. Complete mitochondrial genome of the ragworm annelid *Hediste diversicolor* (of Müller, 1776) (Annelida: Nereididae). *Mitochondrial DNA Part B* 6, 2849–2851. <https://doi.org/10.1080/23802359.2021.1970644>.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mao, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. <https://doi.org/10.1038/nbt.1883>.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. <https://doi.org/10.1038/nprot.2013.084>.
- Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Masién, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Han, Y., Li, J., Lau, N., Lau, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Luo, Y.J., Takeuchi, T., Koyanagi, R., et al., 2015. The *Lingula* genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nat. Commun.* 6, 8301. <https://doi.org/10.1038/ncomms9301>.
- Malzahn, A.M., Villena-Rodríguez, A., Monroig, Ö., Johansen, A., Castro, L.F.C., Navarro, J.C., Hagemann, A., 2023. Diet rather than temperature determines the biochemical composition of the ragworm *Hediste diversicolor* (OF Müller, 1776) (Annelida: Nereididae). *Aquaculture* 569, 739368. <https://doi.org/10.1016/j.aquaculture.2023.739368>.
- Moutinho Cabral, I., Madeira, C., Grosso, A.R., Costa, P.M., 2022. A drug discovery approach based on comparative transcriptomics between two toxin-secreting marine annelids: *Glycera alba* and *Hediste diversicolor*. *Mol. Omics* 18, 731–744. <https://doi.org/10.1039/D2MO00138A>.
- Neuhoff, H.G., 1979. Influence of temperature and salinity on food conversion and growth of different *Nereis* species (Polychaeta, Annelida). *Mar. Ecol. Prog. Ser.* 1, e262.
- Newell, R.C., Branch, G.M., 1980. The influence of temperature on the maintenance of metabolic energy balance in marine invertebrates. *Adv. Mar. Biol.* 17, 329–396. [https://doi.org/10.1016/S0065-2881\(08\)60304-1](https://doi.org/10.1016/S0065-2881(08)60304-1).
- O'Leary, N.A., Wright, M.W., Brister, J.R., Chao, S., Haddad, D., McVeigh, R., Rajput, B., Robert, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetverin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V.S., Kodali, V.K., Li, W., Maglott, D., Masterson, P., McGarvey, K.M., Murphy, M.R., O'Neill, K., Pujar, S., Rangwala, S.H., Rausch, D., Riddick, L.D., Schoch, C., Shkeda, A., Storz, S.S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R.E., Vatsan, A. R., Wallin, C., Webb, D., Wu, W., Landrum, M.J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T.D., Pruitt, K.D., 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. <https://doi.org/10.1093/nar/gkv1189>.
- Olive, P.J.W., Garwood, P.R., 1981. Gametogenic cycle and population structure of *Nereis* (*Hediste*) *diversicolor* and *Nereis* (*Nereis*) *pelagica* from north-east England. *J. Mar. Biol. Assoc. U. K.* 61, 193–213. <https://doi.org/10.1017/S0025315400046014>.
- Ozoh, P.T.E., 1992. The importance of adult *Hediste* (*Nereis*) *diversicolor* in managing heavy metal pollution in shores and estuaries. *Environ. Monit. Assess.* 21, 165–171. <https://doi.org/10.1007/BF00399685>.
- Pombo, A., Baptista, T., Granada, L., Ferreira, S.M.F., Gonçalves, S.C., Anjos, C., Sã, E., Chainho, P., Cancela da Fonseca, L., Fidalgo e Costa, P., Costa, J.L., 2020. Insight into aquaculture's potential of marine annelid worms and ecological concerns: a review. *Rev. Aquac.* 12, 107–121. <https://doi.org/10.1111/raq.12307>.
- Scaps, P., 2002. A review of the biology, ecology and potential use of the common ragworm *Hediste diversicolor* (O.F. Müller) (Annelida: Polychaeta). *Hydrobiologia* 470, 203–218. <https://doi.org/10.1023/A:1015681605656>.
- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Smith-Unna, R.D., Boursnell, C., Patro, R., Hibberd, J.M., Kelly, S., Street, D., Brook, S., Road, S.P., 2016. TransRate: reference free quality assessment of de novo transcriptome assemblies. *Genome Res.* 26, 1134–1144. <https://doi.org/10.1101/gr.196469.115>.
- Song, L., Florea, L., 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *Gigascience* 4, 48. <https://doi.org/10.1186/s13742-015-0089-y>.
- Virgilio, M., Abbiati, M., 2004. Habitat discontinuity and genetic structure in populations of the estuarine species *Hediste diversicolor* (Polychaeta: Nereididae). *Estuar. Coast. Shelf Sci.* 61, 361–367. <https://doi.org/10.1016/j.ECSS.2004.06.005>.
- Wolff, W.J., 1973. The estuary as a habitat: an analysis of data on the soft-bottom macrofauna of the estuarine area of the rivers Rhine, Meuse, and Scheldt. *Zoologische Verhandlungen* 126, 1–242.
- Zhang, J., Chiodini, R., Badr, A., Zhang, G., 2011. The impact of next-generation sequencing on genomics. *J. Genet. Genom.* 38, 95–109. <https://doi.org/10.1016/J.JGG.2011.02.003>.

6. Chapter 6 – General Discussion and Future Perspectives

The aim of this thesis was to develop an evolutionary framework to address the genetic composition of the chemical defensome in Metazoa. By applying genomic, transcriptomic and comparative approaches, I expected to cover tree metazoan branches for which information is scarcer or absent. Additionally, I sought to implement two independent approaches, genomic and transcriptomic, as part of a global outline for the study of the chemical defensome across vertebrates and invertebrate species. Specifically, this framework encompassed both a general strategy for producing resources for non-model organisms and a species-specific approach to understand the interplay of chemical components with other biological systems. This was achieved through the application of diverse methodologies in four non-model species and described in the previous chapters. Furthermore, several catalogues of CD components were produced, multiple analytical frameworks were implemented, and the interactions between CD systems and other systems were highlighted. Overall, these resources, together with the genomic and transcriptomic approaches developed in this thesis, represent a comprehensive framework with practical application in chemical defensome research, and pave the way for further development.

6.1. The Chemical defensome: From its origin through concept to current frameworks in Metazoans

Early approaches of ecotoxicology and environmental toxicology typically studied the impact of chemicals in organisms by applying classical biochemical and molecular approaches. These approaches often relied on the use of individual detoxification enzymes as biomarkers: such as cytochromes P450, glutathione S-transferases or membrane transporters (e.g. ABC transporters). Researchers analysed linear cause-effects relationships, where exposure to a particular chemical was linked to a specific enzymatic activity (Dean et al., 2001; Guengerich, 2001; Hayes et al., 2005). Although foundational, these studies were largely narrow and reductionist in scope, neglecting other levels of biology organisation. The perspectives on chemical responding elements in organisms shifted with the integration of molecular biology subfields, particularly genomics and transcriptomics and the development of bioinformatics. This gave rise to fields of molecular ecotoxicology or mechanistic ecotoxicology. Instead of focusing solely on enzymatic activities as responses to chemical challenge, these approaches began to

explore gene induction and transcriptional regulation (Anderson et al. 1994; Sarkar et al. 2006). The integration of these subfields dramatically increased the quantity and accessibility of molecular information. Researches gained the capacity to describe entire gene catalogues, infer protein products from these genes, and functionally test such catalogues on large scale through genomic and transcriptomic high-throughput approaches (Anderson et al., 1994; Snell et al., 2003). These approaches were naturally complemented with other omics strategies, including lipidomic, metabolomic or proteomic. However, the rational and direction were primarily shaped by transcriptomic and genomic high-throughput methodologies. The impact of next-generation sequencing, allowed the expansion upon earlier biochemical findings to catalogue, discover, and experimentally test entire gene families, often using a combination of genomic, transcriptomic and evolutionary approaches (Dean & Annilo, 2005; Francis et al., 2003; Laudet et al., 1992). This extension produced such large datasets of information, that hierarchical organization became essential, with genes grouped by similarity, evolutionary relatedness and functions. In the midst of this era, the concept of chemical defensome emerged (Goldstone et al., 2006). This framework arose from the need systematize and categorize how organisms organize and orchestrate their molecular toolkit to response to novel and adverse environments, frequently shaped by the release of anthropogenic chemicals. When Goldstone and colleagues proposed the CD, they emphasised two core elements: the molecular toolkit employed in response to a chemical exposure, and the toxic effects of chemicals on the organism (Goldstone et al., 2006). Importantly, the authors established this framework by incorporating a hierarchical organization, grouping genes into four functional categories, encompassing the three classical categories of drug metabolism: phase I (oxidation/reduction), phase II metabolism (conjugation), phase III transport (efflux pumps) and a fourth phase involving regulatory elements were also considered. Interestingly, while Goldstone et al. (Goldstone et al., 2006) postulates this vision, it encompassed not only regulatory elements but also an evolutionary coherent perspective. This contrasts with the classical view of drug metabolism, which originally contemplated only the biochemical phases. In addition, by establishing this umbrella concept, the authors provided an integrative and complete view of the CD pathways, encompassing chemical sensing, biotransformation, and the excretion of chemicals. An interesting consideration was the inclusion of life cycle stages, a critical aspect to highlight the dynamics of the CD. Rather than simply identifying the presence or absence of genes, analysing expression levels provided a global overview of CD in *Strongylocentrotus purpuratus*, while aiming for a broad systematic coverage (Goldstone et al., 2006). In the years that followed, this conceptual framework of categorization and data production at genomic and transcriptomic level (as

applied in **Chapter 2, 3, 4.1, 4.2 and 5**), was rapidly expanded to other metazoan species, taking advantage from the large-scale NGS projects (De Marco et al., 2017; Eide et al., 2021; Franco, Araújo, et al., 2025; Goldstone, 2008; Goldstone et al., 2006; Marrone et al., 2012; Roncalli et al., 2025; Stegeman et al., 2010).

The expansion of molecular resources across metazoans enabled researchers to conduct comparative genomics analyses (as in **Chapter 2 and 3**) as well as toxicogenomic and toxic-transcriptomic studies under challenge conditions across taxa (as in **Chapter 4.2**) Such approaches answer key questions such as: How plastic is the repertoire, and what are its limits of action? Under which environmental conditions does a given repertoire of defensome elements constitute an ecological advantage or disadvantage? How is the CD related to the pollutant concentration and exposure levels? How do the different mechanism of adaptation interplay with chemical defensome? To what extent does chemical defensome repertoire influence the colonization or invasion of new environments? And more broadly, how plastic is this repertoire in enabling species to tolerate chemical changes in environment, whether driven by Anthropocene pollution or species-specific processes? Finally, at species level, how lineages gain key innovation by modifying chemical defensome elements, in several organisms? (Franco, Araújo, et al., 2025; Franco, Hollender, et al., 2025; Naumann et al., 2002; Oziolor et al., 2019; Reid & Whitehead, 2016; Sehlmeier et al., 2010; Whitehead et al., 2017).

Following from these questions, contemporary research increasingly shifted away from analysing repertoires in isolation or in relation to specific pollutants. Instead, the study of the chemical defensome has progressively evolved towards systems biology and network-based approaches, integrating levels of evidences—not only genes, but also protein, metabolites, environmental chemicals and other molecular and ecological factors. Furthermore, several authors have recently argued that the pre-established organizational levels are insufficient to encompass the full complexity currently known about organisms, recommending an expansion of the concept of the chemical defensome to include novel organizational levels such behaviour plasticity. Following this line of thought, Franco et al. (2025) proposed broadening the classical view of CD by incorporating behaviour into a novel CD framework (Franco, Araújo, et al., 2025). In their study authors advocated for the inclusion of avoidance behaviour as fifth functional category of CD framework alongside the existing four—transcription factors, biotransformation enzymes, transporters and stress response pathways—since it acts as primary defence mechanism of organism and plays a crucial role in detection and mitigation of chemical pollutants (Franco, Araújo, et al., 2025). In addition, other authors have gone further and although not explicitly referring to CD, suggested the inclusion of

social and collective behaviour as part of contaminant assessment in organisms (Michelangeli et al., 2022).

6.2. The NGS as game changer in Chemical Defensome

In the early 2000s the emergence of NGS technologies transformed the research paradigm across multiple fields, including molecular ecotoxicology. Prior NGS era, research on chemical defensome faced several constraints, both conceptual and methodological. These limitations often led to fragmented views of biological processes, restricted gene discoveries and neglect of key aspects of the chemical defensome, such as redundant pathways or compensatory mechanisms.

Narrow approaches focused on individual CD elements or small gene families, biased gene discovery restricted to model organisms, and assumptions of linear cause-effect in experimental design, and the reliance on limited technologies such as RT-PCR or microarrays constrained the capacity to identify genome-wide responses and evolutionary patterns across taxa (Stegeman et al., 2010). Many of these limitations were overcome with the emergence of NGS and integration of omic approaches—such as genomics, transcriptomics, proteomics—into chemical defensome research.

Genomics approaches allowed the large-scale cataloguing of CD elements, revealing both conserved elements and lineage-specific innovations. Transcriptomics, in turn, enabled to highlight the functionality of these catalogues under chemical and biological challenges, demonstrating that chemical defensome act as an orchestrated network rather than an isolated and independent element. In parallel, with advances in NGS, the co-evolution of bioinformatics and the development of methods for data analysing, processing, management and storage proved equally essential.

The genomic resources rely primarily on the organization of DNA elements, providing a comprehensive catalogue of the components present in the genome of a given species. Over the past three decades, the sequencing and genome assemblies across multiple organisms—at contig, scaffold, chromosome and recently telomeres to telomere scale—have reshaped research in many fields, including the chemical defensome (Franco, Araújo, et al., 2025; Rhie et al., 2021).

In this thesis, I applied multiple genomic approaches (outlined in **Chapter 2** and **3**) to identify CD elements in non-model species for which genomic information is scarce. The building of high contiguous genome assemblies both at scaffold and chromosome level are essentially for capturing nearly complete repertoires of CD elements. This is particularly critical for xeno-sensor or biotransformation gene families, such as NRs and

CYPs owing to their repetitive sequence nature, large family sizes, and evolutionary origins from whole-genome or local duplications and high similarity.

Importantly, in **Chapter 2 and 3**, both genomes were produced using long-read approaches (PacBio HiFi), which overcome several limitations associated to short read approaches described across literature (Rhie et al. 2021). In **Chapter 2.3**, additional chromatin conformation capture (3C) methods (Hi-C) were applied to achieve chromosome-level resolution for the genome assembly of the chimaeriform, *Hydrolagus colliei*. Both enabled the identification of well-resolved CD clusters, conserved synteny, lineage specific expansions and gene losses, as well as potential structural rearrangements.

Additionally, both genome assemblies were structurally and functionally annotated using several bioinformatic pipelines, including BRAKER and Interproscan, alongside publicly available databases such as the NCBI Reference sequence (RefSeq). Gene model selection was guided by integrating expression datasets (RNA-Seq, generated as part of this work) and proteins from more than 400 organisms as supporting evidence. For the chemical defensome the use of multiple sources of evidence was crucial. This was particularly the case for CD genes families, given their evolutionary characteristics and the tendency of ab-initio predictors to misannotated gene models in such contexts (Brúna et al., 2021; Hoff et al., 2015, 2019). While RNA sequencing provided transcriptional support for gene models, protein homology ensured functional validation based on conserved motifs. Functional annotation was performed using a combination of approaches, including homology- and motif-based searches, with reference to multiple databases (Buchfink et al., 2014; Quevillon et al., 2005).

The transcriptomic approaches represent the next natural step in the evaluation of the chemical defensome, as they add a functional layer to the genes and transcriptional elements identified during the building of defensome repertoire from genomic resources. Importantly, this technology is the natural successor of classical techniques such quantitative real-time PCR (qRT-PCR), but with the added advantage of scale. Among the existing omics, transcriptomic is one of the most multifaceted. In the context of chemical defensome, it has been applied to: uncover the functionality of defensome elements under specific conditions, both in laboratory and in natural environments; to identify novel components; and to determine gene expression profiles across different species and/or tissues (as applied in **Chapters 4.1 and 4.2**). This is the most common application, enabling the identification of mobilised elements in response to chemical or biological agents, or the detection characteristic gene expression signatures in particular tissues.

However, the potential of this technology extends beyond these applications. It is frequently employed to build CD repertoires in the absence of genomic resource, through the *de novo* assembly of transcriptomes (as applied in **Chapter 5**); to perform relative gene expression analyses in a non-quantitative manner (as applied in **Chapter 3**); or, to support other omic approaches such genomics in the process of structural and functional annotation (as applied in **Chapter 2**). It can also be applied in combination of several approaches and applications (e.g. **Chapter 4.1 and 4.2**), where references built from transcriptomic dataset are subsequently used for gene expression analyses. Each application requires specific methods and offer varying levels of accuracy depending on the intending output. For instance, reference guided genomic approaches, using splice-aware aligners enable the precise mapping of transcripts into annotated genomes (applied in **Chapter 3**), thereby achieving greater accuracy in gene expression analyses (Dobin et al., 2013; Kim et al., 2019). By contrast, if the primary goal is to generate exploratory defensome repertoires, capture novel isoforms, or complementing genomic sequencing and annotation by filling gaps, *de novo* assemblies is a cost effective and efficient solution (Grabherr et al., 2011; Haas et al., 2013; Robertson et al., 2010). Although less accurate than reference-guided approaches (largely due the expense of generating high-quality genomes), the combined use of *de novo* assembly and subsequent differential gene expression analyses is particularly valuable in defensome research, given the diversity and lineage specific variants, that may play different roles across organisms.

Within the context of transcriptomic approaches, it was possible to perform functional enrichment analyses in **Chapter 4.1 and 4.2**, using several system-level tools, such as g:profiler or PHANTER. These enabled integration of differentially expressed genes (DEG) into biological pathways and provided functional interpretations (Raudvere et al., 2019; Thomas et al., 2003). Categorization into pathways, such as oxidative stress, immune signalling, xenobiotic or lipid metabolism was critical for gene discover, cataloguing, and the transformation of large-scale datasets into meaningful biological insights.

Overall, both genomic and transcriptomic strategies are complementary. When research questions concern the presence or absence of CD elements, both technologies valuable perspectives. However, when the aim is to construct catalogues and repertoires, transcriptomic resolution can be limited and frequently leads to ambiguous assignments, particularly within large gene families such as NRs. Moreover, the diversity of alternative splicing isoforms also can have strong impact in identification and reconstruction of full-lengths transcripts, resulting in complex and time-consuming

analyses as encountered in **Chapter 4.1** and **4.2** (Conesa et al., 2016). Lastly, although transcriptomic predictions have a functional role in gene expression, they generally do not account for other biological parameters at protein level (e.g. enzymatic activity). Consequently, validation through other omics approaches, such as metabolomics or proteomics, is often essential.

In addition, comparative genomic approaches were applied using the datasets in **Chapter 3**, and **Chapter 4.1**. Comparative genomics is a powerful mean of inferring defensome function, as it enables the identification of conserved genes as well as lineage specific adaptations (Eide et al., 2021; Goldstone, 2008; Goldstone et al., 2006). Conserved defensome elements typically represent critical and essential detoxification mechanisms, while expansions or contractions may reflect adaptation to a specific ecological niche and ecosystems (Reid & Whitehead, 2016). From a different perspective, comparative analyses also allowed functional inference when supported by evolutionary approaches such as phylogeny and/or synteny across several taxa, and by inclusion of characterized elements of model species (Hahn et al., 2008).

In summary, the NGS-based approaches have changed the study of chemical defensome, allowing the transition from fragmented perspectives to system-level understanding. The high-quality assemblies and in-depth transcriptomic analyses performed during this thesis highlighted the role of some chemical defensome elements on studied species. Furthermore, enrichment and comparative genomics were pivotal to situate these CD components not only within biological pathways but also in an evolutionary framework.

6.3. Evolution of chemical defensome in metazoans

Over the last 780-635 million years, metazoans have evolved to adapt, survive, reproduce, and prosper in a wide range of environments (Erwin, 2015; Sebé-Pedrós et al., 2017). The organism survival is highly dependent of the adaptive capacity and frequently can be achieved by phenotypic plasticity, adaptation, distributional shifts or rapid evolution of adequate traits to the new environment/stressor. These changes can be observed at behavioural, physiological, morphological and/or genetic level (Ross et al., 2023). Remarkably, metazoans display a large and diverse range of strategies, involving adaptive capacity processes, and some of the most fascinating ecological adaptations among eukaryotes (Erwin, 2015; Erwin et al., 2011). The understanding of these strategies, integrated with the surrounding environment are crucial to highlight specialized systems such as CD, in specific groups of species.

6.3.1. From Phenotype to chemical defensome adaptations in Metazoans

The integration of anthropogenic chemicals in wild habitats has exerted selective pressure on organisms, driving them to evolve at an unprecedented rate. By artificially increasing environmental complexity and introducing stressors, taxa were placed under a novel source of selection (Brady et al., 2017; McCulloch & Waters, 2023; Whitehead, 2014). In this context, the understanding of the composition, plasticity and response of the CD in Metazoa lineages is paramount to understand resilience through physiological, morphological or behavioural modifications (Forsman, 2014; Fox et al., 2019; Hamilton et al., 2017; Reid & Whitehead, 2016; Sumpter, 2005; Whitehead, 2014). All these modifications/adaptations are even more critical in aquatic environment where species are continuously exposed to these compounds (Fusco & Minelli, 2010; Hofmann & Todgham, 2009; Livingstone, 2001; Sumpter, 2005; Wendelaar Bonga, 1997).

In the presence of chemical aggressors, physiological strategies are deployed at different levels. Among these, the drug metabolism and antioxidant systems are primarily activated, followed or combined with energy reallocation, metabolic shifts, hormonal and endocrine adjustments, osmoregulatory and ion homeostasis regulation (Wendelaar Bonga 1997; Livingstone 2001; Hofmann and Todgham 2009; Fusco and Minelli 2010). In addition to physiological strategies, behavioural changes can function as short-term lines of action when organisms face chemical stressors. In agreement, several species have been documented as modulating their behaviour patterns, in order to minimise or avoid locations with chemical stressors. For example, the teleost *Danio rerio* has been reported to alter its behaviour in the presence of certain chemical, avoiding

concentrations above a certain threshold (Araújo et al., 2014; Silva et al., 2018). These behavioural changes includes and impacts feeding, mating, migration, social, parental, avoidance and defence patterns (Araújo et al., 2016; Franco, Araújo, et al., 2025; Saaristo et al., 2018).

Importantly all above-mentioned modifications are shaped by the organism's molecular repertoire. The molecular alterations, in gene repertoire or function, particularly through expansion, loss, sub- or neofunctionalization of CD elements are essential to chemical ecologic adaptation. These changes not only allowed the increase of CD elements, such transporters, detoxification enzymes and/or stress-response proteins, but also permit to optimise CD molecular mechanism either at energetic and resilience level, while increase subtract selectivity and specificity (Dean & Annilo, 2005; Goldstone, 2008; Goldstone et al., 2006, 2007; Nelson et al., 2013). Across metazoans its quite frequent observe one or multiple adaptations integrated at same time, mainly in protostomes and deuterostomes species models. For instance, the lophotrochozoan *Hediste diversicolor*, a estuarine worm, constantly exposed to sediments rich in natural toxins and metals, optimized their CD repertoire to process algal toxins, metals and plant-derived phenols (Breton & Prentiss, 2019). In deuterostomes this phenomenon occurs transversally. For instance, the echinoderm, *Paracentrotus lividus* use the internal CD diversity to handle environmental chemicals such as heavy metals, phytotoxins or microbial products (Marrone et al., 2012).

Whitin this dissertation we approached some of CD gene families in non-model species. In 2006, Goldstone defined the chemical defensome and postulated that the major part of CD gene families are conserved across Deuterostomia species. Moreover, the presence of many CD genes families in Protostomia, suggested an origin of the CD in the Bilateria ancestor. Here, we focused on the evolutionary pathway of two CD gene families, nuclear receptors and flavin-containing monooxygenases while following the functional framework postulated by Goldstone and colleagues in 2006 (Goldstone et al., 2006).

6.3.2. Transcription factors and biotransformation enzymes in metazoa species

Transcription factors such as NRs, emerged >600-700 MYA in early metazoans, and have been reported as exclusive of this kingdom (Owen & Zelent, 2000). From an evolutionary perspective, several studies suggested that modern NRs descended from a unique ligand-dependent NR common ancestor, with ability to activate or repress gene transcription by biding DNA regulatory regions to small lipophilic molecules (Bridgham et al., 2010). NRs evolution in early divergent metazoans was driven by local duplications

as well as species specific expansions and speciation events, while the vertebrates NR evolution was largely modulated by rounds of WGD events (Laudet et al., 1992; Owen & Zelen, 2000). Importantly, NRs diversification is not only supported by local and whole genome duplication events (that led to the origin of multiple subfamilies) but also by the domain shuffling's and subtler tinkering that allowed NRs speciation (Bridgham et al., 2010; Escrivá García et al., 2003; Escrivá et al., 2000, 2004). Currently, it is believed that NRs super family was massively expanded and diversified in two globally evolutionary moments, in the emergence of Bilaterians (via local gene duplications), and in the two rounds of WGD (2R WGD) occurred between in the origin of chordates and diversification of gnathostomes (Bertrand et al., 2004; Dehal & Boore, 2005; Escrivá et al., 2004; Ohno, 2013). Furthermore, some specific lineages incurred in additional rounds of WGDs increasing even more the NRs portfolio (3R WGD – teleost specific WGD; 4R WGD - salmonoid-specific WGD) (Jatllon et al., 2004; Lien et al., 2016; Taylor et al., 2003). Several studies have described the NRs portfolios across metazoans accompanying these events: the NR expansion from non-bilaterian to bilaterian was consolidated by identifying the NRs portfolio of cnidarians, *Nematostella vectensis* (17 NRs identified) and the Nematoda, *Caenorhabditis elegans* (>200 NRs identified); while the 2R, and 3R WGD events, are mirrored in the increasing number of NR from early-branching chordate lineages, such as *Ciona intestinalis* (17 NRs identified), to mammals such as *Homo sapiens* (2R WGD - 48 NRs identified), or teleosts such *Danio rerio* (3R WGD - 73 NRs) (Reitzel et al., 2008; Reitzel & Tarrant, 2009; Sluder et al., 1999; Yagi et al., 2003; Zhao et al., 2015)

Differently from NR transcription factors, flavin containing-monooxygenases emerged long before pre-bilaterian species. Indeed, these enzymes are classified as one of the most ancient groups of flavin/NADPH dependent enzymes in the Baeyer—Villiger monooxygenases (BV) class, and have been described in all kingdoms of life (Cashman, 2018; Krueger & Williams, 2005; Mascotti et al., 2015; Phillips & Shephard, 2008). The origin of FMOs as well other BV enzymes such as NMOs (N-hydroxylating monooxygenases) and BVMOs (Bayer-Villiger monooxygenases) has been reported to predate the consolidation of three domains of life, Bacteria, Archaea and Eukarya (Mascotti et al., 2015).

This CD gene family is critical for the biotransformation phase of drug metabolism and is mainly responsible for oxidation of nucleophilic heteroatoms (Cashman, 1995; Cashman & Zhang, 2006; Ziegler, 2002). Across metazoans, the FMOs repertoire of genes is relatively reduced, when comparing with other biotransformation gene families as cytochrome P450 superfamily. Additionally, the repertoire of FMOs across metazoans

seem constant and highly conserved, despite their functional divergence. Evolutionarily, this family has been shaped by small scale local gene duplications, and variable gene loss/retaining (Goldstone, 2008; Huang et al., 2021; Reitzel et al., 2008).

By examining the evolution roadmap of FMOS, we observe that early-branching metazoans such the cnidarian *Nematostella vectensis* typically possess between two and eight elements, commonly involved to redox balance through detoxification of reactive oxygen species (Goldstone, 2008; Huang et al., 2021; Reitzel et al., 2008). In contrast, protostomes display a variable number of FMOS. For instance, in Ecdysozoa species such as *Drosophila melanogaster*, two FMO-like have been identified, whereas in Lepidoptera species, three FMO-like members have been reported (Scharf et al., 2004; Sehlmeier et al., 2010). Several authors have described the FMO gene duplication as key innovations in lepidopterans, as these genes enable organisms not only to detoxify toxic pyrrolizidine alkaloids but also to sequester these plant toxins and use it in their own chemical defence (Naumann et al., 2002; Sehlmeier et al., 2010).

Similarly, others studies in Lophotrochozoans, such as molluscs, have reported equivalent number of FMOs, while demonstrating their role in resistance to paralytic shellfish toxins produced by dinoflagellates (Kong et al., 2021). Overall, both studies showed the functional diversity of FMOs in protostomes while showed lineage-specific FMO expansions in response to environmental pressure and ecologic adaptations.

The FMOs evolution in non-chordate Deuterostomes follows similar lineages specific expansion patterns while maintaining functional divergence. For instance, in 2006, Goldstone and co-workers found an expansion of 16 FMO genes in the echinoderm *Strongylocentrotus purpuratus* (Goldstone et al., 2006). Initially, these genes were thought to be predominantly related to detoxification mechanisms of dietary metabolites, algal toxins and anthropogenic pollutants. Later, several authors demonstrated that some of these FMOs genes are involved in echinoids pigmentation pathways, while others are expressed in innate immune cells (coelomocytes) (Perillo et al., 2020; Wessel et al., 2020; Wise et al., 2024).

Differently, in other early Deuterostomes, such as *Ciona intestinalis* the FMO repertoire appear reduced (Hao et al., 2009). Interestingly, FMOs in invertebrate deuterostomes seems to evolved through lineage-specific gene expansions, whereas in vertebrates the FMO gene duplications, retentions and loss are more variable. Nevertheless, the FMO repertoire and evolutionary pathway in tetrapods is relatively well consolidated into five clades (FMO1-5), likely originating from a single FMO sequence following two rounds of whole-genome duplication in early vertebrates. Evidences of

large-scale duplications and loss events is limited. However, the patterns of FMO evolution in early-gnathostomes and teleosts remains poorly understood, with few studies available (Bailleul et al., 2023; Cashman, 1995; Hao et al., 2009; Krueger & Williams, 2005; Phillips & Shephard, 2017, 2019). For instance, in Chondrichthyes only a single study on flavin containing monooxygenases is available (Schlenk & Li-Schlenk, 1994). In teleosts, despite the existence of several functional studies on species of interest, only recently has it been possible to obtain an overview of FMO repertoire in teleosts (between 2 and 5 in five analysed species) (Eide et al., 2021).

In this thesis, our approaches were in two complementary perspectives. The first was a broad approach in which we sequenced, annotated and mined the nuclear genome of two vertebrates—the teleost *Scomber colias* (**Chapter 2**) and the Chimaeriforme *Hydrolagus colliei* (**Chapter 3**)—in order to expand chemical defensome repertoire in non-classical model species. The second was species-specific approach, in which we focused on transcription factors and biotransformation from a functional perspective. Rather than quantifying and cataloguing, we aimed to understand the interplay between the regulation of lipid metabolism (**Chapter 4.1**), the immune system (**Chapter 4.2**), and the chemical defensome system, using the *Paracentrotus lividus* species as model. This species was chosen because it is phylogenetic related to *Strongylocentrotus purpuratus*, a heavily studied species with respect to the CD, with multiple levels of functional data. Finally, in the last **Chapter 5** we sequence, annotate and prepared transcriptomic datasets from another group of organisms, Annelida, with goal of later analyse and understand the repertoire of NRs and FMOs in the whole transcriptome of *Hediste diversicolor*.

In *Scomber colias*, the number of NRs observed was concordant to that previously established for teleosts (Eide et al., 2021; Jatllon et al., 2004). In the same study, we identified PXR, the master regulator of xenobiotic metabolism, as well as other NR1 receptors, including, RXRs, PPARs, LXRs and VDR, which are highly engaged in drug metabolism (see **Chapter 2.1** for details). Importantly, the patterns of presence/absence of PXR are not consistent across teleosts or other vertebrate groups (Eide et al., 2018; Fonseca et al., 2019). Eide et al. (2018), reported that, among 77 teleosts, the gene was absent in 34 species, suggesting other receptors, such as AhR receptor, may have acquired a compensatory regulatory role (Eide et al., 2018). Fonseca et al. (2019) extended this analysis across vertebrates, finding PXR absent in some Chondrichthyes, birds and reptiles (Fonseca et al., 2019). In Chondrichthyes the pattern was variable: PXR was absent in Elasmobranchii but present in the Chimaeriformes, *Hydrolagus affinis* and *Callorinchus milli*. These authors inspected also CAR and VDR, confirming CAR's

emergence in tetrapod and the universal presence of VDR in all species analysed (Fonseca et al., 2019).

These results were reconfirmed in **Chapter 2.1**, and in a chromosome-level dataset of *Hydrolagus colliei* produced in **Chapter 2.2**. In addition to the PXR, VDR and CAR presence/absence assessment in *H. colliei*, RNA-Seq confirmed the gene expression, particularly in gill tissue. Regarding FMOs in *S. colias*, five potential FMOs, were identified: three in tandem at the same locus, likely resulting from gene duplication and two others at separate genomic locations (**Chapter 2.1**). These findings are consistent with FMOs number and genomic location reported in other teleosts (Eide et al., 2021). While few studies on FMOs in teleosts exists, some research in salmonids has linked FMOs to osmoregulation and xenobiotic detoxification (El-Alfy et al., 2002; Larsen & Schlenk, 2001b, 2001a; Lavado et al., 2013). In chimaeriform *H. colliei*, two FMO gene were identified, as in other Chimaeriformes species such as elephant shark and *H. affinis* (**Chapter 2.2**) (Fonseca et al., 2020). Both FMO genes were found expressed predominantly in liver and, to a lesser extent, in gill, suggesting a primary role in systemic xenobiotic and endogenous amines metabolism, with residual expression in gill providing localized protection complementary to CYP enzymes against waterborne compounds, as observed in other aquatic organisms (Jönsson et al., 2006; Larsen & Schlenk, 2001a).

Overall, both genome-wide approaches provide the first insights into CD NRs and FMOs in both species, but also contributed to the understanding of evolutionary phenomena such as duplications, conservation and loss of key CD elements in early-vertebrate species. Additionally, analyses of presence, genomic organization and tissue expression allowed the associate both NRs in and FMOs to chemical processes. Despite their phylogenetic relevance, genomic and functional data regarding CD in these clades remain scarce, making these findings valuable. Finally, this genome-centred approach setting a baseline for future studies on detoxification and chemical defensome in less explored marine vertebrates.

In **Chapter 4.1** and **4.2** several analyses were performed on gonads and immune cells of *Paracentrotus lividus* to evaluate presence or absence of several sex, lipidic, immune related genes and to correlate these gene portfolios with chemical defensome composition. Additionally, gene expression analyses were performed between males and female gonads and immune cells of *P. lividus*. Beyond immune, lipid and sex genes, several CD elements were identified as differentially expressed in males and females.

In **Chapter 4.1** we highlight lipid metabolism and the chemical defensome in gonads of *P. lividus*. Xeno-regulation in *P. lividus* as well as in other invertebrate deuterostomes is rooted in evolutionary older NR elements such as VDR-like, FXRs, LXRs, PPARs and RXRs. This is because the most common xenosensors, PXR and CAR, emerged in post 2R-WGD and are considered vertebrate innovations (Bridgham et al., 2010; Krasowski et al., 2005; Mathäs et al., 2011). Among findings in **Chapter 2.3**, the PPAR as well RXR were found to be up regulated in female gonads, suggesting an active state of heterodimer PPAR-RXR. These results are in line with previous study of Capitão et al. in 2020, where authors functionally characterized PPAR in *P. lividus* gonadal tissue, while verified the ability of Tributilyn, a model lipid homeostasis modulator, to interact with lipid metabolism (Capitão et al., 2020).

Additionally, and although not differentially expressed, we found evidences of LXR, FXR and the ligand-independent nuclear receptor-like gene (HNF4) expression in gonads. In vertebrates, the HNF4 is described as an important FXR co-regulator, suggesting a conserved co-regulatory role for lipid metabolism and chemical defensome regulations in reproductive tissues of *P. lividus* (Eloranta & Kullak-Ublick, 2005). Importantly, we also found expression evidences of nuclear receptor coactivator-6 (NCOA6), which, although it does not directly interact with DMEs has been described as indirectly regulating them, including CYP2C9, by modulating the transcriptional activity of several NRs such as PXR, CAR, PPAR, HNF4 in vertebrates (Surapureddi et al., 2008, 2011). In relation to oxidative stress, we found evidences of expression of NR4 family members, namely NUR77 and NURR1, which are generally involved in oxidative stress and autophagy of multiple organism (Chen et al., 2020; Lee et al., 2014). In addition to NRs, both members of AhR-ARNT heterodimer—the xenobiotic sensor AhR and his co-activator ARNT—were detected, as well as CNC-bZIP family member NRF2, in gonadal tissue.

Turning into the flavin-containing monooxygenase enzymes, previous studies have reported the expansion of FMOs in echinoids, with the sea urchin *S. purpuratus* described as containing about 16 FMO sequences (Goldstone et al., 2006). While most part of these genes are associated with larval stages, some are expressed in specialized adult cells, such as pigmented cells or coelomocytes, with many lacking apparent functions, and often linked to the canonical role in chemical detoxification (Bhat et al., 2024; Perillo et al., 2020; Wessel et al., 2020; Wise et al., 2024). In **Chapter 4.1** we found several FMO genes expressed in gonads of *P. lividus*, with two specifically up-regulated in female gonads, suggesting a potential engagement in sex-specific and reproductive functions. Although studies on the role of FMOs in gonads are limited,

detoxification of endogenous compounds and maintenance of redox homeostasis are essential in all tissues, implying that FMO may participate in this processes *P. lividus* gonadal tissue (Shih et al., 2015). In vertebrates, several studies have linked FMOs to lipid metabolism and homeostasis, primarily via indirect regulation of PPAR xeno-sensor (Shih et al., 2015). Furthermore, as these elements are specifically differentially expressed in females, FMOs may be engaged in lipid catabolism and storage associated to female process such as gametogenesis, oocyte maturing and egg provisioning.

All these components not only contribute to the transcriptomic signatures of adult male and female gonads, but also highlight the role of nuclear receptors such PPAR or RXR as key elements in lipid homeostasis and reproductive physiology in *P. lividus*, while actively coordinating metabolic and chemical defensome detoxification processes in concert with other gonadal NRs. Moreover, the finding of several heterodimerizations and potential coactivator leverage the idea that lipid-signalling pathways have a role in fine-tuning expression of xenobiotic-responsive genes, while other xeno-sensors and pathways ensure and complement chemical defence guarantying the oxidative stress sensing responses and gonadal function. Notably, the expression of several FMOs in both males and females, as well the two females specific FMOs, allowed to correlate it with sex-specific processes such as gametogenesis, in addition to lipid metabolism. Overall, these findings on *P. lividus* gonads contribute to reinforce the reproductive, lipid and chemical regulation and resilience provided by NRs and FMOS in reproductive tissues, while highlight the sex-specific dimension of chemical defensome elements.

In **Chapter 4.2** we focused on understanding of interplay between immune system and CD members in *P. lividus* when submitted to a bacterial challenge. Generally, when organisms are biologically threatened, they deploy a wide range of immune mechanisms to counteract infections. While innate immune system is the first line of defence against biological aggressors, the chemical defensome acts as an internal chemical stabilizer mitigating the collateral damage. After exposure to *Vibrio anguillarum*, the immune cells of *P. lividus*, the so-called coelomocytes, were sequenced, assembled and differential gene expression was performed. Although the results primarily catalogue innate immune genes, it was possible to identify significant interactions between CD and immune system.

Regarding NRs, PPARs were found up regulated in organisms exposed to *Vibrio anguillarum*. Beyond their xenobiotic-sensor role in drug metabolism, these genes have been reported to modulate and regulate anti-inflammatory responses by controlling several genes engaged in immune system such as Toll-like receptors (Clark, 2002; Straus & Glass, 2007). Additionally, also Toll-like receptors themselves were up

regulated, indicating that *P. lividus* potentially orchestrates and integrated chemical-immune defence network when subject to biological threats. Supporting this chemical-immune system relationship, other CD central nuclear receptors—including LXR or RXR and NR4 members such as NUR77—were expressed in coelomocytes. These have been described as immune modulators or associated to oxidative stress generated by inflammation (Bono et al., 2016; Kidani & Bensinger, 2012; Lee et al., 2014). Furthermore, several nuclear receptors co-activators members such NCOA4 and NCOA6 were found expressed in the dataset; these genes are commonly engaged in immune modulation as well as in chemical defensome regulation (Chen et al., 2020; Clark, 2002; Lee et al., 2014; Milito et al., 2020; Straus & Glass, 2007; Surapureddi et al., 2008, 2011). Interestingly, before our study, *P. lividus* chemical-immune relationship had already been evaluated, however from chemical perspective. In 2020, Milito and colleagues, assessed the antioxidant and immune responses of *P. lividus* exposed to polluted environments, and observed similar global patterns (Milito et al., 2020). In addition to sensing mechanisms, FMO biotransformation enzymes also contributed to the immune defence. Previous studies investigating proteins involved in bacterial responses identified common immune system proteins but did not detect FMO enzymes (Inguglia et al., 2020). In a different echinoid model, *S. purpuratus*, Perillo et al. (2020) conducted a comprehensive study associating specific FMOs and FMO groups with developmental stages and adult tissues, including spines, tube feet and coelomocytes (Perillo et al., 2020). While previous studies linked FMOs to pigment and developmental, none had correlated global FMO portfolio with both developmental and adult stages.

In *P. lividus* we identified identical patterns and multiple FMOs candidates expressed in both control and challenged coelomocytes, including as specific FMO up regulated under bacterial condition. These new findings, suggest that FMO mechanisms are conserved across echinoids, demonstrate a direct correlation between immune system and CD elements. Moreover, the differential expression of FMO elements under bacterial challenge provide functional evidence that certain FMOs are not merely expressed in immune cells, but also actively contributed to organism defence against biological stressors. Thus, while Perillo and co-workers perform an exhaustive and foundational work in identification, mapping and cataloguing of FMOs in *S. purpuratus*, our study reveals a previously uncharacterized layer of FMO functional specialization in echinoid immunity, linking the tissue-specific expression to biological challenge and highlighting role of CD in immune system (Perillo et al. 2020).

Among these systems, NRs and their co-activators play an important role, linking xenometabolism, FMOs, and other chemical defensome elements in modulation of immune

signalling. The integration of these chemical and immune response elements allowed organism to limit cellular damage during the biological challenge. Additionally, while Toll-like receptors and other innate immune components seems to ensure a rapid recognition and response, CD components, including NRs, co-activators and FMOs, actively contributed to maintaining homeostasis, by potentially regulation anti-inflammatory processes, while mitigating oxidative stress. Overall, this unified chemical-immune network exemplify a multilayer strategy adopted by echinoids to survive bacterial aggressors, while reveal a previously uncharacterized functional specialization of FMOs in immune defence and highlighting the crosstalk between chemical defensome and immune system.

6.4. Future perspectives

The future of chemical defensome studies in metazoans will be increasingly shaped by our capacity to integrate the impact of chemicals across different organisational levels within a coherent framework: from the molecular and cellular level to tissue/organ levels, and from there to population and ecosystem scales as a whole (Franco et al. 2025). While current frameworks already recognise some of these organisational levels, their practical application remains limited due to high costs, gaps in foundational knowledge at multiple levels (e.g., limited CD repertoires in metazoa species) and technical challenges in linking transversal processes (Franco et al., 2025).

The advent of next-generation sequencing in the early 2000s was revolutionary, democratising genomics and enabling the cataloguing of multiple chemical defensome repertoires, mainly in model species (Goldstone et al. 2006; Goldstone 2008; Stegeman et al. 2010). However, despite 25 years of NGS development and application, the genomic and transcriptomic data remains scarce for many non-model organisms (Blaxter, 2022; Stephan et al., 2022). Indeed, despite significant advances in NGS technologies, the routine production of genomes at the chromosome and telomere scale has only recently become standardised (Li & Durbin, 2024; Miga et al., 2020; Rhie et al., 2021). Therefore, it is expected that the production of high-quality, and contiguous genome assemblies across a broad taxonomic spectrum will increasingly be applied to the identification of gene gains, losses, and expansions in CD pathways, as well as to resolving gene clusters, regulatory regions, and chromosome architecture, in the near future. Furthermore, such resources are expected to facilitate comparative genomic and evolutionary approaches, as demonstrated by Eide and colleagues in 2021 (Eide et al., 2021). This level of genomic scrutiny will enable the detection of evolutionary processes, such as convergent evolution and adaptative expansions in CD gene families, as well as

lineage-specific events, and innovations in response to ecological and anthropogenic pressures. This has already been demonstrated in several studies (Franco, Hollender, et al., 2025; Naumann et al., 2002; Perillo et al., 2020; Whitehead, 2014)

Complementary functional studies will also be essential to assess and functionally characterise genes and resulting proteins. This can be achieved by employing transcriptomic approaches. In addition to bulk RNA-seq, single-cell and spatial transcriptomics are expected to facilitate tissues- and cell-type-specific expression and co-expression, as demonstrated in Perrillo and colleagues (2020) (Perillo et al., 2020). These complementary approaches will lay the groundwork for establishing baseline expression maps of the CD repertoire, identifying expression networks across tissues, and validating subcellular localisation. This will highlighting the context-dependent activation of CD genes. Acting at this level, it is expected that complex cascades of pathways will be revealed across cells and tissues, and that they will be linked to organism physiology. Beyond genes, efforts should be carried out to explore how resulting protein sequence changes may impact structure and function over evolutionary timescales.

Moving up the biological scale, CD research should continue to include *in vitro* and *in vivo* analyses of chemical-cell and chemical-organism interactions based on cause-and-effect evaluations. Such an approaches would reinforce predictive toxicology frameworks and facilitate the identification of stress biomarkers at the cellular and organism, and would sustain prediction at population levels. Moreover, these analyses provide foundational evidence of organism responses under adverse conditions when exposed to chemical or biological derived aggressors, particularly when combined with NGS approaches.

From a systematic, integrative, and bioinformatic perspective, integrating multiple omics approaches remains challenging but essential. The integration of these datasets—including genomics, transcriptomics, metabolomics, lipidomics and other omics—will increasingly rely on the construction of computational models and pipelines, as well as the development of artificial intelligence and machine learning systems capable of handling massive, and multidimensional datasets. These systems and models will enable the prediction of organism responses in diverse environmental contexts, while also revealing the interactions between the chemical defensome, immune networks and other stress response systems.

7. Conclusion

The chemical defensome repertoire of metazoans is diverse and extensive. Over millions of years, it has enabled organisms to adapt and evolve in order to overcome environmental bottlenecks in Earth's natural habitats. However, since the onset of the Anthropocene, organisms have faced their ultimate challenge: human activity. Consequently, they have been forced to evolve and adapt at an unprecedented rate in order to survive and overcome the new obstacles arising from human actions, such as pollution and the unbridled release of man-made chemicals. In order to understand both the effects of these chemicals on organisms, and the molecular mechanisms underlying resistance and tolerance, researchers have developed frameworks for studying organisms, most frequently using model organisms, at multiple levels, including the molecular level—the “defensome”. The work presented in this thesis is rooted in, and reproduces a myriad of bioinformatic approaches that are commonly employed in the omic studies of molecular defensome mechanisms. These approaches were applied to four non-model species, including vertebrate and invertebrate deuterostomes. Additionally, we analysed two critical gene families of the chemical defensome: the transcription factors (nuclear receptors) and the biotransformation enzymes (flavin-containing monooxygenases) across non-model species. To achieve this, we employed both global and species-specific approaches, expanding the knowledge of CD in vertebrates and highlighting the interplay between defensome families and other biological systems, such as the immune and lipid systems. Examining the global repertoire in non-model vertebrates revealed the presence of critical xeno-sensors, such as PXR, and other NRs in the analysed teleosts, as well as the presence of FMOs in vertebrates of both species. In terms of invertebrates, we identified links between gonadal and immune cell tissues, as well as between these tissues and CD elements. We also demonstrated, for the first time in these species, how some of these elements respond to biological agents. Overall, this work established a two-level (genomic-transcriptomic) framework for non-model organisms, beginning with sequence and reference construction, progressing to gene identification, and culminating in functional analysis.

References

- Abdelraheem, E., Thair, B., Varela, R. F., Jockmann, E., Popadić, D., Hailes, H. C., Ward, J. M., Iribarren, A. M., Lewkowicz, E. S., Andexer, J. N., Hagedoorn, P. L., & Hanefeld, U. (2022). Methyltransferases: Functions and Applications. *ChemBioChem*, 23(18), e202200212. <https://doi.org/10.1002/CBIC.202200212>
- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., ... Craig Venter, J. (2000). The genome sequence of *Drosophila melanogaster*. *Science*, 287(5461), 2185–2195. <https://doi.org/D0I:10.1126/science.287.5461.2185>
- Adoutte, A., Balavoine, G., Lartillot, N., & De Rosa, R. (1999). Animal evolution - The end of the intermediate taxa? *Trends in Genetics*, 15(3), 104–108. [https://doi.org/10.1016/S0168-9525\(98\)01671-0](https://doi.org/10.1016/S0168-9525(98)01671-0)
- Adoutte, A., Balavoine, G., Lartillot, N., Lespinet, O., Prud'homme, B., & De Rosa, R. (2000). The new animal phylogeny: Reliability and implications. *Proceedings of the National Academy of Sciences*, 97(9), 4453–4456. <https://doi.org/10.1073/PNAS.97.9.4453>
- Adrian-Kalchhauser, I., Blomberg, A., Larsson, T., Musilova, Z., Peart, C. R., Pippel, M., Solbakken, M. H., Suurväli, J., Walser, J. C., Wilson, J. Y., Alm Rosenblad, M., Burguera, D., Gutnik, S., Michiels, N., Töpel, M., Pankov, K., Schloissnig, S., & Winkler, S. (2020). The round goby genome provides insights into mechanisms that may facilitate biological invasions. *BMC Biology*, 18(1), 1–33. <https://doi.org/10.1186/s12915-019-0731-8>
- Alam, A., & Locher, K. P. (2023). Structure and Mechanism of Human ABC Transporters. *Annual Review of Biophysics*, 52(Volume 52, 2023), 275–300. <https://doi.org/10.1146/annurev-biophys-111622-091232>
- Allen, L. G., & Horn, M. H. (2006). *The ecology of marine fishes: California and adjacent waters*. Univ of California Press. <https://doi.org/https://doi.org/10.1525/california/9780520246539.001.0001>
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389–3402. <https://doi.org/10.1093/NAR/25.17.3389>

- Anderson, S., Sadinski, W., Shugart, L., Brussard, P., Depledge, M., Ford, T., Hose, J., Stegeman, J., Suk, W., Wirgin, I., & Wogan, G. (1994). Genetic and Molecular Ecotoxicology: A Research Framework. *Environmental Health Perspectives*, 102(Suppl 12), 3. <https://doi.org/10.1289/EHP.94102S123>
- Anway, M. D., & Skinner, M. K. (2006). Epigenetic Transgenerational Actions of Endocrine Disruptors. *Endocrinology*, 147(6), s43–s49. <https://doi.org/10.1210/EN.2005-1058>
- Aranguren-Abadía, L., Lille-Langøy, R., Madsen, A. K., Karchner, S. I., Franks, D. G., Yadetie, F., Hahn, M. E., Goksøyr, A., & Karlsen, O. A. (2020). Molecular and Functional Properties of the Atlantic Cod (*Gadus morhua*) Aryl Hydrocarbon Receptors Ahr1a and Ahr2a. *Environmental Science and Technology*, 54(2), 1033–1044. <https://doi.org/10.1021/acs.est.9b05312>
- Araújo, C. V. M., Moreira-Santos, M., & Ribeiro, R. (2016). Active and passive spatial avoidance by aquatic organisms from environmental stressors: A complementary perspective and a critical review. *Environment International*, 92–93, 405–415. <https://doi.org/10.1016/J.ENVINT.2016.04.031>
- Araújo, C. V. M., Shinn, C., Mendes, L. B., Delello-Schneider, D., Sanchez, A. L., & Espíndola, E. L. G. (2014). Avoidance response of *Danio rerio* to a fungicide in a linear contamination gradient. *Science of The Total Environment*, 484(1), 36–42. <https://doi.org/10.1016/J.SCITOTENV.2014.03.037>
- Ashlock, P. D. (1971). Monophyly and Associated Terms. *Systematic Biology*, 20(1), 63–69. <https://doi.org/10.1093/SYSBIO/20.1.63>
- Ashurst, J. L., Chen, C. K., Gilbert, J. G. R., Jekosch, K., Keenan, S., Meidl, P., Searle, S. M., Stalker, J., Storey, R., Trevanion, S., Wilming, L., & Hubbard, T. (2005). The Vertebrate Genome Annotation (Vega) database. *Nucleic Acids Research*, 33(suppl_1), D459–D465. <https://doi.org/10.1093/NAR/GKI135>
- Aurade, R. M., Jayalakshmi, S. K., & Sreeramulu, K. (2010). P-glycoprotein ATPase from the resistant pest, *Helicoverpa armigera*: purification, characterization and effect of various insecticides on its transport function. *Biochimica et Biophysica Acta*, 1798(6), 1135–1143. <https://doi.org/10.1016/J.BBAMEM.2010.02.019>
- Bailleul, G., Nicoll, C. R., Mascotti, M. L., Mattevi, A., & Fraaije, M. W. (2021). Ancestral reconstruction of mammalian FMO1 enables structural determination, revealing unique features that explain its catalytic properties. *Journal of Biological Chemistry*, 296, 100221. <https://doi.org/10.1074/jbc.RA120.016297>

- Bailleul, G., Yang, G., Nicoll, C. R., Mattevi, A., Fraaije, M. W., & Mascotti, M. L. (2023). Evolution of enzyme functionality in the flavin-containing monooxygenases. *Nature Communications*, 14(1), 1042. <https://doi.org/10.1038/S41467-023-36756-X>
- Bard, S. M. (2000). Multixenobiotic resistance as a cellular defense mechanism in aquatic organisms. *Aquatic Toxicology*, 48(4), 357–389. [https://doi.org/10.1016/S0166-445X\(00\)00088-6](https://doi.org/10.1016/S0166-445X(00)00088-6)
- Barghi, N., Hermisson, J., & Schlötterer, C. (2020). Polygenic adaptation: a unifying framework to understand positive selection. *Nature Reviews Genetics* 2020 21:12, 21(12), 769–781. <https://doi.org/10.1038/s41576-020-0250-z>
- Bariami, V., Jones, C. M., Poupardin, R., Vontas, J., & Ranson, H. (2012). Gene amplification, ABC transporters and cytochrome P450s: unraveling the molecular basis of pyrethroid resistance in the dengue vector, *Aedes aegypti*. *PLoS Neglected Tropical Diseases*, 6(6). <https://doi.org/10.1371/JOURNAL.PNTD.0001692>
- Barnett, L. A. K., Earley, R. L., Ebert, D. A., & Cailliet, G. M. (2009). Maturity, fecundity, and reproductive cycle of the spotted ratfish, *Hydrolagus coliei*. *Marine Biology*, 156(3), 301–316. <https://doi.org/10.1007/s00227-008-1084-y>
- Barnett, L. A. K., Ebert, D. A., & Cailliet, G. M. (2012). Evidence of stability in a chondrichthyan population: case study of the spotted ratfish *Hydrolagus coliei* (Chondrichthyes: Chimaeridae). *Journal of Fish Biology*, 80(5), 1765–1788. <https://doi.org/10.1111/J.1095-8649.2011.03216.X>
- Barros, S., Montes, R., Quintana, J. B., Rodil, R., Oliveira, J. M. A., Santos, M. M., & Neuparth, T. (2017). Chronic effects of triclocarban in the amphipod *Gammarus locusta*: Behavioural and biochemical impairment. *Ecotoxicology and Environmental Safety*, 135, 276–283. <https://doi.org/10.1016/J.ECOENV.2016.10.013>
- Berenbaum, M. R. (1995). The chemistry of defense: theory and practice. *Proceedings of the National Academy of Sciences*, 92(1), 2–8. <https://doi.org/10.1073/PNAS.92.1.2>
- Bergman, Å., Heindel, J. J., Jobling, S., Kidd, K. A., & Zoeller, R. T. (2013). Endocrine disrupting chemicals-2012. *Geneva: United Nations Environment Programme and the World Health Organization*.
- Bertrand, S., Brunet, F. G., Escriva, H., Parmentier, G., Laudet, V., & Robinson-Rechavi, M. (2004). Evolutionary genomics of nuclear receptors: From twenty-five ancestral genes to derived endocrine systems. *Molecular Biology and Evolution*, 21(10), 1923–1937. <https://doi.org/10.1093/molbev/msh200>

- Bhat, A., Carranza, F. R., Tuckowski, A. M., & Leiser, S. F. (2024). Flavin-containing monooxygenase (FMO): Beyond xenobiotics. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 46(7), e2400029. <https://doi.org/10.1002/BIES.202400029>
- Blais, J. M., Macdonald, R. W., Mackay, D., Webster, E., Harvey, C., & Smol, J. P. (2007). Biologically mediated transport of contaminants to aquatic systems. *Environmental Science and Technology*, 41(4), 1075–1084. <https://doi.org/https://doi.org/10.1021/es061314a>
- Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., & Shao, Y. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, 277(5331), 1453–1462. <https://doi.org/10.1126/science.277.5331.1453>
- Blaxter, M. (2009). Nematodes (Nematoda). *The Timetree of Life*, 247–250. <https://doi.org/10.1093/OSO/9780199535033.003.0028>
- Blaxter, M. (2022). Sequence locally, think globally: The Darwin Tree of Life Project. *Proceedings of the National Academy of Sciences of the United States of America*, 119(4), e2115642118. <https://doi.org/https://doi.org/10.1073/pnas.2115642118>
- Blazer, V. S., Iwanowicz, L. R., Henderson, H., Mazik, P. M., Jenkins, J. A., Alvarez, D. A., & Young, J. A. (2011). Reproductive endocrine disruption in smallmouth bass (*Micropterus dolomieu*) in the Potomac River basin: spatial and temporal comparisons of biological effects. *Environmental Monitoring and Assessment*, 184(7), 4309. <https://doi.org/10.1007/S10661-011-2266-5>
- Blum, M. (2012). *Chemical defenses of arthropods*. Elsevier.
- Bonnafé, E., Sroda, S., Budzinski, H., Valière, A., Pedelluc, J., Marty, P., & Geret, F. (2015). Responses of cytochrome P450, GST, and MXR in the mollusk *Corbicula fluminea* to the exposure to hospital wastewater effluents. *Environmental Science and Pollution Research*, 22(14), 11033–11046. <https://doi.org/doi:10.1007/s11356-015-4309-x>
- Bono, M. R., Tejon, G., Flores-Santibañez, F., Fernandez, D., Roseblatt, M., & Sauma, D. (2016). Retinoic Acid as a Modulator of T Cell Immunity. *Nutrients 2016, Vol. 8, Page 349*, 8(6), 349. <https://doi.org/10.3390/NU8060349>
- Bookout, A. L., Jeong, Y., Downes, M., Yu, R. T., Evans, R. M., & Mangelsdorf, D. J. (2006). Anatomical Profiling of Nuclear Receptor Expression Reveals a Hierarchical

Transcriptional Network. *Cell*, 126(4), 789–799.
<https://doi.org/10.1016/J.CELL.2006.06.049>

Borchiellini, C., Manuel, M., Alivon, E., Boury-Esnault, N., Vacelet, J., & Le Parco, Y. (2001). Sponge paraphyly and the origin of Metazoa. *Journal of Evolutionary Biology*, 14(1), 171–179. <https://doi.org/10.1046/J.1420-9101.2001.00244.X>

Borst, P., Evers, R., Kool, M., & Wijnholds, J. (2000). A Family of Drug Transporters: the Multidrug Resistance-Associated Proteins. *JNCI: Journal of the National Cancer Institute*, 92(16), 1295–1302. <https://doi.org/10.1093/JNCI/92.16.1295>

Bracquemond, D., Mahdizadeh, S. J., Brunet, L., Thomas, Q.-D., Bouclier, C., Colomb, M., Bourdel, S., Fabbriozio, E., Tourriere, H., Papon, L., Quantin, X., Santos, A., Ferrer, I., Henry, S., Grosgeorges, M., Perna, V., Lozano, M. D., Solórzano, J. L., Bossowski, J. P., ... Maraver, A. (2025). FMO4 drives lung adenocarcinoma by stabilizing the MAT2A/MAT2B complex and hindering ferroptosis. *BioRxiv*, 2025.03.31.646284. <https://doi.org/10.1101/2025.03.31.646284>

Brady, S. P., Monosson, E., Matson, C. W., & Bickham, J. W. (2017). Evolutionary toxicology: Toward a unified understanding of life's response to toxic chemicals. *Evolutionary Applications*, 10(8), 745. <https://doi.org/10.1111/EVA.12519>

Breton, T. S., & Prentiss, N. K. (2019). Metal stress-related gene expression patterns in two marine invertebrates, *Hediste diversicolor* (Annelida, Polychaeta) and *Littorina littorea* (Mollusca, Gastropoda), at a former mining site. *Comparative Biochemistry and Physiology. Toxicology & Pharmacology: CBP*, 225, 108588. <https://doi.org/10.1016/J.CBPC.2019.108588>

Bridgham, J. T., Eick, G. N., Larroux, C., Deshpande, K., Harms, M. J., Gauthier, M. E. A., Ortlund, E. A., Degnan, B. M., & Thornton, J. W. (2010). Protein Evolution by Molecular Tinkering: Diversification of the Nuclear Receptor Superfamily from a Ligand-Dependent Ancestor. *PLOS Biology*, 8(10), e1000497. <https://doi.org/10.1371/JOURNAL.PBIO.1000497>

Brown, A. R., Owen, S. F., Peters, J., Zhang, Y., Soffker, M., Paull, G. C., Hosken, D. J., Abdul Wahab, M., & Tyler, C. R. (2015). Climate change and pollution speed declines in zebrafish populations. *Proceedings of the National Academy of Sciences of the United States of America*, 112(11), E1237–E1246. <https://doi.org/10.1073/pnas.1416269112>

Bruijns, B., Tiggelaar, R., & Gardeniers, H. (2018). Massively parallel sequencing techniques for forensics: A review. *ELECTROPHORESIS*, 39(21), 2642–2654. <https://doi.org/10.1002/ELPS.201800082>

- Brûna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, 3(1), lqaa108. <https://doi.org/10.1093/nargab/lqaa108>
- Brûna, T., Lomsadze, A., & Borodovsky, M. (2020). GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics and Bioinformatics*, 2(2). <https://doi.org/10.1093/NARGAB/LQAA026>
- Brunet, T., & King, N. (2017). The origin of animal multicellularity and cell differentiation. *Developmental Cell*, 43(2), 124. <https://doi.org/10.1016/J.DEVCEL.2017.09.016>
- Brusca, R. C., Giribet, G., & Moore, W. (2022). Phylum Cnidaria. *Invertebrates*. <https://doi.org/10.1093/HESC/9780197554418.003.0007>
- Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 2014 12:1, 12(1), 59–60. <https://doi.org/10.1038/nmeth.3176>
- Budd, G. E. (2008). The earliest fossil record of the animals and its significance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1496), 1425. <https://doi.org/10.1098/RSTB.2007.2232>
- Budd, G. E., & Telford, M. J. (2009). The origin and evolution of arthropods. *Nature* 2009 457:7231, 457(7231), 812–817. <https://doi.org/10.1038/nature07890>
- Burke, R. D., Moller, D. J., Krupke, O. A., & Taylor, V. J. (2014). Sea urchin neural development and the metazoan paradigm of neurogenesis. *Genesis (New York, N.Y. : 2000)*, 52(3), 208–221. <https://doi.org/10.1002/DVG.22750>
- Burris, T. P., Solt, L. A., Wang, Y., Crumbley, C., Banerjee, S., Griffett, K., Lundasen, T., Hughes, T., & Kojetin, D. J. (2013). Nuclear Receptors and Their Selective Pharmacologic Modulators. *Pharmacological Reviews*, 65(2), 710. <https://doi.org/10.1124/PR.112.006833>
- Butcher, N. J., Boukouvala, S., Sim, E., & Minchin, R. F. (2002). Pharmacogenetics of the arylamine N-acetyltransferases. *The Pharmacogenomics Journal*, 2(1), 30–42. <https://doi.org/10.1038/SJ.TPJ.6500053>
- Cameron, C. B., Garey, J. R., & Swalla, B. J. (2000). Evolution of the chordate body plan: new insights from phylogenetic analyses of deuterostome phyla. *Proceedings of the National Academy of Sciences of the United States of America*, 97(9), 4469–4474. <https://doi.org/10.1073/PNAS.97.9.4469>

- Campbell, M. S., Holt, C., Moore, B., & Yandell, M. (2014). Genome Annotation and Curation Using MAKER and MAKER-P. *Current Protocols in Bioinformatics*, 2014, 4.11.1-4.11.39. <https://doi.org/10.1002/0471250953.bi0411s48>
- Cannon, J. T., Vellutini, B. C., Smith, J., Ronquist, F., Jondelius, U., & Hejnol, A. (2016). Xenacoelomorpha is the sister group to Nephrozoa. *Nature*, 530(7588), 89–93. <https://doi.org/10.1038/nature16520>
- Capitão, A., Lopes-Marques, M., Páscoa, I., Ruivo, R., Mendiratta, N., Fonseca, E., Castro, L. F. C., & Santos, M. M. (2020). The Echinodermata PPAR: Functional characterization and exploitation by the model lipid homeostasis regulator tributyltin. *Environmental Pollution*, 263, 114467. <https://doi.org/10.1016/J.ENVPOL.2020.114467>
- Carlberg, C., Bendik, I., Wyss, A., Meier, E., Sturzenbecker, L. J., Grippo, J. F., & Hunziker, W. (1993). Two nuclear signalling pathways for vitamin D. *Nature* 1993 361:6413, 361(6413), 657–660. <https://doi.org/10.1038/361657a0>
- Carthew, R. W., & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell*, 136(4), 642–655. <https://doi.org/10.1016/J.CELL.2009.01.035>
- Casals-Casas, C., & Desvergne, B. (2011). Endocrine disruptors: From endocrine to metabolic disruption. *Annual Review of Physiology*, 73, 135–162. <https://doi.org/10.1146/annurev-physiol-012110-142200>
- Cashman, J. (1995). Structural and Catalytic Properties of the Mammalian Flavin-Containing Monooxygenase. *Chemical Research in Toxicology*, 8(2), 165–181. <https://doi.org/10.1021/tx00044a001>
- Cashman, J. (2002). Human flavin-containing monooxygenase (form 3): Polymorphisms and variations in chemical metabolism. *Pharmacogenomics*, 3(3), 325–339. <https://doi.org/10.1517/14622416.3.3.325>
- Cashman, J. (2005). Human Flavin-Containing Monooxygenase Substrate Specificity and Role in Drug Metabolism. *Current Drug Metabolism*, 1(2), 181–191. <https://doi.org/10.2174/1389200003339135>
- Cashman, J. (2018). Monoamine Oxidases and Flavin-Containing Monooxygenases. *Comprehensive Toxicology: Third Edition*, 10–15, 87–125. <https://doi.org/10.1016/B978-0-12-801238-3.64091-6>
- Cashman, J., & Zhang, J. (2006). Human flavin-containing monooxygenases. *Annual Review of Pharmacology and Toxicology*, 46(Volume 46, 2006), 65–100. <https://doi.org/10.1146/annurev.pharmtox.46.120604.141043>

- Catalano, B., Moltedo, G., Martuccio, G., Gastaldi, L., Virno-Lamberti, C., Lauria, A., & Ausili, A. (2012). Can *Hediste diversicolor* (Nereidae, Polychaete) be considered a good candidate in evaluating PAH contamination? A multimarker approach. *Chemosphere*, 86(9), 875–882. <https://doi.org/10.1016/J.CHEMOSPHERE.2011.10.040>
- Cavalier-Smith, T. (1981). Eukaryote kingdoms: Seven or nine? *Biosystems*, 14(3–4), 461–481. [https://doi.org/10.1016/0303-2647\(81\)90050-2](https://doi.org/10.1016/0303-2647(81)90050-2)
- Cavalier-Smith, T. (1998). A revised six-kingdom system of life. *Biological Reviews*, 73(3), 203–266. <https://doi.org/10.1111/J.1469-185X.1998.TB00030.X>
- Chawta, A., Repa, J. J., Evans, R. M., & Mangelsdorf, D. J. (2001). Nuclear receptors and lipid physiology: Opening the x-files. *Science*, 294(5548), 1866–1870. <https://doi.org/10.1126/science.294.5548.1866>
- Chen, L., Bao, Y., Piekos, S. C., Zhu, K., Zhang, L., & Zhong, X. B. (2018). A Transcriptional Regulatory Network Containing Nuclear Receptors and Long Noncoding RNAs Controls Basal and Drug-Induced Expression of Cytochrome P450s in HepaRG Cells. *Molecular Pharmacology*, 94(1), 749–759. <https://doi.org/10.1124/MOL.118.112235>
- Chen, L., Fan, F., Wu, L., & Zhao, Y. (2020). The nuclear receptor 4A family members: mediators in human disease and autophagy. *Cellular and Molecular Biology Letters*, 25(1), 1–15. <https://doi.org/10.1186/s11658-020-00241-w>
- Chen, M., Xu, P., Zeng, G., Yang, C., Huang, D., & Zhang, J. (2015). Bioremediation of soils contaminated with polycyclic aromatic hydrocarbons, petroleum, pesticides, chlorophenols and heavy metals by composting: Applications, microbes and future research needs. *Biotechnology Advances*, 33(6), 745–755. <https://doi.org/10.1016/J.BIOTECHADV.2015.05.003>
- Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* 2021 18:2, 18(2), 170–175. <https://doi.org/10.1038/s41592-020-01056-5>
- Cho, B. K., Zengler, K., Qiu, Y., Park, Y. S., Knight, E. M., Barrett, C. L., Gao, Y., & Palsson, B. O. (2009). The transcription unit architecture of the *Escherichia coli* genome. *Nature Biotechnology* 2009 27:11, 27(11), 1043–1049. <https://doi.org/10.1038/nbt.1582>
- Choi, B. S., Kim, D. H., Kim, M. S., Park, J. C., Lee, Y. H., Kim, H. J., Jeong, C. B., Hagiwara, A., Souissi, S., & Lee, J. S. (2021). The genome of the European estuarine calanoid copepod *Eurytemora affinis*: Potential use in molecular ecotoxicology. *Marine Pollution Bulletin*, 166, 112190. <https://doi.org/10.1016/J.MARPOLBUL.2021.112190>

- Clark, R. B. (2002). The role of PPARs in inflammation and immunity. *Journal of Leukocyte Biology*, 71(3), 388–400. <https://doi.org/10.1189/JLB.71.3.388>
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17(1), 13. <https://doi.org/10.1186/s13059-016-0881-8>
- Consortium, T. C. elegans S. (1998). Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science*, 282(5396), 2012–2018. <https://doi.org/10.1126/science.282.5396.20>
- Correia, A. T., Moura, A., Triay-Portella, R., Santos, P. T., Pinto, E., Almeida, A. A., Sial, A. N., & Muniz, A. A. (2021). Population structure of the chub mackerel (*Scomber colias*) in the NE Atlantic inferred from otolith elemental and isotopic signatures. *Fisheries Research*, 234, 105785. <https://doi.org/10.1016/J.FISHRES.2020.105785>
- Cortés-Miranda, J., Veliz, D., Rojas-Hernández, N., Rico, C., Gutiérrez, C., & Vega-Retter, C. (2024). Chemical-defensome and whole-transcriptome expression of the silverside fish *Basilichthys microlepidotus* in response to chronic pollution in the Maipo River basin, Central Chile. *Aquatic Toxicology*, 277, 107159. <https://doi.org/10.1016/J.AQUATOX.2024.107159>
- Crespi, E., Burnap, R., Chen, J., Das, M., Gassman, N., Rosa, E., Simmons, R., Wada, H., Wang, Z. Q., Xiao, J., Yang, B., Yin, J., & Goldstone, J. V. (2022). Resolving the Rules of Robustness and Resilience in Biology Across Scales. *Integrative and Comparative Biology*, 61(6), 2163–2179. <https://doi.org/10.1093/ICB/ICAB183>
- Cruzeiro, C., Lopes-Marques, M., Ruivo, R., Rodrigues-Oliveira, N., Santos, M. M., Rocha, M. J., Rocha, E., & Castro, L. F. C. (2016). A mollusk VDR/PXR/CAR-like (NR1J) nuclear receptor provides insight into ancient detoxification mechanisms. *Aquatic Toxicology*, 174, 61–69. <https://doi.org/10.1016/J.AQUATOX.2016.02.007>
- Curpan, A. S., Impellitteri, F., Plavan, G., Ciobica, A., & Faggio, C. (2022). Review: *Mytilus galloprovincialis*: An essential, low-cost model organism for the impact of xenobiotics on oxidative stress and public health. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, 256, 109302. <https://doi.org/10.1016/J.CBPC.2022.109302>
- Dahl, T. W., Hammarlund, E. U., Anbar, A. D., Bond, D. P. G., Gill, B. C., Gordon, G. W., Knoll, A. H., Nielsen, A. T., Schovsbo, N. H., & Canfield, D. E. (2010). Devonian rise in atmospheric oxygen correlated to the radiations of terrestrial plants and large predatory

fish. *Proceedings of the National Academy of Sciences of the United States of America*, 107(42), 17911–17915. <https://doi.org/10.1073/pnas.1011287107>

De Marco, L., Sasser, D., Epis, S., Mastrantonio, V., Ferrari, M., Ricci, I., Comandatore, F., Bandi, C., Porretta, D., & Urbanelli, S. (2017). The choreography of the chemical defensome response to insecticide stress: insights into the *Anopheles stephensi* transcriptome using RNA-Seq. *Scientific Reports* 2017 7:1, 7(1), 1–12. <https://doi.org/10.1038/srep41312>

Dean, M., & Annilo, T. (2005). Downloaded from www.annualreviews EVOLUTION OF THE ATP-BINDING CASSETTE (ABC) TRANSPORTER SUPERFAMILY IN VERTEBRATES *. *Annu. Rev. Genomics Hum. Genet*, 6, 123–165. <https://doi.org/10.1146/annurev.genom.6.080604.162122>

Dean, M., Rzhetsky, A., & Allikmets, R. (2001). The human ATP-binding cassette (ABC) transporter superfamily. *Genome Research*, 11(7), 1156–1166. <https://doi.org/10.1101/GR.184901>

Dehal, P., & Boore, J. L. (2005). Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. *PLOS Biology*, 3(10), e314. <https://doi.org/10.1371/JOURNAL.PBIO.0030314>

Dejong, C. A., & Wilson, J. Y. (2014). The Cytochrome P450 Superfamily Complement (CYPome) in the Annelid *Capitella teleta*. *PLOS ONE*, 9(11), e107728. <https://doi.org/10.1371/JOURNAL.PONE.0107728>

Denison, M. S., & Nagy, S. R. (2003). Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annual Review of Pharmacology and Toxicology*, 43(1), 309–334.

DeSalle, R., Schierwater, B., & Hadrys, H. (2017). MtDNA: The small workhorse of evolutionary studies. *Frontiers in Bioscience (Landmark Edition)*, 22(5), 873–887. <https://doi.org/10.2741/4522>

Diamanti-Kandarakis, E., Bourguignon, J. P., Giudice, L. C., Hauser, R., Prins, G. S., Soto, A. M., Zoeller, R. T., & Gore, A. C. (2009). Endocrine-Disrupting Chemicals: An Endocrine Society Scientific Statement. *Endocrine Reviews*, 30(4), 293. <https://doi.org/10.1210/ER.2009-0002>

Didier, D. A., Kemper, J. M., & Ebert, D. A. (2012). Phylogeny, Biology, and Classification of Extant Holocephalans. *Biology of Sharks and Their Relatives: Second Edition*, 97–122. <https://doi.org/10.1201/b11867>

- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Dolphin, C. T., Beckett, D. J., Janmohamed, A., Cullingford, T. E., Smith, R. L., Shephard, E. A., & Phillips, I. R. (1998). The Flavin-containing Monooxygenase 2 Gene (FMO2) of Humans, but Not of Other Primates, Encodes a Truncated, Nonfunctional Protein. *Journal of Biological Chemistry*, 273(46), 30599–30607. <https://doi.org/10.1074/JBC.273.46.30599>
- Dolphin, C. T., Janmohamed, A., Smith, R. L., Shephard, E. A., & Phillips, I. R. (1997). Missense mutation in flavin-containing mono-oxygenase 3 gene, FMO3, underlies fish-odour syndrome. *Nature Genetics* 1997 17:4, 17(4), 491–494. <https://doi.org/10.1038/ng1297-491>
- Döring, B., & Petzinger, E. (2014). Phase 0 and phase III transport in various organs: combined concept of phases in xenobiotic transport and metabolism. *Drug Metabolism Reviews*, 46(3), 261–282. <https://doi.org/10.3109/03602532.2014.882353>
- Dos Reis, M., Thawornwattana, Y., Angelis, K., Telford, M. J., Donoghue, P. C. J., & Yang, Z. (2015). Uncertainty in the Timing of Origin of Animals and the Limits of Precision in Molecular Timescales. *Current Biology*, 25(22), 2939–2950. <https://doi.org/10.1016/J.CUB.2015.09.066>
- Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., & Aiden, E. L. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science*, 356(6333), 92–95. <https://doi.org/10.1126/science.aal3327>
- Dunin-Horkawicz, S., Czerwoniec, A., Gajda, M. J., Feder, M., Grosjean, H., & Bujnicki, J. M. (2006). MODOMICS: a database of RNA modification pathways. *Nucleic Acids Research*, 34(suppl_1), D145–D149. <https://doi.org/10.1093/NAR/GKJ084>
- Edi, C. V., Djogbénu, L., Jenkins, A. M., Regna, K., Muskavitch, M. A. T., Poupardin, R., Jones, C. M., Essandoh, J., Kétoh, G. K., Paine, M. J. I., Koudou, B. G., Donnelly, M. J., Ranson, H., & Weetman, D. (2014). CYP6 P450 Enzymes and ACE-1 Duplication Produce Extreme and Multiple Insecticide Resistance in the Malaria Mosquito Anopheles gambiae. *PLOS Genetics*, 10(3), e1004236. <https://doi.org/10.1371/JOURNAL.PGEN.1004236>
- Eide, M., Rydbeck, H., Tørresen, O. K., Lille-Langøy, R., Puntervoll, P., Goldstone, J. V., Jakobsen, K. S., Stegeman, J., Goksøyr, A., & Karlsen, O. A. (2018). Independent losses

of a xenobiotic receptor across teleost evolution. *Scientific Reports* 2018 8:1, 8(1), 1–13. <https://doi.org/10.1038/s41598-018-28498-4>

Eide, M., Zhang, X., Karlsen, O. A., Goldstone, J. V., Stegeman, J., Jonassen, I., & Goksøyr, A. (2021). The chemical defensome of five model teleost fish. *Scientific Reports*, 11(1), 1–13. <https://doi.org/10.1038/s41598-021-89948-0>

El-Alfy, A., Larsen, B., & Schlenk, D. (2002). Effect of cortisol and urea on flavin monooxygenase activity and expression in rainbow trout, *Oncorhynchus mykiss*. *Marine Environmental Research*, 54(3–5), 275–278. [https://doi.org/10.1016/S0141-1136\(02\)00183-6](https://doi.org/10.1016/S0141-1136(02)00183-6)

Eloranta, J. J., & Kullak-Ublick, G. A. (2005). Coordinate transcriptional regulation of bile acid homeostasis and drug metabolism. *Archives of Biochemistry and Biophysics*, 433(2), 397–412. <https://doi.org/10.1016/J.ABB.2004.09.019>

Encerrado-Manriquez, A., Spooner, Z., Truong, T., Fine, J., Nicklisch, S., Encerrado-Manriquez, A. M., Spooner, Z. T., Truong, T. T., Fine, J. D., & Nicklisch, S. C. (2025). Developmental and Caste-specific Expression Patterns of ATP-Binding Cassette (ABC) Transporters in Honey bees (*Apis mellifera*). *Authorea Preprints*. <https://doi.org/10.22541/AU.173746687.74569861/V1>

Erwin, D. H. (2015). Early metazoan life: divergence, environment and ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1684). <https://doi.org/10.1098/RSTB.2015.0036>

Erwin, D. H., Laflamme, M., Tweedt, S. M., Sperling, E. A., Pisani, D., & Peterson, K. J. (2011). The Cambrian conundrum: Early divergence and later ecological success in the early history of animals. *Science*, 334(6059), 1091–1097. <https://doi.org/10.1126/science.1206375>

Escrivá García, H., Laudet, V., & Robinson-Rechavi, M. (2003). Nuclear receptors are markers of animal genome evolution. *Journal of Structural and Functional Genomics*, 3(1–4), 177–184. <https://doi.org/10.1023/A:1022638706822>

Escriva, H., Bertrand, S., & Laudet, V. (2004). The evolution of the nuclear receptor superfamily. *Essays in Biochemistry*, 40, 11–26. <https://doi.org/10.1042/BSE0400011>

Escriva, H., Delaunay, F., & Laudet, V. (2000). Ligand binding and nuclear receptor evolution. *BioEssays*, 22, 717–727. [https://doi.org/10.1002/1521-1878\(200008\)22:8<717::AID-BIES5>3.0.CO;2-I](https://doi.org/10.1002/1521-1878(200008)22:8<717::AID-BIES5>3.0.CO;2-I)

- Espinosa, E., Bautista, R., Larrosa, R., & Plata, O. (2024). Advancements in long-read genome sequencing technologies and algorithms. *Genomics*, 116(3), 110842. <https://doi.org/10.1016/J.YGENO.2024.110842>
- Evans, R. M., & Mangelsdorf, D. J. (2014). Nuclear receptors, RXR, and the big bang. *Cell*, 157(1), 255–266. <https://doi.org/10.1016/j.cell.2014.03.012>
- Ewing, B., & Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Research*, 8(3), 186–194. <https://doi.org/10.1101/GR.8.3.186>
- Ewing, B., Hillier, L. D., Wendl, M. C., & Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Research*, 8(3), 175–185. <https://doi.org/10.1101/GR.8.3.175>
- Fernández, R., & Gabaldón, T. (2020). Gene gain and loss across the metazoan tree of life. *Nature Ecology and Evolution*, 4(4), 524–533. <https://doi.org/10.1038/s41559-019-1069-x>
- Fernández-Boo, S., Machado, A., Castro, L. F. C., Azeredo, R., & Costas, B. (2024). Unravelling the main immune repertoire of *Paracentrotus lividus* following *Vibrio anguillarum* bath challenge. *Fish & Shellfish Immunology*, 147, 109431. <https://doi.org/10.1016/J.FSI.2024.109431>
- Ferrada, E., & Superti-Furga, G. (2022). A structure and evolutionary-based classification of solute carriers. *IScience*, 25(10). <https://doi.org/10.1016/j.isci.2022.105096>
- Ferreira, I., Gomes-Bispo, A., Lourenço, H., Matos, J., Afonso, C., Cardoso, C., Castanheira, I., Motta, C., Prates, J. A. M., & Bandarra, N. M. (2020). The chemical composition and lipid profile of the chub mackerel (*Scomber colias*) show a strong seasonal dependence: Contribution to a nutritional evaluation. *Biochimie*, 178, 181–189. <https://doi.org/10.1016/j.biochi.2020.09.022>
- Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G., & Ysebaert, M. (1976). Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* 1976 260:5551, 260(5551), 500–507. <https://doi.org/10.1038/260500a0>
- Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research*, 39(SUPPL. 2), W29–W37. <https://doi.org/10.1093/nar/gkr367>

- Fiorentini, F., Geier, M., Binda, C., Winkler, M., Faber, K., Hall, M., & Mattevi, A. (2016). Biocatalytic Characterization of Human FMO5: Unearthing Baeyer-Villiger Reactions in Humans. *ACS Chemical Biology*, 11(4), 1039–1048. <https://doi.org/10.1021/acscchembio.5b01016>
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L. I., Glodek, A., ... Venter, J. C. (1995). Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd. *Science*, 269(5223), 496–512. <https://doi.org/10.1126/SCIENCE.7542800>
- Fonseca, E. F., Ruivo, R., Machado, A. M., Conrado, F., Tay, B.-H. B. H., Venkatesh, B., Santos, M. M. M., & Castro, L. F. C. (2019). Evolutionary plasticity in detoxification gene modules: The preservation and loss of the pregnane X receptor in chondrichthyes lineages. *International Journal of Molecular Sciences*, 20(9), 2331. <https://doi.org/10.3390/ijms20092331>
- Fonseca, E., Machado, A. M., Vilas-Arrondo, N., Gomes-dos-Santos, A., Veríssimo, A., Esteves, P., Almeida, T., Themudo, G., Ruivo, R., Pérez, M., da Fonseca, R., Santos, M. M., Froufe, E., Román-Marcote, E., Venkatesh, B., & Castro, L. F. C. (2020). Cartilaginous fishes offer unique insights into the evolution of the nuclear receptor gene repertoire in gnathostomes. *General and Comparative Endocrinology*, 295, 113527. <https://doi.org/10.1016/j.ygcen.2020.113527>
- Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., Ciofi, C., Crottini, A., Godoy, J. A., Höglund, J., Malukiewicz, J., Mouton, A., Oomen, R. A., Paez, S., Palsbøll, P. J., Pampoulie, C., Ruiz-López, M. J., Svoldal, H., Theofanopoulou, C., ... Zammit, G. (2022). The era of reference genomes in conservation genomics. *Trends in Ecology & Evolution*, 37(3), 197–202. <https://doi.org/https://doi.org/10.1016/j.tree.2021.11.008>
- Forsman, A. (2014). Rethinking phenotypic plasticity and its consequences for individuals, populations and species. *Heredity* 2015 115:4, 115(4), 276–284. <https://doi.org/10.1038/hdy.2014.92>
- Fox, R. J., Donelson, J. M., Schunter, C., Ravasi, T., & Gaitán-Espitia, J. D. (2019). Beyond buying time: the role of plasticity in phenotypic adaptation to rapid environmental change. *Philosophical Transactions of the Royal Society B*, 374(1768). <https://doi.org/10.1098/RSTB.2018.0174>

- Francis, G. A., Fayard, E., Picard, F., & Auwerx, J. (2003). Nuclear Receptors and the Control of Metabolism. *Annual Review of Physiology*, 65(Volume 65, 2003), 261–311. <https://doi.org/10.1146/annurev.physiol.65.092101.142528>
- Franco, M. E., Araújo, C. V. M., Cervený, D., Koubová, A., Danneels, B., Goksøyr, A., & Bertram, M. G. (2025). The extended chemical defensome: emphasizing mechanisms of defense as key research avenues to tackle priority questions in environmental toxicology. *Environmental Toxicology and Chemistry*. <https://doi.org/10.1093/ETOJNL/VGAF190>
- Franco, M. E., Hollender, J., & Schirmer, K. (2025). Differential biotransformation ability may alter fish biodiversity in polluted waters. *Environment International*, 195, 109254. <https://doi.org/10.1016/J.ENVINT.2025.109254>
- Fretland, A. J., & Omiecinski, C. J. (2000). Epoxide hydrolases: biochemistry and molecular biology. *Chemico-Biological Interactions*, 129(1–2), 41–59. [https://doi.org/10.1016/S0009-2797\(00\)00197-6](https://doi.org/10.1016/S0009-2797(00)00197-6)
- Fullam, E., Westwood, I. M., Anderton, M. C., Lowe, E. D., Sim, E., & Noble, M. E. M. (2008). Divergence of Cofactor Recognition across Evolution: Coenzyme A Binding in a Prokaryotic Arylamine N-Acetyltransferase. *Journal of Molecular Biology*, 375(1), 178–191. <https://doi.org/10.1016/J.JMB.2007.10.019>
- Fusco, G., & Minelli, A. (2010). Phenotypic plasticity in development and evolution: facts and concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1540), 547–556. <https://doi.org/10.1098/RSTB.2009.0267>
- Gaines, R. R., Hammarlund, E. U., Hou, X., Qi, C., Gabbott, S. E., Zhao, Y., Peng, J., & Canfield, D. E. (2012). Mechanism for Burgess Shale-type preservation. *Proceedings of the National Academy of Sciences of the United States of America*, 109(14), 5180–5184. <https://doi.org/10.1073/pnas.1111784109>
- Gamage, N., Barnett, A., Hempel, N., Duggleby, R. G., Windmill, K. F., Martin, J. L., & McManus, M. E. (2006). Human Sulfotransferases and Their Role in Chemical Metabolism. *Toxicological Sciences*, 90(1), 5–22. <https://doi.org/10.1093/TOXSCI/KFJ061>
- Gamito, R., Pita, C., Teixeira, C., Costa, M. J., & Cabral, H. N. (2016). Trends in landings and vulnerability to climate change in different fleet components in the Portuguese coast. *Fisheries Research*, 181, 93–101. <https://doi.org/10.1016/j.fishres.2016.04.008>
- García-Alonso, J., Ayoola, J. A. O., Crompton, J., Rebscher, N., & Hardege, J. D. (2011). Development and maturation in the nereidid polychaetes *Platynereis dumerilii* and *Nereis succinea* exposed to xenoestrogens. *Comparative Biochemistry and Physiology Part C*:

Toxicology & Pharmacology, 154(3), 196–203.
<https://doi.org/10.1016/J.CBPC.2011.05.007>

Gauthier, J., Vincent, A. T., Charette, S. J., & Derome, N. (2019). A brief history of bioinformatics. *Briefings in Bioinformatics*, 20(6), 1981–1996.
<https://doi.org/10.1093/BIB/BBY063>

George Lai, W., Farah, N., Moniz, G. A., & Nancy Wong, Y. (2011). A Baeyer-Villiger Oxidation Specifically Catalyzed by Human Flavin-Containing Monooxygenase 5. *Drug Metabolism and Disposition*, 39(1), 61–70. <https://doi.org/10.1124/DMD.110.035360>

Gerhart, J. (2006). The deuterostome ancestor. *Journal of Cellular Physiology*, 209(3), 677–685. <https://doi.org/10.1002/JCP.20803>

Germain, P., Altucci, L., Bourguet, W., Rochette-Egly, C., & Gronemeyer, H. (2003). Nuclear receptor superfamily: principles of signaling. *Pure and Applied Chemistry*, 75(11–12), 1619–1664. <https://doi.org/10.1351/pac200375111619>

Giacomini, K. M., Huang, S. M., Tweedie, D. J., Benet, L. Z., Brouwer, K. L. R., Chu, X., Dahlin, A., Evers, R., Fischer, V., Hillgren, K. M., Hoffmaster, K. A., Ishikawa, T., Keppler, D., Kim, R. B., Lee, C. A., Niemi, M., Polli, J. W., Sugiyama, Y., Swaan, P. W., ... Zhang, L. (2010). Membrane transporters in drug development. *Nature Reviews Drug Discovery* 2010 9:3, 9(3), 215–236. <https://doi.org/10.1038/nrd3028>

Giani, A. M., Gallo, G. R., Gianfranceschi, L., & Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal*, 18, 9–19.
<https://doi.org/10.1016/J.CSBJ.2019.11.002>

Gibbs, P. E., & Bryan, G. W. (1986). Reproductive Failure in Populations of the Dog-Whelk, *Nucella Lapillus*, Caused by Imposex Induced by Tributyltin from Antifouling Paints. *Journal of the Marine Biological Association of the United Kingdom*, 66(4), 767–777. <https://doi.org/10.1017/S0025315400048414>

Gibbs, R. A. (2020). The Human Genome Project changed everything. *Nature Reviews Genetics* 2020 21:10, 21(10), 575–576. <https://doi.org/10.1038/s41576-020-0275-3>

Gibson, J., & Harwood, C. S. (2002). Metabolic diversity in aromatic compound utilization by anaerobic microbes. *Annual Review of Microbiology*, 56(Volume 56, 2002), 345–369.
<https://doi.org/10.1146/annurev.micro.56.012302.160749>

Giribet, G., Distel, D. L., Polz, M., Sterrer, W., & Wheeler, W. C. (2000). Triploblastic Relationships with Emphasis on the Acoelomates and the Position of Gnathostomulida, Cyclophora, Plathelminthes, and Chaetognatha: A Combined Approach of 18S rDNA

Sequences and Morphology. *Systematic Biology*, 49(3), 539–562.
<https://doi.org/10.1080/10635159950127385>

Glatt, H. (2000). Sulfotransferases in the bioactivation of xenobiotics. *Chemico-Biological Interactions*, 129(1–2), 141–170. [https://doi.org/10.1016/S0009-2797\(00\)00202-7](https://doi.org/10.1016/S0009-2797(00)00202-7)

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., & Oliver, S. G. (1996). Life with 6000 Genes. *Science*, 274(October), 546–567. <https://doi.org/10.1126/science.274.5287.546>

Goldstone, J. V. (2008). Environmental sensing and response genes in cnidaria: The chemical defensome in the sea anemone *Nematostella vectensis*. *Cell Biology and Toxicology*, 24(6), 483–502. <https://doi.org/10.1007/s10565-008-9107-5>

Goldstone, J. V., Goldstone, H. M. H., Morrison, A. M., Tarrant, A., Kern, S. E., Woodin, B. R., & Stegeman, J. J. (2007). Cytochrome P450 1 Genes in Early Deuterostomes (Tunicates and Sea Urchins) and Vertebrates (Chicken and Frog): Origin and Diversification of the CYP1 Gene Family. *Molecular Biology and Evolution*, 24(12), 2619–2631. <https://doi.org/10.1093/MOLBEV/MSM200>

Goldstone, J. V., Hamdoun, A., Cole, B. J., Howard-Ashby, M., Nebert, D. W., Scally, M., Dean, M., Epel, D., Hahn, M. E., & Stegeman, J. J. (2006). The chemical defensome: Environmental sensing and response genes in the *Strongylocentrotus purpuratus* genome. *Developmental Biology*, 300(1), 366–384. <https://doi.org/10.1016/j.ydbio.2006.08.066>

Gomes-dos-Santos, A., Hagemann, A., Valente, L., Malzahn, A. M., Monroig, Ó., Froufe, E., & Castro, L. F. C. (2021). Complete mitochondrial genome of the ragworm annelid *Hediste diversicolor* (of Müller, 1776) (Annelida: Nereididae). *Mitochondrial DNA Part B*, 6(10), 2849–2851. <https://doi.org/10.1080/23802359.2021.1970644>

Gonzalez, F. J., & Yu, A.-M. (2006). Cytochrome P450 and xenobiotic receptor humanized mice. *Annu. Rev. Pharmacol. Toxicol.*, 46(1), 41–64. <https://doi.org/10.1146/annurev.pharmtox.45.120403.100007>

Goodwin, B., Jones, S. A., Price, R. R., Watson, M. A., McKee, D. D., Moore, L. B., Galardi, C., Wilson, J. G., Lewis, M. C., Roth, M. E., Maloney, P. R., Willson, T. M., & Kliewer, S. A. (2000). A Regulatory Cascade of the Nuclear Receptors FXR, SHP-1, and LRH-1 Represses Bile Acid Biosynthesis. *Molecular Cell*, 6(3), 517–526. [https://doi.org/10.1016/S1097-2765\(00\)00051-4](https://doi.org/10.1016/S1097-2765(00)00051-4)

- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* 2016 17:6, 17(6), 333–351. <https://doi.org/10.1038/nrg.2016.49>
- Gore, A. C., Chappell, V. A., Fenton, S. E., Flaws, J. A., Nadal, A., Prins, G. S., Toppari, J., & Zoeller, R. T. (2015). EDC-2: the endocrine society's second scientific statement on endocrine-disrupting chemicals. *Endocrine Reviews*, 36(6), E1–E150. <https://doi.org/10.1210/er.2015-1010>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644–652. <https://doi.org/10.1038/nbt.1883>
- Gronemeyer, H., Gustafsson, J. Å., & Laudet, V. (2004). Principles for modulation of the nuclear receptor superfamily. *Nature Reviews Drug Discovery* 2002 3:11, 3(11), 950–964. <https://doi.org/10.1038/nrd1551>
- Guengerich, F. P. (1999). Cytochrome P-450 3A4: regulation and role in drug metabolism. *Annual Review of Pharmacology and Toxicology*, 39, 1–17. <https://doi.org/10.1146/ANNUREV.PHARMTOX.39.1.1>
- Guengerich, F. P. (2001). Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chemical Research in Toxicology*, 14(6), 611–650. <https://doi.org/10.1021/tx0002583>
- Guo, J. U., Agarwal, V., Guo, H., & Bartel, D. P. (2014). Expanded identification and characterization of mammalian circular RNAs. *Genome Biology*, 15(7), 1–14. <https://doi.org/10.1186/s13059-014-0409-z>
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., ... Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8), 1494–1512. <https://doi.org/10.1038/nprot.2013.084>
- Hahn, M. E., Allan, L. L., & Sherr, D. H. (2008). Regulation of Constitutive and Inducible AHR Signaling: Complex Interactions Involving the AHR Repressor. *Biochemical Pharmacology*, 77(4), 485. <https://doi.org/10.1016/J.BCP.2008.09.016>

- Hamilton, P. B., Rolshausen, G., Uren Webster, T. M., & Tyler, C. R. (2017). Adaptive capabilities and fitness consequences associated with pollution exposure in fish. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1712). <https://doi.org/10.1098/RSTB.2016.0042>
- Han, Z., Jiao, S., Kong, D., Shan, Z., & Zhang, X. (2011). Effects of β -endosulfan on the growth and reproduction of zebrafish (*Danio rerio*). *Environmental Toxicology and Chemistry*, 30(11), 2525–2531. <https://doi.org/10.1002/ETC.646>
- Hao, D. C., Chen, S. L., Mu, J., & Xiao, P. G. (2009). Molecular phylogeny, long-term evolution, and functional divergence of flavin-containing monooxygenases. *Genetica*, 137(2), 173–187. <https://doi.org/10.1007/s10709-009-9382-y>
- Haussler, M. R., Whitfield, G. K., Haussler, C. A., Hsieh, J. C., Thompson, P. D., Selznick, S. H., Dominguez, C. E., & Jurutka, P. W. (1998). The Nuclear Vitamin D Receptor: Biological and Molecular Regulatory Properties Revealed. *Journal of Bone and Mineral Research*, 13(3), 325–349. <https://doi.org/10.1359/JBMR.1998.13.3.325>
- Hayes, J. D., Flanagan, J. U., & Jowsey, I. R. (2005). Glutathione transferases. *Annual Review of Pharmacology and Toxicology*, 45(Volume 45, 2005), 51–88. <https://doi.org/10.1146/annurev.pharmtox.45.120403.095857>
- He, L., He, T., Farrar, S., Ji, L., Liu, T., & Ma, X. (2017). Antioxidants Maintain Cellular Redox Homeostasis by Elimination of Reactive Oxygen Species. *Cellular Physiology and Biochemistry*, 44(2), 532–553. <https://doi.org/10.1159/000485089>
- Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107(1), 1–8. <https://doi.org/10.1016/J.YGENO.2015.11.003>
- Hedges, S. B., Blair, J. E., Venturi, M. L., & Shoe, J. L. (2004). A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evolutionary Biology*, 4(1), 1–9. <https://doi.org/10.1186/1471-2148-4-2>
- Hediger, M. A., Cl  men  on, B., Burrier, R. E., & Bruford, E. A. (2013). The ABCs of membrane transporters in health and disease (SLC series): Introduction. *Molecular Aspects of Medicine*, 34(2–3), 95–107. <https://doi.org/10.1016/J.MAM.2012.12.009>
- Hediger, M. A., Romero, M. F., Peng, J. Bin, Rolfs, A., Takanaga, H., & Bruford, E. A. (2004). The ABCs of solute carriers: Physiological, pathological and therapeutic implications of human membrane transport proteins. *Pflugers Archiv European Journal of Physiology*, 447(5), 465–468. <https://doi.org/10.1007/s00424-003-1192-y>
- Hein, D. W. (2002). Molecular genetics and function of NAT1 and NAT2: role in aromatic amine metabolism and carcinogenesis. *Mutation Research/Fundamental and Molecular*

Mechanisms of Mutagenesis, 506–507, 65–77. [https://doi.org/10.1016/S0027-5107\(02\)00153-7](https://doi.org/10.1016/S0027-5107(02)00153-7)

Heindel, J. J., Newbold, R., & Schug, T. T. (2015). Endocrine disruptors and obesity. *Nature Reviews Endocrinology* 2015 11:11, 11(11), 653–661. <https://doi.org/10.1038/nrendo.2015.163>

Hejnal, A. (2010). A Twist in Time—The Evolution of Spiral Cleavage in the Light of Animal Phylogeny. *Integrative and Comparative Biology*, 50(5), 695–706. <https://doi.org/10.1093/ICB/ICQ103>

Hejnal, A., & Martindale, M. Q. (2008). Acoel development supports a simple planula-like urbilaterian. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1496), 1493–1501. <https://doi.org/10.1098/RSTB.2007.2239>

Hejnal, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G. W., Edgecombe, G. D., Martinez, P., Baguñà, J., Bailly, X., Jondelius, U., Wiens, M., Müller, W. E. G., Seaver, E., Wheeler, W. C., Martindale, M. Q., Giribet, G., & Dunn, C. W. (2009). Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proceedings of the Royal Society B: Biological Sciences*, 276(1677), 4261–4270. <https://doi.org/10.1098/RSPB.2009.0896>

Hejnal, A., & Pang, K. (2016). Xenacoelomorpha's significance for understanding bilaterian evolution. *Current Opinion in Genetics & Development*, 39, 48–54. <https://doi.org/10.1016/J.GDE.2016.05.019>

Henderson, M. C., Krueger, S. K., Siddens, L. K., Stevens, J. F., & Williams, D. E. (2004). S-oxygenation of the thioether organophosphate insecticides phorate and disulfoton by human lung flavin-containing monooxygenase 2. *Biochemical Pharmacology*, 68(5), 959–967. <https://doi.org/10.1016/j.bcp.2004.05.051>

Hernandez, D., Janmohamed, A., Chandan, P., Phillips, I. R., & Shephard, E. A. (2004). Organization and evolution of the flavin-containing monooxygenase genes of human and mouse: identification of novel gene and pseudogene clusters. *Pharmacogenetics*, 14(2), 117–130. <https://doi.org/10.1097/00008571-200402000-00006>

Herrera, A., Štindlová, A., Martínez, I., Rapp, J., Romero-Kutzner, V., Samper, M. D., Montoto, T., Aguiar-González, B., Packard, T., & Gómez, M. (2019). Microplastic ingestion by Atlantic chub mackerel (*Scomber colias*) in the Canary Islands coast. *Marine Pollution Bulletin*, 139, 127–135. <https://doi.org/10.1016/J.MARPOLBUL.2018.12.022>

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2015). BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, 32(5), 767–769. <https://doi.org/10.1093/bioinformatics/btv661>

- Hoff, K. J., Lomsadze, A., Borodovsky, M., & Stanke, M. (2019). Whole-genome annotation with BRAKER. *Methods in Molecular Biology*, 1962, 65–95. https://doi.org/10.1007/978-1-4939-9173-0_5
- Hoffman, P. F., Abbot, D. S., Ashkenazy, Y., Benn, D. I., Brocks, J. J., Cohen, P. A., Cox, G. M., Creveling, J. R., Donnadieu, Y., Erwin, D. H., Fairchild, I. J., Ferreira, D., Goodman, J. C., Halverson, G. P., Jansen, M. F., Le Hir, G., Love, G. D., Macdonald, F. A., Maloof, A. C., ... Warren, S. G. (2017). Snowball Earth climate dynamics and Cryogenian geology-geobiology. *Science Advances*, 3(11). <https://doi.org/10.1126/sciadv.1600983>
- Hoffman, P. F., Kaufman, A. J., Halverson, G. P., & Schrag, D. P. (1998). A neoproterozoic snowball earth. *Science*, 281(5381), 1342–1346. <https://doi.org/10.1126/science.281.5381.1342>
- Hofmann, G. E., & Todgham, A. E. (2009). Living in the now: Physiological mechanisms to tolerate a rapidly changing environment. *Annual Review of Physiology*, 72(Volume 72, 2010), 127–145. <https://doi.org/10.1146/annurev-physiol-021909-135900>
- Höglund, P. J., Nordström, K. J. V., Schiöth, H. B., & Fredriksson, R. (2011). The Solute Carrier Families Have a Remarkably Long Evolutionary History with the Majority of the Human Families Present before Divergence of Bilaterian Species. *Molecular Biology and Evolution*, 28(4), 1531–1541. <https://doi.org/10.1093/MOLBEV/MSQ350>
- Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, 30(1), 62–82. <https://doi.org/10.1111/MEC.15720>
- Holland, L., & Anderson, P. (2015). Evolution of basal deuterostome nervous systems. *Journal of Experimental Biology*, 218(4), 637–645. <https://doi.org/10.1242/JEB.109108>
- Holland, L. Z., Carvalho, J. E., Escrava, H., Laudet, V., Schubert, M., Shimeld, S. M., & Yu, J. K. (2013). Evolution of bilaterian central nervous systems: A single origin? *EvoDevo*, 4(1), 1–20. <https://doi.org/10.1186/2041-9139-4-27>
- Holland, P. (2011). *The animal kingdom: a very short introduction* (Vol. 293). Oxford University Press. <https://doi.org/10.1093/actrade/9780199593217.001.0001>
- Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R., & Zamir, A. (1965). Structure of a Ribonucleic Acid. *Science*, 147(3664), 1462–1465. <https://doi.org/10.1126/SCIENCE.147.3664.1462>
- Hu, J., Wang, Z., Sun, Z., Hu, B., Ayoola, A. O., Liang, F., Li, J., Sandoval, J. R., Cooper, D. N., Ye, K., Ruan, J., Xiao, C. Le, Wang, D., Wu, D. D., & Wang, S. (2024). NextDenovo:

an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biology*, 25(1), 1–19. <https://doi.org/10.1186/s13059-024-03252-4>

Huang, S., Howington, M. B., Dobry, C. J., Evans, C. R., & Leiser, S. F. (2021). Flavin-Containing Monooxygenases Are Conserved Regulators of Stress Resistance and Metabolism. *Frontiers in Cell and Developmental Biology*, 9, 151. <https://doi.org/10.3389/fcell.2021.630188>

Huang, W., Xu, F., Li, J., Li, L., Que, H., & Zhang, G. (2015). Evolution of a novel nuclear receptor subfamily with emphasis on the member from the Pacific oyster *Crassostrea gigas*. *Gene*, 567(2), 164–172. <https://doi.org/10.1016/J.GENE.2015.04.082>

Huber, D. R., Dean, M. N., & Summers, A. P. (2008). Hard prey, soft jaws and the ontogeny of feeding mechanics in the spotted ratfish *Hydrolagus coliei*. *Journal of The Royal Society Interface*, 5(25), 941–953. <https://doi.org/10.1098/RSIF.2007.1325>

Hutchison, C. A. (2007). DNA sequencing: bench to bedside and beyond. *Nucleic Acids Research*, 35(18), 6227–6237. <https://doi.org/10.1093/NAR/GKM688>

Inguglia, L., Chiaramonte, M., Arizza, V., Turiák, L., Vékey, K., Drahos, L., Pitonzo, R., Avellone, G., & Di Stefano, V. (2020). Changes in the proteome of sea urchin *Paracentrotus lividus* coelomocytes in response to LPS injection into the body cavity. *PLOS ONE*, 15(2), e0228893. <https://doi.org/10.1371/JOURNAL.PONE.0228893>

Iyanagi, T. (2007). Molecular Mechanism of Phase I and Phase II Drug-Metabolizing Enzymes: Implications for Detoxification. *International Review of Cytology*, 260, 35–112. [https://doi.org/10.1016/S0074-7696\(06\)60002-8](https://doi.org/10.1016/S0074-7696(06)60002-8)

Jackman, S. D., Coombe, L., Warren, R. L., Kirk, H., Trinh, E., MacLeod, T., Pleasance, S., Pandoh, P., Zhao, Y., Coope, R. J., Bousquet, J., Bohlmann, J., Jones, S. J. M., & Birol, I. (2020). Complete Mitochondrial Genome of a Gymnosperm, Sitka Spruce (*Picea sitchensis*), Indicates a Complex Physical Structure. *Genome Biology and Evolution*, 12(7), 1174–1179. <https://doi.org/10.1093/GBE/EVAA108>

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Tyson, J. R., Beggs, A. D., Dilthey, A. T., Fiddes, I. T., Malla, S., Marriott, H., Nieto, T., O’Grady, J., Olsen, H. E., Pedersen, B. S., Rhie, A., Richardson, H., Quinlan, A. R., ... Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology* 2018 36:4, 36(4), 338–345. <https://doi.org/10.1038/nbt.4060>

Jancova, P., Anzenbacher, P., & Anzenbacherova, E. (2010). Phase II drug metabolizing enzymes. *Biomedical Papers of the Medical Faculty of the University Palacky, Olomouc, Czechoslovakia*, 154(2), 103–116. <https://doi.org/10.5507/BP.2010.017>

Jatllon, O., Aury, J. M., Brunet, F., Petit, J. L., Stange-Thomann, N., Maucell, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., ... Roest Crolius, H. (2004). Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 2004 431:7011, 431(7011), 946–957. <https://doi.org/10.1038/nature03025>

Johnson, A. G., & Horton, H. F. (1972). *Length-weight relationship, food habits, parasites, and sex and age determination of the ratfish, Hydrolagus colliei (Lay and Bennett)*. 421–429.

Jonker, J. W., Buitelaar, M., Wagenaar, E., Van der Valk, M. A., Scheffer, G. L., Scheper, R. J., Plösch, T., Kuipers, F., Oude Elferink, R. P. J., Rosing, H., Beijnen, J. H., & Schinkel, A. H. (2002). The breast cancer resistance protein protects against a major chlorophyll-derived dietary phototoxin and protoporphyria. *Proceedings of the National Academy of Sciences of the United States of America*, 99(24), 15649–15654. <https://doi.org/10.1073/pnas.202607599>

Jönsson, E. M., Abrahamson, A., Brunström, B., & Brandt, I. (2006). Cytochrome P4501A induction in rainbow trout gills and liver following exposure to waterborne indigo, benzo[a]pyrene and 3,3',4,4',5-pentachlorobiphenyl. *Aquatic Toxicology*, 79(3), 226–232. <https://doi.org/10.1016/J.AQUATOX.2006.06.006>

Jou, W. M., Haegeman, G., Ysebaert, M., & Fiers, W. (1972). Nucleotide Sequence of the Gene Coding for the Bacteriophage MS2 Coat Protein. *Nature* 1972 237:5350, 237(5350), 82–88. <https://doi.org/10.1038/237082a0>

Kadmiel, M., & Cidlowski, J. A. (2013). Glucocorticoid receptor signaling in health and disease. *Trends in Pharmacological Sciences*, 34(9), 518–530. <https://doi.org/10.1016/j.tips.2013.07.003>

Kapli, P., Yang, Z., & Telford, M. J. (2020). Phylogenetic tree building in the genomic age. *Nature Reviews Genetics* 2020 21:7, 21(7), 428–444. <https://doi.org/10.1038/s41576-020-0233-0>

Karakoltsidis, P. A., Zotos, A., & Constantinides, S. M. (1995). Composition of the commercially important mediterranean finfish, crustaceans, and molluscs. *Journal of Food Composition and Analysis*, 8(3), 258–273. <https://doi.org/10.1006/jfca.1995.1019>

Kaul, S., Koo, H. L., Jenkins, J., Rizzo, M., Rooney, T., Tallon, L. J., Feldblyum, T., Nierman, W., Benito, M. I., Lin, X., Town, C. D., Venter, J. C., Fraser, C. M., Tabata, S., Nakamura, Y., Kaneko, T., Sato, S., Asamizu, E., Kato, T., ... Somerville, C. (2000).

Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000 408:6814, 408(6814), 796–815. <https://doi.org/10.1038/35048692>

Kaur, S., Jobling, S., Jones, C. S., Noble, L. R., Routledge, E. J., & Lockyer, A. E. (2015). The Nuclear Receptors of *Biomphalaria glabrata* and *Lottia gigantea*: Implications for Developing New Model Organisms. *PLOS ONE*, 10(4), e0121259. <https://doi.org/10.1371/JOURNAL.PONE.0121259>

Kayal, E., Roure, B., Philippe, H., Collins, A. G., & Lavrov, D. V. (2013). Cnidarian phylogenetic relationships as revealed by mitogenomics. *BMC Evolutionary Biology*, 13(1), 1–18. <https://doi.org/10.1186/1471-2148-13-5>

Keller, A. A., Horness, B. H., Tuttle, V. J., Wallace, J. R., Simon, V. H., Fruh, E. L., Bosley, K. L., & Kamikawa, D. J. (2006). *The 2002 US West Coast upper continental slope trawl survey of groundfish resources off Washington, Oregon, and California: Estimates of distribution, abundance, and length composition*. <https://doi.org/10.13140/RG.2.2.24116.17284>

Khan, A. A., & Quigley, J. G. (2011). Control of intracellular heme levels: Heme transporters and heme oxygenases. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1813(5), 668–682. <https://doi.org/10.1016/J.BBAMCR.2011.01.008>

Kidani, Y., & Bensinger, S. J. (2012). Liver X receptor and peroxisome proliferator-activated receptor as integrators of lipid homeostasis and immunity. *Immunological Reviews*, 249(1), 72–83. <https://doi.org/10.1111/J.1600-065X.2012.01153.X>

Kim, B.-M., Amores, A., Kang, S., Ahn, D.-H., Kim, J.-H., Kim, I.-C., Lee, J. H., Lee, S. G., Lee, H., Lee, J., Kim, H.-W., Desvignes, T., Batzel, P., Sydes, J., Titus, T., Wilson, C. A., Catchen, J. M., Warren, W. C., Schartl, M., ... Park, H. (2019). Antarctic blackfin icefish genome reveals adaptations to extreme environments. *Nature Ecology & Evolution*, 3(3), 469–478. <https://doi.org/10.1038/s41559-019-0812-7>

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357–360. <https://doi.org/10.1038/nmeth.3317>

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8), 907–915. <https://doi.org/10.1038/s41587-019-0201-4>

King, N., Westbrook, M. J., Young, S. L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., Marr, M., Pincus, D., Putnam, N., Rokas, A., Wright, K. J., Zuzow, R., Dirks, W., Good, M., Goodstein, D., ... Rokhsar, D. (2008). The genome

of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 2008 451:7180, 451(7180), 783–788. <https://doi.org/10.1038/nature06617>

Kivimäki, M., Bartolomucci, A., & Kawachi, I. (2022). The multiple roles of life stress in metabolic disorders. *Nature Reviews Endocrinology* 2022 19:1, 19(1), 10–27. <https://doi.org/10.1038/s41574-022-00746-8>

Kliwer, S. A., Goodwin, B., & Willson, T. M. (2002). The Nuclear Pregnane X Receptor: A Key Regulator of Xenobiotic Metabolism. *Endocrine Reviews*, 23(5), 687–702. <https://doi.org/10.1210/ER.2001-0038>

Kliwer, S. A., Moore, J. T., Wade, L., Staudinger, J. L., Watson, M. A., Jones, S. A., McKee, D. D., Oliver, B. B., Willson, T. M., Zetterström, R. H., Perlmann, T., & Lehmann, J. M. (1998). An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway. *Cell*, 92(1), 73–82. [https://doi.org/10.1016/S0092-8674\(00\)80900-9](https://doi.org/10.1016/S0092-8674(00)80900-9)

Knoll, A. H., Bergmann, K. D., & Strauss, J. V. (2016). Life: the first two billion years. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1707). <https://doi.org/10.1098/RSTB.2015.0493>

Knoll, A. H., Walter, M. R., Narbonne, G. M., & Christie-Blick, N. (2006). The Ediacaran period: A new addition to the geologic time scale. *Lethaia*, 39(1), 13–30. <https://doi.org/10.1080/00241160500409223>

Kong, L., Liu, P., Li, M., Wang, H., Shi, J., Hu, J., Li, Y., & Hu, X. (2021). Transcriptional Responses of Flavin-Containing Monooxygenase Genes in Scallops Exposed to PST-Producing Dinoflagellates Implying Their Involvements in Detoxification. *Frontiers in Marine Science*, 8, 732000. <https://doi.org/10.3389/fmars.2021.732000>

Koonin, E. V. (2005). Orthologs, paralogs, and evolutionary genomics. *Annual Review of Genetics*, 39, 309–338. <https://doi.org/10.1146/annurev.genet.39.073003.114725>

Koukouritaki, S. B., Simpson, P., Yeung, C. K., Rettie, A. E., & Hines, R. N. (2002). Human hepatic flavin-containing monooxygenases 1 (FMO1) and 3 (FMO3) developmental expression. *Pediatric Research*, 51(2), 236–243. <https://doi.org/10.1203/00006450-200202000-00018>

Krampis, K., Booth, T., Chapman, B., Tiwari, B., Bicak, M., Field, D., & Nelson, K. E. (2012). Cloud BioLinux: Pre-configured and on-demand bioinformatics computing for the genomics community. *BMC Bioinformatics*, 13(1), 1–8. <https://doi.org/10.1186/1471-2105-13-42>

Krasowski, M. D., Yasuda, K., Hagey, L. R., & Schuetz, E. G. (2005). Evolutionary selection across the nuclear hormone receptor superfamily with a focus on the NR1I

subfamily (vitamin D, pregnane X, and constitutive androstane receptors). *Nuclear Receptor*, 3(1), 1–20. <https://doi.org/10.1186/1478-1336-3-2>

Krueger, S. K., Siddens, L. K., Henderson, M. C., Andreasen, E. A., Tanguay, R. L., Pereira, C. B., Cabacungan, E. T., Hines, R. N., Ardlie, K. G., & Williams, D. E. (2005). Haplotype and functional analysis of four flavin-containing monooxygenase isoform 2 (FMO2) polymorphisms in Hispanics. *Pharmacogenetics and Genomics*, 15(4), 245. <https://doi.org/10.1097/01213011-200504000-00008>

Krueger, S. K., & Williams, D. E. (2005). Mammalian flavin-containing monooxygenases: Structure/function, genetic polymorphisms and role in drug metabolism. In *Pharmacology and Therapeutics* (Vol. 106, Issue 3, pp. 357–387). Pergamon. <https://doi.org/10.1016/j.pharmthera.2005.01.001>

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessi eres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., ... Danchin, A. (1997). The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* 1997 390:6657, 390(6657), 249–256. <https://doi.org/10.1038/36786>

Kuraku, S., Nishimura, O., Rozewicki, J., Yamaguchi, K., Tatsumi, K., Ohishi, Y., Ohta, T., Yagura, M., Niwa, T., Tanegashima, C., Teramura, A., Hirase, S., Kawaguchi, A., Tan, M., D’Aniello, S., Castro, F., Machado, A., Koyanagi, M., Terakita, A., ... Uno, Y. (2022). Squalomix: shark and ray genome analysis consortium and its data sharing platform. *F1000Research*, 11, 1077. <https://doi.org/10.12688/F1000RESEARCH.123591.1>

Lang, D., Yeung, C., Peter, R., Ibarra, C., Gasser, R., Itagaki, K., Philpot, R., & Rettie, A. (1998). Isoform specificity of trimethylamine N-oxygenation by human flavin-containing monooxygenase (FMO) and P450 enzymes: selective catalysis by FMO3. *Biochemical Pharmacology*, 56(8), 1005–1012. [https://doi.org/10.1016/s0006-2952\(98\)00218-4](https://doi.org/10.1016/s0006-2952(98)00218-4)

Larsen, B. K., & Schlenk, D. (2001a). Effect of salinity on flavin-containing monooxygenase expression and activity in rainbow trout (*Oncorhynchus mykiss*). *Journal of Comparative Physiology. B, Biochemical, Systemic, and Environmental Physiology*, 171(5), 421–429. <https://doi.org/10.1007/S003600100192>

Larsen, B. K., & Schlenk, D. (2001b). Effect of urea and low temperature on the expression and activity of flavin-containing monooxygenase in the liver and gill of rainbow trout (*Oncorhynchus mykiss*). *Fish Physiology and Biochemistry*, 25(1), 19–29. <https://doi.org/10.1023/A:1019769026824>

- Laudet, V., Hänni, C., Coll, J., Catzeflis, F., & Stéhelin, D. (1992). Evolution of the nuclear receptor gene superfamily. *The EMBO Journal*, 11(3), 1003. <https://doi.org/10.1002/J.1460-2075.1992.TB05139.X>
- Laumer, C. E., Fernández, R., Lemer, S., Combosch, D., Kocot, K. M., Riesgo, A., Andrade, S. C. S., Sterrer, W., Sørensen, M. V., & Giribet, G. (2019). Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proceedings of the Royal Society B*, 286(1906). <https://doi.org/10.1098/RSPB.2019.0831>
- Lavado, R., Aparicio-Fabre, R., & Schlenk, D. (2013). Effects of salinity acclimation on the pesticide-metabolizing enzyme flavin-containing monooxygenase (FMO) in rainbow trout (*Oncorhynchus mykiss*). *Comparative Biochemistry and Physiology. Toxicology & Pharmacology: CBP*, 157(1), 9–15. <https://doi.org/10.1016/J.CBPC.2012.08.004>
- Lavinsky, R. M., Jepsen, K., Heinzl, T., Torchia, J., Mullen, T. M., Schiff, R., Del-Rio, A. L., Ricote, M., Ngo, S., Gemsch, J., Hilsenbeck, S. G., Osborne, C. K., Glass, C. K., Rosenfeld, M. G., & Rose, D. W. (1998). Diverse signaling pathways modulate nuclear receptor recruitment of N-CoR and SMRT complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 95(6), 2920–2925. <https://doi.org/10.1073/pnas.95.6.2920>
- Lee, S. O., Jin, U. H., Kang, J. H., Kim, S. B., Guthrie, A. S., Sreevalsan, S., Lee, J. S., & Safe, S. (2014). The Orphan Nuclear Receptor NR4A1 (Nur77) Regulates Oxidative and Endoplasmic Reticulum Stress in Pancreatic Cancer Cells. *Molecular Cancer Research: MCR*, 12(4), 527. <https://doi.org/10.1158/1541-7786.MCR-13-0567>
- Lefebvre, P., Cariou, B., Lien, F., Kuipers, F., & Staels, B. (2009). Role of bile acids and bile acid receptors in metabolic regulation. *Physiological Reviews*, 89(1), 147–191. <https://doi.org/10.1152/physrev.00010.2008>
- Levene, H. J., Korlach, J., Turner, S. W., Foquet, M., Craighead, H. G., & Webb, W. W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science*, 299(5607), 682–686. <https://doi.org/10.1126/science.1079700>
- Lewin, H. A., Richards, S., Aiden, E. L., Allende, M. L., Archibald, J. M., Bálint, M., Barker, K. B., Baumgartner, B., Belov, K., Bertorelle, G., Blaxter, M. L., Cai, J., Caperello, N. D., Carlson, K., Castilla-Rubio, J. C., Chaw, S. M., Chen, L., Childers, A. K., Coddington, J. A., ... Zhang, G. (2022). The Earth BioGenome Project 2020: Starting the clock. *Proceedings of the National Academy of Sciences of the United States of America*, 119(4), e2115635118. <https://doi.org/10.1073/pnas.2115635118>

- Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., Durbin, R., Edwards, S. V., Forest, F., Gilbert, M. T. P., Goldstein, M. M., Grigoriev, I. V., Hackett, K. J., Haussler, D., Jarvis, E. D., Johnson, W. E., Patrinos, A., Richards, S., Castilla-Rubio, J. C., ... Zhang, G. (2018). Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences of the United States of America*, 115(17), 4325–4333. <https://doi.org/10.1073/pnas.1720115111>
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., & Durbin, R. (2024). Genome assembly in the telomere-to-telomere era. *Nature Reviews Genetics* 2024 25:9, 25(9), 658–670. <https://doi.org/10.1038/s41576-024-00718-w>
- Li, X., Schuler, M. A., & Berenbaum, M. R. (2007). Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annual Review of Entomology*, 52(Volume 52, 2007), 231–253. <https://doi.org/10.1146/annurev.ento.51.110104.151104>
- Li, Y., Meng, Q., Yang, M., Liu, D., Hou, X., Tang, L., Wang, X., Lyu, Y., Chen, X., Liu, K., Yu, A. M., Zuo, Z., & Bi, H. (2019). Current trends in drug metabolism and pharmacokinetics. *Acta Pharmaceutica Sinica. B*, 9(6), 1113. <https://doi.org/10.1016/J.APSB.2019.10.001>
- Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., Hvidsten, T. R., Leong, J. S., Minkley, D. R., Zimin, A., Grammes, F., Grove, H., Gjuvsland, A., Walenz, B., Hermansen, R. A., Von Schalburg, K., Rondeau, E. B., Di Genova, A., Samy, J. K. A., ... Davidson, W. S. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature* 2016 533:7602, 533(7602), 200–205. <https://doi.org/10.1038/nature17164>
- Lin, C. Y., Marlétaz, F., Pérez-Posada, A., Martínez-García, P. M., Schloissnig, S., Peluso, P., Conception, G. T., Bump, P., Chen, Y. C., Chou, C., Lin, C. Y., Fan, T. P., Tsai, C. T., Skarmeta, J. L. G., Tena, J. J., Lowe, C. J., Rank, D. R., Rokhsar, D. S., Yu, K., & Su, Y. H. (2024). Chromosome-level genome assemblies of 2 hemichordates provide new insights into deuterostome origin and chromosome evolution. *PLOS Biology*, 22(6), e3002661. <https://doi.org/10.1371/JOURNAL.PBIO.3002661>

- Lin, L., Yee, S. W., Kim, R. B., & Giacomini, K. M. (2015). SLC transporters as therapeutic targets: emerging opportunities. *Nature Reviews Drug Discovery* 2015 14:8, 14(8), 543–560. <https://doi.org/10.1038/nrd4626>
- Liu, X. (2019). SLC Family Transporters. *Advances in Experimental Medicine and Biology*, 1141, 101–202. https://doi.org/10.1007/978-981-13-7647-4_3
- Livingstone, D. R. (2001). Contaminant-stimulated Reactive Oxygen Species Production and Oxidative Damage in Aquatic Organisms. *Marine Pollution Bulletin*, 42(8), 656–666. [https://doi.org/10.1016/S0025-326X\(01\)00060-1](https://doi.org/10.1016/S0025-326X(01)00060-1)
- Lowe, C. J., Wu, M., Salic, A., Evans, L., Lander, E., Stange-Thomann, N., Gruber, C. E., Gerhart, J., & Kirschner, M. (2003). Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. *Cell*, 113(7), 853–865. [https://doi.org/10.1016/s0092-8674\(03\)00469-0](https://doi.org/10.1016/s0092-8674(03)00469-0)
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics technologies. *PLOS Computational Biology*, 13(5), e1005457. <https://doi.org/10.1371/JOURNAL.PCBI.1005457>
- Lozano-Bilbao, E., Jurado-Ruzafa, A., Lorenzo, J. M., González, J. A., Hardisson, A., González-Weller, D., Paz, S., Rubio, C., & Gutiérrez, Á. J. (2024). Analysis of two coastal pelagic fish species (*Sardinella aurita* and *Scomber colias*) for early detection of the impact of anthropogenic metals released into the marine environment. *Regional Studies in Marine Science*, 74, 103524. <https://doi.org/10.1016/J.RSMA.2024.103524>
- Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics*, 14(5), 265–279. <https://doi.org/10.1016/J.GPB.2016.05.004>
- Lu, N., Li, Y., Zhang, Y., Yan, Z., Wang, X., & Wang, Y. (2025). Transcriptome analysis reveals the different toxic mechanism of three HBCD diastereoisomers to *Brachionus plicatilis* based on chemical defensome. *Journal of Hazardous Materials*, 487, 137224. <https://doi.org/10.1016/J.JHAZMAT.2025.137224>
- Machado, A. M., Felício, M., Fonseca, E., da Fonseca, R. R., & Castro, L. F. C. (2018). A resource for sustainable management: De novo assembly and annotation of the liver transcriptome of the Atlantic chub mackerel, *Scomber colias*. *Data in Brief*, 18, 276–284. <https://doi.org/10.1016/j.dib.2018.03.013>
- Machado, A. M., Fernández-Boo, S., Nande, M., Pinto, R., Costas, B., & Castro, L. F. C. (2022). The male and female gonad transcriptome of the edible sea urchin, *Paracentrotus*

lividus: Identification of sex-related and lipid biosynthesis genes. *Aquaculture Reports*, 22, 100936. <https://doi.org/10.1016/J.AQREP.2021.100936>

Machado, A. M., Gomes-dos-Santos, A., Fonseca, M. M., Fonseca, R. R. da, Veríssimo, A., Felício, M., Capela, R., Alves, N., Santos, M., Salvador-Caramelo, F., Domingues, M., Ruivo, R., Froufe, E., Castro, L. F. C., Gomes-dos-Santos, A., Fonseca, M. M., Fonseca, R. R. da, Veríssimo, A., Felício, M., ... Castro, L. F. C. (2022). A genome assembly of the Atlantic chub mackerel (*Scomber colias*): a valuable teleost fishing resource. *Gigabyte*, 2022, 1–21. <https://doi.org/10.46471/GIGABYTE.40>

Machado, A. M., Samico, R., Domingues, M., Hagemann, A., Valente, L. M. P., Malzahn, A. M., Gomes-dos-Santos, A., Ruivo, R., Navarro, J. C., Monroig, Ó., & Castro, L. F. C. (2024). A whole-body transcriptome assembly of the annelid worm *Hediste diversicolor*. *Marine Genomics*, 74, 101084. <https://doi.org/10.1016/J.MARGEN.2024.101084>

Maddison, W. P., Donoghue, M. J., & Maddison, D. R. (1984). Outgroup Analysis and Parsimony. *Systematic Biology*, 33(1), 83–103. <https://doi.org/10.1093/SYSBIO/33.1.83>

Mángano, M. G., & Buatois, L. A. (2014). Decoupling of body-plan diversification and ecological structuring during the Ediacaran–Cambrian transition: evolutionary and geobiological feedbacks. *Proceedings of the Royal Society B: Biological Sciences*, 281(1780). <https://doi.org/10.1098/RSPB.2014.0038>

Mangelsdorf, D. J., & Evans, R. M. (1995). The RXR heterodimers and orphan receptors. *Cell*, 83(6), 841–850. [https://doi.org/10.1016/0092-8674\(95\)90200-7](https://doi.org/10.1016/0092-8674(95)90200-7)

Mannervik, B., Helena Danielson, U., & Ketterer, B. (1988). Glutathione transferases-structure and catalytic activit. *Critical Reviews in Biochemistry and Molecular Biology*, 23(3), 283–337. <https://doi.org/10.3109/10409238809088226>

Marrone, V., Piscopo, M., Romano, G., Ianora, A., Palumbo, A., & Costantini, M. (2012). Defensome against Toxic Diatom Aldehydes in the Sea Urchin *Paracentrotus lividus*. *PLoS ONE*, 7(2), e31750. <https://doi.org/10.1371/JOURNAL.PONE.0031750>

Marrs, K. A. (1996). The functions and regulation of glutathione s-transferases in plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, 47(1), 127–158. <https://doi.org/10.1146/annurev.arplant.47.1.127>

Marshall, C. R. (2006). Explaining the Cambrian “explosion” of animals. *Annual Review of Earth and Planetary Sciences*, 34(Volume 34, 2006), 355–384. <https://doi.org/10.1146/annurev.earth.33.031504.103001>

Martindale, M. Q. (2005). The evolution of metazoan axial properties. *Nature Reviews Genetics* 2005 6:12, 6(12), 917–927. <https://doi.org/10.1038/nrg1725>

- Martín-Durán, J. M., Janssen, R., Wennberg, S., Budd, G. E., & Hejnol, A. (2012). Deuterostomic Development in the Protostome *Priapulus caudatus*. *Current Biology*, 22(22), 2161–2166. <https://doi.org/10.1016/J.CUB.2012.09.037>
- Martín-Durán, J. M., Passamanek, Y. J., Martindale, M. Q., & Hejnol, A. (2016). The developmental basis for the recurrent evolution of deuterostomy and protostomy. *Nature Ecology & Evolution* 2016 1:1, 1(1), 1–10. <https://doi.org/10.1038/s41559-016-0005>
- Martins, M. M., Skagen, D., Marques, V., Zwolinski, J., & Silva, A. (2013). Changes in the abundance and spatial distribution of the Atlantic chub mackerel (*Scomber colias*) in the pelagic ecosystem and fisheries off Portugal. *Scientia Marina*, 77(4), 551–563.
- Mascotti, M. L., Juri Ayub, M., Furnham, N., Thornton, J. M., & Laskowski, R. A. (2016). Chopping and Changing: the Evolution of the Flavin-dependent Monooxygenases. *Journal of Molecular Biology*, 428(15), 3131–3146. <https://doi.org/10.1016/J.JMB.2016.07.003>
- Mascotti, M. L., Lapadula, W. J., & Ayub, M. J. (2015). The Origin and Evolution of Baeyer—Villiger Monooxygenases (BVMOs): An Ancestral Family of Flavin Monooxygenases. *PLoS ONE*, 10(7), e0132689. <https://doi.org/10.1371/JOURNAL.PONE.0132689>
- Mathäs, M., Burk, O., Qiu, H., Nußhag, C., Gödtel-Armbrust, U., Baranyai, D., Deng, S., Rmer, K., Nem, D., Windshügel, B., & Wojnowski, L. (2011). Evolutionary History and Functional Characterization of the Amphibian Xenosensor CAR. *Molecular Endocrinology*, 26(1), 14. <https://doi.org/10.1210/ME.2011-1235>
- Mathews, C. P. (1975). Some observations on the ecology of *Galeus piperatus* Springer and Wagner, a little known shark endemic to the Northern Gulf of California*. *Journal of Fish Biology*, 7(1), 77–82. <https://doi.org/10.1111/J.1095-8649.1975.TB04577.X>
- Matthiessen, P., & Gibbs, P. E. (1998). Critical appraisal of the evidence for tributyltin-mediated endocrine disruption in mollusks. *Environmental Toxicology and Chemistry*, 17(1), 37–43. <https://doi.org/10.1002/ETC.5620170106>
- McCulloch, G. A., & Waters, J. M. (2023). Rapid adaptation in a fast-changing world: Emerging insights from insect genomics. *Global Change Biology*, 29(4), 943–954. <https://doi.org/10.1111/GCB.16512>
- McDonnell, D. P., & Norris, J. D. (2002). Connection and regulation of the human estrogen receptor. *Science*, 296(5573), 1642–1644. <https://doi.org/10.1126/science.1071884>

- McIntosh, B. E., Hogenesch, J. B., & Bradfield, C. A. (2009). Mammalian Per-Arnt-Sim proteins in environmental adaptation. *Annual Review of Physiology*, 72(Volume 72, 2010), 625–645. <https://doi.org/10.1146/annurev-physiol-021909-135922>
- Mehrmohamadi, M., Sepehri, M. H., Nazer, N., & Norouzi, M. R. (2021). A Comparative Overview of Epigenomic Profiling Methods. *Frontiers in Cell and Developmental Biology*, 9. <https://doi.org/10.3389/FCELL.2021.714687>
- Meyer, U. A. (1996). Overview of enzymes of drug metabolism. *Journal of Pharmacokinetics and Biopharmaceutics*, 24(5), 449–459. <https://doi.org/10.1007/BF02353473>
- Michelangeli, M., Martin, J. M., Pinter-Wollman, N., Ioannou, C. C., McCallum, E. S., Bertram, M. G., & Brodin, T. (2022). Predicting the impacts of chemical pollutants on animal groups. *Trends in Ecology & Evolution*, 37(9), 789–802. <https://doi.org/10.1016/J.TREE.2022.05.009>
- Miga, K. H., Koren, S., Rhie, A., Vollger, M. R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G. A., Schneider, V. A., Potapova, T., Wood, J., Chow, W., Armstrong, J., Fredrickson, J., Pak, E., Tigyi, K., Kremitzki, M., ... Phillippy, A. M. (2020). Telomere-to-telomere assembly of a complete human X chromosome. *Nature* 2020 585:7823, 585(7823), 79–84. <https://doi.org/10.1038/s41586-020-2547-7>
- Milito, A., Murano, C., Castellano, I., Romano, G., & Palumbo, A. (2020). Antioxidant and immune response of the sea urchin *Paracentrotus lividus* to different re-suspension patterns of highly polluted marine sediments. *Marine Environmental Research*, 160, 104978. <https://doi.org/10.1016/J.MARENVRES.2020.104978>
- Minelli, A. (2007). Animal Evolution: Interrelationships of the Living Phyla. *Zoological Journal of the Linnean Society*, 151(1), 216–217. <https://doi.org/10.1111/J.1096-3642.2007.00296.X>
- Mithöfer, A., & Boland, W. (2012). Plant defense against herbivores: Chemical aspects. *Annual Review of Plant Biology*, 63(Volume 63, 2012), 431–450. <https://doi.org/10.1146/annurev-arplant-042110-103854>
- Møbjerg, N., Halberg, K. A., Jørgensen, A., Persson, D., Bjørn, M., Ramløv, H., & Kristensen, R. M. (2011). Survival in extreme environments – on the current knowledge of adaptations in tardigrades. *Acta Physiologica*, 202(3), 409–420. <https://doi.org/10.1111/J.1748-1716.2011.02252.X>

- Möglich, A., Ayers, R. A., & Moffat, K. (2009). Structure and Signaling Mechanism of Per-ARNT-Sim Domains. *Structure*, 17(10), 1282–1294. <https://doi.org/10.1016/j.str.2009.08.011>
- Mohammed, H., Russell, I. A., Stark, R., Rueda, O. M., Hickey, T. E., Tarulli, G. A., Serandour, A. A. A., Birrell, S. N., Bruna, A., Saadi, A., Menon, S., Hadfield, J., Pugh, M., Raj, G. V., Brown, G. D., D'Santos, C., Robinson, J. L. L., Silva, G., Launchbury, R., ... Carroll, J. S. (2015). Progesterone receptor modulates ER α action in breast cancer. *Nature* 2015 523:7560, 523(7560), 313–317. <https://doi.org/10.1038/nature14583>
- Motohashi, H., O'Connor, T., Katsuoka, F., Engel, J. D., & Yamamoto, M. (2002). Integration and diversity of the regulatory network composed of Maf and CNC families of transcription factors. *Gene*, 294(1–2), 1–12. [https://doi.org/10.1016/S0378-1119\(02\)00788-6](https://doi.org/10.1016/S0378-1119(02)00788-6)
- Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., Kravitz, S. A., Mobarry, C. M., Reinert, K. H. J., Remington, K. A., Anson, E. L., Bolanos, R. A., Chou, H. H., Jordan, C. M., Halpern, A. L., Lonardi, S., Beasley, E. M., Brandon, R. C., Chen, L., ... Venter, J. C. (2000). A whole-genome assembly of *Drosophila*. *Science*, 287(5461), 2196–2204. <https://doi.org/10.1126/science.287.5461.2196>
- Na, L., & Kiessling, W. (2015). Diversity partitioning during the Cambrian radiation. *Proceedings of the National Academy of Sciences of the United States of America*, 112(15), 4702–4706. <https://doi.org/10.1073/pnas.1424985112>
- Nagata, K., & Yamazoe, Y. (2000). Pharmacogenetics of sulfotransferase. *Annual Review of Pharmacology and Toxicology*, 40, 159–176. <https://doi.org/10.1146/ANNUREV.PHARMTOX.40.1.159>
- Nanglu, K., Caron, J. B., & Gaines, R. R. (2020). The Burgess Shale paleocommunity with new insights from Marble Canyon, British Columbia. *Paleobiology*, 46(1), 58–81. <https://doi.org/10.1017/PAB.2019.42>
- Naumann, C., Hartmann, T., & Ober, D. (2002). Evolutionary recruitment of a flavin-dependent monooxygenase for the detoxification of host plant-acquired pyrrolizidine alkaloids in the alkaloid-defended arctiid moth *Tyria jacobaeae*. *Proceedings of the National Academy of Sciences of the United States of America*, 99(9), 6085. <https://doi.org/10.1073/PNAS.082674499>
- Nebert, D. W., & Dalton, T. P. (2006). The role of cytochrome P450 enzymes in endogenous signalling pathways and environmental carcinogenesis. *Nature Reviews. Cancer*, 6(12), 947–960. <https://doi.org/10.1038/NRC2015>

- Nebert, D. W., & Russell, D. W. (2002). Clinical importance of the cytochromes P450. *The Lancet*, 360(9340), 1155–1162. [https://doi.org/10.1016/S0140-6736\(02\)11203-7](https://doi.org/10.1016/S0140-6736(02)11203-7)
- Nebert, D. W., & Vasiliou, V. (2004). Analysis of the glutathione S-transferase (GST) gene family. *Human Genomics*, 1(6), 460. <https://doi.org/10.1186/1479-7364-1-6-460>
- Nelson, D. R., Goldstone, J. V., & Stegeman, J. J. (2013). The cytochrome P450 genesis locus: the origin and evolution of animal cytochrome P450s. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1612). <https://doi.org/10.1098/RSTB.2012.0474>
- Nelson, J. (2016). *Fishes of the World* (Fifth edition). John Wiley & Sons. <https://doi.org/10.1002/9781119174844>
- Nguyen, T., Sherratt, P. J., & Pickett, C. B. (2003). Regulatory Mechanisms Controlling Gene Expression Mediated by the Antioxidant Response Element. *Annual Review of Pharmacology and Toxicology*, 43(Volume 43, 2003), 233–260. <https://doi.org/10.1146/annurev.pharmtox.43.100901.140229>
- Nicoll, C. R., Bailleul, G., Fiorentini, F., Mascotti, M. L., Fraaije, M. W., & Mattevi, A. (2019). Ancestral-sequence reconstruction unveils the structural basis of function in mammalian FMOs. *Nature Structural & Molecular Biology* 2019 27:1, 27(1), 14–24. <https://doi.org/10.1038/s41594-019-0347-2>
- Nicoll, C. R., & Mascotti, M. L. (2023). Investigating the biochemical signatures and physiological roles of the FMO family using molecular phylogeny. *BBA Advances*, 4, 100108. <https://doi.org/10.1016/J.BBADVA.2023.100108>
- Nielsen, C. (2005). Trochophora larvae: cell-lineages, ciliary bands and body regions. 2. Other groups and general discussion. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 304B(5), 401–447. <https://doi.org/10.1002/JEZ.B.21050>
- Nielsen, C. (2012). *Animal evolution: interrelationships of the living phyla*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199606023.001.0001>
- Nielsen, C. (2017). Evolution of deuterostomy – and origin of the chordates. *Biological Reviews*, 92(1), 316–325. <https://doi.org/10.1111/BRV.12229>
- Nielsen, C. (2019). Blastopore Fate: Amphistomy, Protostomy or Deuterostomy. *Encyclopedia of Life Sciences*, 1–8. <https://doi.org/10.1002/9780470015902.A0027481>
- Nielsen, C., Scharff, N., & Eibye-Jacobsen, D. (1996). Cladistic analyses of the animal kingdom. *Biological Journal of the Linnean Society*, 57(4), 385–410. <https://doi.org/10.1111/J.1095-8312.1996.TB01857.X>

- Nigam, S. K. (2014). What do drug transporters really do? *Nature Reviews Drug Discovery* 2014 14:1, 14(1), 29–44. <https://doi.org/10.1038/nrd4461>
- Niture, S. K., Khatri, R., & Jaiswal, A. K. (2014). Regulation of Nrf2—an update. *Free Radical Biology and Medicine*, 66, 36–44. <https://doi.org/10.1016/J.FREERADBIOMED.2013.02.008>
- Noguera-Oviedo, K., & Aga, D. S. (2016). Lessons learned from more than two decades of research on emerging contaminants in the environment. *Journal of Hazardous Materials*, 316, 242–251. <https://doi.org/10.1016/J.JHAZMAT.2016.04.058>
- Nurk, S., Walenz, B. P., Rhie, A., Vollger, M. R., Logsdon, G. A., Grothe, R., Miga, K. H., Eichler, E. E., Phillippy, A. M., & Koren, S. (2020). HiCanu: Accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Research*, 30(9), 1291–1305. <https://doi.org/10.1101/gr.263566.120>
- Ohno, S. (2013). *Evolution by gene duplication*. Springer Science & Business Media.
- Owen, G. I., & Zelent, A. (2000). Origins and evolutionary diversification of the nuclear receptor superfamily. *Cellular and Molecular Life Sciences*, 57(5), 809–827. <https://doi.org/10.1007/s000180050043>
- Oziolor, E. M., Reid, N. M., Yair, S., Lee, K. M., VerPloeg, S. G., Bruns, P. C., Shaw, J. R., Whitehead, A., & Matson, C. W. (2019). Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science*, 364(6439), 455–457. <https://doi.org/10.1126/science.aav4155>
- Pachura-Bouchet, S., Blaise, C., & Vasseur, P. (2006). Toxicity of nonylphenol on the cnidarian *Hydra attenuata* and environmental risk assessment. *Environmental Toxicology*, 21(4), 388–394. <https://doi.org/10.1002/TOX.20201>
- Paez, S., Kraus, R. H. S., Shapiro, B., Gilbert, M. T. P., & Jarvis, E. D. (2022). Reference genomes for conservation. *Science*, 377(6604), 364–366. <https://doi.org/10.1126/science.abm8127>
- Papalou, O., Kandaraki, E. A., Papadakis, G., & Diamanti-Kandarakis, E. (2019). Endocrine disrupting chemicals: An occult mediator of metabolic disease. *Frontiers in Endocrinology*, 10(MAR), 429401. <https://doi.org/10.3389/FENDO.2019.00112/XML/NLM>
- Paps, J., Xu, F., Zhang, G., & Holland, P. W. H. (2015). Reinforcing the Egg-Timer: Recruitment of Novel Lophotrochozoa Homeobox Genes to Early and Late Development in the Pacific Oyster. *Genome Biology and Evolution*, 7(3), 677–688. <https://doi.org/10.1093/GBE/EVV018>

- Parente, T. E., Moreira, D. A., Magalhães, M. G. P., de Andrade, P. C. C., Furtado, C., Haas, B. J., Stegeman, J. J., & Hahn, M. E. (2017). The liver transcriptome of suckermouth armoured catfish (*Pterygoplichthys anisitsi*, Loricariidae): Identification of expansions in defensome gene families. *Marine Pollution Bulletin*, 115(1–2), 352–361. <https://doi.org/10.1016/J.MARPOLBUL.2016.12.012>
- Parfrey, L. W., Lahr, D. J. G., Knoll, A. H., & Katz, L. A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences of the United States of America*, 108(33), 13624–13629. <https://doi.org/10.1073/pnas.1110633108>
- Park, Y. C., Lee, S., & Cho, M. H. (2014). The simplest flowchart stating the mechanisms for organic xenobiotics-induced toxicity: Can it possibly be accepted as a “central dogma” for toxic mechanisms? *Toxicological Research*, 30(3), 179–184. <https://doi.org/10.5487/TR.2014.30.3.179>
- Patterson, A. D., Gonzalez, F. J., & Idle, J. R. (2010). Xenobiotic metabolism: A view through the metabolometer. In *Chemical Research in Toxicology* (Vol. 23, Issue 5, pp. 851–860). American Chemical Society. <https://doi.org/10.1021/tx100020p>
- Paul, C. E., Eggerichs, D., Westphal, A. H., Tischler, D., & van Berkel, W. J. H. (2021). Flavoprotein monooxygenases: Versatile biocatalysts. *Biotechnology Advances*, 51, 107712. <https://doi.org/10.1016/J.BIOTECHADV.2021.107712>
- Peet, D. J., Turley, S. D., Ma, W., Janowski, B. A., Lobaccaro, J. M. A., Hammer, R. E., & Mangelsdorf, D. J. (1998). Cholesterol and bile acid metabolism are impaired in mice lacking the nuclear oxysterol receptor LXR alpha. *Cell*, 93(5), 693–704. [https://doi.org/10.1016/S0092-8674\(00\)81432-4](https://doi.org/10.1016/S0092-8674(00)81432-4)
- Pehr, K., Love, G. D., Kuznetsov, A., Podkovyrov, V., Junium, C. K., Shumlyanskyy, L., Sokur, T., & Bekker, A. (2018). Ediacara biota flourished in oligotrophic and bacterially dominated marine environments across Baltica. *Nature Communications* 2018 9:1, 9(1), 1–10. <https://doi.org/10.1038/s41467-018-04195-8>
- Pelis, R. M., & Wright, S. H. (2014). SLC22, SLC44, and SLC47 Transporters—Organic Anion and Cation Transporters: Molecular and Cellular Properties. *Current Topics in Membranes*, 73, 233–261. <https://doi.org/10.1016/B978-0-12-800223-0.00006-2>
- Perillo, M., Oulhen, N., Foster, S., Spurrell, M., Calestani, C., & Wessel, G. (2020). Regulation of dynamic pigment cell states at single-cell resolution. *ELife*, 9, 1–25. <https://doi.org/10.7554/ELIFE.60388>

- Peterson, K. J., Lyons, J. B., Nowak, K. S., Takacs, C. M., Wargo, M. J., Mcpeek, M. A., King, N., Carroll, S., Arnone, I., Zoologica, S., Dohrn, A., Hinman, V., Davidson, E., & Budd, G. (2004). Estimating metazoan divergence times with a molecular clock. *Proceedings of the National Academy of Sciences*, 101(17), 6536–6541. <https://doi.org/10.1073/PNAS.0401670101>
- Phillips, I. R., & Shephard, E. A. (2008). Flavin-containing monooxygenases: mutations, disease and drug response. *Trends in Pharmacological Sciences*, 29(6), 294–301. <https://doi.org/10.1016/j.tips.2008.03.004>
- Phillips, I. R., & Shephard, E. A. (2017). Drug metabolism by flavin-containing monooxygenases of human and mouse. *Expert Opinion on Drug Metabolism & Toxicology*, 13(2), 167–181. <https://doi.org/10.1080/17425255.2017.1239718>
- Phillips, I. R., & Shephard, E. A. (2019). Endogenous Roles of Mammalian Flavin-Containing Monooxygenases. *Catalysts* 2019, Vol. 9, Page 1001, 9(12), 1001. <https://doi.org/10.3390/CATAL9121001>
- Phillips, I. R., Veeravalli, S., Khadayate, S., & Shephard, E. A. (2023). Metabolomic and transcriptomic analyses of Fmo5^{-/-} mice reveal roles for flavin-containing monooxygenase 5 (FMO5) in NRF2-mediated oxidative stress response, unfolded protein response, lipid homeostasis, and carbohydrate and one-carbon metabolism. *PLOS ONE*, 18(6), e0286692. <https://doi.org/10.1371/JOURNAL.PONE.0286692>
- Polinski, J. M., Zimin, A. V., Clark, K. F., Kohn, A. B., Sadowski, N., Timp, W., Ptitsyn, A., Khanna, P., Romanova, D. Y., Williams, P., Greenwood, S. J., Moroz, L. L., Walt, D. R., & Bodnar, A. G. (2021). The American lobster genome reveals insights on longevity, neural, and immune adaptations. *Science Advances*, 7(26), 8290–8313. <https://doi.org/10.1126/sciadv.abe8290>
- Ponder, W. F., Lindberg, D. R., & Ponder, J. M. (2019). Biology and Evolution of the Mollusca, Volume 1. In *Biology and Evolution of the Mollusca* (1st Edition). CRC Press. <https://doi.org/10.1201/9781351115667>
- Poulsen, L. L., & Ziegler, D. M. (1979). The liver microsomal FAD-containing monooxygenase. Spectral characterization and kinetic studies. *Journal of Biological Chemistry*, 254(14), 6449–6455. [https://doi.org/10.1016/s0021-9258\(18\)50388-4](https://doi.org/10.1016/s0021-9258(18)50388-4)
- Poulsen, L. L., & Ziegler, D. M. (1995). Multisubstrate flavin-containing monooxygenases: applications of mechanism to specificity. *Chemico-Biological Interactions*, 96(1), 57–73. [https://doi.org/10.1016/0009-2797\(94\)03583-T](https://doi.org/10.1016/0009-2797(94)03583-T)

- Previšić, A., Vilenica, M., Vučković, N., Petrović, M., & Rožman, M. (2021). Aquatic Insects Transfer Pharmaceuticals and Endocrine Disruptors from Aquatic to Terrestrial Ecosystems. *Environmental Science and Technology*, 55(6), 3736–3746. <https://doi.org/10.1021/acs.est.0c07609>
- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., & Lopez, R. (2005). InterProScan: Protein domains identifier. *Nucleic Acids Research*, 33(SUPPL. 2), W116–W120. <https://doi.org/10.1093/nar/gki442>
- Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, 35(9), 833–844. <https://doi.org/10.1038/NBT.3935>
- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, 47(W1), W191–W198. <https://doi.org/10.1093/NAR/GKZ369>
- Regnault, C., Usal, M., Veyrenc, S., Couturier, K., Batandier, C., Bulteau, A. L., Lejon, D., Sapin, A., Combourieu, B., Chetiveaux, M., Le May, C., Lafond, T., Raveton, M., & Reynaud, S. (2018). Unexpected metabolic disorders induced by endocrine disruptors in *Xenopus tropicalis* provide new lead for understanding amphibian decline. *Proceedings of the National Academy of Sciences of the United States of America*, 115(19), E4416–E4425. <https://doi.org/10.1073/pnas.1721267115>
- Reid, N. M., & Whitehead, A. (2016). Functional genomics to assess biological responses to marine pollution at physiological and evolutionary timescales: Toward a vision of predictive ecotoxicology. *Briefings in Functional Genomics*, 15(5), 358–364. <https://doi.org/10.1093/bfpg/elv060>
- Reif, W. E. (2002). Evolution of organ systems: Phylogeny, function and reciprocal illumination. *Senckenbergiana Lethaea*, 82(1), 356–366. <https://doi.org/10.1007/BF03043794>
- Reitzel, A. M., Sullivan, J. C., Traylor-Knowles, N., & Finnerty, J. R. (2008). Genomic survey of candidate stress-response genes in the estuarine anemone *Nematostella vectensis*. *The Biological Bulletin*, 214(3), 233–254. <https://doi.org/10.2307/25470666>
- Reitzel, A. M., & Tarrant, A. M. (2009). Nuclear receptor complement of the cnidarian *Nematostella vectensis*: Phylogenetic relationships and developmental expression patterns. *BMC Evolutionary Biology*, 9(1), 1–13. <https://doi.org/10.1186/1471-2148-9-230>

- Rendic, S., & Guengerich, F. P. (2015). Survey of human oxidoreductases and cytochrome P450 enzymes involved in the metabolism of xenobiotic and natural chemicals. *Chemical Research in Toxicology*, 28(1), 38–42. <https://doi.org/10.1021/tx500444e>
- Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., ... Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, 592(7856), 737–746. <https://doi.org/10.1038/s41586-021-03451-0>
- Rieger, P. G., Meier, H. M., Gerle, M., Vogt, U., Groth, T., & Knackmuss, H. J. (2002). Xenobiotics in the environment: Present and future strategies to obviate the problem of biological persistence. *Journal of Biotechnology*, 94(1), 101–123. [https://doi.org/10.1016/S0168-1656\(01\)00422-9](https://doi.org/10.1016/S0168-1656(01)00422-9)
- Rinn, J. L., & Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annual Review of Biochemistry*, 81(Volume 81, 2012), 145–166. <https://doi.org/10.1146/annurev-biochem-051410-092902>
- Robaire, B., Delbes, G., Head, J. A., Marlatt, V. L., Martyniuk, C. J., Reynaud, S., Trudeau, V. L., & Mennigen, J. A. (2022). A cross-species comparative approach to assessing multi- and transgenerational effects of endocrine disrupting chemicals. *Environmental Research*, 204, 112063. <https://doi.org/10.1016/J.ENVRES.2021.112063>
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., Mungall, K., Lee, S., Okada, H. M., Qian, J. Q., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y. S., Newsome, R., Chan, S. K., She, R., Varhol, R., ... Birol, I. (2010). De novo assembly and analysis of RNA-seq data. *Nature Methods*, 7(11), 909–912. <https://doi.org/10.1038/nmeth.1517>
- Robinson-Rechavi, M., Carpentier, A. S., Duffrais, M., & Laudet, V. (2001). How many nuclear hormone receptors are there in the human genome? *Trends in Genetics*, 17(10), 554–556. [https://doi.org/10.1016/S0168-9525\(01\)02417-9](https://doi.org/10.1016/S0168-9525(01)02417-9)
- Roncalli, V., Ascione, D., Lauritano, C., & Carotenuto, Y. (2025). The Chemical Defensome: A Survey of Environmental Sensing and Response Genes in Copepods. *International Journal of Molecular Sciences*, 26(4), 1546. <https://doi.org/10.3390/IJMS26041546/S1>
- Ross, P. M., Scanes, E., Byrne, M., Ainsworth, T. D., Donelson, J. M., Foo, S. A., Hutchings, P., Thiyagarajan, V., & Parker, L. M. (2023). SURVIVING THE

ANTHROPOCENE: THE RESILIENCE OF MARINE ANIMALS TO CLIMATE CHANGE. *Oceanography and Marine Biology: An Annual Review, Volume 61*, 61, 35–80. <https://doi.org/10.1201/9781003363873-3>

Rota-Stabelli, O., Kayal, E., Gleeson, D., Daub, J., Boore, J. L., Telford, M. J., Pisani, D., Blaxter, M., & Lavrov, D. V. (2010). Ecdysozoan Mitogenomics: Evidence for a Common Origin of the Legged Invertebrates, the Panarthropoda. *Genome Biology and Evolution*, 2(1), 425–440. <https://doi.org/10.1093/GBE/EVQ030>

Rowland, A., Miners, J. O., & Mackenzie, P. I. (2013). The UDP-glucuronosyltransferases: Their role in drug metabolism and detoxification. *The International Journal of Biochemistry & Cell Biology*, 45(6), 1121–1132. <https://doi.org/10.1016/J.BIOCEL.2013.02.019>

Ruan, J., & Li, H. (2019). Fast and accurate long-read assembly with wtdbg2. *Nature Methods* 17:2, 17(2), 155–158. <https://doi.org/10.1038/s41592-019-0669-3>

Ruggiero, M. A., Gordon, D. P., Orrell, T. M., Bailly, N., Bourgoin, T., Brusca, R. C., Cavalier-Smith, T., Guiry, M. D., & Kirk, P. M. (2015). A Higher Level Classification of All Living Organisms. *PLOS ONE*, 10(4), e0119248. <https://doi.org/10.1371/JOURNAL.PONE.0119248>

Ruiz-Trillo, I., Baguñà, J., Riutort, M., Lf, U., Ondelius, J., Ñaki, I., Uiz -T Rillo, R., Aume, J., Aguiñà, B., Arta, M., & Iutort, R. (2002). The Nemertodermatida are basal bilaterians and not members of the Platyhelminthes. *Zoologica Scripta*, 31(2), 201–215. <https://doi.org/10.1046/J.1463-6409.2002.00090.X>

Ruiz-Trillo, I., Paps, J., Loukota, M., Ribera, C., Jondelius, U., Baguñà, J., & Riutort, M. (2002). A phylogenetic analysis of myosin heavy chain type II sequences corroborates that Acoela and Nemertodermatida are basal bilaterians. *Proceedings of the National Academy of Sciences of the United States of America*, 99(17), 11246–11251. <https://doi.org/10.1073/pnas.172390199>

Saaristo, M., Brodin, T., Balshine, S., Bertram, M. G., Brooks, B. W., Ehlman, S. M., McCallum, E. S., Sih, A., Sundin, J., Wong, B. B. M., & Arnold, K. E. (2018). Direct and indirect effects of chemical contaminants on the behaviour, ecology and evolution of wildlife. *Proceedings of the Royal Society B: Biological Sciences*, 285(1885), 20181297. <https://doi.org/10.1098/RSPB.2018.1297>

Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94(3), 441–448. [https://doi.org/10.1016/0022-2836\(75\)90213-2](https://doi.org/10.1016/0022-2836(75)90213-2)

- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463–5467. <https://doi.org/10.1073/PNAS.74.12.5463>
- Santos, M. M., Ten Hallers-Tjabbes, C. C., Santos, A. M., & Vieira, N. (2002). Imposex in *Nucella lapillus*, a bioindicator for TBT contamination: re-survey along the Portuguese coast to monitor the effectiveness of EU regulation. *Journal of Sea Research*, 48(3), 217–223. [https://doi.org/10.1016/S1385-1101\(02\)00166-1](https://doi.org/10.1016/S1385-1101(02)00166-1)
- Satoh, N., Rokhsar, D., & Nishikawa, T. (2014). Chordate evolution and the three-phylum system. *Proceedings of the Royal Society B: Biological Sciences*, 281(1794). <https://doi.org/10.1098/RSPB.2014.1729>
- Satoh, N., Tagawa, K., & Takahashi, H. (2012). How was the notochord born? *Evolution & Development*, 14(1), 56–75. <https://doi.org/10.1111/J.1525-142X.2011.00522.X>
- Scharf, M. E., Scharf, D. W., Bennett, G. W., & Pittendrigh, B. R. (2004). Catalytic activity and expression of two flavin-containing monooxygenases from *Drosophila melanogaster*. *Archives of Insect Biochemistry and Physiology*, 57(1), 28–39. <https://doi.org/10.1002/ARCH.20012>
- Schierwater, B., Eitel, M., Jakob, W., Osigus, H. J., Hadrys, H., Dellaporta, S. L., Kolokotronis, S. O., & DeSalle, R. (2009). Concatenated Analysis Sheds Light on Early Metazoan Evolution and Fuels a Modern “Urmetazoon” Hypothesis. *PLOS Biology*, 7(1), e1000020. <https://doi.org/10.1371/JOURNAL.PBIO.1000020>
- Schlenk, D. (1998). Occurrence of flavin-containing monooxygenases in non-mammalian eukaryotic organisms. *Comparative Biochemistry and Physiology. Part C, Pharmacology, Toxicology & Endocrinology*, 121(1–3), 185–195. [https://doi.org/10.1016/S0742-8413\(98\)10060-9](https://doi.org/10.1016/S0742-8413(98)10060-9)
- Schlenk, D., & Li-Schlenk, R. (1994). Characterization of liver flavin-containing monooxygenase of the dogfish shark (*Squalus acanthias*) and partial purification of liver flavin-containing monooxygenase of the silky shark (*Carcharhinus falciformis*). *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, 109(4), 655–664. [https://doi.org/10.1016/0305-0491\(94\)90128-7](https://doi.org/10.1016/0305-0491(94)90128-7)
- Seaver, E. C. (2003). Segmentation: mono- or polyphyletic? *The International Journal of Developmental Biology*, 47(7–8), 583–595. <https://doi.org/10.1387/IJDB.14756334>
- Sebastiano, M., Bustamante, P., Costantini, D., Eulaers, I., Malarvannan, G., Mendez-Fernandez, P., Churlaud, C., Blévin, P., Hauselmann, A., Dell’Omo, G., Covaci, A., Eens, M., & Chastel, O. (2016). High levels of mercury and low levels of persistent organic

pollutants in a tropical seabird in French Guiana, the Magnificent frigatebird, *Fregata magnificens*. *Environmental Pollution*, 214, 384–393. <https://doi.org/10.1016/J.ENVPOL.2016.03.070>

Sebé-Pedrós, A., Degnan, B. M., & Ruiz-Trillo, I. (2017). The origin of Metazoa: a unicellular perspective. *Nature Reviews Genetics* 2017 18:8, 18(8), 498–512. <https://doi.org/10.1038/nrg.2017.21>

Sedlazeck, F. J., Lee, H., Darby, C. A., & Schatz, M. C. (2018). Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nature Reviews Genetics* 2018 19:6, 19(6), 329–346. <https://doi.org/10.1038/s41576-018-0003-4>

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G. P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Foote, A. D., Franchini, P., Jiggins, C. D., ... Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics* 2014 15:3, 15(3), 176–192. <https://doi.org/10.1038/nrg3644>

Sehlmeyer, S., Wang, L., Langel, D., Heckel, D. G., Mohagheghi, H., Petschenka, G., & Ober, D. (2010). Flavin-Dependent Monooxygenases as a Detoxification Mechanism in Insects: New Insights from the Arctiids (Lepidoptera). *PLOS ONE*, 5(5), e10435. <https://doi.org/10.1371/JOURNAL.PONE.0010435>

Seidegård, J., & Ekström, G. (1997). The role of human glutathione transferases and epoxide hydrolases in the metabolism of xenobiotics. *Environmental Health Perspectives*, 105(SUPPL. 4), 791–799. <https://doi.org/10.1289/EHP.105-1470052>

Shan, J., Sun, X., Li, R., Zhu, B., Liang, P., & Gao, X. (2021). Identification of ABCG transporter genes associated with chlorantraniliprole resistance in *Plutella xylostella* (L.). *Pest Management Science*, 77(7), 3491–3499. <https://doi.org/10.1002/PS.6402>

Shapiro, J. A., & Von Sternberg, R. (2005). Why repetitive DNA is essential to genome function. *Biological Reviews*, 80(2), 227–250. <https://doi.org/10.1017/S1464793104006657>

Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2017). DNA sequencing at 40: past, present and future. *Nature* 2017 550:7676, 550(7676), 345–353. <https://doi.org/10.1038/nature24286>

Shih, D. M., Wang, Z., Lee, R., Meng, Y., Che, N., Charugundla, S., Qi, H., Wu, J., Pan, C., Mark Brown, J., Vallim, T., Bennett, B. J., Graham, M., Hazen, S. L., & Lusi, A. J. (2015). Flavin containing monooxygenase 3 exerts broad effects on glucose and lipid

metabolism and atherosclerosis. *Journal of Lipid Research*, 56(1), 22–37. <https://doi.org/10.1194/jlr.M051680>

Shih, J. C., Chen, K., & Ridd, M. J. (1999). Monoamine oxidase: From genes to behavior. *Annual Review of Neuroscience*, 22(Volume 22, 1999), 197–217. <https://doi.org/10.1146/annurev.neuro.22.1.197>

Shimada, T., & Fujii-Kuriyama, Y. (2004). Metabolic activation of polycyclic aromatic hydrocarbons to carcinogens by cytochromes P450 1A1 and 1B1. *Cancer Science*, 95(1), 1–6. <https://doi.org/10.1111/J.1349-7006.2004.TB03162.X>

Shinzato, C., Hamada, M., Shoguchi, E., Kawashima, T., & Satoh, N. (2012). The repertoire of chemical defense genes in the coral *Acropora digitifera* genome. *Zoological Science*, 29(8). <https://doi.org/10.2108/zsj.29.510>

Silva, D., Araújo, C., Marassi, R., Cardoso-Silva, S., Neto, M., Silva, G., Ribeiro, R., Silva, F., Paiva, T., & Pompêo, M. (2018). Influence of interspecific interactions on avoidance response to contamination. *Science of The Total Environment*, 642, 824–831. <https://doi.org/10.1016/J.SCITOTENV.2018.06.127>

Silva, A. P. A. da, Oliveira, C. D. L. de, Quirino, A. M. S., Silva, F. D. M. da, Saraiva, R. de A., & Silva-Cavalcanti, J. S. (2018). Endocrine Disruptors in Aquatic Environment: Effects and Consequences on the Biodiversity of Fish and Amphibian Species. *Aquatic Science and Technology*, 6(1), 35–51. <https://doi.org/10.5296/AST.V6I1.12565>

Sim, E., Walters, K., & Boukouvala, S. (2008). Arylamine N-acetyltransferases: From Structure to Function. *Drug Metabolism Reviews*, 40(3), 479–510. <https://doi.org/10.1080/03602530802186603>

Simakov, O., Kawashima, T., Marlétaz, F., Jenkins, J., Koyanagi, R., Mitros, T., Hisata, K., Bredeson, J., Shoguchi, E., Gyoja, F., Yue, J. X., Chen, Y. C., Freeman, R. M., Sasaki, A., Hikosaka-Katayama, T., Sato, A., Fujie, M., Baughman, K. W., Levine, J., ... Gerhart, J. (2015). Hemichordate genomes and deuterostome origins. *Nature* 2015 527:7579, 527(7579), 459–465. <https://doi.org/10.1038/nature16150>

Sluder, A. E., Mathews, S. W., Hough, D., Yin, V. P., & Maina, C. V. (1999). The Nuclear Receptor Superfamily Has Undergone Extensive Proliferation and Diversification in Nematodes. *Genome Research*, 9(2), 103–120. <https://doi.org/10.1101/GR.9.2.103>

Smith, E., Nelson, L., Strange, M., Eyster, A., Rowland, S., Schrag, D., & Macdonald, F. (2016). The end of the Ediacaran: Two new exceptionally preserved body fossil assemblages from Mount Dunfee, Nevada, USA. *Geology*, 44(11), 911–914. <https://doi.org/10.1130/G38157.1>

Smith, M., Warmolts, D., Thoney, D., & Hueter, R. (2004). The elasmobranch husbandry manual: captive care of sharks, rays and their relatives. *Special Publication of the Ohio Biological Survey*, 589(145), 4.

Snell, T. W., Brogdon, S. E., & Morgan, M. B. (2003). Gene Expression Profiling in Ecotoxicology. *Ecotoxicology*, 12(6), 475–483. <https://doi.org/10.1023/B:ECTX.0000003033.09923.a8>

Sohn, J. II, & Nam, J. W. (2018). The present and future of de novo whole-genome assembly. *Briefings in Bioinformatics*, 19(1), 23–40. <https://doi.org/10.1093/BIB/BBW096>

Spjuth, O., Bongcam-Rudloff, E., Hernández, G. C., Forer, L., Giovacchini, M., Guimera, R. V., Kallio, A., Korpelainen, E., Kañdula, M. M., Krachunov, M., Kreil, D. P., Kulev, O., Łabaj, P. P., Lampa, S., Pireddu, L., Schönherr, S., Siretskiy, A., & Vassilev, D. (2015). Experiences with workflows for automating data-intensive bioinformatics. *Biology Direct*, 10(1), 1–12. <https://doi.org/10.1186/s13062-015-0071-8>

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research*, 34(suppl_2), W435–W439. <https://doi.org/10.1093/NAR/GKL200>

Stanke, M., Steinkamp, R., Waack, S., & Morgenstern, B. (2004). AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Research*, 32(WEB SERVER ISS.), W309–W312. <https://doi.org/10.1093/nar/gkh379>

Stegeman, J. J., Goldstone, J. V., & Hahn, M. E. (2010). Perspectives on zebrafish as a model in environmental toxicology. *Fish Physiology*, 29(C), 367–439. [https://doi.org/10.1016/S1546-5098\(10\)02910-9](https://doi.org/10.1016/S1546-5098(10)02910-9)

Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology* 2017 35:11, 35(11), 1026–1028. <https://doi.org/10.1038/nbt.3988>

Steinmetz, P. R. H., Kostyuchenko, R. P., Fischer, A., & Arendt, D. (2011). The segmental pattern of otx, gbx, and Hox genes in the annelid *Platynereis dumerilii*. *Evolution & Development*, 13(1), 72–79. <https://doi.org/10.1111/J.1525-142X.2010.00457.X>

Stephan, T., Burgess, S. M., Cheng, H., Danko, C. G., Gill, C. A., Jarvis, E. D., Koepfli, K. P., Koltes, J. E., Lyons, E., Ronald, P., Ryder, O. A., Schriml, L. M., Soltis, P., VandeWoude, S., Zhou, H., Ostrander, E. A., & Karlsson, E. K. (2022). Darwinian genomics and diversity in the tree of life. *Proceedings of the National Academy of Sciences of the United States of America*, 119(4), e2115644119. <https://doi.org/10.1073/pnas.2115644119>

- Straus, D. S., & Glass, C. K. (2007). Anti-inflammatory actions of PPAR ligands: new insights on cellular and molecular mechanisms. *Trends in Immunology*, 28(12), 551–558. <https://doi.org/10.1016/j.it.2007.09.003>
- Strobel, A., Lille-Langøy, R., Segner, H., Burkhardt-Holm, P., Goksøyr, A., & Karlsen, O. A. (2022). Xenobiotic metabolism and its physiological consequences in high-Antarctic Notothenioid fishes. *Polar Biology*, 45(2), 345–358. <https://doi.org/10.1007/s00300-021-02992-4>
- Sullivan, A. P., Bird, D. W., & Perry, G. H. (2017). Human behaviour as a long-term ecological driver of non-human evolution. *Nature Ecology & Evolution* 2017 1:3, 1(3), 1–11. <https://doi.org/10.1038/s41559-016-0065>
- Sumpter, J. P. (2005). Endocrine Disrupters in the Aquatic Environment: An Overview. *Acta Hydrochimica et Hydrobiologica*, 33(1), 9–16. <https://doi.org/10.1002/AHEH.200400555>
- Surapureddi, S., Rana, R., & Goldstein, J. A. (2011). NCOA6 differentially regulates the expression of the CYP2C9 and CYP3A4 genes. *Pharmacological Research : The Official Journal of the Italian Pharmacological Society*, 63(5), 405. <https://doi.org/10.1016/J.PHRS.2011.01.013>
- Surapureddi, S., Rana, R., Reddy, J. K., & Goldstein, J. A. (2008). Nuclear receptor coactivator 6 mediates the synergistic activation of human cytochrome P-450 2C9 by the constitutive androstane receptor and hepatic nuclear factor-4alpha. *Molecular Pharmacology*, 74(3), 913–923. <https://doi.org/10.1124/MOL.108.048983>
- Tan, H. M., & Low, W. Y. (2018). Rapid birth-death evolution and positive selection in detoxification-type glutathione S-transferases in mammals. *PLoS ONE*, 13(12), e0209336. <https://doi.org/10.1371/JOURNAL.PONE.0209336>
- Taylor, J. S., Braasch, I., Frickey, T., Meyer, A., & Van de Peer, Y. (2003). Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Research*, 13(3), 382–390. <https://doi.org/10.1101/GR.640303>
- Testa, B., & Krämer, S. D. (2007a). The biochemistry of drug metabolism—an introduction: Part 2. Redox reactions and their enzymes. *Chemistry & Biodiversity*, 4(3), 257–405. <https://doi.org/10.1002/cbdv.200790032>
- Testa, B., & Krämer, S. D. (2007b). The biochemistry of drug metabolism—an introduction: part 3. Reactions of hydrolysis and their enzymes. *Chemistry & Biodiversity*, 4(9), 2031–2122. <https://doi.org/10.1002/cbdv.200790169>

- Thomas, P. D., Kejariwal, A., Campbell, M. J., Mi, H., Diemer, K., Guo, N., Ladunga, I., Ulitsky-Lazareva, B., Muruganujan, A., Rabkin, S., Vandergriff, J. A., & Doremieux, O. (2003). PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Research*, 31(1), 334. <https://doi.org/10.1093/NAR/GKG115>
- Thompson, R. P., & Walsh, D. (2014). Evolutionary Biology: Conceptual, Ethical, and Religious Issues. *Evolutionary Biology: Conceptual, Ethical, and Religious Issues*, 1–243. <https://doi.org/10.1017/CBO9781139208796>
- Thorp, J. H., Lovell, L. L., Bain, B. A., Brinkhurst, R. O., Gelder, S. R., Moser, W. E., Govedich, F. R., Ocegüera-Figueroa, A., Richardson, D. J., & Phillips, A. J. (2016). Phylum Annelida. *Thorp and Covich's Freshwater Invertebrates: Keys to Nearctic Fauna*, 223–263. <https://doi.org/10.1016/B978-0-12-385028-7.00012-3>
- Tonelli, C., Chio, I. I. C., & Tuveson, D. A. (2018). Transcriptional Regulation by Nrf2. *Antioxidants and Redox Signaling*, 29(17), 1727–1745. <https://doi.org/10.1089/ars.2017.7342>
- Townsend, D. M., & Tew, K. D. (2003). The role of glutathione-S-transferase in anti-cancer drug resistance. *Oncogene* 2003 22:47, 22(47), 7369–7375. <https://doi.org/10.1038/sj.onc.1206940>
- Tukey, R. H., & Strassburg, C. P. (2000). Human UDP-glucuronosyltransferases: Metabolism, expression, and disease. *Annual Review of Pharmacology and Toxicology*, 40(Volume 40, 2000), 581–616. <https://doi.org/10.1146/annurev.pharmtox.40.1.581>
- Tyler, C. R., Jobling, S., & Sumpter, J. P. (1998). Endocrine disruption in wildlife: A critical review of the evidence. In *Critical Reviews in Toxicology* (Vol. 28, Issue 4, pp. 319–361). Informa Healthcare. <https://doi.org/10.1080/10408449891344236>
- van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends in Genetics*, 34(9), 666–681. <https://doi.org/10.1016/j.tig.2018.05.008>
- van Soest, R. W. M., Boury-Esnault, N., Vacelet, J., Dohrmann, M., Erpenbeck, D., de Voogd, N. J., Santodomingo, N., Vanhoorne, B., Kelly, M., & Hooper, J. N. A. (2012). Global Diversity of Sponges (Porifera). *PLOS ONE*, 7(4), e35105. <https://doi.org/10.1371/JOURNAL.PONE.0035105>
- Vannier, J., Farke, A. A., & Alf, R. M. (2012). Gut Contents as Direct Indicators for Trophic Relationships in the Cambrian Marine Ecosystem. *PLOS ONE*, 7(12), e52200. <https://doi.org/10.1371/JOURNAL.PONE.0052200>

- Vasiliou, V., Pappa, A., & Estey, T. (2004). Role of human aldehyde dehydrogenases in endobiotic and xenobiotic metabolism. *Drug Metabolism Reviews*, 36(2), 279–299. <https://doi.org/10.1081/dmr-120034001>
- Vasiliou, V., Pappa, A., & Petersen, D. R. (2000). Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism. *Chemico-Biological Interactions*, 129(1–2), 1–19. [https://doi.org/10.1016/S0009-2797\(00\)00211-8](https://doi.org/10.1016/S0009-2797(00)00211-8)
- Veeravalli, S., Karu, K., Scott, F., Fennema, D., Phillips, I. R., & Shephard, E. A. (2018). Effect of flavin-containing monooxygenase genotype, mouse strain, and gender on trimethylamine N-oxide production, plasma cholesterol concentration, and an index of atherosclerosis. *Drug Metabolism and Disposition*, 46(1), 20–25. <https://doi.org/10.1124/dmd.117.077636>
- Velasco, E. M., del Arbol, J., Baro, J., & Sobrino, I. (2011). Age and growth of the Spanish chub mackerel *Scomber colias* off southern Spain: a comparison between samples from the NE Atlantic and the SW Mediterranean. *Revista de Biología Marina y Oceanografía*, 46(1), 27–34.
- Walker, D. M., & Gore, A. C. (2011). Transgenerational neuroendocrine disruption of reproduction. *Nature Reviews Endocrinology* 2011 7:4, 7(4), 197–207. <https://doi.org/10.1038/nrendo.2010.215>
- Wallace, R. L., Ricci, C., & Melone, G. (1996). A Cladistic Analysis of Pseudocoelomate (Aschelminth) Morphology. *Invertebrate Biology*, 115(2), 104. <https://doi.org/10.2307/3227041>
- Wang, H., Chen, J., Hollister, K., Sowers, L. C., & Forman, B. M. (1999). Endogenous bile acids are ligands for the nuclear receptor FXR/BAR. *Molecular Cell*, 3(5), 543–553. [https://doi.org/10.1016/s1097-2765\(00\)80348-2](https://doi.org/10.1016/s1097-2765(00)80348-2)
- Wang, H., Faucette, S., Sueyoshi, T., Moore, R., Ferguson, S., Negishi, M., & LeCluyse, E. L. (2003). A novel distal enhancer module regulated by pregnane X receptor/constitutive androstane receptor is essential for the maximal induction of CYP2B6 gene expression. *Journal of Biological Chemistry*, 278(16), 14146–14152. <https://doi.org/10.1074/jbc.M212482200>
- Wang, J., Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H., Xiong, Z., Que, H., Xie, Y., Holland, P. W. H., Paps, J., Zhu, Y., Wu, F., Chen, Y., ... Wang, J. (2012). The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 2012 490:7418, 490(7418), 49–54. <https://doi.org/10.1038/nature11413>

- Warner, K. D., Hajdin, C. E., & Weeks, K. M. (2018). Principles for targeting RNA with drug-like small molecules. *Nature Reviews Drug Discovery* 2018 17:8, 17(8), 547–558. <https://doi.org/10.1038/nrd.2018.93>
- Watson, J. D., & Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 1953 171:4356, 171(4356), 737–738. <https://doi.org/10.1038/171737a0>
- Webster, B. L., Copley, R. R., Jenner, R. A., Mackenzie-Dodds, J. A., Bourlat, S. J., Rota-Stabelli, O., Littlewood, D. T. J., & Telford, M. J. (2006). Mitogenomics and phylogenomics reveal priapulid worms as extant models of the ancestral Ecdysozoan. *Evolution & Development*, 8(6), 502–510. <https://doi.org/10.1111/J.1525-142X.2006.00123.X>
- Weigmann, S. (2016). Annotated checklist of the living sharks, batoids and chimaeras (Chondrichthyes) of the world, with a focus on biogeographical diversity. *Journal of Fish Biology*, 88(3), 837–1037. <https://doi.org/10.1111/jfb.12874>
- Weiner, I. D., Mitch, W. E., & Sands, J. M. (2014). Urea and Ammonia Metabolism and the Control of Renal Nitrogen Excretion. *Clinical Journal of the American Society of Nephrology : CJASN*, 10(8), 1444. <https://doi.org/10.2215/CJN.10311013>
- Weinshilboum, R. M., Otterness, D. M., & Szumlanski, C. L. (1999). Methylation pharmacogenetics: Catechol O-methyltransferase, thiopurine methyltransferase, and histamine N-methyltransferase. *Annual Review of Pharmacology and Toxicology*, 39(Volume 39, 1999), 19–52. <https://doi.org/10.1146/annurev.pharmtox.39.1.19>
- Weinshilboum, R., & Wang, L. (2004). Pharmacogenomics: bench to bedside. *Nature Reviews Drug Discovery* 2004 3:9, 3(9), 739–748. <https://doi.org/10.1038/nrd1497>
- Wendelaar Bonga, S. E. (1997). The stress response in fish. *Physiological Reviews*, 77(3), 591–625. <https://doi.org/10.1152/PHYSREV.1997.77.3.591>
- Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P. C., Hall, R. J., Concepcion, G. T., Ebler, J., Functammasan, A., Kolesnikov, A., Olson, N. D., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C. S., Phillippy, A. M., Schatz, M. C., Myers, G., DePristo, M. A., ... Hunkapiller, M. W. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology* 2019 37:10, 37(10), 1155–1162. <https://doi.org/10.1038/s41587-019-0217-9>
- Wessel, G. M., Kiyomoto, M., Shen, T. L., & Yajima, M. (2020). Genetic manipulation of the pigment pathway in a sea urchin reveals distinct lineage commitment prior to metamorphosis in the bilateral to radial body plan transition. *Scientific Reports*, 10(1). <https://doi.org/10.1038/S41598-020-58584-5>

- Whitehead, A. (2014). *Evolutionary Genomics of Environmental Pollution*. 321–337. https://doi.org/10.1007/978-94-007-7347-9_16
- Whitehead, A., Clark, B. W., Reid, N. M., Hahn, M. E., & Nacci, D. (2017). When evolution is the solution to pollution: Key principles, and lessons from rapid repeated adaptation of killifish (*Fundulus heteroclitus*) populations. *Evolutionary Applications*, 10(8), 762–783. <https://doi.org/10.1111/EVA.12470>
- Wise, M., Silvia, M., Reyes, G., Dunn, R., Onorato, T. M., Pieplow, C., Furze, A., Hebert, E., Oulhen, N., Ritschoff, D., McClay, D. R., & Wessel, G. (2024). A molecular basis for spine color morphs in the sea urchin *Lytechinus variegatus*. *Scientific Reports*, 14(1), 28518. <https://doi.org/10.1038/S41598-024-79312-3>
- Won, E.-J., Han, J., Kim, D.-H., Dahms, H.-U., & Lee, J.-S. (2017). *Rotifers in Ecotoxicology*. 149–176. https://doi.org/10.1007/978-981-10-5635-2_10
- Wray, G. A., Levinton, J. S., & Shapiro, L. H. (1996). Molecular Evidence for Deep Precambrian Divergences Among Metazoan Phyla. *Science*, 274(5287), 568–581. <https://doi.org/10.1126/SCIENCE.274.5287.568>
- Wu, R., & Kaiser, A. D. (1968). Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *Journal of Molecular Biology*, 35(3), 523–537. [https://doi.org/10.1016/S0022-2836\(68\)80012-9](https://doi.org/10.1016/S0022-2836(68)80012-9)
- Wu, R., & Taylor, E. (1971). Nucleotide sequence analysis of DNA: II. Complete nucleotide sequence of the cohesive ends of bacteriophage λ DNA. *Journal of Molecular Biology*, 57(3), 491–511. [https://doi.org/10.1016/0022-2836\(71\)90105-7](https://doi.org/10.1016/0022-2836(71)90105-7)
- Wu, W., Niles, E. G., Hirai, H., & LoVerde, P. T. (2007). Evolution of a novel subfamily of nuclear receptors with members that each contain two DNA binding domains. *BMC Evolutionary Biology*, 7. <https://doi.org/10.1186/1471-2148-7-27>
- Xia, J., Niu, C., & Pei, X. (2010). Effects of chronic exposure to nonylphenol on locomotor activity and social behavior in zebrafish (*Danio rerio*). *Journal of Environmental Sciences (China)*, 22(9), 1435–1440. [https://doi.org/10.1016/S1001-0742\(09\)60272-2](https://doi.org/10.1016/S1001-0742(09)60272-2)
- Xiao, S., Muscente, A. D., Chen, L., Zhou, C., Schiffbauer, J. D., Wood, A. D., Polys, N. F., & Yuan, X. (2014). The Weng'an biota and the Ediacaran radiation of multicellular eukaryotes. *National Science Review*, 1(4), 498–520. <https://doi.org/10.1093/NSR/NWU061>
- Xin, F., Susiarjo, M., & Bartolomei, M. S. (2015). Multigenerational and transgenerational effects of endocrine disrupting chemicals: A role for altered epigenetic regulation?

Seminars in Cell & Developmental Biology, 43, 66–75.
<https://doi.org/10.1016/J.SEMCDB.2015.05.008>

Xu, C., Li, C., & Kong, A. (2005). Induction of phase I, II and III drug metabolism/transport by xenobiotics. *Archives of Pharmacal Research* 28:3, 28(3), 249–268.
<https://doi.org/10.1007/BF02977789>

Yadatie, F., Brun, N. R., Giebichenstein, J., Dmoch, K., Hylland, K., Borgå, K., Karlsen, O. A., & Goksøyr, A. (2022). Transcriptome responses in copepods *Calanus finmarchicus*, *Calanus glacialis* and *Calanus hyperboreus* exposed to phenanthrene and benzo[a]pyrene. *Marine Genomics*, 65, 100981.
<https://doi.org/10.1016/J.MARGEN.2022.100981>

Yadatie, F., Butcher, S., Førde, H. E., Campsteijn, C., Bouquet, J. M., Karlsen, O. A., Denoeud, F., Metpally, R., Thompson, E. M., Manak, J. R., Goksøyr, A., & Chourrout, D. (2012). Conservation and divergence of chemical defense system in the tunicate *Oikopleura dioica* revealed by genome wide response to two xenobiotics. *BMC Genomics*, 13(1), 1–17. <https://doi.org/10.1186/1471-2164-13-55/FIGURES/7>

Yagi, K., Satou, Y., Mazet, F., Shimeld, S. M., Degnan, B., Rokhsar, D., Levine, M., Kohara, Y., & Satoh, N. (2003). A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. III. Genes for Fox, ETS, nuclear receptors and NFκB. *Development Genes and Evolution*, 213(5–6), 235–244. <https://doi.org/10.1007/s00427-003-0317-9>

Yamamoto, A., Liu, M. Y., Kurogi, K., Sakakibara, Y., Saeki, Y., Suiko, M., & Liu, M. C. (2015). Sulphation of acetaminophen by the human cytosolic sulfotransferases: a systematic analysis. *Journal of Biochemistry*, 158(6), 497.
<https://doi.org/10.1093/JB/MVV062>

Yu, P. H., Wright, S., Fan, E. H., Lun, Z. R., & Gubisne-Harberle, D. (2003). Physiological and pathological implications of semicarbazide-sensitive amine oxidase. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1647(1–2), 193–199.
[https://doi.org/10.1016/S1570-9639\(03\)00101-8](https://doi.org/10.1016/S1570-9639(03)00101-8)

Zallen, D. T. (2003). Despite Franklin's work, Wilkins earned his Nobel. *Nature* 2003 425:6953, 425(6953), 15–15. <https://doi.org/10.1038/425015b>

Zanger, U. M., & Schwab, M. (2013). Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacology & Therapeutics*, 138(1), 103–141.
<https://doi.org/10.1016/J.PHARMTHERA.2012.12.007>

- Zhang, L., Wang, Q., Liu, W., Liu, F., Ji, A., & Li, Y. (2018). The Orphan Nuclear Receptor 4A1: A Potential New Therapeutic Target for Metabolic Diseases. *Journal of Diabetes Research*, 2018(1), 9363461. <https://doi.org/10.1155/2018/9363461>
- Zhang, T., Zhou, J., Gao, W., Jia, Y., Wei, Y., & Wang, G. (2022). Complex genome assembly based on long-read sequencing. *Briefings in Bioinformatics*, 23(5), 1–11. <https://doi.org/10.1093/BIB/BBAC305>
- Zhao, Y., & Bruemmer, D. (2010). NR4A orphan nuclear receptors: transcriptional regulators of gene expression in metabolism and vascular biology. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 30(8), 1535–1541. <https://doi.org/10.1161/ATVBAHA.109.191163>
- Zhao, Y., Zhang, K., Giesy, J. P., & Hu, J. (2015). Families of Nuclear Receptors in Vertebrate Models: Characteristic and Comparative Toxicological Perspective. *Scientific Reports 2015 5:1*, 5(1), 1–9. <https://doi.org/10.1038/srep08554>
- Zhou, C., Carotenuto, Y., Vitiello, V., Wu, C., Zhang, J., & Buttino, I. (2018). De novo transcriptome assembly and differential gene expression analysis of the calanoid copepod *Acartia tonsa* exposed to nickel nanoparticles. *Chemosphere*, 209, 163–172. <https://doi.org/10.1016/j.chemosphere.2018.06.096>
- Ziegler, D. M. (1988). Flavin-containing monooxygenases: Catalytic mechanism and substrate specificities. *Drug Metabolism Reviews*, 19(1), 1–32. <https://doi.org/10.3109/03602538809049617>
- Ziegler, D. M. (2002). An overview of the mechanism, substrate specificities, and structure of FMOs. *Drug Metabolism Reviews*, 34(3), 503–511. <https://doi.org/10.1081/DMR-120005650>
- Ziegler, D. M., Poulsen, L. L., & Mckee, E. M. (1971). Interaction of primary amines with a mixed-function amine oxidase isolated from pig liver microsomes. *Xenobiotica*, 1(4–5), 523–531. <https://doi.org/10.3109/00498257109041521>
- Żur, J., Piński, A., Marchlewicz, A., Hupert-Kocurek, K., Wojcieszynska, D., & Guzik, U. (2018). Organic micropollutants paracetamol and ibuprofen-toxicity, biodegradation, and genetic background of their utilization by bacteria. *Environmental Science and Pollution Research International*, 25(22), 21498–21524. <https://doi.org/10.1007/S11356-018-2517-X>

Attachments

Attachment 1: Authorization of use of PDF paper

This article is available in an open-source format under a CC BY 4.0 license.

Cite this article as...

André M. Machado, André Gomes-dos-Santos, Miguel M. Fonseca, Rute R. da Fonseca, Ana Veríssimo, Mónica Felício, Ricardo Capela, Nélon Alves, Miguel Santos, Filipe Salvador-Caramelo, Marcos Domingues, Raquel Ruivo, Elsa Froufe, L. Filipe C. Castro, A genome assembly of the Atlantic chub mackerel (*Scomber colias*): a valuable teleost fishing resource, *Gigabyte*, 2022 <https://doi.org/10.46471/gigabyte.40>

 Copy citation

Article Information

Journal title: Gigabyte

Publisher name: GigaScience Press

Publisher location: Sha Tin, New Territories, Hong Kong SAR

Subject Areas

Genetics and Genomics, Evolutionary Biology, Marine Biology

Received: 22 November 2021, **Accepted:** 09 February 2022, **Published:** 14 February 2022

Preprint submitted to: <https://doi.org/10.1101/2021.11.19.468211>

Copyright © The Author(s) 2022.

<https://creativecommons.org/licenses/by/4.0/>

Attachment 2: Authorization of use of PDF paper

Re: Granting permission to use in PhD thesis [250930-003003] - An... <https://mail.ciimar.up.pt/owa/#viewmodel=ReadMessageItem&Item...>

Re: Granting permission to use in PhD thesis [250930-003003]

Permissions Helpdesk <permissionshelpdesk@elsevier.com>

seg 29-09-2025 22:48

Para: André Machado <andre.machado@ciimar.up.pt>;

⚠️ EXTERNAL E-MAIL ⚠️

--

Dear Customer

Thank you for contacting Elsevier's Permissions Helpdesk.

This is an automated acknowledgement to confirm we have received your query. Ticket number 250930-003003 has been opened on your behalf and we aim to respond within two business days.

Regards,

Permissions Helpdesk

From: André Machado

Date: 29/09/2025 10.48 PM

I am writing to request your assistance in obtaining a document granting permission to use an article that I, as the first author, published in Marine Genomics, in my Ph.D. thesis.

The article in question, titled "A whole-body transcriptome assembly of the annelid worm *Hediste diversicolor*", of which I am the first author holds significant relevance to my research and has played a crucial role in shaping the direction of my thesis.

As the main author of this thesis, it is imperative that I secure official permission from the journal to include the article in my thesis. This document of permission would not only ensure compliance with copyright and intellectual property guidelines but also demonstrate the appropriate acknowledgment and recognition of the journal's contribution to our research.

The inclusion of this article in my thesis would greatly enhance the scientific rigor and comprehensiveness of their work. It would provide essential context, support their arguments, and reinforce the findings of my research.

I kindly request your assistance in providing a formal document that explicitly grants permission to reproduce the article, including any figures, tables, or supplementary materials associated with it, in my Ph.D. thesis. This document will be submitted to the academic committee of our university, ensuring that all necessary permissions are obtained and the proper attribution is given.

If there are any specific requirements or procedures to follow in order to obtain this permission, please kindly advise. I am committed to fulfilling all necessary obligations and complying with any applicable guidelines associated with the use of the article.

Thank you for your attention to this matter. I truly appreciate your support and cooperation in helping

'Re: Granting permission to use in PhD thesis [250930-003003] - An... <https://mail.ciimar.up.pt/owa/#viewmodel=ReadMessageItem&Item...>

me obtain the necessary permission to include the article in my thesis. I look forward to receiving the requested document.

Warm regards,
André Machado

This email is for use by the intended recipient and contains information that may be confidential. If you are not the intended recipient, please notify the sender by return email and delete this email from your inbox. Any unauthorized use or distribution of this email, in whole or in part, is strictly prohibited and may be unlawful. Any price quotes contained in this email are merely indicative and will not result in any legally binding or enforceable obligation. Unless explicitly designated as an intended e-contract, this email does not constitute a contract offer, a contract amendment, or an acceptance of a contract offer.

Elsevier Limited. Registered Office: 125 London Wall, London, EC2Y 5AS, Registration No. 1982084, Registered in England and Wales. [Privacy Policy](#)

Attachment 3: Authorization of use of PDF paper

¹Re: Granting permission to use in PhD thesis [250930-003068] - An... <https://mail.ciimar.up.pt/owa/#viewmodel=ReadMessageItem&Item...>

Re: Granting permission to use in PhD thesis [250930-003068]

Permissions Helpdesk <permissionshelpdesk@elsevier.com>

seg 29-09-2025 22:58

Para: André Machado <andre.machado@ciimar.up.pt>;

⚠️ EXTERNAL E-MAIL ⚠️

--

Dear Customer

Thank you for contacting Elsevier's Permissions Helpdesk.

This is an automated acknowledgement to confirm we have received your query. Ticket number 250930-003068 has been opened on your behalf and we aim to respond within two business days.

Regards,

Permissions Helpdesk

From: André Machado

Date: 29/09/2025 10.58 PM

I am writing to request your assistance in obtaining a document granting permission to use an article that I, as the shared first author, published in Fish & Shellfish Immunology, in my Ph.D. thesis. The article in question, titled "Unravelling the main immune repertoire of *Paracentrotus lividus* following *Vibrio anguillarum* bath challenge", of which I am a shared first author holds significant relevance to my research and has played a crucial role in shaping the direction of my thesis.

As the main author of this thesis, it is imperative that I secure official permission from the journal to include the article in my thesis. This document of permission would not only ensure compliance with copyright and intellectual property guidelines but also demonstrate the appropriate acknowledgment and recognition of the journal's contribution to our research.

The inclusion of this article in my thesis would greatly enhance the scientific rigor and comprehensiveness of their work. It would provide essential context, support their arguments, and reinforce the findings of my research.

I kindly request your assistance in providing a formal document that explicitly grants permission to reproduce the article, including any figures, tables, or supplementary materials associated with it, in my Ph.D. thesis. This document will be submitted to the academic committee of our university, ensuring that all necessary permissions are obtained and the proper attribution is given.

If there are any specific requirements or procedures to follow in order to obtain this permission, please kindly advise. I am committed to fulfilling all necessary obligations and complying with any applicable guidelines associated with the use of the article.

Thank you for your attention to this matter. I truly appreciate your support and cooperation in helping

Re: Granting permission to use in PhD thesis [250930-003068] - An... <https://mail.ciimar.up.pt/owa/#viewmodel=ReadMessageItem&Item...>

me obtain the necessary permission to include the article in my thesis. I look forward to receiving the requested document.

Warm regards,
André Machado

This email is for use by the intended recipient and contains information that may be confidential. If you are not the intended recipient, please notify the sender by return email and delete this email from your inbox. Any unauthorized use or distribution of this email, in whole or in part, is strictly prohibited and may be unlawful. Any price quotes contained in this email are merely indicative and will not result in any legally binding or enforceable obligation. Unless explicitly designated as an intended e-contract, this email does not constitute a contract offer, a contract amendment, or an acceptance of a contract offer.

Elsevier Limited. Registered Office: 125 London Wall, London, EC2Y 5AS, Registration No. 1982084, Registered in England and Wales. [Privacy Policy](#)

Attachment 4: Authorization of use of PDF paper

'Re: Granting permission to use in PhD thesis [250930-003108] - An... <https://mail.ciimar.up.pt/owa/#viewmodel=ReadMessageItem&Item...>

Re: Granting permission to use in PhD thesis [250930-003108]

Permissions Helpdesk <permissionshelpdesk@elsevier.com>

seg 29-09-2025 23:02

Para: André Machado <andre.machado@ciimar.up.pt>;

⚠ EXTERNAL E-MAIL ⚠

--

Dear Customer

Thank you for contacting Elsevier's Permissions Helpdesk.

This is an automated acknowledgement to confirm we have received your query. Ticket number 250930-003108 has been opened on your behalf and we aim to respond within two business days.

Regards,

Permissions Helpdesk

From: André Machado

Date: 29/09/2025 11.02 PM

I am writing to request your assistance in obtaining a document granting permission to use an article that I, as the first author, published in Aquaculture Reports, in my Ph.D. thesis.

The article in question, titled " The male and female gonad transcriptome of the edible sea urchin, *Paracentrotus lividus*: Identification of sex-related and lipid biosynthesis genes," of which I am a first author holds significant relevance to my research and has played a crucial role in shaping the direction of my thesis.

As the main author of this thesis, it is imperative that I secure official permission from the journal to include the article in my thesis. This document of permission would not only ensure compliance with copyright and intellectual property guidelines but also demonstrate the appropriate acknowledgment and recognition of the journal's contribution to our research.

The inclusion of this article in my thesis would greatly enhance the scientific rigor and comprehensiveness of their work. It would provide essential context, support their arguments, and reinforce the findings of my research.

I kindly request your assistance in providing a formal document that explicitly grants permission to reproduce the article, including any figures, tables, or supplementary materials associated with it, in my Ph.D. thesis. This document will be submitted to the academic committee of our university, ensuring that all necessary permissions are obtained and the proper attribution is given.

If there are any specific requirements or procedures to follow in order to obtain this permission, please kindly advise. I am committed to fulfilling all necessary obligations and complying with any applicable guidelines associated with the use of the article.

'Re: Granting permission to use in PhD thesis [250930-003108] - An... <https://mail.ciimar.up.pt/owa/#viewmodel=ReadMessageItem&Item...>

Thank you for your attention to this matter. I truly appreciate your support and cooperation in helping me obtain the necessary permission to include the article in my thesis. I look forward to receiving the requested document.

Warm regards,
André Machado

This email is for use by the intended recipient and contains information that may be confidential. If you are not the intended recipient, please notify the sender by return email and delete this email from your inbox. Any unauthorized use or distribution of this email, in whole or in part, is strictly prohibited and may be unlawful. Any price quotes contained in this email are merely indicative and will not result in any legally binding or enforceable obligation. Unless explicitly designated as an intended e-contract, this email does not constitute a contract offer, a contract amendment, or an acceptance of a contract offer.

Elsevier Limited. Registered Office: 125 London Wall, London, EC2Y 5AS, Registration No. 1982084, Registered in England and Wales. [Privacy Policy](#)