MtDNA phylogeny and phylogeography of Atlantic-Mediterranean species of genus *Felimare* (Chromodorididae, Nudibranchia) and obtainance of genomic markers from RNA-seq data

Cristina P. Rodrigues
Mestrado em Biodiversidade, Genética e Evolução
Departamento de Biologia, Universidade do Porto
Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO)
2016

Orientadora
Sara Rocha, Post-Doc Researcher, Department of Biochemistry, Genetics and Immunology
School of Biology, Universidade de Vigo

Co-orientador
David Posada, Full Professor, Department of Biochemistry, Genetics and Immunology
School of Biology, Universidade de Vigo

30 de Maio de 2016

Relatório da Unidade Curricular de Dissertação do 2º ano do Mestrado em Biodiversidade, Genética e Evolução

Candidato: Cristina Polónia Rodrigues, 200906162, @fc.up.pt

Orientação Científica: Sara Rocha

Mestrado em Biodiversidade, Genética e Evolução

Departamento de Biologia/Centro de Investigação em Biodiversidade e Recursos Genético

Faculdade de Ciências da Universidade do Porto/ CIBIO

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas foram efetuadas.

O Presidente do Júri,

Porto, / /

Agradecimentos

A realização deste trabalho é o resultado de um fantástico esforço colaborativo, envolvendo cientistas de diversas áreas de estudo, e mergulhadores profissionais/não profissionais e voluntários, e sem os quais não teria as oportunidades nem os devidos recursos para abordar os principais temas do meu trabalho.

Passo a citar e a agradecer todos aqueles sem os quais este trabalho não teria sido possível.

Em primeiro lugar gostaria de deixar um especial agradecimento á minha orientadora, Sara Rocha, pela escolha deste tema de Tese, que me possibilitou o contacto com o trabalho de campo e o tratamento de dados NGS. Devo realçar também toda a disponibilidade, paciência, dedicação e profissionalismo, ao logo de todo o trabalho. Sara, obrigada por seres tão paciente comigo, e não me deixares desistir! Obrigada por trabalhares comigo e me ensinares a fazê-lo também. Obrigada por partilhares sempre bons conselhos!

Gostaria, também, de agradecer o contributos do meu co-orientador, David Posada, e professores Carlos Canchaya e Jesus Troncoso, da Universidade de Vigo. Ao David, que me facultou todas as ferramentas logísticas, materiais e bioinformáticas, bem como a oportunidade de integrar uma equipa de profissionais entusiastas, que partilhavam as suas competências e participavam ativamente nos trabalhos do grupo! Obrigada David, pelo curso de mergulho, por um espaço no teu laboratório, pelas sugestões, e pela tua paciência comigo!

Carlos tenho que dizer-te que os teus alunos têm muita sorte! Nunca esquecerei a tua boa-disposição, o facto de interromperes o teu trabalho para me acudires quando eu não compreendia os comandos. Obrigada pelas opiniões, e por te preocupares comigo tantas vezes!

Jesus, foi uma honra trabalhar com tão prestigiado ecologista e apaixonado dos Opistobranchios. Sempre ocupado, mas sempre disponível!

Não poderia esquecer o Miguel. Foi o Miguel quem me sugeriu acrescentar o capítulo do genoma mitocondrial. E muito nos ajudaste (a mim e á Sara) com a árdua tarefa da contaminação!

Quero também agradecer a toda a equipa (Merche, Ramón, Maria, Leonardo, e Diego) porque me receberam tão bem, e me apoiarem nas mais diversas dificuldades!

Gostaria ainda de reconhecer e agradecer o trabalho de todos os mergulhadores que, voluntariamente e dedicadamente, se predispuseram a recolher, preparar e enviar, com descrições detalhadas dos mergulhos, organismos para este trabalho, ao longo da costa Espanhola. Fizeram-no somente em nome da ciência, e na espectativa de que estudos como este contribuam para a expansão do conhecimento e proteção das espécies marinhas. Bem sei que para eles não foi fácil o sacrifício de todos os pequenos seres, mas é louvável a sua compreensão. São eles Enric Madrenas, sócio fundador do grupo GROC (Opisthobranchs Research Group) e VIMAR (Marine Life), bem como mergulhador/fotógrafo experiente no grupo OPK-Opistobranquis; Luis Naya Garmendia, instrutor de mergulho em San Sebastián (Espanha), e autor do livro “Nudibranquios de la Costa Vasca: el pequeño Cantábrico multicolor”; José Acuesta e amigos de Andalucía, José e David, pelas amostragens de *Felimare* no Sul de Espanha; e finalmente Carlos Fernandez Cid (médico oftalmologista de profissão), mergulhador/fotógrafo amador, que nos apresentou os principais mergulhadores acima mencionados, bem como a sua abertura em projetos deste tipo. Um muito obrigado por terem sido tão pacientes á chegada destes resultados.

Não menos importante foi o contributo do Professor Lucas Cervera, professor na Universidade de Cadiz (Espanha), que nos cedeu amostras essenciais do Atlântico Ibérico e costas de Marrocos; e de investigadores das zonas centro e este do mar Mediterrânio, favorecendo a minha base de dados com valiosíssima informação genética proveniente das costas de Itália, Montenegro e Grécia. Refiro-me a Giulia Furfaro e Paolo Mariottini, ambos investigadores na Universidade de estudos de Roma (Itália), e Dimitris Poursanidi, investigador no Instituto de Matemáticas Aplicadas de Creta (Grécia).

Termino agradecendo aos meus Pais, Rosa e Laurindo, que tiveram um papel fundamental na minha educação. Apesar de todas as minhas incertezas e fraquezas foram perseverantes e acreditaram sempre em mim, e nas minhas capacidades, e nunca permitiram que eu desistisse fosse do que fosse!

Acknowledgments

This work has resulted from collaboration among scientist of distinct areas of research, and professional divers, who kindly enriched my work by gathering fresh/preserved material around Iberian, African and Italian coasts. Without these people this work would not be possible.

I now cite and show my gratitude to all these effortless people:

First and foremost I thank my adviser, Sara Rocha, for her guidance, support, professionalism and patience during my graduate internship at Vigo University. It was Sara who first suggested that I considered to focus my research of these beautiful and enigmatic gastropods, always with the best interest in mind. The choice of this work allowed me to combine the contact with field work and the bioinformatics data analyses from NGS technologies. Sara, thank You very much for always being so tolerant, and never letting me quit! Thank You for working late with me, e to teach me to be persistent. I will always remember your good advises!

I would like to thank the contribution of my co-adviser, David Posada (Leader of the Phylogenomics lab), and Professors Carlos Canchaya and Jesus Troncoso, from University of Vigo. To David, who's hosted me for one year, provided me with all the logistic, material and bioinformatics tools, and gave me the opportunity of being part of his team! Thank you, David for the Diving course, for a seat in your lab and all your comments and suggestions regarding my work, and also for being patient.

Carlos, I must tell your students are very lucky! I will never forget your gaiety, the fact that for many times you have interrupted your deeds to help me with the command lines and my doubts. Thank You for your wise opinions and good advises.

Jesus, it has been an honour to work with such respected ecologist and affectionate of the Opisthobranchs. Always so busy, yet always available!

I could not forget Miguel. It was Miguel whom suggested the chapter of the mitochondrial genome. I very much appreciate your availability with the laborious task of contamination.

I am also very thankful to the remaining elements of the Phylogenomic's team, Merche, Ramón, Maria, Leonardo martins, e Diego Mallo, whohave received me so friendly, and supported me with the most diverse difficulties or doubts.

Additionally, I must recognize and express my gratitude to all those people that had volunteered and dedicated their time to the collection, preparation, and send the samples many times with detailed descriptions of each dive session. Most of the times, they didn't have any monetary remuneration, and did it in the name of science, and expecting that studies like this will make a difference headed for the knowledge and protection of marine species. I am very much aware that scarifying these little animals is not easy, and therefore I appreciate Your comprehension on such delicate subject. I am referring to Enric Madrenas, founder partner at GROC (Opstobranchs Research Group) and VIMAR (Marine Life), expert diver and Photographer at OPK-Opistobranquis; Luis Naya Garmendia, (autor do livro "Nudibranquios de la Costa Vasca: el pequeño Cantábrico multicolor") ; José Acuesta and friends from Andalucía, José e David, for all the sampling efforts on *Felimare*, at the South of Spain; and finally to Carlos Fernandez Cid (professional as an ophthalmologist) an amateur diver who introduce us to the previously mentioned divers, as all his openness to projects of this kind. Thank You all, also for your patience to the arrival of these results.

Not of inferior relevance was the contribution of professor Lucas Cervera, professor at the University of Cadiz (Spain), and proficient at the study of Opisthobranchs, for providing us with important samples from the Atlantic (Iberia Peninsula and coasts of Morocco); and researchers from central and eastern sides of the Mediterranean Sea, contributing in the same way with precious genetic information from coasts of Italy, Montenegro and Greece. I am referring to investigators Giulia Furfaro e Paolo Mariottini, both at the University di Studi di Roma (Italy), ans Dimitris Poursanidi, researcher at the Institute Applied and Computational Maths of Crete (Greece).

I end this section by acknowledging my gratitude to my parents, Rosa and Laurindo, who had a crucial role during the course of my education. Regardless all my uncertainties and weakness, they stood firm and believed me and my aptitudes, without ever letting me give up!

Resumo

Os Opistobranquios pertencentes ao género *Felimare* são facilmente reconhecidos pelos seus padrões característicos de coloração: fundo azul, marcado por um ou mais linhas/manchas que variam entre o branco e o laranja. Contudo, a diversidade destes padrões e também a semelhança entre várias espécies dificultam sua identificação no campo e geram suspeitas da existência de espécies crípticas. Urge, então, uma melhor compreensão dos padrões filogeográficos e evolutivos deste grupo, tarefa dificultada pela falta de marcadores moleculares específicos.

Na primeira parte deste trabalho, eu inicio o estudo filogeográfico de sete espécies descritas do género *Felimare*, no Este Atlântico, (EA) com o objectivo de (i) descrever os padrões espaciais de distribuição da variação genética e (ii) identificar possíveis barreiras oceanográficas e os seus efeitos na diversidade genética de diferentes espécies co-distribuídas.

Os padrões de diversidade genética mitocondrial revelaram uma aparente ausência de estrutura nas espécies *F. cantábrica*, ao longo da costa Atlântica, e *F. tricolor*, desde a costa Atlântica ao interior do mar Mediterrânico. Por outro lado, as restantes cinco espécies compõem potenciais complexos de espécies. As barreiras marinhas a atuar nestas espécies são na maior parte dos casos comuns, quer para espécies de desenvolvimento direto (*F. villafranca*), ou planktotróficas (*F. picta, F. fontandraui, F. orsinii*, *F. bilineata*). As barreiras incluem o Estreito de Gibraltar, a frente de Alborán-Orán, e possivelmente a frente Balear. A diferenciação intraespecífica mais alta foi encontrada em *F. picta*, entre populações Atlânticas/Mediterrânicas e de Cabo Verde, coincidindo com outros resultados recentes que argumentam esta espécie ser de facto de um complexo de espécies. Possíveis casos de hibridação são discutidos em detalhe.

Quanto às relações filogenéticas entre as espécies, inferidas a partir de um marcador mitocondrial, a maioria, sobretudo ao nível basal, é pouco suportada. Contudo, foram identificados 12 clados bem suportados (PP> 0.90) e diferenciados. Foi consistente a divisão do género *Felimare* em dois grupos, o primeiro incluindo espécies oriundas do Oceano Pacífico (EP) e Caraíbas; *F. porterae* e *F. kempfi*, e um segundo grupo, que inclui todas as outras espécies do EP e EA. Usando uma taxa de evolução calculada para o fragmento aqui usado noutros grupos de gastrópodes, estima-se que a origem do grupo remonta ao Oligoceno. No entanto, grande parte da diversificação interespecífica parece ter lugar durante o Mioceno, e apenas a divergência entre *F. fontandraui* and *F. tricolor*, data do Pleistoceno. Pelo menos duas colonizações transatlânticas necessitam ser

invocadas para explicar a distribuição atual das espécies deste género, mas uma maior representação das espécies do Oceano Pacífico será necessária para uma melhor compreensão da biogeografia deste grupo.

A aplicação de técnicas de ultra-sequenciação (NGS) têm vindo progressivamente a ganhar destaque, nomeadamente em estudos de filogenia. Mais recentemente, dados de transcriptoma provaram a sua utilidade para a inferência de filogenias de gastrópodes. Ao mesmo tempo também as NGS se têm demonstrado úteis para estudos de reconstrução dos mitogenomas, de forma rápida e mais facilitada.

Na segunda parte deste trabalho procurei (i) caracterizar os transcriptomas de duas espécies do género *Felimare*, o que englobou a sua reconstrução e anotação, e a procura de fragmentos ortólogos entre os mesmos, e (ii) reconstruir os genomas mitocondriais respetivos (*F. cantabrica* e *F. villafranca*). Uma das finalidades do transcriptoma será a identificação de uma bateria de novos marcadores moleculares que possam ser usados em futuros estudos filogenéticos. Conjuntamente, e uma vez que a problemática da contaminação afetou os dados obtidos, eu avalio e discuto a mesma, sugerindo passos adicionais para a sua prevenção.

Para ambas as espécies (*F. cantabrica F. villafranca*, respectivamente) obtive 38 e 35 milhões de pequenos fragmentos RNA-seq (reads) de alta qualidade a partir de tecnologia de sequenciamento Illumina HiSeq. Depois de filtrados, os mesmos eram compostos por 54.6K e 58.5K transcritos, sendo que 16,915 e 18,664 de pelo menos 500 pares de base (bp) respectivamente. Desses, 4,275 e 2,952 transcritos, respetivamente, foram atribuídas diferentes funções biológicas (GO terms) de modo similar a estudos recentes de RNA-seq de outros gastrópodes.

Durante o processo de “assembly” dos mitogenomas detectei a existência de transcritos com alta percentagem de identidade com DNA mitocondrial de anfíbios, o que se confirmou como resultado de uma contaminação laboratorial. De um modo exploratório, e após filtrar os conjuntos de contigs devido à contaminação detectada, foi possível ainda a obtenção de 728 putativos ortólogos entre as duas espécies. Estes fragmentos são um ponto de partida para o desenvolvimento de novos marcadores nucleares, úteis não só para o género *Felimare*, mas também para outros Heterobrânquios/Opistobrânquios.

Os genomas mitocôndriais foram reconstruidos em cerca de 14Kb, tendo uma organização génica idêntica á de outros nudibrânquios. Porém, em nenhuma das espécies o genoma mitocondrial foi completamente assemblado, independentemente da metodologia adoptada (alinhamento manual de transcritos, ou pelo assembly directo dos “reads” usando o programa MITObim).

Palavras-Chave

Heterobranchia, Opisthobranchia, Chromodorididae, filogeografia *Felimare,* filogenia *Felimare*, transcriptoma, NGS molluscs, contaminação NGS, mitogenome molluscs

Abstract

Opisthobranchs from the genus *Felimare* are easily recognized for their dark-bluish background with one or more white or yellow lines and/or dots. Nevertheless, the diversity of these patterns together with morphologic similarities between species makes their identification in the field often challenging and casts doubts respect to the existence of cryptic species. A comprehensive understanding of the biogeography and evolutionary patterns in the genus is hindered by the lack of detailed molecular studies between and within species.

In the first part of this thesis I assessed the genetic structure of seven described species of the genus *Felimare* in the East Atlantic (EA), aiming to (i) describe the spatial patterns of the distribution of genetic diversity and, (ii) identify potential marine barriers and their effects on the genetic diversity of co-distributed species.

Patterns of mitochondrial genetic diversity showed the absence of genetic structure in *F. cantabrica* throughout the Atlantic coast, and in *F. tricolor* from the Atlantic coast to the inner Mediterranean Sea. Conversely, the remaining five species seem to represent potential species-complexes. Most of the marine barriers acting on these species seem to be common regardless of the direct (*F. villafranca*) or planktotrophic (*F. picta, F. fontandraui, F. orsinii, F. bilineata*) mode of development. These barriers include the Strait of Gibraltar, the Alboran-Oran Front, and possibly the Balearic Front. The highest intraspecific differentiation was found between *F. picta* from Iberian/Mediterranean and Cabo Verde/S. Tomé islands, in agreement with recent studies that argue *F. picta* to be in fact a complex of cryptic species. Potential cases of hybridization are discussed in detail.

The phylogenetic relationships between species, inferred from a single mitochondrial marker, are poorly supported, especially at the basal level. Nonetheless, twelve well-supported (PP>0.90) and highly differentiated clades were identified. *Felimare* was consistently divided into two groups; one containing species from the Eastern Pacific (EP), *F. porterae* and Caribbean *F. kempfi*, and another including species inhabiting the EP and Eastern Atlantic (EA). Divergence time estimates, obtained using a rate of evolution for the COI gene estimated from other gastropods, date the origin of the group in the Oligocene. Yet, most diversification within the genus seems to occur in the Miocene, and only the divergence between *F. fontandraui* and *F. tricolor,* was estimated as Pleistocenic. At least two transatlantic colonisations must be invoked to explain the present distribution of the genus.

However, and for a better comprehension of the biogeography of the group, more species from the EP and Western Atlantic should be included in the analyses.

The application of high throughput sequencing has been progressively increasing, with next generation sequencing data being recently used to infer gastropod phylogenies. At the same time, NGS techniques are proving useful when applied to mitogenome reconstruction studies. In the second part of my thesis I aimed at (i) characterize the transcriptomes of two *Felimare* species, by assembling and annotating them, as well as using them to search for potential molecular marker fragments, and (ii) assemble their mitochondrial genomes. One of the main goals of obtaining transcriptomic data was to obtain a battery of molecular markers to support future phylogenetic studies. Additionally, I discuss the problematic of contamination in high throughput datasets, and recommend additional cleaning steps to be included in any type of workflow.

For both species (*F. cantabrica F. villafranca*, respectively) I obtained 38 and 35 million high-quality RNA-seq reads using Illumina HiSeq. After filtering, the assembly was composed of 54.6K and 58.5K transcripts, being 16,915 and 18,664 of at least 500bp long, respectively. From those, biologic functions (GO terms) were assigned to 4,275 and 2,952 transcripts respectively, similarly to recent RNA-seq studies in other gastropods. During the process of assembling the mitochondrial genomes of the two species, I detected sequences from amphibian origin, later confirmed as contamination from the laboratory. After filtering the detected contamination from both assemblies I could still obtain a total of 728 potential ortholog fragments both studied species. These fragments represent a starting point for the development of nuclear markers not only for studies in the genus *Felimare*, but also in Heterobranchia/Opisthobranchia.

The two reconstructed mitochondrial genomes were approximately 14Kb long, with the same arrangement of genes seen in other nudibranchs. However, none of the two mitochondrial genomes was completely assembled regardless of the methodology (contigs manual alignments or from the algorithm MITObim).

Keywords

Heterobranchia, Opisthobranchia, Chromodorididae, phylogeography *Felimare,* phylogeny *Felimare*, transcriptome, NGS molluscs, NGS contamination, mitogenome molluscs.

Contents

List of Figures

List of tables

List of Abreviations

AC – Azores Current

AOF – Alboran-Oran Front

BI – Bayesian Inference

BF – Balearic Front

CaC – Canary Current

Cegma – Core Eucaryotic Genes Mapping Approach

EAM – Eastern Atlantic and Mediterranean

IC – Ibiza Channel

ITCZ – Tropical Convergence Zone

LGM – Last Glacial Maximum

MCC – Aximum Clade Credibility Tree

MSC – Messian Salinity Crisis

MDFs – Mantle Dermal Formations

ML – Maximum Likilihood

MAW - Modified Atlantic Waters

MY - Million years

NAC – North Atlantic Current

NADC – North Atlantic Drift Current

NACW – North Atlantic Central Water

NEC – North Equatorial Current

NECC - North Equatorial Counter Current

NE - North Eastern

NGS – Next Generation Sequencing

PC – Portugal Current

PP – Posterior Probability

QC - Quality Control

SACW – South Atlantic Central Water

SE – South Eastern

SEC – South Equatorial Current

SST – Sea Surface Temperature

SE - South Eastern

SG – Strait Gibraltar

Chapter 1 General Introduction

1.1 GENETIC STRUCTURE OF MARINE POPULATIONS

1.1.1 Life strategies and dispersal in the ocean

For decades the ocean realm was considered a vast and homogeneous system, with no obvious barriers to organisms' dispersal (McGowan 1971; Benzie 1999; Finlay 2002). Additionally, marine organisms often present large effective population sizes, high fertility, high dispersal abilities at least in their pelagic forms (larvae and eggs), and high levels of gene flow (Hedgecock, 1986; Palumbi, 1994; Palumbi, 1996). Under such conditions, high connectivity among geographically separated populations was predicted, and opportunities for allopatric divergence, i.e., divergence due to extrinsic barriers (de Queiroz 2005), were considered to be rare. However, some marine species exist that are widely distributed and exhibit genetic homogeneity (Pujolar et al. 2002; Gilbert-Horvath et al. 2006). Importantly, recent research has been leading to a distinct perception of global open-water population structure, dominated by isolation-by-distance patterns (e.g. Viñas et al. 2004; Sherman et al. 2008) or even sharp genetic breaks (i.e., areas of restricted gene-flow) between populations (Barber et al. 2002; Lourie et al. 2005; Mantelatto et al. 2014).

Connectivity patterns, seem to be, at a great extent, species-specific. Marine organisms display a wide range of reproductive strategies, which may give us some clues about their dispersal abilities and the spatial scale at which gene flow occurs (Palumbi & Warner 2003). Consequently, by studying species' life history traits (such as spawning and egg type, larval duration and behaviour, type and period of development, etc) together with recent and historical oceanographic features of the region (e.g. short tidal currents, jets, gyres; rafting) (Leis & Fisher 2006; Diehl et al. 2007; Nikula et al. 2010; Fraser et al. 2011) one might hypothesize on species' phylogeographic structure, as well as macro evolutionary patterns (Hedgecock 1986; Arndt & Smith 1998; Pechenik 1999; Collin 2001; Jeffery et al. 2003).

Benthic species, and particularly those with sedentary life-style or limited mobility (e.g. marine invertebrates and some non-migratory fishes) (e.g. Shulman & Bermingham 1995; Thorpe et al. 2000) may disperse large distances as adults by rafting, drifting or even human mediated-transport (Martel 1991; Sherman et al. 2008; Rastorgueff et al. 2014). However the main dispersal mechanisms are still their early life pelagic stages (gametes, larvae, eggs, or propagules) (Gopurenko & Hughes 2002; Jeffery & Emlet 2003; Thiel & Gutow 2005).

The majority of the sedentary or sessile marine organisms produce long-living and free-swimming larvae (plankthotrophic), or short-living (lecithotrophic) veliger larvae (Castelin et al. 2012), expected to behave as passive, long distance, dispersers (e.g. Grosberg & Cunningham 2001;

Cárdenas et al. 2009). Nonetheless the broad geographic ranges that characterize such populations separated by hundreds of kilometres often reveal areas of restricted gene-flow (Palumbi 2003; Sotka et al. 2004).

The diversity of genetic patterns regarding planktonic dispersers is highly variable and genetic breaks are often congruent with present-day barriers to dispersal. Factors such as larvae retention through vertical movements and habitat preferences (Hohenlohe 2003); high mortality (Todd et al. 1998; Sanford et al. 2006; Weiss et al. 2007), presence of significant oceanographic barriers that retain the offspring near their natal habitats (Hoffman et al. 2011) or strong local adaptation (Sotka 2005; Levin 2006; Marshall et al. 2010), limit connectivity in planktonic dispersers. Other organisms lack the larval stage and produce reduced forms of adults, often from brood pouches or benthic egg capsules (direct development). In these cases juveniles persist in the parental environment most of their time, decreasing dispersal and promoting local adaptation (reviewed in Bohonak, 1999; Sanford & Kelly, 2011). Direct developers have generally smaller distribution ranges (Soares et al. 2006) and more marked genetic structure than planktonic organisms (Collin 2001; Teske et al. 2007; Pelc et al. 2009; Kelly & Palumbi 2010). Their phylogeographic patterns generally show more structured populations, although without necessarily resulting in their complete isolation (e.g. Bohonak, 1999; Collin, 2001; Bell, 2008), neither completely preventing ranges as wide as species with pelagic larvae. Exceptions exist, with some direct developers showing widespread populations without apparent genetic structure (Oosthuizen et al. 2004; Zealand 2004; Teske et al. 2007).

As barriers shift, species with low dispersal are expected to reflect historical patterns over long time-scales, because low levels of gene flow allow the signal of historical separation to persist for many generations (Hellberg et al. 2002; Pelc et al. 2009). On the other hand phylogeographic breaks regarding planktonic dispersers are usually congruent with contemporary oceanographic features (Pelc et al. 2009).

1.1.2 Cryptic speciation

Cryptic species can be defined as discrete species that are difficult, or sometimes impossible, to distinguish morphologically and thus have been incorrectly classified as a single taxon (Beheregaray & Caccone 2007). Recently, Pante et al. (2014) proposed a distinction between cryptic species and pseudo-cryptic species. The former refers to two or more species that are taxonomically classified as one single nominal species and lack obvious morphological features that diagnose differentiation, while the latter refers to situations where there are subtle morphological differences, but calls upon other lines of evidence (such as DNA) to show what morphology could not easily discriminate. Obviously, their existence has implications on the estimation of biodiversity.

Morphological characters are generally the first elements to be considered in studies of taxonomy and systematics (e.g. Gosliner et al. 1999; Pola et al. 2008). Nevertheless characters related to mating behaviour, chemical signals, physiology or habitat preferences are equally important (Bickford et al. 2007; Vrijenhoek 2009): i) species may recognize themselves by nonvisual mating signals, such as chemical recognition (Malenke et al. 2009; Stanhope et al. 1992; Landry et al. 2003) or mating calls (Henry 1994; Kingston et al. 2001; Amorim et al. 2010); ii) strong selection or settlement in severe environments (e.g. deep sea, under-water karst, glacial shelters) may have strong effects on both behavioural and physiological patterns, without necessarily implying the evolution of morphologic distinctive characters (de Vargas et al. 1999; Rothschild & Mancinelli 2001; Lefébure et al. 2006; Maan & Cummings 2012), and iii) morphological identification is often lagging and requires technologic sophisticated anatomical studies (Jörger et al. 2012).

The boost of molecular markers (alloenzymes, nucleotide sequences) has challenged previous morphology/ecology based assumptions on both population structure and species boundaries in the ocean (Knowlton 2000), fostering the knowledge in this field. Over the past two decades research on cryptic species has largely interested the scientific community, fuelled in large part by the increasing availability of DNA sequences (Bickford et al. 2007). Molecular phylogenies are often used together with morphological characters in taxonomic studies (Knowlton 2000; Blanquer & Uriz 2007). As a result several marine species previously considered cosmopolitan have shown to be in fact cryptic/pseudo-cryptic species complexes of a wide range of organisms including algae (Wolf et al. 2012), crustaceans (Cook et al. 2008; Trontelj et al. 2009; Mantelatto et al. 2014), molluscs (Carmona et al. 2011), bryozoans (McGovern & Hellberg 2003; Nikulina et al. 2007), ascidians, sponges (Solé-Cava et al. 1991) and fishes (von der Heyden et al. 2011; Thomas et al. 2014).

1.2 THE NORTH-EASTERN ATLANTIC OCEAN AND THE MEDITERRANEAN SEA

The North-eastern Atlantic Ocean and the Mediterranean Sea are interesting areas where to study the differentiation of marine organisms. Overall, in these areas, genetic differentiation can be studied across apparently continuous coastal habitats, such as the Iberian shores, across habitats separated by deep water extents, like the African shores or the Macaronesia (and other) islands, across areas separated by ocean fronts (like Almeria-Oran), or by major topographic features (as the Gibraltar and Sicily Straits).

1.2.1 Mediterranean Sea

The current configuration of the Mediterranean Sea results from the collision of two major continental plates, the Eurasian (Laurasia) and the African plates (Gondwana), and further African microplates; the Iberia, Apulia and Arabia. The western side of the Mediterranean basin was formed by the collision of the Iberian and Apulian plates with Eurasia in the early Tertiary (Eocene to early Miocene). The eastern side basin was formed later, from the collision of the Arabian Promontory with Eurasia, around 16 MY (Sá-Pinto 2008). At this point the Mediterranean becomes an enclosured sea that communicates with the Atlantic Ocean by two narrow connections: the Betic and the Rifian corridors, in southern Iberia and northern Morocco, respectively (Rögl 1999; Carminati & Doglioni 2004). Their later closure triggered the Messinian salinity Crisis that started around 5.96 MY (Krijgsman et al. 1999), and lasted about 0.5 million years, during which the Mediterranean Sea turned into a shallow hyper saline basin. After that period Atlantic waters flooded again the Mediterranean basins, through the Strait of Gibraltar, allowing contact with the North Atlantic, as we know it (Krijgsman et al. 1999). In its current configuration, the Mediterranean further contacts with the Black Sea through the Marmara Sea, with the Adriatic Sea through the Pelagosa archipelago and with the Red Sea through the Suez channel. Separating the western and eastern sides of the Mediterranean Sea are the Strait of Sicily (average depth of 330 m and 130 km wide) and the narrow and shallow (300 m wide and average depth 80m) Strait of Messina, between Calabria and Sicily (Billard 1974; Battaglia et al. 2010). Both sides can be further subdivided into smaller regions based on distinct oceanographic features (e.g. depth, currents, and gyres).

Water mass circulation in the Mediterranean Sea occurs through eastward surface and westward subsurface currents. Circulation is driven by winds stress, and the balance between evaporation loss and net cooling (Billard 1974). When the West Atlantic current enters the Mediterranean Sea dense waters it forms what‘s called the Modified Atlantic Waters (MAW) and further originates the formation of two anti-cyclonic gyres, west and east of the Alboran Sea (Figure 1). The MAW may then either re-enter in the Alboran Sea gyres or rather move in the Algerian Current, eastward to the African coast towards the Sicily Strait. In the northern area of the western Mediterranean the Algerian Current forms part of larger-scale cyclonic gyres dominating the surface circulation, between Majorca islands and Sardinia Island, as well as above the coast of Sicily. At the coasts of Sicily the MAW currents split and continue eastwards, increasing in salinity, and flowing through the Ionian Sea and Levantine basin respectively. Generally such increase in salinity is correlated with higher temperatures in the summer. Several cyclonic and anti-cyclonic gyres existing in eastern basins result from currents bifurcations and are interconnected by several jets. During the winter season extreme cooling causes surface water to sink and spread westwards by Levantine Intermediate waters. This Mediterranean outflow is constituted by deep and intermediate depth mixing

waters, which return back to the Atlantic Ocean through the Gibraltar Strait as the Western Intermediate Waters, with colder temperatures around 13.0–13.5 ◦C and salinity of 38.0–38.5 *psu* (Millot & Taupier-Letage 2005). Once released in the Atlantic Ocean the Mediterranean outflow descends below 1000m in the Gulf of Cadiz, and can be traced as discrete subsurface "lenses" of salty and warmer water, named "Meddies". The Meddies are polarized northward, with a portion forming a poleward undercurrent along the slope of Iberian Peninsula (Mason et al. 2006).

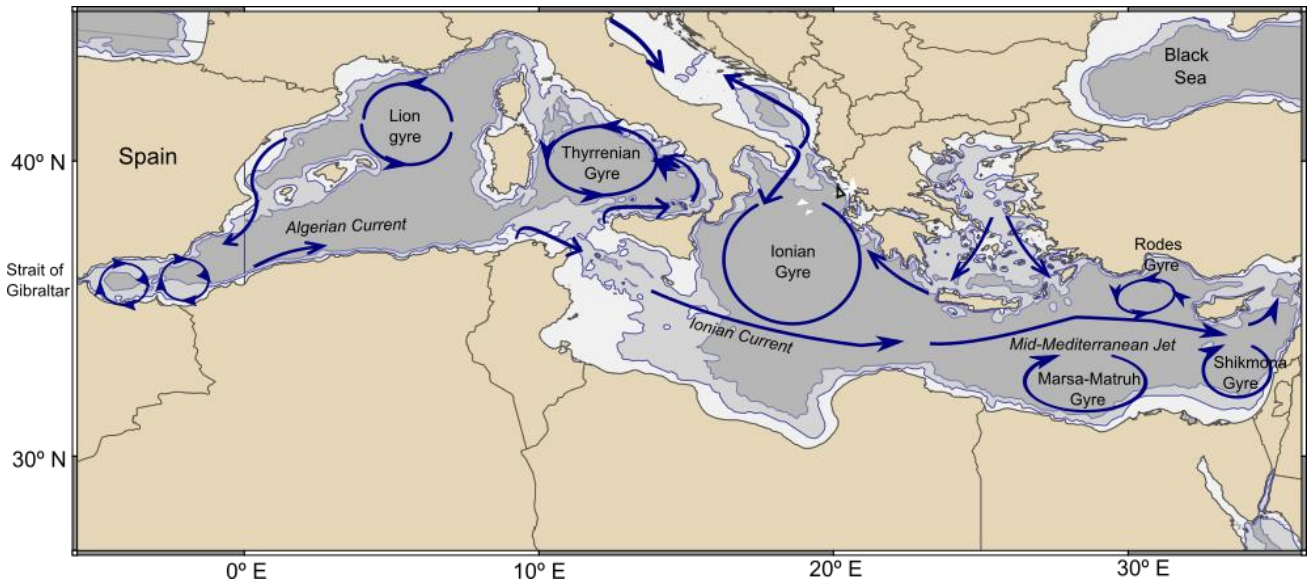


Figure 1 Hydrography, bathymetry major sub-divisions of the Mediterranean Sea. The 200m and 1000m contours are shown and areas in between shaded from lighter to darker grey. Present patterns of surface water circulation are shown in blue (adapted from Rohling et al. 2009).

1.2.2 Northeastern Atlantic Ocean

The eastern boundary of the North Atlantic subtropical gyre spreads from the northern tip of the Iberian Peninsula at 43°N to south of Senegal at 10ºN (Chen 2007). Within this area, this work covers also the islands of Madeira and Cabo Verde, from the Macaronesian region. The area of study includes additionally the islands of São Tomé e Príncipe, located southwards, at the Eastern central Atlantic region.

1.2.3 The Macaronesian archipelagos

Macaronesia comprises five archipelagos, located off the western coasts of Europe and Africa. These insular landmasses include the archipelagos of the Azores, Madeira and Selvagens, the Canary Islands and finally the archipelago of Cabo Verde. They are the result of distinct hotspots of volcanic activity, and have strong ontogenetic affinities with the Atlantic coast of the Iberian Peninsula and the north-western fringes of Africa; nonetheless they were never in contact with any of the continental landmasses.

The archipelago of Madeira comprises three islands located between 33º06' and 32º22'N and between 16º17' and 17º15'W, 635 kms away from the nearest continental coast (western Morocco). Porto Santo is the oldest island, with an estimated age of 14.3 MY (Mitchell-Thomé 1985; Geldmacher et al. 2001).

Cabo Verde is southernmost archipelago, located 570 km away from the nearest mainland point (distance from the island of Boavista to Dakar, Senegal) and site between 17°13' and 14°46'N and between 25°21' and 22°40'W. Within it, Sal is the most ancient island with an estimated age of 25.6 MY (Mitchell-Thomé 1985; Cunha et al. 2005), and Brava and Santo Antão, at its easternmost region, are the most recent, with estimated ages of 5.9 MY and 7.6 MY, respectively (Cunha et al. 2005).

1.2.4 Seamounts

Seamounts (Figure 2) are upward projections of the deep sea floor, with elevations of 100 m or more (Hillier & Watts 2007), and important at geological, oceanographic, and biological levels. Biologically these underwater features are considered as hotspots of pelagic biodiversity and endemism (de Forges et al. 2000; Samadi et al. 2006), showing consistently higher species richness than shore bathymetric clines or open oceanic areas (Morato et al. 2010), and are at some extent important aggregating locations for migratory fishes (e.g. Klimley et al. 2005; Rodríguez-Cabello et al. 2009).

Recent studies contest the validity of these submerged mountains as biodiversity hotspots. Despite several studies advocating that seamounts can have distinct composition of species from the continental slope, it is equally true that they harbour comparable levels of faunal composition, with strong affinities with the continental slopes (Samadi et al. 2006; Hall-Spencer & Rogers 2007; McClain et al. 2009; Howell et al. 2010; Pollock et al. 2014). Furthermore, levels of endemism in these places are highly variable: for example, data for South Pacific seamounts indicates > 30% of the species are potentially endemic (Parin et al. 1997; de Forges et al. 2000) while < 3% has been estimated for North Eastern (NE) Atlantic seamounts (Hall-Spencer & Rogers 2007).

Seamounts are also thought to play important roles on the biogeography of species, acting either as 'stepping stones' (Rowden et al. 2010; Packmor et al. 2014), or ''trapping stones'' for dispersal (George 2013) of small benthic organisms (Packmor et al. 2014). Genetic structure of populations of distinct seamounts in the world (including in the NE Atlantic), such as ophiuroids, corals, and symphurine tonguefish (Cho & Shank 2010; Miller et al. 2010; Tunnicliffe et al. 2010) demonstrate that different patterns of dispersal and migration can exist, with different seamount populations acting either as 'sources' or as 'sinks' depending on the species studied, and the interaction between physical factors and life-history traits.

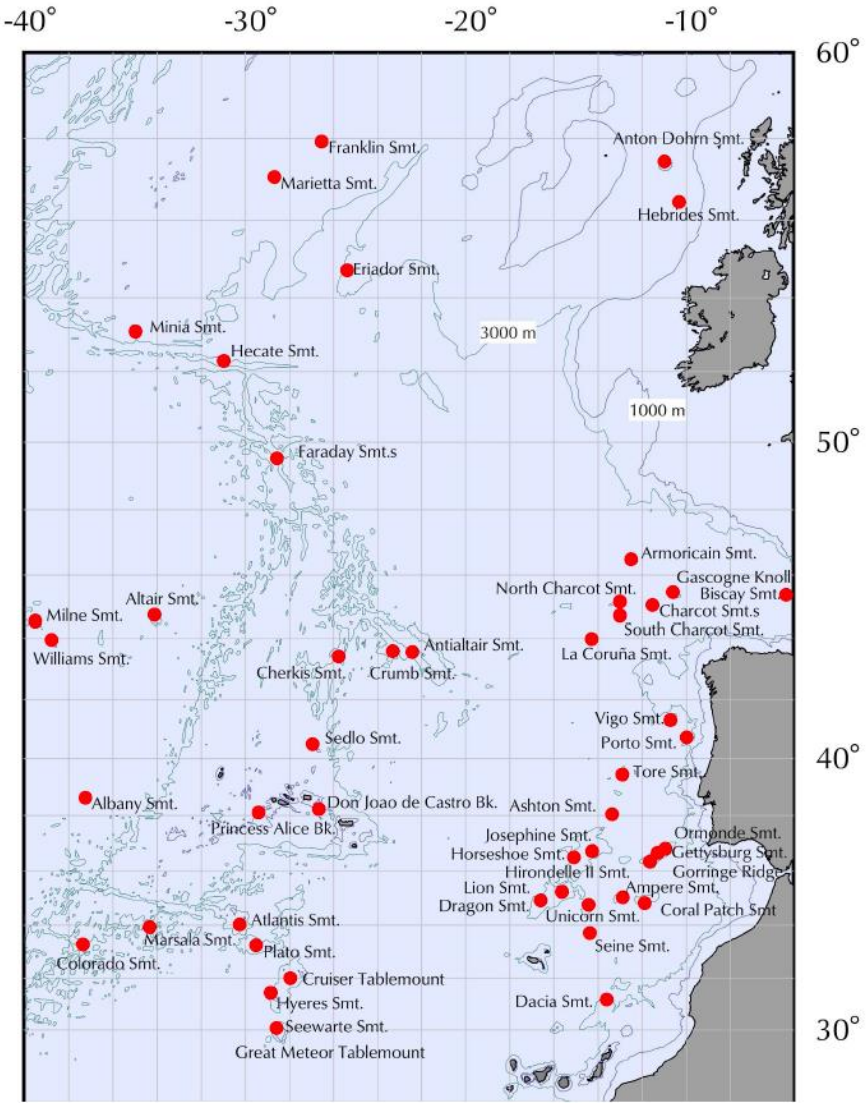


Figure 2 Distribution of principal seamounts (red dots) in the Northeast Atlantic and Mediterranean (adapted from Schmidt & Christiansen 2004).

1.2.5 Northeastern Atlantic Ocean currents

The gradual closure of the Panama seaway, around 4.5 – 2.8 MY (Schmittner et al. 2004), considerably shaped the present day sea-currents in the Northeastern Atlantic, with the interruption of the communication between the Pacific and Atlantic Oceans. In the eastern region of the Atlantic Ocean, surface water circulation to both Poles is largely driven by wind regimes (van Aken 2002; Huthnance et al. 2002), while subsurface currents, like poleward flows, may be regulated by larger-scale mechanisms, such as the decline of the sea level (poleward), the sea level rising toward the coast, but also from regional buoyancy-driven coastal currents (Mason et al. 2006). Surface currents in the Northeastern Atlantic region are to a large extent dominated by the North Atlantic Current (NAC) and the Azores Current (AC) (Figure 3), which are formed between the Grand Banks and the Mid Atlantic Ridge by the branching of the Gulf Stream (Mason et al. 2006). The NAC then splits to form the North Atlantic Drift current (NADC), spreading between Iceland and the British Islands, while the remainder continues to east and northern Europe (Mason et al. 2006). In turn, the AC, at first flowing south and then eastwards, then splits in two branches: 1) northern flowing towards the gulf of Cádiz, and posteriorly feeding Canary Current (CaC), and 2) southern flowing south-eastwards, passing west of Madeira, towards the Canary islands (Johnson & Stevens 2000; Barton et al. 2007; Pingree 2009). The CaC then flows southwards along the African coast, from where it separates in the region of Cap Blanc, and at the proximity of Cabo Verde all the flow then turns westwards, supplying the North Equatorial Current (NEC, Mason et al. 2005). Also relevant to the understanding of the circulation patterns, is to consider the two Atlantic water fronts: the North Atlantic Central Water (NACW), dominated by high rates of salinity above the 600m, and the South Atlantic Central Water (SACW), a southward current slightly warmer and driven by the CaC, at south of Cap Blanc (Figure 3).

The surface currents at western and northwestern coasts of the Iberia Peninsula are dominated by the Portugal Current (PC), which flows south during all the year, between Azores and continental Portugal, in between the North Atlantic current and the Canary current (Peliz & Fiuza 1999). The circulation pattern close to the continental platform is much more complex, exhibiting a marked seasonality defined by topography and coastal wind regimes (Huthnance et al. 2002; van Aken 2002). During spring and summer seasons, winds predominate from northeast giving rise to the Portuguese Coastal Current flowing southward on surface (<100m) and to the Portuguese Coastal Underwater Current, flowing northward by the coastal platform (Wooster et al. 1976; Bakun & Nelson 1991). On the other way, during the rest of the year, winds from southeast predominate, causing the inversion of the direction of the surface current, originating the Portuguese Coastal Counter Current (also known as Iberian Poleward Current), a water mass of 1500m deep, which also leads to the spread of Mediterranean waters to the west and north of the Iberian Peninsula (Wooster et al. 1976; Frouin & Fiúza 1990; Haynes & Barton 1990; Relvas et al. 2007).

Contrary to the Iberian coast, the Northeastern African littoral, from Mauritania to Cap Blanc, receives great influence from the CaC. This current flows towards the Ecuador through coastal line until Cabo Blanco, and thereafter sorts out in to the ocean towards west of Cabo Verde islands (Hughes & Barton 1974).

The region of the Gulf of Guinea, on its turn, is subject to seasonal upwelling, influenced by local and remote trade winds and thermocline oscillations that come from the West Atlantic Ocean, and South and Northeastern Atlantic Ocean (Merle & Arnault 1985; Andrié et al. 1986; Fontaine et al. 1999; Wiafe et al. 2008). Two main coastal upwelling areas, at the northern and eastern boundaries of the Gulf, and one equatorial upwelling area, approximately south from the equator line and extending out to the Atlantic occur in this region from July to September, because the winds are more intense (Lefèvre 2009). The equatorial upwelling brings cold water, rich in nutrient and carbon, to the surface (Andrié et al. 1986). The rest of the year is a non-upwelling season, also known as the minor hydrographic season (Wiafe et al. 2008). This period is characterized by the low salinity due to the presence of the Inter Tropical Convergence Zone (ITCZ), high precipitations and a surface layer of warm fresh water mass, coming from NECC (Merle et al. 2013) which flows westwards, through the South Equatorial Current (SEC) (Andrié et al. 1986; Wiafe et al. 2008). The separation of both the low salinity Guinean water and the northern high salinity equatorial undercurrent cold water (Picaut 1983) is marked by a frontal area dominated by the South Equatorial Current.

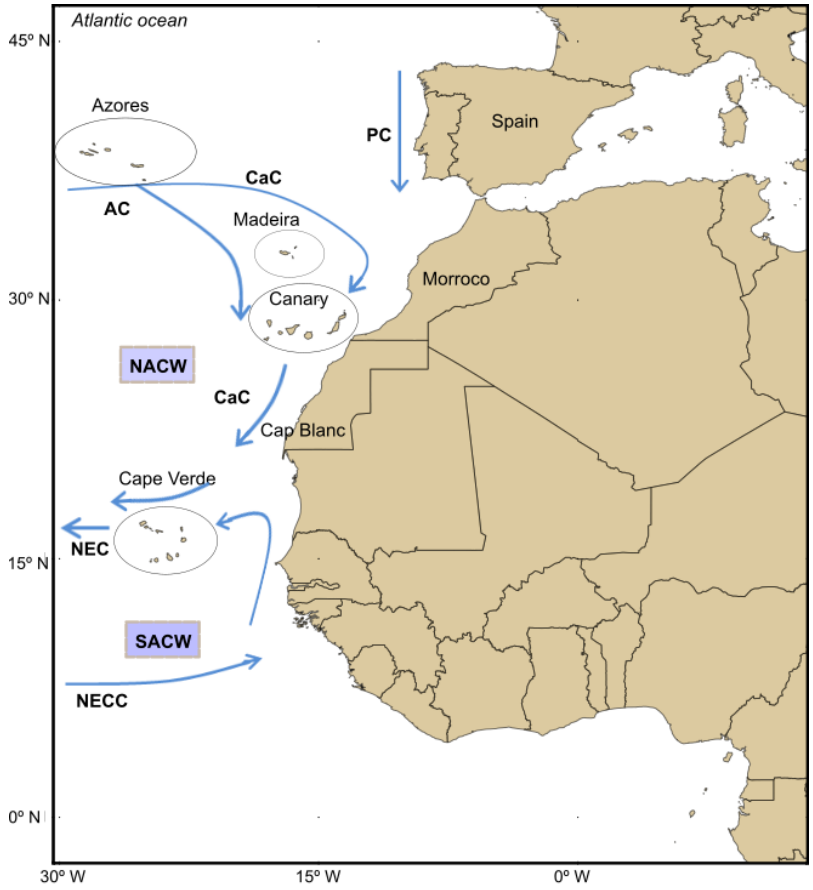


Figure 3 Northeastern Atlantic area and Macaronesian Archipelagos. Schematic representation of the North Atlantic gyre and major surface currents throughout the area. Major extant currents are represented by the blue arrows, from their average position and direction: AC – Azores current, CaC – Canaries current, PC – Portugal current, NEC – North equatorial current; NECC – North equatorial counter current. NACW – North-Atlantic central waters, SACW – South-Atlantic central waters (adapted from Roura 2013; Isabel & Pinto 2008).

1.3 BIOGEOGRAPHIC PATTERNS OF MARINE BIOTA IN THIS REGION

Biogeographic regions are often described based on the overlapping ranges of many species. The boundaries between these regions are usually the result of historical discontinuities (such as the ones caused by events like the Messinian salinity crisis or the Pleistocene glaciations cycles) (Krijgsman et al. 1999; Lambeck et al. 2002) or present-day environmental differences (such as currents, temperature or salinity) (Riginos & Nachman 2001). The analysis of the patterns of diversity and population structure of marine taxa allows the distinction between the effects of historic and present-day processes (e.g. Avise 1992; Grant & Bowen 1998; Provan & Bennett 2008). In the Northeastern Atlantic and Mediterranean regions there are several patterns emerging regarding the biogeographic affinities of their taxa.

The closure of the Tethys Sea, around 20 MY (Steininger & Rögl 1984), isolating the Atlantic and Indo-Pacific realms, as well as the one of the Isthmus of Panama (2.7–3.5 MY) (Keigwin 1982;

Coates et al. 1992; Collins et al. 1996) were the main events that started shaping the biogeography of this region (Avise 2000). This area became thus a transitional region between the tropics and boreal regions, whose climate has been very dynamic since the Pleistocene. While to a great extent it harbours warm and cold temperate species, it is also the northern limit of some tropical species and the southern limit of some boreal species (Pontarotti 2014).

More than 8500 species of macroscopic organisms have been reported for the semi-enclosed Mediterranean Sea (Bianchi & Morri 2000), representing 4-8% of the world's marine biodiversity – a value which is rather high for a basin representing only 0.82% and 0.32% of surface area and volume, respectively, of the world's oceans (Patarnello et al. 2007). The root of this high biodiversity is certainly in big part due to its complex geological history. Its present day biota is largely the result of colonization, mostly from the Atlantic Ocean (Almada et al. 2001; Domingues et al. 2005) and to a minor extent from the Red Sea. In fact, after an isolation and desiccation period that lasted about 0.5 million years – the Messinian salinity crisis (MSC) - re-flooding of the Mediterranean basin was possible because of the inflow of Atlantic waters through the newly opened Strait of Gibraltar. The MSC (5.96–5.33 MY) (Krijgsman et al. 1999) was characterized by the evaporation and subsequent increased salinity of very large parts of the Mediterranean Sea, which turned into shallow, hypersaline, isolated lakes, causing a mass extinction among its marine species, as well as many marine taxa speciation events (e.g. Duggen et al. 2003; Carreras-Carbonell et al. 2005; Palero et al. 2008). It thus drove the pre-existing Indo-Pacific biota to extinction, with the few exceptions of taxa of Miocene origin that have survived in shallow-water refuges, like for example killifishes (Hrbek & Meyer 2003). Interestingly, in these surviving older taxa, it is still possible to see the genetic signatures of the first events shaping the Mediterranean basin, such as the generation of the Siculo-Tunisian Strait, which separated Western and Eastern Mediterranean lineages coalescing back in the Mesozoic period (Rastorgueff et al. 2014). This shallow saddle between Sicily and Tunisia, has been proposed to be a major historical barrier (Nikula 2003; Rastorgueff et al. 2014).

The geographic history of the Mediterranean Sea suggests that species communities contemporarily inhabiting the Mediterranean waters can be divided into the following biogeographical categories: (i) temperate Atlantic-Mediterranean species; (ii) cosmopolitan/panoceanic species; (iii) endemic species, including palaeo-endemic (Miocenic) and neo-endemic (Pliocenic) species; (iv) subtropical Atlantic species (interglacial remnants); (v) boreal Atlantic species (glacial remnants); (vi) Red Sea invasive (Lessepsian) species entering through the Suez Canal; and (vii) eastern Atlantic invasive species (Bianchi & Morri 2000) identified 10 biogeographical regions in the Mediterranean according to the relative abundance of each of the aforementioned categories. Transition from one biogeographical zone to another results from a

combination of geological (orogenesis and hydrogeology), physical (present-day hydrography, coastal and seabed profile) and biological factors (species biology and evolutionary history), all contributing to shape the intra- and interspecific diversity (Figure 4).

As well as other areas of the globe, the late Pliocene and Pleistocene glaciations – the occurrence of major ice-ages interleaved with short and warmer interglacial periods - largely affected this northern hemisphere area and shaped the current's biogeographic patterns of its biota (Hewitt 2000; Hewitt 2011). The first major northern hemisphere glaciation probably occurred around 2.37 MY (Shackleton et al. 1984), and subsequent glacial cycles played an important role in shaping the geographic distribution of intraspecific diversity and promoting population divergence in both marine and terrestrial taxa. Sea-level changes were one of the most obvious impacts: during cold periods great quantities of water became trapped in the Polar Regions, causing major sea-level drops around the world. (Record & Change 2005) reports sea levels lower than 120m below present sea-level during LGM. These sea-level drops caused the emergence of land masses that in some cases (Mediterranean area) restricted or totally prevented the gene-flow between previously connected areas, causing population divergence. Further, glacial cycles also deeply changed sea surface water temperatures causing great changes in intertidal habitats and shifting species distributions, by locally extinguishing and/or deeply reducing many populations. The lowering of the sea-level deeply changed the Atlantic and Mediterranean shorelines configuration, exposing shallow areas of (nowadays) underwater seamounts as well as the English Channel, the North Sea, the Messina Strait, and the northern area of the Adriatic Sea. Slightly deeper, Gibraltar (248m) and the Sicily (330m) Straits remained opened waterways during glacial periods, allowing water exchanges between the Atlantic and the two Mediterranean basins, although through a much-reduced cross-section. During these glacial cycles, the Mediterranean presented a milder environment with colder (but ice-free) conditions and even warmer water pockets, especially in the south (Thiede 1978; Francisco et al. 2014a). Hence, the role of the Mediterranean as a potential glacial refugia has been profusely suggested and reported (Olsen et al. 2004; Maggs et al. 2008), although it's far from consensual (Patarnello et al. 2007).

In the Atlantic shore, these climate oscillations led to drastic changes in sea surface temperature (SST) (Lambeck et al. 2002) and in the location of the polar front at the Western Portuguese coast during glaciations. At the last glacial maximum (LGM), the European ice sheet extended southwards with the polar front being at the Bay of Biscay (Zaragosi et al. 2001), or even far south at the latitude of Lisbon, according to more conservative works (CLIMAP Project Members et al. 1984; Dias 1997). Species ranges became fragmented, with the populations of some species driven south and being restricted to isolated refugia during glacial maxima (Hewitt 2000). Although these refugia were predominantly located in the south, where the climate was relatively buffered against glacial cycles, some populations are thought to have persisted in northern refugia (for a review see Maggs et al.

2008). Populations of different species vary in their phylogeographic patterns, and the picture of how European marine fauna evolved during the Pleistocene is still incomplete. Yet, in a recent review based on the teleost fishes and crustaceans inhabiting Northeastern Atlantic and adjacent seas, (Francisco et al. 2014b) a third of the fish species analysed had a peak of genetic diversity at their southern distribution limit and a substantial fraction of the species showed little or no latitudinal genetic variation. Further, genetic structure varied widely among species, from cases where Atlantic, North Sea and Mediterranean seem to correspond to distinct populations, to cases where no structure could be detected across their entire range. Regarding the putative age of the faunal assemblies, the origin of populations (especially in the Atlantic) was dated mainly from the Lower to Middle Pleistocene. Populations with origin estimated after the LGM occur primarily in the North Sea (for cold water species) and Macaronesia, particularly Azores (for warm water species) (Francisco et al. 2014b). For thermophilic species, data seems to support the Azores colonization from Madeira (e.g. Santos et al. 1995) with more migrants detected in this than in the reverse direction (Domingues et al. 2008). Madeira is in turn biogeographically connected in the Canaries, and the Canaries to Mauritania, and it is argued that this colonization route (together with SST) could explain why there are several fish species present in the tropics and in the Macaronesia islands, but not in Europe (Francisco et al. 2014a). Actually, it is known that despite the fact that dominant average circulation reaches the Macaronesian archipelagos from the west, their marine littoral fauna share affinities with the Mediterranean and western European and African coasts (Francisco et al. 2014a). This is probably related to the fact that this multi-branch oceanographic system also contains coastal upwelling, filaments and eddies, and seasonal variations in the mean directions adding even more complexity to the system (Santos et al. 1995). Further, for some taxa, Azores also seems to have acted as a glacial refugium (Chevolot et al. 2006; Domingues et al. 2007; Maggs et al. 2008). According to several authors, the drop in SST throughout glaciations in the Azorean archipelago was only moderate (Morton & Britton 2000), allowing for the survival of temperate organisms.

For long, the Atlantic-Mediterranean seaway was thought to be a potential barrier for gene-flow. However, as phylogeographic studies accumulate, it is showing to be effective only for some species, while others, often closely related, apparently cross it without restrictions (for reviews see (Patarnello et al. 2007; Kettle et al. 2011; Francisco et al. 2014a).

Overall, if one is to sum current evidences for major phylogeographic breaks throughout this area (and in their border regions), they would be: 1) the separation from the Atlantic and Mediterranean (occurring either at the Gibraltar Strait and/or at the Almeria-Oran Front – which per-se define an additional area); 2) the one between the Baltic and the North seas (Olsen et al. 2004); 3) the Siculo-Tunisian Strait (separating West and East Mediterranean); 4) the Bosphorus-

Dardanelles sill (separating the Black Sea from the Mediterranean); 4) the Pelagosa sill (dividing the central Adriatic Sea from the Eastern Mediterranean) and 5) the Peloponnese Peninsula, acting as a barrier between the Aegean-Ionian Sea and adjacent Mediterranean waters (Patarnello et al. 2007). The temporal Balearic Front (BF), and the Ibiza channel (IC) have only recently been studied, but have shown to be potential barriers for genetic flow in littoral fish species and the red gorgonian (Galarza et al. 2009; Schunter et al. 2011; Mokhtar-Jama et al. 2011).

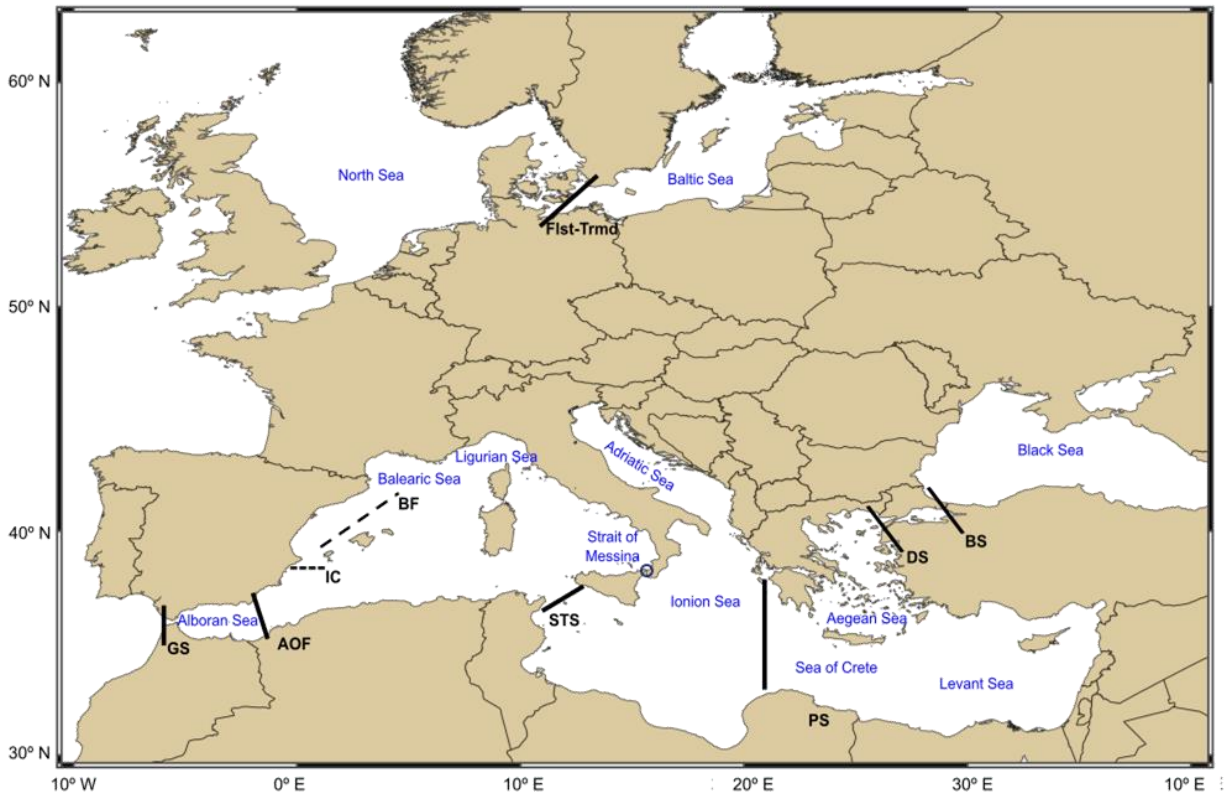


Figure 4 Atlantic and Mediterranean major biogeographic discontinuities and biogeographic sectors (following (Bianchi & Morri 2000; Nikula 2003; Johannesson & André 2006). Biogeographic sectors within the Mediterranean appear in blue on the map (in Green the respective sub-regions). Thick black lines represent major well-defined historical and present day discontinuities (Flst-Trmd – Falsterbo-Travemunde, GS – Gibraltar Strait, AOF – Almeria-Oran Front, STS – Siculo-Tunisian Strait, PS – Peloponnese Strait, DS – Dardanelles Strait, BS – Bosphorus Strait) while dashed lines represent weaker but potential discontinuities (BF – Balearic Front, IC – Ibiza Channel).

1.4 "OPISTHOBRANCHS" DIVERSITY AND PHYLOGENETIC RELATIONSHIPS

1.4.1 Opisthobranchia: a convoluted taxonomic history

Opisthobranchs are a large and diverse group of specialized complex gastropods (Figure 5) within Heterobranchia. "*Opisthobranch"* means "gills behind" (the heart), in contrast to Prosobranch (gills in front of the heart), as these animals are characterized by two pairs of tentacles

and a single gill behind and to the right of the heart. Previously united in the subclass Opisthobranchia (in green in Figure 5), these are now known to be paraphyletic respect to the land snails and slugs (Pulmonata) plus a few other groups (Bouchet et al. 2005; Schrodl et al. 2011, Jorger et al. 2011; Wagele et al. 2014). It is now clear that both Opistobranchia and Pulmonata (in yellow in Figure 5) are informal and obsolete groups. Traditional Opisthobranchia have been allocated with Lower Heterobranchia, lower Euthyneura, Euopistobranchia, and Panpulmonata (Dinapoli & Klussmann 2010).

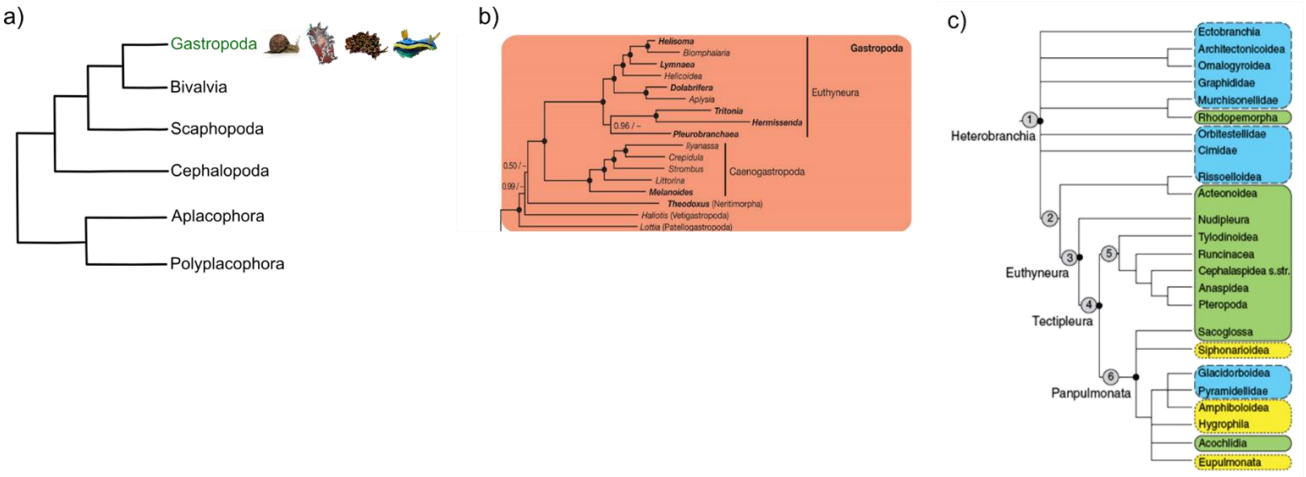


Figure 5 a) Cladogram of the most recent deep Molluscan relationships, based on transcriptomic data, as inferred by Kocot et al. (2011); b) Phylogenetic relationships within Gastropoda. Euthyneura is here represented within a major group Heterobranchia (adapted from Kocot et al. 2011); c) (Wägele et al. 2014) presents a consensus tree for Heterobranch's clades, with special focus in traditional Opisthobranch clades. In green are groups traditionally known as the Opisthobranchs, while in blue and yellow are the lower Heterobranch and Panpulmunata, respectively. Recently the rise of Euopisthobranchia (by Jörger et al. 2010) includes five (Cephalaspidae, Pteropoda, Anaspidae, Runcinacea and Umbraculoidea) of the traditional Opisthobranchia groups (in green). Numbers indicate important putative apomorphies (see Wägele et al. 2014 for detailed descriptions of apomorphic characters).

Heterobranchia are characterized by absence of shell (or presence of a reduced one), allowing elaborate forms of other body parts like the head, foot or the mantle (Grande et al. 2008; Schrödl et al. 2011). It is further supported by other autopomorphies such as pigmented mantel organ, a medial position of the eyes, lack a true ctenidium, simple esophagus, distinctive sperm ultrastructure and the most important, a sinistral larval shell at their planktotrophic veliger stages (Haszprunar 1985; Ponder & Lindberg 1997). The phylogenetic relationships within the group only now start to become more consistent (Jörger et al. 2010; Schrödl et al. 2011; Wägele et al. 2014). Within Heterobranchia, the clade Euthyneura (Figure 5) includes the "traditional" groups "Opisthobranchia" and "Pulmonata" (Grande et al. 2004; Klussmann-Kolb et al. 2008), and has consistently been supported by 'standard markers' such as nuclear 18S and 28S rRNA and mitochondrial 16S rRNA and COI (Klussmann-Kolb

et al. 2008; Dinapoli & Klussmann-Kolb 2010; Jörger et al. 2010; Dinapoli et al. 2011). Nonetheless Euthyneura has recently been rejected by new transcriptomic data (Zapata et al. 2014). Euopistobranchia (group 5 in Figure 5) defines now the group of five monophyletic Opistobranch lineages, leaving out some "traditional" Opisthobranchs including the Nudipleura, Sacoglossa and the Acochlidiacea (Schrödl et al. 2011; Zapata et al. 2014).

According to Jörger et al. (2010) Euthyneura already occurred in the Palaeozoic, with major radiations occurring throughout the early Mesozoic.

To facilitate references to older works I'll still use throughout this thesis the term Opistobranchia in its "broad sense".

1.4.2 Opisthobranchs diversity and ecology

The origin of the Opisthobranchia is estimated between the Permian and the Triassic, with a rapid radiation of the higher taxa in the early to mid-Mesozoic (Klussmann-Kolb et al. 2008; Jörger et al. 2010; Stöger & Schrödl 2013). This is probably the case of Nudipleura, a group including Nudibranchia sea slugs. Yet, as the fossil record of Nudipleura is problematic due to lacking hard body parts, no reliable fossil record for Nudibranchia (true sea slugs) (Schrödl et al. 2011) is yet known. Opisthobranchs have a great diversity of body shapes and striking colour patterns, being great models for some topics of biological research (Todd et al. 2001) including neurophysiology (Gosliner et al. 2008), life-history traits (Wagner et al. 2009), ecology (Wägele & Klussmann-Kolb 2005; da Cruz et al. 2011), chemo-ecology (Avila 1995; Fontana et al. 2001; Blunt et al. 2012), pharmacology (Kijjoa & Sawangwong 2004; Fontana et al. 1999), foraging and/or defence adaptive mechanisms (Greenwood & Mariscal 1984; Rumpho et al. 2001; Frick 2003; Martin 2003; Aguado & Marin 2007), feeding adaptations and evolutionary trends (e.g. Wagner et al. 2004).

Opisthobranchs lack structural defences (i.e. shell), and live in habitats full of potential predators. Despite their apparent fragility, they take advantage on sophisticated defence and feeding mechanisms that allow them to exploit prey hardly used by other species, while at the same time predators learn to avoid them (Tullrot, 1994; Wagner & Daniel, 1998). Such mechanisms include functional hard structures (eg. gizzard plates); physiological (such as the incorporation and usage of intact chloroplasts from algal cells, in Sacoglossans, or as the storage of intact cnidocysts - cells with one giant secretory organelle from Cnidarians by Aeolidoidea); but also biochemical, such as synthesizing or transforming ingested toxic compounds (Faulkner & Ghiselin 1983; Wägele 2004; Wägele et al. 2006; Cortesi & Cheney 2010). They feed on a wide variety of prey such as tunicates, bryozoans, hydrozoans, sponges, crustacean and other Opisthobranchs (McDonald & Nybakken 1997; Wollscheid-Lengeling et al. 2001; Carriglio et al. 2004) and their diet has long been discussed

as a factor leading to adaptive radiation and speciation in many groups (Rizzuti et al. 2004; Cimino & Ghiselin 1999; Mikkelsen 2002; Wägele 2004; Wägele & Klussmann-Kolb 2005).

The particular ability to predate upon chemically defended species, and efficiently storing secondary metabolites in special organs (mantle dermal formations; MDFs) located in the mantle characterizes the most speciose family of Nudibranch dorids, the Chromodorididae (Figure 6). These Nudibranchs are highly stenophagus – they specialize on feeding on particular species or groups, and are capable of biotransformation (Cimino et al. 1993; Fontana et al. 1999) and *de novo* synthesis of the ingested toxic compounds (Cimino et al. 1983; Cimino & Ghiselin 1999). While some authors render MDFs the role of storage of biochemical compounds from sponges in Chromodorids, and further claim them as key characters in their evolution and clues to their ecologic adaptive radiation (Gosliner 2001; Wägele 2004), others show evidence that this apparent synapomorphy is widely spread in other Opistobranchs, with distinct diet preferences (Wagele, 1997; Wagele, 2004, Wägele et al. 2006). Characteristic of this family, and resulting from storage of toxic substances, are the conspicuous patterns and bright aposematic colours, which are interesting models for co-evolutionary studies on predator-prey associations (Gosliner & Behrens 1990; Rudman 1991; Pfennig et al. 2007; Pfennig & Mullen 2010). Nevertheless it is still uncertain if the use of new sources of food, followed by morphological adaptations, was the key factor in Opisthobranchs radiation or vice versa.

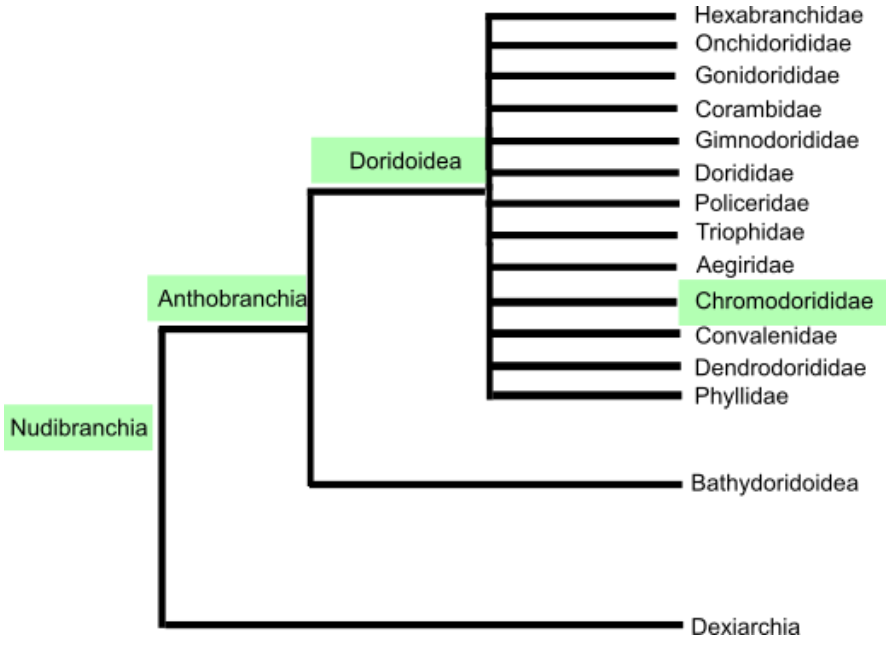


Figure 6 Most recent phylogenetic status of Nudibranchia sea slugs based on morphology data (cladogram with higher taxa names used) (adapted from Wägele & Willan 2000; Wollscheid-Lengeling et al. 2001).

1.4.3 Chromodorididae

The colourful Chromodorids are soft-bodied marine gastropods that completely lost their shell in the adult stage and accumulate chemical (mostly toxic) products obtained from their prey (e.g. cnidarians and sponges), in a series of defensive mantle glands, as their defence mechanisms (Conabio 2006). Chromodorids now include 16 genera and stand for the most speciose family within Nudibranchia, with over 300 described species (Gosliner & Draheim 1996). The group has representatives across nearly all world's oceans, with the highest diversity in the Indo-Pacific basins (Turner & Wilson 2008).

The family Chromodorididae was first described by Bergh (1891) based on the denticulate radicular teeth and unarmed penis. The first comprehensive revision of the family, though, was performed by Rudman (1984), using reproductive characters, external colouration and radular morphology.

For many years relationships within Chromodorididae were controversial (Rudman 1984; Rudman 1991; Ortea 1996; Valdés & Gosliner 1999; Gosliner & Johnson 1999; Turner & Wilson 2008; Rudman & Bergquist 2007). Several studies focused on exploring the monophyly of Chromodorididae, their sister group, mainly respect to *Cadlina spp.*, as well as the relationships within. The majority of the molecular evidence and re-evaluated morphological data suggests *Cadlina spp.* should not be considered a member of the Chromodorididae (Turner & Wilson 2008; Johnson 2010), and they are now considered to be a separate family (Cadlinidae), remaining Chromodorididae being monophyletic. Recent studies using mtDNA data (Johnson & Gosliner 2012) and including several representatives of all currently described genera (14) within Chromodorididae, confirmed its monophyly and its sister group relationship with the Actinocyclidae. They further indicate that all the traditional Chromodorididae genera are either non-monophyletic, or render another genus paraphyletic, and that a few monotypic genera were nested within other clades, leading to a major re-organization of species-level relationships within this group, which is now considered to be composed of 17 lineages (genera), 13 of which are very well supported (Figure 7).

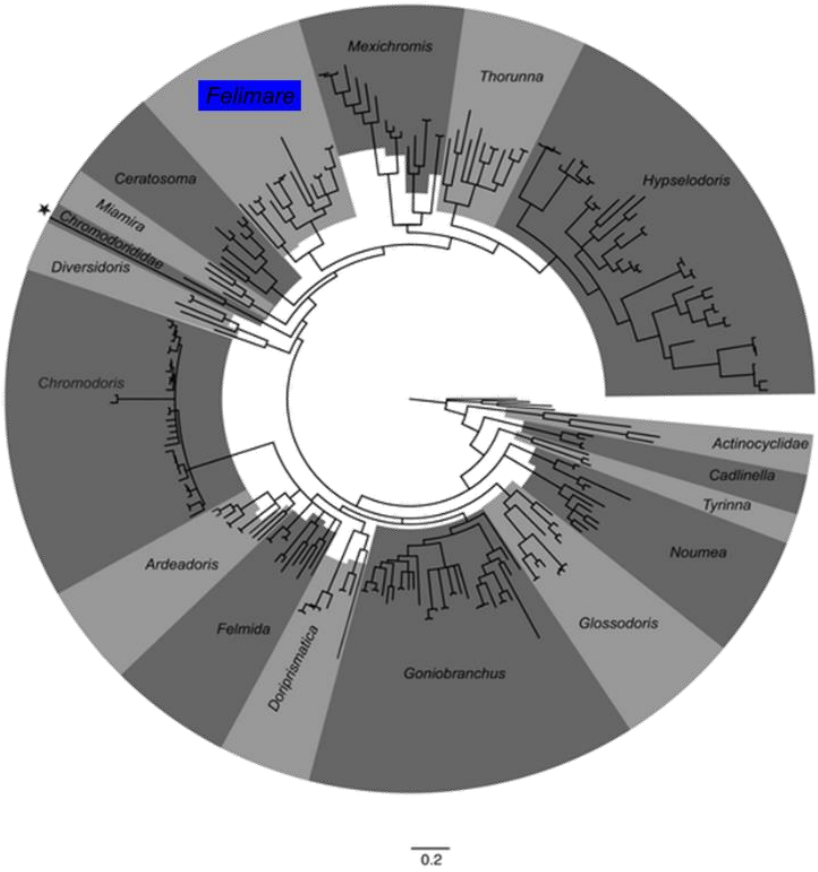


Figure 7 Current phylogenetic relationships within Chromodorididae (tree from Johnson & Gosliner, 2012). *Felimare* is now the monophyletic group highlighted in blue, while remaining Hypselodoris (on the right) are clearly not their sister-taxa. Alternating greys remaining *Hypselodoris* (on the right) are clearly not their sister-taxa. Alternating greys represent distinct genera within Chromodorididae, with the current genera.

This new phylogeny strongly supports a biogeographic scenario where Chromodorids diversified rapidly from the tropical Thethyan realm, a pattern that has also been found in other gastropod groups (Schmittner et al. 2004; Latiolais et al. 2006; Malaquias & Reid 2009; Johnson & Gosliner 2012). The Chromodorids were likely widely distributed and different lineages diversified in isolation following vicariant events. Most of the members of the group are still found in the Indo-Pacific region, while some diversified both in the Atlantic-Mediterranean and eastern Pacific regions. Evidences for some trans-Pacific dispersals, and from there to Western Atlantic exist, but need to be looked into in more detail. Yet, overall, relationships between Chromodorididae clades are unresolved, and current phylogenetic estimates are based only on mitochondrial genes. The addition of slowly evolving nuclear genes is fundamental in helping to resolve poorly supported nodes.

1.4.4 Genus Felimare

One of the genera that was found to be paraphyletic within Chromodorididae was *Hypselodoris (sensu)*, now subdivided in *Hypselodoris* and *Felimare* (Johnson & Gosliner 2012). The *Felimare* clade now includes all eastern Pacific, Atlantic and Mediterranean species of *Hypselodoris* and two species of *Mexichromis* - *M. porterae* and *M. kempfi* from the eastern Pacific and Caribbean respectively, and was found not to be sister taxa of remaining *Hypselodoris* (Johnson & Gosliner 2012). There are thus two eastern Pacific and Atlantic splits in this clade: the eastern Pacific *F. porterae* and Caribbean *F. kempfi* are potentially sister species, and are sister to another larger clade of eastern Pacific, Caribbean and eastern Atlantic (and Mediterranean) *Felimare* species. This pattern seems to suggest that *Felimare* represents an invasion of eastern Pacific from the Indo-Pacific region, and that species within *Felimare* result from eastern Pacific – Atlantic vicariance and transatlantic colonization(s), but further sampling is needed to further disentangle biogeographic patterns within this clade (molecular data is available for less than 10 species of the 42 described.

In the Atlantic-Mediterranean region a group of species within the genus *Felimare*: *F. villafranca*, *F. cantabrica*, *F. tricolor*, *F. fontandraui, F. bilineata*, *F. orsinii* and *F. picta* which are known as the 'blue mimetic circle' (Haber et al. 2010; Blunt et al. 2012). These species have attracted interest as models to study colour evolution, aposematic mimetism (Lindström et al. 2004); predator-prey interactions (da Cruz et al. 2011); biotransformation of prey toxic compounds (Cimino et al. 1993; Feliciano 2009) and reproductive ecology (e.g. Sánchez-Tocino, 2003; Coelho & Calado, 2010). The "blue mimetic circle" is an example of Müllerian mimicry (where all the mimic species have a defense mechanism). Underlying the obvious aposematism, is the presence of food-derived, feeding-deterrent terpenes in the mantle formation glands (MDFs), or simply in the mantle border of some species, most probably derived from sponges of the genus *Dysidea*, upon which they feed (Haber et al. 2010).

Felimare species are characterized by their dark-bluish background dyed by one or more lines and/or dots of white and/or yellow colours (Ortea 1996; Sánchez-Tocino 2003) (Figure 8).

Figure 8 a) Morphologic characterization of six adult species of "blue" *Felimare* (from Calado & Silva, 2012). For each species distinctive characters such as the rhinophores, dorsum, lateral banding and branchias are shown (adapted from (Calado & Silva 2012).

Exceptions to the typical coloration patterns are not rare in this group, generally varying with size of the individual and geographic location (Sánchez-Tocino 2003). For instance, *F. picta* from the Mediterranean coasts are light-blue pale, while *F. picta* from Azores islands are of dark-green and violet background (Ortea et al. 1996). Additionally, ontogenetic changes occur, for example in *F. fontandraui,* with colours becoming more intense, from white, to yellow and in some exemplars even orange, as the individual becomes more mature (Ortea 1996; Sánchez-Tocino 2003). Additional morphological aspects, such as the size and distribution of the mantle glands, and their large seminal receptacle (Gosliner & Johnson 1999) are also useful to distinguish these species from their Western-Atlantic congeners (Ortea 1996; Sánchez-Tocino 2003).

These species inhabit in partial sympatry and are typically found in the rocky bottoms (up to 30m bellow sea-level but mostly at shallow depths) of the Atlantic and Mediterranean marine coasts, including the Macaronesian archipelagos.

Little is known about the ecology of these organisms, except that they feed mainly upon sponges of genus *Dysidae* (the only exception being *F. orsinii*, which feeds upon the sponge *Cacospongia mollior*) (Ortea 1996; Sánchez-Tocino 2003). Similarly to other nudibranchs, these organisms are

simultaneous hermaphrodites with internal fertilization, releasing large amounts of benthic egg strings, with a general slow embryonic development (Rudman & Willan 1998) and short term free swimming veliger larvae (Coelho & Calado 2010). Exceptions are *F. villafranca*, which is a direct developer (Coelho & Calado 2010), and *F. orsinii*, whose reproductive strategy is still unknown.

The distribution of these species varies, from widely distributed - *F. picta*; inhabiting both sides of the Atlantic plus the Macaronesian islands - to more restricted ranges - *F. orsinii;* only present in the Mediterranean. Their approximate distributions are shown in Figure 9. Especially concerning some most widely distributed species (e.g. *F. picta* and *F. bilineata*), it has been hypothesized that they may in fact harbour distinct, cryptic species (eg. *Felimare lajensis*) described like *F. picta lajensis* by Troncoso *et al*. 1998 for Southwest Atlantic), (Ortea et al. 1996, Sánchez-Tocino, 2003). Additionally, some phenotypes of unclear affinities have been observed such as *F. malacitana* – a species described from Southwest Spain by Ortea et al. (1996), proposed to be related to *F. cantabrica* but bearing also resemblances to *F. bilineata*. Further, these species constitute interesting models for studies of phylogeography in marine waters as they 1) have limited mobility as adults, and therefore potential factors affecting patterns of genetic structure will be expected to partially act upon larval dispersal; 2) despite limited dispersal abilities, species are capable of colonizing vast areas and apparently cross major oceanographic barriers; 3) they present both planktonic and direct development reproductive strategies; 4) they are found conspicuously in the East Atlantic and Mediterranean rocky habitats inhabiting from shallow shores to deeper areas; 5) they have largely sympatric distributions, allowing for multiple comparisons across main biogeographic barriers.

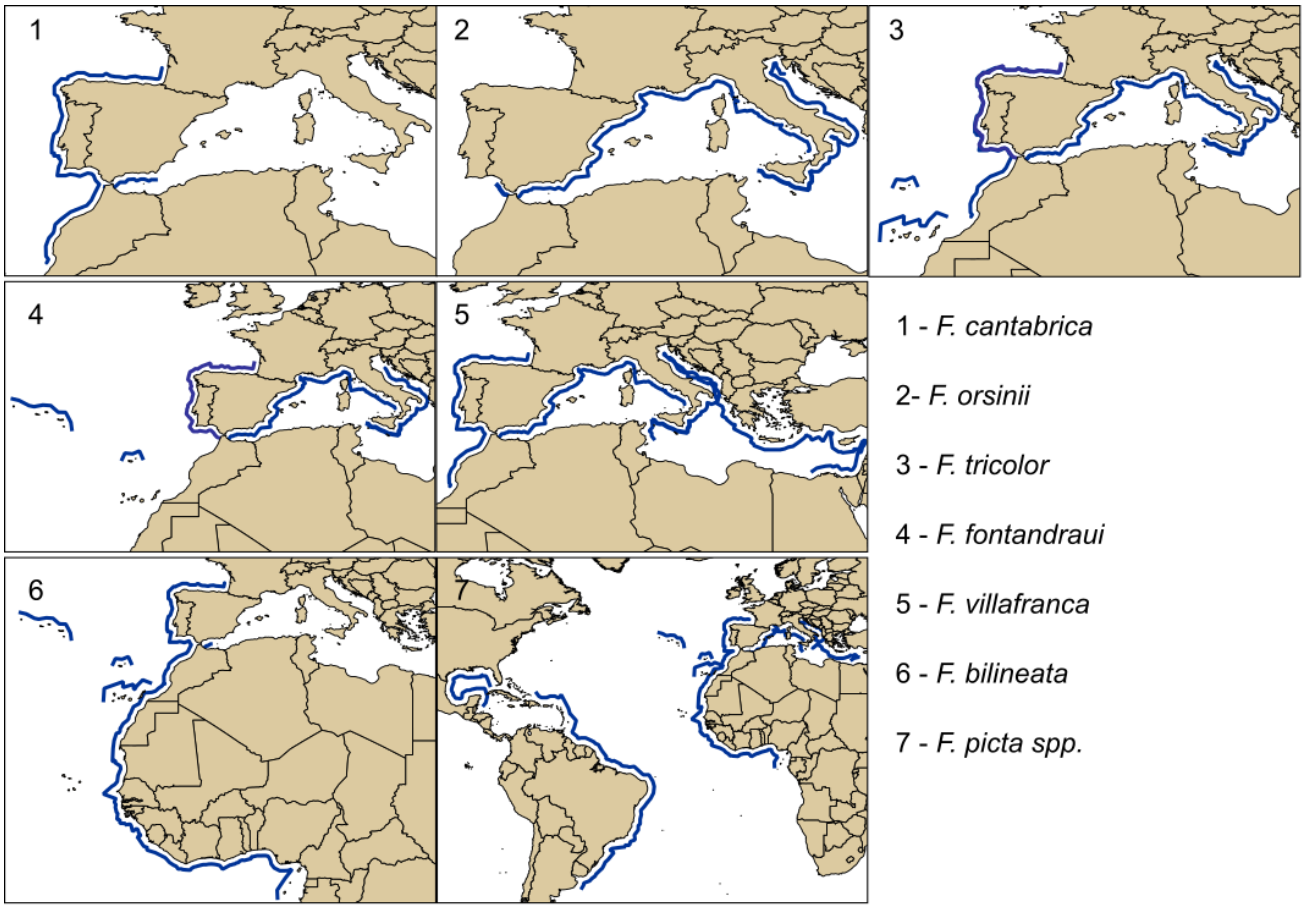


Figure 9 Approximate distribution (blue line) of the seven EAM species of the genus *Felimare* (based on (Ortea 1996; Sánchez-Tocino et al. 2014) Crocetta et al. 2015; Oskars et al. 2015).

1.5 THE LACK OF GENOMIC RESOURCES FOR MOLLUSCA

Over 95% of metazoan species are invertebrates and yet, few invertebrate genomes have been sequenced. For molluscs, a major invertebrate group, there is a surprising lack of genomic resources. In fact, non-insect, non-nematode (which have their own dedicated genome sequencing consortiums) invertebrates are recently being paid more attention (GIGA community of scientists, 2013), highlighting the fact that only a few genomes from a few groups are available. Gastropod molluscs, with approximately 62 000 listed species (the highest diversity Mollusca clade), and very old divergences are surprisingly under-represented. There are five currently finished/ongoing mollusc genome projects, mostly bivalves. *Aplysia californica*, the single gastropod genome sequenced, may be more than 250 MY old divergent from Chromodorid Nudibranchs such as our study group (Jörger et al. 2010). Thus, genomic data is a valuable tool to foster exploration and comparative analyses within this group.

1.5.1 The need of multiple markers for phylogenetic / phylogeographic inference

A major step for accurate delimitation of species and correct estimation of phylogeny is the choice of a useful marker or set of markers. Mitochondrial markers are classically the first to be used, because of their easy amplification and because they provide insights into species or populations relationships, as their maternal inheritance results in smaller effective population sizes and hence a faster fixation of neutral mutations, generally telling a recent dispersal history. At the same time mitochondrial DNA can also be useful at deeper phylogenies, because it's assumed lack of recombination (Ballard & Whitlock 2004). On the other hand nuclear markers represent independent (unlinked), not sex biased, and recombining sites; which provide powerful information respect to population sizes, ages of divergence, and speciation history (Edwards & Beerli 2000).

Phylogeneticists and phylogeographers have been working to collect multilocus data ever since a series of theoretical papers demonstrated that estimates of key demographic parameters improve as the number of loci increases (Edwards & Beerli 2000; Hey & Nielsen 2004; Felsenstein 2006; Carling & Brumfield 2007; Ilves et al. 2010). Also in phylogenetics, where the tradition of equating an estimated gene tree with the history of species divergence has for long predominated, the lessons learned in phylogeography about the inherent benefits of explicitly considering the stochasticity of genetic processes, led to the development of coalescent based approaches (Liu & Pearl 2007; Heled & Drummond 2008; Kubatko et al. 2009), with the focus changing to obtain data from multiple loci and multiple individuals per species or population, and using this data to obtain a direct estimate of the history of divergence - the species-tree – as opposed to focusing on the idiosyncrasies of individual gene trees (Carstens & Knowles 2007; Carstens & Richards 2007; Belfiore et al. 2008; Brumfield et al. 2008; Kubatko et al. 2009). Multigene analyses have now become the baseline in phylogeny and phylogeography, with both the number of loci and individuals dramatically increasing over the past few years (Garrick et al. 2010).

Yet, phylogenetic (and many phylogeographic) studies in molluscs are still mainly based on single or a few set of mtDNA and nuclear (mostly COI, 16S rDNA, trnL and trnV for mtDNA and 18S, 28S, and H3-histon for nuclear) (Grande et al. 2004; Malaquias & Reid 2009; Schrödl et al. 2011) while a few (mostly recent) studies include complete mitochondrial genomes (Medina et al. 2011) and RNA-seq data (Kocot et al. 2013; Zapata et al. 2014; Williams et al. 2014). In fact, recent analyses of multiple nuclear protein-coding genes yielded well-supported topologies, and further convincing frameworks of mollusc evolution, development and anatomy (Kocot et al. 2011; Smith et al. 2011; Vinther et al. 2012). A major pitfall in these studies is that only deep molluscan relationships are reviewed, and yet no study has focused on resolving relationships between lower-level taxa.

1.5.2 Next-Generation sequencing for phylogenetics and phylogeography

Next Generation Sequencing (NGS) is nowadays the standard technology for the obtention of large-scale data for comparative genomic analysis (Künstner et al. 2010), and has made the acquisition of multi-marker datasets highly feasible (Thomson et al. 2010). New approaches allow highly covered, fast and easy reconstruction of small genomes (Groenenberg et al. 2012), as well as a representative part of whole genomes and transcriptomes (e.g. Hou et al. 2011; Zhang et al. 2012), overcoming laborious techniques in the laboratory (J. A. Terrett et al. 1996; Grande et al. 2008; Medina et al. 2011). Moreover high throughput sequencing allows us to obtain all kinds of molecular markers, and in a large number (e.g. SNPs, SSRs, coding and non-coding regions) (Davey & Blaxter 2010; Hohenlohe et al. 2010; Hou et al. 2011; Zhang et al. 2012). Techniques of genome reduction, such as restriction-based, target enrichment or transcriptome sequencing are particularly useful when the goal is a matrix of loci/variable positions for phylogenetic and phylogeographic studies (Mccormack et al. 2013). Choosing a suitable technique though, highly depends on the number of loci desired, bioinformatic capability available, and whether a specific set of loci of known function is desired (Rocha et al. 2013).

In molluscs the application of high throughput sequencing has been progressively increasing, with next generation sequencing data being recently used to confirm both old (Euthyneura) and newer (e.g., Nudipleura, Tectipleura) groups (Zapata et al. 2014). At the same time, Williams et al. (2014) challenged the idea that mitochondrial genome data was unable to resolve deep relationships within Mollusca (e.g. Stöger & Schrödl 2013), using powerful new models of amino acid evolution, and increased taxon sampling, from high throughput data.

1.5.3 Transcriptome sequencing and marker obtainance

Transcriptome sequencing is an efficient means of genome reduction, and especially useful in groups where other “genomic” information is inexistent (McCormack et al. 2013). As the name indicates, RNA-seq is a next-generation sequencing technique that only targets transcripts. It begins with RNA extraction, followed by reverse transcription of RNA into cDNA, and DNA sequencing for the organisms/group/tissue of interest (Rocha et al. 2013). It offers notorious advantages over genome data, from their tractable size (significantly smaller than genomes) to facile assembly with intuitive software (Zerbino & Birney 2008; Simpson et al. 2009; Shipp 2010; Henschel et al. 2012). Moreover, the presence of fewer repetitive elements in transcripts, and the availability of functional information at public databases makes their assembly easier (Grabherr et al. 2011).

Hundreds of orthologous genes can thus be recovered without necessarily having an available genome of a closely-related species, enabling robust phylogenomic analysis. Loci matrixes built directly from RNA-seq data have been used to resolve deep mollusc phylogenies (Kocot et al. 2011; Smith et al. 2011) and the enigmatic position of Myzostomida in the tree of life (Hartmann et al. 2012), for example. They have also been used for lower-level phylogenies, either using assembled transcripts directly (Hittinger et al. 2010), or for development of population-level amplicon markers (Schultheis et al. 2014), as well as having being mined for SNP's in a wide variety of species (e.g. Barbazuk & Schnable 2011; Geraldes et al. 2011). On the other side, the applicability of using directly the transcripts for inference of shallow phylogenies seems not so well explored.

Further, transcriptomes also potentiate analyses at functional level, such as gene discovery and expression (Domingues et al. 2008; Clark et al. 2010), microRNA and piRNA detection (Martin & Wang 2011), detection of loci under selection (Elmer & Meyer 2010), and their functional annotation (Schwarz et al. 2009; Rocha et al. 2013), with important downstream implications. For example, by sequencing ten transcriptomes across five invertebrate phyla, Riesgo et al. (2012) recently shown how effectively low coverage transcriptome data can recover gene-sequences, compared to available genomes of closely related species, and further provided evidence for the existence of multiple paralogues in all phyla and in nearly all gene pathways, including in housekeeping genes that are traditionally used in phylogenetics for their purported single-copy nature.

1.5.4 Optimizing de-novo assembly of short-read RNA-seq data for phylogenomics

Using short-read RNA-seq data either directly for phylogenomic inference, or for marker survey, has its unique challenges, both analytical and computational: error, incompleteness and redundant assembled transcripts (due to skewed expression and/or the assembly of different isoforms of the same gene) cause large amounts of missing data in any aligned matrix (Zhao et al. 2011; Yang & Smith 2013).

As the study of non-model organisms becomes more and more popular, *de novo* transcriptome assemblers are also becoming more common, and exist on a variety of software packages, each one with their own particularities for assembling and filtering data (Robertson et al. 2010; Yang & Smith 2013). Choosing the best assembler for the intended downstream applications is thus of utmost importance (Yang & Smith 2013). Trinity, for example, is a *de novo* assembly package for short-read RNA-seq data that is user-friendly, and works well both on small and large datasets across various conditions, recovering full-length transcripts with good statistical confidence. Conversely it has been found to be time- and memory-consuming, and standard parameters cannot be changed when compared to other assemblers like Oases or Trans-ABySS (Zhao et al. 2011).

Several steps are to be taken for preparing short read RNA-seq data to phylogenomic inference. Besides the standard pre-assembly steps (to eliminate low quality reads and remainder primer and barcodes from flanking regions), some post-assembly filtering steps are needed, especially focused on diminishing the amount of redundancy in the assembly. The required post-processing may differ depending on the assembly methodology used (Yang & Smith 2013).

Basically, most transcriptome assemblers are “hierarchical”, resulting in a set of “loci” or “components”, consisting of one or more “transcripts” or “isoforms”. Biologically a locus or a component can each contain one gene or several paralogs, and a single gene can have fragments distributed among multiple loci or components. All published *de novo* transcriptome assemblers are optimized for building references for comparing gene expression levels, identifying splice variants, and determining gene-fusion events (Robertson et al. 2010; Grabherr et al. 2013). For phylogenomic purposes however, only one representative transcript per gene is required. Splice variants are not only unused as they complicate the detection of true paralogs for the construction of the phylogenomic matrix (Yang & Smith 2013). Common strategies to deal with this problem are choosing the isoform with the highest geometric mean read coverage (Smith et al. 2011), the highest (overall) coverage or the longest one (*trinity* online material). Another issue in *de novo* transcriptome assembly is the creation of chimeras (either resulting from the misassembly of short-reads or PCR-induced recombination during library preparation – false chimeras – or being real biological products from gene-fusion or trans-splicing events).

Yang & Smith (2013) recently examined the extent of these problems in several *de novo* assemblers, providing some strategies for optimizing transcript choice for phylogenomic analyses, that were mostly followed along this work. Shortly, they involve: 1) selecting highly covered isoforms for reducing redundancy; 2) further merge “loci” based on overlap and percentage of identity; and 3) blast against reference coding sequences for chimera removal. The importance of this last step depends on the assembler used, as the type and frequency of chimeras varies dramatically among assembly strategies.Overall, at current standard sequencing depths, *de novo* assembly of short read RNA-seq data is shown to be capable of recovering up to half of the total expressed genes to more than 200 bp, and that after appropriate filtering steps, the products of most of the currently used assemblers result in clean assemblies with lower redundancy and chimera rates.

1.6 THESIS AIMS

A group of Chromodorididae sea slugs, recently rised to the genus level - *Felimare* - inhabits the Eastern Pacific, Atlantic and Mediterranean waters being particularly abundant and known for their chemical defensive strategies and beautiful coloration patterns that fascinate the curious divers. Problems of identification in this group exist potential cryptic species have been suggested based on morphological data, although have never been confirmed. Moreover, no reliable estimate exists of their phylogenetic relationships. Because to resolve phylogenetic relationships between closely related species most often requires a considerable amount of multilocus data, goals of this thesis were two-folded:

1) to obtain and use mtDNA data to infer their mtDNA phylogeny and explore phylogeographic patterns of the Eastern Atlantic-Mediterranean species of the genus, and;

2) to generate RNA-seq data and perform *de novo* transcriptome assembly of representative species within the genus, so they can be later mined for nuclear markers to be employed on future phylogenetic inferences.

More specifically, goals of this work were:

1) To Characterize these species phylogeographic patterns;
2) To use mtDNA to infer the phylogeny of the group;
3) To access the quality of these newly obtained transcriptomes, annotate them, and compare them with published ones of other mollusc species;
4) To obtain a set of ortholog loci that can be further used for nuclear markers development.

Chapter 2 Material & Methods

2.1 MITOCHONDRIAL DNA PHYLOGENY AND PHYLOGEOGRAPHY OF ATLANTIC-MEDITERRANEAN SPECIES OF *FELIMARE*

2.1.1 Sample collection, DNA extraction, amplification and sequencing

Sample collection was conducted by standard scuba diving in open sea between 2008 and 2014, both by our team and collaborators, across different seasons. Samples were collected at several points along the Iberian and Mediterranean shores, plus some Atlantic islands, trying to cover as much as possible the distribution of the species of interest in this region (Figure 10). Specimens were identified by external morphological characters (mainly dorsal coloration patterns), in agreement with Ortea et al. (1996). Dorsal and dorso-lateral photos of all individuals were taken whenever possible. All specimens were preserved in RNAlater® and 99% ethanol (a tissue aliquot).

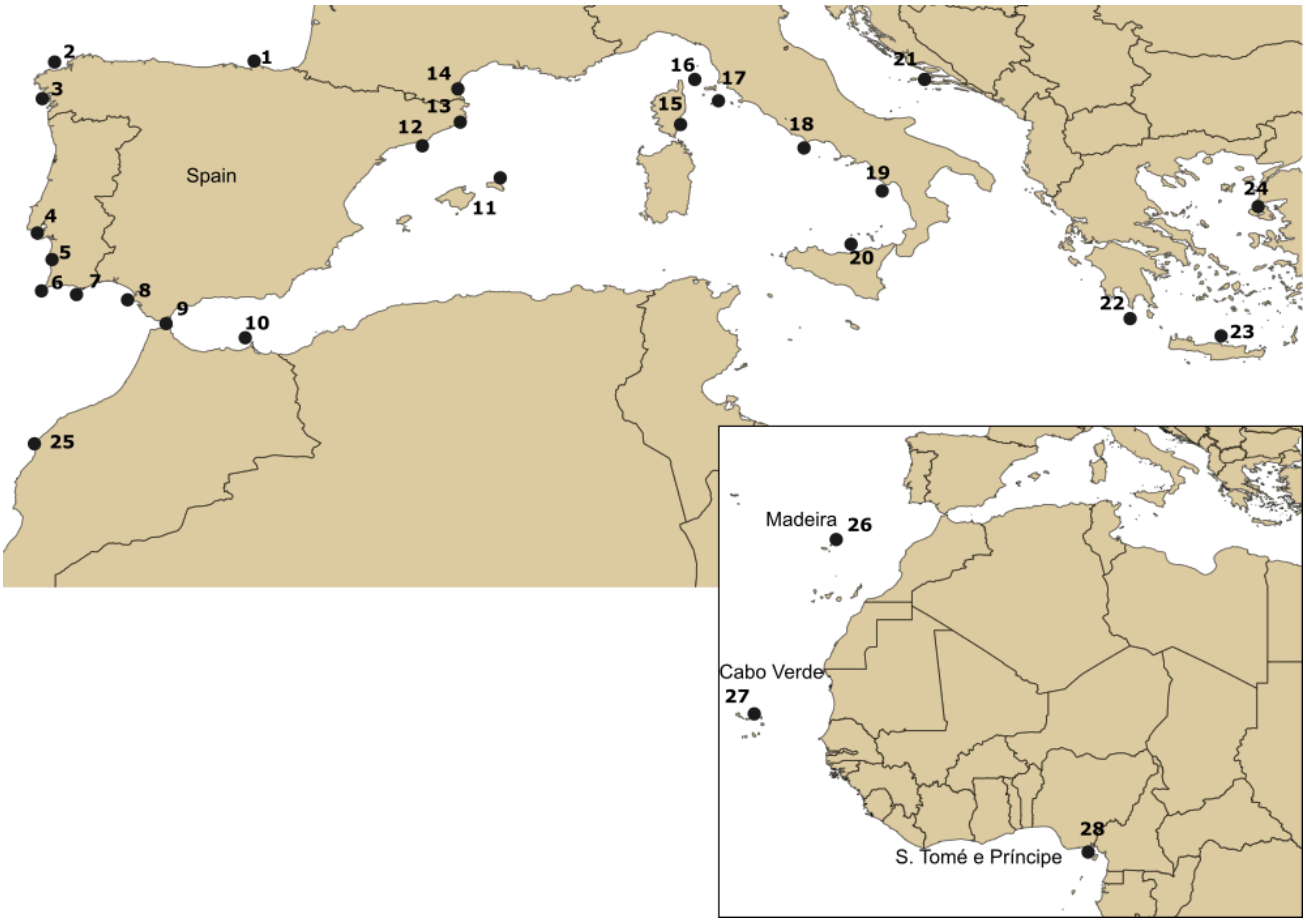


Figure 10 Map showing the sampling localities. Black dots represent localities where samples were collected: 1 San Sebastián; 2 Ferrol; 3 Vigo; 4 Sesimbra; 5 Alentejo; 6 Sagres; 7 Ria Formosa; 8 Cádiz; 9 Gibraltar, 10 Chafarinas; 11 Menorca island; 12 Barcelona; 13 Girona;

14 SW France; 15 - 20 Italy; 21 Kotor bay, Montenegro; 22-24 Greece; 25 Morocco; 26 Madeira; 27 Cabo Verde; 28 S. Tomé e Príncipe. In some localities more than one sampling point exists (see details in Supplementary Table 1).

DNA was isolated from a small piece of muscle tissue using the NZY Tissue gDNA isolation kit (NZYTech) following manufacturer's instructions. A 658 bp fragment of mtDNA cytochrome oxidase I (COI) was initially targeted using universal primers LCO1490 and HCO2198 (Yao et al. 2010). However, the resulting amplification products were not always specific, so a new set of internal primers (605 bp) were designed specificcally for *F. cantabrica* and *F. villafranca* (LCO-Fel 5' GCT TYT KGG TAT RTG GTG TGG and HCO-Fel 5' AAATCA AAA CAG ATG TTG RTA RAG AA). These primers also worked well across all the *Felimare* species included in our study.

PCR amplifications were carried out with an initial denaturation step at 95 ºC for 5 min, followed by 35 cycles of denaturation at 95 ºC for 30s; annealing at 50 ºC ("old" primers) or 54 ºC ("new" primers) for 30s; extension at 72 ºC for 45s; and a final extension step at 72 ºC for 5 min. Amplification products were directly purified using a standard enzyme procedure and sequenced on an ABI 3130xL automated capillary DNA sequencer. DNA extractions, amplifications and sequencing were outsourced to AllGenetics laboratory SA (Coruña, Spain). Amplified fragments were sequenced in both strands, and electroferograms were checked by eye, in Chromas 2.22 (Technelysium Pty. Ltd., Helensvale, Australia), edited and manually aligned using GeniousPro v.6.0.8 (http://www.geneious.com, Kearse et al. 2012). Sequences were translated to confirm the absence of premature stop-codons that could indicate nuclear copies or erroneous sequencing.

2.1.2 Phylogenetic analyses: gene-tree and species-tree inference

The 605bp COI fragment was amplified for a total of 302 individuals. Sequences from our newly sampled specimens were aligned with seven more sequences of WA *Felimare* species and eight more from *Hypselodoris*, *Mexichromis*, and *Roboastra* (outgroups), already available on GenBank (Supplementary Table 2). The final dataset yielded 317 individuals.

Two representatives of genera *Mexichromis* and *Hypselodoris*, (sensu Johnson & Gosliner 2012) plus *Roboastra europea* were included in the analyses (See Supplementary Table 2), as outgroups.

For COI gene-tree inference, the dataset was collapsed to haplotypes using ALTER (Glez-Peña et al. 2010). PartitionFinder (Lanfear et al. 2012) was used to choose the optimum partitioning scheme scheme - using codon positions as possible data-blocks - and substitution models with, under the AICc criterion. Maximum-likelihood (ML) and Bayesian (BI) phylogenetic analyses were conducted in PhyML (Guindon & Gascuel, 2003) and MrBayes v.3.2.2 (Ronquist et al. 2012) respectively. A single

model was applied to the whole alignment (GTR+I+G) according to PartitionFinder results. ML was carried out with 1000 bootstrap replicates, and BI markov chains were run for 11 million generations. In both analyses model parameter values were treated as unknown and estimated along the runs. Two runs with four independent chains (with default heating values) were implemented, and checked for convergence and consistency using AWTY (Nylander et al. 2008). The analysis of clade's posterior probabilities (PP) across generations was used to decide the appropriate burn-in, and both runs were summarized in a single 50% majority-rule consensus tree.

I also estimated a species-tree after removal from the whole dataset of a few individuals (marked with asterisk, at Supplementary Table 2) for which mtDNA did not match morphology plus the outgroup species, *Roboastra europea*, in order to avoid out-group rooting. The species-tree was estimated under the multispecies-coalescent model as implemented in the module *BEAST (Star-Beast) (Heled & Drummond 2010) of the BEASTv.2.1.3 package (Bouckaert et al. 2014). *BEAST estimates jointly the gene and species-trees, and even for single-gene analyses it can provide an accurate assessment of uncertainty in the species-tree estimate, including better estimates of species divergence times, than the gene-tree itself (Drummond et al. 2012). The model of sequence substitution used was the same estimated for the dataset of haplotypes. Divergence times were estimated assuming an uncorrelated lognormal relaxed molecular clock model (Drummond et al. 2006), a Yule tree prior on the species-tree, as well as a population size model of continuous growth with a constant root. Because no specific mitochondrial molecular clock estimate exists for nudibranch molluscs (Hallas & Gosliner 2015), I obtained rough estimates of divergence times between sister groups in years using a teguline gastropod COI molecular clock calibrated upon the Isthmus of Panama (Hellberg & Vacquier 1999). The mtDNA substitution rate was thus set to a normal distribution of mean 1% and standard deviation of 0.1%. Default values were used for all other parameters. Two runs of 50 million generations each were performed, sampled each 5000. To test the influence of the priors on the posterior estimates, an additional run of identical length was made without data, sampling only from the prior. Tracer v1.6 was used to visualize the results of each run, to check the effective sample size of each parameter and to choose appropriate burn-in values. After discarding 10% of burn-in samples, a consensus species-tree (Maximum Clade Credibility tree), with median node heights was obtained using TreeAnnotator v1.8.2 (from the BEAST package) and visualized using FigTree v1.4.2 (Rambaut 2009). Species-tree distributions were also visualized using Densitree (Bouckaert 2010).

2.1.3 Intraspecific DNA variability and genetic structure

Because the level of divergence within each species was generally low, intra-specific diversity for the Atlantic-Mediterranean species was represented using median-joining (MJ) haplotype networks (Bandelt et al. 1999) obtained with the NETWORK software (Fluxus Engineering, Suffolk, UK). Networks of interconnected haplotypes represent the evolutionary relationships and gene genealogies within species usually better than the bifurcating patterns usually recovered by methods of phylogenetic inference (Posada & Crandall 2001).

Additionally, DnaSP (Rozas et al. 2003) was employed to calculate genetic variability across and within species, following geographic criteria, when more than two sequences were present, regarding each partition. Samples from Morroco and Menorca

A series of summary statistics were calculated: haplotype and nucleotide diversity, number of segregating sites, and population mutation parameter θw, for each geographic partition of the data. I also tested for non-neutral evolution by computing Tajima's D (Tajima 1989), as well as signals of demographic expansions by calculating (Fu 1997) and R2 (Ramos-Onsins & Rozas 2002), regarding each geographic partition. Their significance was accessed with 10.000 coalescent simulations, under the hypothesis of selective neutrality and population equilibrium.

The average distance between geographic partitions (Dxy) was calculated within species, when more than two sequences were present in each partition. The uncorrected *p-distance* was applied every time the minimum genetic distance between two species had to be estimated. These analyses were performed in MEGA v.6.0 (Tamura et al. 2013).

2.2 TRANSCRIPTOME ASSEMBLY FOR MARKERS SURVEY IN TWO NON-MODEL SPECIES: *F. VILLAFRANCA* AND *F. CANTABRICA*

Transcriptomes were obtained from single *F. cantabrica* (OP0014) and *F. villafranca* (OP0031) individuals.

2.2.1 RNA extraction

Individuals were frozen after collection, and small pieces of the foot (muscle) were cut about 0.5 cm of length and stored in RNAlater® according to manufacturer's instructions.

RNA extractions and cDNA library construction were outsourced to AllGenetics & Biology laboratory SL (A Coruña, Spain). In short, total RNA was extracted from each sample, using the TRIzol method (Gayral et al. 2011), according to the manufacturer's instructions. Samples were separately disrupted using a mortar and a pestle under liquid nitrogen, and 1 mL of TRIzol solution was added to the ground sample. 200 µL of chloroform were then added and the suspension centrifuged at 12000 rpm for 15 min. The upper phase containing the RNA was recovered, mixed with cold isopropanol and centrifuged at 12000 rpm for 10 min. The supernatant was discarded and the pellet washed with 75% ethanol. Finally, the pellet was re-suspended in 25 µL of nuclease-free water and incubated at 58ºC for 15 min to facilitate re-suspension. The RNA solution was treated with DNase and samples were run on an Agilent 2100 Bioanalyzer to check for RNA integrity and concentration.

cDNA libraries were constructed for each individual using the Illumina TruSeq RNA Sample Preparation v2 Kits following the manufacturer instructions. The procedure includes the purification of the mRNA of the mRNA by oligo-dT hybridization, fragmentation and reverse-transcriptase (RT) PCR. The new cDNA strands were then subjected to an end repair process, with the addition of single "A" nucleotide to which Illumina's specific adaptors and indexes were added. cDNA libraries (300 bp) were amplified and sequenced at Macrogen Inc. on the Illumina HiSeq 2000 platform, for around 70M of 150bp lenght paired-end reads each sample. The paired-end reads were obtained through the Illumina Pipeline (CASA-VA) v1.8.2. The Fastq quality encoding method used was Sanger Quality (ASCII Character Code = Phred Quality Value + 33).

2.2.2 QC and filtering of sequencing data

The quality of the reads generated by Illumina sequencing was assessed with the FastQC software v.0.10.1 (Andrews 2010). Based on a visual inspection of the FASTQ files (two per sample), raw reads were then cleaned using Trimmomatic v. 0.27 (Bolger et al. 2014) removing left-over adaptors and primers and reads smaller than 65 bp. After the quality control step, only paired reads were maintained for subsequent analyses.

2.2.3 Transcriptome assembly and post-processing

For each species *de novo* assembly of the transcriptome was carried out with the Trinity software v.1.7.1.14 beta (Grabherr et al. 2011), using default parameters and the protocol from Haas et al. (2013). Trinity was specially developed for *de novo* transcriptome assembly of short-read RNA-seq data, with or without reference genome. This assembler is comprised of four consecutive modules (or software programs), respectively: Jellyfish, Inchworm, Chrysalis and Butterfly (Figure 11), running

in series through a single Perl script (*Trinity.pl,* by Yang & Smith, 2013), and efficiently recovering transcripts (=”contigs”), clustered in “components” (Grabherr et al. 2011; Zhao et al. 2011). Components can thus contain more than one transcript (or contig) which may represent allelic variation, paralogs and/or isoforms.

Trinity stat’s script (Grabherr et al. 2011) was used to report the number of transcripts, number of components, the transcripts average contig length and the N50 value (defined as the maximum lenght whereby at least 50% of the total assembled sequence is equal of bigger than that). The largest and smallest transcripts, as well as the total, median and average sizes were also calculated.

As a proxy to access the quality and completeness of the transcriptome assembly, CEGMA (Core Eukaryotic Genes Mapping Approach) (Parra et al. 2007) was applied. This procedure identifies a protein set consisting of 248 core proteins that generally lack paralogs in eukaryotes (Parra et al. 2007; Tatusov et al. 2003). Default CEGMA parameters were used for the analysis.

To reduce the redundancy of the dataset and to identify representative isoforms within each component (~loci), I mapped the raw reads to the assembled transcripts using RSEM (RNA-seq by Expectation Maximization) (Li & Dewey 2011). From the transcripts coverage by raw reads, RSEM can accurately estimate the relative abundances of isoforms within single genes (here equivalent to components) (Li & Dewey 2011). Following Yang & Smith (2013), only the isoforms with the highest coverage (highest *IsoPct*) within each sub-component were retained. These are the ones that better correlate with real transcripts, with the longest ones often being chimeras (Yang & Smith 2013). A combination of supplied (*Trinity-pickH.pl*; by Yang & Smith, 2013) and in-house scripts were used to estimate these sequences and collect them into a single fasta file.

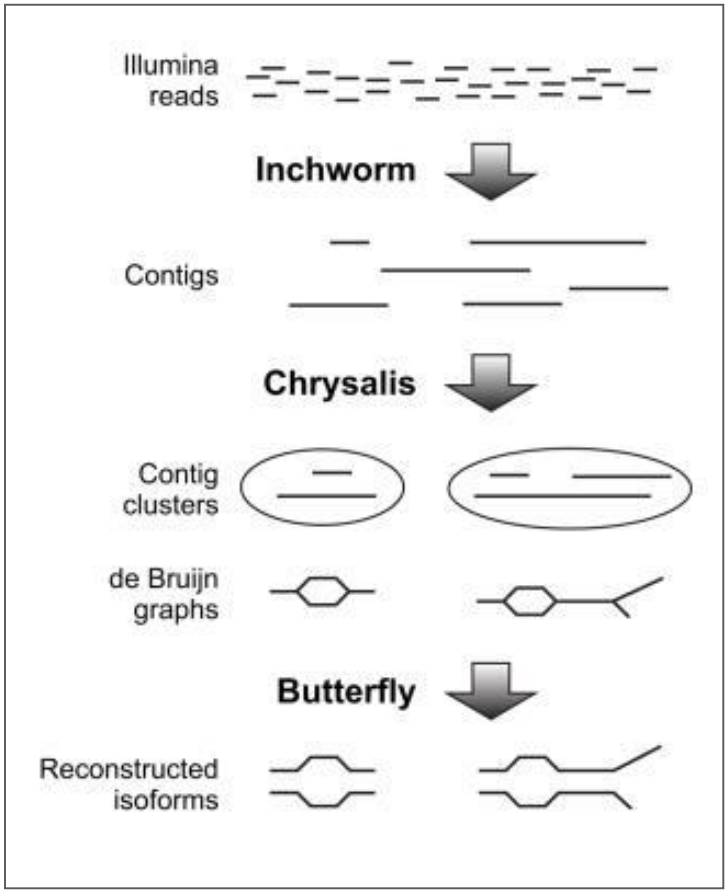


Figure 11 Trinity workflow. Trinity is comprised of four modules: Jellyfish (a fast and memory-efficient counting of k-mers, also known as substrings of length k), Inchworm (reads k-mers, constructs a k-mer dictionary and sorts k-mers by abundance, extending sequences until dictionary is exhausted), Chrysalis (clusters minimally overlapping contigs into sets of components, and constructs de Bruijn graphs for each component) and Butterfly (reconstructs feasible transcripts which in most cases will correspond to alternative splicing of the gene product). Image adapted from Haas et al. (2013). Butterfly (reconstructs feasible transcripts which in most cases will correspond to alternative splicing of the gene product, sometimes giving rise to additional divisions of components, i.e., subcomponents).

2.2.4 Annotation

Accessing the validity, coverage and quality of transcriptomes requires the comparison of the assembled contigs to other genomes. Contigs of all sizes from each assembly were first blasted against the NCBI nr database, with a threshold E-value of $1e^{-6}$. Contigs with blast hits were then functionally annotated using Blast2GO v.2.5.0 (Conesa et al. 2005). Because reliable inference of function depends on the length and quality of the query sequence, only sequences over 200bp and with an Isoform percentage (IsoPct) over 15% were considered. The maximum length allowed by Blast2GOv.2.5.0 is 8000 bp, so sequences above this size were translated into all 6 possible reading frames, and consequently blasted by hand.

2.2.5 Mitochondrial genome reconstruction

I attempted to recover and fully assemble the mitochondrial genome of both species after the main assembly by Trinity using two approaches: 1) blast to other mollusc mitochondrial genomes to recover relevant contigs followed by their "manual" assembly and 2) using MITObim (Hahn et al. 2013).

For the manual assembly *in-house* scripts were used to collect a list of all existing mollusc complete mitochondrial genomes available in Genbank. The resulting fasta file was used to build a (target) database to which the assembled contigs of both transcriptomes were blasted (blastn; E-value threshold of 10^{-6}). Positive hits were first separated into tRNAs and coding regions, and aligned to the complete mitochondrial genome of *Chromodoris magnifica*, using GeniousPro v.6.0.8 (http://www.geneious.com, Kearse et al. 2012). The MITOS web server (Bernt et al. 2013) was used in all contigs with hits to transfer RNAs (tRNAs) regions, and for some unspecific alignments of genes, in order to confirm their annotation as tRNAs. The MITOS web server (Bernt et al. 2013) was used in all contigs with hits to transfer RNAs (tRNAs) regions, and for some unspecific alignments of genes, to confirm their annotation as tRNAs. Nucleotide alignments were generated using the amino acid alignment as a template in GeniousPro v.6.0.8 (http://www.geneious.com, Kearse et al. 2012). The nucleotide and amino acid composition were also estimated in GeniousPro v.6.0.8 (http://www.geneious.com, Kearse et al. 2012).

For the MITObim assembly QC filtered read pools were used, converted to "interleaved" format (where forward and reverse reads appear successively in the same file) and partitioned into smaller files of 10 million reads each. For each species, 3 files of 10 million reads were randomly chosen and used, at a time, to automatically reconstruct the mitochondrial genomes. As part of the MITObim pipeline both transcriptomes were used independently as inputs to MIRA v.4.0.2 (Chevreux et al. 1999) with *C. magnifica* mitochondrial genome as reference. Default parameters were used. The Tablet (Milne et al. 2013) software was used to check the read coverage along the assembled mitogenome. Coding and non-coding regions matching the *C. magnifica* reference genome were consequently confirmed with the MITOS web server (Bernt et al. 2013).

The assembled mitogenome was then manually inspected for repeats at the beginning and end of the assembly to infer circularity.

2.2.6 Detection of mtDNA contamination and further contamination filtering

In the process of recovering and assembling mitochondrial contigs, Amphibian (*Epidalea calamita*) mtDNA was detected (see Results). In order to determine the extent of the contamination in the RNA-seq data, and (tentatively) filter it out, a series of blasts searches and filters were applied.

Based on the blast hits species distribution from the Blast2GO analyses (see Results) a database was built (Table 1) comprising the coding regions of the genomes of species for which a high number of hits existed in both *Felimare* transcriptomes (which were largely the same), the complete genome of the Amphibian *Xenopus laevis*, two complete mitochondrial genomes of genera closely related to *Felimare* (genera *Notodoris* and *Chromodoris*), and also two species of *Bufo* (now subdivided in several genera, including *Epidalea* – Frost *et al.*, 2006), as well as several 18S and 28S sequences of representatives of Euthyneura (18S) and Opistobranchia (28S). Unpublished housekeeping (Beta-actin; Elongation-factor-2; Annexin-2 and GAPDH) and 18S sequences of *E. calamita* (Ylenia Chiari and Nina Séren, *pers. comm.*) were also included in the database, as well as *Homo sapiens* complete coding regions. Additionally, mitochondrial contigs from *Felimare* (OP0014) and *Epidalea* (both recovered from the *Felimare* transcriptome assemblies) were used as controls. All data was downloaded from NCBI or Ensemble on July 2014, and the most recent available coding regions annotations were used in the case of genomes.

Both blastN (against nucleotide databases) and tblastX (translated queries against protein databases) searches were performed, using both transcriptomes separately as queries, recovering all hits of each query sequence. Minimum E-values were set to 1e-6 for both analyses. The number of hits per query sequence was analyzed, as well as the distribution of the hits per species and taxa group. Contigs with blast hits unique to mollusc/non-vertebrata species were recovered from the assemblies to be used for marker survey.

Table 1 List of species and respective markers used to evaluate the existence/absence of contamination on the assembled transcriptomes.

Genetic marker		Genus	Species	Genome ID/ Genbank Acession Number
Contig (control)		*Epidalea*	*E. calamita*	OP14_35204_c0_seq2
		Felimare	*F. cantabrica*	OP14_61774_c0_seq1
		Felimare	*F. villafranca*	NCBI AR[b] 100
Genome (cDNA)		*Aplysia*	*A. californica*	NCBI AR[b] 100
		Saccoglossus	*S. kowallevski*	NCBI AR[b] 101
		Lottia gigantea	*L. gigantea*	GCA _000327385.1 (Ensembl)
		Crassostrea gigas	*C. gigas*	GCA_000297895.1 (Ensembl)
		Capitella	*C. teleta*	GCA_000328365.1 (Ensembl)
		Strongylocentrotus	*S. purpuratus*	GCA_000002235.2 (Ensembl)
		Danio	*D. rerio*	NCBI AR[b] 103
		Oreochromis	*O. niloticus*	NCBI AR[b] 101
		Chrysemis	*Chrysemis picta*	NCBI AR[b] 100
		Branchiostoma	*B. floridae*	NCBI NZ_ABEP00000000[c]
		Xenopus	*X. tropicalis*	NCBI AR[b] 101
		Homo	*H. sapiens*	NCBI AR[b] 100
Complete mitochondrial genomes		*Chromodoris*	*C. magnifica*	DQ991931
		Notodoris	*N. gardineri*	NC_015111
		Bufo	*B. japonicus*	NC_009886
		Bufo	*B. melanocosticus*	NC_005794
Housekeeping genes	Elongation factor 2	*Epidalea*	*E. calamita*	
	Annexin 2	*Epidalea*	*E. calamita*	
	GAPDH	*Epidalea*	*E. calamita*	
rRNA	**18S**	*Epidalea*	*E. calamita*	
		Chromodoris	*C. magnifica*	
	28S	*Chromodoris*	*C. magnifica*	

[a]: refers to sequence ID from current assembly, GenBank or Ensembl accession numbers
[b]: refers to annotation release version (NCBI) of RNA files used
[c]: In the case of *Brachiostoma floridae* only aminoacids (predicted coding regions) fasta was available at the time.

2.2.7 Identification of polymorphic orthologs for future PCR primer design

As much as a conservative method to exclude potential contamination off both datasets, I reserved all the contigs which mapped only to molluks, and excluded all the remaining dubious ones. I then used BLAST to compare the *F. cantabrica* and *F. villafranca* transcriptomes against each other in a reciprocal fashion known as Reciprocal BLAST Hits (RBH) (Wall et al. 2003; Lechner et al. 2011; Kristensen et al. 2011).

Chapter 3 Results

3.1 PHYLOGEOGRAPHY AND PHYLOGENY OF SEVEN EA AND MEDITERRANEAN SPECIES OF *FELIMARE*

3.1.1 Sampling

Species of the genus *Felimare* were surveyed across the Eastern Atlantic and Mediterranean (EAM) shores. Our team sampled locations in the Atlantic, across the Iberian Peninsula (Vigo, Sesimbra and Faro), Morocco and Cabo Verde, while the remaining samples (=individuals) were provided by collaborators. A total of 290 individuals were obtained of which 272 were successfully sequenced, and 30 more were added to the dataset from Genbank or from sequences provided by collaborators. Their distribution is represented in Figure 12, and numbers are as follows: *F. orsinii* (9, from 2 localities); *F. fontandraui* (19, from 7 localities); *F. tricolor* (113, from 10 localities); *F. villafranca* (75, from 12 localities, and including genbank acquisition F*. picta elegans*); *F. picta* (14, from 12 localities); *F. bilineata* (20 individuals, 3 localities) and *F. cantabrica* (52 individuals, 7 localities). Detailed individual's information is given in Supplementary Tables 1 and 2.

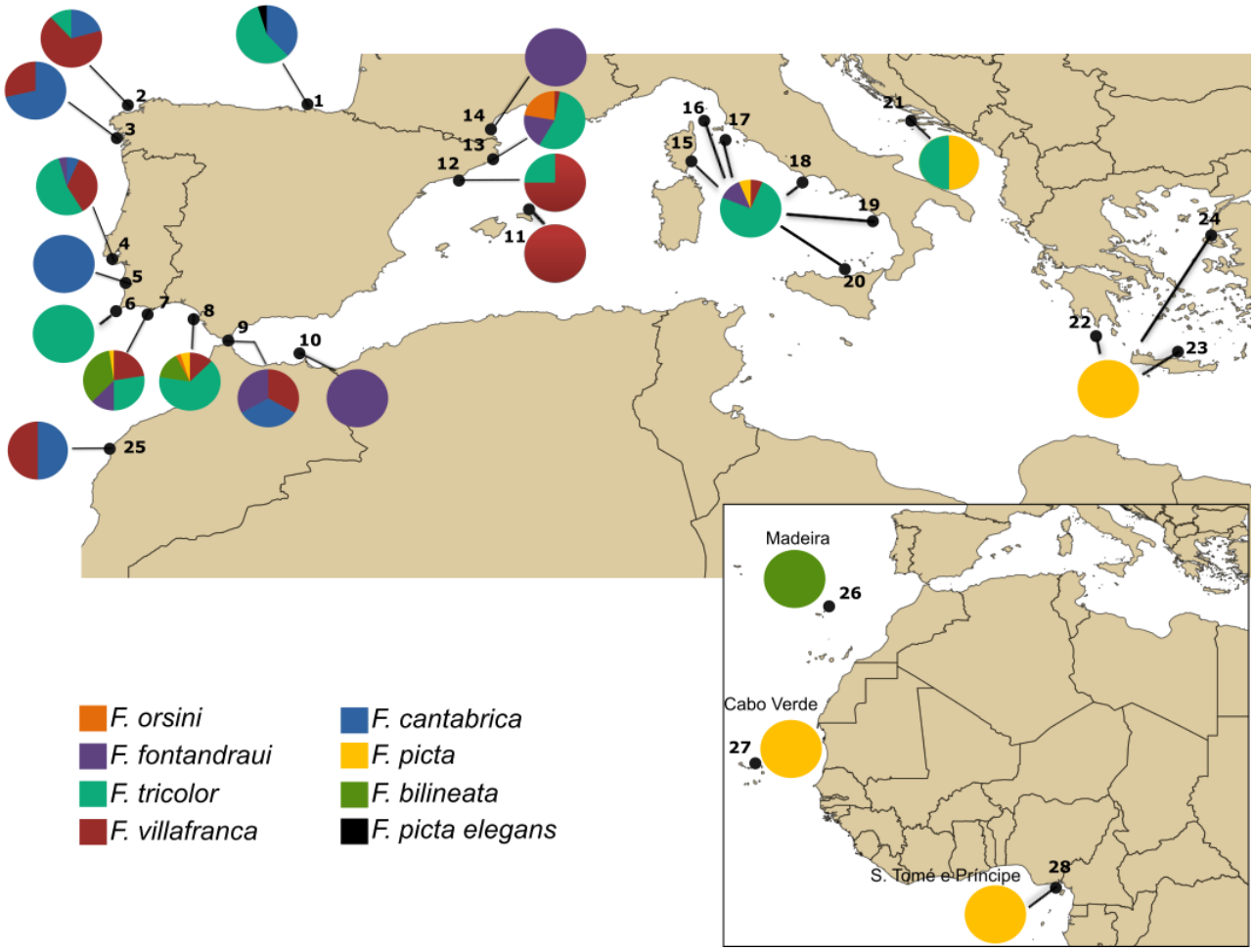


Figure 12 Distribution of the individuals captured at each sampling site (plus Genbank sequances. With exception of two individuals of *F. tricolor* (from Cádiz), *F. tricolor* (Barcelona), and F. cantabrica (Morocco), individuals are here labeled according their mtDNA (see below). Black dots represent localities where samples were collected: 1 San Sebastián; 2 Ferrol; 3 Vigo; 4 Sesimbra; 5 Alentejo; 6 Sagres; 7 Ria Formosa; 8 Cádiz; 9 Gibraltar, 10 Chafarinas; 11 Menorca island; 12 Barcelona; 13 Girona; 14 SW France; 15-20 Italy; 21 Kotor Bay, Montenegro; 22-24 Greece; 25 Morocco; 26 Madeira; 27 Cabo Verde; 28 S. Tomé e Príncipe. In some localities more than one sampling point exists (see details in Supplementary Table 1).

3.1.2 Phylogeny

The alignment of the COI sequences was unambiguous and lacked stop-codons, suggesting that amplification of nuclear copies did not take place. The 317 sequences corresponded to 174 unique haplotypes. Identical haplotypes always belonged to individuals of the same species, with only four exceptions (see below).

All currently recognized species defined well-supported clades but there were a few cases for which the COI haplotype did not match the clade of the species the individual had been identified as: Genbank *F. picta elegans* sequence (AF249787) was identical to *F. villafranca* individuals from North West (NW) Iberia, and one individual from Morocco identified as *F. bilineata* had mtDNA of *F. cantabrica*. Additionally, individual OP0131, identified as *F. tricolor* had *F. fontandraui* mtDNA, and

two individuals identified as *F. fontandraui* (OP0243 and OP0245) had *F. tricolor* mtDNA. These exceptional cases were excluded from subsequent analyses (but see Discussion).

The genus *Felimare* formed a monophyletic group in both ML and BI trees, only differing in their support, i.e, the group yielded very high PP – 97.8% - in the BI tree, while in the ML its support was less than 50%. For both ML and BI, outgroups genera *Mexichromis* and *Hypselodoris* also defined well-supported, and separate, clades. Basal relationships within *Felimare* were unresolved. Relationships among species within the genus were poorly supported in most cases in the ML tree, and in some cases in the BI tree, but in both estimates some aspects could be noted: the Eastern Pacifc *F. porterae* and the Western Atlantic *F. kempfi* always formed a clade (PP=98.5%), as well as *F. picta* with the Eastern Pacific *F. californiensis* (PP=86.8%) plus *F. agassizii*. Also the Eastern Atlantic *F. villafranca, F. cantabrica* and *F. orsinii* (PP= 99.95%) formed a clade, although with unresolved relationships between them. Deep differentiation within species was present in *F. villafranca*, *F. bilineata*, as well as in *F. picta.*

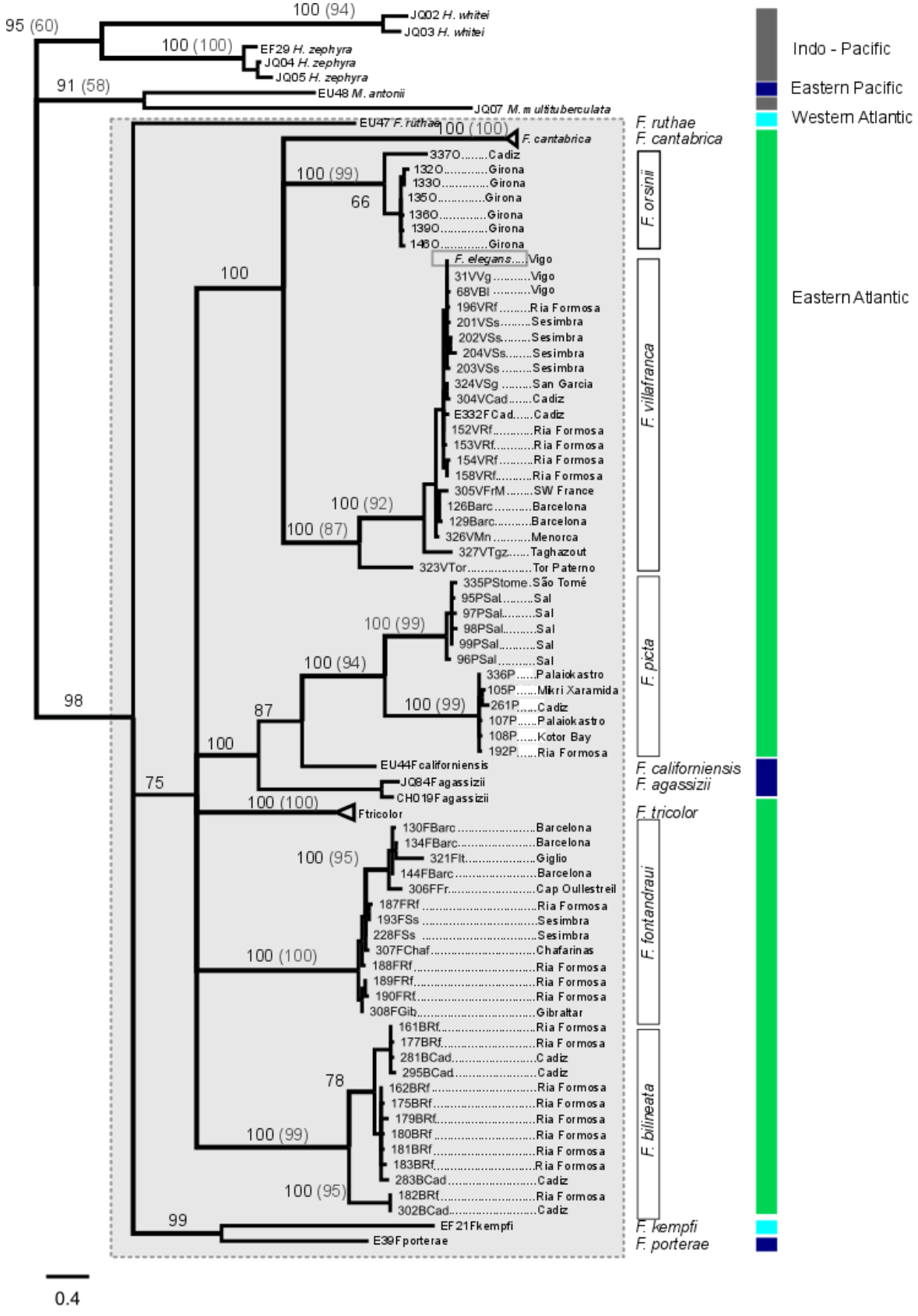


Figure 13 Phylogenetic relationships between Eastern Atlantic *Felimare* inferred from mtDNA COI partial sequence. The tree shown is a 50% majority rule consensus of the BI gene-tree distributions. Posterior probabilities (PP), and bootstrap values (grey, in parethesis) are given in percentages, for values over 50%. Branches with posterior probabilities equal or greater than 90% are in bold. Outgroup (*Roboastra europea*) was removed for graphical purposes. The genus *Felimare* is highlighted with a grey box. Scale is in mean substitutions per site.

3.1.3 Intraspecific variability and genetic structure

The intraspecific diversity for eastern Atlantic and Mediterranean (EAM) species was represented by the COI haplotype median-joining networks in Figures 15 and 16.

3.1.4 F. tricolor and F. cantabrica COI variability

Results across species varied considerably. Neither *F. tricolor* nor *F. cantabrica* showed any signs of population structure, with closely related COI haplotypes distributed and shared across the sampled locations (Figure 14).

F. tricolor was the most abundant species in many sampled sites (Figure 12), both in Atlantic and Mediterranean waters. Despite its wide distribution, *F. tricolor* exhibited a very shallow mitochondrial genealogy. The medium-joining network (Figure 14) was star-like. From the 68 COI haplotypes observed in this species (*F. tricolor*), 90% were singletons, what was reflected in a high haplotype diversity (Hd=0.97) (Table 2), while the nucleotide diversity was the third lowest of all seven species (π=0.0068). There was no evidence of genetic structure, with most central COI haplotypes being shared across widespread localities. For example, the same COI haplotype was observed in S. Sebastián (north Spain), Ria Formosa and Cádiz (South of Iberian Peninsula), Barcelona (Mediterranean, north East Spain) and Italy (central Mediterranean).

F. cantabrica was sampled across almost all its known geographic distribution, from the Cantabrian Sea to Morocco, although it was not possible to collect any individuals from its western distribution limit, in the Mediterranean Sea. Most of the observed COI haplotypes diverged from each other by only one or two substitutions, and singletons were frequent (Figure 14). Thus, and similarly to *F. tricolor*, there was no evidence of differentiation among localities. Haplotype diversity was moderate (Hd= 0.88).

For both *F. tricolor* and *F. cantabrica* neutrality statistics were negative and significant (Fs = -86.11/ D= -2.43, and Fs= -26.91/ D= -2.35, $p < 0.01$), while R_2 values were significantly low (R_2= 0.0216 and 0.0284, $p < 0.01$), suggesting past demographic growth (Table 2).

3.1.5 F. villafranca COI variability

On the other side, considerable diversity was observed within *F. villafranca*, with 19 COI haplotypes, and a clear differentiation among populations from the east Atlantic (West of Gibraltar), West Mediterranean (Barcelona, Girona, France); Central Mediterranean (Italy) and Morocco. Individuals from north and Central western Iberian Peninsula (Ferrol, Vigo, Sesimbra) exhibited a

group of very closely related haplotypes, and shared the central, more abundant one, which was separated from the ones observed in Ria Formosa by at least one mutational step. Similarly, COI haplotypes from Cádiz and Gibraltar also diverged from haplotypes from Ria Formosa by a minimum of one and a maximum of four mutational steps. Despite this low differentiation, there was no haplotype sharing between the northern (Vigo, Ferrol, Sesimbra) and the southern (Ria Formosa, Cádiz and Gibraltar) localities. The individual collected in the southernmost sampled locality (Taghazout, Morocco) exhibited a COI haplotype that was 15 mutational steps away (uncorrected p-distance of 2.6% to 3%), which occurred on the Mediterranean localities of Barcelona and Menorca. These western Mediterranean localities, respectively Menorca Island, Girona (Spain) and SW France (Banylus-sur-mer) exhibited closely related groups of haplotypes, and a single case of haplotype share being between Girona and Banylus-de-mer. The average distance (Dxy) between western Mediterranean and EA lineages (Morocco not included) was 6%. Finally, one individual from the Italian shore was strongly distinct from the rest, with a distance of 38 substitutions to a COI haplotype from Menorca, which translates in an uncorrected p-distance of 6.4% up to 7.1%. Its distance to the Morocco haplotype was roughly similar (7.4%), as well as to the North and Central Iberian haplotypes (7.3% up to 7.6%). Population growth was also inferred for *F. villafranca*, both when considering the species as a single unit (Fs=-2.3649; D=-2.4222, p-value<0.01), as well as when analysing the Atlantic A lineage alone (Fs=-5.6317, p<0.05; D=-1.6057, p<0.05) (Figure 14). Growth was not detected in the Mediterranean populations when analysed separately.

3.1.6 F. fontandraui COI variability

Within *F. fontandraui*, at least three distinct groups could be inferred, with a clear geographical pattern: one at Atlantic + Chafarinas, and two others in the Mediterranean. The genetic distance (Dxy) between Atlantic and Mediterranean groups was 3.8%. Despite the low number of individuals sampled (19), the number of COI haplotypes was high (13), with a very high allelic diversity (Hd= 0.95 and π= 0.015).

The Mediterranean seemed to harbour at least two differentiated lineages (1.8-2.1% uncorrected p-distance), without clear geographic structure: COI haplotypes were shared between Girona and the west Italian coast, and individuals from Girona and SW (Cap Oullestreil), only 200km apart, differed by five mutational steps. Shared haplotypes diverged from an COI haplotype from Tavolara (Sardegna, Italy) by at least 10 mutational steps, yet one haplotype was also shared between Giglio island (Italy west coast) and Girona. We considered all the haplotypes/individuals from the Mediterranean as a single lineage here, for summary diversity statistics calculations, but it they could also represent more than one group.

Both the Atlantic and Mediterranean lineages as here defined, presented high (Hd ≥ 0.90) haplotype diversity. The diversity was observed higher in the Mediterranean (π= 0.012 over π= 0.005).

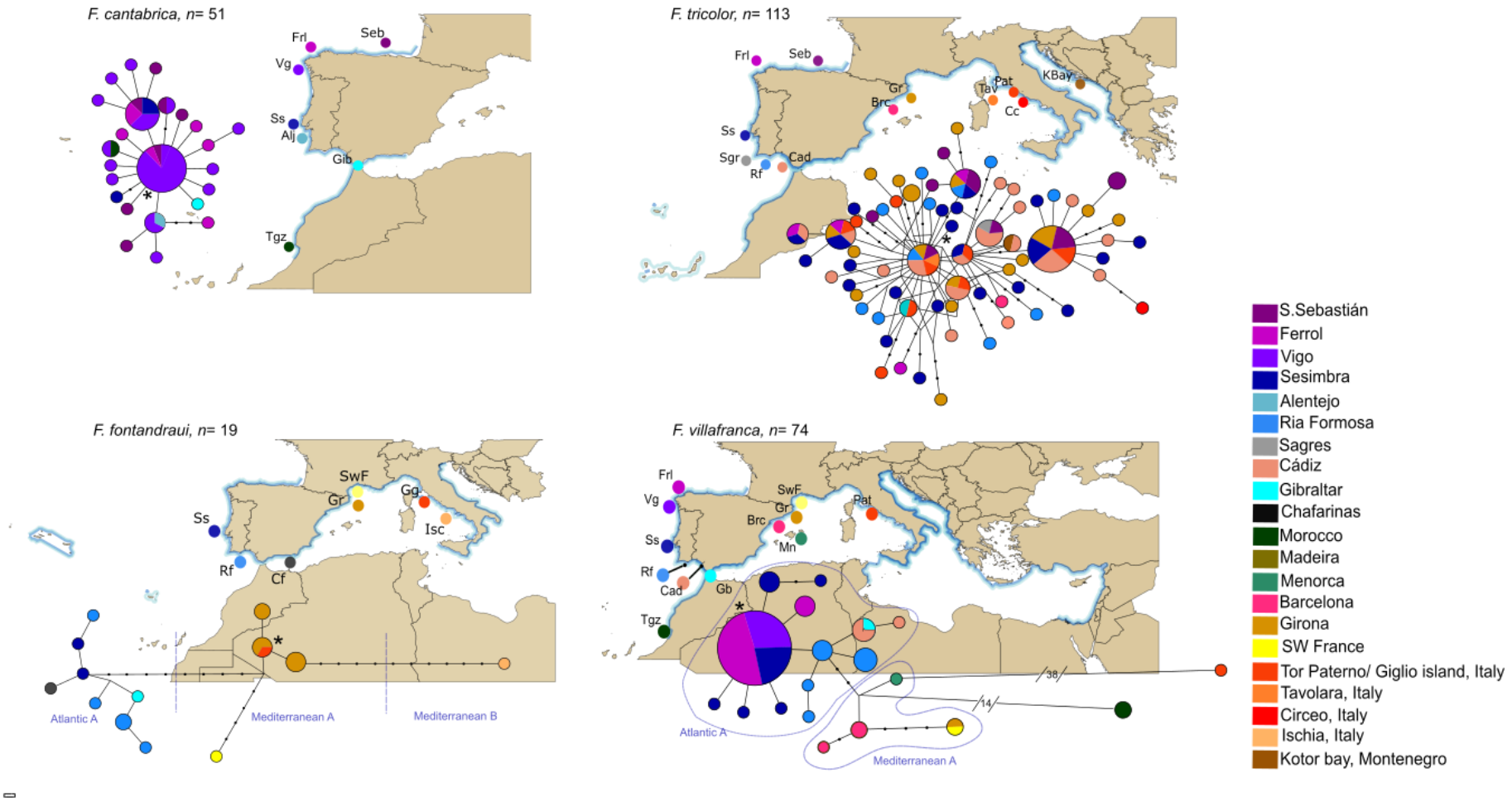


Figure 14 Median-joining networks representing mtDNA (COI, 605 bp) variation within Eastern Atlantic *Felimare* species. For each species, the sampled sites and known distribution range (light blue) are shown. Circle sizes are proportional to haplotype frequencies. Missing haplotypes (unsampled or extinct) are represented by small black circles along the branches (for long branched the total number of substitutions is indicated). The asterisk in *F. cantabrica*, *F. fontandraui* and *F. tricolor* represent the haplotypes exhibited by the specimens with distinct morphological identification mentioned previously. These are here indicated but were not considered for the summary statistics or for any further analyses.

3.1.7 F. bilineata COI variability

With only a small portion of its distribution sampled, *F. bilineata* harboured at least two highly differentiated COI lineages separated by a minimum of 20 substitutions (4.3%), without a clear geographic structure (haplotypes from Ria Formosa and Cádiz are shared by both lineages, Figure 15). Furthermore, considerable differentiation was found within one of the lineages, with two groups of haplotypes separated by a minimum of 7 mutational steps (approximately 1% uncorrected *p-distance*). The other lineage was represented by a single COI haplotype exhibited by two individuals. No relationship between differentiation and geography was found, given that all lineages co-occurred in the southern Iberian Peninsula in Ria Formosa and Cadiz, two of the three localities sampled. Observed haplotype diversity Hd= 0.884 was high, as well as the nucleotide diversity (π= 0.014).

3.1.8 F. picta COI variability

Twelve haplotypes, belonging to two very divergent lineages, one present in Atlantic and Mediterranean waters and the other in the islands of Cabo Verde and S. Tomé e Príncipe, were observed within *F. picta* (Figure 15). One clade included individuals from the Eastern Atlantic Islands of Sal (Cabo Verde) and S. Tomé e Príncipe, while the other harboured individuals from the Atlantic localities of Sesimbra, Ria Formosa and Cádiz, as well as several Mediterranean localities across Italy (west coast), Montenegro and Greece. These lineages had a pairwise Nei's Da distance of 9.6% (49 substitutions). Within each lineage haplotypes were closely related (differing by one to three mutational steps). The overall COI haplotype and sequence diversity within this species was high (Hd= 0.99 and π=0.05049). For each lineage separately, the nucleotide diversity is approximately one order of magnitude less than its "overall" value (Table 2), due to the limited divergence among haplotypes (π=0.004 in the Atlantic A and Mediterranean regions, and π=0,006 for the Atlantic II region) (Table 2). For each *F. picta* clade, Fs and Tajima's D were also negative although not (or only marginally) significant, while R_2 values (0.1352 and 0.0809) were significant ($p<0.05$), suggesting past population growth.

3.1.9 F. orsinii COI variability

F. orsinii (Figure 15) is distributed across the Mediterranean and reaches slightly west of Gibraltar. Again, strong intraspecific divergence was evident, with two divergent (3.6%) COI lineages distributed respectively East and West of the Gibraltar Strait. Individuals from the two sampled localities (Cádiz - Atlantic Ocean - and Barcelona - East Mediterranean), represented thus two distinct lineages, separated by 19 mutational steps. Levels of genetic diversity within this species were high

(Hd= 0.944, π= 0.016), with the majority of it existing in the Mediterranean (Hd= 0.922). High sequence diversity was mainly attributable to the high distance between the two lineages.

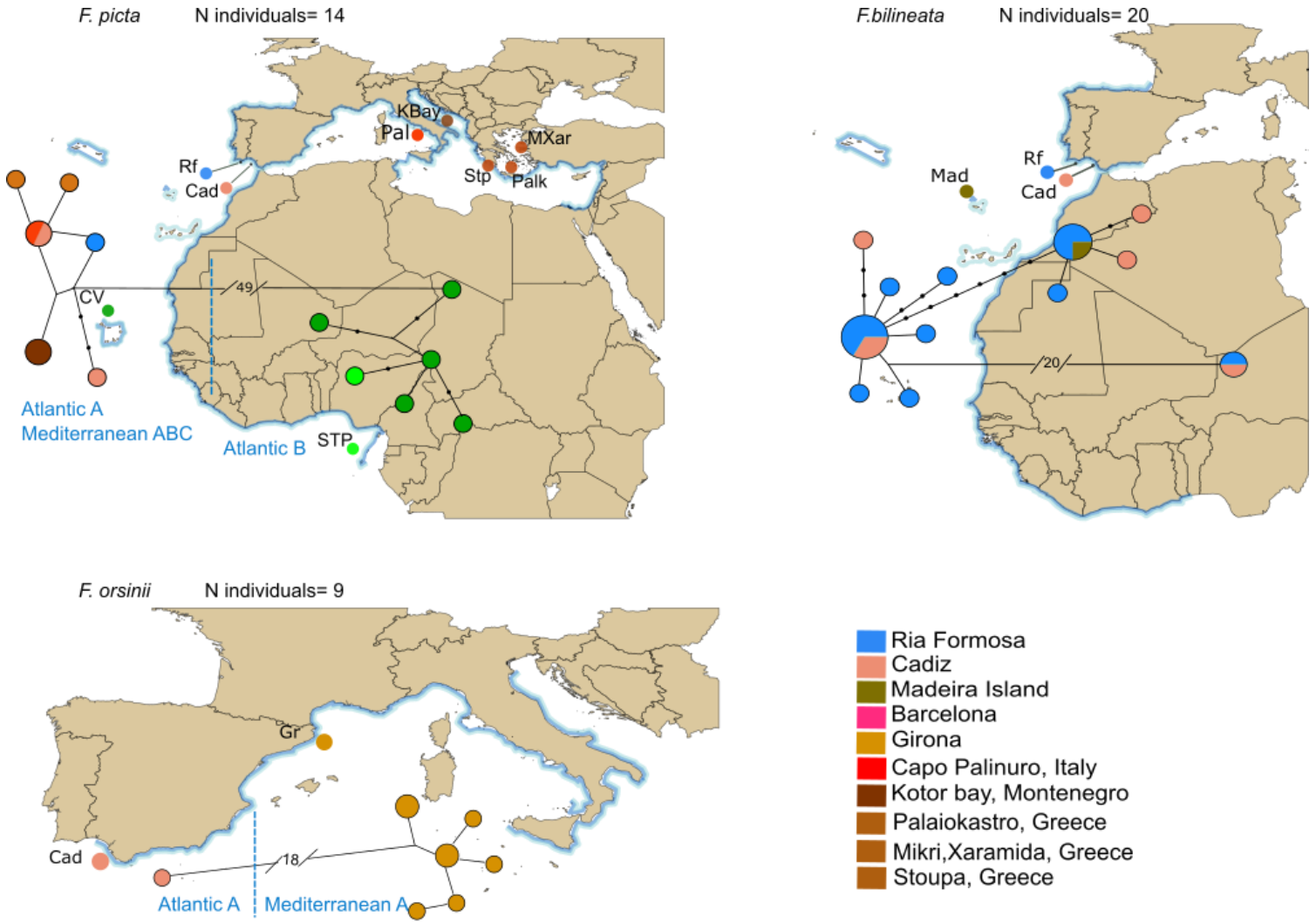


Figure 15 Median-joining networks representing mtDNA (COI, 605 bp) variation within two Eastern Atlantic *Felimare* species. For each species, the sampled sites and known distributions (light blue) are shown. Circle sizes are proportional to haplotype frequencies. Missing haplotypes (not sampled or extinct) are represented by small black circles along the branches (for long branched the total number of substitutions is indicated).

Table 2 Summary diversity statistics, neutrality tests and indexes of population growth for the seven EAM *Felimare* species, based on mtDNA variation.

Species	Partition	N	Polimorphism					Neutrality	Population growth	
			H	S	Hd	π	θ (S)	D	Fs	R_2
F. orsinii	**Total**	9	7	25	0.944	0.01058	0.01578	-1.6400*	-0.7507	0.2362
	Mediterranean A	8	6	7	0.929	0.00378	0.00446	-0.7298	-2.2767	0.1365*
F. fontandraui	**Total**	18	13	35	0.954	0.01528	0.1682	0.3719	-1.8965	0.13443
	Atlantic A	7	6	7	0.952	0.00504	0.00472	0.3451	-2.2004*	0.1717
	Mediterranean AB	11	7	31	0.891	0.01244	0.01749	-1.3404	0.6615	0.1220
F. tricolor	**Total**	111	66	82	0.970	0.00683	0.02618	-2.4329**	-86.111**	0.0216**
F. villafranca	**Total**	74	16	60	0.678	0.00608	0.02051	-2.4222**	-2.3649	0.0671
	Atlantic A	65	10	10	0.583	0.00145	0.00351	-1.6057*	-5.6317**	0.0481
	Mediterranean A	6	4	7	0.867	0.0518	0.00507	0.1284	0.3138	0.1874
F. cantabrica		51	25	28	0.880	0.00295	0.01029	-2.3503**	-26.9186**	0.0284**
F. picta	**Total**	14	12	66	0.987	0.05049	0.03434	1.9907	0.3330	0.14283
	Atl. A Med. ABC	8	6	8	0.929	0.00407	0.0051	-0.9735	-2.0860	0.1352*
	Atlantic B	6	6	10	1.000	0.00584	0.00724	-1.1605*	-2.8976	0.0809**
F. bilineata	**Total**	21	12	40	0.895	0.01438	0.01910	-0.9740	-0.4829	0.1352*

N, number of individuals (mtDNA); H, number of haplotypes; S, number of (polymorphic) segregating sites; Hd, Haplotype diversity; π, nucleotide diversity; θ (from Eta) population mutation parameter (Watterson, 1975). Neutrality test: Tajima's D. Population growth tests: Fu's Fs (1997) Fs; R_2, (Ramos-Onsins and Rozas', 2002). *$p < 0.05$; ** $p < 0.01$. [b] Statistics were not computed for *F. orsinii* and *F. villafranca* populations from Cádiz, Taghazout and Tor Paterno, respectively, as less than 3 individuals were sampled.

3.1.10 Species-Tree Inference and Divergence Times Estimates

All three independent analyses with *BEAST achieved reasonable ESS values for all parameters, but slightly distinct tree topologies. The run with the highest ESS values (and highest posterior) is presented here.

Figure 16 shows a maximum clade credibility (MCC) consensus of the species-trees distribution using median node-heights. Support for interspecific relationships was low in most instances, although often higher than in the ML/BI gene trees. A clade was recovered containing *F. ruthae* and *F. kempfi,* both species from the Caribbean coasts, and *F. porterae*, from the East Pacific coast. The sister-relationship between *F. kempfi* and *F. porterae* was very well supported. Also the clade containing *F. villafranca, F. orsinii* and *F. cantabrica* was well supported, with a PP=93%, although the relationships between these species were not well supported. The remaining six species of *Felimare* defined a clade with a PP= 80%, again, with no internal relationship between these species being highly supported.

Regarding time-estimates, differentiation between species seems to be old, dating mostly to 10 to 25 MY. Highest posterior density intervals (95% HPD) for most nodes largely overlapped.

The common ancestor of the genus *Felimare* was dated 20 - 34 MY (median 27.06 MY), while subsequent diversification events followed throughout the Oligocene and Miocene. The most recent divergence event inferred was between *F. californiensis* (EP) and *F. tricolor* (EA), likely Pleistocenic.

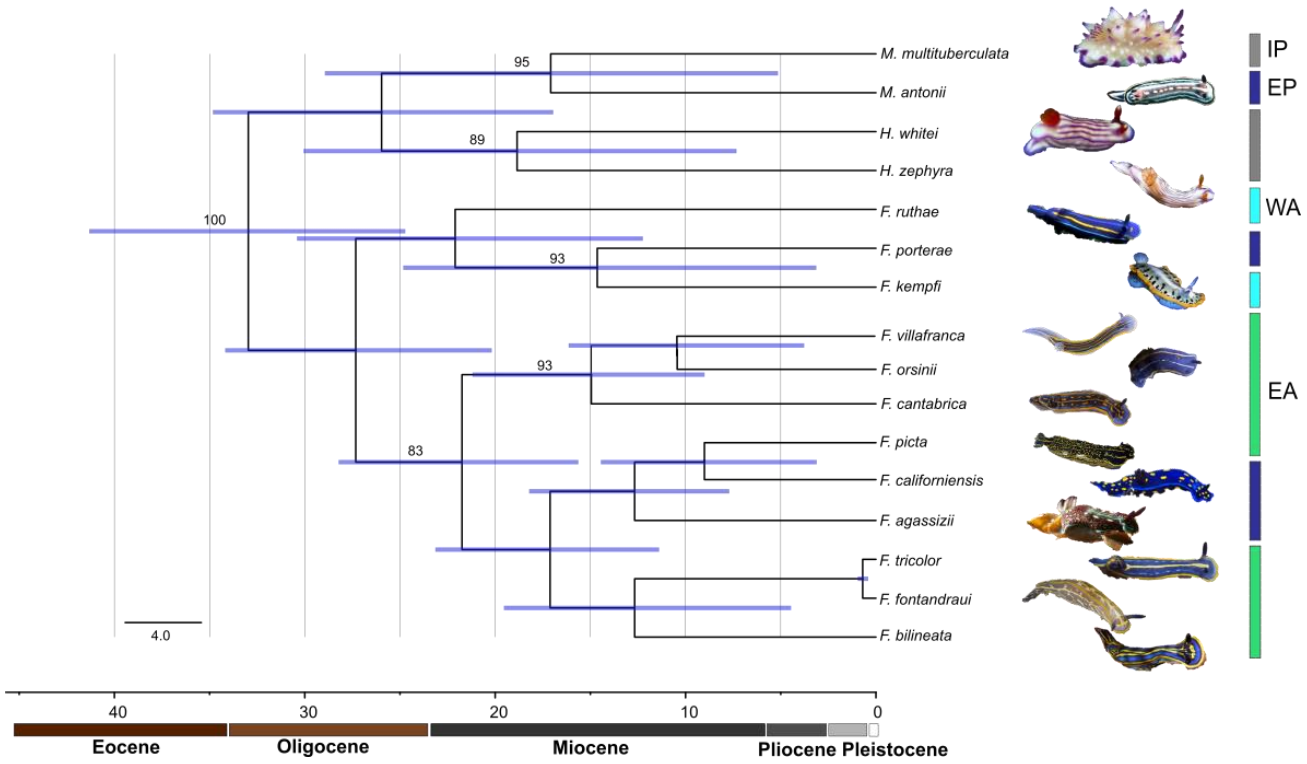


Figure 16 Maximum clade credibility COI tree with median estimates used for node heights. Posterior probabilities above 80% are given above the branches (in %). 95% HPD of node heights is also given. Horizontal axis corresponds to time before present in million years (MY). Geographic category of species distribution is given: IP- Indo-Pacific; EP- Eastern Pacific; WA- Western Atlantic; EA- Eastern Atlantic.

3.2 TRANSCRIPTOME ASSEMBLY FOR MARKERS SURVEY IN *F. CANTABRICA* AND *F. VILLAFRANCA*

3.2.1 Quality control (QC) and filtering of raw sequence data; transcriptome assembly; and post-processing

Approximately 38 million (38,323,175 out of 38,648,449) and 35 million (35,125,908 out of 35,361,951) of 101bp (average length) paired-end reads (PE) obtained for *F. cantabrica* and *F. villafranca*, respectively, met the filtering criteria and were used in the *de-novo* assembly.

Table 3 presents statistics relative to each transcriptome assembly. Final assemblies resulted in 56,776 and 60,484 components (with 78,546 and 82,456 transcripts each), respectively for the *F. cantabrica* and *F. villafranca* individual. Assembled transcript size ranged from 200 bp up to 24,000 bp (Table 3). In both cases less than 40% of the total transcripts were larger than 500bp (Figure 17).

Table 3 Main assembly results for both transcriptomes.

Assembly (Trinity)							
ID	**Reads BT**	**Reads AT**	**Total transcripts**	**Total components**	**N50 transcripts**	**Average contig length**	**Length range**
F. cantabrica	77,296,898	76,646,350	78,546	56,776	899	625.14	[200, 24,100]
F. villafranca	70,723,902	70,251,816	82,456	60,484	950	651.44	[200, 9,200]

BT- before trimming; AT- After trimming. The N50 length is the length of the shortest representative contig from which the sum of contigs of the genome/transcriptome of equal length or longer represents at least 50% of the total length of all contigs.

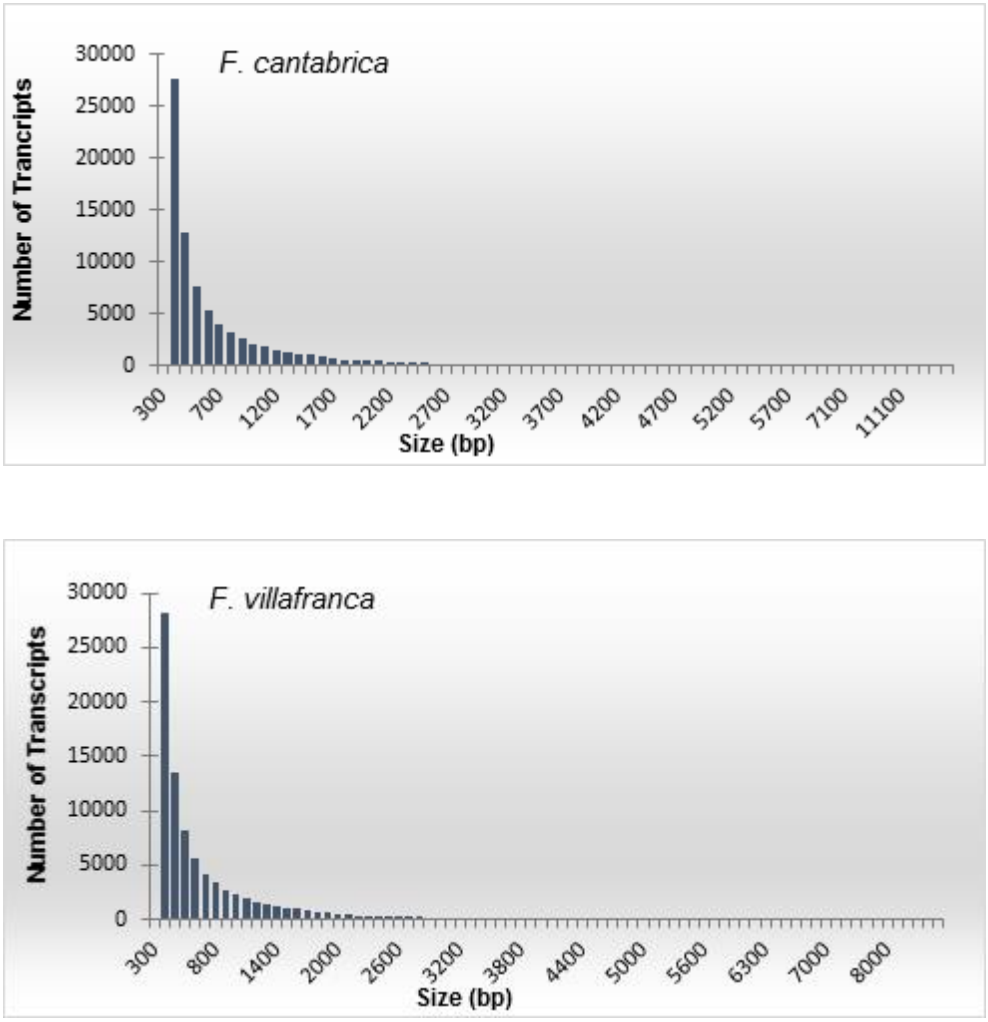


Figure 17 Distribution of transcript length. Histograms show the number of transcripts per size category for F. cantabrica and *F. villafranca* transcriptomes. Each bar/ column, represents 100bp range (starting at 200 bp). Circa of 61% of the transcripts of *F. cantabrica* and 60% of the transcripts of *F. villafranca* are ≤500bp.

Regarding the quality of the assembled transcriptomes, roughly, more than 50% of the CEGMA core set of genes were recovered as “complete” in each transcriptome, while at least 60% for *F. cantabrica* and 65% for *F. villafranca* were recovered as “partial” (Table 4).

Table 4 Completeness report of the percentage of the eukaryotic core genes present in both datasets (CEGMA).

ID	CEGMA (orthologs)	
	% Completeness	
F. cantabrica	**Partial**	61,69
	Complete	53,63
F. villafranca	**Partial**	66,94
	Complete	54,84

The % of core genes which was found to be "complete" (at least 70% of the original protein size is recovered) or only "partial" (less than 70% alignment length but exceeding the minimum alignment score) in each assembly is given.

The presence of redundancy in the assemblies is an important caveat for phylogenomics, as redundant and chimeric contigs may lead to erroneous phylogenetic markers. Redundant contigs can either be related to paralogs or they can be alleles of the same locus. RSEM was used to measure transcripts/isoforms "coverage", and from each component a single representative transcript (the one with the highest "coverage" – here measured by isoform percentage, IsoPct) was chosen (Figure 18).

Main transcripts characteristics after filtering are given in Table 5. As seen in previous plots, most contigs were shorter than 2,000 bp, with 16,915 and 18,664 contigs in *F. cantabrica* and *F. villafranca* respectively being equal or longer than 500 bp. Largest contig size varied greatly; from 24,068 bp to 6,577 bp. None of the three longest (> 8,000 bp) contigs, retrieved a significant hit against NCBI nr database.

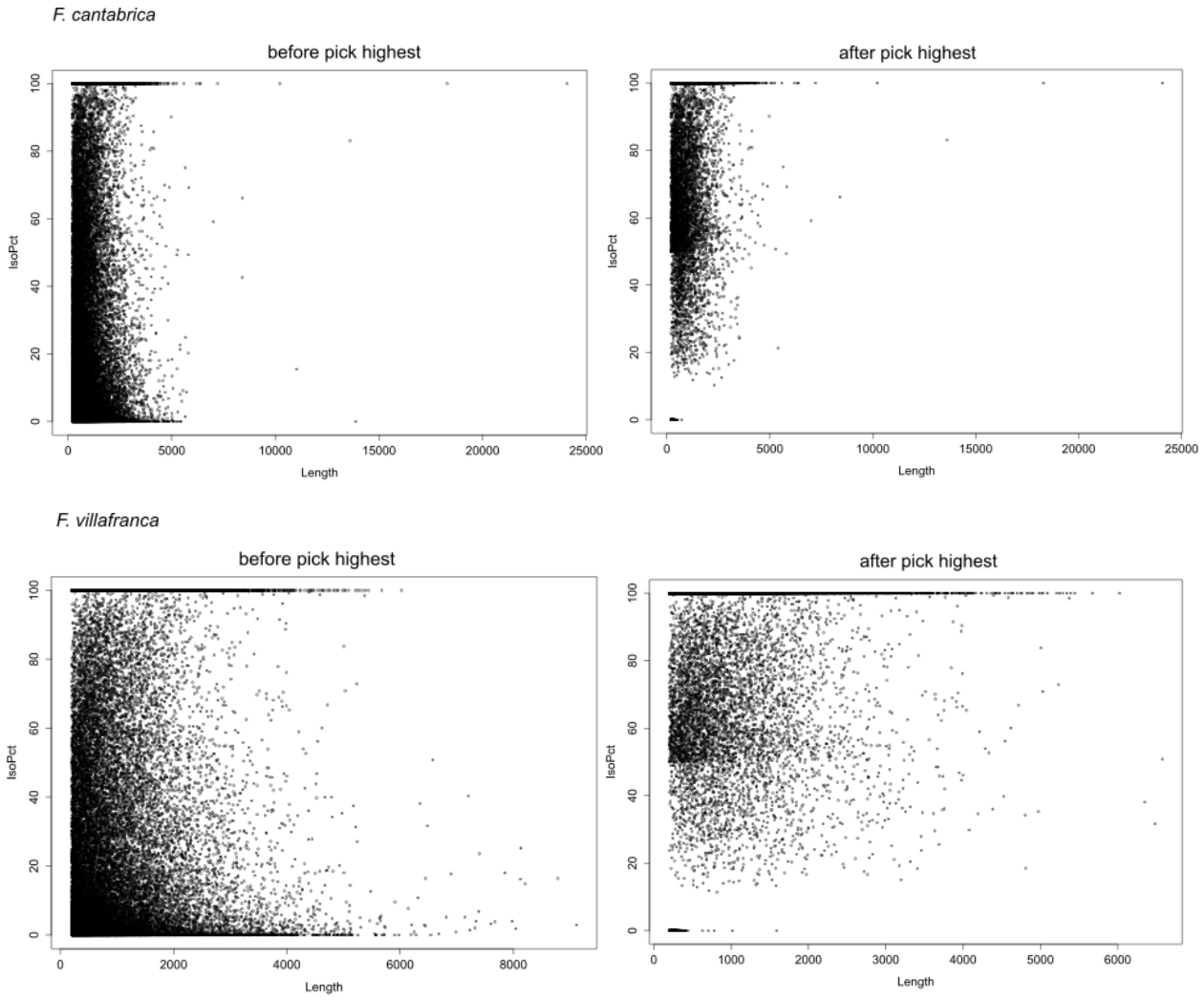


Figure 18 Plots of transcripts length vs. isoform percentage (IsoPct) for each assembly. Each transcript is a dot and the graphs reflect each dataset before and after filtering them by selecting the isoform present in higher percentage (i.e., the most abundant one) in each subcomponent.

Table 5 Assembly statistics after filtering the dataset to a single transcript per subcomponent.

Transcripts characteristics after choosing component representative				
	Total Transcripts	**Total Transcripts IsoPct ≥ 15**	**Average contig length (bp)**	**Transcripts > 500 bp**
F. cantabrica	56,776	54,644	516.01	16,734
F. villafranca	60,484	58,554	528.76	18,443

3.2.2 Transcriptome annotation

Out of the more than 50,000 transcripts of each assembly, 10,005 from *F. cantabrica* and 11,374 from *F. villafranca* had hits against the Genbank nr database (largest protein database available) with e-values above the pre-defined threshold (1e-6) (Figure 19).

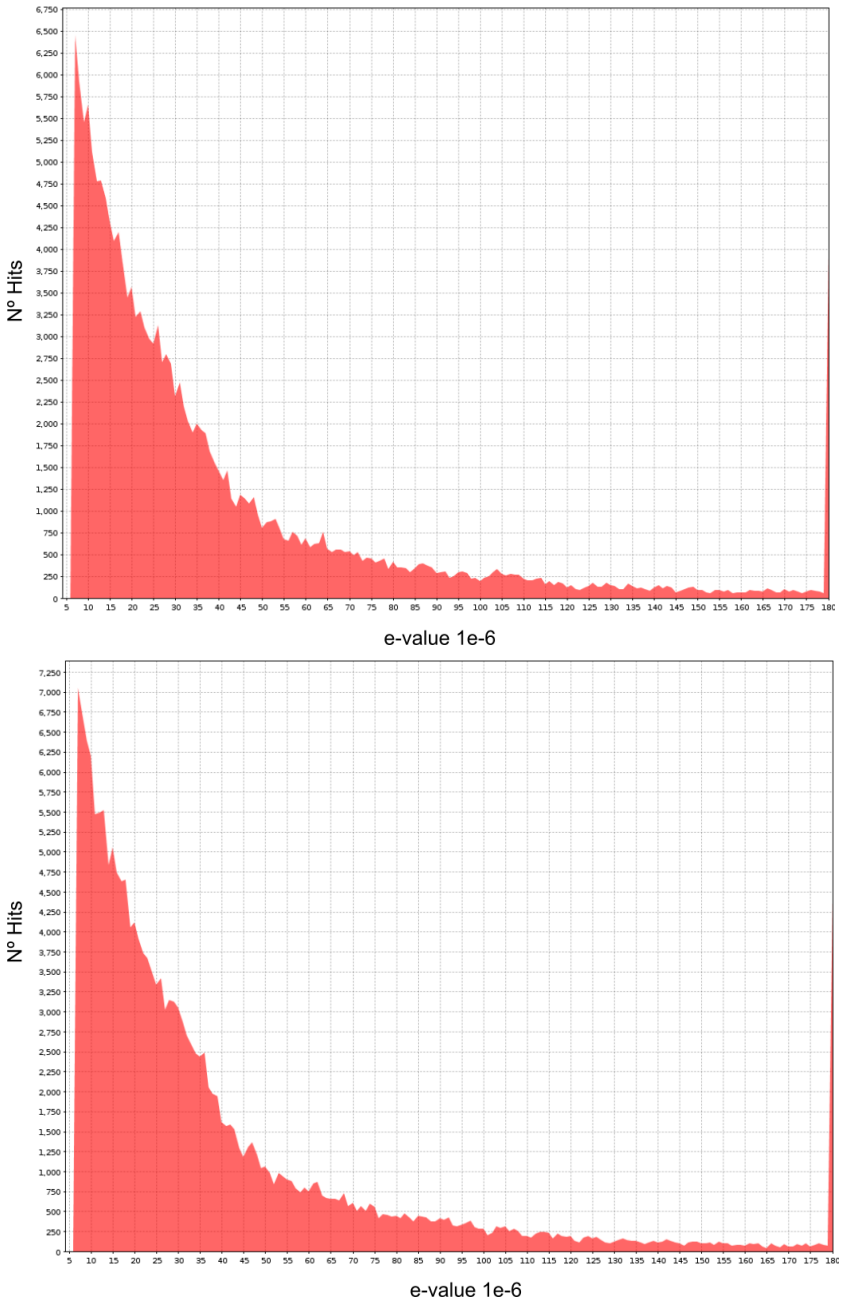


Figure 19 Best hits e-value distribution for *F. cantabrica* (above) and *F. villafranca* (below).

A summary of the annotation statistics is reported in Table 6. From sequences with positive blast hits to the nr database, GO terms were found for 6,620 and 3,871 sequences (mapping), and from those 4,275 and 2,952 were successfully annotated, respectively for *F. cantabrica* and *F. villafranca*.

Table 6 Number of sequences at different annotation stages.

ID	Input Sequences	Blast Hits	Go evidence (Mapping)	Mean GO level	Length (min-max)	Total Annotated (%)
F. cantabrica	54,644	10,005	6,620	5.65 ±1.89	200 - 7,010	4,275 (7,82)
F. villafranca	58,554	11,376	3,873	5.8±1.82	200 - 6,575	2,952 (5,04)

Blast2GO performance: Percentage of sequences with BLAST hits above 200 bp; percentage of sequences with some functional evidence assignment (mapping); mean GO level; and total number of annotated sequences. The number of input sequences includes only s equences of length larger than 200 and shorter than 8000 bp, from the originally 54,644 and 58,554 from Table 5.

The mean GO level (mean number of GO terms between the final GO term to the ontology root term) was of approximately 7.1 and 7.6, respectively, similar in both *F. cantabrica* and *F. villafranca* transcriptomes, with a total of 25,563 and 25,858 sequence annotations assigned, respectively (Figure 20). Of these, assignments to the Biological Process category made up the majority (14,721/13,106 sequences) followed by Molecular Function in *F. cantabrica* (7,209 sequences) and Cellular Component (6,628 sequences) in *F. villafranca* for a total of 27,669/25,858 GO assignments (*F. cantabrica*/*F. villafranca*).

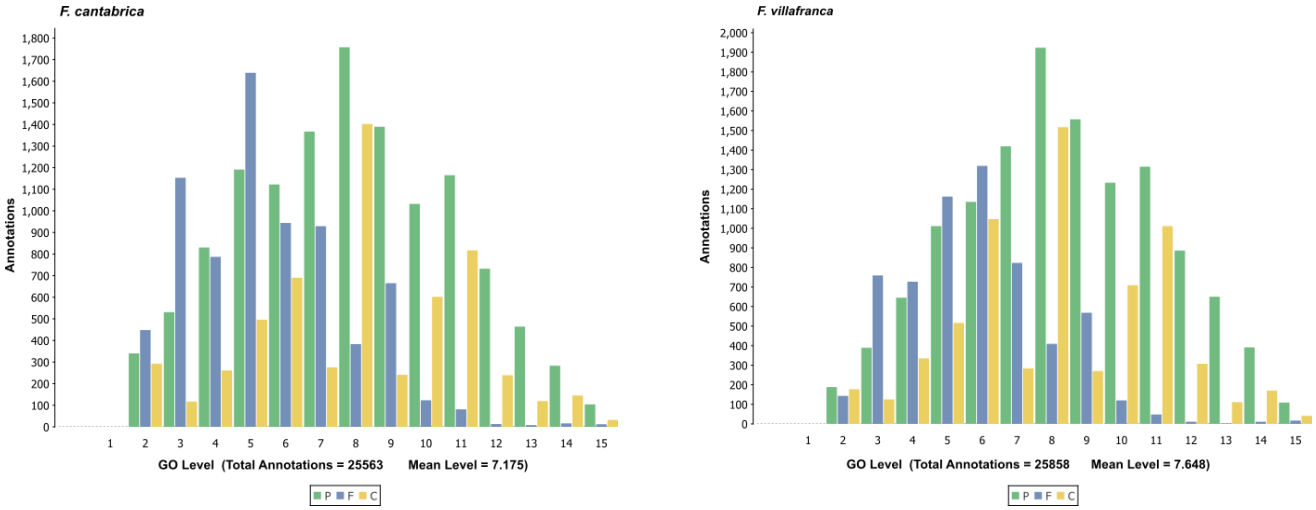


Figure 20 GO Level distribution's chart. Histogram representing the distribution of the multi-level function in the annotated sequences. GO terms representation for Biological Process (P), Molecular Function (F) and Cellular Component (C).

Annotation plots were obtained for both individuals for GO levels 2 and 3 for the top 20 GO terms (Figures 21 and 22). Within level 2, distributions of sequences across GO terms seem more similar across the two individuals, again with Biologic Process being the most common category (14,964/14,721 sequence counts, for *F. cantabrica/F. villafranca*), followed by the Molecular Process and Cellular Component categories. Within the category Biological Process (and respectively, for *F. cantabrica* and *F. villafranca*) "Cellular Process" (18% and 15% of the total sequence counts), "Metabolic Process" (16% and 12%) and "Single-Organism Process" (14% and 13%) are the most prevalent terms. For Cellular Component "Cell" (22% and 19%), "cell part" (22% and 19%) and "organelle" (17% and 16%) were most popular, while for the Molecular Function category most of the sequences were represented by the term "binding activity" (45% and 49%), "catalytic activity" (39% and 34%) and "structural molecule activity" (5% and 6%).

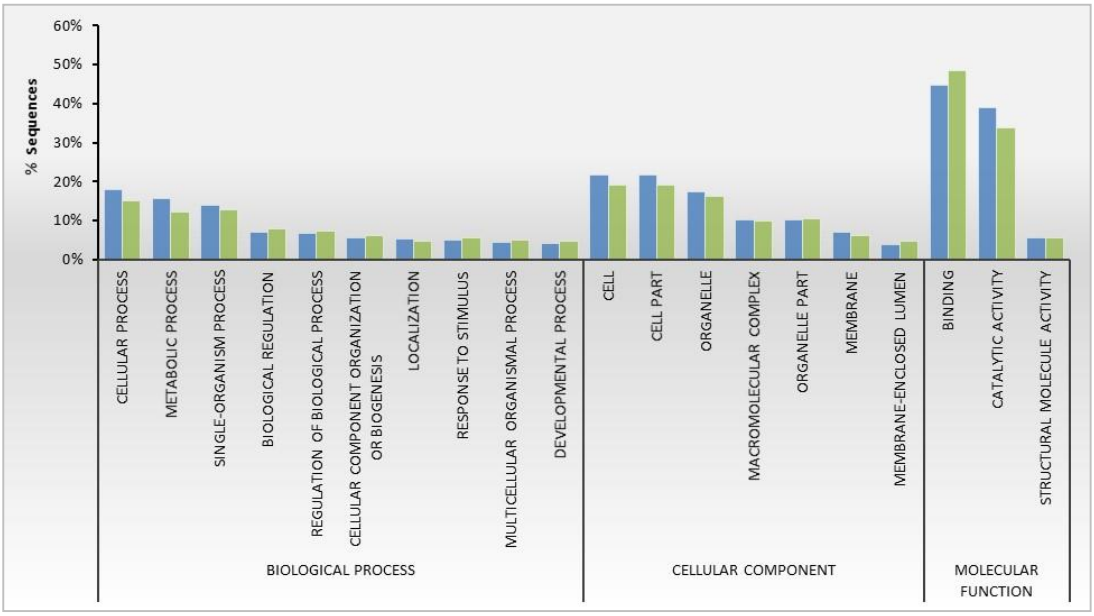


Figure 21 Paired comparison of the sequences mapped to a given gene ontology (GO) term in level 2. For each category (BP, CC, and MF) the top 20 GO terms for each species are represented (Blue – *F. cantabrica*, green – *F. villafranca*).

For level 3 the Molecular Function (3,073 sequences) and Cellular Component (2,946 sequences), were the categories with most assigned sequences for both species, and the Biologic process the least represented one (1,964/1,508 annotated sequences respectively). Here, the most abundant terms were “ATP binding”, “Protein binding” and “structural constituent of ribosome” in Molecular Function, and the least ones “ribosome biogenesis”, “translation” and “oxidaction-reduction process” in the Biologic Process’s category. Additionally, some GO terms could not be found in both transcriptomes. These were “negative regulation of apoptosis process”, extra-celular exosome” and “poly (A) RNA binding”.

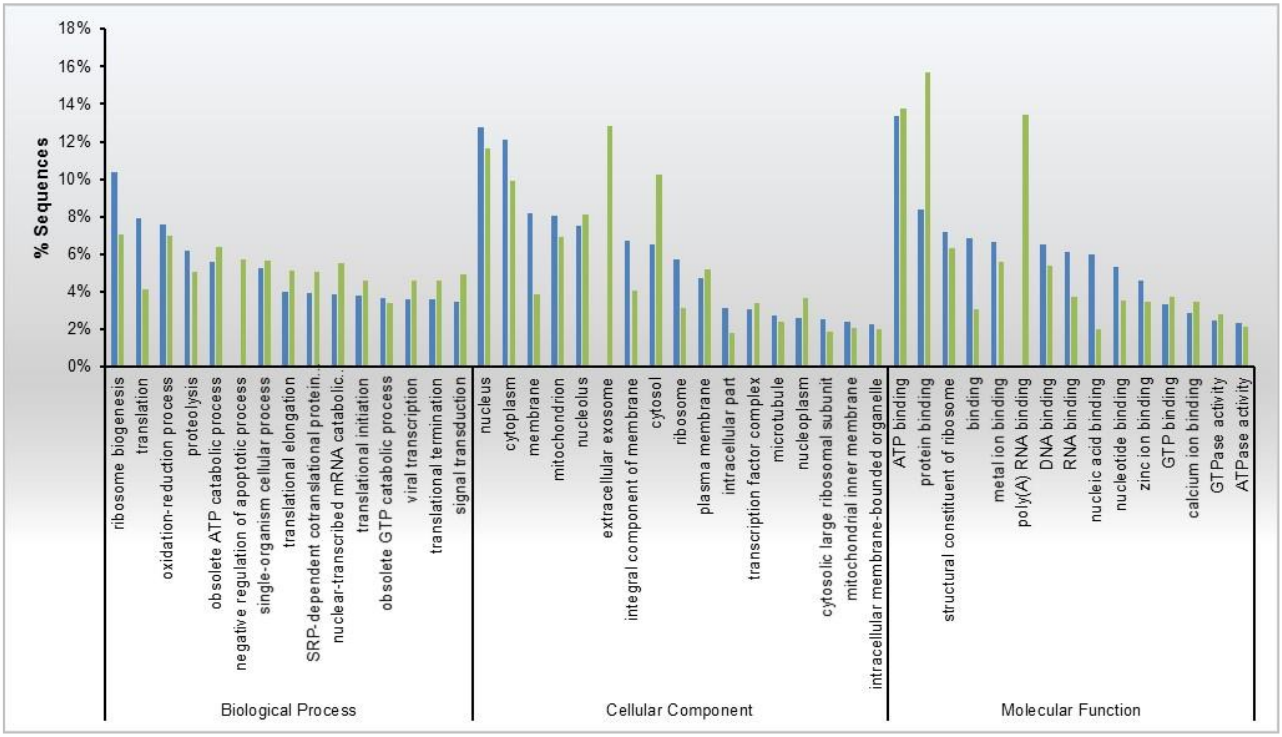


Figure 22 Paired comparison of the sequences mapped to a given gene ontology (GO) term in level 3. For each category (BP, CC, and MF) the top 20 GO terms for each species are represented (Blue – *F. cantabrica;* Green – *F. villafranca*). The five most represented categories are present for both species, and the remaing inlcude all common categories between the two transcriptomes.

Regarding the distribution of hits per species, for both datasets (*F. cantabrica* and *F. villafranca,* respectively) *Aplysia californica* and *Lottia gigantea* were the species with the highest number of contig representatives (Figure 23). Most of the top-hit species in both transcriptomes were other molluscs (*Aplysia*, *Lottia*, *Crassostrea*), yet, this list also included amphibians (*Xenopus sp.*), the sea-urchin *S. pupuratus* and rice plant (*O. sativa)*.

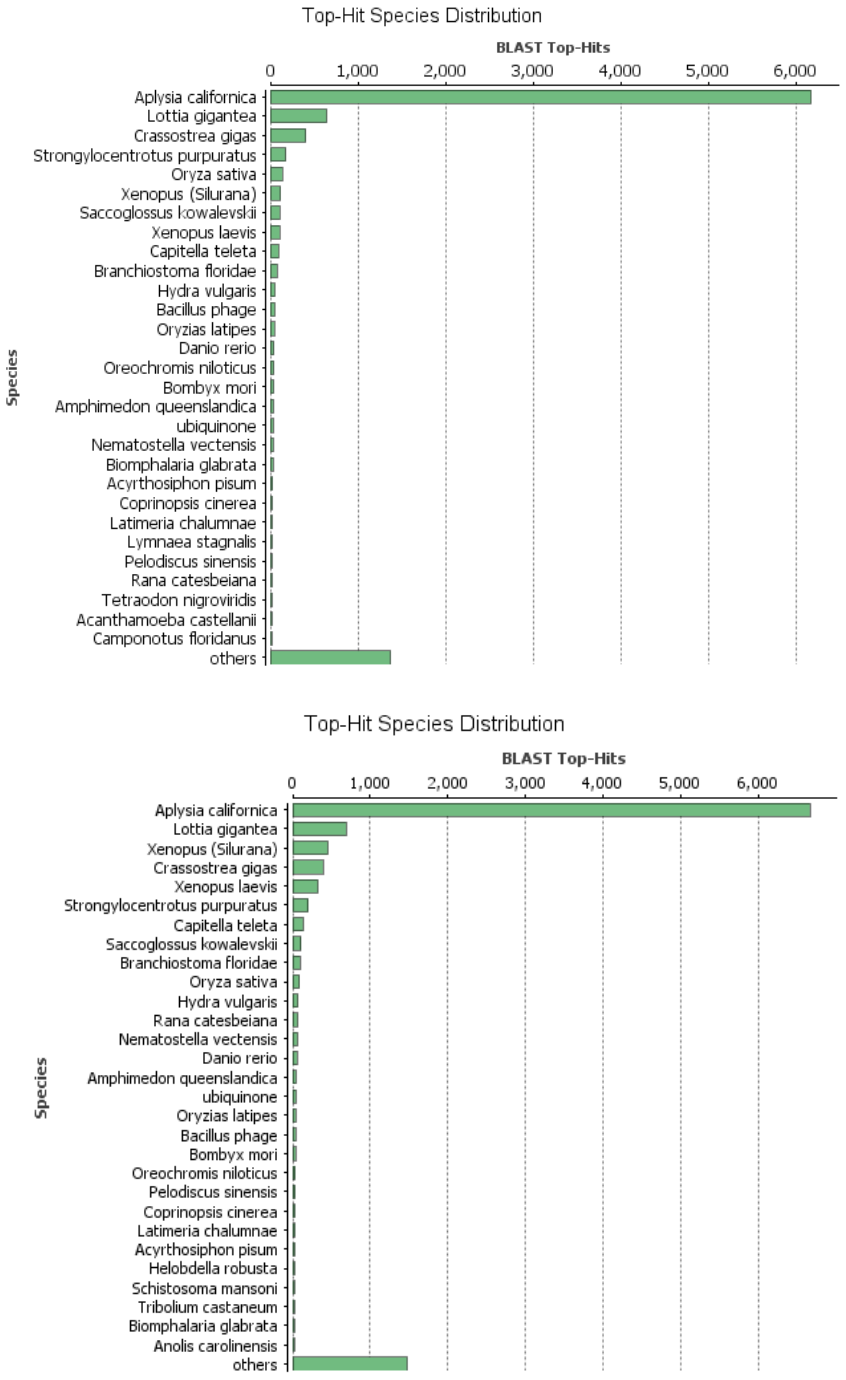


Figure 23 Best-hit species distribution for *F. cantabrica* (above) and *F. villafranca* (below).

3.3 MTDNA RECOVERY AND RECONSTRUCTION

3.3.1 Recovery through BLAST and alignment to a reference genome

The database containing Mollusca mtDNA genomes included, at the time of analysis, a total of 176 full genomes. From the Chromodorididae family the only mitochondrial genome sequenced so far is from *C. magnifica,* being here used as a reference to our species.

A total of 34 and 38 contigs were recovered for *F. cantabrica* and *F. villafranca* respectively. Of those, 20/23 corresponded to 13/13 coding regions, 16/18 tRNA's, and 2/2 ribossomal RNAs, partially or totally recovered for *F. cantabrica* and *F. villafranca*. Only tRNAs Asparagine (Asp), Glycine (Gly), Histone (His), Serine (L chain), in *F. cantabrica* and Valine (Val) and Leucine (Leu), in *F. villafranca,* could not be recovered. The remaining 14/15 contigs were more divergent than expected upon first examinations, and were later identified as non-Mollusca sequences through additional blastn and blastx analyses in the NCBI (see bellow).

Circularization (full molecule overlap) was confirmed for both mitogenomes. The inferred genome organization of the two mitogenomes seemed to be mostly identical to that of *C. magnifica*, as most of the mitogenome could be recovered with only a few contigs. Thus, the mitochondrial gene order of *F. cantabrica* was found to be identical to *C. magnifica* between 16S and tRNA Ser (H strand), ND3 and tRNA Cysteine (Cys), COX2 and tRNA Asp, and CYT B and tRNA Leu (four contigs); whereas for *F. villafranca* single contigs recovered from COI to ND2, COX3 to tRNA Asp, and CYT B to tRNA Proline (Pro). The ID of each region, its strand and length (reference and recovered) are given in Tables 7 and 8, for each mitogenome.

3.3.2 Protein-coding regions

A total of 13,980 bp and 14,208 bp were recovered by blasting *F. cantabrica* and *F. villafranca* (filtered) assemblies to a hand-made database of mitochondrial genomes of molluscs. Base composition was 36.21% of A's, 19.5% C's, 15.4% of G's, 28.4 T's and 33.4% of CG content for *F. cantabrica,* and 37.6% of A's, 18.4% of C's, 14.4% of G's, 29.5% of T's and 31.9% GC content for *F. villafranca*.

Respectively, 13 and 13 mitochondrial protein coding genes were identified comprising 75.4% and 75.3 % of the *F. cantabrica* and *F. villafranca* mitogenome alignment. A major fraction of the assembled coding regions seemed mostly complete. Comparing to the reference genome (Table 7 and 8) *F. cantabrica* assembled mitochondrial genome covers approximately 95.67% of the reference genome, with a nucleotide pairwise identity of 71.1%. *F. villafranca* assembled mitogenome covers

96.9% of the reference genome, with 72.7% pairwise identity. The most frequent start and stop codons were AUG (Met) and UAA, in both mitogenomes. When COI sequences (obtained by PCR amplification and Sanger sequencing) from these same individuals (OP0014 and OP0031) were aligned to the contigs recovered from each assembled transcriptome, 99.8% of identical sites were recovered for *F. cantabrica*, and 85.2% for *F. villafranca*.

3.3.3 Transfer and ribosomal RNAs

A total of 20 and 22 mitochondrial RNAs (tRNAs and rRNAs) were recovered by blast respectively for *F. cantabrica* and *F. villafranca*. Yet, only 18 and 20 of these were later validated by the tRNA/rRNA annotation tools implemented in the MITOS webserver. tRNAs were predicted with lengths ranging from 61 to 72 bp, for *F. cantabrica,* and 55 to 71 bp, for *F. villafranca*. A 166 bp sequence aligning to the one annotated as tRNA (Ala) in *C. magnifica* was recovered for *F. cantabrica,* but only 63 bp of these were validated by MITOS. In fact, this is an homologous size to the tRNA (Ala) from *Roboastra europea* and also, when one examines the published annotation of *C. magnifica* genome, this tRNA (Ala) sequence is found to be overlapping with tRNA(Pro) and part of the ND6 gene, thus probably being an annotation error. As reflected in Figure 2 of Medina et al. (2011), both *C. magnifica* and *R. europea* tRNA's in this region have the same size and order.

Secondary structures of tRNAs as inferred in MITOS are presented in Figures 22 and 23. For most, but not all *Felimare* tRNA genes can be folded into normal cloverleaf secondary structures. The majority of tRNA genes have seven-member aminoacyl stems, four to five member anticodon stems, and seven-member anticodon loops. Exceptions are found for Leu 1 (6 aminoacyl stems), and Thr (9 bp anticodon loops) in *F. cantabrica*, Cys, Leu 1, Arg and Thr (from 3 to 6 aminoacyl stems), His and Thr (9bp anticodon loops) in *F. villafranca*. These results are in accordance with other heterobranchs (Grande et al. 2002).

All the deduced tRNAs could be folded into a cloverleaf secondary structure with the exception of tRNA Ser 2 and tRNA Ser 1 in *F. villafranca,* that lack the DHU arm (at the left of the tertiary structure, D loop contains the base dihydrouracil, for which the arm is named). There were changes from our inferred structure to the common secondary cloverleaf structure of tRNAs Val, Tyr, Cys, Gln, Glu, Thr for *F. cantabrica*, and Tyr, Trp, Asp, Arg, Glu, Ser 1 and Thr with regard to *F. villafranca*. Such alterations involve the creation of slight deformations involving from one or two mismatches (e.g. Cys) to well-formed loops with no matching base pairs (e.g. Arg). Furthermore, tRNAs Trp and Thr of *F. cantabrica*, and Asp of *F. villafranca* presented unusual T arms whose stem does not have any pair of bases preceding the variable TψC-loop. At the same time, for the two His and Gln tRNAs the

variable TψC-loop does not form with our inferred sequence. If these are real differences or errors from assembly needs to be studied further.

The anti-codons of all tRNAs identified, from both mitogenomes, were all documented previously in other Heterobranchia species (e.g. *Pupa strigosa*, *Roboastra europea*).

Ribosomal RNA 16S was assembled with 1,104 bp on *F. cantabrica* and 1,022 bp on *F. villafranca*, while the small-subunit (12S), with 728 bp and 727 bp, respectively.

Table 7 Mitochondrial genome regions recovered by a BLAST search of the transcriptome of F. cantabrica against available mollusc mtDNA genomes. Reference is *C. magnifica* mitochondrial genome.

***F. cantabrica* (OP0014)**									
Gene	**Position (Ref)**		**Size (ungapped)**		**Intergenic Ref. (OP0014)**	**Codon**		**%Differences**	**Strand**
	From	**To**	**Ref. (bp)**	**OP0014 (bp)**		**Start**	**Stop**		
COI	1	1 533	1 553	1 539	8 (8)	AUG (AUG)	UAG (UAA)	20	H
tRNA (Val)	1542	1606	65	65	0 (-1)	-	-	26,9	H
16s rRNA	1608	2698	1 092	1 104	0	-	-	28,7	H
tRNA (Leu2)	2699	2763	65	68	-4 (-3)	-	-	19,7	H
tRNA (Pro)	2706	2924	66	65	-66 (-64)	-	-	15,2	H
tRNA (Ala)	2830	2895	165	63	41 (41)	-	-	25,4	H
ND 6	2896	3366	471	468	16 (16)	UUG (UUG)	UAA (UAA)	29,2	H
ND 5 - like	3383	5007	1 625	1 628	2 (4)	-	-	25,6	H
ND 1*	5010	5924	915	912	3 (3)	GUG (GUG)	UAA (UAG)	25,4	H
tRNA (Tyr)	5928	5994	67	65	-1 (3)	-	-	17,9	H
tRNA (Trp)	5994	6060	67	70	3 (9)	-	-	24,6	H
ND 4L**	6064	6340	277	277	0 (0)	AUA (AUA)	UAA (UA(A))	30,8	H
CYT B**	6341	7454	1 114	1 123	3 (0)	AUG (AUA)	UAA (U(AA))	24,7	H
tRNA (Asp)	7455	7525	71	17	-1	-	-	35,3	H
tRNA (Phe)	7525	7592	68	42	191	-	-	4,7	H
COX 2	7784	8458	675	670	2	AUG (AUG)	UAA (-)	19,9	H
tRNA (Gly)	8461	8525	65	-	2	-	-	-	H
tRNA (His)	8529	8547	19	-	-2	-	-	-	H
tRNA (Cys)	8595	8658	64	65	29 (77)	-	-	20	H
tRNA (Gln)	8750	8688	63	61	5 (6)	-	-	20,3	L
tRNA (Leu1)	8821	8756	66	63	5 (5)	-	-	18,5	L
ATP 8	8991	8827	165	39	0 (1)	AUG (AUG)	UAA (-)	17,9	L
tRNA (Asn)	9059	8992	68	67	4 (4)	-	-	19,1	L
ATP 6*	9726	9064	663	663	0	AUG (AUG)	UAA (UAA)	27,3	L
tRNA (Arg)	9789	9727	63	65	2 (1)	-	-	16,7	L
tRNA (Glu)	9855	9792	64	67	0	-	-	16,2	L
12S rRNA	10586	9856	731	728	1 (1)	-	-	25	L
tRNA (Met)	10652	10588	65	66	-2 (-2)	-	-	17,6	L
ND 3	11004	10651	354	354	4 (1)	AUG (AUG)	UAA (UAA)	27,1	L
tRNA (Ser2)	11065	1009	57	-	6 (0)	-	-	-	L
tRNA (Ser1)	11072	11134	63	48	0	-	-	12,1	H
ND 4**	11135	12483	1 349	1 281	-1 (0)	UAC (GUG)	UAA ((UAA))	30,9	H
tRNA (Thr)	12549	12483	67	72	0 (1)	-	-	23,6	L
COX 3**	13333	12550	784	786	46 (44)	AUU (UUG)	UAA (UAG)	22,2	L
tRNA (Ile)	13380	13445	66	66	0	-	-	16,7	H
ND 2	13446	14381	936	830	-1	GUG (AUG)	UAG (-)	31,4	H
tRNA (Lys)	14381	14445	65	66	1 (1)	-	-	25,8	H

Start and stop codons reference (and *Felimare*) coding genes are given when found. Reference intergenic regions (and intergenic regions recovered) as well as the percentage of differences relative to the reference sequence are also given. Highlighted cells (red) are short (very incomplete) fragments. Codons with lowercase TAA's (in double brackets) are incomplete termination codons, presumably completed by polyadenylation.

Table 8 Mitochondrial genome recovered by blasting the transcriptome of *F. villafranca* against available mollusc mtDNA genomes.

***F. villafranca* (OP0031)**									
Gene	**Position (Ref)**		**Size (ungapped)**		**Intergenic Ref. (OP0031)**	**Codon**		**% Differences**	**Strand**
	From	**To**	**Ref. (bp)**	**OP0031 (bp)**		**Start**	**Stop**		
COI	1	1 533	1 553	1 533	8 (5)	AUG (-)	UAG (UAA)	18,3	H
tRNA (Val)	1 542	1 606	65	-	0	-	-	-	H
16s rRNA	1 608	2 698	1 092	1 022	0	-	-	27,5	H
tRNA (Leu2)	2 699	2 763	65	-	-3	-	-	-	H
tRNA (Pro)	2 830	2 895	66	29	-64 (-64)	-	-	16,7	H
tRNA (Ala)	2 706	2 924	165	58	-4	-	-	23,7	H
ND 6	2 896	3 366	471	468	16 (16)	UUG (UUG)	UAA (UAA)	28,9	H
ND 5 - like	3 383	5 007	1 625	1 625	2 (2)	-	-	24,7	H
ND 1	5 010	5 924	915	735	3 (2)	GUG (-)	UAA (U(AA))	25,5	H
tRNA (Tyr)	5 928	5 994	67	65	-1 (1)	-	-	23,5	H
tRNA (Trp)	5 994	6 060	67	64	3 (9)	-	-	23,3	H
ND 4L	6 064	6 340	277	279	0	AUA (AUA)	UAA (UA(A))	29,6	H
CYT B**	6 341	7 454	1 114	729	0	AUG (AUG)	UAA ((UAA))	21,8	H
tRNA (Asp)	7 455	7 525	71	62	-1 (2)	-	-	24,2	H
tRNA (Phe)	7 525	7 592	68	65	191 (0)	-	-	14,7	H
COX 2	7 784	8 458	675	675	2 (4)	AUG (AUG)	UAA (UAA)	17,3	H
tRNA (Gly)	8 461	8 525	65	63	2 (3)	-	-	40,6	H
tRNA (His)	8 529	8 547	69	63	-2 (5)	-	-	28,1	H
tRNA (Cys)	8 595	8 658	64	67	29 (4)	-	-	22,1	H
tRNA (Gln)	8 750	8 688	63	55	5 (4)	-	-	15,2	L
tRNA (Leu1)	8 821	8 756	66	61	4 (5)	-	-	23,1	L
ATP 8	8 991	8 827	165	39	0	AUG (AUG)	UAA (-)	41,5	L
tRNA (Asn)	9 059	8 992	68	68	4 (10)			18,8	L
ATP 6	9 726	9 064	663	672	0	AUG (AUG)	UAA (UAA)	25,1	L
tRNA (Arg)	9 789	9 727	63	62	2 (1)	-	-	20,6	L
tRNA (Glu)	9 855	9 792	64	67	0 (0)	-	-	13,2	L
12S rRNA	10 586	9 856	731	727	1 (1)	-	-	22	L
tRNA (Met)	10 652	10 588	65	64	-2	-	-	9,1	L
ND 3	11 004	10 651	354	354	4 (4)	AUG (GUG)	UAA (UAA)	26,3	L
tRNA (Ser2)	11 065	1 009	57	59	6 (4)	-	-	23,7	L
tRNA (Ser1)	11 072	11 134	63	61	0	-	-	14,3	H
ND 4	11 135	12 483	1 349	1 281	-1	UAC (GUG)	UAA (-)	30,6	H
tRNA (Thr)	12 549	12 483	67	63	0	-	-	30,2	L
COX 3**	13 333	12 550	775	788	46 (41)	AUU (-)	UAA ((UAA))	26,1	L
tRNA (Ile)	13 380	13 445	66	66	0	-	-	18,2	H
ND 2	13 446	14 381	936	843	-1	GUG (AUG)	UAG (-)	30,4	H
tRNA (Lys)	14 381	14 445	65	161	1 (1)	-	-	23,4	H

Reference is *C. magnifica* mitochondrial genome. Start and stop codons reference (and *Felimare*) coding genes are given when found. Reference intergenic regions (and intergenic regions recovered) as well as the percentage of differences relative to the reference intergenic regions (and intergenic regions recovered) Highlighted cells (red) are short (very incomplete) fragments. Codons with lowercase TAA's (in double brackets) are incomplete termination codons, presumably completed by polyadenylation.

3.4 MTDNA ASSEMBLY DIRECTLY FROM READS USING MITOBIM

The ten million reads used as input to MITObim correspond to 13% and 14.1% of the initial reads of *F. cantabrica* and *F. villafranca*, respectively.

In none of the species was the mitochondrial genome completely assembled (Tables 9 and 10). In total, 14,662 bp and 14,520 bp, were recovered for *F. cantabrica* and *F. villafranca*, respectively. Base composition of the mitogenomes consist of 17.1% of A's, 9.2% C's, 11.1% of G's, 21.7 T's and 18.8% of CG content, for *F. cantabrica,* while *F. villafranca* has 20.8% of A's, 10.1% of C's, 12.5% of G's, 26.1% of T's and 21.9 % of GC content.

Circularization was confirmed for *F. cantabrica* only, by the overlap of the start and end mitogenome regions assembled (ND2, tRNA Lysine – Lys - and COI).

3.4.1 Protein–coding regions

Respectively, 12 and 11 mitochondrial protein coding genes were identified thus comprising 68.6% and 57.9% of the *F. cantabrica* and *F. villafranca* mitogenomes alignment. These numbers include sequences with vast amounts of ambiguities (>50%) and small sparse regions within genes (at ND6, ATP6, ND4L and ND4, in *F. cantabrica*, and ND4 and ND2 in *F. villafranca* - Tables 9 and 10).

For both recovered mitochondrial genomes, ND5 was the longest fragment, with 1,607 and 1,622 bp, respectively for *F. cantabrica* and *F. villafranca*, but it is COI that yields less percentage of ambiguities. ATP 8 was the smallest gene of the mitochondrial genomes, and yet it was never completely recovered, with 112 and 153 bp (out of 165) respectively. The predominant start/stop codons were, respectively, AUG/UAA in both transcriptomes.

PCR amplified sanger-sequenced COI fragments from OP0014 and OP0031 were 49,2% and 40,9% identical to these MITObim-assembled regions.

3.4.2 Transfer and ribosomal RNAs

For each transcriptome a total of 18/2 and 15/1 tRNA/rRNAs (respectively for *F. cantabrica* and *F. villafranca*) were fully or partially assembled with MITObim. From these, respectively 9/1 and 8/1 tRNA/rRNAs were validated by MITOS (Tables 9 and 10). The remaining tRNAs and rRNA could not be found or validated independently as their sequences were too short, or had a big amount of ambiguities (highlighted in red at Tables 9 and 10). Inferred cloverleaf secondary structures of tRNAs are represented in Figures 24 and 25. Inferred secondary structures of tRNAs are presented in

Figures 24 and 25. Most of the *Felimare* tRNAs could be folded into normal cloverleaf secondary structures. The few exceptions are Trp, which lacks the variable TψC-loop, and Ser, that lacks the DHU arm, in *F. villafranca.* Again, I found it (lack of DHU arm) recurrent in other heterobranchs (e.g. Kurabayashi & Ueshima 2000; Grande et al. 2002; Groenenberg et al. 2012). The majority of tRNA genes had seven-member aminoacyl stems, four to five member anticodon stems, and seven-member anticodon loops. Exceptions are found for Tyrosine (Tyr) (2 aminoacyl stems), in both *F. cantabrica* and *F. villafranca*, and Thr (9 member anticodon loop). The anti-codons of all tRNAs, from both mitogenomes, were all documented previously in other Heterobranchia species (e.g. *Pupa strigosa*, *Roboastra europea*). The small-subunit rRNA (12S) was recovered for both species, with 729 bp and 728 bp, respectively in *F. cantabrica* and *F. villafranca,* with more than 70% of similarity to the reference.

Table 9 Assembled regions of mitochondrial genomes of *F. cantabrica* using MITObim.

F. cantabrica (OP0014)									
Gene	Position (Ref.)		Size (ungapped)		Intergenic nucleotide Ref. (OP0014)	% AMB	Codon		Strand
	From	To	Ref. (bp)	Op0014 (bp)			Start	Stop	
COI	1	1 533	1 553	1535	8 (0)	16,11	AUG (AUG)	UAG (UAA)	H
tRNA (Val)	1 546	1 606	65	24	0	0	-	-	H
16S rRNA	1 607	2 698	1 092	960	0	49,3	-	-	H
tRNA (Leu)	2 699	2 763	65	-	-4 (0)	100	-	-	H
tRNA (Ala)	2 760	2 928	66	-	-66 (0)	100	-	-	H
tRNA (Pro)	2 830	2 895	165	-	0	100	-	-	H
ND 6	2 896	3 366	471	456	16 (0)	89,8	UUG (-)	UAA (-)	H
ND 5-like	3 383	5 007	1 625	1 607	2 (0)	30,8	-	-	H
ND 1	5 010	5 924	915	915	3 (1)	17,6	GUG (-)	UAA (UAG)	H
tRNA (Tyr)	5 929	5 994	67	67	-1 (-1)	1	-	-	H
tRNA (Trp)	5 994	6 060	67	70	3 (3)	0	-	-	H
ND 4L	6 064	6 340	277	222	0	85,6	AUA (AUA)	UAA (-)	H
CYT B	6 341	7 450	1 114	1 119	0	17,8	AUG (-)	UAA (-)	H
tRNA (Asp)	7 455	7 525	71	22	-1	0	-	-	H
tRNA (Phe)	7 525	7 592	68	65	191 (82)	0	-	-	H
COX 2	7 784	8 458	675	606	2 (4)	25,3	AUG (-)	UAA (UAA)	H
tRNA (Gly)	8 461	8 525	65	63	2 (2)	0	-	-	H
tRNA (His)	8 528	8 596	19	-	-2	100	-	-	H
tRNA (Cys)	8 595	8 658	64	-	29 (8)	100	-	-	H
tRNA (Gln)	8 688	8 750	63	18	5	0	-	-	H
tRNA (Leu)	8 821	8 756	66	6	5	0	-	-	L
ATP 8	8 991	8 827	165	112	0	0	AUG (AUG)	UAA (-)	L
tRNA (Asn)	9 059	8 992	68	68	4 (12)	1,5	-	-	L
ATP 6	9 726	9 064	663	662	0	77,8	AUG (AUG)	UAA (UAA)	L
tRNA (Arg)	9 789	9 727	63	65	2 (1)	0	-	-	L
tRNA (Glu)	8 995	9 780	64	67	0	1,5	-	-	L
12S rRNA	10 586	9 856	731	729	1	8	-	-	L
tRNA (Met)	10 652	10 588	65	59	-2 (-2)	18,6	-	-	L
ND 3	11 004	10 651	354	354	4 (1)	6,5	AUG (AUG)	UAA (UAA)	L
tRNA (Ser)	11 065	11 009	57	9	6 (6)	0	-	-	L
tRNA (Ser)(2)	11 072	11 134	63	22	0	0	-	-	L
ND 4	11 135	12 483	1 349	1 468	-1 (1)	69,8	AUG (UUG)	UAA ((UAA))	H
tRNA (Thr)	12 549	12 483	67	72	0 (1)	0	-	-	L
COX 3	13 333	12 550	784	801	46 (42)	34,2	AUU (AUN)	UAA ((UAA))	L
tRNA (Ile)	13380	13 445	66	69	0	28,9	-	-	H
ND 2	13 446	14 381	936	936	-1 (-1)	62,4	GUG (-)	UAG (UAA)	H
tRNA (Lys)	14 381	14 445	65	68	1 (1)	0	-	-	H

Start and stop codons of reference (and *Felimare*) coding genes are given when found. Reference intergenic regions (and intergenic regions recovered) as well as the percentage of differences relative to the reference sequence are also given. Highlighted cells (red) are short (very incomplete) fragments. Codons with lowercase TAA's (in double brackets) are incomplete termination codons, presumably completed by polyadenylation.

Table 10 Assembled regions of mitochondrial genome of *F. villafranca* using MITObim.

***F. villafranca* (OP0031)**									
Gene	**Position (Ref.)**		**Size (ungapped)**		**Intergenic Ref. (OP0031)**	**% AMB**	**Codon**		**Strand**
	From	**To**	**Ref. (bp)**	**Op0031 (bp)**			**Start**	**Stop**	
COI	1	1533	1 553	1541	8 (8)	8,7	AUG (AUG)	UAG (UAG)	H
tRNA (Val)	1 546	1606	65	99	0 (1)	12,1	-	-	H
16S rRNA	1 607	2698	1 092	1 287	0	60,6	-	-	H
tRNA (Leu)	2 699	2763	65	-	-4 (0)	100	-	-	H
tRNA (Ala)	2 760	2928	66	-	-66 (0)	100	-	-	H
tRNA (Pro)	2 830	2895	165	-	0	100	-	-	H
ND 6	2 896	3366	471	456	16 (8)	37,5	UUG (-)	UAA (UAA)	H
ND 5-like	3 383	5007	1 625	1 622	2 (4)	4,2	-	-	H
ND 1	5 010	5 924	915	915	3 (1)	3,6	GUG (AUG)	UAA (UAG)	H
tRNA (Tyr)	5 929	5 994	67	67	-1 (-1)	1,5	-	-	H
tRNA (Trp)	5 994	6 060	67	70	3 (3)	0	-	-	H
ND 4L**	6 064	6 340	277	230	0	1,7	AUA (AUA)	UAA (UA(A))	H
CYT B**	6 341	7 450	1 114	1 125	0	8,2	AUG (AWG)	UAG (U(AA))	H
tRNA (Asp)	7 455	7 525	71	45	-1	15	-	-	H
tRNA (Phe)	7 525	7 592	68	47	191 (27)	2,1	-	-	H
COX 2	7 784	8 458	675	675	2	17,8	AUG (AUG)	UAA (-)	H
tRNA (Gly)	8 461	8 525	65	-	2	100	-	-	H
tRNA (His)	8 528	8 596	19	-	-2	100	-	-	H
tRNA (Cys)	8 595	8 658	64	-	29	100	-	-	H
tRNA (Gln)	8 688	8 750	63	-	5	100	-	-	L
tRNA (Leu)	8 821	8 756	66	84	5 (11)	40,5	-	-	L
ATP 8	8 991	8 827	165	156	0 (1)	1,3	AUG (AUG)	UAA (UAA)	L
tRNA (Asn)	9 059	8 992	68	68	4 (19)	0	-	-	L
ATP 6	9 726	9 064	663	665	0	38,2	AUG (AUG)	UAA (UAA)	L
tRNA (Arg)	9 789	9 727	63	66	2 (1)	0	-	-	L
tRNA (Glu)	8 995	9 780	64	67	0	0	-	-	L
12S rRNA	10 586	9 856	731	728	1	1,6	-	-	L
tRNA (Met)	10 652	10 588	65	64	-2 (-2)	9,1	-	-	L
ND 3	11 004	10 651	354	356	4 (31)	2,5	AUG (-)	UAA (UAA)	L
tRNA (Ser)	11 065	11 009	57	59	6 (4)	0	-	-	L
tRNA (Ser)(2)	11 072	11 134	63	63	0	56,4	-	-	H
ND 4	11 135	12 483	1 349	1 354	-1 (-1)	96,55	AUG (-)	UAA (-)	H
tRNA (Thr)	12 549	12 483	67	68	0	11,8	-	-	L
COX 3	13 333	12 550	784	821	46 (44)	39,1	AUU (-)	UAA (-)	L
tRNA (Ile)	13 380	13 445	66	66	0	0	-	-	H
ND 2	13 446	14 381	936	242	-1 (-1)	63,2	GUG (AUG)	UAG (-)	H
tRNA (Lys)	14 381	14 445	65	66	1 (2)	0	-	-	H

Start and stop codons of reference (and *Felimare*) coding genes are given when found. Reference intergenic regions (and intergenic regions recovered) as well as the percentage of differences relative to the reference sequence are also given. Highlighted cells (red) are short (very incomplete) fragments. Codons with lowercase TAA's (in double brackets) are incomplete termination codons, presumably completed by polyadenylation.

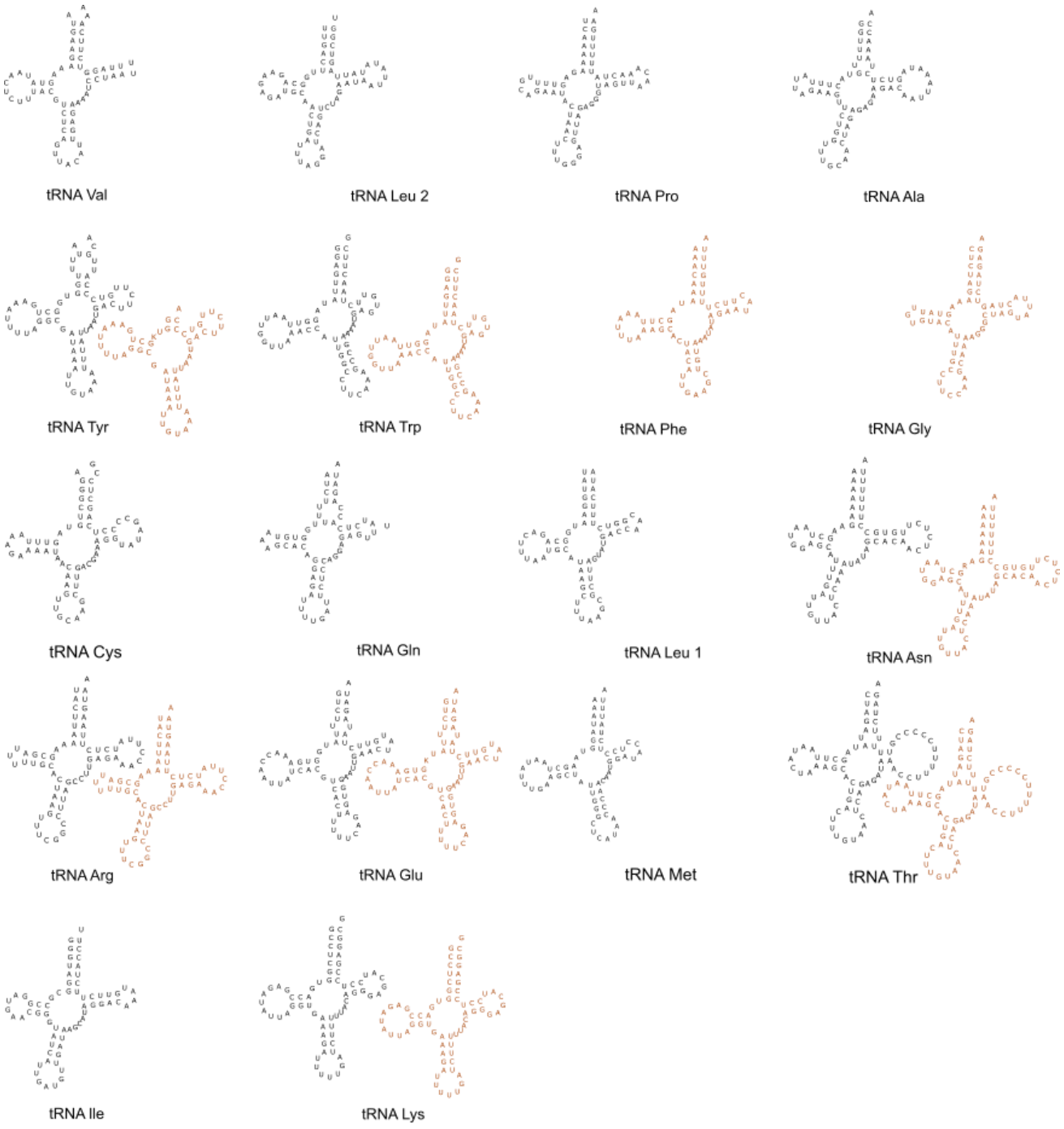


Figure 24 Putative tRNAs cloverleaf secondary structures for *F. cantabrica*. Sixteen (black) and nine (orange) tRNAs were identified, respectively, from blast of transcriptome contigs and reads assembly (MITObim) of *F. cantabrica* using *C. magnifica* mitogenome as reference. tRNAs were inferred with the Internal software module, which is part of the MITOS annotation pipeline.

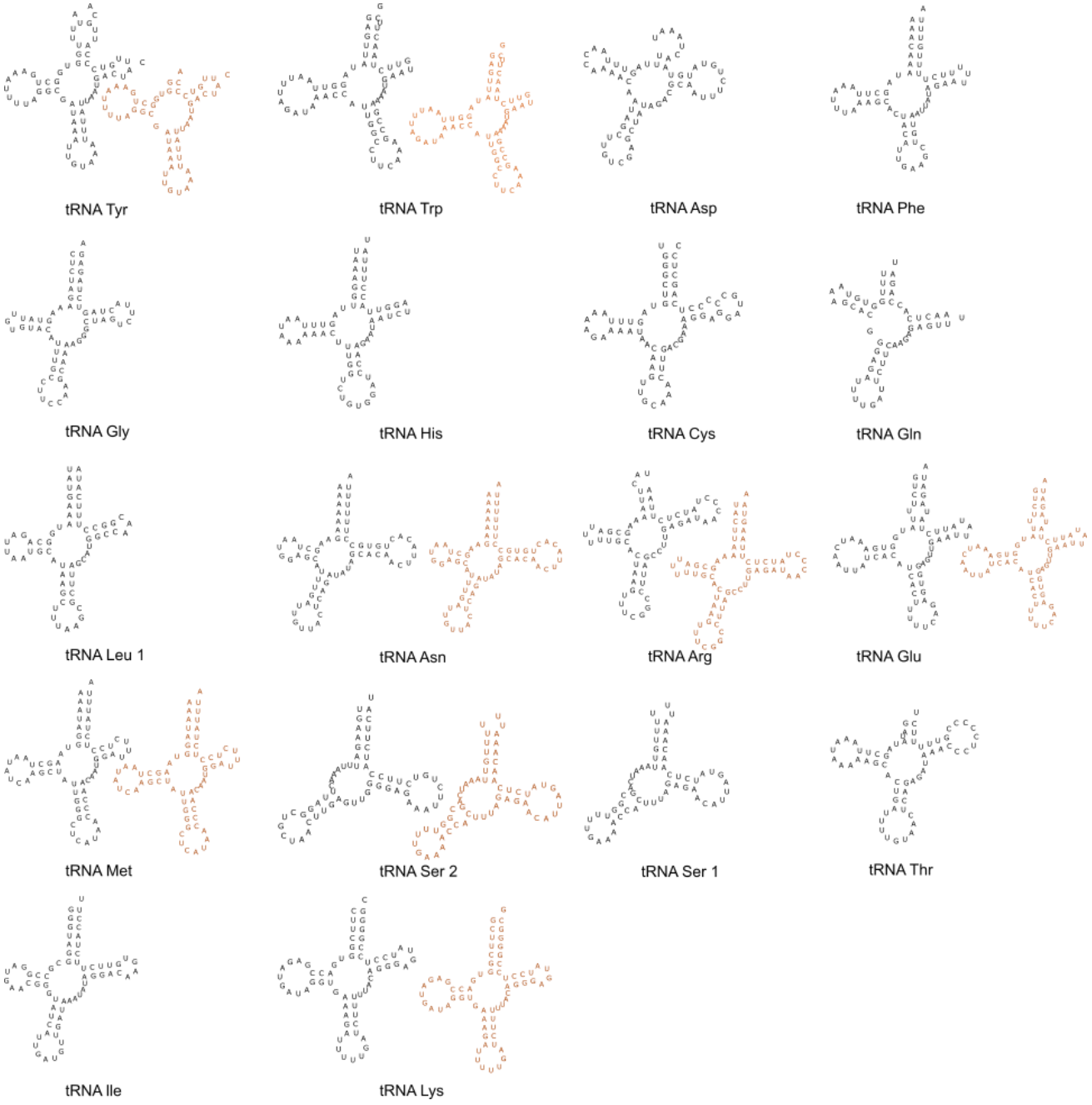


Figure 25 Putative tRNAs cloverleaf secondary structures for *F. villafranca* eighteeb (black) and eight (orange) tRNAs were identified, respectively, from blast of transcriptome contigs and reads assembly (MITObim) of *F. villafranca* using *C. magnifica* mitogenome as reference. tRNAs were inferred with the Internal software module, which is part of the MITOS annotation pipeline.

3.5 DETECTION AND FILTERING OF AMPHIBIAN MTDNA CONTAMINATION

As explained in the Material and Methods section, some contigs were found to be highly divergent from *Chromodoris magnifica* (and other molluscs) mtDNA. Quick BLAST searches against NCBI nr database revealed them to be amphibian mtDNA belonging to the toad *Epidalea calamita*. Of the contigs recovered through the BLAST against the Mollusca mtDNA database, 15 of a total of 34 for *F. cantabrica,* and 20 of a total of 38 in *F. villafranca*, were *E. calamita* mtDNA.

Thus I built a database with several coding/non-coding regions and some additional Nudibranchs and Anura mitochondrial genomes (See section 2.2.6 of Material & Methods), using both taxonomic criteria and previous results from annotation (Figure 23). I used that database against the transcriptome assemblies as an attempt to evaluate the extension of this contamination and, if possible, exclude “amphibian” contigs for further analyses (Table 1).

Using the default BLAST limit on the number of hits (500), most query sequences returned a low (1-5) number of hits at both nucleotide and protein blast, against the previously mentioned database (Figures 26 and 27).

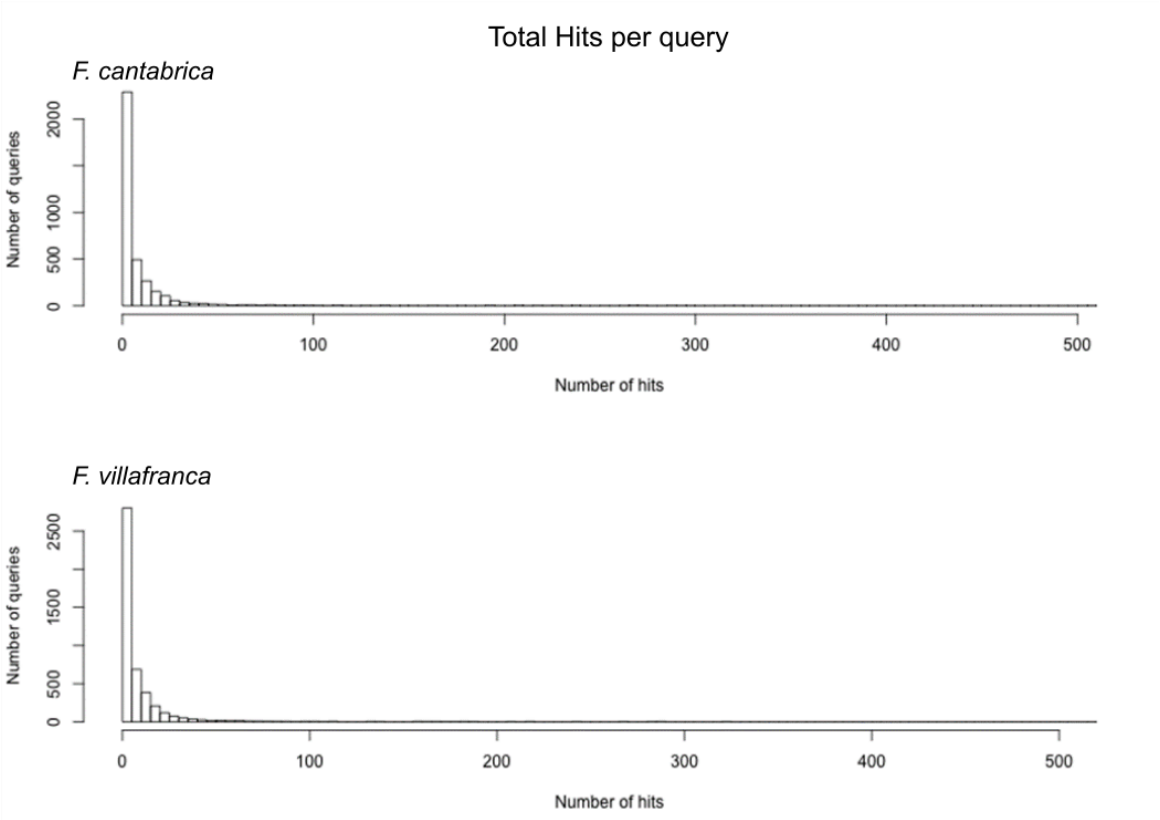


Figure 26 Number of hits per sequence for blastn (nucleotide) analyses against nuclear and mtDNA coding regions across molluscs and other invertebrate and vertebrate species for each transcriptome. (graph truncated as blastp returned hits can be much higher – as they may include several hits in the same target sequence).

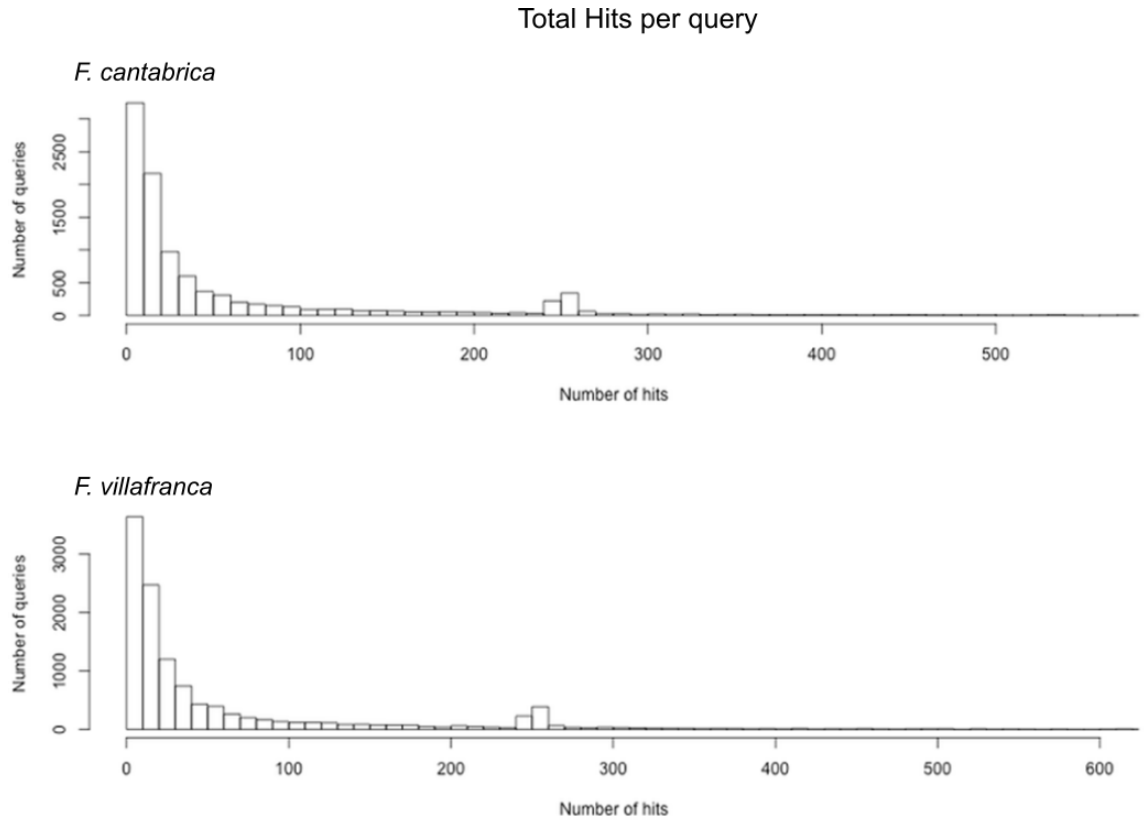


Figure 27 Number of hits per sequence for blastx (protein) analyses against nuclear and mtDNA coding regions across molluscs and other invertebrate and vertebrate species for each transcriptome (graph truncated as blastp returned hits can be much higher – as they may include several hits in the same target sequence).

Table 11 presents the number of sequences from each transcriptome with hits, as well as the best-hit distribution across the genomes in the database. Approximately 6.5% (3,582 sequences) and 18.9% (10,345 sequences) of the total contigs from *F. cantabrica*, as well as 7.7% (4,545 sequences), and 20.34% (11,914 sequences) of *F. villafranca* contigs had hits with sequences in the database (nucleotide and protein sequences, respectively). Contigs matching all taxa (invertebrates and vertebrates) included in the database were also found. Under both strategies and in both (*F. cantabrica* and *F. villafranca)* transcriptomes, the genome with the highest number of nucleotide hits was *Aplysia californica*, followed by *Xenopus tropicalis* and *Lottia gigantea (*Figures 28 and 29). These numbers changed when we look at results from the protein blasts, with an increase of “Mollusca” - homolog sequences retrieved.

Table 11 Number of sequences of transcritpomes with hits, as well as the best-hit distribution across nuclear and mtDNA coding regions of molluscs and other invertebrate and vertebrate species.

	OP0014 blastN	OP0031 blastN	OP0014 blastX	OP0031 blastX	
Input	54664	58554	54664	58554	
Sequences with Hits	3582 (6,5%)	4545 (7,7%)	10345 (18,9%)	11914 (20,34%)	
Hits					**Source**
A. californica	2207	2458	6849	7412	cDNA
Unique Hits	1206	1385	1057	1180	
N. gardineri	6	5	8	4	Mt Genome
C. magnifica	8	7	11	11	Mt Genome
	18	23	-	-	18s/28s
L. gigantea	254	238	825	862	cDNA
C. gigas	175	181	515	518	cDNA
C. teleta	98	87	227	235	cDNA
Unique Hits	29	27	65	47	
S. purpuratus	77	63	715	789	cDNA
S. kowallevski	56	64	215	218	cDNA
Unique Hits	24	27	33	31	
Branchiostoma			161	155	cDNA
S. tropicalis	346	895	340	996	cDNA
Unique Hits	76	243	19	35	
B. japonicus	5	5	5	4	Mt Genome
B. melanocosticus	10	12	7	6	Mt Genome
E. calamita	1	5	-	-	18S
C. picta	97	200	123	234	cDNA
D.rerio	74	89	107	146	cDNA
O. nilotica	73	95	143	174	
H. sapiens	85	128	93	148	cDNA
Control_Felimare	BEST HIT	3	BEST HIT	4	Original contig
Control_Bufo	HIT	1	HIT	2	Original contig
Total Unique Non-Vertebrata	1972	2158	3227	3487	
Total Unique Vertebrata	377	926	143	306	
Total shared Non/Vertebrata	1229	1455	6975	8121	

Mollusca, Gastropoda; Mollusca, Bivalvia; Annelida, Polichaeta; Echinodermata, Echinoidea; Hemichordata, Enteropneusta; Chordata, Cephalochordata; Chordata, Amphibia; Chordata, Reptilia; Chordata, Actinopterygii; Chordata, Actinopterygii, Periciformes; Chordata, Mammalia

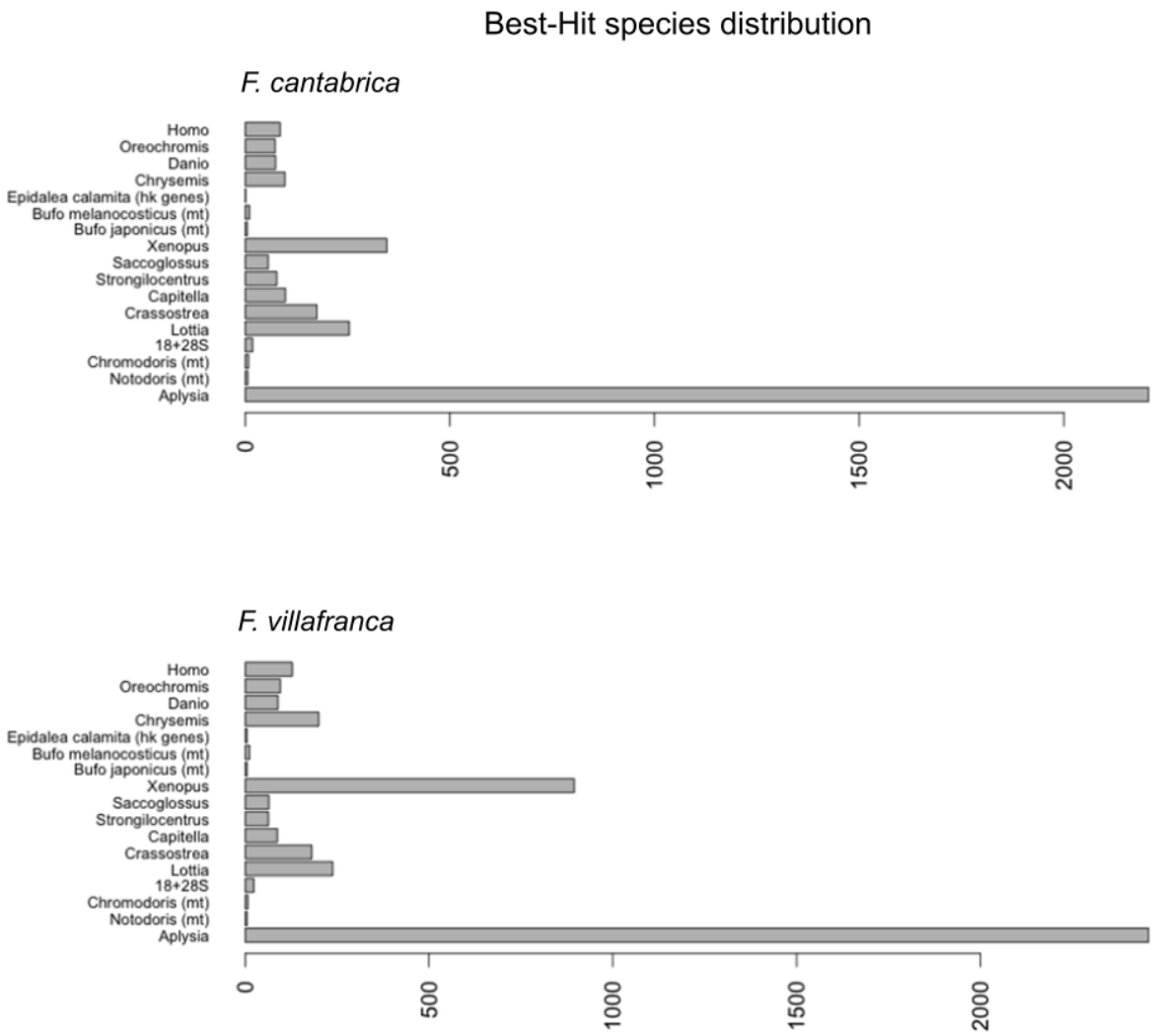


Figure 28 Number of sequences with best hits against genomes in the database (blastn).

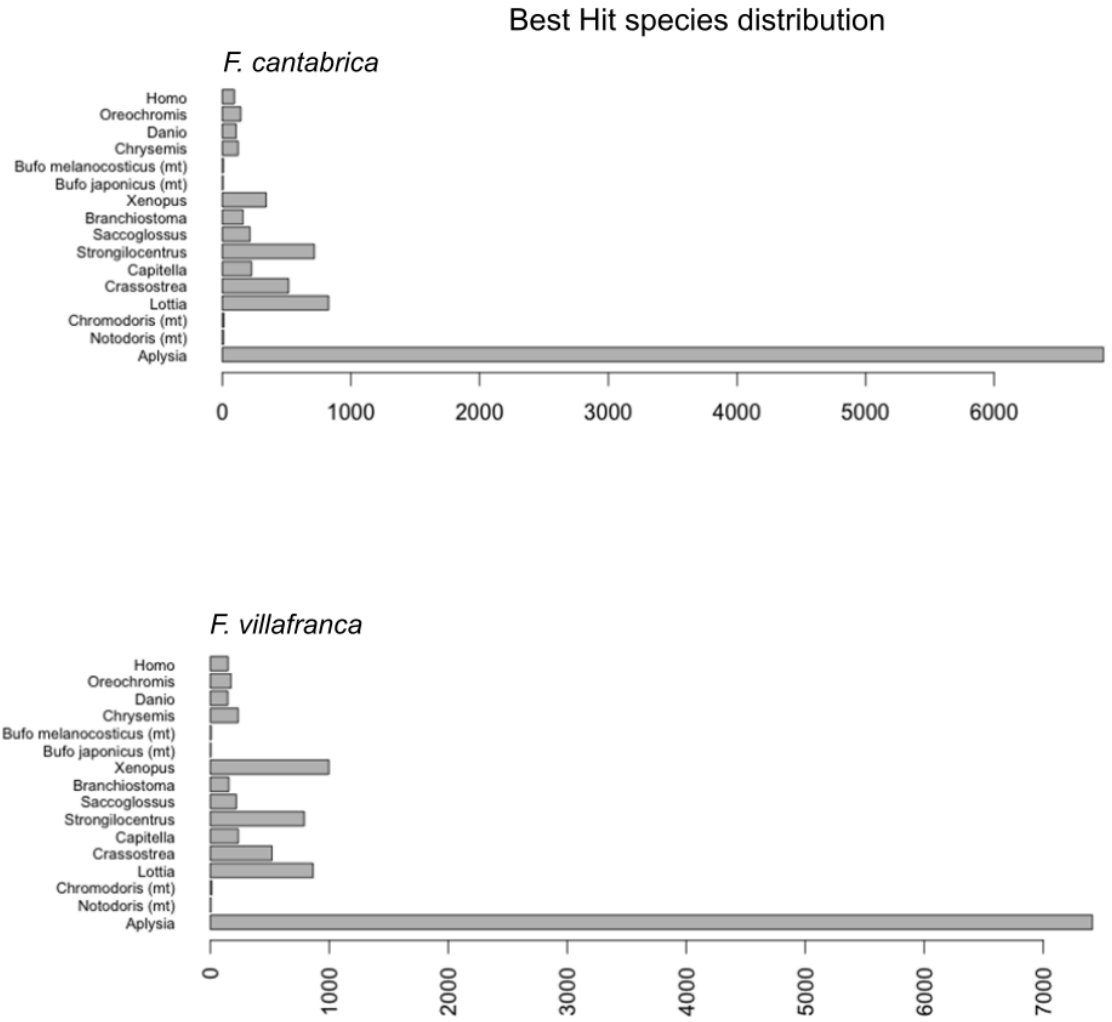


Figure 29 Number of sequences with best hits against genomes in the database (blastx).

Overall, the total number of “unique hits” (sequences for which hits were only observed in a certain taxonomic group) for non-vertebrata was significantly higher than for vertebrata, as expected. There was a great number of sequences that had hits simultaneously with vertebrata and non-vertebrata, a result more evident in the protein blast (blastx), as expected. Several blast parameters (frequency, percentage of identity, alignment length) were plotted across different e-value categories, to better evaluate the reliability of these hits. Overall, the most frequent classes of e-value were the two highest ones ($> 1e^{-50}$), with an overall correlation between relationship was more evident in the protein blast results (Figures 30 and 31). These plots were very similar when using all hits or only the best hits.

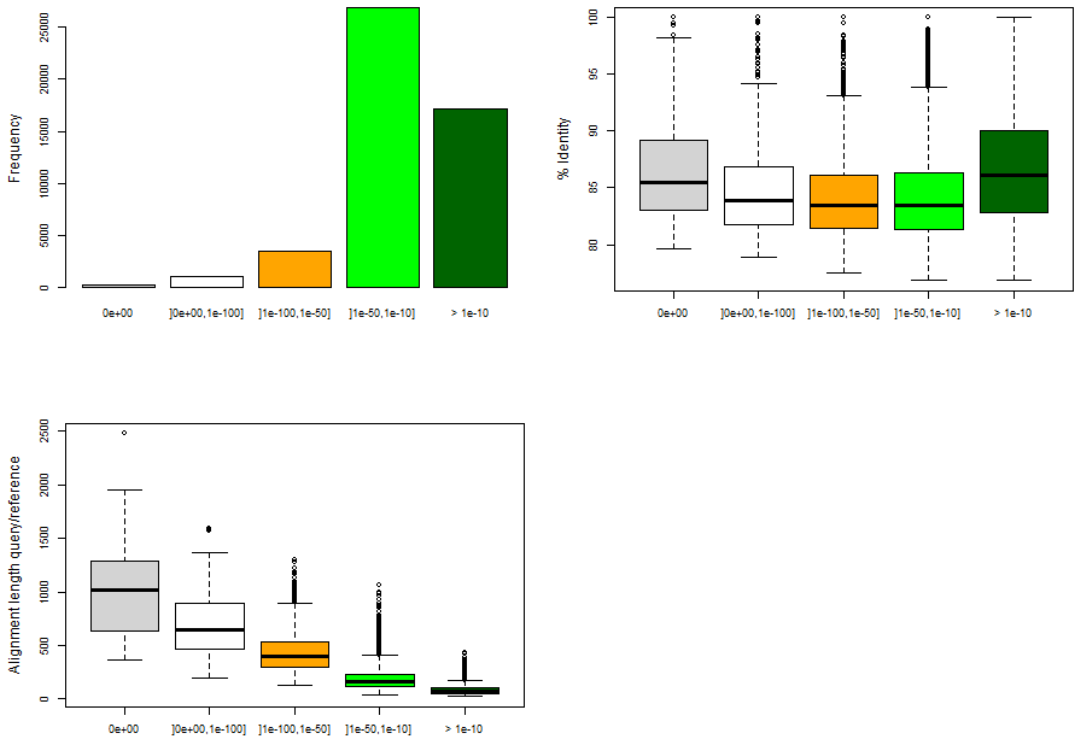


Figure 30 Plots of frequency, % identity and alignment length per class of e-value of *F. cantabrica* contigs with all nucleotide hits.

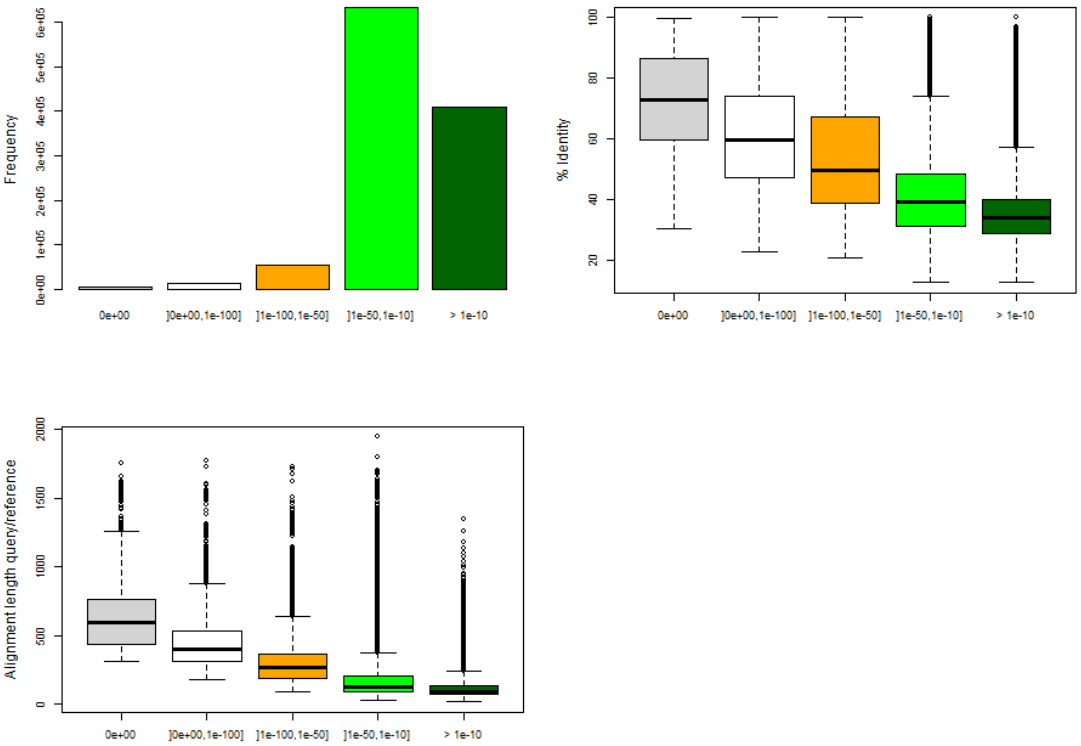


Figure 31 Plots of frequency, % identity and alignment length per class of e-value of *F. cantabrica* contigs with all protein hits.

These same plots but only concerning the sequences with hits (best) retrieved against a specific organism/group (*Aplysia californica*; *Xenopus tropicalis*; *Bufo sp*. mtDNA; nudibranchs mtDNA and *Homo sapiens*) showed similar patterns for the hits with *A. californica* and *Xenopus*, while for the other organisms, as the number of hits were lower, these general trends were not observed. Yet, overall, we can see that the huge majority of the hits are reliable, with a few possible exceptions. This few cases that seem to have both high e-values and low %ID and alignment length were found in the nucleotide hits with *Homo*, and in the protein hits with both *Aplysia* and *Xenopus* (Supplementary Material 3-6, 9-12). Plots regarding best-hits to *A. californica* (full genome coding sequences) and *C. magnifica*, *N. gardineri*, and species of the genus *Bufo,* as well as the remaining analyses are presented in Supplementary Material 3, 7-10).

3.5.1 Battery of orthologs

Using a conservative criterion, with 30% of minimum identity between sequences, only sequences with hits exclusively unique to other Mollusca (unique hits) were considered the orthology search between the two transcriptomes. These represent, respectively 6% (3,342) for *F. cantabrica*, and 6,1% (3,573) for *F. villafranca* contigs.

A reciprocal blast of these two sets of contigs resulted in 728 putative orthologs between the two species, ranging from 113bp up to 5282bp (Figure 32).

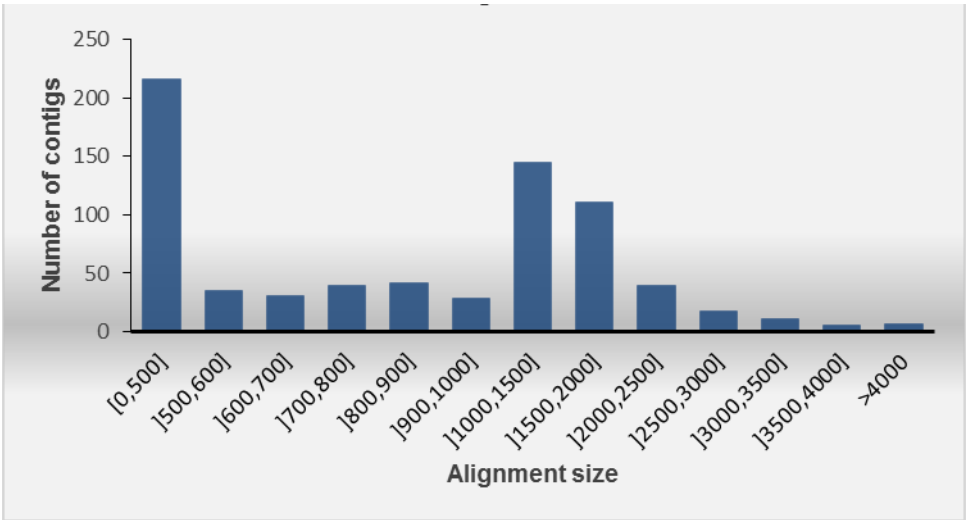


Figure 32 Distribution of the number of putative ortholog transcritps per alignment size categories, after reciprocal blast between clean contigs of *F. cantabrica* and *F. villafranca.*

Chapter 4 Discussion

For “Opisthobranchs”, taxonomy relies mostly on external morphology and anatomy of soft parts (Gosliner et al. 2008). Such information, we know today, is often not sufficient to distinguish between species, therefore other types of studies are necessary (e.g. Eilertsen & Malaquias 2013). In the case of *Felimare*, despite some meticulous descriptions of its life-history traits (e.g. type of habitat; abundance, diet) and of the many colour variants across species, the majority of the species are still described based on anatomical features and colour patterns, with many having high numbers of synonyms – often still used in the literature (Ortea 1996; Sánchez-Tocino 2003). This generates an overall cumbersome taxonomy for this group and many questions thus exist regarding the validity of some of the described species, and the putative existence of species-complexes, which demand the use of molecular data.

About 42 species are currently assigned to the genus *Felimare* (WoRMS, 2015) but there is sequence data (Genbank, mtDNA) only for ten of those, two of which are synonyms (*F. tricolor* and *F. midatlantica*). Equally important is the lack of genomic resources for Chromodorids in general, as well as for other Gastropod molluscs. With this in mind, additionally to sequencing one mtDNA fragment for several populations of our group of interest, I also sequenced and assembled the transcriptomes of two of its species, *F. cantabrica* and *F. villafranca*. Transcriptome sequencing is a fast and efficient mean of genome reduction (McCormack et al. 2013), and an accessible tool for the acquisition of multi-marker datasets (Barley et al. 2010) applicable either in studies of phylogeny or phylogeography. The data here generated can be later used to generate marker sets for these and other closely related organisms.

This work has resulted in the largest molecular dataset (12 species) used for the study of both phylogeny and phylogeography of *Felimare* so far. I obtained molecular data (COI) newly for two species, *F. cantabrica* and *F. orsinii*, as well as for many individuals from several species and several locations across their range. These molecular analyses considerably further our understanding of the genetic diversity within the genus.

4.1 PHYLOGEOGRAPHY AND PHYLOGENY OF SEVEN EA AND MEDITERRANEAN SPECIES OF *FELIMARE*

4.1.1 Sampling and species abundance

The three most abundant species in my dataset (*F. tricolor*, *F. villafranca* and *F. cantabrica*) are also the most abundant throughout the Atlantic and Mediterranean coasts according to the literature (Ortea 1996; Sánchez-Tocino 2003). *F. tricolor*, here sampled all the way from the Cantabrian Sea, Atlantic Iberian waters, west and central Mediterranean, is described as being abundant throughout all year, both in NW and south Iberian coasts. Similarly, *F. villafranca* is described as frequent, not only in the NE/south of the Iberian coasts, but also in its African Atlantic distribution limits, during almost all year, with possible declines in the winter season. *F. cantabrica*, on the other hand, here collected mainly in the Atlantic Iberian coast, and at a minor extent in the Atlantic African coast, is reported as very frequent in the Iberian coast, however scarce and seasonal throughout Alboran-Oran waters. In Granada (South Eastern - SE - Spain), this species was observed punctually by Sánchez-Tocino (2003) only between January and May.

The remaining species were not so frequently encountered in our samplings. However that does not mean that they cannot be equally abundant: our sampling was quite opportunistic with only a few dives performed when sea conditions were favourable. A proper evaluation of species abundance was not performed and was outside the scope of our work, although our sampling was concordant with the abundance trends reported in the literature.

4.2 INTERSPECIFIC RELATIONSHIPS WITHIN *FELIMARE*

Previous molecular phylogenies of Opisthobranchs have included sequences of *Felimare*, but not more than one or two representatives of each species, with many having never been sequenced (Gosliner & Johnson 1999; Valdés 2006; Pola et al. 2007; Johnson & Gosliner 2012; Sánchez-Tocino et al. 2014). I used a mitochondrial protein-coding gene, cytochrome c oxidase I (COI), as homologous sequences (and primers) were already available for some *Felimare* species (Johnson & Gosliner 2012) and also because this fragment allows resolution at species level in a wide variety of taxa (Hebert et al. 2003), being widely used within nudibranchs (Thollesson 2000; Wollscheid-Lengeling et al. 2001; Zardoya et al. 2004; Turner & Wilson 2007). Its fast rate of evolution makes it also an ideal marker to record recent historical events (Wilson et al. 2009), although caution has to be taken when looking to accurately estimate levels of divergence within and between groups (Mueller 2006).

DNA sequences obtained mostly reflected previous morphological identifications, with only a few exceptions, which I comment on more detail bellow (see section 4.4). With this I mean that both Gene tree (GT) inferences (Bayesian and Maximum likelihood) reflect that the previously recognized species are distinct mtDNA lineages defining reciprocally well-supported monophyletic groups (PP/BS > 95%). The genetic distances between these species vary between 10.8% (*F. villafranca*/*F. orsinii*) and 16.4% (*F. cantabrica/F. fontandraui*) distances, comparable to the ones between sister-species in other dorid nudibranchs, although these can be high variable. For example, Pola et al. (2007), Turner & Wilson (2007) and Wilson et al. (2009) highlight sister species to be well supported and with divergences of 3.3 – 5.77%; when compared with the typical interspecific values of 11.1 ± 5.1% of mean uncorrected p-distances (Hebert et al. 2003). On its turn, Malaquias & Reid (2009) estimate uncorrected p-distances of 10 - 22% between species of genus *Bulla,* while other Heterobranch species show distances of around 7% (Valdés et al. 2011). Bayesian and maximum likelihood phylogenetic analyses support the monophyly of *Felimare*, as recovered by Johnson & Gosliner (2012). At the same time, but contrarily to Johnson & Gosliner (2012), *Hypselodoris* and *Mexichromis* were not recovered as a monophyletic group in the COI tree. With additional analyses (Supplementary Figure 2), I verified that perhaps the greater distance between *Roboastra europea* (outgroup) and the remaining species in the tree, could be disrupting those relationships: when using instead *Chromodoris magnifica* as outgroup*, Mexichromis* and *Hypselodoris* were recovered as a clade, with very good support in the Bayesian COI tree, and consistent in the ML COI tree. Relationships within *Felimare,* though, seem unaffected by the outgroup used.

Despite the overall lack of resolution of the tree at basal nodes within *Felimare*, both gene-tree and species-tree analyses were largely congruent in their main aspects: i) the Western Atlantic *F. kempfi* and the Eastern Pacific *F. porterae* are always sister-taxa (well-supported in the BI tree and in the ST), making the Western Atlantic species paraphyletic; ii) also the Eastern Atlantic and Mediterranean species are not a monophyletic group, given that the Eastern Pacific *F. californiensis* and *F. agassizi* cluster within them; iii) of the interspecific relationships, there seems to be some support for a sister-relationship between the EP *F. californiensis* and the Atlantic *F. picta,* with *F. agassizi* (EP) as their basal group - although not especially supported, this topology is always recovered - and iv) the EAM *F. villafranca*, *F. orsinii* and *F. cantabrica* are clearly a well-supported clade, although relationships between them are unresolved. Despite biogeographic patterns remaining largely unclear, and considering the majority of the diversity of this group is supposed to be in the Eastern Pacific region, these phylogenies argue at least for multiple colonisations and divergence of the Western and Eastern Atlantic regions. Until a comprehensive inclusion of the Eastern and Western Pacific diversity of this group though, it is very difficult to make any statement regarding the direction of the colonization of the EAM area – if through the Western Atlantic or from

the Eastern Pacific, via Suez canal. Possible explanations to the lack of phylogenetic resolution at basal level may include lack of power of the molecular marker used or rapid diversification of the group, as has been proposed for this family by Johnson & Gosliner (2012). Potential causes of this radiation in the family were previously attributed to morphological key innovation characters, as the MDF (Wägele & Klussmann-Kolb 2005), or to ecological speciation by resource partitioning (Krug 2011). The first explanation, nowadays, does not seem completely valid, given that MDF occurr in other Opisthobranchia (Turner & Wilson 2008), and may not occur on some species (e.g. *H. fontandraui*) (Haber et al. 2010). The latter explanation, on the other hand, has been proposed to potentiate diversification at more local scales, for species living in sympatry, and where vicariance does not seems reliable.

For the seven EAM species in this study, living in partial sympatry ecological speciation is not easy to corroborate, therefore I suggest future studies on resource utilization, depth zonation, or larvae development, for example, to better understand the co-occurrence of these sister species. In order to resolve these phylogenetic questions the addition of more species, especially from the Eastern Pacific and Western Atlantic regions will be needed, as well as to survey multiple, unlinked nuclear loci and more mtDNA data for the increment of the phylogenetic resolution (as resolution at deeper branches is almost inexistent). Still, our results corroborate *Felimare* as a monophyletic group, point to multiple (at least two) colonisations of the Eastern Atlantic and Mediterranean area, and also to Miocenic divergences between most species. Some of our results regarding phylogenetic relationships and possible instances of cryptic speciation (for *F. villafranca*) are further corroborated in a work published during the writing of this thesis (Furfaro et al. 2016).

4.3 DIVERGENCE TIME ESTIMATES

Our analysis dates the ancestral divergence within the genus in the Oligocene (30 - 40 MY). Subsequent differentiation between species seems to be old and might have occurred mostly during the Miocene, with a few confidence intervals extending into the Pleistocene. This means that multiple opportunities for dispersal across the Atlantic may have existed throughout this time. These old divergences between species estimated from molecular data and based on fossil calibrations seem to be a recurrent pattern in the literature for several groups of molluscs (Williams & Reid 2004; Duda & Kohn 2005; Williams 2007; Williams & Duda 2008; Malaquias & Reid 2008) and also fish (Hrbek & Meyer 2003; Bellwood et al. 2004).

At least two sister relationships between Eastern Pacific and Atlantic are observed, in agreement with previous phylogenetic studies of other dorid nudibranchs, such as genera

Phyllidiopsis (Valdés 2001), *Rostanga* (Garovoy et al. 2001), and *Platydoristhe* (Dorgan et al. 2002), as well as other marine invertebrates, such as gastropods from the periwinkle family *Littorinidae* (Williams & Reid 2004; Reid et al. 2010)*, Nerita* and *Bulla* (Frey & Vermeij 2008; Malaquias & Reid 2008), the fiddler crab genus *Uca* (Sturmbauer et al. 1996), the shrimp genus *Penaeus* (Lavery et al. 2004), the sea urchin genus *Eucidaris* (Lessios et al. 1999), and at some extent the divergence between the live fossil gastropods of *Pleurobranchus* and its closest sister genus (Goodheart et al. 2015). These patterns are usually explained by vicariant events related to the break-up of the Tethys Sea, including the formation of the Arabian Land Bridge 23 MY (Harzhauser et al. 2007), and the East Pacific Barrier about 10 MY (Lessios et al. 2007). Being much older, divergences between these groups do not seem related to the closing of the Isthmus of Panama that isolated the eastern Pacific and the Atlantic only 2.7–3.4 MY (Leigh et al. 2014). The long branches and old divergence ages may also reflect extinction events across lineages, as it is the case in the gastropod genus *Turbo* (subgenus *Marmarostoma*) (Williams 2007). Extinction could lead to artefacts, this is, transisthmian species may not be the true sister-taxa, and the divergence times may have been overestimated (Malaquias & Reid, 2009). These hypotheses are usually very hard to distinguish, especially in the cases of old divergences. Thus, trans-Atlantic faunal exchange has usually a long and complex history (Harzhauser et al. 2002).

A very recent study on *F. picta* complex suggests that this species had an unusual success crossing main Atlantic biogeographic barriers, including the mid-Atlantic barrier (stretching over 3500 km), and proposed the origin of *F. picta* to be posterior to the closing of the Isthmus of Panama (Almada et al. 2015). The sister-relationship between the eastern Pacific *F. californiensis* and the *F. picta* complex are corroborated here but our age-estimates point to a much older divergence than the final closure of the Isthmus and call for a re-interpretation of the biogeographic history of *F. picta* complex. The youngest divergence within *Felimare*, as here estimated is the one between *F. tricolor* and *F. fontandraui*, however, it has to be taken with special caution, as it was particularly inconsistent throughout the several ST analyses (not shown).

All in all, I must stress the need of taking all these age estimates with caution. Additionally to being based on a single locus, estimates of the timing of speciation events are subject to errors of incomplete sampling and poor phylogenetic resolution.

4.4 INCONGRUENCES BETWEEN MORPHOLOGY AND MOLECULAR DATA

Because colour pattern identification is the most straightforward way to identify these species in the field, and because no internal anatomic characters were used for this study, I relied on descriptions, records of geographic distributions, and detailed discussions upon morphological characters from Ortea et al. (1996) and Sánchez-Tocino (2003). Ortea et al. (1996) was the pioneer in the study of the blue Chromodorid group, i.e. all species with a bluish background and their multiple yellow, orange and/or white stained forms. After him only Sánchez-Tocino (2003) attempted a revision on the systematics of the blue Chromodorids of the Atlantic. However, for some species, descriptions are based on one or two individuals (e.g. *F. malacitana*), with some generalizations based on few individuals or on human misidentifications (Sánchez-Tocino 2003), and many times dealing with the existence of many synonyms for many species, resulting in many confusing characters descriptions.

I noticed Important incongruences between morphologic and mitochondrial data in this study. Two individuals from Cádiz, with no photo record available but identified as *F. fontandraui* were assigned to *F. tricolor* according to COI. Conversely, one specimen identified as *F. tricolor* clustered within *F. fontandraui* (Figure 33 below). Another individual from Morocco (Anza) first identified as *F. bilineata*, was found to share its COI haplotype with individuals of *F. cantabrica*. (Figure 34 below). Additionally, an individual labelled in Genbank as *F. picta elegans,* from NW Spain and which has been used in many studies of Chromodorididae phylogenies, shares its haplotype with *F. villafranca* individuals from NW Iberia. This one is probably a case of wrong labelling/sequencing (Heike *Wägele*, *pers. comm.*).

4.4.1 F. fontandraui and F. tricolor

Regarding individuals from Cádiz, because no pictures of these specimens were taken and colour degrades after a few hours in ethanol, I could not re-access the identification of these individuals and address if they were erroneous ones. On the other side, one photograph was taken (Figure 33) of the individual identified as *F. tricolor* from Barcelona, which was later assigned to *F. fontandraui* based on mtDNA. In the photo it is clear the regular contoured dorsal central line and the lack of several linear stains in the border of the mantle (as expected in *F. fontandraui*). Plus, the anchor-like pattern in the head was not observed in the Figure 33 (*personal communication*). As I could not be certain of non-erroneous identifications, I considered these individuals dubious and, for the purpose of this work, re-assigned them to the species of their mtDNA (marked with an asterisk (*) in Supplementary Table 1). Yet, it would be interesting to keep these in mind for future investigations either on the morphological diagnosis of these two species or/and the possibility of hybridization between them.

Figure 33 Specimen from Barcelona originally identified as *F. tricolor* but harbouring *F. fontandraui* mtDNA (and thus re-assigned to this species). This photo was kindly handed by Enric Madrenas from Cataluña.

4.4.2 F. bilineata and F. cantabrica

One specimen collected in Morocco (Anza), shown in Figure 34, was originally identified as *F. bilineata* based on its coloration pattern. Its COI haplotype is shared with *F. cantabrica*, shared with individuals from the north Atlantic. Contrary to the previous case, morphological identification was far from ambiguous.

Figure 34 *Felimare* individual collected at Anza, Morocco (a), first identified as *F. bilineata* (c), and which exhibits a mtDNA haplotype shared with many *F. cantabrica* (b) individuals.

F. malacitana (Figure 34a) is restricted to the Southern coasts of Spain (Málaga and Almería) (Ortea 1996), and was recently reported from deep waters of the East Atlantic shores (Canary Islands and Senegal) (Sánchez-Tocino 2003).

Adults of *F. malacitana* can be confounded with *F. cantabrica* (Ortea 1996). The presence of tubers or the so-called mantle dermal formations (MDF) in the surface mantle is apparently the one visual characteristic that allows this species to be differentiated from the *F. cantabrica* (Tociño et al. 2003). Still this characteristic isn't so easy to see at naked eye, and for sure it is even harder to see when trying to identify species only by photography (e.g. divers often identify species records only based on visual/digital evidence). Juveniles present two central lines in the dorsum that tend to unify right after the line of the rinophors. Such pattern can also be confounded with juveniles of *F. bilineata*. Again the presence/absence of tubers in *F. malacitana* and *F. bilineata*, respectively, appears the most evident distinctive character. As adults their coloration is reported as easy to differentiate (although exceptions are reported at distinguishing *F. bilineata* from Ghana; *http://www.seaslugforum.net/*) (Figure 36), and they also differ concerning to the lip armour (*F. malacitana* has more dentils than *F. bilineata*). *F. malacitana* is known to be hard to identify even for experts with recurrent contact with these Chromodorid nudibranchs (see *http://seaslugforum.net/showall/hypsmala*).

Regarding the individual analysed in this work (Figure 34a), it was identified as *F. bilineata*, and there isn't enough detail in the picture to access if it could be a *F. malacitana*. As I found it to have mtDNA clearly clustering within *F. cantabrica*, and as *F. malacitana* and *F. cantabrica* are also sometimes reported to be difficult to distinguish, I have to hypothesize the potential occurrence of hybridization between *F. cantabrica* and *F. bilineata*, and that the geographically-restricted, phenotypically-variable *F. malacitana* may actually be not a valid species but instead represent hybrid individuals between these two co-occurring species. It would also be interesting to further investigate this issue.

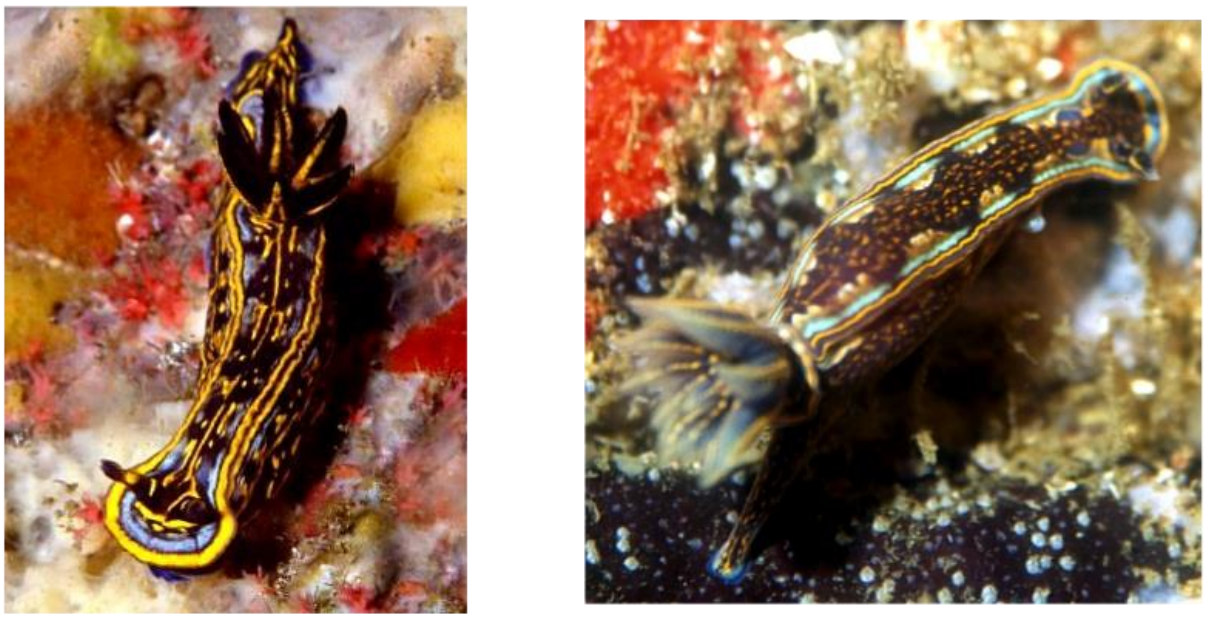


Figure 35 Images taken from http://seaslugforum.net/showall/hypsmala. Specimens in the image have been suggested to be *F. malacitana* by experts on nudibranch's taxonomy (Bill Rudman). On the left: a specimen from El Hierro Island (Canary). On the right: a specimen from "Grand Turiba", Dakar, Senegal.

4.5 INTRA-SPECIFIC VARIABILITY AND GENETIC STRUCTURE

4.5.1 Genetic differentiation

The genetic structure of the seven EAM *Felimare* species as inferred from the mtDNA COI marker has shown to be useful for tracing both intra and inter-specific genealogies of many heterobranch species (Ellingson & Krug 2011; Wilson et al. 2009; Johnson & Gosliner 2012). Bearing in mind the sampling limitations of this study, my results show already some clear patterns of genetic structure across the different species, as well as unveil several cases of potential cryptic species. The seven EAM species exhibit distinct phylogeographic patterns, as well as variable degrees of differentiation within each species. For five of the species, the substantial genetic divergence between clades and reciprocally allopatric distributions point to the existence of species complexes. One of those, *F. picta*, has been subject of further exploration as this thesis was being written (Almada et al. 2015).

4.5.2 Species without signs of population structure

4.5.2.1 *F. cantabrica and F. tricolor*

These two species did not show any signs of population structure, with closely related haplotypes distributed (and shared) across their sampled distributions. In the case of *F. cantabrica*, its distribution seems to encompass only the Cantabrian and Eastern Atlantic coast, so a strong genetic structure would be less expectable. The absence of phylogeographic structure in *F. tricolor*, distributed from the Cantabrian sea to the Eastern Mediterranean is remarkable (Figure 14). These patterns may either indicate the existence of overall gene-flow across this area for these species or an overall recent colonization of their distribution area. To test these hypotheses, the collection of more data and the use of comprehensive multilocus inference frameworks will be needed. Similar patterns of lack of genetic differentiation in this region are observed in other limpet-like species (e.g. *Patella rustica*, *Pollicipes pollicipes*, *Nassarius reticulatus*) and isopods of the genus *Stenosoma* (Couceiro et al. 2007; Campo et al. 2009; Ribeiro et al. 2010; Xavier et al. 2011), as well as for the sea urchin *Arbacia lixula* (Wangensteen et al. 2012).

Their pelagic larvae confer these organisms a high dispersal potential which may offset their small-scale mobility when adults. Both species, *F. cantabrica* and *F. tricolor*, need approximately seven to eight days (six of which are expended as veliger stage) after their period of fertilization to settle in the substrate (Coelho & Calado 2010). These numbers are similar (9 to 15 days) to those of other species such as limpets of the genus *Patella,* that share haplotypes between the Atlantic shores of Africa and Iberian Peninsula and Mediterranean Iberian shores (Sá-Pinto 2008). In the case of *P. rustica*, Ribeiro et al. (2010) hypothesized that the prevailing current patterns during the reproductive season could be responsible for the observed genetic homogeneity. Throughout the year, the mean flow on the surface of the Portuguese coast is southward, but seasonal winds can change this pattern (Martins et al. 2002). The general absence of genetic differentiation throughout the distribution of these two species may mean that regional hydrographic patterns may be promoting their high conectivity, both northwards and southwards, as observed for other species by Xavier et al. (2011).

4.5.3 Species with signs of population structure

4.5.3.1 *F. villafranca*

F. villafranca exhibited a marked genetic structure with clear differentiation between the East Atlantic (Iberian coast, west of Gibraltar), West Mediterranean (Barcelona, Girona, Menorca and France), Central Mediterranean (Italy) and Morocco. In particular, the samples from Italy and Morocco are highly differentiated from the rest (2.6% and 7.1%, respectively). Differentiation between Iberian and Moroccan populations has sometimes been reported in the literature, for other species

(Simon-Bouhet et al. 2006; Xavier et al. 2011; Ouagajjou & Presa 2015), but many other studies on crustaceans and gastropods (e.g. barnacle, squat lobster, limpets *Cymbula nigra*, *P. rustica*) fail to find it (Pannacciulli et al. 1997; Naciri 1999; Sá-Pinto 2008; Campo et al. 2009; Cabezas et al. 2011; Espinosa et al. 2011). With exception of limpets, either barnacles or squat lobsters need at least one to several months to settle to the substract, and therefore are expected to disperse quite well. As for other organisms (e.g. Ouagajjou & Presa 2015), this latitudinal divergence in the Atlantic is likely due to a poor offshore dispersal between the continental coasts of Europe and Africa. Even within the northeastern Atlantic coast some differentiation (1.2% *p-distance*) is seen across localities, with the southern ones (Ria Formosa, Cádiz, Gibraltar) not sharing haplotypes with the more northern localities (Ferrol, Vigo and Sesimbra). Although I still miss sampling from a great part of its known distribution, this species is a the only direct developer in the group and therefore expected to be subdivided. Juveniles from *F. villafranca* hatch already in a benthic dorid-shape (Coelho & Calado 2010), and in fact it is possible that the observed patterns can be mostly explained by isolation-by-distance. Additionally, given what is known about Mediterranean and north African coastal biogeographic areas and barriers to gene-flow, it is likely that further sampling across its distribution range unveils yet more lineages.

4.5.3.2 ***F. fontandraui***

At least three lineages can be seen within this species: Atlantic + Chafarinas, and at least two others in the Mediterranean. Despite this species having planktonic larvae (which disperse more than *F. villafranca*), little haplotype sharing was observed between localities (only in one case). This may be a real pattern or result of the limited sampling. In any case, the geographic distribution of the different clades seems to coincide with the known marine barriers of Almeria-Oran Front (samples west to it, including Chafarinas belong to the same clade) and the Balearic Front and the border between the Ligurean and Thyrrenean Seas (Figures 1 and 4) (which could explain the differentiation between the samples from Girona and South-West France from the diverging Italian one, from Ischia). Yet, the fact that one haplotype is also shared between individuals from Girona and Italy (Giglio island) indicates that further effort is needed to clarify the biogeographic structure of this species in this area. The region that separates the Ligurian Sea and the North Tyrrhenian Sea, between the coast of Tuscany, and the islands of Elba and Corsica, has been hypothesized as a biogeographic barrier (Bianchi 2004), being characterized by divergent currents on both sides (Millot 1999). However very few studies have tested such barrier (Lenfant & Planes 1996; Rastorgueff et al. 2014) and for none of the organisms it was shown to be effective, including one study with gastropods (Villamor et al. 2014). The isolation of the north-western Mediterranean lineages might been maintained by a quasi-circular cyclonic front located in the Ligurian Sea active all year round, and marked seasonal variability

of the Tyrrhenian Current. When active (autumn and winter), the Tyrrhenian current flows close to the Ligurian coast (Astraldi et al. 1995) and could explain some dispersals between the two regions. Sánchez-Tocino (2003) states the winter season and possibly the spring as the reproductive period of *F. fontandraui*. This could facilitate larvae migration during this period between the two basins and explain the currently observed patterns. On the other side, the differences in the thermohaline conditions of the two basins in winter (Astraldi et al. 1995) may act in promoting isolation.

4.5.3.3 ***F. orsinii***

Very few samples of this species, with a mostly Mediterranean distribution, were collected, but, again, the Atlantic-Mediterranean split is very clear, with an Atlantic and a Mediterranean lineage around 7.5% divergent.

4.5.3.4 ***F. bilineata***

F. bilineata has been already suggested to be a species complex, with possible cryptic lineages in Atlantic Iberia, Canary islands and Ghana (Ortea 1996). For this study samples could only be found in the southern Iberian Atlantic coast (Ria Formosa and Cádiz) and Madeira. Remarkably, even just within those two close localities, *F. bilineata* harbours at least two highly differentiated lineages with a minimum p-distance of 4.3%, without any clear geographic structure. Further differentiation was found within one lineage, by a minimum of seven mutational steps (approximately 1% uncorrected p-distance). Such pattern of divergence is not common in gastropods in such a small area, particularly for species with a planktonic larval dispersal. It will be important to further sample across the distribution range of this species in order clarify the distribution and origin of these different lineages (ecological or allopatric differentiation in secondary contact?)

4.5.3.5 ***F. picta***

F. picta showed a pattern of deep genetic differentiation between southern Atlantic populations (Cabo Verde and S. Tomé) and more northern Atlantic and Mediterranean populations. These two where the highest differentiated intraspecific lineages in this stuidy (more than 9% divergent). This strongly suggests that *F. picta* is a species complex, in agreement with Almada et al. (2015), and with Ortea et al. (1996). Studies reporting northern and tropical eastern Atlantic genetic breaks are common in the literature (Muss et al. 2001; Quinteiro et al. 2007; Domingues et al. 2008; Pellerito et al. 2009), and this patterns is not surprising. Yet, in this case, no differentiation was detected between Iberian Atlantic (west of Almeria-Oran) and Mediterranean populations, despite the multiple eastern Mediterranean samples included in our study. In this study there are only a few samples from the northeastern Atlantic region, and no samples from eastern Mediterranean locations or other disjoint *F. picta* populations (Azores, Brasil, the Caribean) but Almada et al. (2015), showed that little

differentiation exists between populations from Iberian Peninsula (south Spain, Cádiz), Madeira, Canary islands, Azores and even Mexico, which would belong to our Atlantic clade. In our samples, haplotypes from both clades are very closely related, or shared (e.g. individuals from Cádiz and Italy shared their haplotypes), likely indicating gene flow across large distances, as also argued by Almada et al. (2015). Based also on their mtDNA phylogeny, which also includes *F. picta* representatives from Brasil and another closely related species from Bermuda (*F. zebra*), these authors argue two eastwards colonizations to explain its amphi-Atlantic distribution, unique among chromodorids. Coelho & Calado (2010) reported that *F. picta* shows the largest egg size and planktotrophic larvae length at hatching reported among nudibranch molluscs, which could help to explain this remarkable pattern. Still, Almada et al. (2015) argue for an origin of this species complex posterior to the closure of the Isthmus of Panama (given its sister-relationship with the eastern Pacific *F. californiensis*), which clearly is contradicted by the genetic distances and divergence time estimates obtained here, and that suggest that the origin of this species complex seems to largely predate this closure.

4.6 ATLANTIC AND MEDITERRANEAN BARRIERS TO GENE-FLOW

As expected, many of the areas/features known to act as barriers to gene flow in other marine species seemed to affect the distribution of *Felimare*. The two biographic discontinuities known to cause differentiation between Atlantic and Mediterranean populations are the Strait of Gibraltar (SG) and Almeria-Oran Front (AOF). These have been widely described as major dispersal barriers for many marine organisms (e.g. Patarnello et al. 2007; Galarza et al. 2009), and are here hypothesized as causes for the observed patterns in several species. Regardless the fact I often do not have a representative sampling to clearly distinguish between the effects of the Strait of Gibraltar or of the AOF it seems that for species as *F. fontandraui* and *F. villafranca* (where samples collected east of the SG - Chafarinas, Gibraltar - cluster within remaining Iberian Atlantic ones) AOF may be the prevailing barrier acting on Atlantic-Mediterranean differentiation. A similar pattern is suggested in Sá-Pinto et al. (2012). *F. fontandraui* and *F. villafranca* populations from the Balearic area (Girona, Southwest France, Balearic islands) also show some slight differentiation, likely reflecting as well the action of the Balearic front. This area is under the farthest influence from the incoming Atlantic water and constitute the most differentiated Mediterranean waters (Palero et al. 2008). Here, the occurrence of the Northern and Balearic Currents form a mesoscale cyclonic gyre (Figure 1), which may favour larval retention.

The role of further eastern proposed barriers in the Mediterranean either was not observed in *F. tricolor and F. picta* and could not be tested in the remaining species. It would be interesting to extend the sampling of the remaining species across their distribution area in the eastern Mediterranean, as further population structure is likely expected. Additionally, some studies (Jaziri & Benazzou 2002) suggest further structure within north African coasts, which would be interesting to explore in the several species with northern African distributions.

4.7 TRANSCRIPTOME ASSEMBLY

4.7.1 Sequence analysis and assembly

From the many short read assemblers available I selected Trinity, a method for *de novo* reconstruction of full-length transcripts in a sample from RNA-seq reads, and broad range of expression levels, high reference coverage, low levels of redundancy, on both small and large datasets (Yang & Smith 2013). Another advantage of using Trinity is the efficiency of its default settings, already optimized for a single k-mer size (25 k-mer), and with efficient thresholds for minimum length of transcripts. Other popular software packages, instead, use a range of k-mer sizes for short-read RNA-seq assembly that generally require additional analyses (e.g. k-mer sizes manipulation, average insert length manipulation). Furthermore Trinity explicitly outputs assembled sequences in a hierarchical mode (e.g. component or genes, subcomponent or homologs and transcripts or isoforms), which is an advantage over other assemblers. In the particular case of transcriptomic data, Trinity has shown to be superior, with sensitivity comparable to methods relying on genome alignments, even in the complete absence of a reference genome (Riesgo et al. 2012; Grabherr et al. 2013; Yang & Smith 2013).

The number of reads collected in this study (close to 40 35 million) is equivalent to other transcriptome analyses of invertebrates molluscs (e.g. cephalopod *Octopus vulgaris*, the chitone *Chiton olivaceus*, and pond snail *Radix balthica*) (Feldmeyer et al. 2011; Riesgo et al. 2012), and significantly larger than some other previous studies (e.g. *Nerita melanotragus*) (Amin et al. 2014). One way of evaluating the transcriptome completeness was by using CEGMA, which searches for 248 core proteins that generally lack paralogs in the eukaryotes (Tatusov et al. 2003; Parra et al. 2007). CEGMA reports their recovery as “complete” or “partial”, where “complete” means that 70% or more of the amino acid sequence was recovered. The values here obtained (50% to 60%) are lower in comparison with other transcriptomic studies on molluscs (e.g. Hu et al. 2011; Cornman et al. 2014), and may reveal incompleteness of the transcriptome. Such scenario might be expected given that only foot tissue was represented in the sequencing libraries. The number of contigs/transcripts

obtained here (80K) fall in a range comparable to studies of molluscs and other invertebrates. For instances, for the two pond snail gastropods analysed by Feldmeyer et al. (2011), contigs varied between 52K up to 57K, while Riesgo et al. (2012), obtained totals from 71K – 207K contigs in a wide set of invertebrates.

The N50 is defined as the shortest sequence length such that half of the total sequence output length is included in sequences. Usually, it is around 1 Kb for genome assemblies, as the average size of an exon for animals (Bolshoy et al. 2011), but at the same time it varies proportionally to both the length and coverage (O'Neil & Emrich 2013). This wasn't primarily aimed to be a reference measure in transcriptomes, yet some studies refer to it as a complement comparative value among similar studies (e.g. Riesgo et al. 2012; Amin et al. 2014). The lowest N50 recovered among my datasets belonged to *F. cantabrica*, with 899 bp, with an average length of 625, while for *F. villafranca* the values are slightly higher, with 950 bp of N50 and average of 651 bp. These values were equivalent to many other transcriptome sequencing studies (e.g. 900 bp for the chickpea; 893 bp for *Oncopeltus*, or 837 for the sea cucumber *Apostichopus japonicus*) (Garg et al. 2011; Ewen-Campen et al. 2011; Zhou et al. 2014) and higher than others (e.g. 258 for *N. melanotragus*, 599 for *O. vulgaris*) (Riesgo et al. 2012; Amin et al. 2014). It is not straightforward to compare transcriptome assemblies and their qualities based on these metrics, though. (O'Neil and Emrich, 2013). Finally, by picking each component more representative isoform (most "abundant"/the one with the highest coverage) in both cases, I attempted to diminish redundancy, obtaining an appropriate dataset for "phylogenomics", a strategy used previously by other authors (e.g. Yang & Smith 2013).

4.8 FUNCTIONAL ANNOTATION

The blast against nr (18% and 19%) and annotation (8%-5%) success for both assemblies were low but similar to other non-models organisms studies (Riesgo et al. 2012). One possible explanation for the limited annotation success is the fact that few reference genome sequences exist for mollusc species (Sadamoto et al. 2012). Amin et al. (2014) highlights that the success of transcriptome representation on gastropod species will require more reference genome sequences, as well as an increase in mollusc protein sequences in public databases. Very short contigs (200–500 bp) rarely returned blast hits or annotations, with more than 90% of these sequences excluded from the final assignments to GO terms. Among the ~10K hits against the nr database, the majority had its best-hit with *Aplysia californica,* the only gastropod genome sequenced within Heterobranchs. The top-hits species distribution showed that the majority of the hits were also to other mollusc species. The fact

that many hits were also present (and especially in *F. villafranca*) for *Xenopus laevis*, did not call our attention in the first moment. This is a well-annotated reference genome, and other distantly related organisms were also present in the hit distribution at similar percentages.

Comparisons annotations among studies are not easy, given the focus on specific biological questions. No functional category of gene function was lacking in any of the *Felimare* transcriptomes, and both transcriptomes shared most of the terms. Despite the limited number of annotated contigs, the assembly captured a broad variety of GO categories for all functional domains. As expected, annotations regarding level 2 were more similar across transcriptomes, with all the top-20 terms being shared by the two species. My results were in fact similar to those of Zhou et al. (2014), for the sea cucumber *A. japonicas*, Sadamoto et al. (2012), for the pond snail *L. stagnalis*, and Amin et al. (2014), for the gastropod *N. melanotragus,* either respect to the top GO terms found, as well as their percentages. Here I provide the first global view of two Chromodorids transcriptomes that may serve as a basis for future studies. Overall both my datasets provide a relatively comprehensive view of each transcriptome, and the relative proportion of different GO terms bears a resemblance to those of other well-annotated transcriptomes.

4.9 DETECTION OF CONTAMINATION

There are numerous challenges for effective NGS data quality control (QC) (Schmieder & Edwards 2011; Zhou et al. 2013), although this is extremely important for downstream analysis. Many NGS studies perform a basic cleaning/trimming of the reads (Parchman et al. 2010; Feldmeyer et al. 2011; Zhou et al. 2014; Senatore et al. 2015). Others perform additional validations by using a reference set of genes/genomes of closely related species (Amin et al. 2014). The quality of data may be affected by several factors regardless of the NGS platform. Contamination by external sources, especially cross-contamination during library preparation, is generally neglected, but can be an important source of error (Zhang et al. 2012; Tosar et al. 2014). A recent tool, QC-Chain (Zhou et al. 2013) claims the ability to both filter low quality reads as well as to detect contaminating reads without prior information on the contamination source, by mapping reads either to rDNA databases or NCBI-NT database. However that was not used in this study.

In this work I only noticed the occurrence of sequences of amphibian origin at a later stage of my analyses (mitochondria reconstruction). In fact, this contamination was detected by chance, as it was not in the initial plans to reconstruct the mitochondrial genomes. At that point of the work, remaking every analysis from the very first step (including obtaining new reads) was not possible. Likewise, I believed I could identify a reasonable number of homologous contigs between the two

species. The strategy adopted was to perform a series of blast analyses to evaluate the extension of the contamination and filter the contaminating contigs. I used available genome data on molluscs and amphibians, as well as other invertebrate and vertebrates (based on previous results from the annotation step), as I felt reasonable for the "exercise" of this thesis. I used a conservative criterion and filtered both assembled transcriptomes to contigs with hits exclusive to available Mollusc coding-genomes, using only these data in downstream analyses.

The real extension of the contamination on both transcriptomes is difficult to determine given that I measured it using the assembled transcripts and not the initial reads. Both amphibian and sea slug's reads may have been assembled together, creating some chimeric (fake) contigs that then would likely retrieve no blast-hit or annotation. Given the fact that most of the assembly statistics were similar to other molluscs' *de novo* transcriptome assemblies, I had no clues of the existence of an external contamination, especially not from amphibian origin. Looking backwards, the top-hits species graph was in fact the first sign of it, as the Western clawed frog (*Xenopus* sp) is represented as a high ranking species in both transcriptomes. At the same time, it represented less than 4% (3,9%) of the total annotated contigs in *F. villafranca*, and 1% in *F. cantabrica*.

The number of unique hits to vertebrates was clearly inferior to the unique hits with non-vertebrates. Taking into account these results, and further analyses given in the supplementary material figures 3-13, I decided to consider only unique hits to Molluscs as a conservative approach to eliminate potential "contaminant" sequences in downstream analyses of marker development.

To fully evaluate the extension of contamination in both datasets and its interference in the assembly, I would need to remove the reads prior to the assembly (by mapping to available amphibian sequences) and then compare the two assemblies for example. Besides using a new version of the fastQC from Zhou et al. (2013) to clean the reads I should also map my reads against a full mitochondrial genome close enough to the species of study. Selecting mtDNA reads prior to assembly seems to be advisable (Ekblom et al. 2014), and it may be a way of prevent/detect mtDNA contamination (as mtDNA reads are more abundant), allowing for more refined and accurate homology testing.

Regardless the contamination affecting this study, identification and exclusion of non molluscs contigs served as test for if we could still get enough information to proceed with orthologous discovery. Meanwhile both of the species transcriptomes are being re-sequenced for further work.

4.10 PRELIMINARY ORTHOLOG SEARCH

Regardless contamination, and after filtering both assemblies, I could still obtain a total of 728 potential ortholog fragments. In a recent study, Schultheis et al. (2014) used a similar approach using two RNAseq samples and a distant reference genome, and obtained around double of the potential orthologs than I did. From those, they ultimately ended with 24 with > 85% and <98% identity suitable for primer design, plus 30 mapping to intron regions of a distant reference genome, from which primers were designed for 14, which revealed to be useful at a population scale. Plus, they argue that >1500 contigs could still be screened for useful molecular markers information. Overall, the putative orthologs I obtained could still be a useful starting point for the development of nuclear markers using a similar same approach, especially considering the lack of phylogenetic markers at several levels within the Heterobranchs/Opistobranchs.

4.11 MITOGENOME ASSEMBLIES

4.11.1 Genome composition and arrangement

Full mitochondrial genomes offer interesting information on gene order and number and are themselves additional sources of phylogenetic information (Simon & Hadrys 2013; Perseke et al. 2013). Although a great number of mitochondrial genomes is available in NCBI for Heterobranchia, only one is available for the Chromodorididae family (*Chromodoris magnifica)*. In this study I tried to obtain the first two mitochondrial genomes from the genus *Felimare,* and consequently two additional Chromodoridid representatives.

The mitochondrial genomes of gastropods are generally known to be very compact, i.e. with absence or mostly short intergenic regions (Boore 1999). Its length in this organisms usually varies from 13 to 17 kb (e.g. 13,670 bp in *Biomphalaria glabrata*, 17,575 bp in *Diodora aspera*), with few exceptions (ribbed limpet *Lottia digitalis*, with a length of 26,835 bp) (Simison et al. 2006). Even though I was not able to recover the complete mitochondrial genome sequences (approximately 14 kb)*,* they seem to be of similar size to other Opistobranch mtDNA genomes (Medina et al. 2011).

The mitochondrial genomes were not completely assembled. With regards to the first approach (manual contig alignment), around 30% of the mtDNA contigs initially recovered by BLAST revealed to be *Bufo/Epidalea* mitochondrial sequences in both genomes. Regarding the contigs homologous to Opistobranchs mtDNA, I obtained 13,980 bp and 14,208 bp, respectively, for *F. cantabrica* and *F. villafranca*, very similar to the *C. magnifica* mitogenome. These mitochondrial genomes had the same arrangement of genes as other nudibranchs (Grande et al. 2002; Medina et al. 2011). More

than one coding and/or noncoding regions of the mitochondrial genomes were recovered within the same contig, generally with the same gene rearrangement of the reference.

The total bp recovered by MITObim was higher in general, yet involving a great amount of ambiguous nucleotides. Given that, and regardless the approximately 14 kb assembled for both species, as well as missing regions within genes, I highlight that the ambiguities in the assembly resulting from MITObim are too high for me to be confident on its recovered mtDNA sequence. Thus, in the following sections I will mostly take in consideration the results obtained from the mtDNA assembly from the contigs.

MITObim has been used to achieve complete assemblies of mitogenome in several distinct species (e.g. insects, fishes, birds, gastropods) (Hahn et al. 2013; Gan et al. 2014; Gan et al. 2015; Guan & Xu 2015; Krzeminska et al. 2015), with or without a reference of a closely related organism, and also to infer rearrangements of the gene order. Unfortunately that was not the case in this study, where this assembler performed very poorly. This assembler has been mostly used for genomic rather than transcriptomic data (Doyle et al. 2014), and is frequently recalled as requiring low-input DNA quantity and of simple usage, being particularly advantageous for new users (Gan et al. 2014). Nonetheless, it was not straightfoward the understanding of the different settings of the progam, and I have verified that for inputs above 10 million reads, the program would generate greater amounts of ambiguities and could not perform properly. I suspect that one of the problems with this approach may have been its default number of allowed mismatches to the reference sequence (15% of the average read length), which I used, and may have been too low for the distant reference mtDNA genome (as can be inferred from the % differences obtained from the assembled contigs). Another hypothesis for the bad perfotmance of the programme could be the existence of mtDNA reads of *Bufo/Epidalea* in the readpool, complicating the MITObim assembly.

4.11.2 Coding regions

As inferred from these assemblies, the gene content for the two *Felimare* mitochondrial genomes is the same typically present in other known nudibranch gastropods (Medina et al. 2011; Sevigny et al. 2015). Overall, for most of the encountered genes, I was able to recover most of the start-codons and stop-codons The presence of incomplete termination codons is fairly common in metazoan mitogenomes (Ojala et al. 1981), including in the Molluscs mitochondrial genomes sequenced to date (Grande et al. 2002; Medina et al. 2011). In such circumstances termination codons are generated via polyadenylation to TAA (Ojala et al. 1981). Indeed, that seemed to be also the case for some of our genes.

4.11.3 ND5-like case

The gene content for the two *Felimare* mitochondrial genomes seems to be the same as in other nudibranch gastropods (Medina et al. 2011; Sevigny et al. 2015). For most of the genes I was able to recover most of the start- and stop-codons The presence of incomplete termination codons is fairly common in metazoan mitogenomes (Ojala et al. 1981), including mollusks (Grande et al. 2002; Medina et al. 2011). In such circumstances, termination codons are generated via polyadenylation to TAA (Ojala et al. 1981). Indeed, that seemed to be also the case for some of our genes.

I found a region homologous to the "ND5-like" of *C. magnifica*. This region is annotated as non-coding in *C. magnifica*, as no valid start or stop codons were found (Medina et al. 2011) in our case, the beginning of the fragment, was TTA/UUA, as in *C. magnifica*. This triplet codes for Leucin, which is also found to be a starting-codon in invertebrates, although only in its TTG/UUG forms, as in *Roboastra europaea* (Grande et al. 2002; Serb & Lydeard 2003). TTA/UUA has not described as a start codon for invertebrates so far. The stop codon, was also homologous to *C. magnifica*. Despite that a stop codon has not been annotated for this gene, a codon TAG/UAG was found in the dataset, exactly in the same position as in *R. europaea* genome, followed by gene (ND1) possibly starting immediately 2 bp after (with ATT/AUU). This is as annotated for *R. europaea,* as also homologous to what we find in the *Felimare* transcriptomes. Further comparisons with other invertebrate mtDNA genomes for this region seem necessary to decipher whether this is a real coding fragment, or an an overlapping gene.

4.11.4 RNAs

MtDNA intergenic noncoding regions are generally very variable and include indels (Grande et al. 2008). This is also observed in these two *Felimare* mitogenomes, with many of the tRNAs (and rRNAs) differing a few bp in length respect to the *C. magnifica* mitogenome.

The identification of tRNAs in molluscs or in other invertebrates (J. Terrett et al. 1996; Hatzoglou et al. 1995; Yamazaki & Ueshima 1997; Sasuga et al. 1999) can be difficult because the standard cloverleaf secondary structure may not be present (T or D arms can be lacking) (Kurabayashi & Ueshima 2000; Grande et al. 2002). Still, pulmonate tRNAs can be functional in these conditions (Sasuga et al. 1999; Boore 1999). The observed lack of a paired D arm is unusual in metazoan organisms (Kurabayashi & Ueshima 2000), however they are found recurrently in heterobranch organisms (Grande et al. 2002; Kurabayashi & Ueshima 2000; Groenenberg et al. 2012).

Success in recovering the tRNA regions was also higher when using the assembled contigs instead MiITObim. Although not all tRNAs were recovered, several were identified, forming secondary structures similar to the ones recovered for *R. europaea*. A less stringent mapping approach with

MITObim, or a *de novo* assembly from a partial mtDNA assembly could be used to try to improve the assembly and to potentially recover additional tRNA and other mitochondrial regions.

Chapter 5 Conclusions & Future Prospects

5.1 CONCLUSIONS

This work has provided relevant information on the patterns of genetic diversity and divergence of seven species of the genus of *Felimare*, throughout the eastern Atlantic Ocean and Mediterranean sea. Furthermore, the COI phylogeny resulted in a more comprehensive understanding of the relationships within the genus, despite the lack of representatives from the Western side of the Atlantic Ocean.

Overall, five species showed highly divergent and often geographically localized lineages, representing potential species complexes. The intraspecific structure patterns appear to be strongly influenced by coastal topography and water circulation, with areas of restricted gene-flow matching the ones also acting on other organisms (e.g Gibraltar, Almeria-Oran front, and possibly the Balearic front), and with some cases of high diversity across relatively short distances. As expected due to its direct development, *F. villafranca* showed a strong population structure, with different lineages with variable distribution ranges (one large western lineage and possibly four smaller eastern ones). Overall, our results suggest that the taxonomic status of some of the (for now intra-specific) lineages need further revision.

The phylogenetic resolution of the deep branches was very low. Still, our results corroborate *Felimare* as a monophyletic group, point to at least two colonisations of the Eastern Atlantic and Mediterranean area, and to Miocenic divergences between most species. Importantly, some instances of disagreement between morphological identifications and mitochondrial diversity were highlighted, which deserve further investigation (especially for *F. bilineata* and *F. cantabrica*)

The two new transcriptomes for *F. cantabrica* and *F. villafranca* represent important genomic resources that will facilitate future research in this genus. Additionally, I was able recover a large portion of the mitochondrial genome of these two species, even in the absence of a close reference.

New nuclear markers can be further developed from the information here provided (the preliminary ortholog battery). These could could potentially encompass population level information, and be of great utility for further phylogenetic/phylogeographic studies in this group solving several of the questions exposed above.

5.2 FUTURE WORK

This work contributed to the knowledge of the phylogeny of this group and the phylogeographic patterns of some species and revealed several instances of possible cryptic species – further work is needed, though, to clarify their distributions, evolutionary history and taxonomic status.

The phylogeny of the group reveals interesting aspects regarding patterns and times of divergence, yet it is still largely unresolved. The use of extra markers (mostly nuclear) will be essential to improve the estimate of phylogeny of the group.

The developed nuclear markers will also be useful in the study of Chromodoridae phylogeny in general, as presently they are largely unresolved and mostly based on mtDNA markers.

Given that some portions in the currently assembled mtDNA genomes are still missing, a better assembly of these must also be performed. Relaxing mapping parameters or providing MITObim with an initiating mtDNA fragment (COI or from previous assembly) may reveal to be a better strategy.

In this thesis, and after the detection of the contamination of the two transcriptomes, I mostly proceeded with the filtering of the contigs and "preliminary" ortholog search as a proof of concept that information useful for marker development could still be collected. To avoid uncertainties about misassembles and increase data size, the sequencing of these two transcriptomes is being repeated, as well as of another representative of Chromodoridae (external to *Felimare*). The new data will be used to survey again for orthologs between these species, as well as mapping the obtained contigs to *A. californica* genome assembly and survey for adequate fragments for marker development (identification of exonic and intronic regions and suitable flanking sequences for primer design) at several levels (phylogenetic and population level) in an approach similar to Schultheis et al. (2014). Given the current lack of nuclear markers for phylogenetic inference in these organisms, this data will be a valuable contribution for further evolutionary studies in these organisms.

Bibliography

Aguado, F. & Marin, A., 2007. Warning coloration associated with nematocyst-based defences in aeolidiodean nudibranchs. *Journal of Molluscan Studies*, 73, pp.23–28.

van Aken, H.M., 2002. Surface currents in the Bay of Biscay as observed with drifters between 1995 and 1999. *Deep Sea Research Part I: Oceanographic Research Papers*, 49(6), pp.1071–1086.

Almada, F., Levy, A. & Robalo, J.I., 2015. Not so sluggish : the success of the Felimare picta complex (Gastropoda , Nudibranchia) crossing Atlantic biogeographic. *PeerJ*, (4), p.e1561.

Almada, V.C. et al., 2001. Patterns of diversity of the north-eastern Atlantic blenniid fish fauna (Pisces: Blenniidae). *Global Ecology & Biogeography*, 10, pp.411– 422.

Amin, S. et al., 2014. Assembly and annotation of a non-model gastropod (Nerita melanotragus) transcriptome: a comparison of De novo assemblers. *BMC Research Notes*, 7(1), p.488.

Amorim, M.C.P. et al., 2010. Stereotypy and variation of the mating call in the Lusitanian toadfish, Halobatrachus didactylus. *Behavioral Ecology and Sociobiology*, 65(4), pp.707–716.

Andrews, S., 2010. FastQC: A quality control tool for high throughput sequence data. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*.

Andrié, C. et al., 1986. CO 2 fluxes in the tropical Atlantic during FOCAL cruises. *Journal of Geophysical Research*, 91(C10), p.11741.

Arndt, A. & Smith, M.J., 1998. Genetic diversity and population structure in two species of sea cucumber: differing patterns according to mode of development. *Molecular Ecology*, 7, pp.1053–1064.

Astraldi, M. et al., 1995. Climatic fluctuations, current variability and marine species distribution: a case study in the Ligurian Sea (north-west Mediterranean). *Oceanologica Acta*, 18(2), pp.139–149.

Avila, C., 1995. Natural products of opisthobranch molluscs: a Bioligical Review. *Oceanography and Marine Biology*, 33, pp.487–559.

Avise, J., 2000. Phylogeography: The history and formation of species. In *Harvard University Press*. p. 447.

Avise, J.C., 1992. Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. *Oikos*, 63(1), pp.62–76.

Bakun, A. & Nelson, C.S., 1991. The Seasonal Cycle of Wind-Stress Curl in Subtropical Eastern Boundary Current Regions. *Journal of Physical Oceanography*, 21(12), pp.1815–1834.

Ballard, J.W.O. & Whitlock, M.C., 2004. The incomplete natural history of mitochondria. *Molecular Ecology*, 13(4), pp.729–744.

Bandelt, H.J., Forster, P. & Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16(1), pp.37–48.

Barbazuk, W.B. & Schnable, P.S., 2011. SNP discovery by transcriptome pyrosequencing. *Methods in Molecular Biology*, 729, pp.225–246.

Barber, P.H. et al., 2002. Sharp genetic breaks among populations of Haptosquilla pulchella (Stomatopoda) indicate limits to larval transport: Patterns, causes, and consequences. *Molecular Ecology*, 11(4), pp.659–674.

Barley, A.J. et al., 2010. Fourteen nuclear genes provide phylogenetic resolution for difficult nodes in the turtle tree of life.

Molecular Phylogenetics and Evolution, 55(3), pp.1189–1194.

Barton, E.D. et al., 2007. Mesoscale structure viewed by SAR and AVHRR near the Canary Islands. *Scientia Marina*, 65(S1), pp.167–175.

Battaglia, P. et al., 2010. Relationships between otolith size and fish size in some mesopelagic and bathypelagic species from the Mediterranean Sea (Strait of Messina, Italy). *Scientia Marina*, 74(3), pp.605–612.

Beheregaray, L.B. & Caccone, A., 2007. Cryptic biodiversity in a changing world. *Journal of Biology*, 6(4), p.9.

Belfiore, N.M., Liu, L. & Moritz, C., 2008. Multilocus phylogenetics of a rapid radiation in the genus Thomomys (Rodentia: Geomyidae). *Systematic biology*, 57(2), pp.294–310.

Bell, J.J., 2008. Similarity in connectivity patterns for two gastropod species lacking pelagic larvae. *Marine Ecology Progress Series*, 357, pp.185–194.

Bellwood, D.R., Van Herwerden, L. & Konow, N., 2004. Evolution and biogeography of marine angelfishes (Pisces: Pomacanthidae). *Molecular Phylogenetics and Evolution*, 33(1), pp.140–155.

Benzie, J.A.H., 1999. Major genetic differences between crown-of-thorns starfish (Acanthasterplanci) populations in the Indian and Pacific Oceans. *Evolution*, 53, pp.1782–1795.

Bernt, M. et al., 2013. A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny. *Molecular Phylogenetics and Evolution*, 69(2), pp.352–364.

Bianchi, C.N. & Morri, C., 2000. Marine Biodiversity of the Mediterranean Sea: Situation, Problems and Prospects for Future Research. *Marine Pollution Bulletin*, 40(5), pp.367–376.

Bickford, D. et al., 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, 22(3), pp.148–155.

Billard, J.B., 1974. The World of the American Indian. In *Ecosystems of the World*. p. 663.

Blanquer, A. & Uriz, M.J., 2007. Cryptic speciation in marine sponges evidenced by mitochondrial and nuclear genes: A phylogenetic approach. *Molecular Phylogenetics and Evolution*, 45(1), pp.392–397.

Blunt, J., Buckingham, J. & Munro, M., 2012. *Handbook of Marine Natural Products* E. Fattorusso, W. H. Gerwick, & O. Taglialatela-Scafati, eds., Dordrecht, Netherlands: Springer.

Bohonak, A.J., 1999. Dispersal, gene flow, and population structure. *Quarterly Review of Biology*, 74(1), pp.21–45.

Bolger, A.M., Lohse, M. & Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, pp.2114–2120.

Bolshoy, A., Kaplunovsky Alexander & Ivashchenko Anatoliy, 2011. Statistical analysis of exon lengths in various eukaryotes. *Open Access Bioinformatics*, pp.1–15.

Boore, J.L., 1999. Animal mitochondrial genomes. *Nucleic Acids Research*, 27(8), pp.1767–1780.

Bouckaert, R. et al., 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology*, 10(4), p.e1003537.

Bouckaert, R.R., 2010. DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics*, 26(10), pp.1372–1373.

Brumfield, R.T. et al., 2008. Comparison of species tree methods for reconstructing the phylogeny of bearded manakins (Aves: Pipridae, Manacus) from multilocus sequence data. *Systematic Biology*, 57(5), pp.719–731.

Cabezas, P., Macpherson, E. & Machordom, A., 2011. Allogalathea (Decapoda: Galatheidae): a monospecific genus of squat lobster? *Zoological Journal of the Linnean Society*, 162, pp.245–270.

Calado, G. & Silva, J., 2012. Sea slugs of the Algarve : opisthobranch guide for the southern coast of Portugal. In Subnauta, ed. p. 164.

Campo, D. et al., 2009. Phylogeography of the European stalked barnacle (Pollicipes pollicipes): Identification of glacial refugia. *Marine Biology*, 157, pp.147–156.

Cárdenas, L., Castilla, J.C. & Viard, F., 2009. A phylogeographical analysis across three biogeographical provinces of the south-eastern Pacific: The case of the marine gastropod Concholepas concholepas. *Journal of Biogeography*, 36(5), pp.969–981.

Carling, M.D. & Brumfield, R.T., 2007. Gene sampling strategies for multi-locus population estimates of genetic diversity (theta). *PloS One*, 2(1), p.e160.

Carminati, E. & Doglioni, C., 2004. Europe - Mediterranean tectonics. *Encyclopedia of Geology*, 2, pp.135–146.

Carmona, L. et al., 2011. Amphi-Atlantic distributions and cryptic species in Sacoglossan sea slugs. *Journal of Molluscan Studies*, 77(4), pp.401–412.

Carreras-Carbonell, J., Macpherson, E. & Pascual, M., 2005. Rapid radiation and cryptic speciation in mediterranean triplefin blennies (Pisces: Tripterygiidae) combining multiple genes. *Molecular Phylogenetics and Evolution*, 37(3), pp.751–761.

Carriglio, D., Fanelli, G. & Rubino, F., 2004. First record of the alien gastropod Melibe fimbriata (Opistobranchia: Tethyidae) in the Taranto seas (Mediterranean Sea). *Journal of the Marine Biological Association of the UK*, 84(5), pp.1067–1068.

Carstens, B.C. & Knowles, L.L., 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from Melanoplus grasshoppers. *Systematic biology*, 56(3), pp.400–411.

Carstens, B.C. & Richards, C.L., 2007. Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, 61(6), pp.1439–1454.

Castelin, M. et al., 2012. Speciation patterns in gastropods with long-lived larvae from deep-sea seamounts. *Molecular Ecology*, 21(19), pp.4828–4853.

Chen, C.-T.A., 2007. The Global Coastal Ocean: Interdisciplinary Regional Studies and Syntheses. In *Eos, Transactions American Geophysical Union*. Harvard University Press, p. 172.

Chevolot, M. et al., 2006. Phylogeography and population structure of thornback rays (Raja clavata L., Rajidae). *Molecular Ecology*, 15(12), pp.3693–3705.

Chevreux, B., Wetter, T. & Suhai, S., 1999. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*, (1995), pp.45–56.

Cho, W. & Shank, T.M., 2010. Incongruent patterns of genetic connectivity among four ophiuroid species with differing coral host specificity on North Atlantic seamounts. *Marine Ecology*, 31(1), pp.121–143.

Cimino, G. et al., 1993. Biotransformation of a dietary sesterterpenoid in the Mediterranean nudibranch Hypselodoris orsini. *Experientia*, 49(6-7), pp.582–586.

Cimino, G. et al., 1983. Dorid Nudibranch Elaborates Its Own Chemical Defense. *Science*, 219(4589), pp.1237–1238.

Cimino, G. & Ghiselin, M.T., 1999. Chemical defense and evolutionary trends in biosynthetic capacity among dorid nudibranchs (Mollusca: Gastropoda: Opisthobranchia). *Chemoecology*, 9, pp.187–207.

Clark, M.S. et al., 2010. Insights into shell deposition in the Antarctic bivalve *Laternula elliptica*: gene discovery in the mantle transcriptome using 454 pyrosequencing. *BMC Genomics*, 11(1), p.362.

CLIMAP Project Members et al., 1984. The last interglacial ocean. *Quaternary Research*, 21(2), pp.123–224.

Coates, A.G. et al., 1992. Closure of the Isthmus of Panama : The near-shore marine record of Costa Rica and western Panama Closure of the Isthmus of Panama : The near-shore marine record of Costa Rica and western Panama. *Geoogical Society of America*, 104(7), pp.814–828.

Coelho, R. & Calado, G., 2010. Spawn and early development of NE Atlantic species of Hypselodoris (Gastropoda: Opisthobranchia). *Sociedad Española de Malacología*, 2(28), pp.63–72.

Collin, R., 2001. The effects of mode of development on phylogeography and population structure of North Atlantic Crepidula (Gastropoda: Calyptraeidae). *Molecular Ecology*, 10, pp.2249–2262.

Collins, L.S. et al., 1996. The late Miocene Panama isthmian strait. *Geology*, 24(8), pp.687–690.

Conabio, 2006. Opistobranquios de México: Guía de babosas marinas del Pacífico, Golfo de California y las islas oceánicas. In p. 143.

Conesa, A. et al., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21, pp.3674–3676.

Cook, B.D., Page, T.J. & Hughes, J.M., 2008. Importance of cryptic species for identifying 'representative' units of biodiversity for freshwater conservation. *Biological Conservation*, 141, pp.2821–2831.

Cornman, R.S. et al., 2014. Transcriptomic analysis of the mussel Elliptio complanata identifies candidate stress-response genes and an abundance of novel or noncoding transcripts. *PloS One*, 9(11), p.e112420.

Cortesi, F. & Cheney, K.L., 2010. Conspicuousness is correlated with toxicity in marine opisthobranchs. *Journal of Evolutionary Biology*, 23, pp.1509–1518.

Couceiro, L. et al., 2007. Genetic isolation by distance among populations of the netted dog whelk Nassarius reticulatus (L.) along the European Atlantic coastline. *Journal of Heredity*, 98(6), pp.603–610.

da Cruz, J.F., Gaspar, H. & Calado, G., 2011. Turning the game around: toxicity in a nudibranch-sponge predator–prey association. *Chemoecology*, 22(1), pp.47–53.

Cunha, R.L. et al., 2005. Patterns of cladogenesis in the venomous marine gastropod genus *Conus* from the Cape Verde islands. *Systematic biology*, 54(4), pp.634–50.

Davey, J.L. & Blaxter, M.W., 2010. RADseq: Next-generation population genetics. *Briefings in Functional Genomics*, 9, pp.416–423.

Dias, A., 1997. A história da evolução do litoral portugués nos úiltimos vinte milénios. In L. Tavares, A.; Tavares Maria; Cardoso, ed. *Evolução geohistórica do litoral português e fenómenos correlativos*. Universidade Aberta, pp. 158–170.

Diehl, J.M., Toonen, R.J. & Botsford, L.W., 2007. Spatial variability of recruitment in the sand crab *Emerita analoga* throughout California in relation to wind-driven currents. *Marine Ecology Progress Series*, 350, pp.1–17.

Dinapoli, A. & Klussmann-Kolb, A., 2010. The long way to diversity - Phylogeny and evolution of the Heterobranchia (Mollusca: Gastropoda). *Molecular Phylogenetics and Evolution*, 55, pp.60–76.

Dinapoli, A., Zinssmeister, C. & Klussmann-Kolb, A., 2011. New insights into the phylogeny of the Pyramidellidae (Gastropoda). *Journal of Molluscan Studies*, 77, pp.1–7.

Domingues, V.S. et al., 2005. Historical colonization and demography of the Mediterranean damselfish, *Chromis chromis*. *Molecular Ecology*, 14, pp.4051–4063.

Domingues, V.S. et al., 2007. Phylogeography and evolution of the triplefin *Tripterygion delaisi* (Pisces, Blennioidei). *Marine Biology*, 150, pp.509–519.

Domingues, V.S. et al., 2008. Tropical fishes in a temperate sea: evolution of the wrasse *Thalassoma pavo* and the parrotfish *Sparisoma cretense* in the Mediterranean and the adjacent Macaronesian and Cape Verde Archipelagos. *Marine Biology*, 154(3), pp.465–474.

Dorgan, K.M., Valdés, Á. & Gosliner, T.M., 2002. Phylogenetic systematics of the genus *Platydoris* (Mollusca, Nudibranchia, Doridoidea) with descriptions of six new species. *Zoologica Scripta*, 31, pp.271–319.

Doyle, J.M. et al., 2014. The genome sequence of a widespread apex predator, the golden eagle (*Aquila chrysaetos*). *PLoS One*, 9(4), p.e95599.

Drummond, A.J. et al., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*,

29(8), pp.1969–73.

Drummond, A.J. et al., 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4(5), pp.699–710.

Duda, T.F. & Kohn, A.J., 2005. Species-level phylogeography and evolutionary history of the hyperdiverse marine gastropod genus *Conus*. *Molecular Phylogenetics and Evolution*, 34, pp.257–272.

Duggen, S. et al., 2003. Deep roots of the Messinian salinity crisis. *Nature*, 422(6932), pp.602–606.

Edwards, S. & Beerli, P., 2000. Perspective: Gene Divergence, Population Divergence, and the Variance in Coalescence Time in Phylogeographic Studies. *Evolution*, 54(6), pp.1839–1854.

Eilertsen, M.H. & Malaquias, M.A.E., 2013. Systematic revision of the genus Scaphander (Gastropoda, Cephalaspidea) in the Atlantic Ocean, with a molecular phylogenetic hypothesis. *Zoological Journal of the Linnean Society*, 167(3), pp.389–429.

Ekblom, R. et al., 2014. Characterization of the house sparrow (*Passer domesticus*) transcriptome: a resource for molecular ecology and immunogenetics. *Molecular ecology resources*, 14(3), pp.636–646.

Ellingson, R.A. & Krug, P.J., 2011. Cryptic Speciation , Evolution of Poecilogony From Planktotrophy : and Larval Development in the Gastropod. *Society*, 60(11), pp.2293–2310.

Elmer, K.R. & Meyer, A., 2010. Sympatric speciation without borders? *Molecular Ecology*, 19(10), pp.1991–1993.

Espinosa, F. et al., 2011. Population genetic structure of the endangered limpet *Cymbula nigra* in a temperate Northern hemisphere region: Influence of palaeoclimatic events? *Marine Ecology*, 32(1994), pp.1–5.

Ewen-Campen, B. et al., 2011. The maternal and early embryonic transcriptome of the milkweed bug *Oncopeltus fasciatus*. *BMC Genomics*, 12(1), p.61.

Faulkner, D. & Ghiselin, M., 1983. Chemical defense and evolutionary ecology of dorid nudibranchs and some other opisthobranch gastropods . *Marine Ecology Progress Series*, 13, pp.295–301.

Feldmeyer, B. et al., 2011. Short read Illumina data for the de novo assembly of a non-model snail species transcriptome (*Radix balthica*, Basommatophora, Pulmonata), and a comparison of assembler performance. *BMC Genomics*, 12(1), p.317.

Feliciano, T.S. dos S., 2009. *Avaliação do perfil químico do nudibrânquio hypselodoris fontandraui e de consórcios bacterianos provenientes de esponjas marinhas*. Ph.D. Dissertation. Universidade de Lisboa.

Felsenstein, J., 2006. Accuracy of coalescent likelihood estimates: Do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution*, 23(3), pp.691–700.

Finlay, B.J., 2002. Global dispersal of free-living microbial eukaryote species. *Science (New York, N.Y.)*, 296, pp.1061–1063.

Fontaine, B., Roucou, P. & Janicot, S., 1999. Coupled Ocean-Atmosphere Surface Variability and Its Climate Impacts in the Tropical Atlantic Region. *Climate Dynamics*, 15, pp.451–473.

Fontana, A. et al., 1999. New scalaranes from the nudibranch Glossodoris atromarginata and its sponge prey. *Journal of Natural Products*, 62(10), pp.1367–1370.

Fontana, A. et al., 2001. Selected chemo-ecological studies of marine opisthobranchs from Indian coasts. *Journal of the Indian Institute of Science*, 81, pp.403–415.

de Forges, B.R., Koslow, J.A. & Poore, G.C., 2000. Diversity and endemism of the benthic seamount fauna in the southwest Pacific. *Nature*, 405(6789), pp.944–947.

Francisco, S.M. et al., 2014a. Phylogeographic pattern and glacial refugia of a rocky shore species with limited dispersal capability: the case of Montagu's blenny (Coryphoblennius galerita, Blenniidae). *Marine Biology*, 161(11), pp.2509–2520.

Francisco, S.M. et al., 2014b. In search of phylogeographic patterns in the northeastern atlantic and adjacent seas. *Evolutionary Biology: Genome Evolution, Speciation, Coevolution and Origin of Life*, pp.323–338.

Fraser, C.I., Nikula, R. & Waters, J.M., 2011. Oceanic rafting by a coastal community. *Proceedings of the Royal Society B: Biological Sciences*, 278(1706), pp.649–55.

Frey, M. & Vermeij, G.J., 2008. Molecular phylogenies and historical biogeography of a circumtropical group of gastropods (Genus: Nerita): implications for regional diversity patterns in the marine tropics. *Molecular Phylogenetics and Evolution*, 48(3), pp.1067–1086.

Frick, K., 2003. Response in Nematocyst Uptake by the Nudibranch Flabellina verrucosa to the Presence of Various Predators in the Southern Gulf of Maine. *Biological Bulletin*, 205(3), pp.367–376.

Frouin, R. et al., 1990. Observations of a poleward surface current off the coasts of Portugal and Spain during winter. *Journal of Geophysical Research*, 95(C1), p.679.

Fu, Y.X., 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147(2), pp.915–925.

Furfaro, G. et al., 2016. A DNA-barcoding approach to the phenotypic diversity of Mediterranean species of Felimare Ev. Marcus & Er. Marcus, 1967 (Mollusca: Gastropoda), with a preliminary phylogenetic analysis. *Italian Journal of Zoology*, pp.1–13.

Galarza, J.A. et al., 2009. The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species. *Proceedings of the National Academy of Sciences of the United States of America*, 106(5), pp.1473–8.

Gan, H.M., Schultz, M.B. & Austin, C.M., 2014. Integrated shotgun sequencing and bioinformatics pipeline allows ultra-fast mitogenome recovery and confirms substantial gene rearrangements in Australian freshwater crayfishes. *BMC Evolutionary Biology*, 14(1), p.19.

Gan, H.Y. et al., 2015. The complete mitogenome of the hermit crab *Clibanarius infraspinatus* (Hilgendorf, 1869), (Crustacea; Decapoda; Diogenidae) – a new gene order for the Decapoda. *Mitochondrial DNA*, pp.1–2.

Garg, R. et al., 2011. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Research*, 18(1), pp.53–63.

Garovoy, J.B., Valdés, Á. & Gosliner, T.M., 2001. Phylogeny of the Genus *Rostanga* (Nudibranchia), With Descriptions of Three New Species From South Africa. *Journal of Molluscan Studies*, 67, pp.131–144.

Garrick, R.C., Sunnucks, P. & Dyer, R.J., 2010. Nuclear gene phylogeography using PHASE: dealing with unresolved genotypes, lost alleles, and systematic bias in parameter estimation. *BMC Evolutionary Biology*, 10, p.118.

Geldmacher, J. et al., 2001. Earlier history of the ≥70-Ma-old Canary hotspot based on the temporal and geochemical evolution of the Selvagen Archipelago and neighboring seamonts in the Eastern North Atlantic. *Journal of Volcanology and Geothermal Research*, 111(1-4), pp.55–87.

George, K.H., 2013. Faunistic research on metazoan meiofauna from seamounts - a review. *Meiofauna Marina*, 20, pp.1–32.

Geraldes, A. et al., 2011. SNP discovery in black cottonwood (*Populus trichocarpa*) by population transcriptome resequencing. *Molecular Ecology Resources*, 11(1), pp.81–92.

Gilbert-Horvath, E.A., Larson, R.J. & Garza, J.C., 2006. Temporal recruitment patterns and gene flow in kelp rockfish (*Sebastes atrovirens*). *Molecular Ecology*, 15(12), pp.3801–3815.

Glez-Peña, D. et al., 2010. ALTER: Program-oriented conversion of DNA and protein alignments. *Nucleic Acids Research*, 38(2), pp.W14-W18.

Goodheart, J. et al., 2015. Systematics and biogeography of *Pleurobranchus* Cuvier, 1804, sea slugs (Heterobranchia: Nudipleura: Pleurobranchidae). *Zoological Journal of the Linnean Society*, 174(2), pp.322–362.

Gopurenko, D. & Hughes, J.M., 2002. Regional patterns of genetic structure among Australian populations of the mud crab, *Scylla serrata* (Crustacea: Decapoda): Evidence from mitochondrial DNA. *Marine and Freshwater Research*, 53(5), pp.849–857.

Gosliner, T. & Behrens, D., 1990. Special resemblance, aposematic coloration and mimicry in opisthobranch gastropods. *Texas A&M University Sea Grant College Program, College Station*, Station 1, pp.127–138.

Gosliner, T.M., 2001. Aposematic coloration and mimicry in opisthobranch mollusks: new phylogenetic and experimental data. *Bollettino Malacologico*, 37(5-8), pp.163–170.

Gosliner, T.M., Cervera, J.L. & Ghiselin, M.T., 2008. Scientific Exploration in the Mediterranean Region Biodiversity of the Mediterranean Opisthobranch Gastropod Fauna : Historical and Phylogenetic Perspectives. *Proceedings of the California Academy of Sciences*, 59(8), pp.117–137.

Gosliner, T.M. & Draheim, R., 1996. Indo-Pacific opisthobranch gastropod biogeography: How do we know what we don't know? *American Malacological Bulletin*, 12, pp.37–43.

Gosliner, T.M. & Johnson, R.F., 1999. Phylogeny of *Hypselodoris* (Nudibranchia : Chromodorididae) with a review of the monophyletic clade of Indo-Pacific species, including descriptions of twelve new species. *Zoological Journal of the Linnean Society*, 125(1), pp.1–114.

Grabherr, M.G. et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), pp.644–52.

Grabherr, M.G. et al., 2013. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, 29(7), pp.644–652.

Grande, C. et al., 2004. Phylogenetic relationships among Opisthobranchia (Mollusca: Gastropoda) based on mitochondrial cox 1, trnV, and rrnL genes. *Molecular Phylogenetics and Evolution*, 33, pp.378–388.

Grande, C. et al., 2002. The Complete Mitochondrial Genome of the Nudibranch *Roboastra europaea* (Mollusca: Gastropoda) Supports the Monophyly of Opisthobranchs. *Molecular Biology and Evolution*, 19(10), pp.1672–1685.

Grande, C., Templado, J. & Zardoya, R., 2008. Evolution of gastropod mitochondrial genome arrangements. *BMC Evolutionary Biology*, 8, p.61.

Grant, W.S. & Bowen, B.W., 1998. Shallow Population Historiesin Deep Evolutionnary Lineages of Marine Fishhes: Insights From Sardines and Anchovies and Lesson for Conservation. *Journal of Heredity*, 89, pp.415–426.

Greenwood, P.G. & Mariscal, R.N., 1984. The utilization of cnidarian nematocysts by aeolid nudibranchs: nematocyst maintenance and release in Spurilla. *Tissue & cell*, 16, pp.719–730.

Groenenberg, D.S.J. et al., 2012. The complete mitogenome of *Cylindrus obtusus* (Helicidae, Ariantinae) using Illumina next generation sequencing. *BMC Genomics*, 13(1), p.114.

Guan, D.-L. & Xu, S.-Q., 2015. Complete mitochondrial genome of the geophilous grasshopper *Trilophidia annulata* (Acrididae: Oedipodinae: Trilophidia). *Mitochondrial DNA*, pp.1–2.

Haas, B.J. et al., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8), pp.1494–1512.

Haber, M. et al., 2010. Coloration and defense in the nudibranch gastropod *Hypselodoris fontandraui*. *The Biological Bulletin*, 218(2), pp.181–8.

Hahn, C., Bachmann, L. & Chevreux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - A baiting and iterative mapping approach. *Nucleic Acids Research*, 41(13), pp.1–9.

Hallas, J.M. & Gosliner, T.M., 2015. Family matters: The first molecular phylogeny of the Onchidoridídae Gray, 1827 (Mollusca, Gastropoda, Nudibranchia). *Molecular Phylogenetics and Evolution*, 88, pp.16–27.

Hall-Spencer, J. & Rogers, A., 2007. Deep-sea coral distribution on seamounts, oceanic islands, and continental slopes in the Northeast Atlantic. *Bulletin of Marine Science*, 81(3), pp.135–146.

Hartmann, S. et al., 2012. Exploiting gene families for phylogenomic analysis of myzostomid transcriptome data. *PLoS One*, 7(1), p.e29843.

Harzhauser, M. et al., 2007. Biogeographic responses to geodynamics: A key study all around the Oligo-Miocene Tethyan Seaway. *Zoologischer Anzeiger*, 246(4), pp.241–256.

Harzhauser, M., Piller, W.E. & Steininger, F.F., 2002. Circum-Mediterranean Oligo-Miocene biogeographic evolution - the gastropods' point of view. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 183, pp.103–133.

Hatzoglou, E., Rodakis, G.C. & Lecanidou, R., 1995. Complete sequence and gene organization of the mitochondrial genome of the land snail *Albinaria coerulea*. *Genetics*, 140(4), pp.1353–66.

Haynes, R. & Barton, E.D., 1990. A poleward flow along the Atlantic coast of the Iberian Peninsula. *Journal of Geophysical Research C: Oceans*, 95(7), pp.11425–11441.

Hebert, P.D., Ratnasingham, S. & DeWaard, J.R., 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(Suppl 1), p.S96.

Hedgecock, D., 1986. Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates? *Bulletin of Marine Science*, 39, pp.550–564.

Heled, J. & Drummond, A.J., 2008. Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, 8, p.289.

Heled, J. & Drummond, A.J., 2010. Bayesian Inference of Species Trees from Multilocus Data. *Molecular Biology and Evolution*, 27(3), pp.570–580.

Hellberg, M.E. et al., 2002. Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science*, 70(1), pp.273–290.

Hellberg, M.E. & Vacquier, V.D., 1999. Rapid evolution of fertilization selectivity and lysin cDNA sequences in teguline gastropods. *Molecular Biology and Evolution*, 16(6), pp.839–848.

Henry, C.S., 1994. Singing and cryptic speciation in insects. *Trends in Ecology and Evolution*, 9(10), pp.388–392.

Henschel, R. et al., 2012. Trinity RNA-Seq assembler performance optimization. *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment on Bridging from the eXtreme to the campus and beyond - XSEDE '12*, p.45.

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), pp.907–913.

Hewitt, G.M., 2011. Quaternary phylogeography: The roots of hybrid zones. *Genetica*, 139(5), pp.617–638.

Hey, J. & Nielsen, R., 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, 167(2), pp.747–60.

von der Heyden, S. et al., 2011. Phylogeographic patterns and cryptic speciation across oceanographic barriers in South African intertidal fishes. *Journal of Evolutionary Biology*, 24(11), pp.2505–2519.

Hillier, J.K. & Watts, A.B., 2007. Global distribution of seamounts from ship-track bathymetry data. *Geophysical Research Letters*, 34(13).

Hittinger, C.T. et al., 2010. Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. *Proceedings of the National Academy of Sciences of the United States of America*, 107(4), pp.1476–81.

Hoffman, J.I. et al., 2011. Strong population genetic structure in a broadcast-spawning antarctic marine invertebrate. *Journal of Heredity*, 102(1), pp.55–66.

Hohenlohe, P.A., 2003. Distribution of sister Littorina species, II: Geographic and tidal-height patterns do not support sympatric spaciation. *The Veliger*, 46, pp.211–219.

Hohenlohe, P.A. et al., 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *Genetics*, 6(2).

Hou, R. et al., 2011. Transcriptome sequencing and de novo analysis for Yesso scallop (*Patinopecten yessoensis*) using 454 GS FLX. *PloS One*, 6(6), p.e21560.

Howell, K.L., Mowles, S.L. & Foggo, A., 2010. Mounting evidence: Near-slope seamounts are faunally indistinct from an adjacent bank. *Marine Ecology*, 31(1), pp.52–62.

Hrbek, T. & Meyer, A., 2003. Closing of the Tethys Sea and the phylogeny of Eurasian killifishes (Cyprinodontiformes: Cyprinodontidae). *Journal of Evolutionary Biology*, 16(1), pp.17–36.

Hu, H. et al., 2011. Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. *BMC genomics*, 12(1), p.60.

Hughes, P. & Barton, E.D., 1974. Stratification and water mass structure in the upwelling area off northwest Africa in April/May 1969. *Deep Sea Research and Oceanographic Abstracts*, 21(8), pp.611–628.

Huthnance, J.M. et al., 2002. Ocean margin exchange - Water flux estimates. *Journal of Marine Systems*, 32(1-3), pp.107–137.

Ilves, K.L. et al., 2010. Colonization and/or mitochondrial selective sweeps across the North Atlantic intertidal assemblage revealed by multi-taxa approximate Bayesian computation. *Molecular ecology*, 19(20), pp.4505–19.

Jablonski, D., 1986. Larval ecology and macroevolution in marine invertebrates. *Bulletin of Marine Science*, 39(2), pp.565–587.

Jaziri, H. & Benazzou, T., 2002. Differenciation allozymique multilocus des populations de moule *Mytilus galloprovincialis* Lmk. des côtes marocaines. *Comptes Rendus - Biologies*, 325(12), pp.1175–1183.

Jeffery, C.H. & Emlet, R.B., 2003. Macroevolutionary Consequences of Developmental Mode in Temnopleurid Echinoids From the Tertiary of Southern Australia. *Evolution*, 57(5), pp.1031–1048.

Jeffery, C.H., Emlet, R.B. & Littlewood, D.T.J., 2003. Phylogeny and evolution of developmental mode in temnopleurid echinoids. *Molecular Phylogenetics and Evolution*, 28(1), pp.99–118.

Johannesson, K. & André, C., 2006. Life on the margin: Genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology*, 15(8), pp.2013–2029.

Johnson, J. & Stevens, I., 2000. A fine resolution model of the eastern North Atlantic between the Azores, the Canary Islands and the Gibraltar Strait. *Deep Sea Research Part I: Oceanographic Research Papers*, 47(5), pp.875–899.

Johnson, R.F., 2010. Breaking family ties: taxon sampling and molecular phylogeny of chromodorid nudibranchs (Mollusca, Gastropoda). *Zoologica Scripta*, 40(2), pp.137–157.

Johnson, R.F. & Gosliner, T.M., 2012. Traditional taxonomic groupings mask evolutionary history: a molecular phylogeny and new classification of the chromodorid nudibranchs. *PloS One*, 7(4), p.e33479.

Jörger, K.M. et al., 2012. Barcoding against a paradox? Combined molecular species delineations reveal multiple cryptic lineages in elusive meiofaunal sea slugs. *BMC Evolutionary Biology*, 12(1), p.245.

Jörger, K.M. et al., 2010. On the origin of Acochlidia and other enigmatic euthyneuran gastropods, with implications for the systematics of Heterobranchia. *BMC Evolutionary Biology*, 10(1), p.323.

Kearse, M. et al., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), pp.1647–1649.

Keigwin, L., 1982. Isotopic paleoceanography of the Caribbean and East pacific: role of panama uplift in late neogene time. *Science (New York, N.Y.)*, 217, pp.350–353.

Kelly, R.P. & Palumbi, S.R., 2010. Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. *PloS One*, 5(1), p.e8594.

Kettle, A.J. et al., 2011. Refugia of marine fish in the northeast Atlantic during the last glacial maximum: Concordant assessment from archaeozoology and palaeotemperature reconstructions. *Climate of the Past*, 7(1), pp.181–201.

Kijjoa, A. & Sawangwong, P., 2004. Drugs and Cosmetics from the Sea. *Marine Drugs*, 2, pp.73–82.

Kingston, T. et al., 2001. Acoustic divergence in two cryptic Hipposideros species: a role for social selection? *Proceedings Biological Sciences*, 268(1474), pp.1381–1386.

Kinlan, B.P., Gaines, S.D. & Lester, S.E., 2005. Propagule dispersal and the scales of marine community process. *Diversity and Distributions*, 11, pp.139–148.

Klimley, A., Richert, J. & Jorgensen, S., 2005. The home of blue water fish. *American Scientist*, 93(1), pp.42–49.

Klussmann-Kolb, A. et al., 2008. From sea to land and beyond--new insights into the evolution of euthyneuran Gastropoda (Mollusca). *BMC Evolutionary Biology*, 8(1), p.57.

Knowlton, N., 2000. Molecular genetic analyses of species boundaries in the sea. *Hydrobiologia*, 420, pp.73–90.

Kocot, K.M. et al., 2011. Phylogenomics reveals deep molluscan relationships. *Nature*, 477(7365), pp.452–456.

Kocot, K.M., Halanych, K.M. & Krug, P.J., 2013. Phylogenomics supports Panpulmonata: Opisthobranch paraphyly and key evolutionary steps in a major radiation of gastropod molluscs. *Molecular Phylogenetics and Evolution*, 69, pp.764–771.

Krijgsman, W. et al., 1999. Chronology, causes and progression of the Messinian salinity crisis. *Nature*, 400(6745), pp.652–655.

Kristensen, D.M. et al., 2011. Computational methods for Gene Orthology inference. *Briefings in Bioinformatics*, 12(5), pp.379–391.

Krug, P.J., 2011. Patterns of speciation in marine gastropods: A review of the phylogenetic evidence for localized radiations in the sea. *American Malacological Bulletin*, 29(1-2), pp.169–186.

Krzeminska, U. et al., 2015. Mitochondrial genomes of the jungle crow *Corvus macrorhynchos* (Passeriformes: Corvidae) from shed feathers and a phylogenetic analysis of genus using mitochondrial protein-coding genes. *Mitochondrial DNA*, pp.1–3.

Kubatko, L.S., Carstens, B.C. & Knowles, L.L., 2009. STEM: Species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics*, 25(7), pp.971–973.

Künstner, A. et al., 2010. Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. *Molecular Ecology*, 19(1), pp.266–276.

Kurabayashi, A. & Ueshima, R., 2000. Complete sequence of the mitochondrial DNA of the primitive opisthobranch gastropod *Pupa strigosa*: systematic implication of the genome organization. *Molecular Biology and Evolution*, 17(2), pp.266–77.

Lambeck, K., Esat, T.M. & Potter, E.K., 2002. Links between climate and sea levels for the past three million years. *Nature*, 419(6903), pp.199–206.

Landry, C. et al., 2003. Recent speciation in the Indo-West Pacific: rapid evolution of gamete recognition and sperm morphology in cryptic species of sea urchin. *Proceedings. Biological sciences / The Royal Society*, 270(1526), pp.1839–1847.

Lanfear, R. et al., 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, 29(6), pp.1695–1701.

Latiolais, J.M. et al., 2006. A molecular phylogenetic analysis of strombid gastropod morphological diversity. *Molecular Phylogenetics and Evolution*, 41(2), pp.436–444.

Lavery, S. et al., 2004. Phylogenetic relationships and evolutionary history of the shrimp genus *Penaeus* s.l. derived from mitochondrial DNA. *Molecular Phylogenetics and Evolution*, 31(1), pp.39–49.

Lechner, M. et al., 2011. Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics*, 12(1), p.124.

Lefébure, T. et al., 2006. Phylogeography of a subterranean amphipod reveals cryptic diversity and dynamic evolution in extreme environments. *Molecular ecology*, 15(7), pp.1797–806.

Lefèvre, N., 2009. Low CO2 concentrations in the Gulf of Guinea during the upwelling season in 2006. *Marine Chemistry*,

113(1-2), pp.93–101.

Leigh, E.G., O'Dea, A. & Vermeij, G.J., 2014. Historical biogeography of the isthmus of panama. *Biological Reviews*, 89(1), pp.148–172.

Leis, J.M. & Fisher, R., 2006. Swimming speed of settlement-stage reef-fish larvae measured in the laboratory and in the field: a comparison of critical speed and in situ speed. *Proceedings of the 10th International Coral Reef Symposium*, 445, pp.438–445.

Lenfant, P. & Planes, S., 1996. Genetic differentiation of white sea bream within the Lion's Gulf and the Ligurian Sea (Mediterranean Sea). *Journal of Fish Biology*, 49(4), pp.613–621.

Lessios, H. a et al., 1999. Phylogeography of the pantropical sea urchin *Eucidaris* in relation to land barriers and ocean currents. *Evolution*, 53(3), pp.806–817.

Lessios, H.A., Kessing, B.D. & Pearse, J.S., 2007. Population Structure and Speciation in Tropical Seas: Global Phylogeography of the Sea Urchin *Diadema*. *Evolution*, 55(5), pp.955–975.

Levin, L.A., 2006. Recent progress in understanding larval dispersal: New directions and digressions. In *Integrative and Comparative Biology*. pp. 282–297.

Li, B. & Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12(1), p.323.

Lindström, L. et al., 2004. the Effect of Alternative Prey on the Dynamics of Imperfect Batesian and Müllerian Mimicries. *Evolution*, 58(6), pp.1294–1302.

Liu, L. & Pearl, D.K., 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Systematic biology*, 56(3), pp.504–514.

Lourie, S.A., Green, D.M. & Vincent, A.C.J., 2005. Dispersal, habitat differences, and comparative phylogeography of Southeast Asian seahorses (Syngnathidae: Hippocampus). *Molecular Ecology*, 14(4), pp.1073–1094.

Maan, M.E. & Cummings, M.E., 2012. Poison Frog Colors Are Honest Signals of Toxicity, Particularly for Bird Predators. *The American Naturalist*, 179(1), pp.E1–E14.

Maggs, C.A. et al., 2008. Evaluating signatures of glacial refugia for north atlantic benthic marine taxa. *Ecology*, 89(11), pp.08–S122.

Malaquias, M.A.E. & Reid, D.G., 2008. Systematic revision of the living species of Bullidae (Mollusca: Gastropoda: Cephalaspidea), with a molecular phylogenetic analysis. *Zoological Journal of the Linnean Society*, 153, pp.453–543.

Malaquias, M.A.E. & Reid, D.G., 2009. Tethyan vicariance, relictualism and speciation: Evidence from a global molecular phylogeny of the opisthobranch genus *Bulla*. *Journal of Biogeography*, 36(9), pp.1760–1777.

Malenke, J.R., Johnson, K.P. & Clayton, D.H., 2009. Host specialization differentiates cryptic species of feather-feeding lice. *Evolution*, 63(6), pp.1427–1438.

Mantelatto, F.L. et al., 2014. Molecular analysis of the commercial deep-sea crabs *Chaceon ramosae* and *Chaceon notialis* (Brachyura, Geryonidae) reveals possible cryptic species in the South Atlantic. *Deep-Sea Research Part I: Oceanographic Research Papers*, 84, pp.29–37.

Marshall, D. et al., 2010. Phenotype-environment mismatches reduce connectivity in the sea. *Ecology Letters*, 13(1), pp.128–140.

Martel, A., 1991. Drifting and dispersal of small bivalves and gastropods with direct development. *Journal of Experimental Marine Biology and Ecology*, 150, pp.131–147.

Martin, J. a. & Wang, Z., 2011. Next-generation transcriptome assembly. *Nature Reviews Genetics*, 12(10), pp.671–682.

Martin, R., 2003. Management of nematocysts in the alimentary tract and in cnidosacs of the aeolid nudibranch gastropod

Cratena peregrina. *Marine Biology*, 143(3), pp.533–541.

Martins, C.S., Hamann, M. & Fiúza, A.F.G., 2002. Surface circulation in the eastern North Atlantic, from drifters and altimetry. *Journal of Geophysical Research*, 107(C12), p.3217.

Mason, E., Coombs, S. & Oliveira, P.B., 2006. An overview of the literature concerning the oceanography of the eastern North Atlantic region. *Relatórios Científicos e Técnicos IPIMAR Série Digital nº 33, IPIMAR, Lisboa.*, 33, p.58.

McClain, C.R. et al., 2009. Endemicity, biogeograhy, composition, and community structure on a northeast pacific seamount. *PloS One*, 4(1), p.e4141.

Mccormack, J.E. et al., 2013. Molecular Phylogenetics and Evolution Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, 66(2), pp.526–538.

McDonald, G.R. & Nybakken, J.W., 1997. List of the Worldwide Food Habits of Nudibranchs. *Veliger*, 40(2).

McGovern, T.M. & Hellberg, M.E., 2003. Cryptic species, cryptic endosymbionts, and geographical variation in chemical defences in the bryozoan *Bugula neritina*. *Molecular Ecology*, 12(5), pp.1207–1215.

McGowan, J.A., 1971. Oceanic biogeography of the Pacific. In *The Micropalaeontology of Oceans*. Cambridge University Press, p. 3.

McKibbin, W.F. et al., 2010. Risk of sperm competition moderates the relationship between men's satisfaction with their partner and men's interest in their partner's copulatory orgasm. *Personality and Individual Differences*, 49(8), pp.961–966.

Medina, M. et al., 2011. Crawling through time: Transition of snails to slugs dating back to the Paleozoic, based on mitochondrial phylogenomics. *Marine genomics*, 4(1), pp.51–9.

Merle, J. & Arnault, S., 1985. Seasonal variability of the surface dynamic topography in the tropical Atlantic Ocean. *Journal of Marine Research*, 43, pp.267–288.

Merle, J., Fieux, M. & Hisard, P., 2013. Annual signal and interannual anomalies of sea surface temperature in the eastern equatorial atlantic ocean. *Oceanography and Surface Layer Meteorology in the B/C Scale: Global Atmospheric Research Program Atlantic Tropical Experiment (Gate)*, 26, pp.77–101.

Mikkelsen, P.M., 2002. Shelled opisthobranchs. *Advances in Marine Biology*, 42, pp.67–136.

Miller, K. et al., 2010. Conflicting estimates of connectivity among deep-sea coral populations. *Marine Ecology*, 31(1), pp.144–157.

Millot, C., 1999. Circulation in the Western Mediterranean Sea. *Journal of Marine Systems*, 20(1-4), pp.423–442.

Millot, C. & Taupier-Letage, I., 2005. Circulation in the Mediterranean Sea. *Environmental Chemistry*, 5(Part K), pp.29–66.

Milne, I. et al., 2013. Using tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics*, 14(2), pp.193–202.

Mitchell-Thomé, R.C., 1985. Radiometric studies in Macaronesia. *Boletim do Museu Municipal do Funchal*, 37(167), pp.52–85.

Mokhtar-Jamai, K. et al., 2011. From global to local genetic structuring in the red gorgonian *Paramuricea clavata*: The interplay between oceanographic conditions and limited larval dispersal. *Molecular Ecology*, 20(16), pp.3291–3305.

Morato, T. et al., 2010. Seamounts are hotspots of pelagic biodiversity in the open ocean. *Proceedings of the National Academy of Sciences*, 107(21), pp.9707–9711.

Morton, B. & Britton, J.C., 2000. *The origins of the coastal and marine flora and fauna of the Azores*.

Mueller, R.L., 2006. Evolutionary rates, divergence dates, and the performance of mitochondrial genes in Bayesian phylogenetic analysis. *Systematic biology*, 55(2), pp.289–300.

Muss, A. et al., 2001. Phylogeography of *Ophioblennius:* the Role of Ocean Currents and Geography in Reef Fish Evolution.

Evolution, 55(3), p.561.

Naciri, M., 1999. Genetic study of the Atlantic/Mediterranean transition in sea bass (*Dicentrarchus labrax*). *Journal of Heredity*, 90(6), pp.591–596.

Nikula, R. et al., 2010. Circumpolar dispersal by rafting in two subantarctic kelp-dwelling crustaceans. *Marine Ecology Progress Series*, 405, pp.221–230.

Nikula, R., 2003. Phylogeography of *Cerastoderma glaucum* (Bivalvia: Cardiidae) across Europe: A major break in the Eastern Mediterranean. *Marine Biology*, 143(2), pp.339–350.

Nikulina, E.A., Hanel, R. & Schäfer, P., 2007. Cryptic speciation and paraphyly in the cosmopolitan bryozoan *Electra pilosa*-Impact of the Tethys closing on species evolution. *Molecular Phylogenetics and Evolution*, 45, pp.765–776.

Nunes, O.C. et al., 1995. Compatible Solutes in the Thermophilic Bacteria *Rhodothermus marinus* and '*Thermus thermophilus*'. *Applied and Environmental Microbiology*, 61(6), pp.2351–7.

Nylander, J.A.A. et al., 2008. AWTY (are we there yet?): A system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics*, 24(4), pp.581–583.

O'Neil, S.T. & Emrich, S.J., 2013. Assessing De Novo transcriptome assembly metrics for consistency and utility. *BMC genomics*, 14(1), p.465.

Ojala, D., Montoya, J. & Attardi, G., 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature*, 290(5806), pp.470–474.

Olsen, J.L. et al., 2004. North Atlantic phylogeography and large-scale population differentiation of the seagrass *Zostera marina*. *Molecular Ecology*, 13(7), pp.1923–1941.

Oosthuizen, A., Jiwaji, M. & Shaw, P., 2004. Genetic analysis of the *Octopus vulgaris* population on the coast of South Africa. *South African Journal of Science*, 100(11-12), pp.603–606.

Ortea, A.V. y C.G.-G.J., 1996. Revisión de las especies atlánticas de la familia Chromodoridídae (Mollusca:Nudibranquia) del grupo cromático azul. *Avicennia*, (1), p.165.

Ouagajjou, Y. & Presa, P., 2015. The connectivity of *Mytilus galloprovincialis* in northern Morocco: A gene flow crossroads between continents. *Estuarine, Coastal and Shelf Science*, 152, pp.1–10.

Packmor, J., Müller, F. & George, K., 2014. *Copepoda Harpacticoida* (Crustacea)–Shallow-water Paramesochridae Lang, 1944 from the North-East Atlantic Ocean, including the (re-) description of three species. *Progress in Oceanography*, (131), pp.59–81.

Palero, F. et al., 2008. Phylogeography of the European spiny lobster (*Palinurus elephas*): Influence of current oceanographical features and historical processes. *Molecular Phylogenetics and Evolution*, 48(2), pp.708–717.

Palumbi, S.R., 1994. Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics*, 25, pp.547–572.

Palumbi, S.R., 1992. Marine speciation on a small planet. *Trends in Ecology & Evolution*, 7(4), pp.114–118.

Palumbi, S.R., 2003. Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications*, 13(1), pp.146–158.

Palumbi, S.R. & Warner, R.R., 2003. Why Gobies Are Like Hobbits. *Science (New York, N.Y.)*, 299(5603), pp.51–52.

Pannacciulli, F.G., Bishop, J.D.D. & Hawkins, S.J., 1997. Genetic structure of populations of two species of *Chthamalus* (Crustacea: Cirripedia) in the north-east Atlantic and Mediterranean. *Marine Biology*, 128(1), pp.73–82.

Pante, E. et al., 2015. Species are hypotheses: Avoid connectivity assessments based on pillars of sand. *Molecular Ecology*, 24(3), pp.525–544.

Parchman, T.L. et al., 2010. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics*, 11, p.180.

Parin, N. V, Mironov, A.N. & Nesis, K.N., 1997. Biology of the Nazca and Sala y Gomez Submarine Ridges, an Outpost of the Indo- West Pacific Fauna in the Eastern Pacific Ocean: Composition and Distribution of the Fauna, its Communities and History. *Advances in Marine Biology*, 32, pp.145–242.

Parra, G., Bradnam, K. & Korf, I., 2007. CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, 23(9), pp.1061–1067.

Patarnello, T., Volckaert, F. a M.J. & Castilho, R., 2007. Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break? *Molecular Ecology*, 16(21), pp.4426–44.

Pechenik, J.A., 1999. On the advantages and disadvantages of larval stages in benthic marine invertebrate life cycles. *Marine Ecology Progress Series*, 177, pp.269–297.

Pelc, R.A., Warner, R.R. & Gaines, S.D., 2009. Geographical patterns of genetic structure in marine species with contrasting life histories. *Journal of Biogeography*, 36, pp.1881–1890.

Peliz, Á. & Fiuza, A.F.G., 1999. Temporal and spatial variability of CZCS-derived phytoplankton pigment concentrations off the western Iberian Peninsula. *International Journal of Remote Sensing*, 20(7), pp.1363–1403.

Pellerito, R., Arculeo, M. & Bonhomme, F., 2009. Recent expansion of Northeast Atlantic and Mediterranean populations of *Melicertus* (Penaeus) *kerathurus* (Crustacea: Decapoda). *Fisheries Science*, 75, pp.1089–1095.

Perseke, M. et al., 2013. The impact of mitochondrial genome analyses on the understanding of deuterostome phylogeny. *Molecular Phylogenetics and Evolution*, 66(3), pp.898–905.

Pfennig, D.W. et al., 2007. Population differences in predation on Batesian mimics in allopatry with their model: Selection against mimics is strongest when they are common. *Behavioral Ecology and Sociobiology*, 61, pp.505–511.

Pfennig, D.W. & Mullen, S.P., 2010. Mimics without models: causes and consequences of allopatry in Batesian mimicry complexes. *Proceedings. Biological sciences*, 277(1694), pp.2577–2585.

Picaut, J., 1983. Propagation of the Seasonal Upwelling in the Eastern Equatorial Atlantic. *Journal of Physical Oceanography*, 13(1), pp.18–37.

Pingree, R.D., 2009. The Eastern Subtropical Gyre (North Atlantic): Flow Rings Recirculations Structure and Subduction. *Journal of the Marine Biological Association of the United Kingdom*, 77(03), p.573.

Pola, M., Cervera, J.L. & Gosliner, T.M., 2008. Description of the first Roboastra species (nudibranchia, polyceridae, nembrothinae) from the Western Atlantic. *Bulletin of Marine Science*, 83(2), pp.391–399.

Pola, M., Cervera, J.L. & Gosliner, T.M., 2007. Phylogenetic relationships of Nembrothinae (Mollusca: Doridacea: Polyceridae) inferred from morphology and mitochondrial DNA. *Molecular phylogenetics and evolution*, 43(3), pp.726–42.

Pollock, L.J. et al., 2014. Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM) J. McPherson, ed. *Methods in Ecology and Evolution*, 5(5), pp.397–406.

Pontarotti, P., 2014. Evolutionary biology: Genome evolution, speciation, coevolution and origin of life. In P. Pontarotti, ed. *Evolutionary Biology: Genome Evolution, Speciation, Coevolution and Origin of Life*. Cham: Springer International Publishing, pp. 1–398.

Posada, D. & Crandall, K., 2001. Intraspecific gene genealogies: trees grafting into networks. *Trends in ecology & evolution*, 16(1), pp.37–45.

Provan, J. & Bennett, K.D., 2008. Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution*, 23(10), pp.564–571.

Pujolar, J.M., Roldán, M.I. & Pla, C., 2002. A genetic assessment of the population structure of swordfish (*Xiphias gladius*) in the Mediterranean Sea. *Journal of Experimental Marine Biology and Ecology*, 276(1-2), pp.19–29.

de Queiroz, K., 2005. Ernst Mayr and the modern concept of species. In *Proceedings of the National Academy of Sciences*

of the United States of America. pp. 6600–6607.

Quinteiro, J., Rodríguez-Castro, J. & Rey-Méndez, M., 2007. Population genetic structure of the stalked barnacle *Pollicipes pollicipes* (Gmelin, 1789) in the northeastern Atlantic: Influence of coastal currents and mesoscale hydrographic structures. *Marine Biology*, 153(1), pp.47–60.

Rambaut, A., 2009. FigTree v1.3.1: Tree figure drawing tool.

Ramos-Onsins, S.E. & Rozas, J., 2002. Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, 19(12), pp.2092–2100.

Rastorgueff, P.A. et al., 2014. Cryptic habitats and cryptic diversity: Unexpected patterns of connectivity and phylogeographical breaks in a Mediterranean endemic marine cave mysid. *Molecular Ecology*, 23(11), pp.2825–2843.

Record, T.P. & Change, G.S., 2005. The Phanerozoic Record of Global Sea-Level Change. *Science*, 310(5752), pp.1293–1298.

Reid, D.G., Dyal, P. & Williams, S.T., 2010. Global diversification of mangrove fauna: a molecular phylogeny of *Littoraria* (Gastropoda: Littorinidae). *Molecular Phylogenetics and Evolution*, 55(1), pp.185–201.

Relvas, P. et al., 2007. Physical oceanography of the western Iberia ecosystem: Latest views and challenges. *Progress in Oceanography*, 74(2-3), pp.149–173.

Ribeiro, P.A. et al., 2010. Recent changes in the distribution of a marine gastropod, *Patella rustica*, across the Iberian Atlantic coast did not result in diminished genetic diversity or increased connectivity. *Journal of Biogeography*, 37(9), pp.1782–1796.

Ricqlès, A. de, 2000. Palaeohistology of the bones of pterosaurs (Reptilia: Archosauria): anatomy, ontogeny, and biomechanical implications. *Zoological Journal of the Linnean Society*, 129, pp.349–385.

Riesgo, A. et al., 2012. Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. *Frontiers in Zoology*, 9(1), pp.1–33.

Riginos, C. & Nachman, M.W., 2001. Population subdivision in marine environments: The contributions of biogeography, geographical distance and discontinuous habitat to genetic differentiation in a blennioid fish, *Axoclinus nigricaudus*. *Molecular Ecology*, 10(6), pp.1439–1453.

Rizzuti, A.M., Cohen, A.D. & Stack, E.M., 2004. Using hydraulic conductivity and micropetrography to assess water flow through peat-containing wetlands. *International Journal of Coal Geology*, 60(1), pp.1–16.

Robertson, G. et al., 2010. De novo assembly and analysis of RNA-seq data. *Nature Methods*, 7(11), pp.909–912.

Rocha, L.A. et al., 2005. Ecological speciation in tropical reef fishes. *Proceedings of the Royal Society Biological sciences*, 272(1563), p.573 579.

Rocha, L.A. et al., 2013. Massively parallel DNA sequencing: the new frontier in biogeography. *Frontiers of Biogeography*, 26(1), pp.217–220.

Rodríguez-Cabello, C. et al., 2009. Does Le Danois Bank (El Cachucho) influence albacore catches in the Cantabrian Sea? *Continental Shelf Research*, 29(8), pp.1205–1212.

Rögl, F., 1999. Mediterranean and Paratethys. Facts and hypotheses of an Oligocene to Miocene paleogeography (short overview). *Geologica Carpathica*, 50(4), pp.339–349.

Ronquist, F. et al., 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), pp.539–542.

Rothschild, L.J. & Mancinelli, R.L., 2001. Life in extreme environments. *Nature*, 409(6823), pp.1092–101.

Roura, Á., 2013. *Ecología de paralarvas planctónicas de cefalópodos en áreas de afloramiento costero*. Ph.D. Dissertation. Universidade de Vigo.

Rowden, A.A. et al., 2010. Paradigms in seamount ecology: Fact, fiction and future. *Marine Ecology*, 31(1), pp.226–241.

Rozas, J. et al., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, 19(18), pp.2496–2497.

Rudman, W.B., 1991. Purpose in pattern: The evolution of colour in chromodorid nudibranchs. *Journal of Molluscan Studies*, 57, pp.5–21.

Rudman, W.B., 1984. The Chromodorididae (Opisthobranchia: Mollusca) of the Indo-West Pacific: a review of the genera. *Zoological Journal of the Linnean Society*, 81(1987), pp.115–273.

Rudman, W.B. & Bergquist, P.R., 2007. A review of feeding specificity in the sponge-feeding Chromodorididae (Nudibranchia: Mollusca). *Molluscan Research*, 27(2), pp.60–88.

Rudman, W.B. & Willan, R.C., 1998. Opisthobranchia. The southern synthesis. Fauna of Australia. In *Mollusca: The Southern synthesis. Fauna of Australia*. Melbourne, Australia: In Beesly P.L., Ross G.J.B., Wells A., pp. 915–1035.

Rumpho, M.E. et al., 2001. Mollusc/algal chloroplast symbiosis: how can isolated chloroplasts continue to function for months in the cytosol of a sea slug in the absence of an algal nucleus? *Zoology (Jena, Germany)*, 104(3-4), pp.303–312.

Sadamoto, H. et al., 2012. De novo sequencing and transcriptome analysis of the central nervous system of Mollusc Lymnaea stagnalis by deep RNA sequencing. *PLoS ONE*, 7(8), pp.1–13.

Samadi, S. et al., 2006. Seamount endemism questioned by the geographic distribution and population genetic structure of marine invertebrates. *Marine Biology*, 149(6), pp.1463–1475.

Sánchez-Tocino, L., 2003. *Tocino 2003 Aspectos taxonómicos y biológicos de los doridoidea (mollusca; nudibranchia) del litoral granadino*.

Sánchez-Tocino, L., Tierno de Figueroa, J.M. & Cervera, J.L., 2014. Ontogenetic changes in the spicule formation and their possible role in chromodorid opisthobranchs (Mollusca, Chromodorididae). *Marine Biology Research*, 10(4), pp.357–373.

Sanford, E. et al., 2006. Larval tolerance, gene flow, and the northern geographic range limit of fiddler crabs. *Ecology*, 87(11), pp.2882–2894.

Sanford, E. & Kelly, M.W., 2011. Local adaptation in marine invertebrates. *Annual Review of Marine Science*, 3, pp.509–535.

Santos, R.S. et al., 1995. Marine research, resources and conservation in the Azores. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 5(4), pp.311–354.

Sá-Pinto, A. et al., 2012. Barriers to Gene Flow in the Marine Environment: Insights from Two Common Intertidal Limpet Species of the Atlantic and Mediterranean. *PLoS One*, 7(12), p.e50330.

Sá-Pinto, I.A., 2008. *Evolution and Speciation in the Marine Realm : the genus Patella as a case study Evolution and speciation in the marine realm : the genus Patella as a case study* . Ph.D. Dissertation. Universidade do Porto.

Sasuga, J. et al., 1999. Gene Contents and Organization of a Mitochondrial DNA Segment of the Squid *Loligo bleekeri*. *Journal of Molecular Evolution*, 48(6), pp.692–702.

Schmieder, R. & Edwards, R., 2011. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One*, 6(3), p.e17288.

Schmittner, a. et al., 2004. Global impact of the Panamanian seaway closure. *Eos, Transactions American Geophysical Union*, 85(49), p.526.

Schrödl, M. et al., 2011. Bye bye 'opisthobranchia'! a review on the contribution of mesopsammic sea slugs to euthyneuran systematics. *Thalassas*, 27(2), pp.101–112.

Schultheis, A., Davis, N. & Page, J., 2014. Comparative transcriptomics allows for rapid development of population-level nuclear markers in *Hesperoperla pacifica* (Plecoptera: Perlidae). *Freshwater Science*, 33(1), pp.364–373.

Schunter, C. et al., 2011. Matching genetics with oceanography: Directional gene flow in a Mediterranean fish species. *Molecular Ecology*, 20(24), pp.5167–5181.

Schwarz, D. et al., 2009. Sympatric ecological speciation meets pyrosequencing: sampling the transcriptome of the apple maggot *Rhagoletis pomonella*. *BMC Genomics*, 10(1), p.633.

Senatore, A., Edirisinghe, N. & Katz, P.S., 2015. Deep mRNA sequencing of the *Tritonia diomedea* brain transcriptome provides access to gene homologues for neuronal excitability, synaptic transmission and peptidergic signalling. *PloS one*, 10(2), p.e0118321.

Serb, J.M. & Lydeard, C., 2003. Complete mtDNA sequence of the North American freshwater mussel, *Lampsilis ornata* (Unionidae): an examination of the evolution and phylogenetic utility of mitochondrial genome organization in Bivalvia (Mollusca). *Molecular Biology and Evolution*, 20(11), pp.1854–66.

Sevigny, J.L. et al., 2015. The Mitochondrial Genomes of the Nudibranch Mollusks, *Melibe leonina* and *Tritonia diomedea*, and Their Impact on Gastropod Phylogeny. *Plos One*, 10(5), p.e0127519.

Shackleton, N.J. et al., 1984. Oxygen isotope calibration of the onset of ice-rafting and history of glaciation in the North Atlantic region. *Nature*, 307(5952), pp.620–623.

Sherman, C.D.H., Hunt, A. & Ayre, D.J., 2008. Is life history a barrier to dispersal? Contrasting patterns of genetic differentiation along an oceanographically complex coast. *Biological Journal of the Linnean Society*, 95, pp.106–116.

Shipp, A.E., 2010. Book Review: Born Digital: Understanding the First Generation of Digital Natives. In *Professional School Counseling*. pp. 270–271.

Shulman, M. & Bermingham, E., 1996. Early life histories, ocean currents, and the population genetics of Caribbean reef fishes. *Oceanographic Literature Review*, 49(5), pp.897–910.

Simison, W.B., Lindberg, D.R. & Boore, J.L., 2006. Rolling circle amplification of metazoan mitochondrial genomes. *Molecular Phylogenetics and Evolution*, 39(2), pp.562–567.

Simon, S. & Hadrys, H., 2013. A comparative analysis of complete mitochondrial genomes among Hexapoda. *Molecular Phylogenetics and Evolution*, 69(2), pp.393–403.

Simon-Bouhet, B., Garcia-Meunier, P. & Viard, F., 2006. Multiple introductions promote range expansion of the mollusc *Cyclope neritea* (Nassariidae) in France: evidence from mitochondrial sequence data. *Molecular Ecology*, 15(6), pp.1699–711.

Simpson, J.T. et al., 2009. ABySS: A parallel assembler for short read sequence data. *Genome Research*, 19(6), pp.1117–1123.

Smith, S. a et al., 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature*, 480(7377), pp.364–7.

Solé-Cava, A.M. et al., 1991. Genetic evidence for cryptic speciation in allopatric populations of two cosmopolitan species of the calcareous sponge genus *Clathrina*. *Marine Biology*, 111(3), pp.381–386.

Sotka, E.E., 2005. Local adaptation in host use among marine invertebrates. *Ecology Letters*, 8, pp.448–459.

Sotka, E.E. et al., 2004. Strong genetic clines and geographical variation in gene flow in the rocky intertidal barnacle *Balanus glandula*. *Molecular Ecology*, 13, pp.2143–2156.

Stanhope, M.J., Connelly, M.M. & Hartwick, B., 1992. Evolution of a crustacean chemical communication channel: Behavioral and ecological genetic evidence for a habitat-modified, race-specific pheromone. *Journal of Chemical Ecology*, 18(10), pp.1871–1887.

Steininger, F.F. & Rögl, F., 1984. Paleogeography and palinspastic reconstruction of the Neogene of the Mediterranean and Paratethys. *Geological Society London, Special Publications*, 17, pp.659–668.

Stöger, I. & Schrödl, M., 2013. Mitogenomics does not resolve deep molluscan relationships (yet?). *Molecular Phylogenetics*

and Evolution, 69(2), pp.376–392.

Sturmbauer, C., Levinton, J.S. & Christy, J., 1996. Molecular phylogeny analysis of fiddler crabs: test of the hypothesis of increasing behavioral complexity in evolution. *Proceedings of the National Academy of Sciences*, 93(20), pp.10855–10857.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, pp.585–595.

Tamura, K. et al., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12), pp.2725–2729.

Tatusov, R.L. et al., 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, 4(1), p.1.

Taylor, M.S. & Hellberg, M.E., 2003. Genetic evidence for local retention of pelagic larvae in a Caribbean reef fish. *Science (New York, N.Y.)*, 299, pp.107–109.

Terrett, J. a, Miles, S. & Thomas, R.H., 1996. Complete DNA sequence of the mitochondrial genome of *Cepaea nemoralis* (Gastropoda: Pulmonata). *Journal of Molecular Evolution*, 42(2), pp.160–8.

Teske, P.R. et al., 2007. Implications of life history for genetic structure and migration rates of southern African coastal invertebrates: Planktonic, abbreviated and direct development. *Marine Biology*, 152, pp.697–711.

Thatje, S., Hillenbrand, C.-D. & Larter, R., 2005. On the origin of Antarctic marine benthic community structure. *Trends in Ecology & Evolution*, 20, pp.534–540.

Thiede, J., 1978. A Glacial Mediterranean. *Nature*, 276, pp.680–683.

Thiel, M. & Gutow, L., 2005. The ecology of rafting in the marine environment. I. The floating substrata. *Oceanography and Marine Biology,* 42, pp.181-264.

Thollesson, M., 2000. Increasing fidelity in parsimony analysis of dorid nudibranchs by differential weighting, or a tale of two genes. *Molecular Phylogenetics and Evolution*, 16(2), pp.161–72.

Thomas, R.C. et al., 2014. Hidden diversity in sardines: Genetic and morphological evidence for cryptic species in the goldstripe sardinella, *Sardinella gibbosa* (Bleeker, 1849). *PLoS One*, 9(1).

Thomson, R.C., Wang, I.J. & Johnson, J.R., 2010. Genome-enabled development of DNA markers for ecology, evolution and conservation. *Molecular Ecology*, 19(11), pp.2184–95.

Thorpe, J.P., Solé-Cava, A.M. & Watts, P.C., 2000. Exploited marine invertebrates: Genetics and fisheries. *Hydrobiologia*, 420, pp.165–184.

Todd, C.D., Lambert, W.J. & Thorpe, J.P., 1998. The genetic structure of intertidal populations of two species of nudibranch molluscs with planktotrophic and pelagic lecithotrophic larval stages: Are pelagic larvae 'for' dispersal? *Journal of Experimental Marine Biology and Ecology*, 228, pp.1–28.

Tosar, J.P. et al., 2014. Mining of public sequencing databases supports a non-dietary origin for putative foreign miRNAs: underestimated effects of contamination in NGS. *RNA*, 20(6), pp.754–7.

Trontelj, P. et al., 2009. A molecular test for cryptic diversity in ground water: How large are the ranges of macro-stygobionts? *Freshwater Biology*, 54(4), pp.727–744.

Tunnicliffe, V. et al., 2010. Flatfish at seamount hydrothermal vents show strong genetic divergence between volcanic arcs. *Marine Ecology*, 31, pp.158–167.

Turner, L.M. & Wilson, N.G., 2008. Polyphyly across oceans: A molecular phylogeny of the Chromodorididae (Mollusca, Nudibranchia). *Zoologica Scripta*, 37(1), pp.23–42.

Valdés, Á., 2001. Depth-related adaptations, speciation processes and evolution of color in the genus *Phyllidiopsis* (Mollusca: Nudibranchia). *Marine Biology*, 139(3), pp.485–496.

Valdés, Á. et al., 2011. Multi-Data Set Revision of Two Uncommon Species of Chromodorididae (Nudibranchia) from the Gulf of Mexico. *American Malacological Bulletin*, 29(1-2), pp.51–62.

Valdés, Á. & Gosliner, T., 1999. Reassessment of the systematic status of Miamira Bergh, 1875 and Orodoris Bergh, 1875 (Nudibranchia: Chromodorididae) in light of phylogenetic analysis. *Journal of molluscan studies*, 1875, pp.33–45.

Valdés, A.&, 2006. Phylogeny and Biogeography of the Atlantic and Eastern Pacific *Hypselodoris* Stimpson, 1855 (Nudibranchia, Chromodorididae) With the Description of a New Species From the Caribbean Sea. *Journal of Molluscan Studies*, 72(2), pp.189–198.

de Vargas, C. et al., 1999. Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proceedings of the National Academy of Sciences of the United States of America*, 96(6), pp.2864–2868.

Villamor, A., Costantini, F. & Abbiati, M., 2014. Genetic structuring across marine biogeographic boundaries in rocky shore invertebrates. *PLoS One*, 9(7), p.e101135.

Viñas, J., Alvarado Bremer, J. & Pla, C., 2004. Phylogeography of the Atlantic bonito (*Sarda sarda*) in the northern Mediterranean: The combined effects of historical vicariance, population expansion, secondary invasion, and isolation by distance. *Molecular Phylogenetics and Evolution*, 33, pp.32–42.

Vinther, J. et al., 2012. A molecular palaeobiological hypothesis for the origin of aplacophoran molluscs and their derivation from chiton-like ancestors. *Proceedings of the Royal Society B: Biological Sciences*, 279(1732), pp.1259–1268.

Vrijenhoek, R.C., 2009. Cryptic species, phenotypic plasticity, and complex life histories: Assessing deep-sea faunal diversity with molecular markers. *Deep Sea Research Part II: Topical Studies in Oceanography*, 56(19-20), pp.1713–1723.

Wägele, H. et al., 2014. Flashback and foreshadowing - A review of the taxon Opisthobranchia. *Organisms Diversity and Evolution*, 14(1), pp.133–149.

Wägele, H., 2004. Potential key characters in Opisthobranchia (Gastropoda, Mollusca) enhancing adaptive radiation. *Organisms Diversity and Evolution*, 4(3), pp.175–188.

Wägele, H., Ballesteros, M. & Avila, C., 2006. Defensive glandular structures in opisthobranch molluscs - From histology to ecology. In *Oceanography and Marine Biology - an Annual Review, Vol 44*. pp. 197–276.

Wägele, H. & Klussmann-Kolb, A., 2005. Opisthobranchia (Mollusca, Gastropoda) - more than just slimy slugs. Shell reduction and its implications on defence and foraging. *Frontiers in zoology*, 2(1), p.3.

Wagner, D., Kahng, S.E. & Toonen, R.J., 2009. Observations on the life history and feeding ecology of a specialized nudibranch predator (Phyllodesmium poindimiei), with implications for biocontrol of an invasive octocoral (Carijoa riisei) in Hawaii. *Journal of Experimental Marine Biology and Ecology*, 372, pp.64–74.

Wall, D.P., Fraser, H.B. & Hirsh, A.E., 2003. Detecting putative orthologs. *Bioinformatics*, 19(13), pp.1710–1711.

Wangensteen, O.S. et al., 2012. Natural or Naturalized? Phylogeography Suggests That the Abundant Sea Urchin *Arbacia lixula* Is a Recent Colonizer of the Mediterranean. *PLoS One*, 7(9), p.e45067.

Weiss, J.M. et al., 2007. Temperature control of larval dispersal and the implications for marine ecology, evolution, and conservation. *Proceedings of the National Academy of Sciences*, 104(4), pp.1266–1271.

Wiafe, G. et al., 2008. Impact of climate change on long-term zooplankton biomass in the upwelling region of the Gulf of Guinea. *ICES Journal of Marine Science*, 65(3), pp.318–324.

Williams, S.T., 2007. Origins and diversification of Indo-West Pacific marine fauna: Evolutionary history and biogeography of turban shells (Gastropoda, Turbinidae). *Biological Journal of the Linnean Society*, 92(3), pp.573–592.

Williams, S.T. & Duda, T.F., 2008. Did tectonic activity stimulate Oligo-Miocene speciation in the Indo-West Pacific? *Evolution*, 62, pp.1618–1634.

Williams, S.T., Foster, P.G. & Littlewood, D.T.J., 2014. The complete mitochondrial genome of a turbinid vetigastropod from MiSeq Illumina sequencing of genomic DNA and steps towards a resolved gastropod phylogeny. *Gene*, 533(1), pp.38–47.

Williams, S.T. & Reid, D.G., 2004. Speciation and Diversity on Tropical Rocky Shores: a Global Phylogeny of Snails of the Genus *Echinolittorina*. *Evolution*, 58(10), pp.2227–2251.

Wilson, N.G., Schroedl, M. & Halanych, K.M., 2009. Ocean barriers and glaciation: Evidence for explosive radiation of mitochondrial lineages in the Antarctic sea slug *Doris kerguelenensis* (Mollusca, Nudibranchia). *Molecular Ecology*, 18(5), pp.965–984.

Wolf, M.A. et al., 2012. *Ulva* (Chlorophyta, Ulvales) Biodiversity in the North Adriatic Sea (Mediterranean, Italy): Cryptic Species and New Introductions. *Journal of Phycology*, 48(6), pp.1510–1521.

Wollscheid-Lengeling, E. et al., 2001. The phylogeny of Nudibranchia (Opisthobranchia, Gastropoda, Mollusca) reconstructed by three molecular markers. *Organisms, Diversity & Evolution*, 1, pp.241–256.

Wooster, W.S., Bakun, A. & Mclain, D.R., 1976. Seasonal Upwelling Cycle Along Eastern Boundary of North-Atlantic. *Journal of Marine Research*, 34(2), pp.131–141.

Xavier, R. et al., 2011. Phylogeography of the marine isopod *Stenosoma nadejda* (Rezig, 1989) in North African Atlantic and western Mediterranean coasts reveals complex differentiation patterns and a new species. *Biological Journal of the Linnean Society*, 104(2), pp.419–431.

Xavier, R., Murias Santos, A. & Madalena, B., 2011. MtDNA and nuclear data reveal patterns of low genetic differentiation for the isopods *Stenosoma lancifer* and *Stenosoma acuminatum*, with low dispersal ability along the northeast Atlantic coast. *Scientia Marina*, 76(1), pp.133–140.

Yamazaki, N. & Ueshima, R., 1997. Evolution of pulmonate gastropod mitochondrial genomes: Comparisons of gene organizations of. *Genetics*, 145(3), p.749.

Yang, Y. & Smith, S. a, 2013. Optimizing de novo assembly of short-read RNA-seq data for phylogenomics. *BMC genomics*, 14(1), p.328.

Yao, H. et al., 2010. Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS One*, 5(10), pp.294–299.

Zapata, F. et al., 2014. Orthogastropoda Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. *Proceedings of the Royal Society B: Biological Sciences*, 281, p.20141739.

Zaragosi, S. et al., 2001. Initiation of the European deglaciation as recorded in the northwestern Bay of Biscay slope environments (Meriadzek Terrace and Trevelyan Escarpment): A multi-proxy approach. *Earth and Planetary Science Letters*, 188(3-4), pp.493–507.

Zardoya, R. et al., 2004. Differential population structuring of two closely related fish species, the mackerel (*Scomber scombrus*) and the chub mackerel (*Scomber japonicus*), in the Mediterranean Sea. *Molecular Ecology*, 13(7), pp.1785–98.

Zealand, N., 2004. Out of Africa : The Slow Train to Australasia. *Systematic Biology*, 53(1), pp.18–24.

Zerbino, D.R. & Birney, E., 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18, pp.821–829.

Zhang, L. et al., 2012. Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. *Cell Research*, 22(1), pp.273–274.

Zhao, Q.-Y. et al., 2011. Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinformatics*, 12(14), p.S2.

Zhou, Q. et al., 2013. QC-Chain: Fast and Holistic Quality Control Method for Next-Generation Sequencing Data. *PLoS One*, 8(4), p.e60234.

Zhou, Z.C. et al., 2014. Transcriptome sequencing of sea cucumber (*Apostichopus japonicus*) and the identification of gene-associated markers. *Molecular Ecology Resources*, 14(1), pp.127–138.

Supplementary Material

Supplementary Table 1 - Detailed information on the sampling localities

Site ID	Locality	Area
1	S. Sebastián	S. Sebastián
		Guetaria
2	Ferrol	Enseada de Nande
		Rabo da Porca
3	Vigo	Limens, Vigo
		Cies, Vigo Islands
4	Sesimbra	Sesimbra
5	Alentejo	Alentejo
6	Sagres	Sagres
7	Ria Formosa	Ria Formosa
8	Cádiz	Cádiz
9	Gibraltar	Tarifa
		Punta San Garcia
10	Chafarinas	Chafarinas
11	Menorca	Menorca Island
12	Barcelona	Mataró, Barcelona
13	Girona	Roca Roja
		L'Escala
		L'Estartit
		Sá Tuna
		Caials
14	SW France	Banyuls-Sur-mer
		Cap Oulestreil
15	Italy	Tor Paterno
16		Tavolara
17		Circeo
18		Ischia
19		Capo Palinuro
20		Giglio
21	Kotor bay	Kotor bay
22	Greece	Stoupa
23		Palaikastro
24		Lesvos
25	West Morroco	Anza
		Tagahzout
26	Madeira	Madeira island
27	Cape Verde	Sal island
28	São Tomé	São Tome e Príncipe

Supplementary Table 2 – Detailed information of the individuals used in this study (ID and GPS diving coordinates for the newly collected specimens and accession numbers for those from Genbank.

Species	Individual Code	Locality	Country	GPS Coordinates		Genbank accession
F. cantabrica	Op0006	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0007	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0008	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0014	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0015	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0016	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0017	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0019	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0020	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0021	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0022	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0023	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0024	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0025	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0026	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0028	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0029	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0030	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0038	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	
	Op0039	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W	

	Op0040	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W
	Op0041	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W
	Op0042	Cies islands, Vigo	Spain	42˚93'23.99"N	8˚45'26.41"W
	Op0050	Liméns, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0051	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0052	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0053	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0054	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0055	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0057	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0060	Enseada de Nande, Ferrol	Spain	43˚27'36.28"N	8˚17'0.74"W
	Op0063	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W
	Op0064	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W
	Op0065	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W
	Op0089	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W
	Op0090	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W
	Op0091	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W
	Op0114	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	OP0115	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	Op0119	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	Op0123	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	Op0195	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0226	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0227	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0329	Alentejo	Portugal	38˚0'0"N	9˚6'3.6"W
	Op0240	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W

	Op0241	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	Op0242	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	Op0330	Guetaria, S.Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	Op0301	Tarifa, Gibraltar	Spain	36˚8'26.70"N	5˚21'12.90"W
	Op0303*	Anza	Morroco	30˚27'15.50"N	9˚39'13.04"W
	Op0331	Tagahzout	Morroco	30˚32'42.02"N	9˚42'31.09"W
F. villafranca	Op0031	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0032	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0033	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0034	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0035	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0036	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0037	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0043	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0048	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0049	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0058	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0059	Limens, Vigo	Spain	42˚15'53.60"N	8˚49'7.22"W
	Op0066	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W
	Op0067	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W
	Op0068	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W
	Op0069	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W
	Op0070	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W
	Op0071	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W
	Op0072	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W
	Op0073	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W

Op0074	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
Op0075	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
Op0076	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
Op0077	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
Op0078	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
Op0079	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
Op0080	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
Op0081	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
Op0082	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
Op0083	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
Op0084	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
Op0085	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
Op0086	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
Op0087	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
Op0088	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
Op0332	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W	AJ223266
Op0300	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W	
Op0325	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W	KJ911291
Op0299	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W	
Op0304	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W	
Op0324	Punta San Garcia	Spain	36˚6'18.68"N	5˚25'50.44"W	KJ911292
Op0326	Illes de Porros, Menorca	Spain	40˚5'12.73"N	4˚5'2.69"E	KJ911290

Op0327	Tagahzout	Morroco	30˚32'42.02"N	9˚42'31.09"W	KJ911288
Op0328	Tagahzout	Morroco	30˚32'42.02"N	9˚42'31.09"W	KJ911289
Op0126	Mataró, Barcelona	Spain	41˚32'17.33"N	2˚26'41.33"E	
Op0128	Mataró, Barcelona	Spain	41˚32'17.33"N	2˚26'41.33"E	
Op0129	Mataró, Barcelona	Spain	41˚32'17.33"N	2˚26'41.33"E	
Op0152	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
Op0153	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
Op0154	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
Op0155	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
Op0156	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
Op0157	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
Op0158	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
Op0159	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
Op0160	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
Op0196	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W	
Op0197	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W	
Op0198	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W	
Op0201	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W	
Op0202	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W	
Op0203	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W	
Op0204	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W	
Op0205	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W	
Op0231	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W	
Op0233	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W	
Op0234	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W	
Op0235	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W	
Op0236	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W	
Op0237	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W	
Op0238	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W	

	Op0247	L'Estartit, Girona	Spain	42˚3'4.62"N	3˚11'25.86"E	
	Op0323	Tor Paterno	Italy	41˚39'22.34"N	12˚24'58.09"E	
	Op0305	Banyuls-Sur-mer	France	42˚22'0.34"N	3˚7'43.94"E	
F. tricolor	Op0333	Martinhal Islands, Sagres	Portugal	37˚1'0.585"N	8˚56'26.13"W	JQ727898
	Op0061	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
	Op0092	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
	Op0093	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
	Op0094	Rabo da Porca, Ferrol	Spain	43˚27'40.50"N	8˚18'15.63"W	
	Op0104	Kotor Bay	MonteNegro	42˚23'34.93"N	18˚33'36.10"E	
	Op0117	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W	
	Op0127	Mataró, Barcelona	Spain	41˚32'17.33"N	2˚26'41.33"E	
	Op0141	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E	
	Op0142	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E	
	Op0145	L'Estartit, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E	
	Op0148	L'Estartit, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E	
	Op0149	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E	
	Op0164	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0165	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0166	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0167	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0168	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0169	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0170	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0171	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	

	Op0172	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W
	Op0173	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W
	Op0174	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W
	Op0194	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0199	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0200	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0206	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0207	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0208	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0210	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0211	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0212	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0213	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0214	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0215	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0216	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0217	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0218	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0219	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0220	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0221	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0222	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0223	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0224	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0225	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0229	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0230	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0248	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0244	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0246	L'Estartit, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E

	Op0249	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0250	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0251	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0252	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0253	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0254	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0255	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0256	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0257	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0258	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0259	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0263	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0264	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0265	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0266	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0272	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0273	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0275	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0276	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0277	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0278	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0279	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0280	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0284	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0285	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0286	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W

	Op0287	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0288	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0289	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0290	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0291	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0292	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0293	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0294	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0274	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0309	Circeo	Italy	41˚13'60"N	13˚2'59.99"E
	Op0310	Tor Paterno	Italy	41˚39'22.34"N	12˚24'58.09"E
	Op0311	Tor Paterno	Italy	41˚39'22.34"N	12˚24'58.09"E
	Op0312	Tor Paterno	Italy	41˚39'22.34"N	12˚24'58.09"E
	Op0313	Tor Paterno	Italy	41˚39'22.34"N	12˚24'58.09"E
	Op0314	Tor Paterno	Italy	41˚39'22.34"N	12˚24'58.09"E
	Op0315	Tor Paterno	Italy	41˚39'22.34"N	12˚24'58.09"E
	Op0316	Tor Paterno	Italy	41˚39'22.34"N	12˚24'58.09"E
	Op0317	Tor Paterno	Italy	41˚39'22.34"N	12˚24'58.09"E
	Op0318	Tor Paterno	Italy	41˚39'22.34"N	12˚24'58.09"E
	Op0319	Tor Paterno	Italy	41˚39'22.34"N	12˚24'58.09"E
	Op0320	Tavolara	Italy	40˚54'22.32"N	9˚42'47.88"E
	Op0110	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	Op0111	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	Op0112	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	Op0113	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	Op0116	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	Op0118	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	Op0120	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	Op0121	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	Op0122	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	Op0124	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W

	Op0125	S. Sebastián	Spain	43˚19'6"N	1˚58'52.43"W
	Op0243*	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0245*	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0297	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
	Op0298	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W
F. fontandraui	Op0130	Sa Tuna, Girona	Spain	41˚57'38.02"N	3˚13'43.36"E
	Op0134	L'Escala, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0140	Roca Roja, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0143	Roca Roja, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0144	Roca Roja, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0150	Roca Roja, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E
	Op0187	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W
	Op0188	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W
	Op0189	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W
	Op0190	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W
	Op0191	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W
	Op0193	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0228	Sesimbra	Portugal	38˚26'39.18"N	9˚6'5.36"W
	Op0307	Chafarinas	Spain	35˚14'7.48"N	3˚58'51.91"W
	Op0306	Cap Oulestreil	France	42˚30"0"N	3˚7'59.99"E
	Op0321	Ischia	Italy	40˚44'20.86"N	13˚57'3.58"E
	Op0322	Giglio	Italy	42˚22'27"N	10˚52'47"E
	Op0131*	Caials, Girona	Spain	42˚42'19.41"N	2˚56'59"E
	Op0308	Tarifa, Gibraltar	Spain	36˚8'26.70"N	5˚21'12.90"W
F. bilineata	Op0161	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W
	Op0162	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W
	Op0163	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W

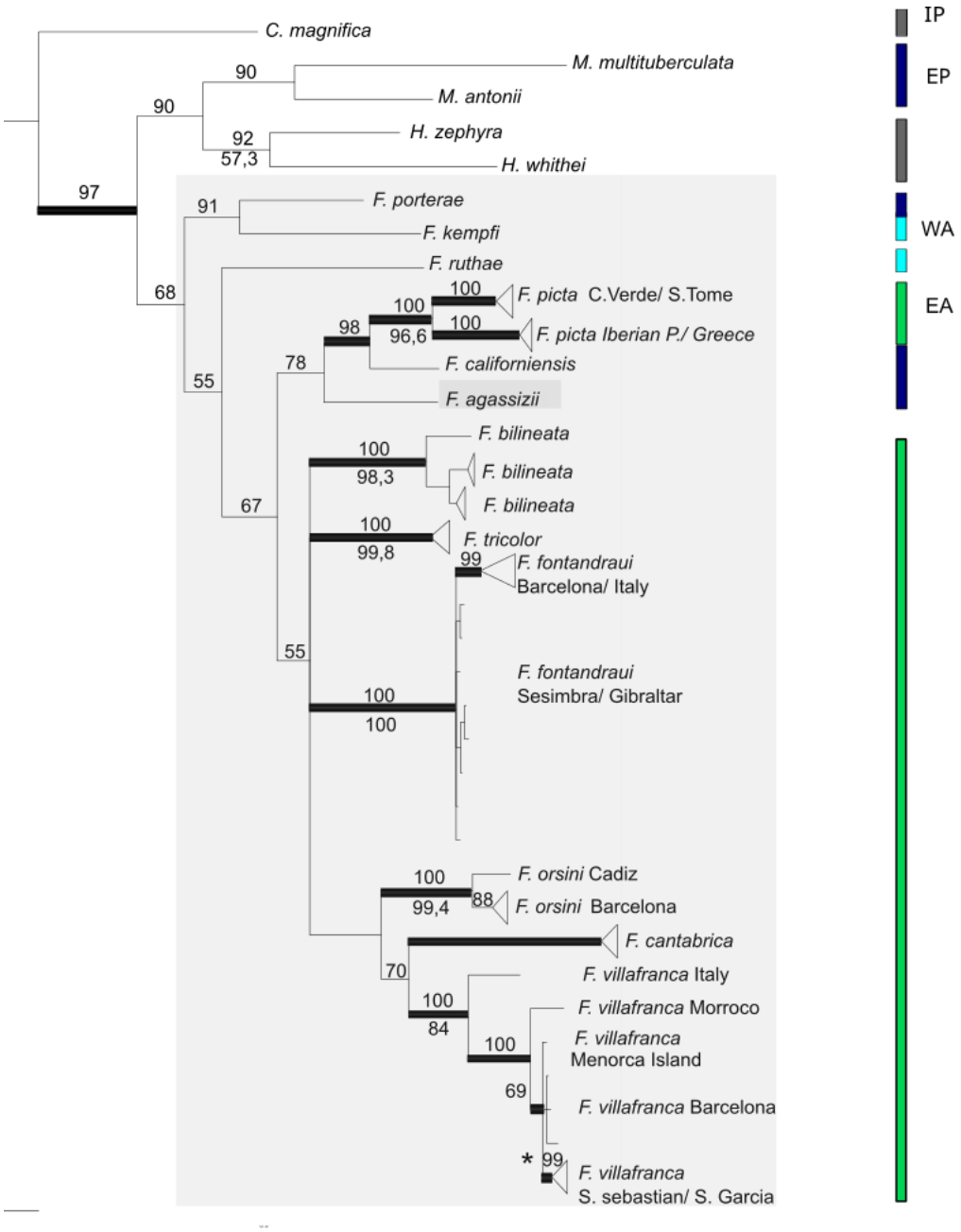
	Op0175	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0176	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0177	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0178	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0179	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0180	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0181	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0182	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0183	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0185	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0281	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W	
	Op0282	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W	
	Op0283	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W	
	Op0285	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W	
	Op0296	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W	
	Op0334	Madeira	Portugal	32˚45'38.54"N	16˚57'34.1"W	
	Op0302	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W	
F. orsinii	Op0337	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W	
	Op0136	Roca Roja, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E	
	Op0132	Caials, Girona	Spain	42˚42'19.41"N	2˚56'59"E	
	Op0133	Caials, Girona	Spain	42˚42'19.41"N	2˚56'59"E	
	Op0135	Roca Roja, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E	
	Op0138	Roca Roja, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E	
	Op0139	Roca Roja, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E	
	Op0146	Roca Roja, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E	
	Op0147	Roca Roja, Girona	Spain	42˚7'11.46"N	3˚7'21.97"E	
F. picta	Op0335	Sao Tome	São Tomé e Príncepe	0˚15'36.92"N	6˚37'15.81"W	HM162685
	Op0095	Sal	Cape Verde	16˚55'35.81"N	22˚55'46.95"W	

	Op0096	Sal	Cape Verde	16˚55'35.81"N	22˚55'46.95"W	
	Op0097	Sal	Cape Verde	16˚55'35.81"N	22˚55'46.95"W	
	Op0098	Sal	Cape Verde	16˚55'35.81"N	22˚55'46.95"W	
	Op0099	Sal	Cape Verde	16˚55'35.81"N	22˚55'46.95"W	
	Op0105	Miki Xaramida, Lesvos	Greece	39˚20'18.76"N	26˚8'21.14"E	
	Op0106	Stoupa	Greece	36˚50'40.92"N	22˚15'35.13"E	
	Op0107	Palaikastro	Greece	35˚11'52.56"N	26˚15'15.44"E	
	Op0108	Kotor bay	Montenegro	42˚23'34.93"N	18˚33'36.10"E	
	Op0192	Ria Formosa	Portugal	36˚58'8.74"N	7˚52'10.01"W	
	Op0267	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W	
	Op0270	Cádiz	Spain	36˚31'37.42"N	6˚17'18.94"W	
	LN715204	Capo Palinuro	Italy			LN715204
H. picta elegans*	AF249787	Vigo	Spain			AF249787
R. europea*	AY083457	-	-			AY083457
F. kempfi	EF535121	Florida Keys	USA			EF535139
F. porterae	EF535139	Palos Verdes	California, USA			EF535139
F. ruthae	EU982747	Bocas del Toro	Panama			EU982747
	JQ727883					JQ727883
F. agassizii	JQ727884	Baja California	Mexico			JQ727884
	CH019	Marina de la Paz	Bolivia			CH019
F. californiensis	EU98744	Punta Carbon	Costa Rica			EU982744
	JQ727902		Madagascar			JQ727902
H. whitei	JQ727903	Queensland	Australia			JQ727903
H. zephyra	EF535129	Queensland	Australia			EF535129

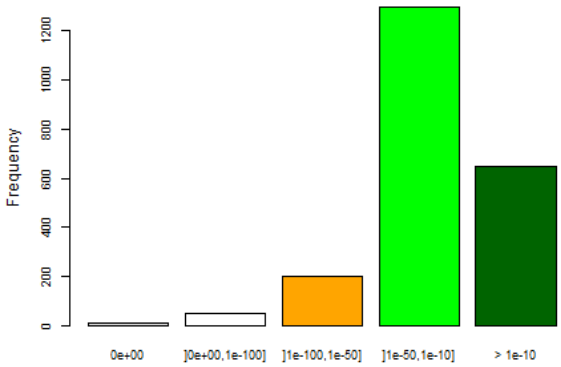
	JQ727904	Bali	Indonesia		JQ727904
	JQ727905	Iles de Radama,	Madagascar		JQ727905
M. multituberculata	JQ727907	Batangas, Twin Rocks	Philippines		JQ727906
M. antonii	EU982748	Playa Real	Costa Rica		EU982748

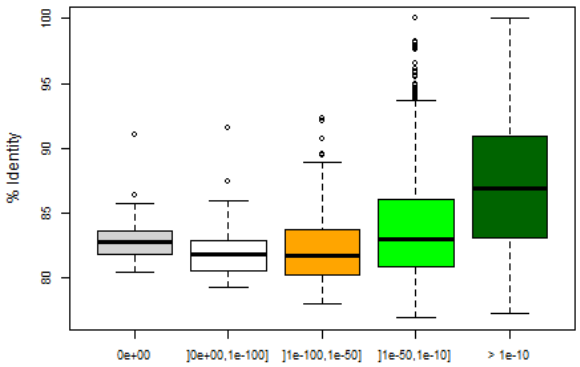
Letters above the Individual code (a-e), indicate individuals excluded for the "species-tree" analyses: a) the individual was gathered in Anza (Morocco) as a *F. bilineata*, however mitochondrial DNA recognizes it as *F. cantabrica*; b) first identified as *F. fontandraui,* these specimens fall in the *F. tricolor* clade (digital records are not clear for any of the species); c) First identified as *F. tricolor*, it falls in *F. fontandraui* clade (and once more picture is lacks important morphologic features characterizing both species); d) originally from genebank, the specimen classified as *F. picta elegans* detains the mtDNA of a *F.villafranca* from Vigo's (no photo record is available); e) outgroups included in gene-tree inferences.

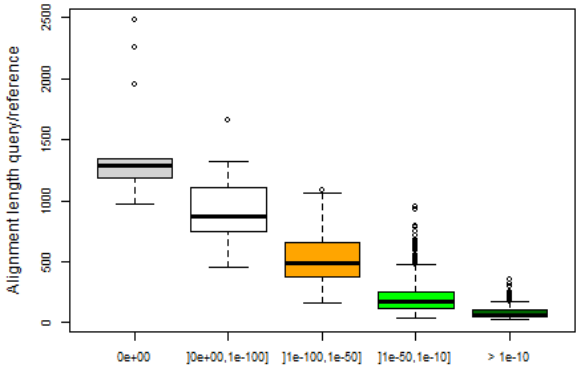
Supplementary Figure 1 Maximum likelihood phylogeny of mtDNA COI partial sequence. It includes seven species of the Eastern Atlantic and Mediterranean Sea, and Genbank sequences from the Eastern Pacific and Western Atlantic *Felimare* species. Outgroup species include two *Hypselodoris* from Indo-Pacific Ocean (*H. zephyra and H. whitei*), two *Mexichromis,* from Eastern Pacific (*M. antonii*) and Indo-Pacific (M. *multituberculata*), and the more distant *Roboastra europea*, from Eastern Atlantic. Bootstrap values above 50% (from 1000) are indicated. Branches with support equal or greater than 95% are bolded.

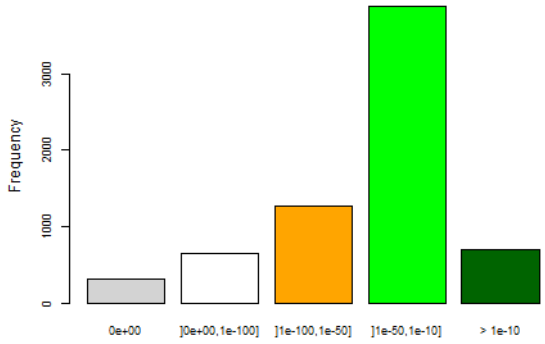


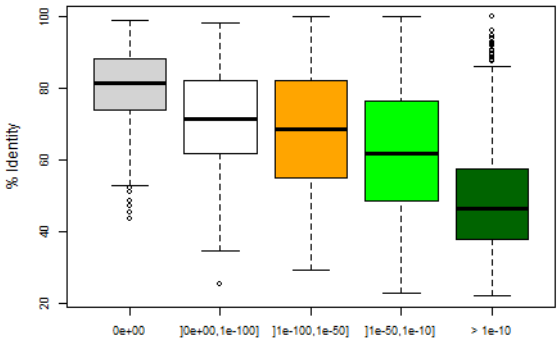
Supplementary Figure 2 – Consensus phylogeny resulting from Bayesian and ML analyses of mtDNA COI partial sequence. Analysis includes 7 species of the Eastern Atlantic and Mediterranean Sea, and Genebank sequences from the Eastern Pacific and Western Atlantic oceans of the genus Felimare. Other species include two *Hypselodoris* from Indo-Pacific Ocean (*H. zephyra and H. Whitei*), two Mexichromis from Eastern Pacific (*M. antonii*) and Indo-Pacific (M. *multituberculata*), one Chromodrid from Indo-Pacific (*Chromodoris magnifica*. The tree has been rooted with *Roboastra europea* (not visivel in the image), from Eastern Atlantic. This is a majority-rule of 11 Million generation (MCMC) with 2 runs. Posterior probabilities (above) and bootstrap values above 50, in percentage, from 1000 bootstrap of ML tree (below) are indicated as the support for each node. Branches for probabilities equal or greater than 95% are bolded

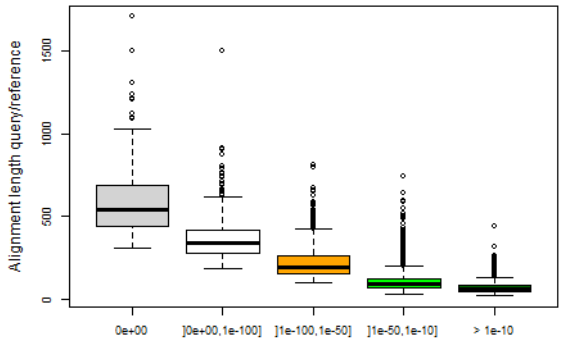

Frequency
0
200
400
600
800
1000
1200
0e+00
]0e+00,1e-100]
]1e-100,1e-50]
]1e-50,1e-10]
> 1e-10


% Identity
80
85
90
95
100
0e+00
]0e+00,1e-100]
]1e-100,1e-50]
]1e-50,1e-10]
> 1e-10

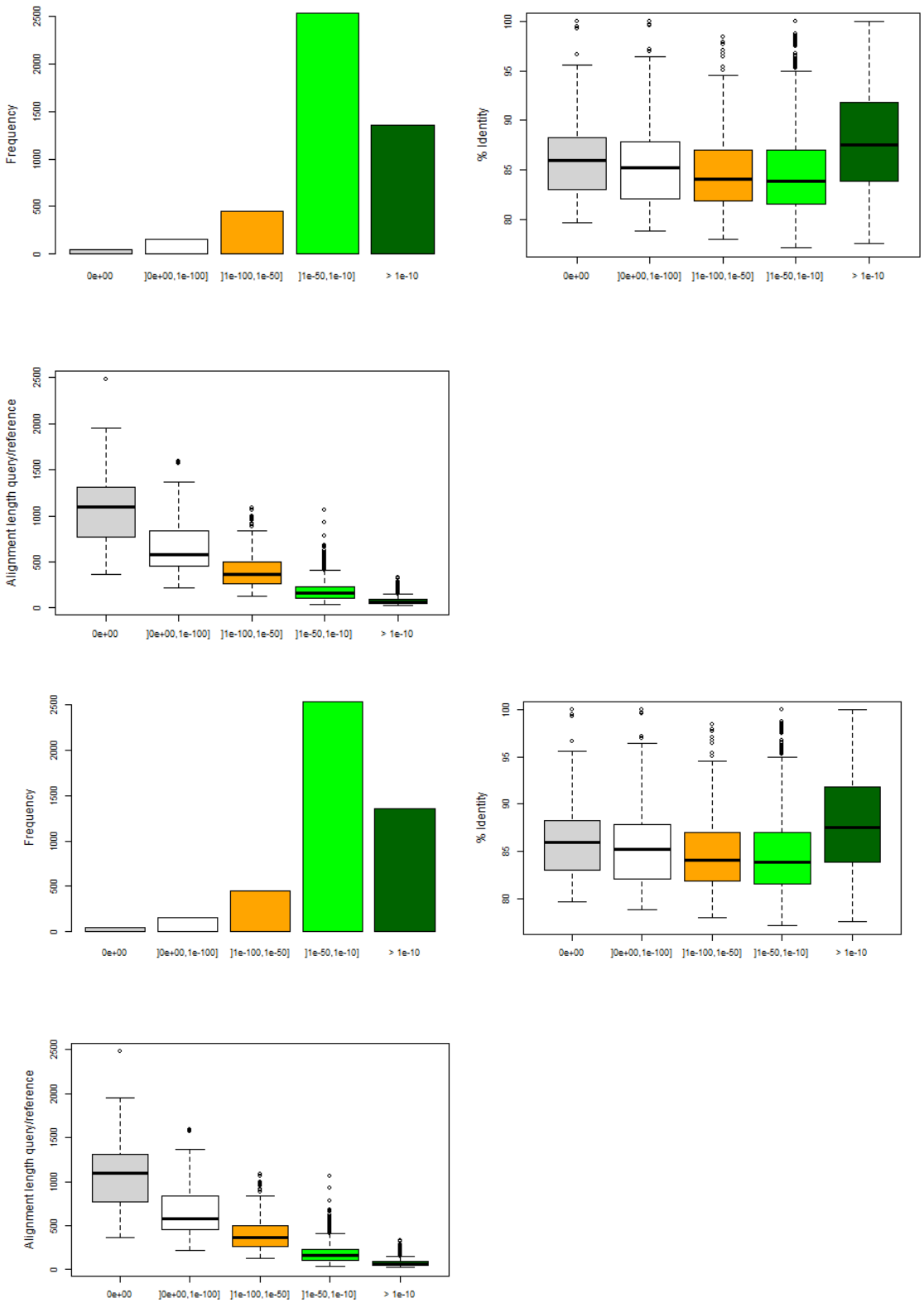

Alignment length query/reference
0
500
1000
1500
2000
2500
0e+00
]0e+00,1e-100]
]1e-100,1e-50]
]1e-50,1e-10]
> 1e-10


Frequency
0
1000
2000
3000
0e+00
]0e+00,1e-100]
]1e-100,1e-50]
]1e-50,1e-10]
> 1e-10

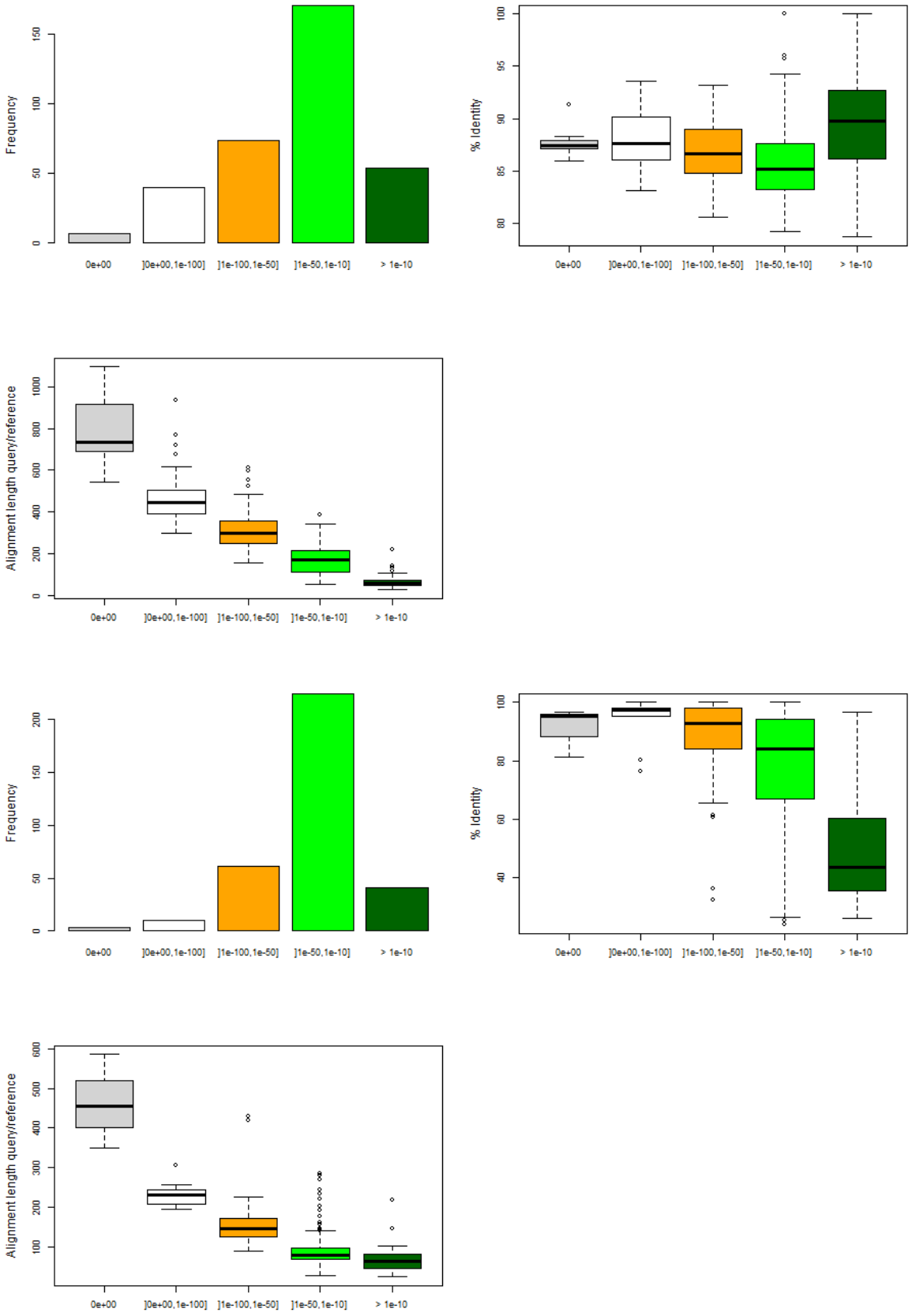

% Identity
20
40
60
80
100
0e+00
]0e+00,1e-100]
]1e-100,1e-50]
]1e-50,1e-10]
> 1e-10


Alignment length query/reference
0
500
1000
1500
0e+00
]0e+00,1e-100]
]1e-100,1e-50]
]1e-50,1e-10]
> 1e-10

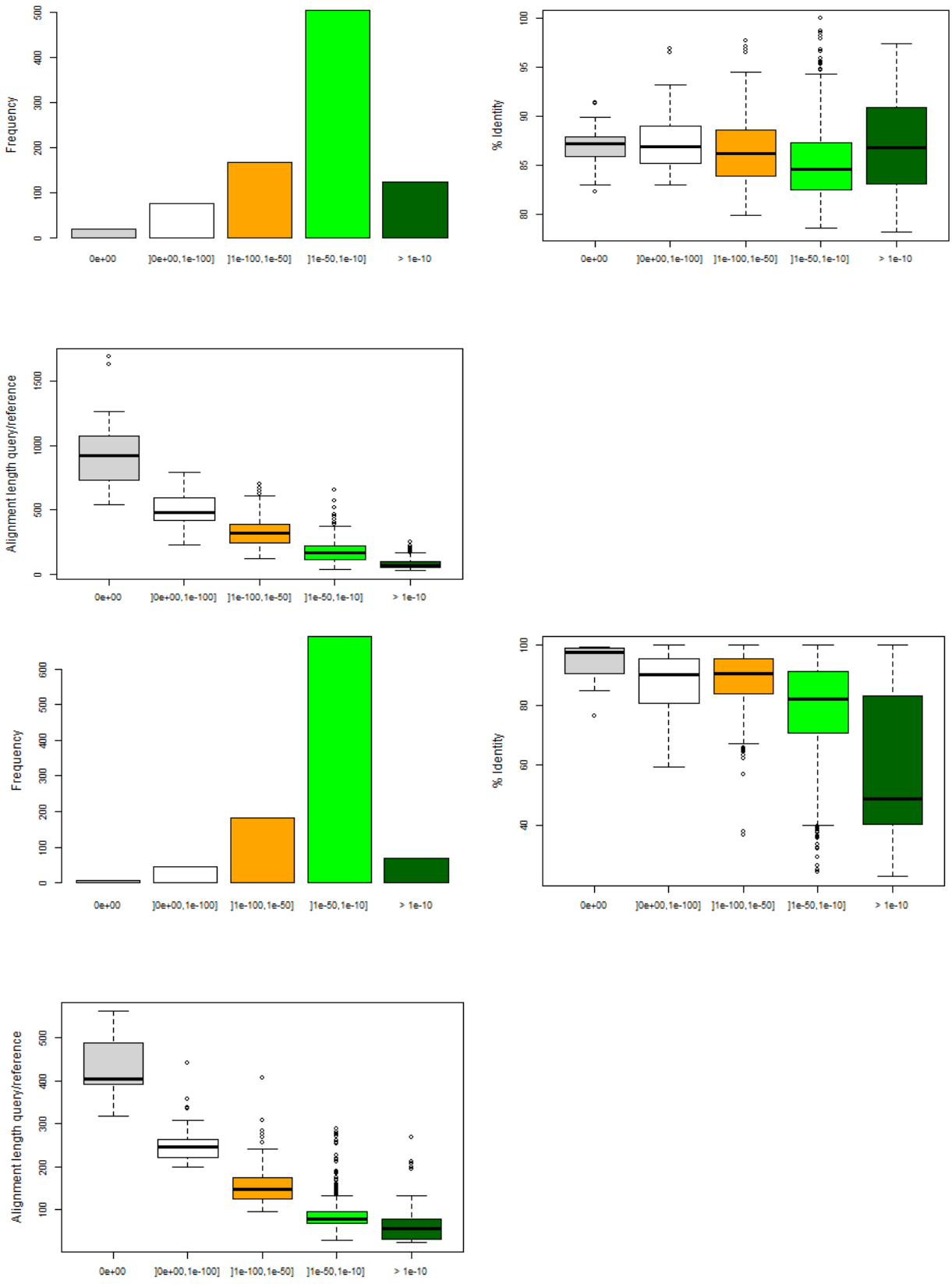
Supplementary Figure 3 Plots of frequency, % identity and alignment length per class of e-value of *F. cantabrica* against *A. californica* contigs with only the best nucleotide (above) and protein (below) hits considered.



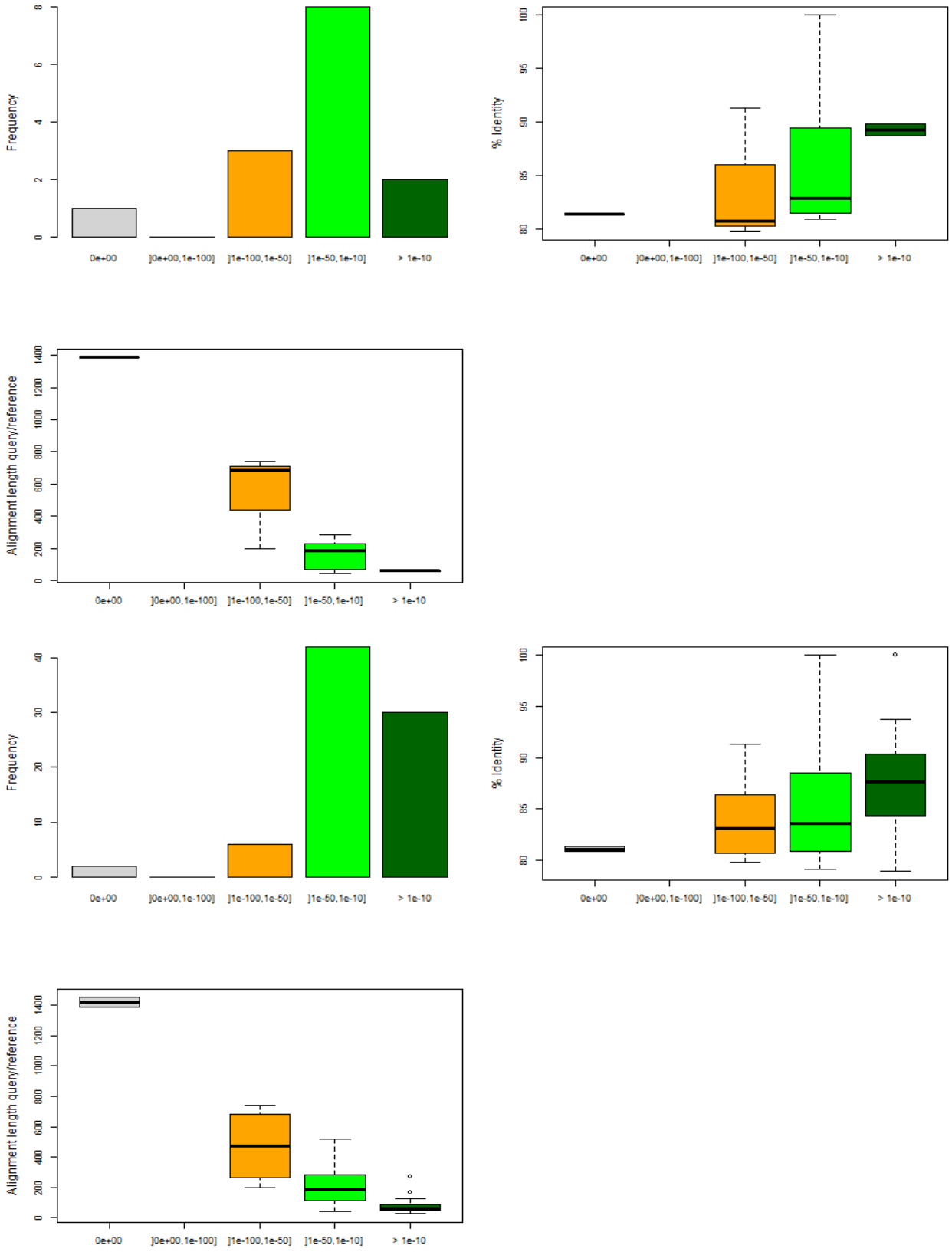
Supplementary Figure 4 Plots of frequency, % identity and alignment length per class of e-value of *F. villafranca* against *A. californica* contigs with only the best nucleotide (above) and protein (below) hits considered.



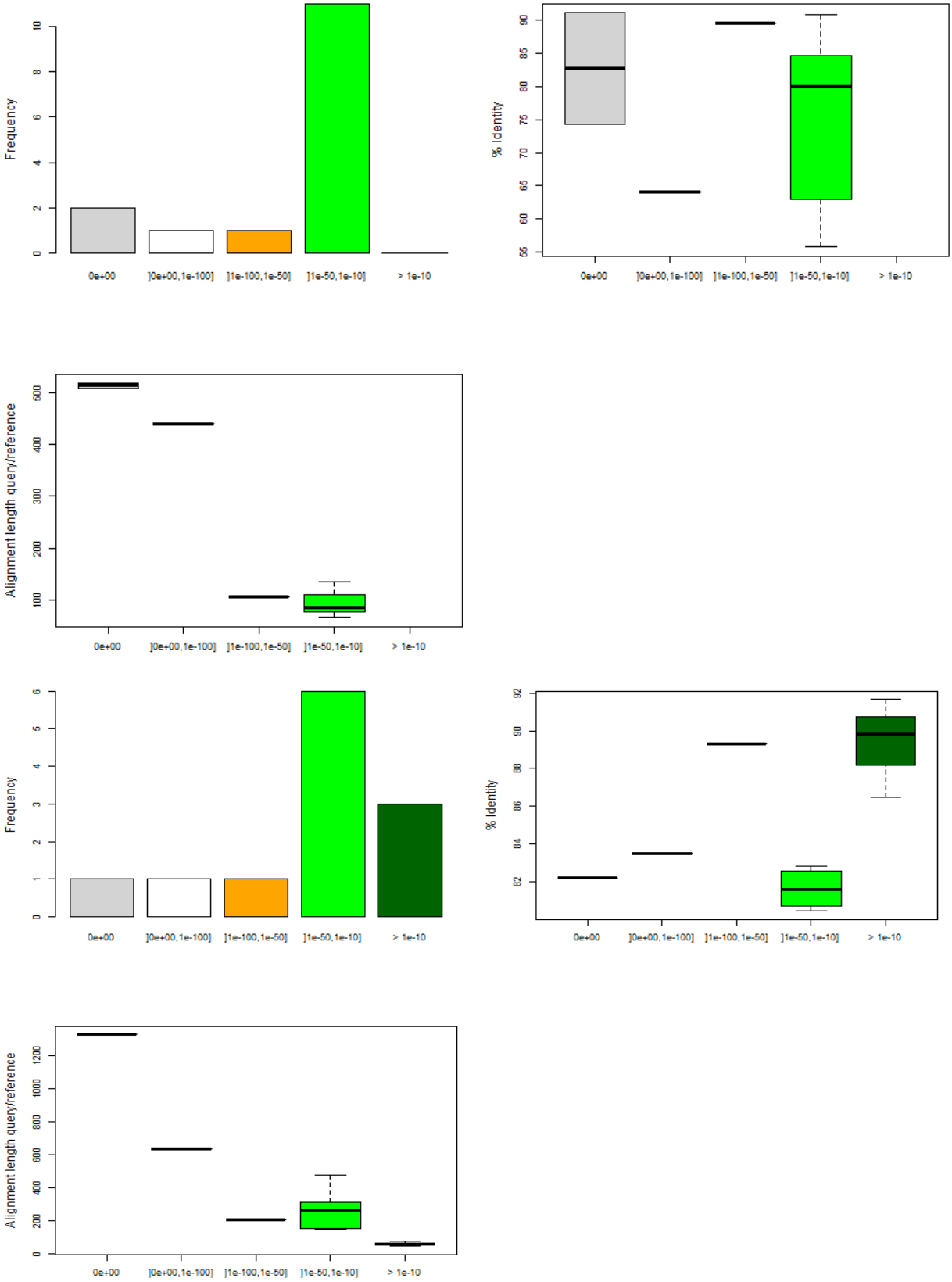
Supplementary Figure 5 Plots of frequency, % identity and alignment length per class of e-value of *F. cantabrica* against *X. silurana* contigs with only the best nucleotide (above) and protein (below) hits considered.



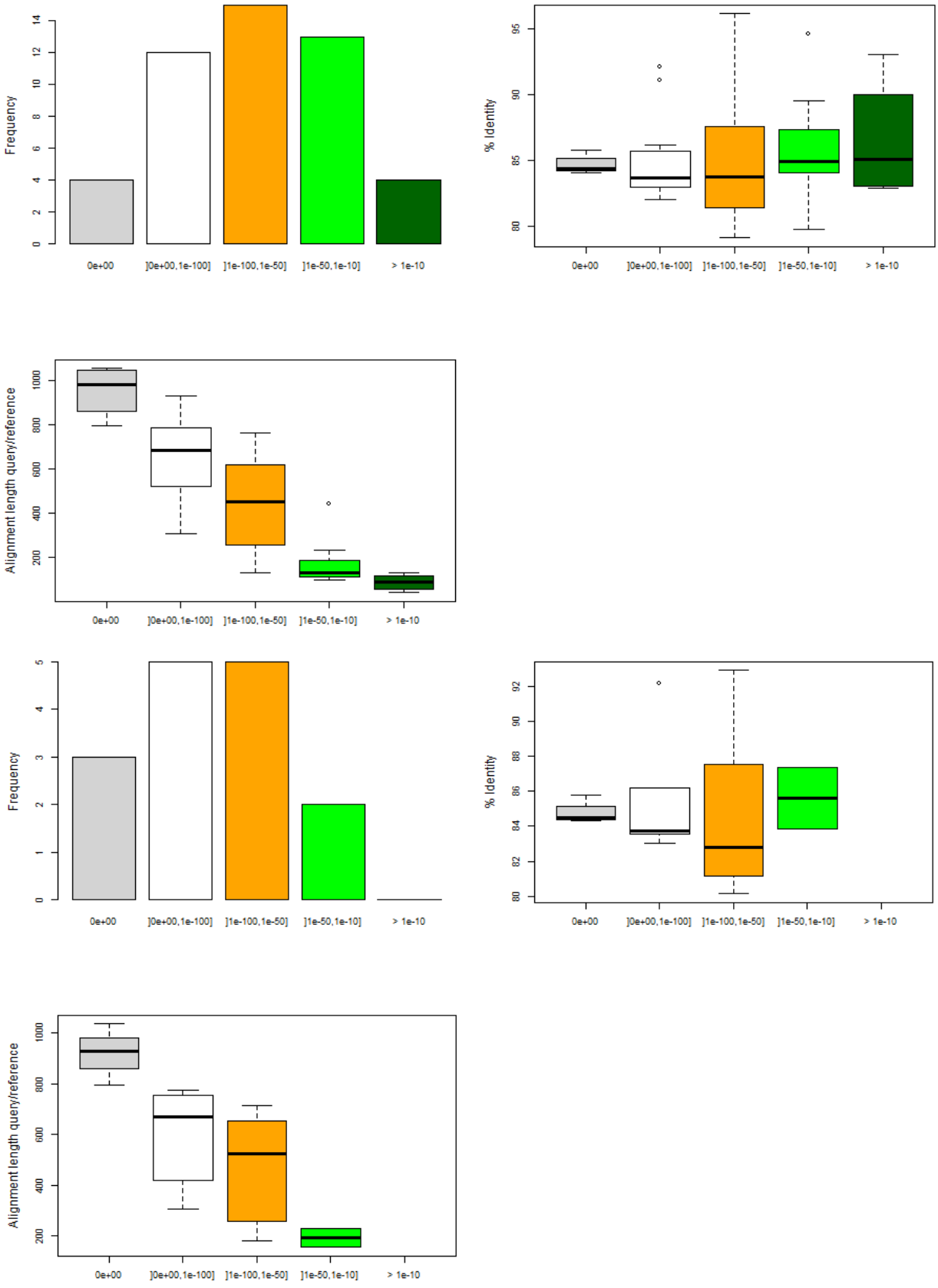
Supplementary Figure 6 - Plots of frequency, % identity and alignment length per class of e-value of *F. villafranca* against *X. silurana* contigs with only the best nucleotide (above) and protein (below) hits considered.



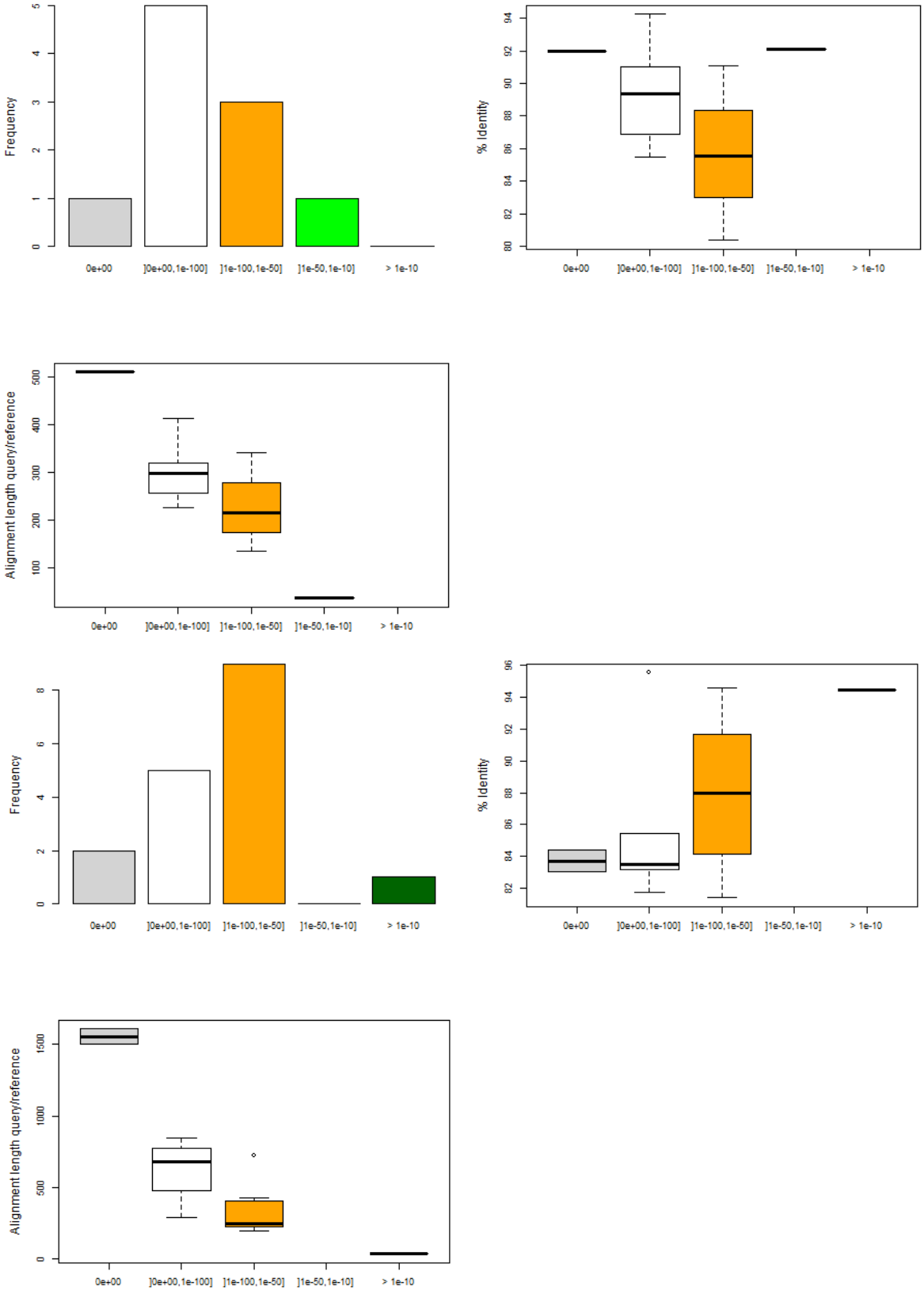
Supplementary Figure 7 - Plots of frequency, % identity and alignment length per class of e-value of *F. cantabrica* against nudibranchs contigs with only the best nucleotide (above) and protein (below) hits considered.



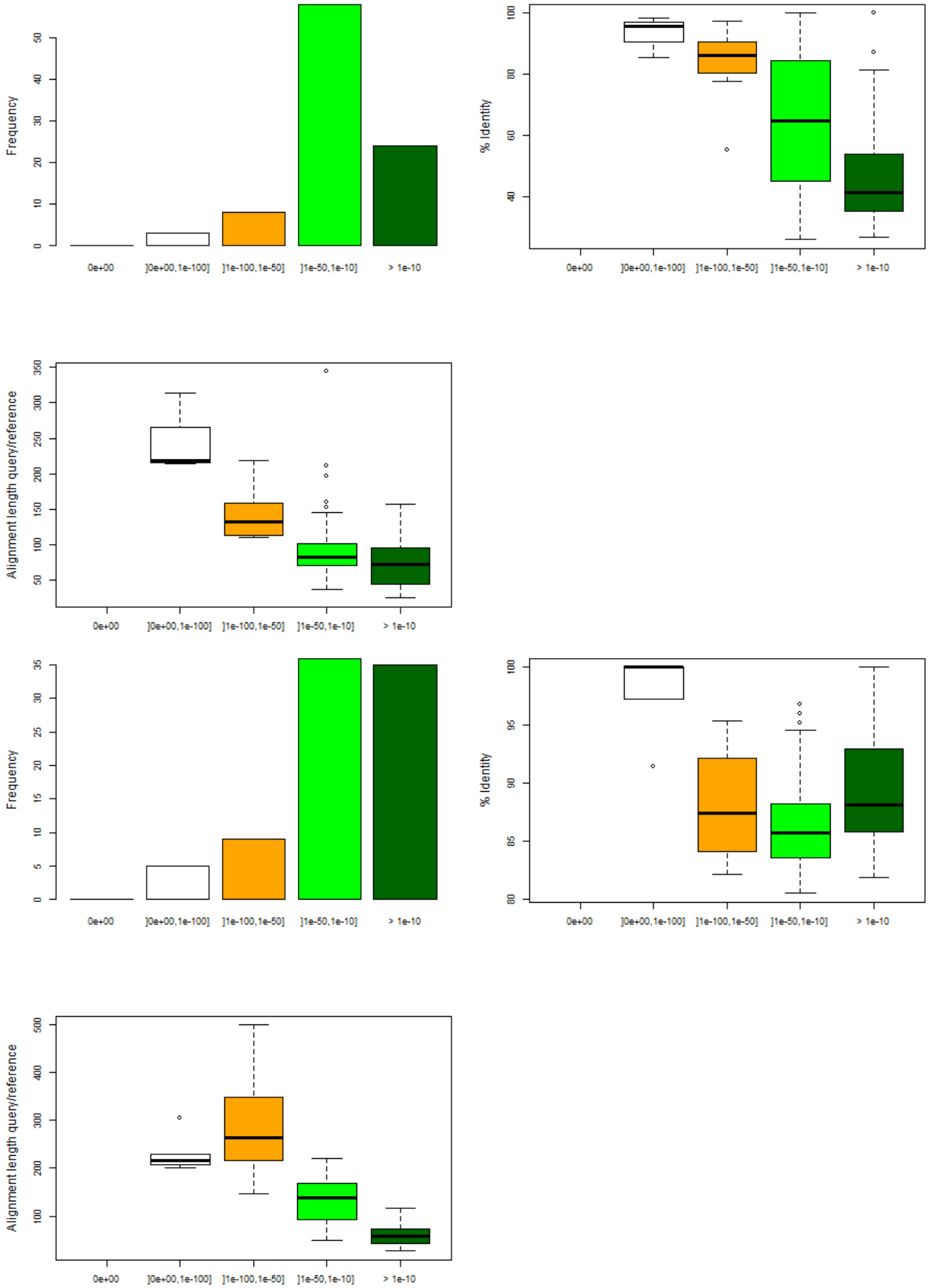
Supplementary Figure 8 - Plots of frequency, % identity and alignment length per class of e-value of *F. villafranca* against nudibranchs contigs with only the best nucleotide (above) and protein (below) hits considered.



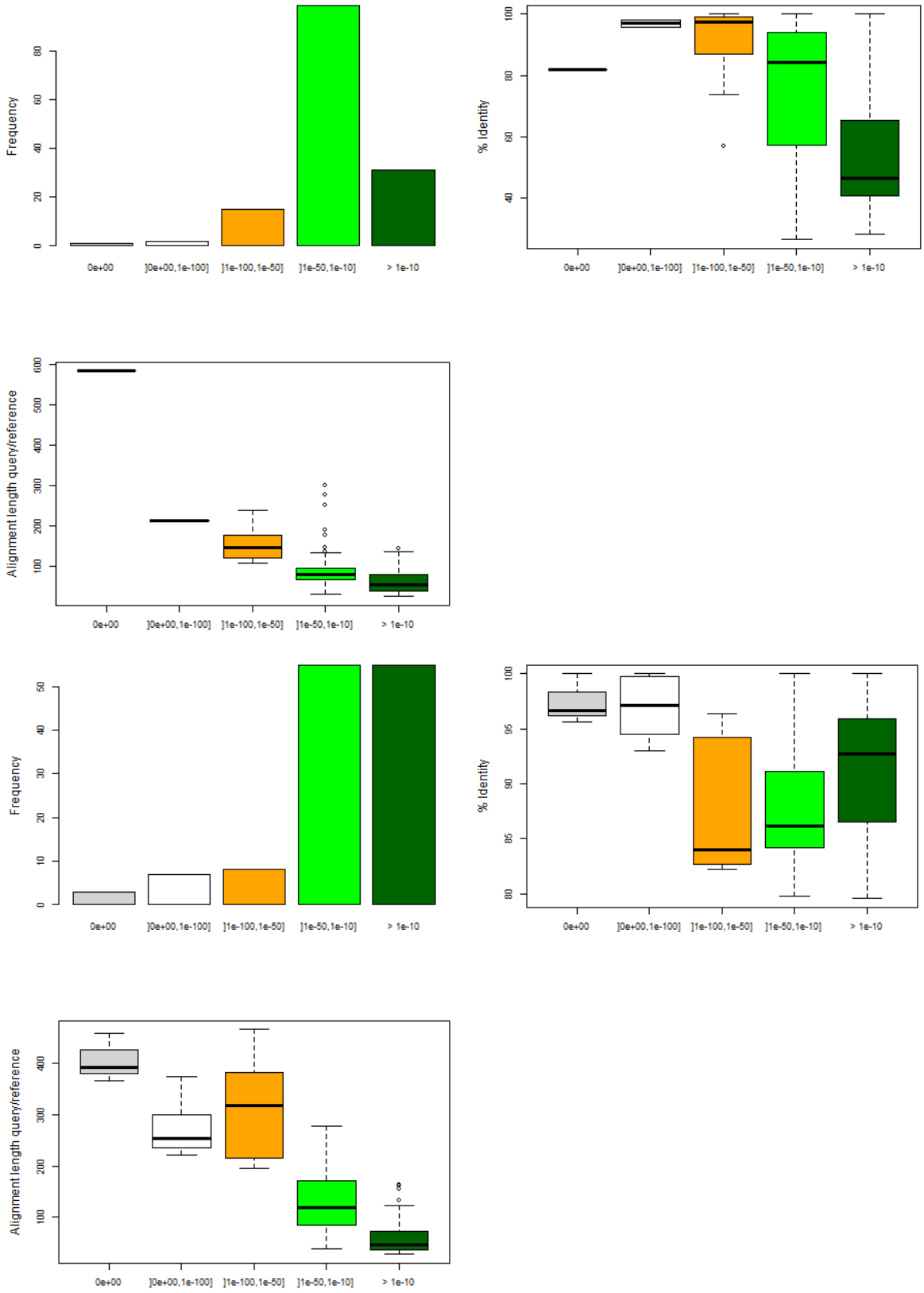
Supplementary Figure 9- Plots of frequency, % identity and alignment length per class of e-value of *F. cantabrica* against Bufo species contigs with only the best nucleotide (above) and protein (below) hits considered.



Supplementary Figure 10 - Plots of frequency, % identity and alignment length per class of e-value of *F. villafranca* against Bufo species contigs with only the best nucleotide (above) and protein (below) hits considered.



Supplementary Figure 11 - Plots of frequency, % identity and alignment length per class of e-value of *F. cantabrica* against *H. sapiens* contigs with only the best nucleotide (above) and protein (below) hits considered.



Supplementary Figure 12 - Plots of frequency, % identity and alignment length per class of e-value of *F. villafranca* against *H. sapiens* contigs with only the best nucleotide (above) and protein (below) hits considered.