U. PORTO

DOUTORAMENTO

BIOLOGIA MOLECULAR E CELULAR

The novel *Drosophila melanogaster* ascorbic acid synthesis and transport mechanisms

Pedro Miguel Barros Duque

D
2024

U. PORTO

Pedro Miguel Barros Duque. The novel *Drosophila melanogaster* ascorbic acid synthesis and transport mechanisms

D.ICBAS **2024**

The novel *Drosophila melanogaster* ascorbic acid synthesis and transport mechanisms

Pedro Miguel Barros Duque

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

PEDRO MIGUEL BARROS DUQUE

THE NOVEL *DROSOPHILA MELANOGASTER* ASCORBIC ACID SYNTHESIS AND TRANSPORT MECHANISMS

Tese de Candidatura ao grau de Doutor em

Biologia Molecular e Celular;

Programa Doutoral da Universidade do Porto

(Instituto de Ciências Biomédicas Abel Salazar e

Faculdade de Ciências)

Orientador: Doutor Jorge Manuel de Sousa Basto Vieira

Investigador Principal

i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto

IBMC - Instituto de Biologia Molecular e Celular

Co-orientadora: Doutora Cristina Alexandra Gonçalves Paula Vieira

Investigadora Auxiliar

i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto

IBMC - Instituto de Biologia Molecular e Celular

Acknowledgements / Agradecimentos

Although this thesis is only authored by myself, the PhD work and writing process conducive to this manuscript would never have occurred without the support of many great people.

Firstly, this work would not be possible without the knowledge and friendship provided by the Phenotypic Evolution group members. I would like to give a special thank you to my supervisors Jorge and Cristina Vieira for their patience, work and life teachings, and great feedback along this wondrous experience. In addition, I want to express my sincere gratitude to all the colleagues I had the chance to work with, especially to André Sousa, Pedro Ferreira and Ricardo Soares for their constant humorous and uplifting mood.

Furthermore, I would also like to thank all the members of the Yeast Signalling Networks for the excellent work environment provided in the lab, and particularly Vítor Teixeira, Telma Martins and Clara Pereira for the constant positive energy and considerate personality they have always shown. In addition, many thanks to Bruno Cavadas for the insightful opinions given and the helpful mindset consistently displayed.

Also my thanks to ICBAS, FCUP, i3S, IBMC and the MCBiology doctoral program for allowing the opportunity to perform this work. Within the i3S, I would like to give a particular thank you to the members of the Cell Culture and Genotyping Platform (CCGen) and the Lab Support Unit for the important help with experimental work and fly stock maintenance.

My family was always supportive of my academic pursuits, motivating me to become resilient and hard working. My parents in particular never ceased to believe that, even during my wavering moments, I had the potential to do whatever I sought out to do in all aspects of my life. I could have never asked for a better source of strength and support than that, and a simple thank you will never be enough to express my gratitude. I hope that in time I will get a chance to return their vote of confidence somehow.

Since my friends are always making fun of me for working with flies despite my efforts to explain to them how these organisms are a valuable experimental asset, I considered excluding them from the acknowledgements many times (kidding!). Seriously, I think that friends are the family we are able to freely choose during our lives, and thus unsurprisingly play an important role in the personal development and success of an individual. Sincerely, it is invaluable to know that they always wish for me to improve and prosper, and although I might not express it as many times as I should, I really appreciate the unconditional support. For that, you all deserve a huge thank you.

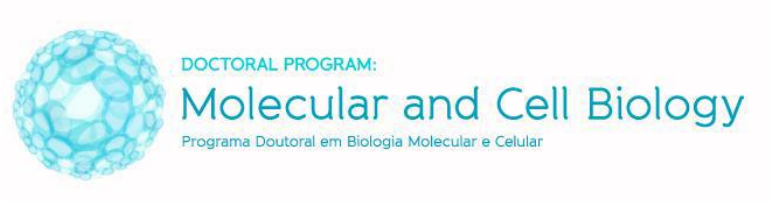
Financial support

This work was financed by the National Funds through FCT - Fundação para a Ciência e a Tecnologia, I.P., under the project UIDB/04293/2020. Pedro Duque was supported by a PhD fellowship from FCT (SFRH/BD/145515/2019; https://doi.org/10.54499/SFRH/BD/145515/2019), cofinanced by the European Social Fund through the Norte Portugal Regional Operational Programme (NORTE 2020).

















O autor esclarece que nesta tese foram incluídos os artigos publicados abaixo indicados, assim como resultados presentes em manuscritos submetidos para publicação. O autor declara também ter participado ativamente em todos os passos conducentes à elaboração da tese.

The author declares that the published articles mentioned below are included in this thesis, along with results included in manuscripts submitted for publication. The author further states that he participated actively in all steps conducive to the thesis writing.

Ao abrigo do disposto do no 2, alínea a) do artigo 31° do Decreto-Lei n.º 115/2013 de 7 de agosto, fazem parte integrante desta tese de doutoramento os seguintes trabalhos já publicados ou submetidos para publicação:

List of publications:

Duque, P., Vieira, C. P., Bastos, B., & Vieira, J. (2022). The evolution of vitamin C biosynthesis and transport in animals. *BMC Ecology and Evolution*, *22*(1), 84. https://doi.org/10.1186/s12862-022-02040-7

Duque, P., Vieira, C. P., & Vieira, J. (2022). Advances in Novel Animal Vitamin C Biosynthesis Pathways and the Role of Prokaryote-Based Inferences to Understand Their Origin. *Genes*, *13*(10), 1917. https://doi.org/10.3390/genes13101917

López-Fernández, H., Pinto, M., Vieira, C.P., **Duque, P.**, Reboiro-Jato, M., & Vieira, J. (2023). Auto-phylo: A Pipeline Maker for Phylogenetic Studies. In: Rocha, M., Fdez-Riverola, F., Mohamad, M.S., Gil-González, A.B. (eds) *Practical Applications of Computational Biology and Bioinformatics, 17th International Conference (PACBB 2023)*. PACBB 2023. Lecture Notes in Networks and Systems, vol 743. Springer, Cham. https://doi.org/10.1007/978-3-031-38079-2_3

Glez-Peña, D., López-Fernández, H., **Duque, P.**, Vieira, C. P., & Vieira, J. (2024). Inferences on the evolution of the ascorbic acid synthesis pathway in insects using Phylogenetic Tree Collapser (PTC), a tool for the automated collapsing of phylogenetic trees using taxonomic information. *Journal of Integrative Bioinformatics*. Under second round of revision.

Table of Contents

Abbreviations viii

Abstract x

Resumo xiii

Chapter 1. Introduction 1–1

1.1. Misconceptions regarding vitamin C auxotrophy in protostomian species and potential experimental applications of novel VC synthesis pathways 1–2

1.1.1. References 1–6

1.2. Review article: Advances in Novel Animal Vitamin C Biosynthesis Pathways and the Role of Prokaryote-Based Inferences to Understand Their Origin 1–12

1.2.1. Supporting information 1–42

1.2.2. Addendum: Further information on *C. elegans* VC synthesis and considerations on VC transport mechanisms 1–47

1.2.2.1. Review article synopsis 1–47

1.2.2.2. E03H4.3 and F54D5.12 as Regucalcin and GULO functional substitutes in *C. elegans* 1–47

1.2.2.3. Vitamin C transport: an uncharacterized mechanism in basal animal species 1–48

1.2.2.4. References 1–51

1.3. Thesis objectives and outlines 1–54

Chapter 2. The conservation of VC synthesis and transport mechanisms in animals: from the global perspective to the case of *D. melanogaster* 2–56

2.1. The evolution of vitamin C biosynthesis and transport in animals 2–57

2.1.1. Supplementary material 2–78

2.2. Addendum: Further experimental work and considerations in the context of "The evolution of vitamin C biosynthesis and transport in animals" article 2–103

2.2.1. Article synopsis 2–103

2.2.2. VC supplementation of *Regucalcin* and *Dca* RNAi crossings 2–104

2.2.3. Additional insights regarding the VC synthesis capabilities of *C. elegans* 2–104

2.2.4. References 2–105

Chapter 3. Vitamin C synthesis in *D. melanogaster*: bioinformatics and molecular biology approaches for the identification of novel pathways 3–106

3.1. Introduction 3–107

3.2. Results and discussion 3–110

3.2.1. Identification of gene homologs related to described VC synthesis pathways in *D. melanogaster* 3–110

3.2.2. VC colorimetric assays of *D. melanogaster* chromosomal deletion and RNAi stocks 3–126

3.2.2.1. *D. melanogaster* chromosomal deletion stocks screening 3–126

3.2.2.2. *D. melanogaster* RNAi crossings screening 3–131

3.3. Conclusion 3–135

3.4. Materials and methods 3–135

3.4.1. Phylogenetic analyses 3–135

3.4.2. *D. melanogaster* stock maintenance and crossings 3–138

3.4.3. VC colorimetric assays 3–139

3.4.4. Gene expression analyses 3–139

3.5. References 3–141

3.6. Supplementary material 3–146

Chapter 4. Phylogenetic inference of AOs presence within bacteria using auto-phylo 4–159

Chapter 5. General discussion 5–170

5.1. Initial background and objectives synopsis 5–171

5.2. The identification of genes related to VC synthesis in *D. melanogaster* 5–172

5.2.1. *Regucalcin* and *Dca* characterization efforts 5–172

5.2.2. Extended search for other gene candidates: The identification of *Miox* as a contributor to VC synthesis in *D. melanogaster* 5–173

5.2.3. Future perspectives on the characterization of the *D. melanogaster* VC synthesis pathways. 5–175

5.3. Identification of SVCT homologs as possible VC transporters in protostomian species 5–176

5.3.1. Future perspectives on the characterization of the *D. melanogaster* CG6293 5–176

5.4. Bacterial AOs as indication of possible endogenous VC synthesis in prokaryotes: the first large-scale step and future perspectives 5–177

5.5. References 5–178

Abbreviations

2R-WGD - Two rounds of whole genome duplication

ADOPS - Automatic Detection Of Positively Selected sites

AKR1A1 - Aldo-keto reductase family 1 member A1

Akr1B - Aldo-keto reductase 1B

AKR1B1 - Aldo-keto reductase family 1 member B1

AO - Aldonolactone oxidoreductase

BDSC - Bloomington Drosophila Stock Center

CDS - Coding sequence

Dca - Drosophila cold acclimation

DHCR24 - 24-Dehydrocholesterol Reductase

GLUT - Sodium-independent facilitative glucose transporter

GULO - L-Gulonolactone Oxidase

IMPA - Inositol monophosphatase

Inos - Myo-inositol-1-phosphate synthase

ISYNA1 - Inositol-3-phosphate synthase 1

MIOX - Myo-inositol oxygenase

NAT - Nucleobase-Ascorbate Transporter

NCBI - National Center for Biotechnology Information

PGM1 - Phosphoglucomutase 1

PGM2 - Phosphoglucomutase 2

RNAi - Ribonucleic acid interference

RPM - Rotations per minute

RT-PCR - Real-Time Polymerase Chain Reaction

SEDA - SEquence DAtaset builder

sgl - sugarless

SVCT - Sodium-dependent VC transporter

UGDH - UDP-Glucose 6-dehydrogenase

UGP - UDP-glucose pyrophosphorylase

UGP2 - UDP-Glucose pyrophosphorylase 2

UGUR - UDP-Glucuronidase

VAO/PCMH - Vanillyl alcohol oxidase/para-cresol methylhydroxylase

VC - Vitamin C

WGD - Whole genome duplication

Abstract

L-Gulonolactone oxidase (GULO) catalyses the final step of D-Glucose-6P conversion to ascorbic acid, also known as vitamin C (VC), in animal species. The loss or pseudogenization of *GULO* has been referred as the cause of VC auxotrophy in vertebrate species such as humans and teleost fishes. Nevertheless, the absence of this gene is not always indicative of an inability to synthesize this nutrient in basal animals, given that protostomes such as *Drosophila melanogaster*, *Bombyx mori* and *Caenorhabditis elegans* are capable of endogenous VC production. In mammalian species, VC has been known to intervene in neurodevelopment, neuromodulation and immune response, and could also be related to the attenuation of visceral fat deposition. However, the molecular mechanisms behind the roles of VC in such processes are mainly unknown. Coincidently, *D. melanogaster* and *C. elegans* are model organisms recurrently used in neuroscience, innate immunity, and obesity studies. Thus, the characterization of endogenous VC synthesis pathways in these species can open an unprecedented opportunity to mitigate the current limitations in the understanding of the functions of VC in many research topics of notable concern to human health. Previous novel VC synthesis pathway identification efforts revealed that the production of this nutrient in *B. mori* is associated with the functional adaptation of a 24-Dehydrocholesterol Reductase homolog to replace GULO. In *C. elegans*, the production of this nutrient is partly achieved by the replacement of GULO and the enzyme that precedes it in the characterized animal VC synthesis pathway, Regucalcin, by F54D5.12 and E03H4.3, respectively. Furthermore, although commonly disregarded, the conversion of myo-inositol to D-Glucuronate catalysed by the myo-inositol oxygenase (MIOX) has also been proposed as an alternative contributor to VC synthesis in animal species. However, the molecular mechanisms behind the capacity for VC production in *D. melanogaster* remains unknown.

The homeostasis of VC is crucial for optimal cellular function in animal species. In vertebrates, this is mainly achieved through active VC transport mediated by the sodium-dependent VC transporters (SVCTs) 1 and 2 during the dietary uptake and whole-body distribution of this nutrient, respectively. However, the SVCT protein family also contains the SVCT3 with yet unknown functions and the SVCT4 capable of nucleobase transport, which, despite the nomenclature, do not have affinity for VC. Although *SVCT* homologs similar to the vertebrate *SVCT4* were previously found in protostomes, none similar to the *SVCT1* or *SVCT2* was identified in these species, raising questions regarding the existence of dedicated VC transport in basal animal lineages. The absence of such mechanism would implicate that this nutrient is not required for several essential biological processes in a large subset of animal species, which is unlikely considering, for example, that *B. mori*

synthesizes VC but still relies on extra dietary uptake of this nutrient during larval stages 3L-E to 5L-E for adequate development. Therefore, *B. mori*, and inherently other protostomian species, likely have functional VC transporters. If so, then the identified invertebrate SVCTs might possess generalized nucleobase and VC transport capabilities, while the vertebrate SVCT4 possibly suffered the loss of the VC transport function throughout evolution. However, if such an evolutionary model can relate to the known existence of two rounds of whole genome duplication (2R-WGD) in vertebrates remains to be addressed. If verified, this would strengthen the hypothesis that the biological roles of the vertebrate-exclusive SVCT1, SVCT2, SVCT3 and SVCT4 are likely the result of subfunctionalization of a single ancestral SVCT. In this context, the identification of protostomian SVCTs in species such as *D. melanogaster* and *C. elegans* could ultimately translate to the development of novel experimental setups in which to verify the functional relevance of VC in several biological processes.

Despite often disregarded, the divergence of aldonolactone oxidoreductases (AOs) from prokaryotic sources has recently emerged as a possible evolutionary scenario that could explain the existence of the distinct eukaryote VC synthesis pathways. Nevertheless, the distribution of AOs in prokaryotes was never explored in a large-scale genomic context. Thus, the phylogenetic analysis of large prokaryote coding sequence datasets can help infer the existence of VC synthesis mechanisms in these organisms, a finding that might ultimately contribute to future improvements in the costly two-step fermentation process currently used in the industrial production of VC.

Following on this background, the work performed in this thesis focused mainly on the characterization of the *D. melanogaster* VC synthesis pathway and the identification of possible VC transporters present in basal animal species. Additionally, the distribution pattern of AOs in bacterial lineages was also addressed. A combination of bioinformatics methodologies, *D. melanogaster* deficiency stocks, colorimetric assays and RNAi-derived gene suppression using the GAL4-UAS system were used in order to pursue the objectives.

The results here presented show that the RNAi-mediated suppression of the homolog of *MIOX* in *D. melanogaster*, *Miox*, is associated with a maximum 25% reduction in total VC levels. Thus, VC synthesis is partly derived from the conversion of myo-inositol to D-Glucuronate in this organism. If D-Glucose-6P metabolism also synergistically contributes to D-Glucuronate formation in the context of VC synthesis in this species remains, however, to be addressed. The phylogenetic inferences regarding the evolutionary history of the *SVCT* gene lineage demonstrate for the first time that vertebrate *SVCT*s can derive from the known 2R-WGD events that affected species from this taxonomic group.

This evidence supported the hypothesis that the biological roles of SVCT1, SVCT2, SVCT3 and SVCT4 can be the result of subfunctionalization of a single ancestral SVCT after duplication, and highlighted that protostomian SVCTs are thus likely capable of nucleobase and VC transport. In this context, CG6293 emerges as the main VC transporter candidate in *D. melanogaster.* Regarding the distribution of AOs in prokaryotes, the presented phylogenetic data depicts for the first time that *GULO*-lineage homologs exist in 2073 bacteria. These AOs are particularly prevalent in the analysed Actinomycetota representatives, and seemingly crucial for the majority of Streptomycetales species. This evidence hints that a completely functional VC synthesis pathway might be present in a large number of bacteria, and that VC is required for optimal biological function in many of these organisms. Such findings provide a new perspective on the origins and presence of VC synthesis mechanisms in the different life kingdoms.

Keywords: vitamin C, evolution, *Regucalcin*, *SVCT*, aldonolactone oxidoreductase, myo-inositol, *D. melanogaster*, prokaryotes

Resumo

A enzima L-Gulonolactona oxidase (GULO) catalisa o passo final de conversão de D-Glucose-6P em ácido ascórbico, também conhecido como vitamina C (VC), em espécies animais. A perda ou pseudogenização do gene *GULO* foi referida como a causa de auxotrofia em VC em vertebrados como humanos e peixes teleósteos. No entanto, a perda deste gene não está sempre associada à perda de capacidade para síntese deste nutriente em animais basais, uma vez que protostómios como *Drosophila melanogaster*, *Bombyx mori* e *Caenorhabditis elegans* conseguem produzir VC. Em mamíferos, a VC intervém em processos de neurodesenvolvimento, neuromodelação e resposta imune, e pode também estar relacionada com a atenuação da deposição de gordura visceral. Porém, os mecanismos moleculares relacionados com o papel da VC nestes processos são maioritariamente desconhecidos. Coincidentemente, *D. melanogaste*r e *C. elegans* são organismos modelo usados recorrentemente em estudos de neurociência, imunidade inata e obesidade. Portanto, a caracterização da via de síntese endógena de VC nestas espécies pode constituir uma oportunidade sem precedentes para mitigar as limitações atuais na compreensão das funções da VC em vários tópicos de investigação com importância para a saúde humana. Esforços prévios na identificação de novas vias de síntese de VC revelaram que este nutriente é produzido em *B. mori* através da adaptação funcional de um homólogo da enzima 24-dehidrocolesterol redutase para substituição da GULO. Em *C. elegans,* a produção deste nutriente ocorre parcialmente pela substituição da GULO e da enzima que a precede na via de síntese de VC caracterizada em animais, Regucalcina, pelas enzimas F54D5.12 e E03H4.3, respetivamente. Além disso, ainda que habitualmente desvalorizada, a conversão de myo-inositol em D-Glucuronato catalisada pela enzima myo-inositol oxigenase (MIOX) foi também proposta como contribuidora alternativa para a síntese de VC em animais. No entanto, os mecanismos moleculares responsáveis pela produção de VC em *D. melanogaster* são ainda desconhecidos.

A homeostasia de VC é crucial para a completa funcionalidade das células em animais. Em vertebrados, este equilíbrio é mantido primariamente através do transporte ativo de VC mediado pelos transportadores de VC dependentes de sódio (SVCTs) 1 e 2 durante a absorção intestinal de nutrientes ou a distribuição corporal deste nutriente, respetivamente. Porém, a família de proteínas SVCT contém também o SVCT3 com funções ainda desconhecidas e o SVCT4 capaz de transportar bases azotadas, que, apesar da sua denominação, não têm afinidade para VC. Ainda que homólogos semelhantes ao *SVCT4* de vertebrados tenham sido encontrados em protostómios, não foram encontrados homólogos comparáveis aos *SVCT1* e *SVCT2* nestas espécies, originando dúvidas sobre a existência de transporte especifico de VC em linhagens de

animais basais. A ausência de mecanismos do género implicaria que este nutriente não é necessário para vários processos biológicos essenciais num considerável número de animais, o que é improvável considerando, por exemplo, que *B. mori* sintetiza VC, mas necessita da absorção deste nutriente através da dieta durante os estados larvares 3L-E a 5L-E para o seu desenvolvimento adequado. Posto isto, *B. mori*, e inerentemente outros protostómios, possivelmente possuem transportadores de VC funcionais. Se tal se verifica, então os SVCTs identificados em invertebrados podem ter uma capacidade generalizada para transporte de bases azotadas e VC, enquanto o SVCT4 encontrado em vertebrados perdeu a capacidade de transporte de VC durante o processo de evolução. No entanto, continua por averiguar se este modelo evolutivo é compatível com a existência de dois eventos de duplicação completa de genoma (2R-WGD) em vertebrados. Em caso de correlação, a hipótese que sugere que as funções dos transportadores exclusivos de vertebrados SVCT1, SVCT2, SVCT3 e SVCT4 derivam de subfuncionalização de um SVCT ancestral seria fortalecida. Neste contexto, a identificação de SVCTs em protostómios como *D. melanogaster* e *C. elegans* pode eventualmente contribuir para o desenvolvimento de novas metodologias experimentais nas quais a verificação de funções da VC em vários processos biológicos seja possível.

Apesar de regulamente desconsiderada, a divergência evolutiva de oxidorredutases de aldonolactonas (AOs) de origem procariótica emergiu recentemente como possível explicação para a existência de distintas vias de síntese de VC em eucariontes. Ainda assim, a distribuição de AOs em procariontes nunca foi explorada num contexto genómico alargado. Portanto, a análise filogenética de uma grande quantidade de sequências codificantes procarióticas pode ajudar a determinar a existência de mecanismos de síntese de VC nestes organismos, uma descoberta que poderá contribuir para o melhoramento do caro processo de fermentação em dois passos atualmente usado na produção industrial de VC.

Neste contexto, o trabalho executado nesta tese incidiu primariamente na caracterização da via de síntese de VC em *D. melanogaster* e na identificação de potenciais transportadores de VC em animais basais. Adicionalmente, o padrão de distribuição de AOs em linhagens de bactérias foi também observado. Metodologias de bioinformática, uma coleção de deficiências cromossómicas existentes em *D. melanogaster*, ensaios colorimétricos e a supressão da expressão de genes mediada por RNAi usando o sistema GAL4-UAS contribuíram para alcançar estes objetivos.

Os resultados aqui apresentados demonstram que a supressão mediada por RNAi do homólogo do gene *MIOX* em *D. melanogaster*, *Miox*, está associada a uma redução

máxima de 25% nos níveis de VC total. Portanto, a síntese de VC deriva parcialmente da conversão de myo-inositol em D-Glucuronato neste organismo. Continua, no entanto, por verificar se o metabolismo de D-Glucose-6P contribui de forma sinérgica para a formação de D-Glucuronato no contexto de síntese de VC nesta espécie. As inferências filogenéticas relativas à história evolutiva da linhagem do gene *SVCT* demonstram pela primeira vez que os *SVCT*s presentes em vertebrados podem derivar dos 2R-WDG que afetaram as espécies deste grupo taxonómico. Este resultado suporta a hipótese que sugere que as funções dos transportadores SVCT1, SVCT2, SVCT3 e SVCT4 derivam da subfuncionalização de um único SVCT ancestral, e evidencia que os SVCTs presentes em protostómios são possivelmente capazes de transportar bases azotadas e VC. Neste contexto, a proteína CG6293 surge como candidato primário a transportador de VC em *D. melanogaster*. Relativamente à distribuição de AOs em procariontes, os dados filogenéticos apresentados revelam que homólogos da linhagem do gene *GULO* existem em 2073 espécies de bactérias. Estes AOs são particularmente prevalentes nos representantes de Actinomycetota analisados, e parecem ser cruciais para a maioria das espécies de Streptomycetales. Esta evidência sugere que uma via funcional de síntese de VC pode estar presente num grande número de bactérias, e que VC é necessária para a completa funcionalidade biológica de muitos destes organismos. Estes resultados fornecem uma nova perspetiva sobre a origem e presença de mecanismos de síntese de VC nos diferentes reinos de seres vivos.

Palavras-chave: vitamina C, evolução, *Regucalcina*, *SVCT*, aldonolactona oxidorredutase, myo-inositol, *D. melanogaster*, procariontes

Chapter 1. Introduction

This chapter includes the following published review article:

Duque, P., Vieira, C. P., & Vieira, J. (2022). Advances in Novel Animal Vitamin C Biosynthesis Pathways and the Role of Prokaryote-Based Inferences to Understand Their Origin. *Genes*, *13*(10), 1917. https://doi.org/10.3390/genes13101917

This article is here reproduced under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1.1. Misconceptions regarding vitamin C auxotrophy in protostomian species and potential experimental applications of novel VC synthesis pathways

Vitamin C (VC), also known as L-ascorbic acid or L-ascorbate, is a crucial antioxidant and cofactor in the cellular physiology context (Doseděl *et al.* 2021, Li *et al.* 2023). However, the ability to endogenously synthesize VC is not ubiquitous in animal species. Humans (*Homo sapiens*) are commonly known VC auxotrophs (Davey *et al.* 2002, Dresen *et al.* 2023, Montel-Hagen *et al.* 2008) susceptible to defects in collagen hydroxylation under prolonged absence of VC supplementation (Hirschmann and Raugi 1999). Similarly, non-human primates, the guinea pig, numerous bats and birds, and teleost fishes have also lost this trait, due to the complete or partial loss of the *L-Gulonolactone oxidase* (*GULO*) gene (Drouin *et al.* 2011, Mellidou and Kanellis 2017, Yang 2013). GULO catalyzes the final oxidation step of the currently characterized vertebrate VC synthesis pathway (Figure 1.1), and is known to only have this function in vertebrates (Drouin *et al.* 2011, Linster and Van Schaftingen 2007).

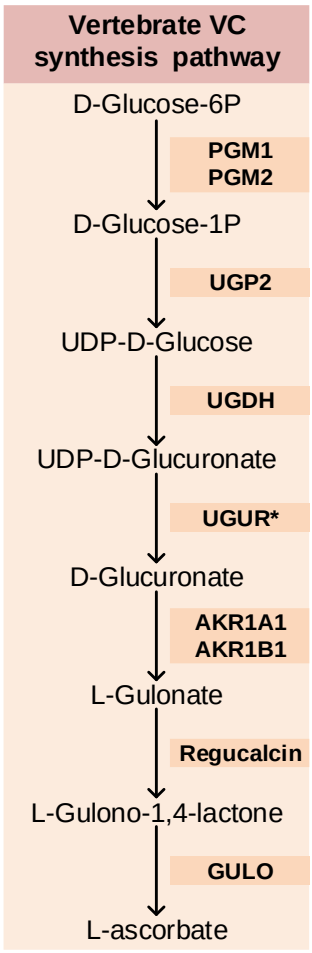


Figure 1.1. Graphical illustration of the characterized vertebrate AA synthesis pathway, adapted from the MetaCyc database (Caspi *et al.* 2020). The enzymes that catalyze each reaction step are shown between the relevant substrates and products. (Continues next page)

The enzyme acronyms represent the following nomenclature: Phosphoglucomutases 1 and 2 (PGM1 and PGM2), UDP-Glucose pyrophosphorylase 2 (UGP2), UDP-Glucose 6-dehydrogenase (UGDH), UDP-Glucuronidase (UGUR), the A1/B1 members of the aldo-keto reductase family 1 (AKR1A1/AKR1B1), and GULO. The asterisk highlights the only substrate conversion that remains hypothetical in the literature.

Within Euteleostomi species, *GULO* was lost multiple times independently and in a considerable different time scale, since the oldest loss event is likely to have occurred within teleost fishes around 200 to 210 million years ago (Cho *et al.* 2007, Dabrowski 1994, Dabrowski and Hinterleitner 1989, Moreau and Dabrowski 2005) while some more recent events may have happened 61 or even 14 million years ago in Haplorrhini primates and the guinea pig, respectively (Lachapelle and Drouin 2011). Even though the *GULO* loss events have been thoroughly identified within vertebrate species, the reasons behind this phenomenon in the affected taxonomic groups are not yet understood (Drouin *et al.* 2011, Fernie and Tohge 2015, Yang 2013). The same can be said from the various independent *GULO* gene reactivation events reported in some bats and passeriform birds so far (Cui *et al.* 2011, Drouin *et al.* 2011, Martinez del Rio 1997).

Extensive data on patterns of *GULO* loss or presence in non-bilaterian and protostomian species was only explored in the last decade. Initially, Wheeler *et al.* provided bioinformatics evidence of *GULO* presence in the non-bilaterian poriferan *Amphimedon queenslandica*, placozoan *Trichoplax adhaerens* and cnidarian *Nematostella vectensis* (Wheeler *et al.* 2015). In addition, the authors also pointed that this gene was observed within protostomians, namely in the annelid *Capitella teleta*, the acari tick *Ixodes scapularis*, the centipede *Strigamia maritima* and the gastropods *Aplysia californica*, *Haliotis discus hannai* and *Lottia gigantea* (Wheeler *et al.* 2015). Unfortunately, excluding the extrapolation that the ability to synthesize VC via *GULO* is the likely ancestral trait of the Ophistokonta taxonomic group (animal and fungi) that was after lost in specific lineages (Wheeler *et al.* 2015), there was no further discussion on these findings. Following on this knowledge gap, several subsequent analyses were performed in the host laboratory to clarify the pattern of *GULO* distribution within non-bilaterian and protostomian species. To achieve this purpose, López-Fernández *et al.* performed a large-scale phylogenetic analysis using representative animal coding sequences (CDS) provided by annotated genomes (López-Fernández *et al.* 2019), which was subsequently refined by the addition of non-annotated genome analyses (Henriques *et al.* 2019). The phylogenetic results from the annotated animal genomes revealed the presence of *GULO* in one non-bilaterian (Anthozoa) and four protostomian lineages (Araneae, Priapulida, Brachiopoda and Gastropoda), as well as the loss of this gene in five protostomian lineages (Insects, Crustacea, Nematoda, Platyhelminthes and

Bivalvia) (López-Fernández *et al.* 2019). Nevertheless, the results were inconclusive for more than half of the total analysed non-bilaterian and protostomian groups (11 out of 21) (López-Fernández *et al.* 2019). The analyses on non-annotated animal genomes performed by Henriques *et al.* provided additional insights on the subject, revealing that *GULO* was also present in two non-bilaterian (Placozoa and Myxozoa) and two protostomian (Acari and Annelida) lineages, but was lost in the non-Insecta Hexapoda species (Figure 1.2) (Henriques *et al.* 2019). Nevertheless, *GULO* presence or loss could not be inferred in the remaining six lineages due to limitations with the total number of species that had available genomic data (Figure 1.2) (Henriques *et al.* 2019).

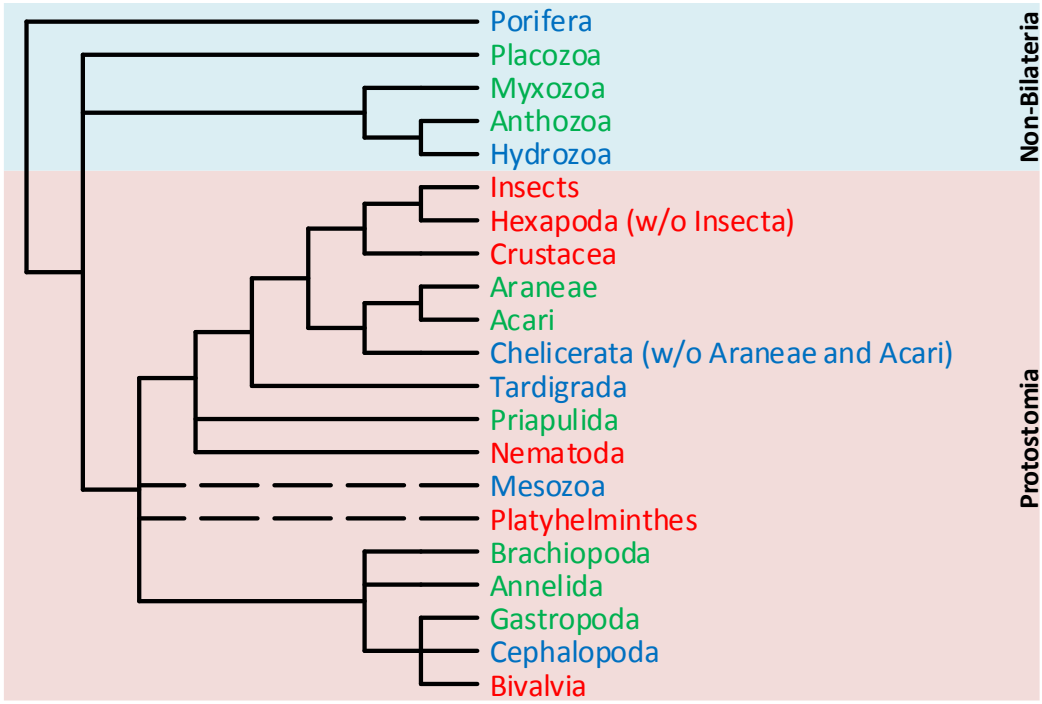


Figure 1.2. Distribution of putative *GULO* genes in protostomian and non-bilaterian lineages. The *GULO* gene is present in the green highlighted lineages, absent in the red and currently uncharacterized in the blue. Dashed lines represent uncertain taxonomic relationships. The depicted taxonomy was based on the Tree of Life Web Project (http://tolweb.org/tree/; (Maddison *et al.* 2007)). The figure and description were adapted from Henriques *et al.* (Henriques *et al.* 2019).

These phylogenetic results were compatible with the previous preconceptions of global insect VC auxotrophy (Chatterjee 1973, Gillespie 1980, Gupta *et al.* 1972), as the lack of *GULO* in these species correlated well with their reported inability to synthesize this nutrient. Nevertheless, although the pseudogenization or even complete loss of *GULO* is often considered the primary cause of VC synthesis inability in vertebrates (Drouin *et al.* 2011, Lachapelle and Drouin 2011, Yang 2013), it is not the case in some protostomian species. Despite absence of *GULO*, functional research has so far shown that endogenous

VC synthesis occurs in the insects *Drosophila melanogaster* (Henriques *et al.* 2019, Massie *et al.* 1991) and *Bombyx mori* (Hou *et al.* 2021), and the nematode *Caernorhabditis elegans* (Patananan *et al.* 2015, Yabuta *et al.* 2020), indicating that yet undescribed VC synthesis mechanisms have emerged within basal animal species.

VC has been the subject of research in many scientific fields with human relevance, such as human nutrition (Chaturvedi *et al.* 2022, Yin *et al.* 2022), animal feed supplementation (Hieu *et al.* 2022, Pappenberger and Hohmann 2014), agricultural and industrial waste management (Chaudhry *et al.* 2022, Sheng *et al.* 2021), and sustainable nanomaterial synthesis (Basnayake Pussepitiyalage and Hemmati 2022). Recent reports and reviews on VC are mainly centred on potential novel applications or functions of this nutrient (e.g. (Lavillegrand *et al.* 2022, Li *et al.* 2022, Shahbaz *et al.* 2022, Taira *et al.* 2023)), while findings on novel endogenous VC synthesis mechanisms have become less frequent since the characterization of the three main pathways occurring in animals, plants and photosynthetic protists. This research discrepancy is arguably not surprising, considering that the Smirnoff-Wheeler plant VC synthesis pathway (Wheeler *et al.* 1998) is the one predominant in embryophyte species of nutritional interest (Fenech *et al.* 2019), and VC already has a considerable industrial production reliant on optimized 2-keto-L-gulonic fermentation (Luo *et al.* 2021, Pappenberger and Hohmann 2014). Nevertheless, further exploration on alternative VC synthesis pathways can also translate into applications of anthropological significance, and thus should not be disregarded. VC has key functions in mammalian neurodevelopment (Coker *et al.* 2022), neuromodulation (Zylinska *et al.* 2023), and immune response (Carr and Maggini 2017), although little is known regarding the role of this nutrient in the mediation of such processes (Chen *et al.* 2022, Papri *et al.* 2022). Moreover, VC supplementation has been previously correlated with a decrease of visceral obesity in vertebrate species (Lee *et al.* 2019), but its contribution for human obesity prevention is still a controversial subject in need of further research (Hosseini *et al.* 2017). All of these topics are of notable concern to human health, and thus objectively require continuous scientific scrutiny. *C. elegans* and *D. melanogaster* have been recurrently used in neuroscience (Bessa *et al.* 2013, Calahorro and Ruiz-Rubio 2011, Fischer *et al.* 2023, Guerreiro and Maciel 2023), innate immunity (Alarco *et al.* 2004, Marsh Elizabeth and May Robin 2012, Younes *et al.* 2020), and obesity studies (Eickelberg *et al.* 2022, Jones and Ashrafi 2009, Zheng and Greenway 2012). However, such research was never performed while accounting for possible VC self-sufficiency in these organisms, and thus, could possibly present an unknowingly limited theoretical framework. As previously stated, *C. elegans* and *D. melanogaster* are capable of endogenous VC synthesis despite the loss of GULO (Henriques *et al.* 2019, Massie *et al.* 1991, Patananan *et al.* 2015, Yabuta *et al.*

2020), potentially due to the presence of partially adapted or completely novel metabolic pathways. The existence of such pathways in these species has opened an unprecedented opportunity to mitigate the current limitations in the understanding of VC-derived molecular mechanisms underlying many biological processes, putatively through the development of novel experimental setups. In fact, both model organisms offer several experimental advantages relative to the vertebrates often used to verify the biological effects of VC deprivation (e.g. mice (Aumailley *et al.* 2016) and guinea pig (Kipp *et al.* 1996)), such as the absence of ethical limitations (Fontana *et al.* 2021) and the fairly simple genetic screening of mutant phenotypes (Kimble and Nüsslein-Volhard 2022). Nonetheless, the potential applications first require the full enzymatic characterization of the VC synthesis systems in these protostomians, a pursuit that is considerably complex when considering the available literature within this research field.

1.1.1. References

Alarco, A.-M., A. Marcil, J. Chen, B. Suter, D. Thomas and M. Whiteway. **2004.** Immune-Deficient *Drosophila melanogaster*: A Model for the Innate Immune Response to Human Fungal Pathogens. *The Journal of Immunology*. 172:5622-5628.

Aumailley, L., A. Warren, C. Garand, M. J. Dubois, E. R. Paquet, D. G. Le Couteur, A. Marette, V. C. Cogger and M. Lebel. **2016.** Vitamin C modulates the metabolic and cytokine profiles, alleviates hepatic endoplasmic reticulum stress, and increases the life span of Gulo-/- mice. *Aging (Albany NY)*. 8:458-483.

Basnayake Pussepitiyalage, V. and S. Hemmati. **2022.** Sustainable, Green, and Continuous Synthesis of Fivefold Palladium Nanorods Using l-Ascorbic Acid in a Segmented Millifluidic Flow Reactor. *Langmuir*. 38:4200-4212.

Bessa, C., P. Maciel and A. J. Rodrigues. **2013.** Using *C. elegans* to Decipher the Cellular and Molecular Mechanisms Underlying Neurodevelopmental Disorders. *Molecular Neurobiology*. 48:465-489.

Calahorro, F. and M. Ruiz-Rubio. **2011.** *Caenorhabditis elegans* as an experimental tool for the study of complex neurological diseases: Parkinson's disease, Alzheimer's disease and autism spectrum disorder. *Invertebrate Neuroscience*. 11:73-83.

Carr, A. C. and S. Maggini. **2017.** Vitamin C and Immune Function. *Nutrients*. 9:1211.

Caspi, R., R. Billington, I. M. Keseler, A. Kothari, M. Krummenacker, P. E. Midford, W. K. Ong, S. Paley, P. Subhraveti and P. D. Karp. **2020.** The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Research*. 48:D445-d453.

Chatterjee, I. B. **1973.** Evolution and the Biosynthesis of Ascorbic Acid. *Science*. 182:1271.

Chaturvedi, S., S. Khan, R. K. Bhunia, K. Kaur and S. Tiwari. **2022.** Metabolic engineering in food crops to enhance ascorbic acid production: crop biofortification perspectives for human health. *Physiology and Molecular Biology of Plants*. 28:871-884.

Chaudhry, A., F. Bashir, S. F. Adil, S. Saif, M. R. Shaik, M. R. Hatshan and B. Shaik. **2022.** Ascorbic acid-mediated Fe/Cu nanoparticles and their application for removal of COD and phenols from industrial wastewater. *Journal of King Saud University - Science*. 34:101927.

Chen, H.-Y., A. Almonte-Loya, F.-Y. Lay, M. Hsu, E. Johnson, E. González-Avalos, J. Yin, R. S. Bruno, Q. Ma, H. E. Ghoneim, D. J. Wozniak, F. E. Harrison and C.-W. J. Lio. **2022.** Epigenetic remodeling by vitamin C potentiates plasma cell differentiation. *eLife*. 11:e73754.

Cho, Y. S., S. E. Douglas, J. W. Gallant, K. Y. Kim, D. S. Kim and Y. K. Nam. **2007.** Isolation and characterization of cDNA sequences of L-gulono-gamma-lactone oxidase, a key enzyme for biosynthesis of ascorbic acid, from extant primitive fish groups. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*. 147:178-190.

Coker, S. J., C. C. Smith-Díaz, R. M. Dyson, M. C. M. Vissers and M. J. Berry. **2022.** The Epigenetic Role of Vitamin C in Neurodevelopment. *International Journal of Molecular Sciences*. 23:1208.

Cui, J., Y. H. Pan, Y. Zhang, G. Jones and S. Zhang. **2011.** Progressive pseudogenization: vitamin C synthesis and its loss in bats. *Molecular Biology and Evolution*. 28:1025-1031.

Dabrowski, K. **1994.** Primitive actimoterigian fishes can synthesize ascorbic acid. *Experientia*. 50:745-748.

Dabrowski, K. and S. Hinterleitner. **1989.** Applications of a simultaneous assay of ascorbic acid, dehydroascorbic acid and ascorbic sulphate in biological materials. *Analyst*. 114:83-87.

Davey, M. W., M. V. Montagu, D. Inzé, M. Sanmartin, A. Kanellis, N. Smirnoff, I. J. J. Benzie, J. J. Strain, D. Favell and J. Fletcher. **2002.** Plant L-ascorbic acid: chemistry, function, metabolism, bioavailability and effects of processing. *Journal of the Science of Food and Agriculture*. 80:825-860.

Doseděl, M., E. Jirkovský, K. Macáková, L. K. Krčmová, L. Javorská, J. Pourová, L. Mercolini, F. Remião, L. Nováková, P. Mladěnka and on behalf of The OEMONOM. **2021.** Vitamin C—Sources, Physiological Role, Kinetics, Deficiency, Use, Toxicity, and Determination. *Nutrients*. 13:615.

Dresen, E., Z.-Y. Lee, A. Hill, Q. Notz, J. J. Patel and C. Stoppe. **2023.** History of scurvy and use of vitamin C in critical illness: A narrative review. *Nutrition in Clinical Practice*. 38:46-54.

Drouin, G., J.-R. Godin and B. Pagé. **2011.** The Genetics of Vitamin C Loss in Vertebrates. *Current Genomics*. 12:371-378.

Eickelberg, V., K. Lüersen, S. Staats and G. Rimbach. **2022.** Phenotyping of *Drosophila melanogaster* - A Nutritional Perspective. *Biomolecules*. 12:221.

Fenech, M., I. Amaya, V. Valpuesta and M. A. Botella. **2019.** Vitamin C Content in Fruits: Biosynthesis and Regulation. *Frontiers in Plant Science*. 9:2006.

Fernie, A. R. and T. Tohge. **2015.** A cross-kingdom history. *eLife*. 4:e07527.

Fischer, F. P., R. A. Karge, Y. G. Weber, H. Koch, S. Wolking and A. Voigt. **2023.** *Drosophila melanogaster* as a versatile model organism to study genetic epilepsies: An overview. *Frontiers in Molecular Neuroscience*. 16:42.

Fontana, F., P. Figueiredo, J. P. Martins and H. A. Santos. **2021.** Requirements for Animal Experiments: Problems and Challenges. *Small*. 17:2004182.

Gillespie, D. S. **1980.** Overview of Species Needing Dietary Vitamin C. *The Journal of Zoo Animal Medicine*. 11:88-91.

Guerreiro, S. and P. Maciel. **2023.** Transition from Animal-Based to Human Induced Pluripotent Stem Cells (iPSCs)-Based Models of Neurodevelopmental Disorders: Opportunities and Challenges. *Cells*. 12:538.

Gupta, S. D., C. R. Chaudhuri and I. B. Chatterjee. **1972.** Incapability of l-ascorbic acid synthesis by insects. *Archives of Biochemistry and Biophysics*. 152:889-890.

Henriques, S. F., P. Duque, H. López-Fernández, N. Vázquez, F. Fdez-Riverola, M. Reboiro-Jato, C. P. Vieira and J. Vieira. **2019.** Multiple independent *L-gulonolactone oxidase* (*GULO*) gene losses and vitamin C synthesis reacquisition events in non-Deuterostomian animal species. *BMC Evolutionary Biology*. 19:126.

Hieu, T. V., B. Guntoro, N. H. Qui, N. T. K. Quyen and F. A. A. Hafiz. **2022.** The application of ascorbic acid as a therapeutic feed additive to boost immunity and antioxidant activity of poultry in heat stress environment. *Vet World*. 15:685-693.

Hirschmann, J. V. and G. J. Raugi. **1999.** Adult scurvy. *Journal of the American Academy of Dermatology*. 41:895-910.

Hosseini, B., A. Saedisomeolia and M. Allman-Farinelli. **2017.** Association Between Antioxidant Intake/Status and Obesity: a Systematic Review of Observational Studies. *Biological Trace Element Research*. 175:287-297.

Hou, M., T. M. Dai, X. Y. Liang, S. X. Zhang, W. Z. Cui, J. F. Qiu, Y. H. Sima, W. Z. Cui and S. Q. Xu. **2021.** *Bombyx mori* can synthesize ascorbic acid through the l-gulose pathway to varying degrees depending on developmental stage. *Archives of Insect Biochemistry and Physiology*. 106:e21783.

Jones, K. T. and K. Ashrafi. **2009.** *Caenorhabditis elegans* as an emerging model for studying the basic biology of obesity. *Disease models & mechanisms*. 2:224-229.

Kimble, J. and C. Nüsslein-Volhard. **2022.** The great small organisms of developmental genetics: *Caenorhabditis elegans* and *Drosophila melanogaster*. *Developmental Biology*. 485:93-122.

Kipp, D. E., M. McElvain, D. B. Kimmel, M. P. Akhter, R. G. Robinson and B. P. Lukert. **1996.** Scurvy results in decreased collagen synthesis and bone density in the guinea pig animal model. *Bone*. 18:281-288.

Lachapelle, M. Y. and G. Drouin. **2011.** Inactivation dates of the human and guinea pig vitamin C genes. *Genetica*. 139:199-207.

Lavillegrand, J.-R., L. Raia, T. Urbina, G. Hariri, P. Gabarre, V. Bonny, N. Bigé, J.-L. Baudel, A. Bruneel, T. Dupre, B. Guidet, E. Maury and H. Ait-Oufella. **2022.** Vitamin C improves microvascular reactivity and peripheral tissue perfusion in septic shock patients. *Critical Care*. 26:25.

Lee, H., J. Ahn, S. S. Shin and M. Yoon. **2019.** Ascorbic acid inhibits visceral obesity and nonalcoholic fatty liver disease by activating peroxisome proliferator-activated receptor α in high-fat-diet-fed C57BL/6J mice. *International Journal of Obesity*. 43:1620-1630.

Li, W.-N., S.-J. Zhang, J.-Q. Feng and W.-L. Jin. **2022.** Repurposing Vitamin C for Cancer Treatment: Focus on Targeting the Tumor Microenvironment. *Cancers*. 14:2608.

Li, X.-Y., L. Meng, L. Shen and H.-F. Ji. **2023.** Regulation of gut microbiota by vitamin C, vitamin E and β-carotene. *Food Research International*. 169:112749.

Linster, C. L. and E. Van Schaftingen. **2007.** Vitamin C. Biosynthesis, recycling and degradation in mammals. *The FEBS Journal*. 274:1-22.

López-Fernández, H., P. Duque, S. Henriques, N. Vázquez, F. Fdez-Riverola, C. P. Vieira, M. Reboiro-Jato and J. Vieira. **2019.** A Bioinformatics Protocol for Quickly Creating Large-Scale Phylogenetic Trees. In: F. Fdez-Riverola, M. S. Mohamad, M. Rocha, J. F. De Paz and P. González, editors. *Practical Applications of Computational Biology and Bioinformatics, 12th International Conference.* Cham. Springer. p. 88-96.

Luo, Z., S. Yu, W. Zeng and J. Zhou. **2021.** Comparative analysis of the chemical and biochemical synthesis of keto acids. *Biotechnology Advances*. 47:107706.

Maddison, D. R., K.-S. Schulz and W. P. Maddison. **2007.** The tree of life web project. *Zootaxa*. 1668:19-40.

Marsh Elizabeth, K. and C. May Robin. **2012.** *Caenorhabditis elegans*, a Model Organism for Investigating Immunity. *Applied and Environmental Microbiology*. 78:2075-2081.

Martinez del Rio, C. **1997.** Can Passerines Synthesize Vitamin C? *The Auk*. 114:513-516.

Massie, H. R., M. E. Shumway, S. J. P. Whitney, S. M. Sternick and V. R. Aiello. **1991.** Ascorbic acid in *Drosophila* and changes during aging. *Experimental Gerontology*. 26:487-494.

Mellidou, I. and A. K. Kanellis. **2017.** Genetic Control of Ascorbic Acid Biosynthesis and Recycling in Horticultural Crops. *Frontiers in Chemistry*. 5.

Montel-Hagen, A., S. Kinet, N. Manel, C. Mongellaz, R. Prohaska, J.-L. Battini, J. Delaunay, M. Sitbon and N. Taylor. **2008.** Erythrocyte Glut1 Triggers Dehydroascorbic Acid Uptake in Mammals Unable to Synthesize Vitamin C. *Cell*. 132:1039-1048.

Moreau, R. and K. Dabrowski. **2005.** Biosynthesis of ascorbic acid by extant actinopterygians. *Journal of Fish Biology*. 57:733-745.

Pappenberger, G. and H.-P. Hohmann. **2014.** Industrial Production of l-Ascorbic Acid (Vitamin C) and d-Isoascorbic Acid. In: H. Zorn and P. Czermak*,* editors. *Biotechnology of Food and Feed Additives*. Berlin, Heidelberg. Springer Berlin Heidelberg. p. 143-188.

Papri, D., V. Akanksha and A. Richa. **2022.** Nutrition influences nervous system development by regulating neural stem cell homeostasis. *Proceedings of the Indian National Science Academy*. 88:482-498.

Patananan, A. N., L. M. Budenholzer, M. E. Pedraza, E. R. Torres, L. N. Adler and S. G. Clarke. **2015.** The invertebrate *Caenorhabditis elegans* biosynthesizes ascorbate. *Archives of Biochemistry and Biophysics*. 569:32-44.

Shahbaz, U., N. Fatima, S. Basharat, A. Bibi, X. Yu, M. I. Hussain and M. Nasrullah. **2022.** Role of vitamin C in preventing of COVID-19 infection, progression and severity. *AIMS Microbiology*. 8:108-124.

Sheng, Y., X. Tan, Y. Gu, X. Zhou, M. Tu and Y. Xu. **2021.** Effect of ascorbic acid assisted dilute acid pretreatment on lignin removal and enzyme digestibility of agricultural residues. *Renewable Energy*. 163:732-739.

Smirnoff, N. **2018.** Ascorbic acid metabolism and functions: A comparison of plants and mammals. *Free Radical Biology and Medicine*. 122:116-129.

Taira, A., K. Palin, A. Kuosmanen, N. Välimäki, O. Kuittinen, O. Kuismin, E. Kaasinen, K. Rajamäki and L. A. Aaltonen. **2023.** Vitamin C boosts DNA demethylation in TET2 germline mutation carriers. *Clinical Epigenetics*. 15:7.

Wheeler, G., T. Ishikawa, V. Pornsaksit and N. Smirnoff. **2015.** Evolution of alternative biosynthetic pathways for vitamin C following plastid acquisition in photosynthetic eukaryotes. *eLife*. 4:e06369.

Wheeler, G. L., M. A. Jones and N. Smirnoff. **1998.** The biosynthetic pathway of vitamin C in higher plants. *Nature*. 393:365.

Yabuta, Y., R. Nagata, Y. Aoki, A. Kariya, K. Wada, A. Yanagimoto, H. Hara, T. Bito, N. Okamoto, S. Yoshida, A. Ishihara and F. Watanabe. **2020.** L-Ascorbate Biosynthesis Involves Carbon Skeleton Rearrangement in the Nematode *Caenorhabditis elegans*. *Metabolites*. 10:334.

Yang, H. **2013.** Conserved or Lost: Molecular Evolution of the Key Gene *GULO* in Vertebrate Vitamin C Biosynthesis. *Biochemical Genetics*. 51:413-425.

Yin, X., K. Chen, H. Cheng, X. Chen, S. Feng, Y. Song and L. Liang. **2022.** Chemical Stability of Ascorbic Acid Integrated into Commercial Products: A Review on Bioactivity and Delivery Technology. *Antioxidants*. 11:153.

Younes, S., A. Al-Sulaiti, E. A. A. Nasser, H. Najjar and L. Kamareddine. **2020.** *Drosophila* as a Model Organism in Host–Pathogen Interaction Studies. *Frontiers in Cellular and Infection Microbiology*. 10:214.

Zheng, J. and F. Greenway. **2012.** *Caenorhabditis elegans* as a model for obesity research. *International journal of obesity*. 36:186-194.

Zylinska, L., M. Lisek, F. Guo and T. Boczek. **2023.** Vitamin C Modes of Action in Calcium-Involved Signaling in the Brain. *Antioxidants*. 12:231.

1.2. Review article: Advances in Novel Animal Vitamin C Biosynthesis Pathways and the Role of Prokaryote-Based Inferences to Understand Their Origin

genes

MDPI

Review

Advances in Novel Animal Vitamin C Biosynthesis Pathways and the Role of Prokaryote-Based Inferences to Understand Their Origin

Pedro Duque [1,2,3,4], Cristina P. Vieira [3,4] and Jorge Vieira [3,4,*]

[1] School of Medicine and Biomedical Sciences (ICBAS), Porto University, Rua de Jorge Viterbo Ferreira, 228, 4050-313 Porto, Portugal
[2] Faculdade de Ciências da Universidade do Porto (FCUP), Rua do Campo Alegre, s/n, 4169-007 Porto, Portugal
[3] Instituto de Investigação e Inovação em Saúde (I3S), Universidade do Porto, Rua Alfredo Allen, 208, 4200-135 Porto, Portugal
[4] Instituto de Biologia Molecular e Celular (IBMC), Rua Alfredo Allen, 208, 4200-135 Porto, Portugal
* Correspondence: jbvieira@ibmc.up.pt; Tel.:+351-226-074-900

Abstract: Vitamin C (VC) is an essential nutrient required for the optimal function and development of many organisms. VC has been studied for many decades, and still today, the characterization of its functions is a dynamic scientific field, mainly because of its commercial and therapeutic applications. In this review, we discuss, in a comparative way, the increasing evidence for alternative VC synthesis pathways in insects and nematodes, and the potential of myo-inositol as a possible substrate for this metabolic process in metazoans. Methodological approaches that may be useful for the future characterization of the VC synthesis pathways of *Caenorhabditis elegans* and *Drosophila melanogaster* are here discussed. We also summarize the current distribution of the eukaryote aldonolactone oxidoreductases gene lineages, while highlighting the added value of studies on prokaryote species that are likely able to synthesize VC for both the characterization of novel VC synthesis pathways and inferences on the complex evolutionary history of such pathways. Such work may help improve the industrial production of VC.

Keywords: ascorbic acid; evolution; synthesis; aldonolactone oxidoreductases; insects; nematodes; prokaryotes

check for updates

Citation: Duque, P.; Vieira, C.P.; Vieira, J. Advances in Novel Animal Vitamin C Biosynthesis Pathways and the Role of Prokaryote-Based Inferences to Understand Their Origin. *Genes* **2022**, *13*, 1917. https://doi.org/10.3390/genes13101917

Academic Editor: Fabrizio Lombardo

Received: 20 September 2022
Accepted: 18 October 2022
Published: 21 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Vitamin C (VC) is a water-soluble compound that, at physiological pH, is mainly found in a reduced ionizable form (L-ascorbate) [1]. The potential applications of VC have been the subject of research since it was first synthesized in 1933 [2] to the present day, in many areas relevant to mankind, such as human nutrition [3,4], animal feed supplementation [5,6], agricultural and industrial waste management [7,8], and sustainable nanomaterial synthesis [9]. Recently, this nutrient gained further relevance as a potential therapeutic agent for COVID-19 prevention and treatment [10,11].

VC plays a role in a remarkable number of metabolic processes across the life kingdoms, but is more commonly known for its crucial functions in oxidative stress response, acting as a primary scavenger of reactive oxygen and nitrogen species which originated from exogenous (e.g., alcohol ingestion) and endogenous (e.g., mitochondrial metabolism) sources [12–16]. In this process, reactive species such as superoxide and hydroxyl radicals are reduced by VC to form one electron oxidized monodehydroascorbate (MDHA), which can be converted to fully oxidized nonionic dehydroascorbate (DHA) by interaction with other MDHA molecules in solution [1,13,17]. The subsequent oxidation steps of VC do not deplete the available VC pool in cells, as this nutrient can be regenerated

from either MDHA or DHA using a combination of enzymatic and non-enzymatic sparing mechanisms in plants, animals, and fungi [16–19], or directly as a by-product of MDHA dismutation [16,17]. However, VC can be irreversibly hydrolyzed to 2,3-diketo-L-Gulonic acid if not reduced in a timely manner by a Glutathione-dependent system [16,20,21]. VC oxidation with consequent formation of MDHA was also found to be responsible for the partial recycling of the lipid-soluble Vitamin E [22,23], an important membrane antioxidant known to act directly against fatty acid peroxyl radicals, thus preventing cell and organelle membrane damage [24–26].

VC is a prerequisite for normal neuromodulation and overall protection against neurodegeneration in vertebrate species [27–31], as well as hepatic and renal protection [32–35]. Furthermore, VC seems to be a key element in cancer prevention and treatment, as the suppression of free radical accumulation can ultimately retard or avoid the development of tumors [36–39]. Paradoxically, VC can act as a pro-oxidant when present in high concentration in the cellular environment [40], a chemical feature that can also be beneficial in potential cancer-treatment therapies [41–43]. In this context, the VC-induced oxidative stress appears to induce cancer cell apoptosis by itself, or synergistically in combination with other therapeutic agents [44–46]. This effect might not be, however, tumor-specific [47,48]. Nevertheless, since a very reduced number of patients treated with high intravenous VC concentrations have side effects [49], such treatment appears to be safe from a human health perspective. VC is also important in collagen biosynthesis, acting as a cofactor for the collagen stabilization enzymes prolyl and lysyl hydroxylases [50,51]. Thus, adequate levels of this nutrient are required to avoid the degradation of connective tissues in animal species [52,53]. This molecule has also been implicated in the modulation of epigenetic modifications such as DNA and histone demethylation, acting as a regulatory cofactor of ten-eleven translocation dioxygenases and Jumonji C domain-containing histone demethylases [54,55], and in the facilitation of non-transferrin-bound iron uptake [56,57].

Metazoa and fungi share a close evolutionary relationship and represent the two major groups of the eukaryotic Opisthokonta [58,59]. Nevertheless, in fungi, knowledge on VC functions is relatively scarce. Fungi species mainly rely on five-carbon VC analogs for optimal metabolic function [60,61]. The most common, D-Erythroascorbate, seemingly displays antioxidant behavior [60]. Nevertheless, the supplementation of VC can also improve fungal enzymatic antioxidant system activity and reduce the overall oxidative stress induced by mechanical lesions, when in moderation [62]. Furthermore, exogenous VC was also shown to confer considerable protection against free radical species originating from paraquat exposure in *Saccharomyces cerevisiae* [63]. In *Sclerotinia minor*, a fungi species where endogenous VC synthesis may occur, this nutrient appears to play a role in sclerotium biogenesis [64], and thus is likely involved in the adaptation of this species to environmental stresses.

In plants, VC plays an active role on overall stress perception and response [65–68]. Indeed, VC has been shown to ameliorate the negative effects of high salinity and heavy metal exposure in agricultural crops [67,69–72]. The protective aspects of VC are mostly relevant to the apoplast, a plant matrix considered to be the primary target of external abiotic or biotic factors in the roots [73,74], the chloroplast [75,76], and the organelle, where important physiological processes such as photosynthesis and lipid metabolism occur [77]. In simplistic terms, in response to stress factors that induce oxidative stress, such as exposure to ozone, VC can be released from the cytoplasm of plant cells to the apoplast, through the plasma membrane, where it becomes enzymatically or non-enzymatically oxidized, while scavenging radical species [73]. In the apoplast, where the conditions are considerably acidic (pH between 5 and 6) and dehydroascorbate reductase is absent, the originated MDHA is very unstable and the capacity to regenerate VC from DHA is impaired, and thus the main metabolite that is present is DHA [73,75]. Although the research is limited, this high concentration of DHA can potentially interfere in relevant physiological mechanisms, such as the regulation of protein activity in the plasma membrane and cell wall [73]. To maintain a regulated VC/DHA ratio in the apoplast, DHA diffuses to the cytoplasm and

is reduced back to VC [73,75]. In the chloroplast, VC intervenes as an electron donor for ascorbate peroxidases, which regulate oxidative stress by reducing hydrogen peroxide that results from the activity of the photosystems I and II through the "water-water cycle" [78–83]. VC is also a cofactor in the conversion of violaxanthin to zeaxanthin, catalyzed by violaxanthin de-epoxidase during the xanthophyll cycle, contributing to the protection of the photosynthetic apparatus from photodamage in excess light conditions [84–86].

Regarding prokaryotes, overall, VC is reported as not being required for normal metabolic functions [17]. Nevertheless, there is evidence for endogenous VC synthesis in these organisms [87–93], and some authors reported that VC can be decomposed and utilised for bacteria growth, particularly in the absence of other carbohydrate sources such as glucose [92]. Furthermore, the presence of VC also enables oxygen uptake in some *Bacterium* species [92], suggesting a protective role against increased oxidative stress. Therefore, the importance of VC for these basal organisms, might be more important than previously thought.

The current knowledge on VC molecular roles is likely to increase in the future, particularly as a way to further explore potential applications for human health and agriculture. Detailed knowledge on VC synthesis, in a wide variety of organisms, is also needed to explore such applications. Therefore, in the following sections, endogenous VC synthesis pathways are reviewed, with emphasis on animal species once thought to be unable to synthesize or even require this nutrient. In addition, the possibility of endogenous VC synthesis in prokaryotes will be addressed.

2. Vitamin C Biosynthetic Pathways: The Current Status

Although the origins of the divergent VC synthesis mechanisms in eukaryotes have not been completely elucidated, three main metabolic pathways have been well-described in photosynthetic protists, algae, plants and animals [94–101]. In addition, fungi species are known to synthesize VC analogs, such as D-erythroascorbate [102]. Due to the comprehensive reviews on the subject [17,68,80,94,103], and lack of significant recent advances, these will be only briefly discussed to highlight the main differences between pathways.

One of the distinctive features of the main VC synthesis pathways is the sugar precursor. Therefore, the main VC synthesis pathways are shown, not according to taxonomic relationships, but in order to highlight this point, as well as the sharing of the last enzyme of the pathway between photosynthetic protists and plants and green algae, and the sharing of the first enzyme of the animal and photosynthetic protists pathways (Figure 1). While plants rely mainly on D-Mannose as the main substrate for VC synthesis, animals and photosynthetic protists utilize D-Glucose [17]. Starting from different sugars, and by the use of divergent enzymatic mechanisms, plants and animals convert these sugars to L-Galactono-1,4-lactone and L-Gulono-1,4-lactone in the penultimate step of the pathway, respectively [17]. In photosynthetic protists, however, VC synthesis relies on a mix of animal and plant enzyme orthologs to confer the hybrid capacity for the conversion of D-Glucose to L-Galactono-1,4-lactone [17]. L-Galactono-1,4-lactone is ultimately metabolized to VC by the action of a L-Galactono-1,4-lactone dehydrogenase (GALDH) in plants and photosynthetic protists, while in animals, L-Gulono-1,4-lactone originates VC through an oxidation step catalyzed by L-Gulonolactone oxidase (GULO) [17]. It should be noted that some of the intermediate sugar substrate conversion steps in the animal and photosynthetic protists pathways are still hypothetical, since the putative enzymes involved in these steps have not yet been isolated and functionally characterized. For instance, the UDP-Glucuronidase that catalyzes the transition of UDP-D-Glucuronate to D-Glucuronate is thought to exist in vertebrate species but has not yet been identified [104]. Fungi species rely on the five carbon D-Arabinose sugar to synthesize D-erythroascorbate after two conversion steps catalyzed by D-Arabinose dehydrogenase and D-Arabinono-1,4-lactone oxidase (ALO), respectively [17].

In plants, alternative mechanisms in which myo-Inositol, L-Gulose and D-Galacturonate are used as precursors of VC synthesis, have been proposed [105–107]. They are based on

species-specific observations, and they have not been yet verified in a wider taxonomic context, and thus are not here discussed.

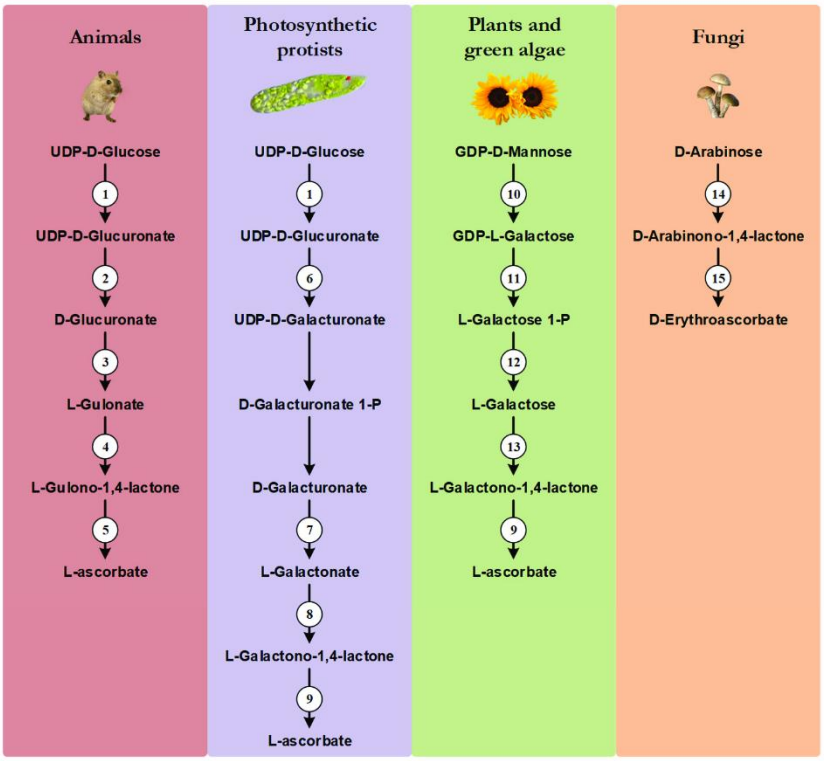


Figure 1. Graphical illustration of the main VC and D-erythroascorbate synthesis pathways, adapted from the MetaCyc database [108]. The numbered circles represent enzymes that are known to catalyze the substrate conversions within the pathways (see the enzyme list presented below the illustration). Reaction steps without an attributed enzyme are performed by unknown enzymes.

3. Alternative Animal VC Synthesis Pathways

The pseudogenization or even complete loss of the *GULO* gene is often reported as the primary cause of VC synthesis inability, in species such as humans, teleost fish, insects and nematodes [109–113]. Nevertheless, recent findings challenged the idea that the absence of a functional *GULO* gene can be used as a marker of VC synthesis inability in all animals [114–116]. In addition, although scarcely mentioned, the notion that D-Glucose is exclusively used for animal VC synthesis has also been challenged [117].

3.1. VC Synthesis in Caenorhabditis Elegans

C. elegans, where both *GULO* and *Regucalcin* (Regucalcin precedes GULO in the animal VC synthesis pathway) genes are missing, synthesizes VC [115]. Indeed, after exposing individuals to ^{13}C-labeled *Escherichia coli* as food source, ^{13}C was found in the VC pool, which led to the conclusion that VC must be synthesized de novo [115]. It was also observed that total VC content did not correlate with the concentration of sugar precursors commonly used by animals or plants, as well as their relevant derivatives [115]. Nevertheless, it is not clear that an increase in such sugar precursors would necessarily lead to

an increase in VC synthesis [115]. In addition, a comprehensive search for homologs of the enzymatic constituents of the plant and animal VC synthesis pathways in *C. elegans* was also performed, revealing that no pathway was present as a complete set [115]. The loss of the *GULO* and *Regucalcin*, is also observed in other nematode species [118]. Given the evidence, it was suggested that VC synthesis in *C. elegans* likely occurs through a yet undescribed metabolic pathway [115]. No enzymatic constituents have been proposed for this pathway [115].

3.2. VC Synthesis in Drosophila Melanogaster

In insects, the need for VC synthesis, or adequate overall VC levels, varies depending on the species group that is considered [119]. In species from the Diptera, Lepidoptera and Coleoptera taxonomic groups, VC plays an important metabolic role [119]. Indeed, VC synthesis has been observed in the dipteran *D. melanogaster* [120] and in the lepidopteran *Bombyx mori* [116]. *Massie* et al. [120] were the first to note that *D. melanogaster* Oregon-R and Swedish-C male flies maintained for over 10 generations in food and water sources deprived of VC, displayed measurable total body VC concentrations. Moreover, by observing the continuous decrease in VC content in flies with age, these authors suggested that VC synthesis happened during the developmental stages, rather than in adult individuals [120]. Nevertheless, the amount of total VC increases in adult flies after exposure to a brief cold shock [120], indicating that in *D. melanogaster* VC synthesis might be also involved in stress response mechanisms. The authors evaluated the impacts of supplementation with L-Gulonolactone, D-Galactonolactone, and Glucose in adult flies, but none of these substrates contributed to increased VC content after the cold shock stimulus [120]. As the authors discussed, the enzymatic mechanisms might simply not be stimulated by dietary precursors [120]. This would suggest that *D. melanogaster* is able to synthesize VC precursors above the limiting rate of consumption, and that this self-sufficiency overshadows their addition. The authors, however, did not address the role of the flies' gut microbiome in the synthesis process. This assessment is essential to determine the source of the VC found in the whole body homogenates, as several microbes have been found to synthesize and supply vitamins to hosts incapable of endogenous synthesis [121], and VC synthesis by insect symbionts has been previously observed [122].

Based on these findings, *Henriques* et al. [114] have demonstrated that axenic *D. melanogaster* Oregon-R flies possess similar levels of VC relative to normal individuals, and that the ex vivo culture of the microbiome did not lead to any VC accumulation in the resulting bacterial pellet and supernatant. Thus, Oregon-R flies' microbiome does not significantly contribute to VC synthesis [114]. The impact of flies' gender and sexual activity, not considered in [120], was also evaluated [114]. VC content in virgin flies had no significant difference relative to mated flies, but females had approximately three times more VC than males, and this observation could not be exclusively explained by the different bodyweight between genders [114]. This result implies that female flies can demonstrate increased amounts of VC due to: (i) the overexpression of the VC synthesis pathway; (ii) the presence of gender-specific structures, such as the ovaries; and (iii) increased VC storage and retention. Plasma VC concentrations are also higher in human and mouse females [123,124], but by a slight margin when in comparison to what is observed in *D. melanogaster*. The VC difference observed in mice may be explained in part by distinct renal VC excretion rates between males and females, since females are able to retain more VC from urine than males [125]. Therefore, gender-specific VC excretion rates along the *D. melanogaster* Malpighian tubule system, as well as the 50% difference in bodyweight between the sexes [114], might account for the observed results. These hypotheses, however, remain to be experimentally addressed.

It has been suggested that the physiological responses of insects can vary considerably depending on the cold exposure conditions [126]. Thus, as a complement to the brief cold shock exposure experiment reported in [120], *Henriques* et al. [114] verified the effects of extended cold exposure on the VC content of *D. melanogaster* Oregon-R. Unlike

what was observed in [120], flies acclimated one day at 15 °C had a 23% decrease in VC levels relative to flies kept at 25 °C [114]. Nevertheless, cold-acclimated flies were able to replenish base VC levels after one day of recovery at 25 °C [114]. Therefore, different cold exposure conditions impose contrasting changes in *D. melanogaster*'s VC metabolism. Interestingly, the gene that encodes the lactonase that precedes GULO in the animal VC synthesis pathway, *Regucalcin*, has a duplicate in Sophophora species such as *D. melanogaster*, called *Drosophila cold acclimation* (*Dca*) [127,128]. The functions of this gene are still being determined, but so far they appear to be correlated with the response to cold exposure and wing size variation [127–130]. In regard to cold exposure, *Dca* has contrasting expression changes depending on the applied stress, since it is upregulated under acclimation conditions [126,128], but downregulated under cold shock [131,132]. This information implies that *Dca* expression changes are likely not correlated with the VC synthesis changes that occur under cold exposure responses in Sophophora species. Nevertheless, without further experimental data, the involvement of Dca in VC synthesis under cold exposure conditions cannot be yet ruled out. A direct involvement of Dca or Regucalcin in VC synthesis would imply that at least the lactonase component of the animal pathway is conserved in *D. melanogaster*. Using an in silico methodology, both *Regucalcin* and *Dca* were inferred to be involved in calcium homeostasis and in the oxidative stress response, but so far, no evidence was found for its involvement in VC synthesis [118]. The involvement in the calcium homeostasis and in the oxidative stress response are expected given the multiple functions of *Regucalcin* in vertebrates (reviewed in [133]). The inference of a divergent protein interaction surface at Dca in comparison with Regucalcin, along with the detection of positively selected amino acids near and within the active site region, suggests that Dca might have also acquired novel functions involving its lactonase activity and overall contribution to protein complexes [118]. The constitutive suppression, under the actin 5C driver, of either *Regucalcin* or *Dca* in *D. melanogaster,* did not clarify the role of these genes in VC synthesis [118]. Indeed, both genes are required for *D. melanogaster* viability, since in both cases over 70% of the individuals with suppressed expression were only able to reach the pupal stage [118], and thus the biological processes in which they are involved could not be determined. The reduced viability could be attributed, for instance, to the dysregulation of calcium homeostasis/signaling, as stated in [118]. Nevertheless, a cascade of metabolic events caused by VC deprivation could also explain the negative effects, especially when considering the previous evidence that VC synthesis occurs mainly during the developmental stages [120], which coincides with the lethal phase observed in the fly RNAi experiments [118]. In *B. mori*, a *Regucalcin*-like gene has been inferred to be involved in VC synthesis [116] (see below).

3.3. VC Synthesis in B. mori

In the lepidopteran *B. mori*, where *GULO* is absent, an alternative pathway for VC synthesis has been proposed, based upon the identification of a protein with GULO-like activity [116]. Evidence for VC synthesis comes from the observation that during silkworm development, the concentration of this nutrient is stable or even increased during stages without an active consumption of VC from mulberry leaves [116]. As in *D. melanogaster* [114], the authors addressed the impact of the microbiome, and found that it does not significantly contribute to VC synthesis [116]. Furthermore, the contribution of VC regeneration mechanisms in VC content fluctuations was also dismissed, since DHA levels remained unaltered during egg development [116]. Thus, it was hypothesised that *B. mori* is capable of endogenous VC synthesis [116]. Using a bioinformatics approach, *Hou* et al. [116] were able to identify two candidate GULO-like sequences (BGIBMGA012624 and BGIBMGA005735). Because BGIBMGA005735 codes for a protein that has a domain with similarity with another aldonolactone oxidoreductase (AO), it was selected for further analyses [116]. Enzymatic activity assays showed that this GULO-like protein is capable of VC synthesis using the same substrate as GULO, and that its activity is in part modulated by physiological changes that occur in the fat body during pupae development, implicating

this tissue in the process [116]. This information is intriguing, as the *B. mori* fat body is considered to be analogous to the vertebrate liver and adipose tissue [134], and VC synthesis is known to occur in the liver of higher vertebrates [109,135]. Nevertheless, it is believed that in vertebrates, VC synthesis transitioned from the kidney to the liver during evolution and not the opposite [109], as now suggested by the observation that in protostomians the tissue responsible for VC synthesis is the structure that is analogous to the liver. It is, however, conceivable that VC synthesis might have transitioned from the kidney analogous tissue to the fat body in insects, in a manner similar, but independent of what is observed in vertebrates. In *B. mori*, the siRNA suppression of the *GULO-like* gene expression in the pupae fat body decreased VC synthesis but was ultimately non-lethal, while in the egg stage it considerably reduced overall survivability and led to delayed hatching [116]. Thus, although not ubiquitously essential for the complete growth of silkworms, this gene appears to have relevant functions implicated in viable transitioning from the egg to the larval stage [116]. The direct correlation between the reduced fitness of silkworm eggs and VC levels is, however, missing in that study. The authors note, however, that VC synthesis associated with the activity of the GULO-like protein studied, is not sufficient to fulfil the nutritional needs of developing individuals without additional supplementation [116]. The *B. mori* GULO-like protein is the ortholog of the vertebrate Delta(24)-sterol reductase (DHCR24) [136,137]. Like the vertebrate GULO, sterol reductases belong to the vanillyl alcohol oxidase/para-cresol methylhydroxylase (VAO/PCMH) flavoprotein family, but are usually known to catalyze distinct reactions [138]. While the DHCR24 protein is known to catalyze the NADPH-dependent reduction of a carbon–carbon double bond during the synthesis of cholesterol [138], using as substrate desmosterol that derives from endogenous [139] or dietary [140] sources, GULO is known to catalyze the oxidation of L-Gulonolactone, using O_2 as an electron receptor, in the synthesis of VC [138,141]. In agreement with the functional divergence, phylogenetic studies also show that sterol reductases and AOs have diverged early in evolution [138]. Thus, it was not expected that a sterol reductase could metabolically replace GULO.

Considering that *B. mori* depends exclusively on mulberry leaves for nutrition purposes [142,143], and the VC content in these tends to differ depending on factors such as the climate or plant variety [144], the novel DHCR24 GULO-like functions might be the result of selective pressures to enhance VC availability during the development stages of this specific organism. Assuming that the reduced viability of silkworm eggs was caused by lower VC levels after *DHCR24* knockout, this evolutionary perspective could explain why VC synthesis mediated by DHCR24 seems to play an important but not essential role in silkworm growth and fitness. This interpretation correlates well with previous data that imply that some enzymes involved in insect sterol metabolism have likely evolved additional functions throughout evolution [145]. Nevertheless, the available data do not rule out the diminished sterol reductase activity as the partial or complete cause of the increased egg lethality.

Hou et al. [116] proposed that *B. mori* synthesizes VC through an adapted animal pathway in which the identified GULO-like enzyme functionally replaced GULO in the catalysis of the final oxidation step. This hypothesis appears to be the most parsimonious, since although GULO was lost in *B. mori*, the genes that encode enzymes that catalyze upstream enzymatic reactions in the animal pathway, such as *Regucalcin*, are present in this organism [118]. It should be noted, however, that this GULO-like enzyme is not present in *D. melanogaster* (by performing a "blastp" search (https://blast.ncbi.nlm.nih.gov/Blast.cgi, accessed on 28 July 2022) using the *B. mori* DHCR24 sequence (XP_004926865.1) as query, provided no homologous sequence hits in *D. melanogaster)* and thus, the *B. mori* VC pathway evolved independently from that of *D. melanogaster*. The nomenclature used in *Hou* et al. [116] can be confusing, since the authors refer to this animal modified pathway as the L-Gulose pathway, and state that it is analogous to the one found in plants. The proposed plant L-Gulose pathway has partial similarity with the animal D-Glucose pathway, since they both rely on L-Gulonate as intermediate substrate [146]. They differ,

however, in the molecular mechanisms that lead to this step, as in the characterized animal pathway this molecule derives from D-Glucuronate and in plants from L-Gulose [146,147]. The metabolism of L-Gulose seems to be restricted to plants and archaea species [148], since the formation of L-Gulose likely involves a reaction catalyzed by a GDP-Mannose 3′,5′-Epimerase (GME) [107,148–150], and such epimerase activity was only found associated with the conversion of GDP-D-Mannose to GDP-L-Galactose in animals [151]. Considering these observations, and that naturally occurring L-Gulose is considerably rare [152], VC synthesis in *B. mori* likely does not rely on this precursor, and thus the nomenclature referred to in [116] should be interpreted with caution. Nevertheless, since plant GME can convert GDP-D-mannose to a mixture of GDP-L-Galactose, GDP-L-Gulose and GDP-D-Altrose [150], the contribution of enzymes with GDP-Mannose 3′,5′-Epimerase activity to protostomian L-Gulose metabolism, such as the GDP-4-keto-6-deoxy-D-mannose 3,5-epimerase/4-reductase of *D. melanogaster* [153] and its homologs in other species, cannot at present be excluded as playing a role in VC synthesis in protostomian species that do not use the described animal synthesis pathway.

3.4. The Role of Myo-Inositol in Animal VC Synthesis

Like in plants, the role of myo-Inositol as a precursor for animal VC synthesis through a partially alternative pathway has been addressed. The animal myo-Inositol Oxygenase (MIOX) converts this sugar to D-Glucuronate [154,155], which is a key molecule in the glucuronate-xylulose pathway [156,157] and an intermediate precursor in the vertebrate VC synthesis process ([17]; Figure 1). Experiments in *Cavia porcellus* (guinea pig) first suggested that myo-Inositol played an important role in VC synthesis or sparing [158], but later studies showed that this was not the case [159]. Indeed, there is no functional *GULO* gene in *C. porcellus*, and this is why this species is highly regarded as a scorbutic model organism when missing a dietary source of this nutrient [53,160,161]. The auxotroph nature of the guinea pig should have been considered a limiting methodologic factor in addressing the role of myo-Inositol in animal VC synthesis, as the absence of a functional VC synthesis pathway could invalidate the study of potential precursors in this organism. This criticism stands even considering that this polyalcohol is naturally present in the guinea pig [162,163] in sufficient concentration for viable VC synthesis [117]. The conclusions referred in [159] were thus, expected. In this sense, teleost species are also scurvy prone in the absence of VC supplementation [164,165], and thus should be avoided as experimental subjects in further explorations on the subject.

Myo-Inositol was found not to be involved in VC synthesis in *Rattus norvegicus* (rat) [166]. Unlike the guinea pig, rats are capable of synthesizing VC using the characterized animal pathway [167]. In placental mammal species capable of doing so, VC synthesis occurs in the liver since it relies on the hepatocyte glycogen reserves [168], and GULO is exclusively found in this tissue [109,135,169]. The D-Glucuronate needed for the process can be obtained as the by-product of the glucuronidation of endobiotic and xenobiotic compounds in the mammals liver [104,170,171], as well as the conversion of UDP-D-Glucuronate to D-Glucuronate by a putative UDP-glucuronidase [104] (see Figure 1). However, myo-Inositol is only metabolized in the vertebrate kidney [172], and, as expected given this observation, MIOX appears to be a kidney-specific enzyme [155,173]. The D-Glucuronate originated by the action of MIOX is likely mostly used in the glucuronate-xylulose pathway that is particularly active in the vertebrate kidney [174]. Therefore, although it is conceivable that myo-Inositol could play a role both in VC synthesis and in the glucuronate-xylulose pathways [175,176], it is likely that only a small amount of D-Glucuronate diffuses from the kidney to the liver, thus playing a negligible contribution in the context of VC metabolism in mammals that synthesise VC in the liver.

The observations made in rats may not apply to all mammal species capable of endogenous VC synthesis. Indeed, in egg-laying mammals (Monotremata) GULO is exclusively found in the kidney [177]. In the Peramelina order (Marsupials) GULO is found in both the liver and kidney, although in species from the Diprotodonta order (Marsupials),

GULO is mainly found in the liver [177]. Therefore, in basal mammal groups, besides D-Glucose, myo-Inositol may also be a relevant substrate for VC synthesis in species where GULO is found in the kidney.

Hänninen et al. [117] compared the capacity for VC synthesis from myo-Inositol in the Leghorn hen (*Gallus domesticus*) kidney and rat liver. In [117], the findings of [166] were replicated, as rat liver extracts displayed no MIOX activity and were unable to synthesize VC from myo-Inositol [117]. However, the analyses performed on the galliform species capable of VC synthesis in the kidney [178] showed opposite results [117]. The authors were able to determine that ^{14}C-labelled myo-Inositol was incorporated into the hen kidney VC pool, in a manner partly dependent on the MIOX ability to convert this sugar to D-Glucuronate [117]. In addition, the results also showed that the concentration of this polyalcohol in hens is sufficient for biologically relevant VC synthesis [117]. Not all birds have, however, GULO in the kidney, an essential requisite for VC synthesis in this organ. For instance, Passeriformes species synthesize VC in the liver [117].

In amphibians [179,180], reptiles [179,180] and basal fish groups [179,181], GULO is only found in the kidney as well. Therefore, although generally disregarded, the contribution of myo-Inositol to VC synthesis in animals, particularly in ancestral vertebrate species, cannot be excluded.

3.5. Summary and Future Perspectives

In this section we put forward hypotheses regarding VC synthesis pathways in nematodes and insects (Figure 2), and present possible methodological approaches for further exploration of this subject. The available information strongly suggests that VC synthesis in *C. elegans* occurs by a very distinct pathway from the one used by most animals, since it may not rely on the known substrates of the described pathways as precursors [115], and the two key enzymes in the process, namely Regucalcin and GULO [115,118], are also missing (see detailed discussion above). The precursor supplementation assays have some experimental limitations [115], that can be overcome using radiolabeled compounds, since this methodology allows the direct visualization of substrate incorporation and excretion rates [182]. In addition to the classical substrates, the assays could also be expanded to include myo-Inositol, since the vertebrate *MIOX* gene has an ortholog in *C. elegans* (*C54E4.5*; Ensembl accession number WBGene00016920), and the D-Glucuronate that is generated could contribute to VC synthesis as it happens in birds (Figure 2). The use of a complete or partial knockout of *C54E4.5* could also be used to address the involvement of this gene in VC synthesis by looking at VC levels that can be easily measured using the methodology described in [114].

The report of an AO function for the *B. mori DHCR24* sterol reductase [116] also raises questions regarding possible adaptations of non-AO VAO/PCMH flavoproteins for VC synthesis (Figure 2). *C. elegans* contains a functional *DHCR24* ortholog [183,184] that should be also studied in detail.

As in *C. elegans*, in flies, the supplementation of food with precursors naturally used in the animal VC synthesis pathway did not contribute to increased VC content [101] (see detailed discussion above). However, as suggested for *C. elegans*, the use of radiolabeled substrates is likely more adequate to address this issue. Such approach could be particularly relevant in deciphering the role of myo-Inositol in *D. melanogaster* VC metabolism, since a *MIOX* ortholog has been identified in this organism (*CG6910*; [185]; Figure 2). In *D. melanogaster,* under cold acclimation conditions, the expression of the *CG6910* gene is reduced [126]. *Dca* is overexpressed under cold acclimation conditions [126], but VC levels decrease during this stimulus [114], which may be due to the suppression of *CG6910* expression during the cold stress response.

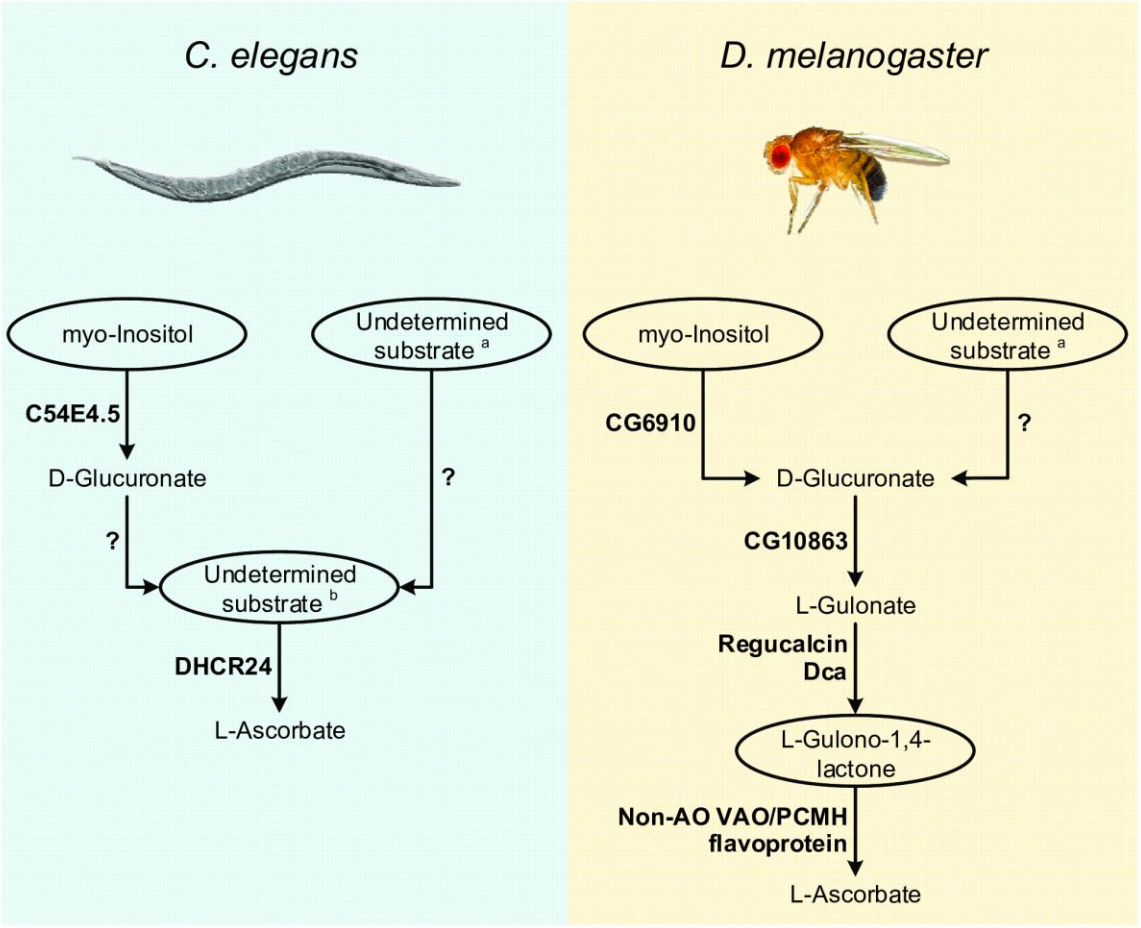


Figure 2. Inferences on enzymatic mechanisms and possible substrates utilized by *C. elegans* and *D. melanogaster* in the VC animal synthesis pathways. The ellipses highlight the main substrates possibly used in the proposed pathways, but that remain to be addressed using appropriate experimental setups. Enzymes that may be correlated with VC synthesis in these species are highlighted in bold next to the metabolic steps they are likely to catalyze. The interrogation marks indicate uncertainty in the steps regarding mechanisms of conversion of uncharacterized substrates, and the relative contribution of these processes to the availability of VC precursors. Common sugars (such as D-mannose, L-Galactose, D-Glucose and D-Arabinose) are assigned with [a] while L-Galactono-1,4-lactone, L-Gulono-1,4-lactone and D-Arabinono-1,4-lactone are assigned with [b].

Both Dca and Regucalcin likely catalyze the conversion of L-Gulonate to L-Gulono-1,4-lactone in *D. melanogaster*, and as such, if Dca is involved in VC synthesis, through a pathway resembling the one found in vertebrates, then the availability of its substrate might be a limiting factor for VC synthesis during cold acclimation conditions. L-Gulonate is obtained by the transformation of D-Glucoronate by the action of the aldo-keto reductases AKR1A1 and AKR1B1 in mammals [186,187] (Figure 1). The catalytical contribution of the two enzymes is very distinct, as AKR1A1 and AKR1B1 are responsible for approximately 85% and 15% of the conversion ratio, respectively [187]. The reduced contribution of AKR1B1 can be explained by its reduced specific activity rather than substrate specificity [188], reflected in the reduced efficiency for D-Glucuronate reduction when compared to AKR1A1 [187]. The human *AKR1A1* gene has no orthologs in *D. melanogaster* (Ensembl accession number ENSG00000117448), but the *AKR1B1* gene displays a one-to-many or-

tholog relationship (Ensembl accession number ENSG00000117448). One of these homologs, namely *CG10863*, encodes an enzyme known to share catalytic similarities to AKR1A1 and AKR1B1 [189]. This gene is under-expressed during cold acclimation, and D-Glucoronate is notably more abundant in cold acclimated flies [126]. Therefore, we hypothesize that under the mentioned stress conditions, VC synthesis might be impaired by the reduced expression levels of both *CG6910* (see above) and *CG10863*, the latter leading to the accumulation of D-Glucoronate in flies. Since, in contrast to *Dca,* that is overexpressed, *Regucalcin* is also under-expressed under cold acclimation conditions [126], when flies are placed in recovery conditions, along with the restoration of normal *CG6910* and *CG10863* expression levels, the role of Dca might be important for the rapid increase in VC synthesis, by providing a boost to the conversion of L-Gulonate to L-Gulono-1,4-lactone. Nevertheless, since Regucalcin and Dca are likely to perform important roles in calcium homeostasis as well [118], they may play also other important roles during cold exposure recovery. The regulation of calcium levels might be important for the maintenance of viable muscle function [126] in a manner unrelated to *D. melanogaster* VC synthesis. Thus, even if Dca does not intervene in VC metabolism, the reduced expression of *CG10863* and *CG6910* and *Regucalcin* under cold acclimation, and subsequent normalized expression on recovery conditions might be sufficient to explain the results obtained in [114]. While further studies on these genes may give insight into the VC synthesis pathway, there is still no candidate for the functional replacement of the *GULO* gene that has been lost in insects, such as *D. melanogaster*. Indeed, this species is a cholesterol auxotroph [190,191], that has lost the ability to convert desmosterol to cholesterol [192], likely due to the loss of the *DHCR24* orthologous gene (the one involved in VC synthesis in *B. mori* [116]).

4. Prokaryote AOs: An Unexplored Concept

4.1. The Role of AOs in VC Synthesis Evolution

The divergent VC synthesis mechanisms here discussed suggest that the majority of species rely on AOs to catalyze the final metabolic step. The distribution pattern of AOs in eukaryotes is complex (Figure 3), and the evolutionary history that led to this scenario is still lacking definition. A very comprehensive study has previously explored the emergence of distinct VC synthesis pathways based upon the eukaryote AOs dispersion, while evaluating the contribution of putative lateral gene transfer events in the process [94]. Although the evolutionary origin of the plant *GALDH* could not be inferred, it was hypothesized that *GULO* (used in this context as the representative of both *GULO* and *ALO*) and *GALDH* could be the result of an AO gene duplication event that occurred in the ancestral eukaryote (Figure 3) [94]. Alternatively, it was also proposed that *GALDH* might have arisen in the Archaeplastida ancestral (Figure 3), and ultimately replaced *GULO* in many non-glaucophyte lineages due to the increased oxidative stress imposed by the primary endosymbiosis event that led to chloroplast emergence [94]. The authors concluded that this was the most parsimonious explanation for *GALDH* emergence, considering that VC synthesis catalyzed by GALDH does not contribute to ROS production, and thus increases the availability of anti-oxidant mechanisms to deal with ROS originated by chloroplast metabolism [94]. Then, possible secondary endosymbiotic gene transfer (EGT) events could have been responsible for the transfer of *GALDH* to the SAR and Euglenida lineages, while horizontal gene transfer (HGT) events were the likeliest cause of *GALDH* in the opisthokont Choanoflagellata [94]. Although a putative *GALDH* was uncovered in the Chytridiomycota fungi *Gonapodya prolifera* [94], there was no discussion on the acquisition mechanism responsible for the presence of this AO in this taxonomic lineage. Nevertheless, considering the proposed hypothesis, this might also be the outcome of a putative HGT event, as the Chytridiomycota are aquatic species inferred to have interacted with streptophyte algae throughout evolution [193].

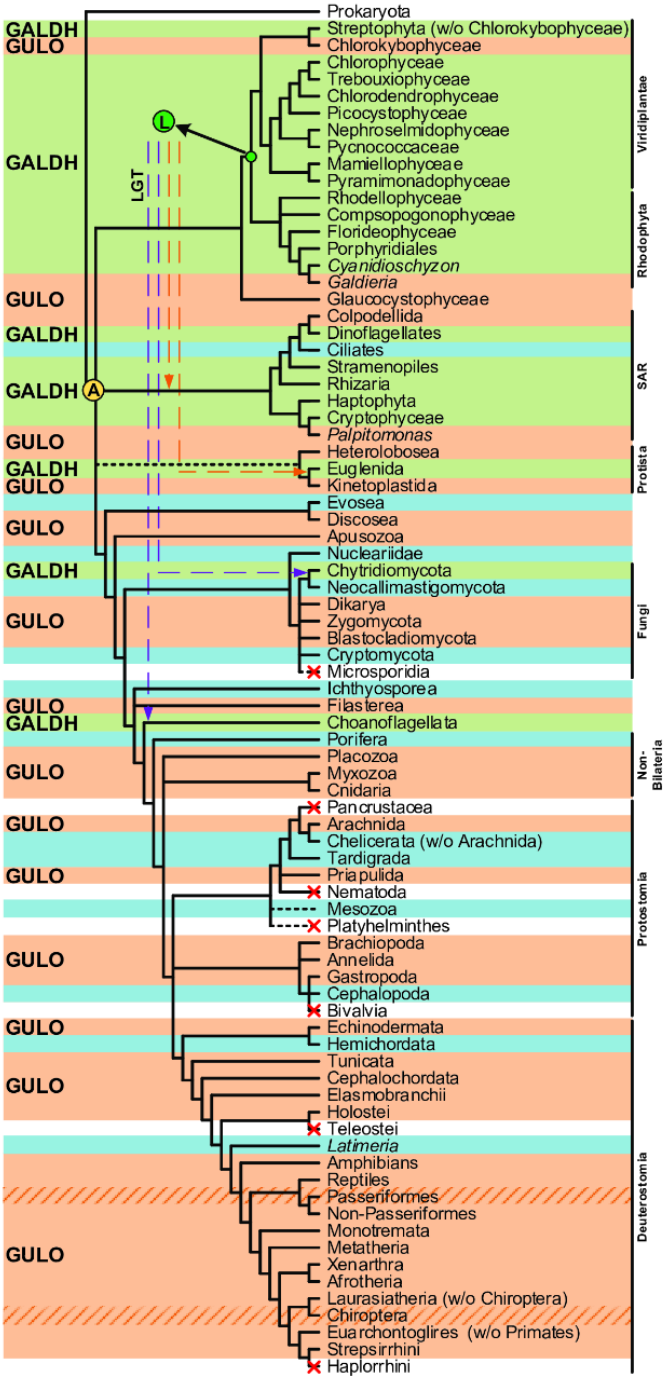


Figure 3. Presence/absence pattern of the *GULO*/*GALDH* lineages in eukaryotes. The taxonomic groups where the *GULO* lineage is present are highlighted by an orange background, while those where the *GALDH* lineage is present are highlighted by a green background. Independent *GULO* loss events within specific vertebrate groups are represented by stripe patterns. The blue background highlights uncertainty regarding either gene lineage presence/absence. These cases represent groups where both gene lineages could not be found in less than three species, or groups where the sequences

identified were likely derived from genome contamination. Taxonomic groups where either gene lineage was lost are not highlighted, and a red cross can be seen in overlapping the branches they represent. Higher taxonomic ranks were added to the right of the figure to facilitate the interpretation. The ancestral paralogy and lateral gene transfer hypotheses proposed in [94] are indicated by the yellow and green filled circles with an A and L acronyms over the relevant tree nodes, respectively. The implied duplication events occurred during the time scale represented by the branches that led to the highlighted nodes. Both gene lineages already existed in scenario A, and the pattern of loss/presence of each was based upon selective constraints of the distinct taxonomic groups. Scenario L implies the origin of the *GALDH* lineage in the common ancestor of the Rhodophyta and Viridiplantae, that was later transferred to other taxonomic groups by endosymbiotic (orange dashed arrows) or horizontal (dashed purple arrows) gene transfer events. The data depicted in the figure was obtained from [94,112,114], and the taxonomic relationships are based on those found in [194,195] and the Tree of Life Web Project (http://tolweb.org/tree/; [196]; last accessed on 12 October 2022).

Both hypotheses consider that the ancestral eukaryote had at least one functional AO to catalyze the last step of VC synthesis. This assumption raises additional questions regarding the emergence of AOs, as it does not clarify if they have prokaryotic origins or are the result of the evolution of de novo genes that arose in the ancestral eukaryote. In an extreme scenario, it could be proposed that *GULO*, *GALDH*, or even both genes, did not exist in the eukaryote ancestor, and were later on acquired through EGT from the alphaproteobacteria and cyanobacteria that are thought to have led to the emergence of mitochondria and chloroplasts, respectively [197,198]. This is a viable hypothesis, considering that some of the genomic information of these organelles has been transferred to nuclear DNA throughout evolution [199–201].

The possibility of endogenous VC synthesis in bacteria is often disregarded by many scientific publications on the subject, and this may be the reason why the acquisition of AO genes in the ancestral eukaryote by EGT is not considered as a current alternative scenario. However, although relatively scarce, evidence for prokaryote VC synthesis has been published before [87–93,122]. In some cases, the authors were able to characterize L-Gulono-1,4-lactone Dehydrogenases (GUDH) as the likely AOs that catalyze the final step of synthesis of this nutrient in bacterial species [87,90]. These GUDH have high substrate specificity to L-Gulono-1,4-lactone and do not metabolize L-Galactono-1,4-lactone [202,203], and therefore have more functional similarities to the animal GULO and fungi ALO proteins than to plant GALDH. However, they are unable to use oxygen as an electron acceptor, but can use cytochrome C to this effect [203], bearing some similarities to the plant GALDH. Based on the hybrid functional characteristics of bacterial GUDH, if such an AO of prokaryote origin is related to the emergence of VC synthesis in the ancestral eukaryote, then the *GULO* and *GALDH* lineages could be the end result of adaptation to the cell conditions present in the distinct taxonomic groups. This hypothesis implies a single gene origin for the various AO lineages observed in eukaryotes and is compatible with the ancient paralogy scenario presented in [94], under the assumption that this AO was duplicated before the emergence of the main eukaryote lineages. However, *Arabidopsis thaliana* has a functional GUDH enzyme with similar characteristics to the GUDH found in *Mycobacterium tuberculosis* [202,203], thus implying that some Streptophyta can have both GUDH and GALDH. This could mean that the *GALDH* lineage might have derived from a duplication of *GUDH* in the Archaeplastida after the divergence from glaucophytes (Figure 4A), thus reinforcing the role of LGT in the evolution of AO [94]. Nevertheless, if *GALDH* and *GULO* derived from a *GUDH* duplication in the ancestral eukaryote, the identified *A. thaliana GUDH* could be a remnant of an ancestral prokaryote gene (Figure 4A), an hypothesis that correlates well with its phylogenetic outgroup positioning relative to the *GALDH* and *GULO* gene clusters in previous phylogenetic analyses [94,203].

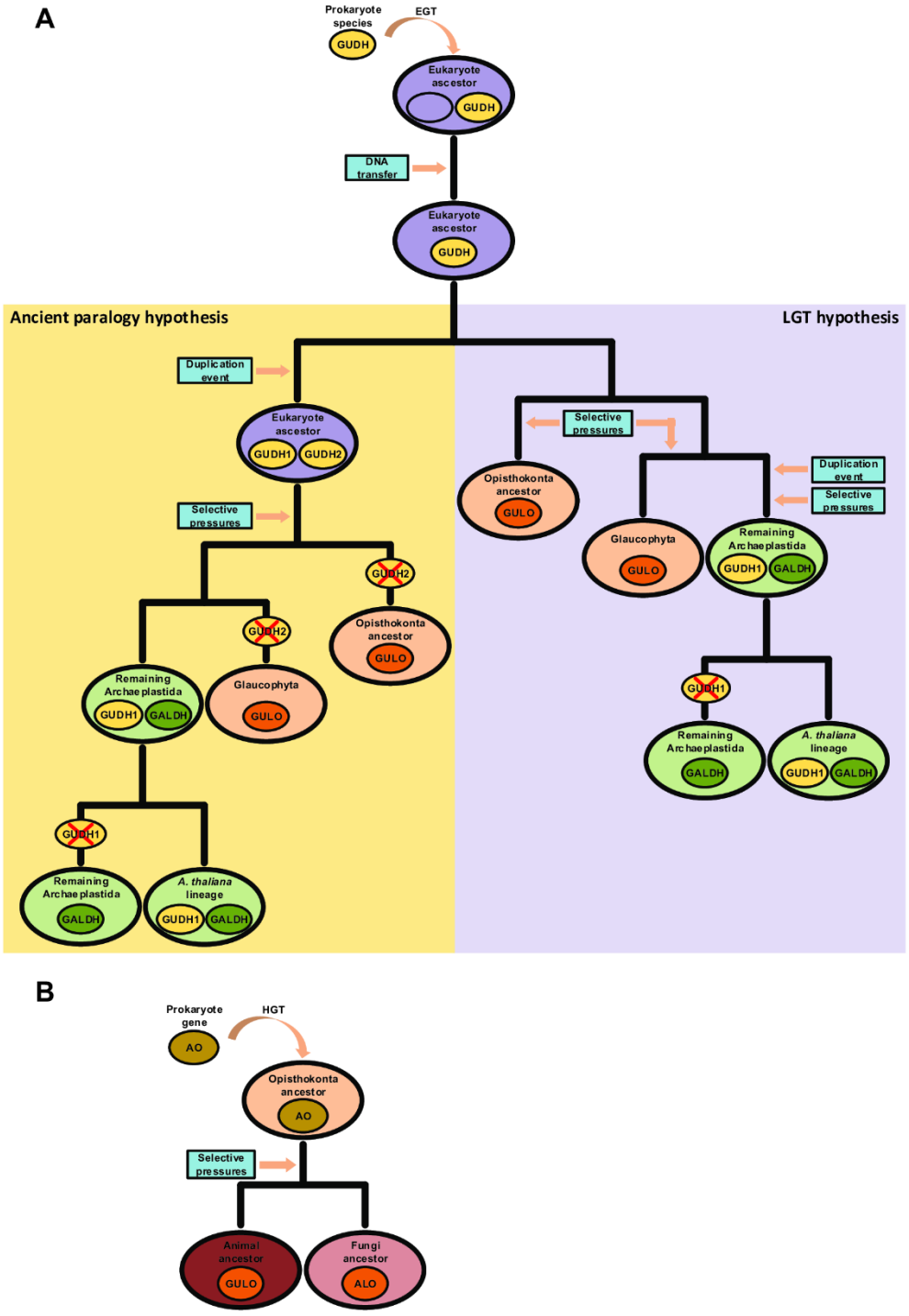


Figure 4. Putative scenarios of eukaryote AO evolution derived from prokaryote LGT events. In panel (**A**), a prokaryote *GUDH* was initially integrated in the genome of the ancestral eukaryote after an EGT event. Two hypotheses for the divergence of *GUDH* in eukaryotes are presented. In the left (yellow background), the *GUDH* present in the ancestral eukaryote was duplicated, originating *GUDH1* and *GUDH2*. Distinct selective pressures led to the loss of the *GALDH* ancestral gene

(*GUDH2*) in glaucophytes and opisthokonts, and thus *GULO* is the result of *GUDH1* evolution. In the non-glaucophyta Archaeplastida, *GALDH* was the result of selective pressures imposed on *GUDH2*, while *GUDH1* maintained the ancestral state. During evolution, *GUDH1* was eventually lost in taxonomic groups that diverged from the one where *A. thaliana* is included. This scenario is compatible with the ancient paralogy hypothesis presented in [94], as indicated in the figure. In the right (purple background), the scenario represents the direct origin of the opisthokont and glaucophyte *GULO* lineages as the result of selective pressures imposed on the ancestral *GUDH*. In this case, a *GUDH* duplication event affected only the Archaeplastida ancestor after the divergence from the glaucophytes, and the *GUDH1* and *GALDH* duplicates are the result of distinct selective pressures. Then, *GUDH1* was lost in taxonomic groups that diverged from the one where *A. thaliana* is included. This scenario is more compatible with the LGT hypothesis presented in [94], as indicated in the figure. In panel (**B**), *GULO* and *ALO*, in animals and fungi, respectively, are the result of selective pressures imposed on an AO acquired through HGT by the ancestral opisthokont.

If AOs had a prokaryotic origin, bacteria to eukaryote HGTs might also have played a role in the emergence of some VC synthesis mechanisms, since HGT events are known to have been important in the evolutionary adaptation of eukaryotes [204,205]. Such events appear to be widespread, as evidence for such phenomenon has been reported in protist [206], plant [207], fungi [208] and animal [209] species. HGT events seem to be more common in fungi than in animals and plants [210], and in some cases, fungi appear to have acquired entire metabolic pathways [211,212]. In the context of VC metabolism, several fungi species are able to synthesize VC analogues, such as D-erythroascorbate, through the two-step conversion of D-arabinose, L-fucose, L-xylose or L-galactose [102,213]. Interestingly, older reports have suggested that bacterial species are also capable of synthesizing a molecule with the antiscorbutic properties of VC from xylose, although without conclusive identification of the compound produced at the time [92,93]. Bacteria are known to have very few introns across their genome [214], and the fungi *S. cerevisiae ALO* gene is encoded by a single exon (EnsemblFungi accession number YML086C). Nevertheless, this species is known to have a particularly low intron presence across the genome when compared to other species [215,216], and thus such information cannot be used as evidence of HGT. However, other examples are available, such as the single exon *ALO* gene of *Scheffersomyces stipites* (EnsemblFungi accession number PICST_88335) and *Aspergillus nidulans* (EnsemblFungi accession number ANIA_00836). These fungi sequences might represent remnants of assimilated prokaryotic DNA. Considering the information presented, it would not be implausible to assume that the fungal D-erythroascorbate biosynthetic pathway could be related with the complete or partial acquisition and conservation of a pre-existent bacterial pathway in the Opisthokonta ancestral, which later on diversified after speciation (Figure 4B). To our knowledge, no study was yet performed to evaluate the presence and distribution of AOs in prokaryotes, and since phylogenetic methods can help track HGT events [217], we find that there is still much more to explore in this respect.

Although speculative, this discussion clearly highlights the limitations imposed by the exclusion of prokaryotes from VC synthesis studies. As a preliminary step, the inference of putative AOs in such basal lineages could provide the foundation for further findings on this subject.

4.2. AOs Presence in Archaea and Bacteria

In [94], the authors state that an extensive homology search led to the inference of only one significant *GULO* homolog in prokaryote species. The initial search was performed using the *M. musculus* GULO and *A. thaliana* GALDH sequences as query in a "blastp" approach [94]. Nevertheless, since *ALO* sequences represent a distinct lineage within the *GULO* clade [94], the use of a representative ALO protein sequence as query can still improve the likelihood of encountering additional homology hits. Moreover, the rapid increase in genome availability for uncultured bacteria [218], as well as the use of

more sensitive methods to detect remote (distant) homologs than BLASTp, may help fill some gaps in prokaryote AOs distribution patterns. Such data are needed to confidently address the HGT hypotheses discussed above, and should the subject of future research. Nevertheless, a "blastp" search, coupled with Bayesian phylogenetic inferences, as here performed, already provides some interesting clues, despite the low posterior credibility values of some branches (Figure 5).

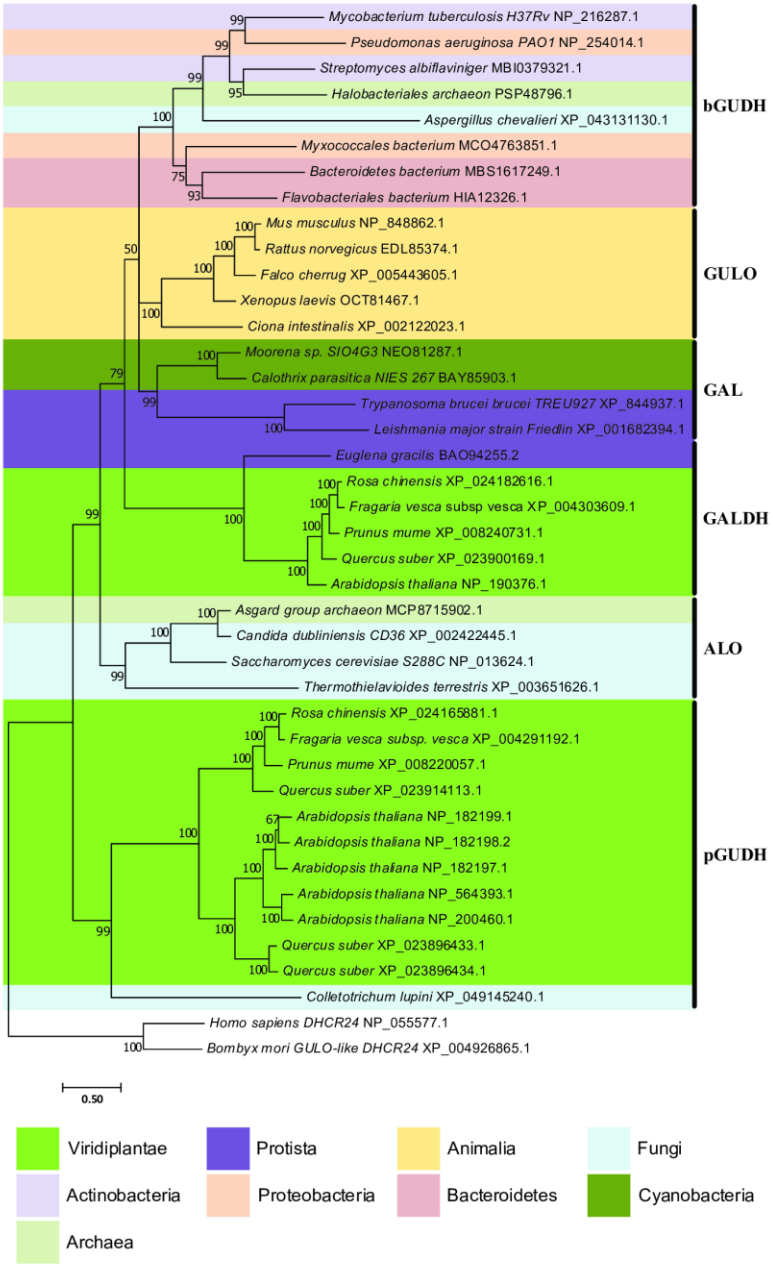


Figure 5. Bayesian phylogenetic analysis of eukaryote and putative prokaryote *AO*s. Two *DHCR24* sequences from *Homo sapiens* and *B. mori* were used to root the tree. The remaining sequences are colored according to the relevant taxonomic groups analyzed, following the legend available below

the phylogeny. The likely type of AO represented in each inferred sequence cluster can be seen to the right of the phylogeny. The GUDH, GULO, GAL, GALDH and ALO acronyms are indicative of L-Gulono-1,4-lactone Dehydrogenase, L-Gulonolactone oxidase, L-Galactono-1,4-lactone oxidase, L-Galactono-1,4-lactone dehydrogenase and D-Arabinono-1,4-lactone oxidase, respectively. The bacterial and plant GUDHs are distinguished by the "b" and "p" prefixes, respectively. The sequences used for the phylogenetic inference were obtained using the "blastp" algorithm integrated in the NCBI database (https://blast.ncbi.nlm.nih.gov/Blast.cgi; accessed on 26 July 2022) and the *S. cerevisiae* S288C ALO sequence (NP_013624.1) as query against the genomes of Actinobacteria, Proteobacteria, Bacteroidetes, Cyanobacteria, Archaea, Viridiplantae, fungi and Euglenozoa species. This genome selection considered bacterial groups previously correlated with VC synthesis [87–93], as well as the complementation of prokaryote taxonomy by the inclusion of Archaea. The Viridiplantae, fungi and Euglenozoa groups were selected to provide methodological validation relative to current literature and overall phylogenetic context. Homologous sequences representative of two species of each bacterial group and Archaea, five species of Viridiplantae, five species of fungi and three species of Euglenozoa were extracted. The "blastp" output produced considerably more results for each taxonomic group than those extracted, but the restrictive selection is representative, and through the use of more elaborated methods that have limitations in relation to the number of sequences they can handle, the quality of the phylogenetic inferences was improved. In addition to these sequences, five animal *GULO* sequences identified in [112] were also included to facilitate the interpretation of the phylogenetic results, as well as two sterol reductases sequences encoding the human and *B. mori* DHCR24 proteins, to be used as outgroup. The SEDA software [219] was used to verify the datasets and remove any anomalous sequence or isoforms that could be present. The Bayesian phylogenetic inference was performed using a MUSCLE [220] alignment as implemented in T-Coffee [221] and 2,000,000 tree generations with a defined burn-in of 25% as parameters for MrBayes [222], within ADOPS [223]. For both SEDA and ADOPS, the Docker images that are available at the pegi3S Bioinformatics Docker Images Project (https://pegi3s.github.io/dockerfiles/; accessed on 26 July 2022; [224]) were used.

One of the most notorious results is the well-supported close relationship between sequences from the cyanobacteria *Moorena sp. SIO4G3* and *Calothrix parasitica NIES-267* and the parasitic protists *Trypanosoma brucei brucei TREU927* and *Leishmania major* strain *Friedlin*. The symbiotic relations of these protists with bacteria provide advantageous nutritional benefits and promote HGT events [225]. Indeed, it has been argued that prokaryote to eukaryote HGTs of genes related to vitamin synthesis and salvage mechanisms are likely to have occurred in Trypanosomatidae species [225]. The observation made here seems to follow the same trend. A previous report indicated that *Trypanosoma cruzi* and *T. brucei* possess a L-Galactono-1,4-lactone oxidase (GAL) capable of synthesizing VC using the primary plant and fungi substrates, namely L-Galactono-1,4-lactone and D-Arabinono-1,4-lactone, but with negligible affinity to the known vertebrate substrate, L-Gulono-1,4-lactone [226,227]. Furthermore, one species of the same genus of *L. major*, *Leishmania donovani*, has an ALO that is also capable of VC synthesis using D-Arabinono-1,4-lactone as substrate, but not L-Gulono-1,4-lactone [228]. Although not yet shown, given the close phylogenetic relationship relative to *Trypanosoma* sequences, it is possible that the *L. donovani* ALO has affinity to L-Galactono-1,4-lactone as well. Therefore, the *Moorena sp. SIO4G3* and *C. parasitica NIES-267* sequences may represent genes capable of encoding AOs that are able to synthesize VC using L-Galactono 1,4-lactone and D-Arabinono-1,4-lactone, but not L-Gulono-1,4-lactone. The close phylogenetic relationship between the sequences of *Asgard group archaeon* and several fungi *ALO*, as well as the considerable sequence clustering of the majority of bacteria, one fungi and one Archaea with the *M. tuberculosis GUDH*, constitute additional indication of possible shared functional characteristics between the encoded enzymes. If so, this would suggest that some Archaea are capable of utilizing all of the common substrates for VC synthesis to a variable degree, while some Proteobacteria, Actinobacteria, Bacteroidetes, Archaea and fungi species could have specialized in the exclusive use of L-Gulono-1,4-lactone. The *M. tuberculosis GUDH* lineage is clearly divergent from the *A. thaliana GUDH*, revealing that although the biological properties are similar

between them, they likely have distinct origins. The latter lineage does not appear to be restricted to *A. thaliana*, indicating that *GUDH* might be present and intervene in VC synthesis in more Viridiplantae and even fungi. The *A. thaliana GUDH* appears to have been duplicated many times, a result consistent with previous observations of a high number of homologous AO sequences in this species [229]. The phylogenetic relationships between the *ALO*, *GULO*, *GUDH* and *GALDH* lineages vary considerably relative to those found in [94]. However, the clustering patterns of the Bacteria and Archaea sequences relative to those lineages are very well supported, implying that they likely represent relevant AOs.

The currently described AOs are characterized in the literature by the presence of amino acid positions thought to be related to their biological functions [202,226,230–232]. Following on these criteria, a search for relevant protein regions was performed (Table 1) using the alignment file that was used for the phylogenetic inference here presented (Supplementary Figure S1). To simplify the visualization, the total number of representative sequences was reduced in the results shown (Table 1). Despite this, the evidence still illustrates a convincing degree of amino acid conservation between putative bacteria and archaea AOs relative to the reference sequences presented. The putative prokaryote AOs that were identified are likely able to covalently bind to flavin adenine dinucleotide (FAD) and have efficient catalytic activity, since the essential histidine (H), tryptophan (W) and lysine (K) residues distributed across the N-terminal FAD-binding domain and HWXK motifs [202] are conserved. Furthermore, the Gluthamine (E)–Arginine (R) pair usually involved in the stabilization of the lower electric potential derived from flavin reduction and AO substrate binding and specificity [202] also appears to be conserved. With the exception of the *M. tuberculosis* GUDH [87], it is thought that the presence of a specific Glycine (G) residue within the amino acid sequence of AOs confers the ability to use oxygen as electron acceptor [202]. If so, the putative prokaryote AOs can act as oxidases, similarly to animal GULO and fungi ALO. Moreover, a cysteine (C) residue linked to substrate specificity of AOs [202] is present in many of the putative prokaryote AOs, with the exception of *Bacteroidetes bacterium*. In agreement with the phylogenetic results, the overall conservation of essential amino acid positions suggests that the putative prokaryote proteins here displayed represent novel AOs.

Table 1. Amino acid regions relevant for AO functions in prokaryote and eukaryote species.

Protein Sequences	FAD-Binding Motif (X[GR]X[GS][HLK][SG]) *	HWXK Motif	EXR Pair	Oxygen Reactivity Marker (G)	Conserved Cysteine (C)
M. tuberculosis H37Rv bGUDH NP_216287.1	VGS-G**H**S	**HWGK**	**E**V**R**	**G**	R
Pseudomonas aeruginosa AO NP_254014.1	VGS-G**H**S	**HWGK**	**E**Y**R**	A	R
B. bacterium AO MBS1617249.1 **	VGT-G**H**S	**HWGK**	**E**I**R**	**G**	D
C. parasitica NIES 267 AO BAY85903.1 **	FGS-G**H**S	**HWAK**	**E**V**R**	**G**	**C**
T. brucei brucei TREU927 GAL XP_844937.1	VGG-GKS	**HWAK**	**E**F**R**	P	**C**
M. musculus GULO NP_848862.1	VGG-G**H**S	**HWAK**	**E**V**R**	**G**	**C**
A. thaliana GALDH NP_190376.1	VGS-GLS	**HWAK**	**E**Q**R**	A	**C**
A. thaliana pGUDH NP_182197.1	VTTRYSH	**HWGK**	LM**R**	P	**C**
Asgard group archaeon AO MCP8715902.1 **	VGS-G**H**S	**HWAK**	**E**V**R**	**G**	**C**
S. cerevisiae S288C ALO NP_013624.1	VGS-G**H**S	**HWAK**	**E**V**R**	**G**	**C**
Euglena gracilis GALDH BAO94255.2	AGA-MLS	**HWAK**	**E**Q**R**	S	**C**
Aspergillus chevalieri bGUDH XP_043131130.1	VGN-G**H**G	**HWNK**	Q**F**R	G	E
C. lupini pGUDH XP_049145240.1	SGK-G**H**M	**HWTK**	NIK	L	V
B. mori DHCR24 XP_004926865.1	CTA-RPT	LYAD	FKI	P	K

* N-terminal FAD-binding motif derived from evidence in [202]. ** Novel prokaryote AOs inferred from the phylogenetic study. Essential amino acids are highlighted in bold, and amino acid regions that do not follow the motifs are underlined.

As previously stated, the analyses performed in this section were primarily focused on the verification of the existence of putative AOs in prokaryote species. Although there are no functional data to support these inferences, the results here presented raise doubts regarding the common assumption of generalized prokaryote VC auxotrophy, and provide a new perspective on the possible distribution of VC synthesis mechanisms in the different life kingdoms (Figure 6). The Asgard archaea lineage has been regarded as the prokaryote origin of the ancestral eukaryotes [233,234], and chloroplasts are organelles derived from cyanobacteria endosymbiosis [198]. In this sense, if functional AOs are present in species from both taxonomic groups as the in silico analyses suggest, VC synthesis evolutionary models must consider them.

The presence of AOs in prokaryotes may open new perspectives for potential industrial applications. VC synthesis usually relies on an initial chemical conversion of D-Glucose to D-Sorbitol followed by two independent fermentation steps using *Gluconobacter oxydans* and a mixed culture of *Ketogulonigenium vulgare* and *Bacillus megaterium*, respectively, to obtain 2-Keto-L-Gulonic acid [235], a VC precursor. This two-step fermentation method is highly efficient, but the hydrogenation of D-Glucose and the maintenance of balanced co-cultures still have economical and labor costs that can, in principle, be avoided [235], using for instance, a direct one-step fermentation process [235,236]. This can be achieved through the genetic integration of plant VC synthesis pathway components, such as the *A. thaliana GALDH*, in fungi [236], as well as the integration of genes involved in D-Sorbitol/L-Sorbose metabolism from different prokaryote sources into a single bacterial chassis [237–239]. Substantial optimization is, however, required since the efficiency of VC synthesis is still dependent on the adequate availability of cofactors and/or compatible electron acceptors [235,236]. Nevertheless, if the species used as chassis already has the majority of the required enzymatic components, such limitations could be overcome. Some γ-proteobacteria species, such as *Erwinia herbicola*, are able to perform the direct oxidation of D-Glucose to 2,5-diketo-D-Gluconic acid [240], and therefore constitute a suitable bacterial chassis candidate for 2-Keto-L-Gulonic acid synthesis with reduced genetic manipulation and kinetic optimization. This possibility has been successfully explored through the integration of an exogenous 2,5-diketo-D-Gluconic acid reductase into *E. herbicola*, that allows the production of 2-Keto-L-Gulonic acid from 2,5-diketo-D-Gluconic acid [240] or after the fusion of a *Corynebacterium sp.* protoplast with *E. herbicola* [241]. The effective yields were low in both cases, but such approaches still remain relevant for potential industrial applications [235]. The phylogenetic results here presented imply the presence of a putative AO in *P. aeruginosa*, which is also a γ-proteobacteria. Therefore, the metabolic components that are present in *E. herbicola*, might also be present in this species. If true, *P. aeruginosa* could represent an additional bacterial chassis that could be explored for the purpose of VC synthesis. Given that a bioengineered *P. aeruginosa* acid phosphatase has already been used to convert VC to a more stable L-Ascorbic acid-2-phosphate derivative resistant to high-temperatures [242], the same bacterial species could be used for both VC synthesis and stabilization, a relevant concern in the recovery process of this nutrient [236].

The current knowledge and hypotheses on novel VC synthesis pathways here discussed (summarized in Figure 6) are expected to promote research on this issue in different organisms.

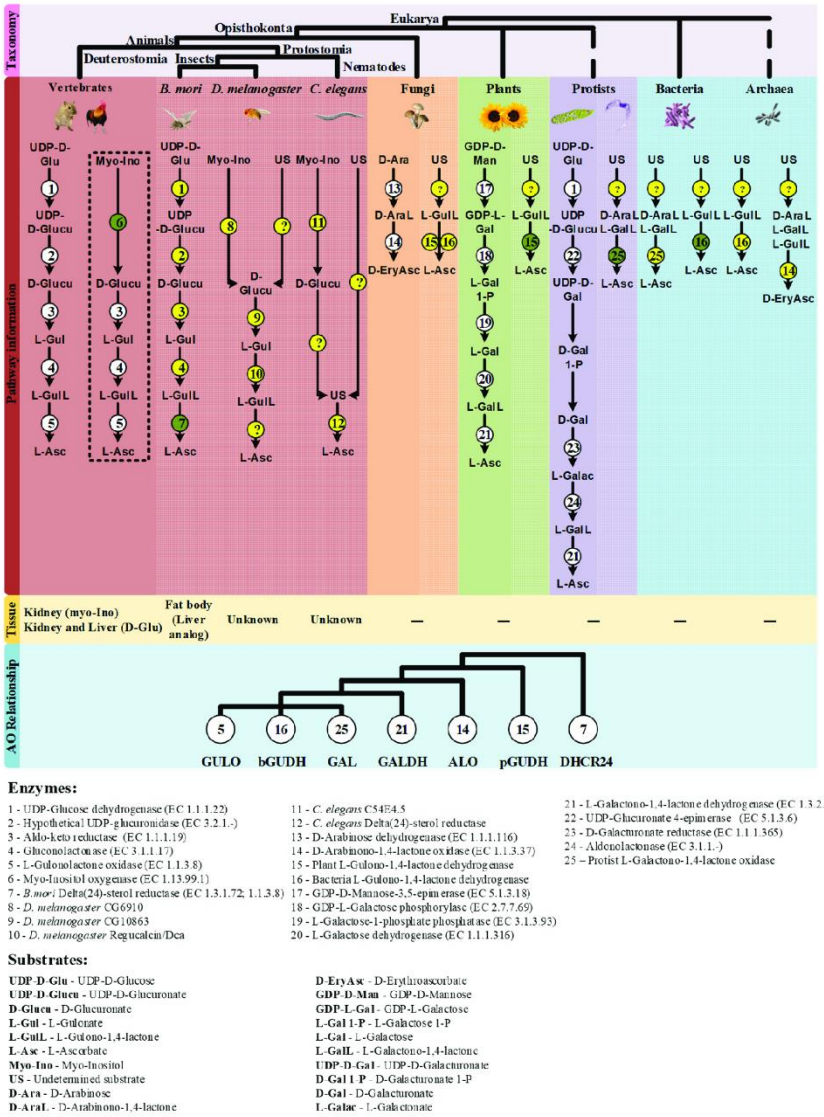


Figure 6. Summary of the findings discussed in this review. The taxonomic relationship between the species mentioned in this work is displayed on the top of the figure. The characterized VC synthesis pathways and the putative alternatives in eukaryote and prokaryote species are presented in the mid part of the figure. The pathways already described in the literature are highlighted with bold colors according to the taxonomic groups mentioned, while the putative VC pathways are presented in the gridline sections. The enzymes that intervene in the substrate conversions are presented by numbered circles according to the following scheme: Circles in white represent enzymatic reactions described in the literature, circles in yellow represent hypothetical enzymatic reactions and circles in green indicate novel enzymatic constituents of putative alternative VC synthesis pathways with available functional characterization. The question marks (?) represent putative enzymatic reactions for which no candidate enzyme can be suggested with current data. The tissues correlated with VC synthesis in animal species are represented below the corresponding proposed pathways when available. The dashed box highlights the myo-Inositol VC synthesis pathway that is exclusively used by species that synthesize VC in the kidney. The phylogenetic relationship inferred between the studied AOs is summarized in a cladogram format at the bottom. The enzyme number and substrate acronym lists are presented below the figure to facilitate the interpretation.

5. Conclusions

Current research has only grasped a fraction of the VC contributions to overall organism function, as well as the complexity behind the endogenous synthesis of this nutrient. Here, we critically review the literature on metazoan VC synthesis and put forward novel hypotheses and discuss methodological approaches that may contribute to the characterization of alternative enzymatic mechanisms behind VC synthesis in *C. elegans* and *D. melanogaster*. A putative prokaryotic origin for eukaryote *AO*s is also hypothesized and discussed in the context of the current evolutionary scenario that led to the emergence of the *GULO* and *GALDH* lineages. The performed phylogenetic analysis indicated that putative AOs are present in bacteria and archaea species. These inferences are discussed in the context of possible applications on the optimization of the industrial production of VC.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes13101917/s1, Figure S1: Complete AO dataset MUSCLE alignment file.

Author Contributions: Conceptualization, P.D., C.P.V. and J.V.; formal analysis, P.D., C.P.V. and J.V.; investigation, P.D., C.P.V. and J.V.; writing—original draft preparation, P.D., C.P.V. and J.V.; writing—review and editing, P.D., C.P.V. and J.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financed by the National Funds through FCT—Fundação para a Ciência e a Tecnologia, I.P., under the project UIDB/04293/2020. PD is supported by a PhD fellowship from Fundação para a Ciência e Tecnologia (SFRH/BD/145515/2019), co-financed by the European Social Fund through the Norte Portugal Regional Operational Programme (NORTE 2020).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Du, J.; Cullen, J.J.; Buettner, G.R. Ascorbic acid: Chemistry, biology and the treatment of cancer. *Biochim. Biophys. Acta* **2012**, *1826*, 443–457. [CrossRef]
2. Bauernfeind, J.C. Ascorbic Acid Technology in Agricultural, Pharmaceutical, Food, and Industrial Applications. In *Ascorbic Acid: Chemistry, Metabolism, and Uses*; Advances in Chemistry; American Chemical Society: Washington, DC, USA, 1982; pp. 395–497.
3. Yin, X.; Chen, K.; Cheng, H.; Chen, X.; Feng, S.; Song, Y.; Liang, L. Chemical Stability of Ascorbic Acid Integrated into Commercial Products: A Review on Bioactivity and Delivery Technology. *Antioxidants* **2022**, *11*, 153. [CrossRef] [PubMed]
4. Chaturvedi, S.; Khan, S.; Bhunia, R.K.; Kaur, K.; Tiwari, S. Metabolic engineering in food crops to enhance ascorbic acid production: Crop biofortification perspectives for human health. *Physiol. Mol. Biol. Plants* **2022**, *28*, 871–884. [CrossRef] [PubMed]
5. Hieu, T.V.; Guntoro, B.; Qui, N.H.; Quyen, N.T.K.; Hafiz, F.A.A. The application of ascorbic acid as a therapeutic feed additive to boost immunity and antioxidant activity of poultry in heat stress environment. *Vet. World* **2022**, *15*, 685–693. [CrossRef] [PubMed]
6. Pappenberger, G.; Hohmann, H.-P. Industrial Production of l-Ascorbic Acid (Vitamin C) and d-Isoascorbic Acid. In *Biotechnology of Food and Feed Additives*; Zorn, H., Czermak, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 143–188.
7. Sheng, Y.; Tan, X.; Gu, Y.; Zhou, X.; Tu, M.; Xu, Y. Effect of ascorbic acid assisted dilute acid pretreatment on lignin removal and enzyme digestibility of agricultural residues. *Renew. Energy* **2021**, *163*, 732–739. [CrossRef]
8. Chaudhry, A.; Bashir, F.; Adil, S.F.; Saif, S.; Shaik, M.R.; Hatshan, M.R.; Shaik, B. Ascorbic acid-mediated Fe/Cu nanoparticles and their application for removal of COD and phenols from industrial wastewater. *J. King Saud Univ.–Sci.* **2022**, *34*, 101927. [CrossRef]
9. Basnayake Pussepitiyalage, V.; Hemmati, S. Sustainable, Green, and Continuous Synthesis of Fivefold Palladium Nanorods Using l-Ascorbic Acid in a Segmented Millifluidic Flow Reactor. *Langmuir* **2022**, *38*, 4200–4212. [CrossRef]
10. Bae, M.; Kim, H. The Role of Vitamin C, Vitamin D, and Selenium in Immune System against COVID-19. *Molecules* **2020**, *25*, 5346. [CrossRef] [PubMed]
11. Carr, A.C.; Rowe, S. The Emerging Role of Vitamin C in the Prevention and Treatment of COVID-19. *Nutrients* **2020**, *12*, 3286. [CrossRef] [PubMed]
12. Peng, Y.; Kwok, K.H.H.; Yang, P.-H.; Ng, S.S.M.; Liu, J.; Wong, O.G.; He, M.-L.; Kung, H.-F.; Lin, M.C.M. Ascorbic acid inhibits ROS production, NF-κB activation and prevents ethanol-induced growth retardation and microencephaly. *Neuropharmacology* **2005**, *48*, 426–434. [CrossRef]

13. Do, H.; Kim, I.-S.; Jeon, B.W.; Lee, C.W.; Park, A.K.; Wi, A.R.; Shin, S.C.; Park, H.; Kim, Y.-S.; Yoon, H.-S.; et al. Structural understanding of the recycling of oxidized ascorbate by dehydroascorbate reductase (OsDHAR) from *Oryza sativa* L. *japonica*. *Sci. Rep.* **2016**, *6*, 19498. [CrossRef]
14. Yimcharoen, M.; Kittikunnathum, S.; Suknikorn, C.; Nak-on, W.; Yeethong, P.; Anthony, T.G.; Bunpo, P. Effects of ascorbic acid supplementation on oxidative stress markers in healthy women following a single bout of exercise. *J. Int. Soc. Sport. Nutr.* **2019**, *16*, 2. [CrossRef] [PubMed]
15. Padayatty, S.J.; Katz, A.; Wang, Y.; Eck, P.; Kwon, O.; Lee, J.-H.; Chen, S.; Corpe, C.; Dutta, A.; Dutta, S.K.; et al. Vitamin C as an Antioxidant: Evaluation of Its Role in Disease Prevention. *J. Am. Coll. Nutr.* **2003**, *22*, 18–35. [CrossRef] [PubMed]
16. May, J.M.; Harrison, F.E. Role of vitamin C in the function of the vascular endothelium. *Antioxid. Redox Signal.* **2013**, *19*, 2068–2083. [CrossRef]
17. Smirnoff, N. Ascorbic acid metabolism and functions: A comparison of plants and mammals. *Free Radic. Biol. Med.* **2018**, *122*, 116–129. [CrossRef] [PubMed]
18. Ferreira, I.C.; Barros, L.; Abreu, R. Antioxidants in wild mushrooms. *Curr. Med. Chem.* **2009**, *16*, 1543–1560. [CrossRef] [PubMed]
19. Bilska, K.; Wojciechowska, N.; Alipour, S.; Kalemba, E.M. Ascorbic Acid—The Little-Known Antioxidant in Woody Plants. *Antioxidants* **2019**, *8*, 645. [CrossRef] [PubMed]
20. Deutsch, J.C. Dehydroascorbic acid. *J. Chromatogr. A* **2000**, *881*, 299–307. [CrossRef]
21. Knight, J.; Madduma-Liyanage, K.; Mobley, J.A.; Assimos, D.G.; Holmes, R.P. Ascorbic acid intake and oxalate synthesis. *Urolithiasis* **2016**, *44*, 289–297. [CrossRef]
22. Chan, A.C. Partners in defense, vitamin E and vitamin C. *Can. J. Physiol. Pharmacol.* **1993**, *71*, 725–731. [CrossRef] [PubMed]
23. Winkler, B.S.; Orselli, S.M.; Rex, T.S. The redox couple between glutathione and ascorbic acid: A chemical and physiological perspective. *Free Radic. Biol. Med.* **1994**, *17*, 333–349. [CrossRef]
24. Evans, W.J. Vitamin E, vitamin C, and exercise. *Am. J. Clin. Nutr.* **2000**, *72*, 647S–652S. [CrossRef] [PubMed]
25. Higgins, M.R.; Izadi, A.; Kaviani, M. Antioxidants and Exercise Performance: With a Focus on Vitamin E and C Supplementation. *Int. J. Environ. Res. Public Health* **2020**, *17*, 8452. [CrossRef] [PubMed]
26. Bentley, D.J.; Ackerman, J.; Clifford, T.; Slattery, K.S. Acute and Chronic Effects of Antioxidant Supplementation on Exercise Performance. In *Antioxidants in Sport Nutrition*; Lamprecht, M., Ed.; CRC Press/Taylor & Francis: Boca Raton, FL, USA, 2015; pp. 141–154.
27. Figueroa-Méndez, R.; Rivas-Arancibia, S. Vitamin C in Health and Disease: Its Role in the Metabolism of Cells and Redox State in the Brain. *Front. Physiol.* **2015**, *6*, 397. [CrossRef]
28. Rong, Y.; Mi, X.; Ni, C.; Liu, T.; Yang, N.; Hong, J.; Li, Y.; Li, Z.; Han, D.; Guo, X. Protective effect of vitamin C on DNA damage in surgery-induced cognitive dysfunction in APP/PS1 mice. *Neurosci. Lett.* **2022**, *784*, 136740. [CrossRef]
29. Covarrubias-Pinto, A.; Acuña, A.I.; Beltrán, F.A.; Torres-Díaz, L.; Castro, M.A. Old Things New View: Ascorbic Acid Protects the Brain in Neurodegenerative Disorders. *Int. J. Mol. Sci.* **2015**, *16*, 28194–28217. [CrossRef]
30. Ballaz, S.J.; Rebec, G.V. Neurobiology of vitamin C: Expanding the focus from antioxidant to endogenous neuromodulator. *Pharmacol. Res.* **2019**, *146*, 104321. [CrossRef]
31. Naseer, M.I.; Ullah, I.; Ullah, N.; Lee, H.Y.; Cheon, E.W.; Chung, J.; Kim, M.O. Neuroprotective effect of vitamin C against PTZ induced apoptotic neurodegeneration in adult rat brain. *Pak. J. Pharm. Sci.* **2011**, *24*, 263–268.
32. Weiskirchen, R. Hepatoprotective and Anti-fibrotic Agents: It's Time to Take the Next Step. *Front. Pharmacol.* **2016**, *6*, 303. [CrossRef]
33. Adikwu, E.; Deo, O. Hepatoprotective effect of Vitamin C (ascorbic acid). *Pharmacol. Pharm.* **2013**, *4*, 84–92. [CrossRef]
34. Abdulrazzaq, A.M.; Badr, M.; Gammoh, O.; Abu Khalil, A.A.; Ghanim, B.Y.; Alhussainy, T.M.; Qinna, N.A. Hepatoprotective Actions of Ascorbic Acid, α Lipoic Acid and Silymarin or Their Combination Against Acetaminophen-Induced Hepatotoxicity in Rats. *Medicina* **2019**, *55*, 181. [CrossRef]
35. Ali, S.; Hussain, S.; Khan, R.; Mumtaz, S.; Ashraf, N.; Andleeb, S.; Shakir, H.A.; Tahir, H.M.; Khan, M.K.A.; Ulhaq, M. Renal toxicity of heavy metals (cadmium and mercury) and their amelioration with ascorbic acid in rabbits. *Environ. Sci. Pollut. Res.* **2019**, *26*, 3909–3920. [CrossRef] [PubMed]
36. Oliveira, C.P.; Kassab, P.; Lopasso, F.P.; Souza, H.P.; Janiszewski, M.; Laurindo, F.R.; Iriya, K.; Laudanna, A.A. Protective effect of ascorbic acid in experimental gastric cancer: Reduction of oxidative stress. *World J. Gastroenterol.* **2003**, *9*, 446–448. [CrossRef]
37. Attia, M.; Essa, E.A.; Zaki, R.M.; Elkordy, A.A. An overview of the antioxidant effects of ascorbic acid and α lipoic acid (in liposomal forms) as adjuvant in cancer treatment. *Antioxidants* **2020**, *9*, 359. [CrossRef]
38. Sunil Kumar, B.; Singh, S.; Verma, R. Anticancer potential of dietary vitamin D and ascorbic acid: A review. *Crit. Rev. Food Sci. Nutr.* **2017**, *57*, 2623–2635. [CrossRef] [PubMed]
39. De Carvalho Melo-Cavalcante, A.A.; da Rocha Sousa, L.; Alencar, M.V.O.B.; de Oliveira Santos, J.V.; da Mata, A.M.o.; Paz, M.F.C.J.; de Carvalho, R.M.; Nunes, N.M.F.; Islam, M.T.; Mendes, A.N.; et al. Retinol palmitate and ascorbic acid: Role in oncological prevention and therapy. *Biomed. Pharmacother.* **2019**, *109*, 1394–1405. [CrossRef] [PubMed]
40. Osiecki, M.; Ghanavi, P.; Atkinson, K.; Nielsen, L.K.; Doran, M.R. The ascorbic acid paradox. *Biochem. Biophys. Res. Commun.* **2010**, *400*, 466–470. [CrossRef] [PubMed]
41. Zhong, B.; Zhao, L.; Yu, J.; Hou, Y.; Ai, N.; Lu, J.-J.; Ge, W.; Chen, X. Exogenous iron impairs the anti-cancer effect of ascorbic acid both in vitro and in vivo. *J. Adv. Res.* **2022**. [CrossRef] [PubMed]

42. Vissers, M.C.M.; Das, A.B. Potential Mechanisms of Action for Vitamin C in Cancer: Reviewing the Evidence. *Front. Physiol.* **2018**, *9*, 809. [CrossRef]
43. González, M.J.; Rosario-Pérez, G.; Guzmán, A.M.; Miranda-Massari, J.R.; Duconge, J.; Lavergne, J.; Fernandez, N.; Ortiz, N.; Quintero, A.; Mikirova, N.; et al. Mitochondria, Energy and Cancer: The Relationship with Ascorbic Acid. *J. Orthomol. Med. Off. J. Acad. Orthomol. Med.* **2010**, *25*, 29–38.
44. Yun, J.; Mullarky, E.; Lu, C.; Bosch, K.N.; Kavalier, A.; Rivera, K.; Roper, J.; Chio, I.I.C.; Giannopoulou, E.G.; Rago, C.; et al. Vitamin C selectively kills KRAS and BRAF mutant colorectal cancer cells by targeting GAPDH. *Science* **2015**, *350*, 1391–1396. [CrossRef]
45. Tian, W.; Wang, Z.; Tang, N.-n.; Li, J.-t.; Liu, Y.; Chu, W.-F.; Yang, B.-F. Ascorbic Acid Sensitizes Colorectal Carcinoma to the Cytotoxicity of Arsenic Trioxide via Promoting Reactive Oxygen Species-Dependent Apoptosis and Pyroptosis. *Front. Pharmacol.* **2020**, *11*, 123. [CrossRef]
46. Gong, E.-Y.; Shin, Y.J.; Hwang, I.-Y.; Kim, J.H.; Kim, S.-M.; Moon, J.-H.; Shin, J.-S.; Lee, D.-H.; Hur, D.Y.; Jin, D.-H.; et al. Combined treatment with vitamin C and sulindac synergistically induces p53- and ROS-dependent apoptosis in human colon cancer cells. *Toxicol. Lett.* **2016**, *258*, 126–133. [CrossRef]
47. Podmore, I.D.; Griffiths, H.R.; Herbert, K.E.; Mistry, N.; Mistry, P.; Lunec, J. Vitamin C exhibits pro-oxidant properties. *Nature* **1998**, *392*, 559. [CrossRef] [PubMed]
48. Duarte, T.L.; Almeida, G.M.; Jones, G.D.D. Investigation of the role of extracellular H_2O_2 and transition metal ions in the genotoxic action of ascorbic acid in cell culture models. *Toxicol. Lett.* **2007**, *170*, 57–65. [CrossRef] [PubMed]
49. Carr, A.C.; Cook, J. Intravenous Vitamin C for Cancer Therapy—Identifying the Current Gaps in Our Knowledge. *Front. Physiol.* **2018**, *9*, 1182. [CrossRef] [PubMed]
50. Pinnell, S.R. Regulation of collagen biosynthesis by ascorbic acid: A review. *Yale J. Biol. Med.* **1985**, *58*, 553–559.
51. Phillips, C.L.; Combs, S.B.; Pinnell, S.R. Effects of Ascorbic Acid on Proliferation and Collagen Synthesis in Relation to the Donor Age of Human Dermal Fibroblasts. *J. Investig. Dermatol.* **1994**, *103*, 228–232. [CrossRef]
52. Agarwal, A.; Shaharyar, A.; Kumar, A.; Bhat, M.S.; Mishra, M. Scurvy in pediatric age group—A disease often forgotten? *J. Clin. Orthop. Trauma* **2015**, *6*, 101–107. [CrossRef]
53. Kipp, D.E.; McElvain, M.; Kimmel, D.B.; Akhter, M.P.; Robinson, R.G.; Lukert, B.P. Scurvy results in decreased collagen synthesis and bone density in the guinea pig animal model. *Bone* **1996**, *18*, 281–288. [CrossRef]
54. Camarena, V.; Wang, G. The epigenetic role of vitamin C in health and disease. *Cell. Mol. Life Sci.* **2016**, *73*, 1645–1658. [CrossRef]
55. Lee Chong, T.; Ahearn, E.L.; Cimmino, L. Reprogramming the Epigenome With Vitamin C. *Front. Cell Dev. Biol.* **2019**, *7*, 128. [CrossRef]
56. Grillet, L.; Ouerdane, L.; Flis, P.; Hoang, M.T.T.; Isaure, M.-P.; Lobinski, R.; Curie, C.; Mari, S. Ascorbate Efflux as a New Strategy for Iron Reduction and Transport in Plants. *J. Biol. Chem.* **2014**, *289*, 2515–2525. [CrossRef]
57. Lane, D.J.R.; Richardson, D.R. The active role of vitamin C in mammalian iron metabolism: Much more than just enhanced iron absorption! *Free Radic. Biol. Med.* **2014**, *75*, 69–83. [CrossRef] [PubMed]
58. Baldauf, S.L.; Palmer, J.D. Animals and fungi are each other's closest relatives: Congruent evidence from multiple proteins. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11558–11562. [CrossRef] [PubMed]
59. Koch, L. The origin of animals and fungi. *Nat. Rev. Genet.* **2022**, *23*, 648–649. [CrossRef]
60. Baroja-Mazo, A.; Valle, P.d.; Rúa, J.; de Cima, S.; Busto, F.; de Arriaga, D.; Smirnoff, N. Characterisation and biosynthesis of d-erythroascorbic acid in *Phycomyces blakesleeanus*. *Fungal Genet. Biol.* **2005**, *42*, 390–402. [CrossRef] [PubMed]
61. Spickett, C.M.; Smirnoff, N.; Pitt, A.R. The biosynthesis of erythroascorbate in *Saccharomyces cerevisiae* and its role as an antioxidant. *Free Radic. Biol. Med.* **2000**, *28*, 183–192. [CrossRef]
62. Chen, H.; Hao, H.; Han, C.; Wang, H.; Wang, Q.; Chen, M.; Juan, J.; Feng, Z.; Zhang, J. Exogenous l-ascorbic acid regulates the antioxidant system to increase the regeneration of damaged mycelia and induce the development of fruiting bodies in *Hypsizygus marmoreus*. *Fungal Biol.* **2020**, *124*, 551–561. [CrossRef] [PubMed]
63. Saffi, J.; Sonego, L.; Varela, Q.D.; Salvador, M. Antioxidant activity of L-ascorbic acid in wild-type and superoxide dismutase deficient strains of *Saccharomyces cerevisiae*. *Redox Rep.* **2006**, *11*, 179–184. [CrossRef] [PubMed]
64. Georgiou, C.D.; Petropoulou, K.P. The role of ascorbic acid role in the differentiation of sclerotia in *Sclerotinia minor*. *Mycopathologia* **2002**, *154*, 71–77. [CrossRef] [PubMed]
65. Akram, N.A.; Shafiq, F.; Ashraf, M. Ascorbic Acid-A Potential Oxidant Scavenger and Its Role in Plant Development and Abiotic Stress Tolerance. *Front. Plant Sci.* **2017**, *8*, 613. [CrossRef] [PubMed]
66. Niu, J.; Chen, Z.; Yu, S.; Wang, Q. Ascorbic acid regulates nitrogen, energy, and gas exchange metabolisms of alfalfa in response to high-nitrate stress. *Environ. Sci. Pollut. Res.* **2022**, *29*, 24085–24097. [CrossRef] [PubMed]
67. Alamri, S.A.; Siddiqui, M.H.; Al-Khaishany, M.Y.Y.; Nasir Khan, M.; Ali, H.M.; Alaraidh, I.A.; Alsahli, A.A.; Al-Rabiah, H.; Mateen, M. Ascorbic acid improves the tolerance of wheat plants to lead toxicity. *J. Plant Interact.* **2018**, *13*, 409–419. [CrossRef]
68. Zheng, X.; Gong, M.; Zhang, Q.; Tan, H.; Li, L.; Tang, Y.; Li, Z.; Peng, M.; Deng, W. Metabolism and Regulation of Ascorbic Acid in Fruits. *Plants* **2022**, *11*, 1602. [CrossRef]
69. Bybordi, A. Effect of Ascorbic Acid and Silicium on Photosynthesis, Antioxidant Enzyme Activity, and Fatty Acid Contents in Canola Exposure to Salt Stress. *J. Integr. Agric.* **2012**, *11*, 1610–1620. [CrossRef]

70. Agami, R.A. Applications of ascorbic acid or proline increase resistance to salt stress in barley seedlings. *Biol. Plant.* **2014**, *58*, 341–347. [CrossRef]
71. Fatima, A.; Farid, M.; Alharby, H.F.; Bamagoos, A.A.; Rizwan, M.; Ali, S. Efficacy of fenugreek plant for ascorbic acid assisted phytoextraction of copper (Cu); A detailed study of Cu induced morpho-physiological and biochemical alterations. *Chemosphere* **2020**, *251*, 126424. [CrossRef]
72. El-Beltagi, H.S.; Ahmad, I.; Basit, A.; Shehata, W.F.; Hassan, U.; Shah, S.T.; Haleema, B.; Jalal, A.; Amin, R.; Khalid, M.A.; et al. Ascorbic Acid Enhances Growth and Yield of Sweet Peppers (*Capsicum annum*) by Mitigating Salinity Stress. *Gesunde Pflanz.* **2022**, *74*, 423–433. [CrossRef]
73. Sharova, E.I.; Medvedev, S.S.; Demidchik, V.V. Ascorbate in the Apoplast: Metabolism and Functions. *Russ. J. Plant Physiol.* **2020**, *67*, 207–220. [CrossRef]
74. Bellini, E.; De Tullio, M.C. Ascorbic Acid and Ozone: Novel Perspectives to Explain an Elusive Relationship. *Plants* **2019**, *8*, 122. [CrossRef] [PubMed]
75. Suekawa, M.; Fujikawa, Y.; Esaka, M. Physiological Role of Ascorbic Acid Recycling Enzymes in Plants. In *Ascorbic Acid in Plant Growth, Development and Stress Tolerance*; Hossain, M.A., Munné-Bosch, S., Burritt, D.J., Diaz-Vivancos, P., Fujita, M., Lorence, A., Eds.; Springer: Cham, Switzerland, 2017; pp. 355–373.
76. Smirnoff, N. Chapter 4—Vitamin C: The Metabolism and Functions of Ascorbic Acid in Plants. In *Advances in Botanical Research*; Rébeillé, F., Douce, R., Eds.; Academic Press: Cambridge, MA, USA, 2011; Volume 59, pp. 107–177.
77. Finkemeier, I.; Leister, D. Plant Chloroplasts and Other Plastids. In *eLS*; John Wiley & Sons: Hoboken, NJ, USA, 2010.
78. Roach, T.; Krieger-Liszkay, A. Photosynthetic Regulatory Mechanisms for Efficiency and Prevention of Photo-Oxidative Stress. In *Annual Plant Reviews Online*; John Wiley & Sons: Hoboken, NJ, USA, 2019; pp. 273–306.
79. Foyer, C.H.; Shigeoka, S. Understanding Oxidative Stress and Antioxidant Functions to Enhance Photosynthesis. *Plant Physiol.* **2011**, *155*, 93–100. [CrossRef] [PubMed]
80. Paciolla, C.; Fortunato, S.; Dipierro, N.; Paradiso, A.; De Leonardis, S.; Mastropasqua, L.; de Pinto, M.C. Vitamin C in Plants: From Functions to Biofortification. *Antioxidants* **2019**, *8*, 519. [CrossRef] [PubMed]
81. Asada, K. The water-water cycle in chloroplasts: Scavenging of Active Oxygens and Dissipation of Excess Photons. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1999**, *50*, 601–639. [CrossRef]
82. Miyake, C. Alternative Electron Flows (Water–Water Cycle and Cyclic Electron Flow Around PSI) in Photosynthesis: Molecular Mechanisms and Physiological Functions. *Plant Cell Physiol.* **2010**, *51*, 1951–1963. [CrossRef]
83. Caverzan, A.; Bonifacio, A.; Carvalho, F.E.L.; Andrade, C.M.B.; Passaia, G.; Schünemann, M.; Maraschin, F.d.S.; Martins, M.O.; Teixeira, F.K.; Rauber, R.; et al. The knockdown of chloroplastic ascorbate peroxidases reveals its regulatory role in the photosynthesis and protection under photo-oxidative stress in rice. *Plant Sci.* **2014**, *214*, 74–87. [CrossRef]
84. Eskling, M.; Arvidsson, P.O.; Åkerlund, H.E. The xanthophyll cycle, its regulation and components. *Physiol. Plant.* **1997**, *100*, 806–816. [CrossRef]
85. Latowski, D.; Grzyb, J.; Strzałka, K. The xanthophyll cycle-molecular mechanism and physiological significance. *Acta Physiol. Plant.* **2004**, *26*, 197–212. [CrossRef]
86. Muüller-Mouleé, P.; Havaux, M.; Niyogi, K.K. Zeaxanthin Deficiency Enhances the High Light Sensitivity of an Ascorbate-Deficient Mutant of Arabidopsis. *Plant Physiol.* **2003**, *133*, 748–760. [CrossRef]
87. Wolucka, B.A.; Communi, D. Mycobacterium tuberculosis possesses a functional enzyme for the synthesis of vitamin C, L-gulono-1,4-lactone dehydrogenase. *FEBS J.* **2006**, *273*, 4435–4445. [CrossRef]
88. Arumugam, M.; Raes, J.; Pelletier, E.; Le Paslier, D.; Yamada, T.; Mende, D.R.; Fernandes, G.R.; Tap, J.; Bruls, T.; Batto, J.M.; et al. Enterotypes of the human gut microbiome. *Nature* **2011**, *473*, 174–180. [CrossRef] [PubMed]
89. Chang, Y.L.; Rossetti, M.; Vlamakis, H.; Casero, D.; Sunga, G.; Harre, N.; Miller, S.; Humphries, R.; Stappenbeck, T.; Simpson, K.W.; et al. A screen of Crohn's disease-associated microbial metabolites identifies ascorbate as a novel metabolic inhibitor of activated human T cells. *Mucosal Immunol.* **2019**, *12*, 457–467. [CrossRef] [PubMed]
90. Sugisawa, T.; Ojima, S.; Matzinger, P.K.; Hoshino, T. Isolation and Characterization of a New Vitamin C Producing Enzyme (L-Gulono-γ-lactone Dehydrogenase) of Bacterial Origin. *Biosci. Biotechnol. Biochem.* **1995**, *59*, 190–196. [CrossRef]
91. Sugisawa, T.; Miyazaki, T.; Hoshino, T. Microbial production of L-ascorbic acid from D-sorbitol, L-sorbose, L-gulose, and L-sorbosone by Ketogulonicigenium vulgare DSM 4025. *Biosci Biotechnol Biochem* **2005**, *69*, 659–662. [CrossRef] [PubMed]
92. Eddy, B.; Ingram, M. Interactions between ascorbic acid and bacteria. *Bacteriol. Rev.* **1953**, *17*, 93–107. [CrossRef] [PubMed]
93. Büsing, K.; Peters, F. tVber die Ascorbinsiurebildung des Bac. pro-digiosus aus Xylose. *Biochem. Z* **1940**, *304*, 134–136.
94. Wheeler, G.; Ishikawa, T.; Pornsaksit, V.; Smirnoff, N. Evolution of alternative biosynthetic pathways for vitamin C following plastid acquisition in photosynthetic eukaryotes. *eLife* **2015**, *4*, e06369. [CrossRef]
95. Wheeler, G.L.; Jones, M.A.; Smirnoff, N. The biosynthetic pathway of vitamin C in higher plants. *Nature* **1998**, *393*, 365–369. [CrossRef]
96. Shigeoka, S.; Nakano, Y.; Kitaoka, S. The biosynthetic pathway of L-ascorbic acid in *Euglena gracilis* Z. *J. Nutr. Sci. Vitaminol.* **1979**, *25*, 299–307. [CrossRef]
97. Nishikimi, M.; Yagi, K. Molecular basis for the deficiency in humans of gulonolactone oxidase, a key enzyme for ascorbic acid biosynthesis. *Am. J. Clin. Nutr.* **1991**, *54*, 1203S–1208S. [CrossRef]

98. Vargas, J.A.; Leonardo, D.A.; D'Muniz Pereira, H.; Lopes, A.R.; Rodriguez, H.N.; Cobos, M.; Marapara, J.L.; Castro, J.C.; Garratt, R.C. Structural Characterization of L-Galactose Dehydrogenase: An Essential Enzyme for Vitamin C Biosynthesis. *Plant Cell Physiol.* **2022**, *63*, 1140–1155. [CrossRef] [PubMed]
99. Gan, L.; Fan, H.; Nie, W.; Guo, Y. Ascorbic acid synthesis and transportation capacity in old laying hens and the effects of dietary supplementation with ascorbic acid. *J. Anim. Sci. Biotechnol.* **2018**, *9*, 71. [CrossRef] [PubMed]
100. Goggin, F.L.; Avila, C.A.; Lorence, A. Vitamin C content in plants is modified by insects and influences susceptibility to herbivory. *BioEssays* **2010**, *32*, 777–790. [CrossRef]
101. Ishikawa, T.; Shigeoka, S. Recent advances in ascorbate biosynthesis and the physiological significance of ascorbate peroxidase in photosynthesizing organisms. *Biosci. Biotechnol. Biochem.* **2008**, *72*, 1143–1154. [CrossRef] [PubMed]
102. Loewus, F.A. Biosynthesis and metabolism of ascorbic acid in plants and of analogs of ascorbic acid in fungi. *Phytochemistry* **1999**, *52*, 193–210. [CrossRef]
103. Linster, C.L.; Van Schaftingen, E. Vitamin C: Biosynthesis, recycling and degradation in mammals. *FEBS J.* **2007**, *274*, 1–22. [CrossRef]
104. Linster, C.L.; Van Schaftingen, E. Glucuronate, the precursor of vitamin C, is directly formed from UDP-glucuronate in liver. *FEBS J.* **2006**, *273*, 1516–1527. [CrossRef]
105. Agius, F.; González-Lamothe, R.; Caballero, J.L.; Muñoz-Blanco, J.; Botella, M.A.; Valpuesta, V. Engineering increased Vitamin C levels in plants by overexpression of a D-galacturonic acid reductase. *Nat. Biotechnol.* **2003**, *21*, 177–181. [CrossRef]
106. Lorence, A.; Chevone, B.I.; Mendes, P.; Nessler, C.L. myo-Inositol Oxygenase Offers a Possible Entry Point into Plant Ascorbate Biosynthesis. *Plant Physiol.* **2004**, *134*, 1200–1205. [CrossRef]
107. Wolucka, B.A.; Van Montagu, M. GDP-mannose 3′,5′-epimerase forms GDP-L-gulose, a putative intermediate for the de novo biosynthesis of vitamin C in plants. *J. Biol. Chem.* **2003**, *278*, 47483–47490. [CrossRef]
108. Caspi, R.; Billington, R.; Keseler, I.M.; Kothari, A.; Krummenacker, M.; Midford, P.E.; Ong, W.K.; Paley, S.; Subhraveti, P.; Karp, P.D. The MetaCyc database of metabolic pathways and enzymes—A 2019 update. *Nucleic Acids Res.* **2020**, *48*, D445–D453. [CrossRef] [PubMed]
109. Drouin, G.; Godin, J.R.; Pagé, B. The genetics of vitamin C loss in vertebrates. *Curr. Genom.* **2011**, *12*, 371–378. [CrossRef] [PubMed]
110. Yang, H. Conserved or Lost: Molecular Evolution of the Key Gene GULO in Vertebrate Vitamin C Biosynthesis. *Biochem. Genet.* **2013**, *51*, 413–425. [CrossRef] [PubMed]
111. Lachapelle, M.Y.; Drouin, G. Inactivation dates of the human and guinea pig vitamin C genes. *Genetica* **2011**, *139*, 199–207. [CrossRef] [PubMed]
112. López-Fernández, H.; Duque, P.; Henriques, S.; Vázquez, N.; Fdez-Riverola, F.; Vieira, C.P.; Reboiro-Jato, M.; Vieira, J. A bioinformatics protocol for quickly creating large-scale phylogenetic trees. In *Practical Applications of Computational Biology and Bioinformatics, 12th International Conference*; Fdez-Riverola, F., Mohamad, M.S., Rocha, M., De Paz, J.F., González, P., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 88–96.
113. Chatterjee, I.B. Evolution and the Biosynthesis of Ascorbic Acid. *Science* **1973**, *182*, 1271–1272. [CrossRef] [PubMed]
114. Henriques, S.F.; Duque, P.; López-Fernández, H.; Vázquez, N.; Fdez-Riverola, F.; Reboiro-Jato, M.; Vieira, C.P.; Vieira, J. Multiple independent L-gulonolactone oxidase (GULO) gene losses and vitamin C synthesis reacquisition events in non-Deuterostomian animal species. *BMC Evol. Biol.* **2019**, *19*, 126. [CrossRef]
115. Patananan, A.N.; Budenholzer, L.M.; Pedraza, M.E.; Torres, E.R.; Adler, L.N.; Clarke, S.G. The invertebrate Caenorhabditis elegans biosynthesizes ascorbate. *Arch. Biochem. Biophys.* **2015**, *569*, 32–44. [CrossRef]
116. Hou, M.; Dai, T.M.; Liang, X.Y.; Zhang, S.X.; Cui, W.Z.; Qiu, J.F.; Sima, Y.H.; Cui, W.Z.; Xu, S.Q. *Bombyx mori* can synthesize ascorbic acid through the l-gulose pathway to varying degrees depending on developmental stage. *Arch. Insect Biochem. Physiol.* **2021**, *106*, e21783. [CrossRef]
117. Hänninen, O.; Raunio, R.; Marniemi, J. Biosynthesis of L-ascorbic acid from myo-inositol. *Carbohydr. Res.* **1971**, *16*, 343–351. [CrossRef]
118. Duque, P.; Vieira, C.P.; Bastos, B.; Vieira, J. The evolution of vitamin C biosynthesis and transport in animals. *BMC Ecol. Evol.* **2022**, *22*, 84. [CrossRef]
119. Kramer, K.J.; Seib, P.A. Ascorbic acid and the growth and development of insects. In *Ascorbic Acid: Chemistry, Metabolism, and Uses*; Advances in Chemistry; American Chemical Society: Washington, DC, USA, 1982; Volume 200, pp. 275–291.
120. Massie, H.R.; Shumway, M.E.; Whitney, S.J.P.; Sternick, S.M.; Aiello, V.R. Ascorbic acid in Drosophila and changes during aging. *Exp. Gerontol.* **1991**, *26*, 487–494. [CrossRef]
121. Steinert, R.E.; Lee, Y.-K.; Sybesma, W. Vitamins for the Gut Microbiome. *Trends Mol. Med.* **2020**, *26*, 137–140. [CrossRef] [PubMed]
122. Pierre, L.L. Synthesis of Ascorbic Acid by the Normal Fat-Body of the Cockroach, *Leucophaea maderae* (F.), and by its Symbionts. *Nature* **1962**, *193*, 904–905. [CrossRef] [PubMed]
123. Travica, N.; Ried, K.; Hudson, I.; Sali, A.; Scholey, A.; Pipingas, A. The Contribution of Plasma and Brain Vitamin C on Age and Gender-Related Cognitive Differences: A Mini-Review of the Literature. *Front. Integr. Neurosci.* **2020**, *14*, 47. [CrossRef]
124. Aumailley, L.; Bourassa, S.; Gotti, C.; Droit, A.; Lebel, M. Vitamin C Differentially Impacts the Serum Proteome Profile in Female and Male Mice. *J. Proteome Res.* **2021**, *20*, 5036–5053. [CrossRef]

125. Kuo, S.-M.; MacLean, M.E.; McCormick, K.; Wilson, J.X. Gender and sodium-ascorbate transporter isoforms determine ascorbate concentrations in mice. *J. Nutr.* **2004**, *134*, 2216–2221. [CrossRef]
126. MacMillan, H.A.; Knee, J.M.; Dennis, A.B.; Udaka, H.; Marshall, K.E.; Merritt, T.J.S.; Sinclair, B.J. Cold acclimation wholly reorganizes the *Drosophila melanogaster* transcriptome and metabolome. *Sci. Rep.* **2016**, *6*, 28999. [CrossRef]
127. Arboleda-Bustos, C.E.; Segarra, C. The Dca gene involved in cold adaptation in *Drosophila melanogaster* arose by duplication of the ancestral regucalcin gene. *Mol. Biol. Evol.* **2011**, *28*, 2185–2195. [CrossRef]
128. Goto, S.G. Expression of Drosophila homologue of senescence marker protein-30 during cold acclimation. *J. Insect Physiol.* **2000**, *46*, 1111–1120. [CrossRef]
129. Lee, S.F.; Chen, Y.; Varan, A.K.; Wee, C.W.; Rako, L.; Axford, J.K.; Good, R.T.; Blacket, M.J.; Reuter, C.; Partridge, L. Molecular basis of adaptive shift in body size in *Drosophila melanogaster*: Functional and sequence analyses of the Dca gene. *Mol. Biol. Evol.* **2011**, *28*, 2393–2402. [CrossRef]
130. McKechnie, S.W.; Blacket, M.; Song, S.; Rako, L.; Carroll, X.; Johnson, T.; Jensen, L.; Lee, S.; Wee, C.; Hoffmann, A.A. A clinally varying promoter polymorphism associated with adaptive variation in wing size in *Drosophila*. *Mol. Ecol.* **2010**, *19*, 775–784. [CrossRef] [PubMed]
131. Qin, W.; Neal, S.J.; Robertson, R.M.; Westwood, J.T.; Walker, V.K. Cold hardening and transcriptional change in *Drosophila melanogaster*. *Insect Mol. Biol.* **2005**, *14*, 607–613. [CrossRef] [PubMed]
132. Sinclair, B.J.; Gibbs, A.G.; Roberts, S.P. Gene transcription during exposure to, and recovery from, cold and desiccation stress in *Drosophila melanogaster*. *Insect Mol. Biol.* **2007**, *16*, 435–443. [CrossRef] [PubMed]
133. Yamaguchi, M. The transcriptional regulation of regucalcin gene expression. *Mol. Cell. Biochem.* **2011**, *346*, 147–171. [CrossRef] [PubMed]
134. Tian, L.; Guo, E.; Wang, S.; Liu, S.; Jiang, R.-J.; Cao, Y.; Ling, E.; Li, S. Developmental Regulation of Glycolysis by 20-hydroxyecdysone and Juvenile Hormone in Fat Body Tissues of the Silkworm, *Bombyx mori*. *J. Mol. Cell Biol.* **2010**, *2*, 255–263. [CrossRef]
135. Gupta, S.D.; Choudhury, P.; Chatterjee, I. Synthesis of l-ascorbic acid from d-glucurono-1, 4-lactone conjugates by different species of animals. *Int. J. Biochem.* **1973**, *4*, 309–314. [CrossRef]
136. Ciufo, L.F.; Murray, P.A.; Thompson, A.; Rigden, D.J.; Rees, H.H. Characterisation of a desmosterol reductase involved in phytosterol dealkylation in the silkworm, *Bombyx mori*. *PLoS ONE* **2011**, *6*, e21316. [CrossRef] [PubMed]
137. Fujimori, H.; Zhou, Y.J.; Fukumura, K.; Matsumoto, S.; Tukamoto, Y.; Nagata, S. Specific distribution of expression and enzymatic activity of cholesterol biosynthetic enzyme DHCR24 orthologs in the phytophagous insect. *Biosci. Biotechnol. Biochem.* **2020**, *84*, 126–133. [CrossRef] [PubMed]
138. Ewing, T.A.; Fraaije, M.W.; Mattevi, A.; van Berkel, W.J. The VAO/PCMH flavoprotein family. *Arch. Biochem. Biophys.* **2017**, *632*, 104–117. [CrossRef]
139. Zerenturk, E.J.; Sharpe, L.J.; Ikonen, E.; Brown, A.J. Desmosterol and DHCR24: Unexpected new directions for a terminal step in cholesterol synthesis. *Prog. Lipid Res.* **2013**, *52*, 666–680. [CrossRef]
140. Takeshima, M.; Ogihara, M.H.; Kataoka, H. Characterization and functional analysis of BmSR-B1 for phytosterol uptake. *Steroids* **2022**, *184*, 109039. [CrossRef] [PubMed]
141. Leferink, N.G.H.; van Berkel, W.J.H. Aldonolactone Oxidoreductases. In *Flavins and Flavoproteins: Methods and Protocols*; Weber, S., Schleicher, E., Eds.; Springer: New York, NY, USA, 2014; pp. 95–111.
142. Kanafi, R.R.; Ebadi, R.; Mirhosseini, S.; Seidavi, A.; Zolfaghari, M.; Etebari, K. A review on nutritive effect of mulberry leaves enrichment with vitamins on economic traits and biological parameters of silkworm *Bombyx mori* L. *Invertebr. Surviv. J.* **2007**, *4*, 86–91.
143. Cappellozza, L.; Cappellozza, S.; Saviane, A.; Sbrenna, G. Artificial diet rearing system for the silkworm *Bombyx mori* (Lepidoptera: Bombycidae): Effect of vitamin C deprivation on larval growth and cocoon production. *Appl. Entomol. Zool.* **2005**, *40*, 405–412. [CrossRef]
144. Etebari, K.; Matindoost, L. Application of Multi-vitamins as Supplementary Nutrients on Biological and Economical Characteristics of Silkworm *Bombyx mori* L. *J. Asia-Pac. Entomol.* **2005**, *8*, 107–112. [CrossRef]
145. Vinci, G.; Xia, X.; Veitia, R.A. Preservation of Genes Involved in Sterol Metabolism in Cholesterol Auxotrophs: Facts and Hypotheses. *PLoS ONE* **2008**, *3*, e2883. [CrossRef] [PubMed]
146. Valpuesta, V.; Botella, M.A. Biosynthesis of L-ascorbic acid in plants: New pathways for an old antioxidant. *Trends Plant Sci.* **2004**, *9*, 573–577. [CrossRef] [PubMed]
147. Wolucka, B.A.; Van Montagu, M. The VTC2 cycle and the de novo biosynthesis pathways for vitamin C in plants: An opinion. *Phytochemistry* **2007**, *68*, 2602–2613. [CrossRef]
148. Yamauchi, N.; Nakayama, Y. Biosynthetic Mechanism for l-Gulose in Main Polar Lipids of Thermoplasma acidophilum and Possible Resemblance to Plant Ascorbic Acid Biosynthesis. *Biosci. Biotechnol. Biochem.* **2013**, *77*, 2087–2093. [CrossRef]
149. Major, L.L.; Wolucka, B.A.; Naismith, J.H. Structure and Function of GDP-Mannose-3′,5′-Epimerase: An Enzyme which Performs Three Chemical Reactions at the Same Active Site. *J. Am. Chem. Soc.* **2005**, *127*, 18309–18320. [CrossRef]
150. Beerens, K.; Gevaert, O.; Desmet, T. GDP-Mannose 3,5-Epimerase: A View on Structure, Mechanism, and Industrial Potential. *Front. Mol. Biosci.* **2021**, *8*, 784142. [CrossRef] [PubMed]

151. Goudsmit, E.M.; Neufeld, E.F. Formation of GDP-L-galactose from GDP-D-mannose. *Biochem. Biophys. Res. Commun.* **1967**, *26*, 730–735. [CrossRef]
152. Woodyer, R.D.; Christ, T.N.; Deweese, K.A. Single-step bioconversion for the preparation of L-gulose and L-galactose. *Carbohydr. Res.* **2010**, *345*, 363–368. [CrossRef] [PubMed]
153. Roos, C.; Kolmer, M.; Mattila, P.; Renkonen, R. Composition of *Drosophila melanogaster* proteome involved in fucosylated glycan metabolism. *J. Biol. Chem.* **2002**, *277*, 3168–3175. [CrossRef] [PubMed]
154. Thorsell, A.-G.; Persson, C.; Voevodskaya, N.; Busam, R.D.; Hammarström, M.; Gräslund, S.; Gräslund, A.; Hallberg, B.M. Structural and biophysical characterization of human myo-inositol oxygenase. *J. Biol. Chem.* **2008**, *283*, 15209–15216. [CrossRef] [PubMed]
155. Brown, P.M.; Caradoc-Davies, T.T.; Dickson, J.M.; Cooper, G.J.; Loomes, K.M.; Baker, E.N. Crystal structure of a substrate complex of *myo*-inositol oxygenase, a di-iron oxygenase with a key role in inositol metabolism. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15032–15037. [CrossRef]
156. Chang, H.-H.; Chao, H.-N.; Walker, C.S.; Choong, S.-Y.; Phillips, A.; Loomes, K.M. Renal depletion of myo-inositol is associated with its increased degradation in animal models of metabolic disease. *Am. J. Physiol. Ren. Physiol.* **2015**, *309*, F755–F763. [CrossRef] [PubMed]
157. Lu, Y.; Agarwal, A. Myo-inositol oxygenase in cadmium-induced kidney injury. *Am. J. Physiol. Ren. Physiol.* **2022**, *322*, F470–F472. [CrossRef] [PubMed]
158. Oggioni, G. Experiments on the antiscorbutic action of inositol. *Boll. Della Soc. Ital. Biol. Sper.* **1953**, *29*, 1421–1422.
159. Anderson, L.; Coots, R.H.; Halliday, J.W. A Critical Evaluation of Myo-Inositol as an Ascorbic Acid-Sparing Agent. *J. Nutr.* **1958**, *64*, 167–175. [CrossRef]
160. Peterkofsky, B. Ascorbate requirement for hydroxylation and secretion of procollagen: Relationship to inhibition of collagen synthesis in scurvy. *Am. J. Clin. Nutr.* **1991**, *54*, 1135S–1140S. [CrossRef]
161. Chatterjee, I.B.; Gupta, S.D.; Majumder, A.K.; Nandi, B.K.; Subramanian, N. Effect of ascorbic acid on histamine metabolism in scorbutic guinea-pigs. *J. Physiol.* **1975**, *251*, 271–279. [CrossRef] [PubMed]
162. Freinkel, N.; Dawson, R.M.C.; Ingbar, S.H.; White, R.W. The Free Myo-Inositol of Thyroid Tissue. *Proc. Soc. Exp. Biol. Med.* **1959**, *100*, 549–551. [CrossRef]
163. Dawson, R.; Freinkel, N. The distribution of free mesoinositol in mammalian tissues, including some observations on the lactating rat. *Biochem. J.* **1961**, *78*, 606. [CrossRef] [PubMed]
164. Halver, J.E.; Hardy, R.W. L-Ascorbyl-2-Sulfate Alleviates Atlantic Salmon Scurvy. *Proc. Soc. Exp. Biol. Med.* **1994**, *206*, 421–424. [CrossRef] [PubMed]
165. Halver, J.E.; Smith, R.R.; Tolbert, B.M.; Baker, E.M. Utilization of ascorbic acid in fish. *Ann. N. Y. Acad. Sci.* **1975**, *258*, 81–102. [CrossRef] [PubMed]
166. Howard, C.F.; Anderson, L. Metabolism of myo-inositol in animals: II. Complete catabolism of myo-inositol-14C by rat kidney slices. *Arch. Biochem. Biophys.* **1967**, *118*, 332–339. [CrossRef]
167. Burns, J.; Evans, C. The synthesis of L-ascorbic acid in the rat from D-glucuronolactone and L-gulonolactone. *J. Biol. Chem* **1956**, *223*, 897–905. [CrossRef]
168. Bánhegyi, G.; Braun, L.; Csala, M.; Puskás, F.; Mandl, J. Ascorbate Metabolism and Its Regulation in Animals. *Free Radic. Biol. Med.* **1997**, *23*, 793–803. [CrossRef]
169. Jenness, R.; Birney, E.C.; Ayaz, K.L.; Buzzell, D.M. Ontogenetic development of l-gulonolactone oxidase activity in several vertebrates. *Comp. Biochem. Physiol. Part B Comp. Biochem.* **1984**, *78*, 167–173. [CrossRef]
170. Walter Bock, K.; Köhle, C. UDP-Glucuronosyltransferase 1A6: Structural, Functional, and Regulatory Aspects. In *Methods in Enzymology*; Academic Press: Cambridge, MA, USA, 2005; Volume 400, pp. 57–75.
171. Strassburg, C.P.; Strassburg, A.; Kneip, S.; Barut, A.; Tukey, R.H.; Rodeck, B.; Manns, M.P. Developmental aspects of human hepatic drug glucuronidation in young children and adults. *Gut* **2002**, *50*, 259. [CrossRef]
172. Kiani, A.; Paolacci, S.; Calogero, A.; Cannarella, R.; Di Renzo, G.; Gerli, S.; Della Morte, C.; Busetto, G.; De Berardinis, E.; Del Giudice, F. From Myo-inositol to D-chiro-inositol molecular pathways. *Eur. Rev. Med. Pharmacol. Sci* **2021**, *25*, 2390. [CrossRef] [PubMed]
173. Arner, R.J.; Prabhu, K.S.; Reddy, C.C. Molecular cloning, expression, and characterization of myo-inositol oxygenase from mouse, rat, and human kidney. *Biochem. Biophys. Res. Commun.* **2004**, *324*, 1386–1392. [CrossRef] [PubMed]
174. Sochor, M.; Baquer, N.Z.; McLean, P. Regulation of pathways of glucose metabolism in kidney. The effect of experimental diabetes on the activity of the pentose phosphate pathway and the glucuronate-xylulose pathway. *Arch. Biochem. Biophys.* **1979**, *198*, 632–646. [CrossRef]
175. Barski, O.A.; Papusha, V.Z.; Ivanova, M.M.; Rudman, D.M.; Finegold, M.J. Developmental expression and function of aldehyde reductase in proximal tubules of the kidney. *Am. J. Physiol. Ren. Physiol.* **2005**, *289*, F200–F207. [CrossRef] [PubMed]
176. Agrawal, N.; Hossain, M.S.; Skelton, A.A.; Muralidhar, K.; Kaushik, S. Unraveling the mechanism of l-gulonate-3-dehydrogenase inhibition by ascorbic acid: Insights from molecular modeling. *Comput. Biol. Chem.* **2018**, *77*, 146–153. [CrossRef] [PubMed]
177. Birney, E.C.; Jenness, R.; Hume, I.D. Ascorbic acid biosynthesis in the mammalian kidney. *Experientia* **1979**, *35*, 1425–1426. [CrossRef] [PubMed]
178. Chaudhuri, C.R.; Chatterjee, I.B. L-Ascorbic Acid Synthesis in Birds: Phylogenetic Trend. *Science* **1969**, *164*, 435–436. [CrossRef]

179. Nishikimi, M.; Yagi, K. Biochemistry and Molecular Biology of Ascorbic Acid Biosynthesis. In *Subcellular Biochemistry: Ascorbic Acid: Biochemistry and Biomedical Cell Biology*; Harris, J.R., Ed.; Springer: Boston, MA, USA, 1996; pp. 17–39.
180. Roy, R.N.; Guha, B.C. Species Difference in regard to the Biosynthesis of Ascorbic Acid. *Nature* **1958**, *182*, 319–320. [CrossRef]
181. Dabrowski, K. Primitive actimoterigian fishes can synthesize ascorbic acid. *Experientia* **1994**, *50*, 745–748. [CrossRef]
182. White, R.E.; Evans, D.C.; Hop, C.E.C.A.; Moore, D.J.; Prakash, C.; Surapaneni, S.; Tse, F.L.S. Radiolabeled mass-balance excretion and metabolism studies in laboratory animals: A commentary on why they are still necessary. *Xenobiotica* **2013**, *43*, 219–225. [CrossRef]
183. Chitwood, D.J. Biochemistry and Function of Nematode Steroids. *Crit. Rev. Biochem. Mol. Biol.* **1999**, *34*, 273–284. [CrossRef] [PubMed]
184. Waterham, H.R.; Koster, J.; Romeijn, G.J.; Hennekam, R.C.M.; Vreken, P.; Andersson, H.C.; FitzPatrick, D.R.; Kelley, R.I.; Wanders, R.J.A. Mutations in the 3β-Hydroxysterol Δ24-Reductase Gene Cause Desmosterolosis, an Autosomal Recessive Disorder of Cholesterol Biosynthesis. *Am. J. Hum. Genet.* **2001**, *69*, 685–694. [CrossRef] [PubMed]
185. Jones, M.K.; Eldon, E.D.; Klig, L.S. myo-Inositol oxygenase identified in *Drosophila melanogaster*. *FASEB J.* **2012**, *26*, 759.13. [CrossRef]
186. Lai, C.W.; Chen, H.L.; Tu, M.Y.; Lin, W.Y.; Röhrig, T.; Yang, S.H.; Lan, Y.W.; Chong, K.Y.; Chen, C.M. A novel osteoporosis model with ascorbic acid deficiency in Akr1A1 gene knockout mice. *Oncotarget* **2017**, *8*, 7357–7369. [CrossRef]
187. Gabbay, K.H.; Bohren, K.M.; Morello, R.; Bertin, T.; Liu, J.; Vogel, P. Ascorbate synthesis pathway: Dual role of ascorbate in bone homeostasis. *J. Biol. Chem.* **2010**, *285*, 19510–19520. [CrossRef]
188. O'Connor, T.; Ireland, L.S.; Harrison, D.J.; Hayes, J.D. Major differences exist in the function and tissue-specific expression of human aflatoxin B1 aldehyde reductase and the principal human aldo-keto reductase AKR1 family members. *Biochem. J.* **1999**, *343*, 487–504. [CrossRef] [PubMed]
189. Di Luccio, E.; Elling, R.A.; Wilson, D.K. Identification of a novel NADH-specific aldo-keto reductase using sequence and structural homologies. *Biochem. J.* **2006**, *400*, 105–114. [CrossRef]
190. Jing, X.; Behmer, S.T. Insect sterol nutrition: Physiological mechanisms, ecology, and applications. *Annu. Rev. Entomol.* **2020**, *65*, 251–271. [CrossRef]
191. Lavrynenko, O.; Rodenfels, J.; Carvalho, M.; Dye, N.A.; Lafont, R.; Eaton, S.; Shevchenko, A. The ecdysteroidome of *Drosophila*: Influence of diet and development. *Development* **2015**, *142*, 3758–3768. [CrossRef]
192. Svoboda, J.A.; Feldlaufer, M.F. Neutral sterol metabolism in insects. *Lipids* **1991**, *26*, 614–618. [CrossRef]
193. Chang, Y.; Wang, S.; Sekimoto, S.; Aerts, A.L.; Choi, C.; Clum, A.; LaButti, K.M.; Lindquist, E.A.; Yee Ngan, C.; Ohm, R.A.; et al. Phylogenomic Analyses Indicate that Early Fungi Evolved Digesting Cell Walls of Algal Ancestors of Land Plants. *Genome Biol. Evol.* **2015**, *7*, 1590–1601. [CrossRef] [PubMed]
194. Gulbrandsen, Ø.S.; Andresen, I.J.; Krabberød, A.K.; Bråte, J.; Shalchian-Tabrizi, K. Phylogenomic analysis restructures the ulvophyceae. *J. Phycol.* **2021**, *57*, 1223–1233. [CrossRef] [PubMed]
195. Yoon, H.S.; Müller, K.M.; Sheath, R.G.; Ott, F.D.; Bhattacharya, D. Defining the major lineages of red algae (Rhodophyta). *J. Phycol.* **2006**, *42*, 482–492. [CrossRef]
196. Maddison, D.R.; Schulz, K.-S.; Maddison, W.P. The tree of life web project. *Zootaxa* **2007**, *1668*, 19–40. [CrossRef]
197. Gray, M.W. Mitochondrial evolution. *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, a011403. [CrossRef] [PubMed]
198. Jensen, P.E.; Leister, D. Chloroplast evolution, structure and functions. *F1000Prime Rep.* **2014**, *6*, 40. [CrossRef] [PubMed]
199. Huang, C.Y.; Ayliffe, M.A.; Timmis, J.N. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* **2003**, *422*, 72–76. [CrossRef]
200. Henze, K.; Martin, W. How do mitochondrial genes get into the nucleus? *Trends Genet.* **2001**, *17*, 383–387. [CrossRef]
201. Kleine, T.; Maier, U.G.; Leister, D. DNA transfer from organelles to the nucleus: The idiosyncratic genetics of endosymbiosis. *Annu. Rev. Plant Biol.* **2009**, *60*, 115–138. [CrossRef] [PubMed]
202. Aboobucker, S.I.; Lorence, A. Recent progress on the characterization of aldonolactone oxidoreductases. *Plant Physiol. Biochem. PPB* **2016**, *98*, 171–185. [CrossRef]
203. Aboobucker, S.I.; Suza, W.P.; Lorence, A. Characterization of Two Arabidopsis L-Gulono-1,4-lactone Oxidases, AtGulLO3 and AtGulLO5, Involved in Ascorbate Biosynthesis. *React. Oxyg. Species* **2017**, *4*, 389–417. [CrossRef]
204. Keeling, P.J.; Palmer, J.D. Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* **2008**, *9*, 605–618. [CrossRef]
205. Husnik, F.; McCutcheon, J.P. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat. Rev. Microbiol.* **2018**, *16*, 67–79. [CrossRef] [PubMed]
206. Berriman, M.; Ghedin, E.; Hertz-Fowler, C.; Blandin, G.; Renauld, H.; Bartholomeu, D.C.; Lennard, N.J.; Caler, E.; Hamlin, N.E.; Haas, B.; et al. The genome of the African trypanosome *Trypanosoma brucei*. *Science* **2005**, *309*, 416–422. [CrossRef] [PubMed]
207. Bock, R. The give-and-take of DNA: Horizontal gene transfer in plants. *Trends Plant Sci.* **2010**, *15*, 11–22. [CrossRef]
208. Richards, T.A.; Leonard, G.; Soanes, D.M.; Talbot, N.J. Gene transfer into the fungi. *Fungal Biol. Rev.* **2011**, *25*, 98–110. [CrossRef]
209. Dunning Hotopp, J.C. Horizontal gene transfer between bacteria and animals. *Trends Genet.* **2011**, *27*, 157–163. [CrossRef] [PubMed]
210. Rosewich, U.L.; Kistler, H.C. Role of horizontal gene transfer in the evolution of fungi. *Annu. Rev. Phytopathol.* **2000**, *38*, 325–363. [CrossRef] [PubMed]
211. Fitzpatrick, D.A. Horizontal gene transfer in fungi. *FEMS Microbiol. Lett.* **2012**, *329*, 1–8. [CrossRef] [PubMed]

212. Schmitt, I.; Lumbsch, H.T. Ancient Horizontal Gene Transfer from Bacteria Enhances Biosynthetic Capabilities of Fungi. *PLoS ONE* **2009**, *4*, e4437. [CrossRef] [PubMed]
213. Hancock, R.D.; Galpin, J.R.; Viola, R. Biosynthesis of L-ascorbic acid (vitamin C) by *Saccharomyces cerevisiae*. *FEMS Microbiol. Lett.* **2000**, *186*, 245–250. [CrossRef] [PubMed]
214. Edgell, D.R.; Belfort, M.; Shub, D.A. Barriers to Intron Promiscuity in Bacteria. *J. Bacteriol.* **2000**, *182*, 5281–5289. [CrossRef] [PubMed]
215. Nielsen, C.B.; Friedman, B.; Birren, B.; Burge, C.B.; Galagan, J.E. Patterns of Intron Gain and Loss in Fungi. *PLoS Biol.* **2004**, *2*, e422. [CrossRef] [PubMed]
216. Lynch, M.; Richardson, A.O. The evolution of spliceosomal introns. *Curr. Opin. Genet. Dev.* **2002**, *12*, 701–710. [CrossRef]
217. Philippe, H.; Douady, C.J. Horizontal gene transfer and phylogenetics. *Curr. Opin. Microbiol.* **2003**, *6*, 498–505. [CrossRef]
218. Hugenholtz, P.; Chuvochina, M.; Oren, A.; Parks, D.H.; Soo, R.M. Prokaryotic taxonomy and nomenclature in the age of big sequence data. *ISME J.* **2021**, *15*, 1879–1892. [CrossRef]
219. López-Fernández, H.; Duque, P.; Vázquez, N.; Fdez-Riverola, F.; Reboiro-Jato, M.; Vieira, C.P.; Vieira, J. SEDA: A Desktop Tool Suite for FASTA Files Processing. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2022**, *19*, 1850–1860. [CrossRef]
220. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef]
221. Notredame, C.; Higgins, D.G.; Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **2000**, *302*, 205–217. [CrossRef]
222. Huelsenbeck, J.P.; Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **2001**, *17*, 754–755. [CrossRef]
223. Reboiro-Jato, D.; Reboiro-Jato, M.; Fdez-Riverola, F.; Vieira, C.P.; Fonseca, N.A.; Vieira, J. ADOPS-Automatic detection of positively selected sites. *J. Integr. Bioinform.* **2012**, *9*, 18–32. [CrossRef]
224. López-Fernández, H.; Ferreira, P.; Reboiro-Jato, M.; Vieira, C.P.; Vieira, J. The pegi3s Bioinformatics Docker Images Project. In *Practical Applications of Computational Biology & Bioinformatics, 15th International Conference (PACBB 2021)*; Rocha, M., Fdez-Riverola, F., Mohamad, M.S., Casado-Vara, R., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 31–40.
225. Klein, C.C.; Alves, J.M.P.; Serrano, M.G.; Buck, G.A.; Vasconcelos, A.T.R.; Sagot, M.-F.; Teixeira, M.M.G.; Camargo, E.P.; Motta, M.C.M. Biosynthesis of Vitamins and Cofactors in Bacterium-Harbouring Trypanosomatids Depends on the Symbiotic Association as Revealed by Genomic Analyses. *PLoS ONE* **2013**, *8*, e79786. [CrossRef] [PubMed]
226. Logan, F.J.; Taylor, M.C.; Wilkinson, S.R.; Kaur, H.; Kelly, J.M. The terminal step in vitamin C biosynthesis in *Trypanosoma cruzi* is mediated by a FMN-dependent galactonolactone oxidase. *Biochem. J* **2007**, *407*, 419–426. [CrossRef] [PubMed]
227. Wilkinson, S.R.; Prathalingam, S.R.; Taylor, M.C.; Horn, D.; Kelly, J.M. Vitamin C biosynthesis in trypanosomes: A role for the glycosome. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 11645–11650. [CrossRef]
228. Biyani, N.; Madhubala, R. Leishmania donovani encodes a functional enzyme involved in vitamin c biosynthesis: Arabino-1,4-lactone oxidase. *Mol. Biochem. Parasitol.* **2011**, *180*, 76–85. [CrossRef]
229. Maruta, T.; Ichikawa, Y.; Mieda, T.; Takeda, T.; Tamoi, M.; Yabuta, Y.; Ishikawa, T.; Shigeoka, S. The contribution of Arabidopsis homologs of L-gulono-1, 4-lactone oxidase to the biosynthesis of ascorbic acid. *Biosci. Biotechnol. Biochem.* **2010**, *74*, 1494–1497. [CrossRef]
230. Mewies, M.; McIntire, W.S.; Scrutton, N.S. Covalent attachment of flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) to enzymes: The current state of affairs. *Protein Sci.* **1998**, *7*, 7–20. [CrossRef]
231. Fraaije, M.W.; van Den Heuvel, R.H.; van Berkel, W.J.; Mattevi, A. Structural analysis of flavinylation in vanillyl-alcohol oxidase. *J. Biol. Chem.* **2000**, *275*, 38654–38658. [CrossRef]
232. Leferink, N.G.; Jose, M.D.; van den Berg, W.A.; van Berkel, W.J. Functional assignment of Glu386 and Arg388 in the active site of L-galactono-γ-lactone dehydrogenase. *FEBS Lett.* **2009**, *583*, 3199–3203. [CrossRef]
233. Imachi, H.; Nobu, M.K.; Nakahara, N.; Morono, Y.; Ogawara, M.; Takaki, Y.; Takano, Y.; Uematsu, K.; Ikuta, T.; Ito, M.; et al. Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* **2020**, *577*, 519–525. [CrossRef]
234. Akıl, C.; Robinson, R.C. Genomes of Asgard archaea encode profilins that regulate actin. *Nature* **2018**, *562*, 439–443. [CrossRef] [PubMed]
235. Wang, P.; Zeng, W.; Xu, S.; Du, G.; Zhou, J.; Chen, J. Current challenges facing one-step production of l-ascorbic acid. *Biotechnol. Adv.* **2018**, *36*, 1882–1899. [CrossRef] [PubMed]
236. Tucaliuc, A.; Cîșlaru, A.; Kloetzer, L.; Blaga, A.C. Strain Development, Substrate Utilization, and Downstream Purification of Vitamin C. *Processes* **2022**, *10*, 1595. [CrossRef]
237. Bremus, C.; Herrmann, U.; Bringer-Meyer, S.; Sahm, H. The use of microorganisms in l-ascorbic acid production. *J. Biotechnol.* **2006**, *124*, 196–205. [CrossRef] [PubMed]
238. Zeng, W.; Wang, P.; Li, N.; Li, J.; Chen, J.; Zhou, J. Production of 2-keto-L-gulonic acid by metabolically engineered *Escherichia coli*. *Bioresour. Technol.* **2020**, *318*, 124069. [CrossRef]
239. Shibata, T.; Ichikawa, C.; Matsuura, M.; Takata, Y.; Noguchi, Y.; Saito, Y.; Yamashita, M. Cloning of a gene for D-sorbitol dehydrogenase from *Gluconobacter oxydans* G624 and expression of the gene in *Pseudomonas putida* IFO3738. *J. Biosci. Bioeng.* **2000**, *89*, 463–468. [CrossRef]

240. Anderson, S.; Marks, C.B.; Lazarus, R.; Miller, J.; Stafford, K.; Seymour, J.; Light, D.; Rastetter, W.; Estell, D. Production of 2-keto-L-gulonate, an intermediate in L-ascorbate synthesis, by a genetically modified *Erwinia herbicola*. *Science* **1985**, *230*, 144–149. [CrossRef]
241. Hongyu, L.; Ceshi, C.; Guanglin, Y. Intergeneric cell fusion of Corynebacterium and Erwinia. *Wei Sheng Wu Xue Tong Bao* **1999**, *26*, 3–6.
242. Song, W.; Zheng, K.; Xu, X.; Gao, C.; Guo, L.; Liu, J.; Chen, X.; Liu, L.; Hu, G.; Wu, J. Enzymatic Production of Ascorbic Acid-2-Phosphate by Engineered Pseudomonas aeruginosa Acid Phosphatase. *J. Agric. Food Chem.* **2021**, *69*, 14215–14221. [CrossRef]

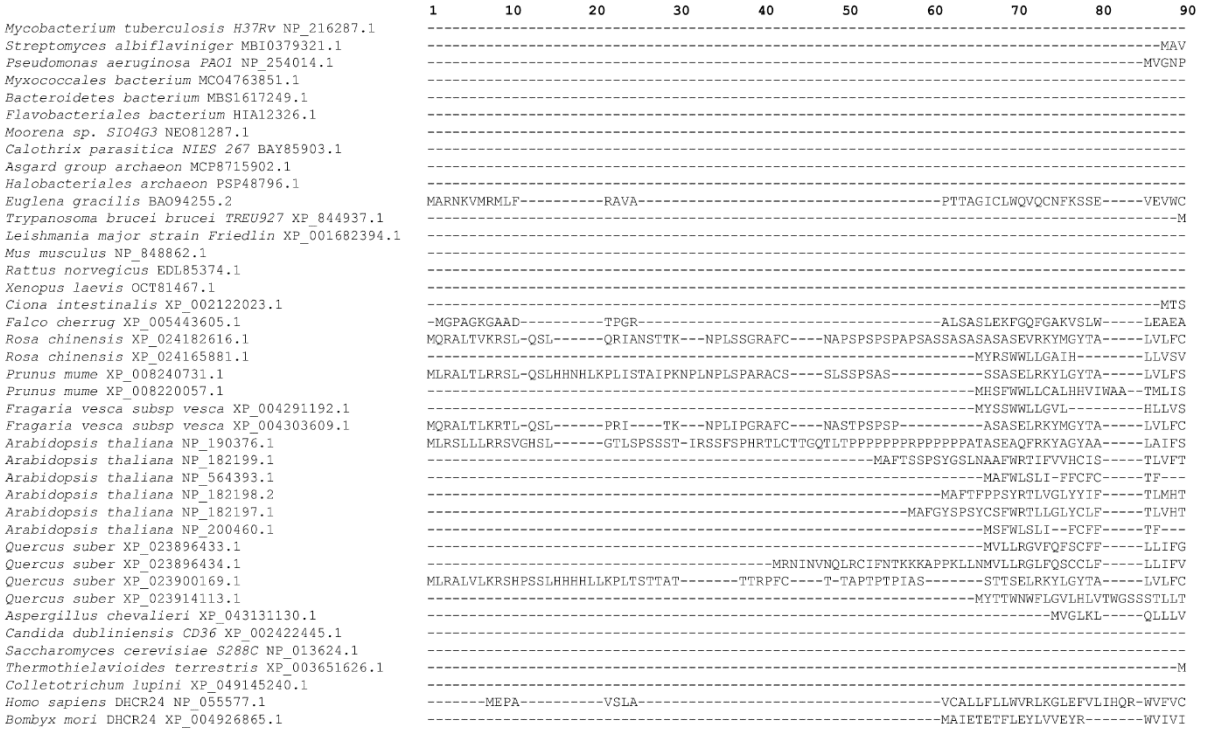
1.2.1. Supporting information

Advances in Novel Animal Vitamin C Biosynthesis Pathways and the Role of Prokaryote-Based Inferences to Understand Their Origin

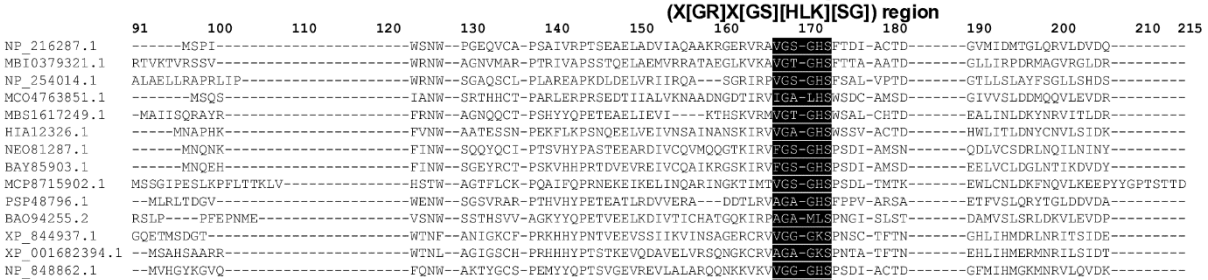
Pedro Duque, Cristina P. Vieira, Jorge Vieira

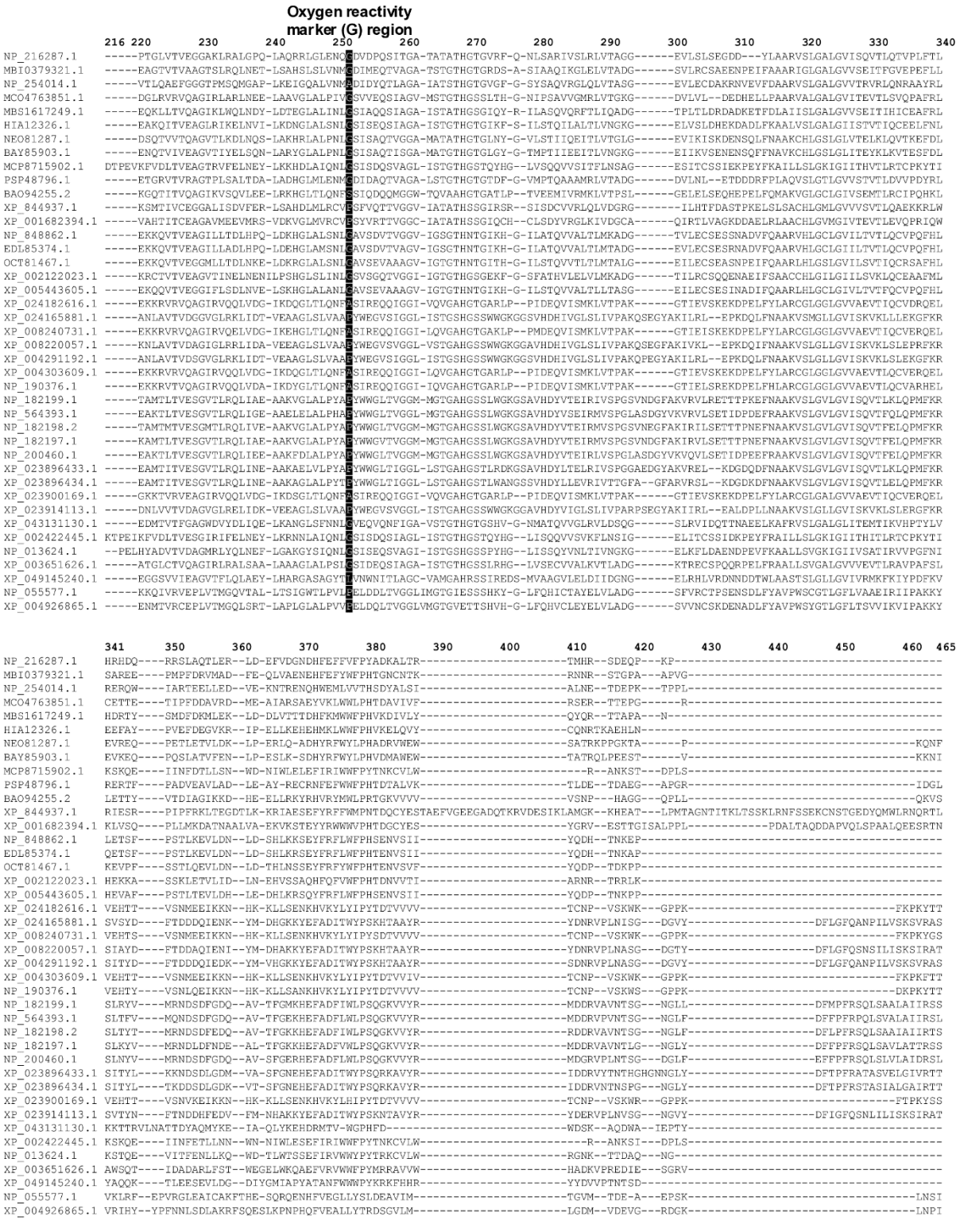
Supplementary Figures

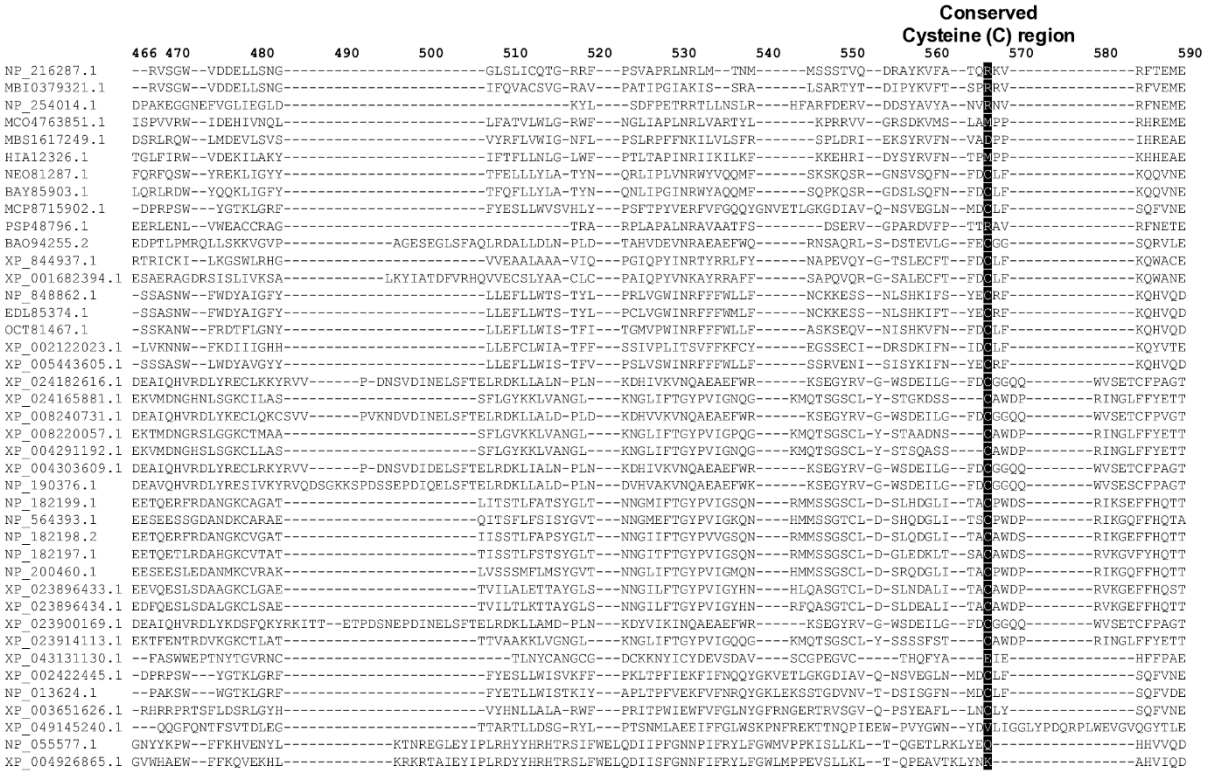
Figure S1 – Complete AO dataset MUSCLE alignment file. The alignment presented was the one used to perform the phylogenetic analyses discussed in the manuscript. The relevant amino acid regions used to extrapolate putative AO functions are tagged at the top of the respective alignment positions, and delimited by the black colour background.



FAD-Binding motif (X[GR]X[GS][HLK][SG]) region




Oxygen reactivity
marker (G) region


Conserved
Cysteine (C) region
Glu-Arg pair (EXR)
region

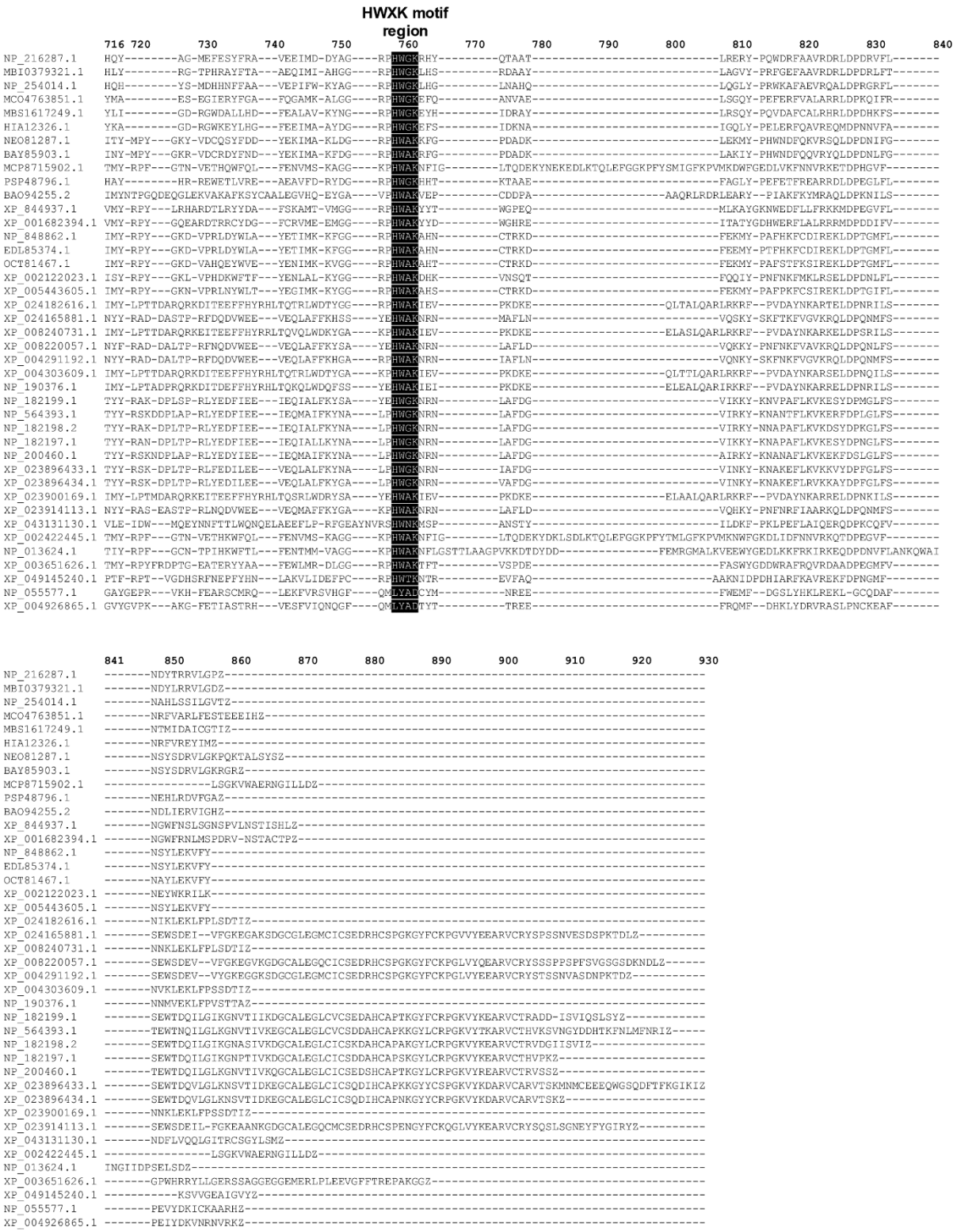


Figure S1

1.2.2. Addendum: Further information on *C. elegans* VC synthesis and considerations on VC transport mechanisms

1.2.2.1. Review article synopsis

The review article just presented summarized the main biological roles of VC in eukaryotes and prokaryotes, showing the notable importance of this nutrient for the optimal function and development of many species (Duque *et al.* 2022b). Furthermore, the article summarized the currently characterized VC synthesis pathways in eukaryotes, highlighting the increasing evidence for alternative VC synthesis pathways in insects and nematodes (Duque *et al.* 2022b). Additionally, several methodological approaches that could be useful for the future characterization of the VC synthesis pathways of *C. elegans* and *D. melanogaster* were also proposed (Duque *et al.* 2022b). Finally, the possibility that VC synthesis in eukaryotes might be derived from prokaryote origins was also preliminarily addressed, using for this purpose a sequence homology search for putative prokaryote aldonolactone oxidoreductases (AOs) (Duque *et al.* 2022b).

1.2.2.2. E03H4.3 and F54D5.12 as Regucalcin and GULO functional substitutes in *C. elegans*

Although the presented literature review on animal VC synthesis only contemplated information from Patananan *et al.* (Patananan *et al.* 2015) in regards to *C. elegans* endogenous VC production (Duque *et al.* 2022b), further research has revealed that this organism is capable of VC synthesis through an adaptation of the characterized animal pathway (Yabuta *et al.* 2020). These findings are here included to provide additional evolutionary context on this research topic. Yabuta *et al.* were initially able to replicate the incorporation of ^{13}C into the available VC pool using a ^{13}C-labeled *Escherichia coli* food source as seen in Patananan *et al.* (Patananan *et al.* 2015), and additionally correlated L-Gulono-1,4-lactone supplementation with increased VC levels (Yabuta *et al.* 2020). This latter result implicated the known animal VC precursor in endogenous VC production in *C. elegans*, a finding that correlated well with the observed UDP-D-Glucose carbon skeleton rearrangement that is known to occur during the mammal and photosynthetic protist VC synthesis process (Yabuta *et al.* 2020). Contradicting the majority of the literature on the subject (Duque *et al.* 2022a, López-Fernández *et al.* 2019, Patananan *et al.* 2015, Wheeler *et al.* 2015), the authors further identified putative homologs of *Regucalcin* and *GULO* (*E03H4.3* and *F54D5.12*, respectively) in *C. elegans* (Yabuta *et al.* 2020). According to the performed functional assays, *E03H4.3* knockdown and *F54D5.12* knockout led to approximately 18% and 19% reductions in total VC levels, respectively, hence implicating both genes in the VC synthesis process in this organism (Yabuta *et al.* 2020). These results

hint that although E03H4.3 and F54D5.12 might be functionally replacing Regucalcin and GULO in *C. elegans*, this metabolic complementation is only partially responsible for total VC levels, and thus, alternative VC synthesis mechanisms are likely present in this species. Nevertheless, in conjunction with the observations of 24-Dehydrocholesterol Reductase (DHCR24) functional adaptation in *B. mori* (Duque *et al.* 2022b, Hou *et al.* 2021), they highlight that selective pressures to reacquire endogenous VC synthesis are possibly prevalent and widespread within basal animal lineages.

1.2.2.3. Vitamin C transport: an uncharacterized mechanism in basal animal species

The homeostasis of VC greatly mediates the metabolic utility of this nutrient in all organisms. However, mechanisms responsible for adequate cellular VC concentrations were not considered within the scope of the original review article. Therefore, information on the main VC transporters in animal species is here presented to offer complementary background in the thesis context.

Independently of its source, VC needs to be in homeostasis for the optimal function of the organism (Burzle *et al.* 2013, Rivas *et al.* 2008, Savini *et al.* 2008). In vertebrates, this equilibrium is mainly maintained due to the presence of sodium-dependent VC transporters (SVCTs) and Sodium-independent facilitative glucose transporters (GLUTs) (Burzle *et al.* 2013, Rivas *et al.* 2008, Savini *et al.* 2008, Wang *et al.* 2023). GLUT1, GLUT3 and GLUT4 are known to passively transport an oxidized form of VC (dehydroascorbic acid ; DHA) rather than VC (Burzle *et al.* 2013, Rivas *et al.* 2008, Rumsey *et al.* 2000, Rumsey *et al.* 1997, Wang *et al.* 2023, Welch *et al.* 1995), and thus are considered VC recycling facilitators (Rivas *et al.* 2008). In contrast, SVCT1 and SVCT2 actively transport VC with high specificity (Burzle *et al.* 2013, Daruwala *et al.* 1999, May 2011, Rivas *et al.* 2008, Savini *et al.* 2008, Tsukaguchi *et al.* 1999), have no significant affinity for DHA (May 2011, Rivas *et al.* 2008, Savini *et al.* 2008, Tsukaguchi *et al.* 1999), and are known to play important roles in VC absorption through the digestive and renal systems and generalized VC distribution, respectively (Burzle *et al.* 2013, Corpe *et al.* 2010, Daruwala *et al.* 1999, Harrison *et al.* 2010, Savini *et al.* 2008, Tsukaguchi *et al.* 1999, Wang *et al.* 2023). SVCT1 and SVCT2 presence appears crucial in animals capable of VC synthesis, and arguably even more so in VC auxotroph species that require the dietary absorption of this nutrient. Indeed, the knockout of *SVCT1* in *Mus musculus* (mice) potentiates increased perinatal mortality and VC excretion levels, despite an unexpected increase in endogenous VC synthesis possibly linked to low body levels of this nutrient (Corpe *et al.* 2010). Furthermore, *SVCT2* knockout in this model organism resulted in a lethal phenotype associated with intracerebral hemorrhages (Harrison *et al.* 2010, Sotiriou *et al.* 2002). However, regardless

of their importance, evidence for the existence of VC transporters in non-bilaterian and protostomians is currently lacking.

SVCTs are surface glycoproteins of the nucleobase-ascorbate transporter (NAT) protein family (Burzle *et al.* 2013, Rivas *et al.* 2008, Savini *et al.* 2008). This protein family is divided into three distinct subgroups according to substrate specificity, namely xanthine/uric acid, uracil or VC (Burzle *et al.* 2013). The VC subgroup is constituted by four SVCTs (SVCT1 to 4), which are thought to be vertebrate-exclusive (Burzle *et al.* 2013, de Koning and Diallinas 2000, Yamamoto *et al.* 2010). Interestingly, only SVCT1 and SVCT2 were directly correlated with VC transport so far (Godoy *et al.* 2007, Tsukaguchi *et al.* 1999, Wang *et al.* 2000), sharing a unique and conserved amino acid motif within the transmembrane domain 10 (SSSP) that is likely related with their substrate specificity (Kourkoulou *et al.* 2018). The SVCT3 transporter function remains a mystery (Burzle *et al.* 2013, Wang *et al.* 2023). This protein is known to be expressed in the S3 segment of renal proximal tubules (kidney) of mice, but also in the kidney, heart and brain of humans (Burzle *et al.* 2013). Due to its high expression in the kidney of both species, it was proposed that the SVCT3 transporter might play a role in the reabsorption of substrates that would otherwise be excreted, such as nucleobases, although no likely substrate was found so far (Burzle *et al.* 2013). Nevertheless, Bürzle *et al.* demonstrated that SVCT3 is not involved in the transport of VC (Burzle *et al.* 2013). This is not surprising since previous phylogenetic analyses indicated that the SVCT3 gene appears to have diverged from SVCT1 and SVCT2 early in evolution, possibly losing the VC transport capacity and specializing in the absorption of other substrates (Burzle *et al.* 2013, Kourkoulou *et al.* 2018). In fact, this transporter has a conserved transmembrane amino acid motif (SS(FIV)(PAS)) similar to the one found in SVCT1 and SVCT2, strengthening the hypothesis of a possible phylogenetic ancestral relationship between them (Kourkoulou *et al.* 2018). The *SVCT4* gene was functionally described in *Rattus norvegicus* (rat) and found to be pseudogenized in humans (Yamamoto *et al.* 2010). The SVCT4 transporter is mainly expressed in the epithelial cells of the rat small intestine and is able to transport many nucleobases, such as xanthine, hypoxanthine, guanine, thymine and uracil (Yamamoto *et al.* 2010). However, similarly to the SVCT3 transporter, it does not seem to contribute to VC uptake (Yamamoto *et al.* 2010). Curiously, previous research hinted that *SVCT* homologs found in basal animal species share similarities with vertebrate *SVCT4* (Kourkoulou *et al.* 2018). These findings suggest that there may be no specific VC transporters in non-bilaterian and protostomian species. Nevertheless, this hypothesis is highly unlikely, as it would implicate that this nutrient is not required for several essential biological processes in a large subset of animal species. Although limited, considering the global taxonomic context, the observation that the

lepidopteran *B. mori* synthesizes VC, but still relies on extra uptake of this nutrient during larval stages 3L-E to 5L-E for adequate development (Hou *et al.* 2021), goes in accordance with this perspective. Therefore, *B. mori*, and inherently other protostomian species, likely have functional VC transporters. If so, then the identified non-vertebrate SVCTs might possess generalized nucleobase and VC transport capabilities, and the vertebrate SVCT4 possibly suffered partial loss of function throughout evolution.

From the evolutionary perspective, it is accepted that *SVCT1* and *SVCT2* were probably the result of a gene duplication event that preceded the divergence of bony fish (Osteichthyes) and tetrapods (Eck *et al.* 2004, Kourkoulou *et al.* 2018, Savini *et al.* 2008), dated around 410 million years ago (Betancur *et al.* 2013). This hypothesis is supported by the similarities in open reading frames between both genes, the almost identical exon-intron borders positioning (Eck *et al.* 2004, Savini *et al.* 2008), and the presence of the previously mentioned conserved amino motif in the resulting encoded proteins (Kourkoulou *et al.* 2018). Nevertheless, the origins of *SVCT3* and *SVCT4* remain to be uncovered. It is possible, however, that the four genes share common ancestry. Gene duplications and genome rearrangement events are particularly important in the appearance of new genes and inherent phenotypes, playing a crucial role in the diversification of vertebrates (Cañestro *et al.* 2013, Tautz and Domazet-Lošo 2011). In this context, whole genome duplications (WGDs) are known to usually result in an improved adaptation capacity to new environmental conditions (Kasahara 2013, Moriyama and Koshiba-Takeuchi 2018). The detection of four sets of paralogous regions that constitute about 25% of the human genome (Dehal and Boore 2005, Kasahara 2013) and the observed quadruple conserved synteny pattern of the human genome relative to the basal deuterostome *Branchiostoma floridae* (amphioxus) (Kasahara 2013, Putnam *et al.* 2008) suggest that vertebrates likely underwent two rounds of WGD. Currently, it is considered that the first round of WGD affected the common ancestor of all vertebrates, while the second took place in the common ancestor of jawed vertebrates, after the separation from jawless vertebrates (as lampreys and hagfish) (Dehal and Boore 2005, Kasahara 2013, Yu *et al.* 2024). Such events could potentially lead to the emergence of four distinct *SVCT*s from a single ancestral *SVCT* lineage during vertebrate speciation, a scenario that is consistent with the existence of *SVCT1*, *SVCT2*, *SVCT3* and *SVCT4* in many vertebrate lineages (Kourkoulou *et al.* 2018). In this context, protostomian and non-bilaterian *SVCT* homologs can likely encode for ancestral transporters with generalized substrate affinity, and the four vertebrate *SVCT*s might have emerged following selective pressures that led to gene subfunctionalization after the two WGD events. This evolutionary setting is compatible with a proposed model where VC transporters evolved from the continuous adaptation of nucleobase-specific

Nucleobase-Ascorbate Transporters (NATs), given the very low affinity of the *Aspergillus nidulans* uric acid-xanthine transporter for VC (Kourkoulou *et al.* 2018). Thus, if verified, such hypothesis might provide valuable insights on VC homeostasis in basal animal species and ultimately translate to the development of novel experimental setups in studies related to VC-mediated biological processes using, for example, model protostomian organisms such as *D. melanogaster* or *C. elegans*.

1.2.2.4. References

Betancur, R. R., R. E. Broughton, E. O. Wiley, K. Carpenter, J. A. Lopez, C. Li, N. I. Holcroft, D. Arcila, M. Sanciangco, J. C. Cureton Ii, F. Zhang, T. Buser, M. A. Campbell, J. A. Ballesteros, A. Roa-Varon, S. Willis, W. C. Borden, T. Rowley, P. C. Reneau, D. J. Hough, G. Lu, T. Grande, G. Arratia and G. Orti. **2013.** The tree of life and a new classification of bony fishes. *PLoS Currents*. 5:ecurrents.tol.53ba26640df26640ccaee26675bb26165c26648c26288.

Burzle, M., Y. Suzuki, D. Ackermann, H. Miyazaki, N. Maeda, B. Clemencon, R. Burrier and M. A. Hediger. **2013.** The sodium-dependent ascorbic acid transporter family SLC23. *Molecular Aspects of Medicine*. 34:436-454.

Cañestro, C., R. Albalat, M. Irimia and J. Garcia-Fernàndez. **2013.** Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. *Seminars in Cell & Developmental Biology*. 24:83-94.

Corpe, C. P., H. Tu, P. Eck, J. Wang, R. Faulhaber-Walter, J. Schnermann, S. Margolis, S. Padayatty, H. Sun, Y. Wang, R. L. Nussbaum, M. G. Espey and M. Levine. **2010.** Vitamin C transporter Slc23a1 links renal reabsorption, vitamin C tissue accumulation, and perinatal survival in mice. *The Journal of Clinical Investigation*. 120:1069-1083.

Daruwala, R., J. Song, W. S. Koh, S. C. Rumsey and M. Levine. **1999.** Cloning and functional characterization of the human sodium-dependent vitamin C transporters hSVCT1 and hSVCT2. *FEBS Letters*. 460:480-484.

de Koning, H. and G. Diallinas. **2000.** Nucleobase transporters. *Molecular Membrane Biology*. 17:75-94.

Dehal, P. and J. L. Boore. **2005.** Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. *PLOS Biology*. 3:e314.

Duque, P., C. P. Vieira, B. Bastos and J. Vieira. **2022a.** The evolution of vitamin C biosynthesis and transport in animals. *BMC Ecology and Evolution*. 22:84.

Duque, P., C. P. Vieira and J. Vieira. **2022b.** Advances in Novel Animal Vitamin C Biosynthesis Pathways and the Role of Prokaryote-Based Inferences to Understand Their Origin. *Genes*. 13:1917.

Eck, P., H. C. Erichsen, J. G. Taylor, M. Yeager, A. L. Hughes, M. Levine and S. Chanock. **2004.** Comparison of the genomic structure and variation in the two human sodium-dependent vitamin C transporters, SLC23A1 and SLC23A2. *Human Genetics*. 115:285-294.

Godoy, A., V. Ormazabal, G. Moraga-Cid, F. A. Zuniga, P. Sotomayor, V. Barra, O. Vasquez, V. Montecinos, L. Mardones, C. Guzman, M. Villagran, L. G. Aguayo, S. A. Onate, A. M. Reyes, J. G. Carcamo, C. I. Rivas and J. C. Vera. **2007.** Mechanistic insights and functional determinants of the transport cycle of the ascorbic acid transporter SVCT2. Activation by sodium and absolute dependence on bivalent cations. *Journal of Biological Chemistry*. 282:615-624.

Harrison, F. E., S. M. Dawes, M. E. Meredith, V. R. Babaev, L. Li and J. M. May. **2010.** Low vitamin C and increased oxidative stress and cell death in mice that lack the sodium-dependent vitamin C transporter SVCT2. *Free Radical Biology and Medicine*. 49:821-829.

Hou, M., T. M. Dai, X. Y. Liang, S. X. Zhang, W. Z. Cui, J. F. Qiu, Y. H. Sima, W. Z. Cui and S. Q. Xu. **2021.** *Bombyx mori* can synthesize ascorbic acid through the l-gulose pathway to varying degrees depending on developmental stage. *Archives of Insect Biochemistry and Physiology*. 106:e21783.

Kasahara, M. **2013.** Impact of whole-genome duplication on vertebrate development and evolution. *Semin Cell Dev Biol*. 24:81-82.

Kourkoulou, A., A. A. Pittis and G. Diallinas. **2018.** Evolution of substrate specificity in the Nucleobase-Ascorbate Transporter (NAT) protein family. *Microbial Cell*. 5:280-292.

López-Fernández, H., P. Duque, S. Henriques, N. Vázquez, F. Fdez-Riverola, C. P. Vieira, M. Reboiro-Jato and J. Vieira. **2019.** A Bioinformatics Protocol for Quickly Creating Large-Scale Phylogenetic Trees. In: F. Fdez-Riverola, M. S. Mohamad, M. Rocha, J. F. De Paz and P. González, editors. *Practical Applications of Computational Biology and Bioinformatics, 12th International Conference.* Cham. Springer. p. 88-96.

May, J. M. **2011.** The SLC23 family of ascorbate transporters: ensuring that you get and keep your daily dose of vitamin C. *British Journal of Pharmacology*. 164:1793-1801.

Moriyama, Y. and K. Koshiba-Takeuchi. **2018.** Significance of whole-genome duplications on the emergence of evolutionary novelties. *Briefings in Functional Genomics*. 17:329–338.

Patananan, A. N., L. M. Budenholzer, M. E. Pedraza, E. R. Torres, L. N. Adler and S. G. Clarke. **2015.** The invertebrate *Caenorhabditis elegans* biosynthesizes ascorbate. *Archives of Biochemistry and Biophysics*. 569:32-44.

Putnam, N. H., T. Butts, D. E. Ferrier, R. F. Furlong, U. Hellsten, T. Kawashima, M. Robinson-Rechavi, E. Shoguchi, A. Terry, J. K. Yu, E. L. Benito-Gutierrez, I. Dubchak, J. Garcia-Fernandez, J. J. Gibson-Brown, I. V. Grigoriev, A. C. Horton, P. J. de Jong, J. Jurka, V. V. Kapitonov, Y. Kohara, Y. Kuroki, E. Lindquist, S. Lucas, K. Osoegawa, L. A.

Pennacchio, A. A. Salamov, Y. Satou, T. Sauka-Spengler, J. Schmutz, I. T. Shin, A. Toyoda, M. Bronner-Fraser, A. Fujiyama, L. Z. Holland, P. W. Holland, N. Satoh and D. S. Rokhsar. **2008.** The amphioxus genome and the evolution of the chordate karyotype. *Nature*. 453:1064-1071.

Rivas, C. I., F. A. Zúñiga, A. Salas-Burgos, L. Mardones, V. Ormazabal and J. C. Vera. **2008.** Vitamin C transporters. *Journal of Physiology and Biochemistry*. 64:357-375.

Rumsey, S. C., R. Daruwala, H. Al-Hasani, M. J. Zarnowski, I. A. Simpson and M. Levine. **2000.** Dehydroascorbic Acid Transport by GLUT4 in *Xenopus* Oocytes and Isolated Rat Adipocytes. *Journal of Biological Chemistry*. 275:28246-28253.

Rumsey, S. C., O. Kwon, G. W. Xu, C. F. Burant, I. Simpson and M. Levine. **1997.** Glucose Transporter Isoforms GLUT1 and GLUT3 Transport Dehydroascorbic Acid. *Journal of Biological Chemistry*. 272:18982-18989.

Savini, I., A. Rossi, C. Pierro, L. Avigliano and M. V. Catani. **2008.** SVCT1 and SVCT2: key proteins for vitamin C uptake. *Amino Acids*. 34:347-355.

Sotiriou, S., S. Gispert, J. Cheng, Y. Wang, A. Chen, S. Hoogstraten-Miller, G. F. Miller, O. Kwon, M. Levine, S. H. Guttentag and R. L. Nussbaum. **2002.** Ascorbic-acid transporter Slc23a1 is essential for vitamin C transport into the brain and for perinatal survival. *Nature Medicine*. 8:514-517.

Tautz, D. and T. Domazet-Lošo. **2011.** The evolutionary origin of orphan genes. *Nature Reviews Genetics*. 12:692.

Tsukaguchi, H., T. Tokui, B. Mackenzie, U. V. Berger, X. Z. Chen, Y. Wang, R. F. Brubaker and M. A. Hediger. **1999.** A family of mammalian Na+-dependent L-ascorbic acid transporters. *Nature*. 399:70-75.

Wang, M., J. He, S. Li, Q. Cai, K. Zhang and J. She. **2023.** Structural basis of vitamin C recognition and transport by mammalian SVCT1 transporter. *Nature Communications*. 14:1361.

Wang, Y., B. Mackenzie, H. Tsukaguchi, S. Weremowicz, C. C. Morton and M. A. Hediger. **2000.** Human vitamin C (L-ascorbic acid) transporter SVCT1. *Biochemical and Biophysical Research Communications*. 267:488-494.

Welch, R. W., Y. Wang, A. Crossman, Jr., J. B. Park, K. L. Kirk and M. Levine. **1995.** Accumulation of vitamin C (ascorbate) and its oxidized metabolite dehydroascorbic acid occurs by separate mechanisms. *Journal of Biological Chemistry*. 270:12584-12592.

Wheeler, G., T. Ishikawa, V. Pornsaksit and N. Smirnoff. **2015.** Evolution of alternative biosynthetic pathways for vitamin C following plastid acquisition in photosynthetic eukaryotes. *eLife*. 4:e06369.

Yabuta, Y., R. Nagata, Y. Aoki, A. Kariya, K. Wada, A. Yanagimoto, H. Hara, T. Bito, N. Okamoto, S. Yoshida, A. Ishihara and F. Watanabe. **2020.** L-Ascorbate Biosynthesis

Involves Carbon Skeleton Rearrangement in the Nematode *Caenorhabditis elegans*. *Metabolites*. 10:334.
Yamamoto, S., K. Inoue, T. Murata, S. Kamigaso, T. Yasujima, J. Y. Maeda, Y. Yoshida, K. Y. Ohta and H. Yuasa. **2010.** Identification and functional characterization of the first nucleobase transporter in mammals: implication in the species difference in the intestinal absorption mechanism of nucleobases and their analogs between higher primates and other mammals. *Journal of Biological Chemistry*. 285:6522-6531.
Yu, D., Y. Ren, M. Uesaka, A. J. S. Beavan, M. Muffato, J. Shen, Y. Li, I. Sato, W. Wan, J. W. Clark, J. N. Keating, E. M. Carlisle, R. P. Dearden, S. Giles, E. Randle, R. S. Sansom, R. Feuda, J. F. Fleming, F. Sugahara, C. Cummins, M. Patricio, W. Akanni, S. D'Aniello, C. Bertolucci, N. Irie, C. Alev, G. Sheng, A. de Mendoza, I. Maeso, M. Irimia, B. Fromm, K. J. Peterson, S. Das, M. Hirano, J. P. Rast, M. D. Cooper, J. Paps, D. Pisani, S. Kuratani, F. J. Martin, W. Wang, P. C. J. Donoghue, Y. E. Zhang and J. Pascual-Anaya. **2024.** Hagfish genome elucidates vertebrate whole-genome duplication events and their evolutionary consequences. *Nature Ecology & Evolution*. 8:519-535.

1.3. Thesis objectives and outlines

The presented background indicates that further model organism characterization, in particular *D. melanogaster* and *C. elegans*, might provide experimental advantages for the understanding of transversal VC-related cellular processes in animals. Moreover, preliminary efforts to verify if the emergence of eukaryote VC synthesis mechanisms could be derived from prokaryote origins have surprisingly revealed that putative AOs are present in bacteria and archaea species. Given that the current industrial production of VC is mainly reliant on a costly two-step fermentation process, the identification of prokaryote species capable of endogenous VC synthesis without genomic manipulation would thus represent an alternative and promising opportunity for the development of efficient one-step fermentation methodologies. Therefore, efforts regarding the identification of prokaryote AOs could lead to potential cost improvements of industrial VC production.

Following this rationale, the work performed in this thesis focused on two main objectives: i) the characterization of the *D. melanogaster* VC synthesis pathway and ii) the identification of possible VC transporters present in basal animal species. Additionally, a third objective regarding the identification of AOs in bacteria species was also pursued. The findings regarding each topic are shown in the following chapters.

Chapter 2 concerns the clarification of the evolutionary history of *Regucalcin*-like and *SVCT*-like genes in the context of previous *GULO* literature, as well as the inference of

the functional role of *Regucalcin* and *Drosophila cold acclimation* (*Dca*) in *D. melanogaster*. The results revealed that *Regucalcin* homologs are likely involved in VC synthesis, calcium homeostasis and the oxidative stress response in all animals, but might have acquired new roles during evolution in some taxonomic groups such as insects. Furthermore, the presence of *SVCT*s in basal animal lineages was generally correlated with intrinsic VC requirements, suggesting that such transporters are required more notably in VC auxotrophs. Moreover, it was uncovered that *Regucalcin* presence might not be indicative of a partially conserved VC synthesis pathway in species where *GULO* is absent but VC synthesis occurs, such as *D. melanogaster*. This latter observation led to the consideration of possible novel VC synthesis mechanisms in this organism, motivating the methodological approach presented in Chapter 3.

Chapter 3 discusses the bioinformatics and molecular biology approaches used to identify genes possibly related with VC synthesis in *D. melanogaster*. In the global context, the phylogenetic analyses highlight that the synthesis of this nutrient in *D. melanogaster* might derive from either D-Glucose-6P or myo-inositol. The use of chromosomal deletion stocks in VC level screenings unexpectedly suggested that 33 genomic regions could be related with VC synthesis in *D. melanogaster*. These regions comprised an unfeasible amount of 1480 genes to functionally address within the time period of the PhD project. Nevertheless, the colorimetric data highlighted that two concurrent VC synthesis pathways reliant in D-Glucose-6P and myo-inositol, respectively, might be present in *D. melanogaster*. The suppression of gene homologs related to these pathways using ribonucleic acid interference (RNAi) hinted so far to a possibly synergistic relation between both synthesis mechanisms, with the conversion of myo-inositol to D-Glucuronate catalyzed by myo-inositol oxygenase (Miox, or CG6910) contributing to 14% to 25% of total VC content in *D. melanogaster* under several experimental conditions.

Chapter 4 displays the use of the auto-phylo pipeline to perform the phylogenetic analysis required in order to identify putative bacterial AOs. The results demonstrate that AOs with homology to GULO are present in several bacteria lineages, having a particularly high distribution in Actinomycetota species. Despite lack of functional data to support these inferences, this observation raises doubts regarding the common assumption of generalized prokaryote VC auxotrophy, and provides a new perspective on the possible distribution of VC synthesis mechanisms in the different life kingdoms.

Lastly, a brief thesis summary, a generalized conclusion regarding the presented findings, and future work perspectives, are provided in Chapter 5.

Chapter 2. The conservation of VC synthesis and transport mechanisms in animals: from the global perspective to the case of *D. melanogaster*

This chapter includes the following published research article:

Duque, P., Vieira, C. P., Bastos, B., & Vieira, J. (2022). The evolution of vitamin C biosynthesis and transport in animals. *BMC Ecology and Evolution*, *22*(1), 84. https://doi.org/10.1186/s12862-022-02040-7

This article is here reproduced under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Duque *et al. BMC Ecology and Evolution* (2022) 22:84
https://doi.org/10.1186/s12862-022-02040-7

BMC Ecology and Evolution

RESEARCH

Open Access

The evolution of vitamin C biosynthesis and transport in animals

Pedro Duque[1,2,3,4], Cristina P. Vieira[3,4], Bárbara Bastos[3,4] and Jorge Vieira[3,4*]

Abstract

Background: Vitamin C (VC) is an indispensable antioxidant and co-factor for optimal function and development of eukaryotic cells. In animals, VC can be synthesized by the organism, acquired through the diet, or both. In the single VC synthesis pathway described in animals, the penultimate step is catalysed by Regucalcin, and the last step by L-gulonolactone oxidase (GULO). The *GULO* gene has been implicated in VC synthesis only, while *Regucalcin* has been shown to have multiple functions in mammals.

Results: Both *GULO* and *Regucalcin* can be found in non-bilaterian, protostome and deuterostome species. *Regucalcin*, as here shown, is involved in multiple functions such as VC synthesis, calcium homeostasis, and the oxidative stress response in both Deuterostomes and Protostomes, and in insects in receptor-mediated uptake of hexamerin storage proteins from haemolymph. In Insecta and Nematoda, however, there is no *GULO* gene, and in the latter no *Regucalcin* gene, but species from these lineages are still able to synthesize VC, implying at least one novel synthesis pathway. In vertebrates, *SVCT1*, a gene that belongs to a family with up to five members, as here shown, is the only gene involved in the uptake of VC in the gut. This specificity is likely the result of a subfunctionalization event that happened at the base of the Craniata subphylum. *SVCT*-like genes present in non-Vertebrate animals are likely involved in both VC and nucleobase transport. It is also shown that in lineages where *GULO* has been lost, *SVCT1* is now an essential gene, while in lineages where *SVCT1* gene has been lost, *GULO* is now an essential gene.

Conclusions: The simultaneous study, for the first time, of *GULO*, *Regucalcin* and *SVCT*s evolution provides a clear picture of VC synthesis/acquisition and reveals very different selective pressures in different animal taxonomic groups.

Keywords: Ascorbic acid, GULO, Regucalcin, SVCT, Animals

Background

In eukaryotes, Vitamin C (VC), also known as L-ascorbic acid or ascorbate, is an essential micronutrient for normal cell function, growth and development, acting as an antioxidant capable of detoxifying exogenous radical species or those generated during mitochondrial metabolism [1, 2]. VC is required for normal neuromodulation [3], protection against lipid peroxidation [4], and collagen biosynthesis, where it acts as a cofactor for collagen stabilization enzymes, namely prolyl and lysyl hydroxylases [5]. Elevated VC levels have been associated with protection against degenerative diseases and cancer, with several reports suggesting beneficial effects in cancer treatment and chemotherapy recovery [6]. Nevertheless, at high concentrations, and in the presence of catalytic metal ions such as ferric iron (Fe^{3+}), VC can also have a pro-oxidative behavior through an indirect contribution in the Fenton reaction [6]. The radical species that are generated can act as triggers for lipid peroxidation [7], a self-proliferating chain reaction that ultimately leads to damage in membranes, proteins and DNA [8]. In humans, lack of VC leads to scurvy [9].

*Correspondence: jbvieira@ibmc.up.pt

[3] Instituto de Investigação e Inovação em Saúde (I3S), Universidade do Porto, Rua Alfredo Allen, 208, 4200-135 Porto, Portugal
Full list of author information is available at the end of the article

© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Being an essential vitamin, it is not surprising that VC is endogenously synthetized by a wide variety of species, ranging from commensal and parasitic bacteria [10–12] to highly complex eukaryotes, such as plants and animals [13–15]. Three main VC biosynthesis pathways have been described (Fig. 1), namely the photosynthetic protists, plant/green algae, and animal pathways [14]. In animals, D-glucose is used as the initial precursor, and after several enzymatic reactions is converted to L-gulonolactone by Regucalcin (in deuterostomians it is also known as Senescence marker protein 30 (SMP30)). L-gulonolactone then participates in a final oxidation reaction catalyzed by L-gulonolactone oxidase (GULO), leading to the formation of 2-keto-L-gulonolactone, which spontaneously enolizes to L-ascorbic acid [16, 17]. Not all animals are, however, capable of synthesizing VC. Within deuterostomes, humans (*Homo sapiens*), non-human primates, the guinea pig (*Cavia porcellus*), numerous bats, various birds, and the teleost fish have lost this ability, due to the complete or partial loss of the *GULO* gene [13]. Although *GULO* is also present in non-bilateria species, it is clear that this gene was lost in several protostomian lineages, such as insects and nematodes [15]. Since only one VC synthesis pathway has been described in animals, species belonging to these lineages were thought not to be able to synthesize VC. Nevertheless, at present, there are two insect species, namely *Drosophila melanogaster* (diptera), and *Bombyx mori* (lepidoptera), as well as one nematode (*Caenorhabditis elegans*), that have been shown to synthesize VC [18–20]. It should be noted that distantly related insect groups, such as the Diptera and the Lepidoptera, may have independently acquired de novo the ability to synthesize VC. Indeed, there is evidence that the ability to synthesize VC without the intervention of *GULO* may have arisen independently multiple times in protostomian evolution, since this gene is absent in nematodes and insects, but present in Arachnida species that are closer to insects [19, 21].

In *B. mori*, it has been suggested that the VC synthesis pathway could be identical to that of mammals and other vertebrates with the exception of the enzyme involved in the last step of the pathway [20]. Indeed, GULO-like activity has been detected in *B. mori*, although at reduced levels compared to other species [20]. Surprisingly, the protein identified as having GULO-like activity is a sterol reductase, in contrast to GULO that is an aldonolactone

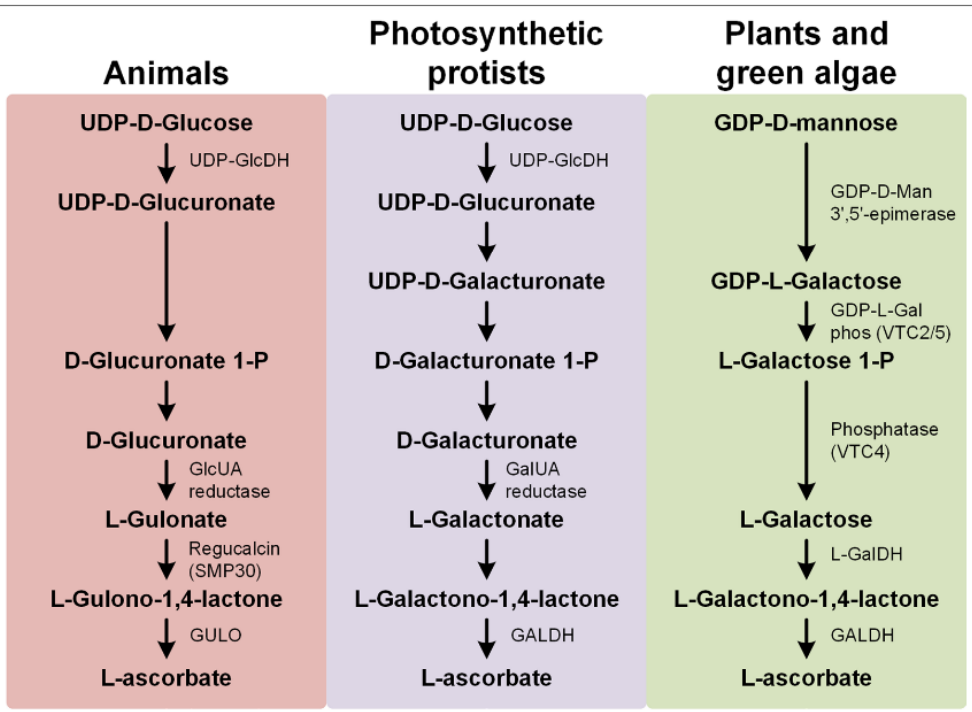


Fig. 1 Graphic display of the currently known VC synthesis pathways. The final oxidation step of the distinct aldono-1,4-lactone to ascorbate is performed by a FAD-linked oxidase or dehydrogenase (GULO or GALDH). Photosynthetic protists appear to possess enzymatic components from animal and plant pathways, and due to this characteristic, the current described pathway for these species likely evolved from a secondary endosymbiosis event regarding a non-photosynthetic ancestor and algae [15]. The figure here presented and the corresponding description were adapted from the literature [14]

oxireductase [20, 22]. Since aldonolactone oxireductases and sterol reductases are distantly related flavoproteins [22], the ability of the *B. mori* sterol reductase to catalyze the oxidation of L-gulonolactone was unexpected.

In deuterostomians, apart from the substrate conversion role in the VC synthesis pathway, Regucalcin is also implicated in several physiological mechanisms, such as intracellular Ca^{2+} homeostasis, proteolysis, and signaling and oxidative stress regulation [23]. Despite the many functional roles played by this protein in deuterostomians, homozygous *Regucalcin* knockout mice are viable [24, 25]. It should be noted that there are no *Regucalcin* paralogs in *Mus musculus* that could compensate *Regucalcin*'s functions in homozygous knockout mice. Nevertheless, although viable, knockout mice still show increased cell senescence in several tissues and shorter life spans [26, 27].

In protostomians, the role of *Regucalcin* is much less clear, but in the dipteran *Sarcophaga peregrina*, the Anterior fat body protein encoded by the *AFP* gene (an homolog of *Regucalcin*) is almost exclusively found in the anterior fat body, and does not exhibit strong affinity towards calcium [28]. AFP may participate in the receptor-mediated uptake of hexamerin storage proteins from insect haemolymph by fat body cells, a mechanism that is only found in the Insecta. In another dipteran, *Calliphora vicina*, there is an *AFP* orthologue that encodes a protein that interacts with an hexamerin receptor [29], and in *D. melanogaster* Regucalcin has been found in the haemolymph [30], findings that strengthen this putative function. Therefore, it has been assumed that *Regucalcin* does not have a role in calcium homeostasis in *D. melanogaster* [31]. Nevertheless, in this species, there is a *Regucalcin* paralog, named *Drosophila cold acclimation* (*Dca*), that is restricted to the subgenus Sophophora [31, 32]. *Dca* shows high sequence identity with *Regucalcin* [31], and thus, could have similar functions. Nevertheless, Dca is currently implicated in the response to cold exposure and wing size variation in flies [25, 31–33], functions that are not easily reconciled with the role of a protein involved in the uptake of storage proteins by fat body cells. In the lineage leading to *Dca*, the nonsynonymous fixation rate is, however, increased, suggesting that *Dca* could have acquired a new function (neofunctionalization) after the duplication event [31]. The expression pattern of these genes is indeed distinct, since *Regucalcin* has a much higher expression level than *Dca* under normal conditions (information available at https://flybase.org/). Moreover, *Dca* is primarily expressed in the adult digestive system tissues and *Regucalcin* in the larval fat body and pupa [31, 34, 35]. There is no data on whether *D. melanogaster Regucalcin* is an essential gene, but *Dca* knockout flies have been shown to be viable [25].

In vertebrates, VC is primarily synthesized in either the liver or the kidney depending on the species being considered [13], but once synthesized, VC must be transported to the cells where it is needed. This process is mainly reliant on Sodium-dependent Vitamin C Transporters (SVCTs) and Sodium-independent facilitative glucose transporters (GLUTs) [6, 36, 37]. The contribution of GLUTs to the maintenance of favorable VC concentrations in cells is, however, small when compared to SVCTs, which are considered the main regulator of VC uptake [6, 38–40]. In vertebrates, SVCTs are a family of surface glycoproteins with four members (SVCT1 to 4), but only SVCT1 and SVCT2 proteins have been implicated in VC homeostasis [36]. These two proteins share a unique amino acid motif in transmembrane domain 10 (SSSP) that may be responsible for the substrate specificity [41]. SVCT1 is essentially expressed in the epithelial membranes of various organs, such as the intestine, kidney and liver, while SVCT2 has a generalized presence throughout the body [36, 42]. Therefore, SVCT1 protein has been implicated in specific VC uptake, whereas the SVCT2 transporter has been implicated in localized cell responses to oxidative stress [36, 40, 43]. No function has yet been attributed to *SVCT3*, but it does not seem to have VC or nucleobase affinity [36]. *SVCT4* is known to transport various nucleobases but not VC [44].

SVCT1 and *SVCT2* genes are probably the result of a duplication event that preceded the divergence of bony fish (Osteichthyes) and tetrapods [37, 41]. The evolutionary history of the four members of the SVCT family is, however, not yet clear. Some authors consider *SVCT3* and *SVCT4* orphan genes [37, 42], while others indicate that *SVCT3* may have diverged early in evolution from *SVCT1* and *SVCT2* [41]. The evolutionary history of SVCTs in non-bilaterian and protostomians is also largely unknown. It has been, however, hypothesized that *SVCT4* is most similar to the *SVCT*s observed in these basal taxonomic groups [41]. Since the vertebrate *SVCT4* is not a VC transporter, if true, it could be that there are no VC transporters in non-bilaterian and protostomian, species, or that the vertebrate *SVCT4* lost its ancient VC transporter activity. The former hypothesis is highly unlikely. For instance, although the lepidopteran *B. mori* is able to synthesize VC, it does not synthesize enough of it during larval stages 3-LE up to 5-LE for its needs, and thus must rely on an external source of this nutrient [20]. Therefore, *B. mori* must have a functional VC transporter.

Using a combination of phylogenetic and protein structure analyses, as well as inferences on positively selected amino acid sites, the evolutionary history of *Regucalcin*-like and *SVCT*-like genes in animals is here clarified. These results are interpreted in the light of the recently

published *GULO* phylogenetic and functional analyses [18–21], revealing a history of pseudogenization, neofunctionalization and sub-functionalization. The joint evolutionary history of *Regucalcin*, *GULO* and *SVCT*s, also explains why some animal lineages lost the ability to synthesize VC while other basal lineages did not.

Results

Contrasting patterns of *Regucalcin* gene duplication and loss in different animal lineages

Regucalcin presence/loss was addressed using restrictive criteria. Indeed, within a given taxonomic group, a minimum of three species showing no evidence for the presence of a gene was required to infer gene loss in that lineage, since most genomes do not have a 100% coverage.

Regucalcin can be found in non-bilaterian Porifera and Anthozoa lineages, some of the most basal lineages of the animal kingdom (Fig. 2 and Additional file 1: Fig. S1). Within the Protostomian insect species, *Regucalcin* appears to have been ancestrally duplicated within the Hemipteroid Assemblage, however the relationship between the identified *H1*, *H2* and *H3* genes could not be resolved (Fig. 2 and Additional file 1: Fig. S2). Gene *H3* is most closely related to the *D. melanogaster Regucalcin* gene chosen as a representative sequence of the remaining insects. The gene *H3* lineage depicts a high prevalence of duplications spread across many insect taxonomic groups, and some duplicates are associated with considerable gene loss events (Additional file 1: Figs. S2 to S7). Species from the dipteran Sophophora subgenus have *Regucalcin* and a known paralog of this gene, *Dca* [31, 32]. The phylogenetic approach here implemented replicates these findings and also depicts the expected *Dca* duplicates in *Drosophila ananassae* [31]. Unexpectedly, the results also show additional and yet undescribed *Dca* duplications in *Drosophila bipectinata*. On a particular note, the presence of two *Regucalcin*-like copies in *B. mori* (XP_012549788.1 and XP_004930722.1) fits the hypothesis of a novel VC synthesis pathway, identical to that of vertebrates, with the exception of the enzyme involved in the last step of the pathway [20]. Nevertheless, given the multiple functions associated with *Regucalcin*, many of them unrelated to synthesis of VC (see Background), one must be cautious regarding this interpretation.

The frequent observation of *Regucalcin* gene duplications and losses in Insecta (Additional file 1: Fig. S7) raises the issue of whether *Regucalcin* homologs are mostly non-essential genes, involved in local species adaptation. This could be the case for *D. melanogaster Dca*, since this gene has been involved in the response to cold exposure and wing size variation [25, 31–33]. Moreover, the *D. melanogaster Dca* gene has been reported as a non-essential gene [25]. Nevertheless, the *daughterless*-GAL4 (*da*-GAL4) driver used in that study to supress the expression of *D. melanogaster Dca* [25] might not have a high constitutive expression [45]. Therefore, in order to elucidate whether in insects *Regucalcin* homologs are in general non-essential genes, *Regucalcin* and *Dca* RNAi knockouts were performed using an *Actin 5 C*-GAL4 driver with ubiquitous and high constitutive expression [46]. The viability of the progeny that expresses either *Regucalcin* or *Dca* RNAi was greatly diminished (Additional file 1: Table S1), with at least 71% of the individuals not transitioning from the pupal stage. This reduced viability was particularly accentuated in the male progeny when compared to the female. Therefore, under our experimental conditions, both *Regucalcin* and *Dca* seem to be essential genes. The discrepancy between our results for *Dca* and those reported in [25], is likely explained by the distinct driver or/and the different *Dca* RNAi strains used. Given these observations, the dynamic evolution of *Regucalcin* homologs in insects is surprising.

Contrary to the recurrent gene duplications observed within insects, the *Regucalcin* gene does not seem to have been duplicated ancestrally in the remaining Non-Lophotrochozoa although it can be seen specifically duplicated in Collembola, Araneae, Platyhelminthes and Acari species (Fig. 2 and Additional file 1: Fig. S8). The presence of *Regucalcin* in the Platyhelminthes is restrained to one out of nine species analysed (*Macrostomum lignano*). This evidence suggests an event of gene loss soon after the appearance of this taxonomic group. Other *Regucalcin* loss events likely affected the Crustacea and Nematoda. Since nematodes are able to synthesize VC [18], this observation implies that VC synthesis in *C. elegans* does not occur through GULO replacement in the known animal VC biosynthetic pathway, but rather by a novel pathway that does not involve either *Regucalcin* and *GULO*. The loss of *Regucalcin* in Nematoda, further suggests that this gene does not perform other essential functions unrelated to VC synthesis, or alternatively, that this gene might have been functionally replaced by others in this group of species. Although the results suggest up to four ancestral gene duplications within the Lophotrochozoa groups (Fig. 2 and Additional file 1: Fig. S9), and subsequent multiple independent gene losses within groups, such a complex scenario is supported by a single species representative of each group. Therefore, at present, we propose a scenario similar to that for Non-Lophotrochozoa species, in which the variation in the number of gene copies is due to group specific gene duplications. These show a particularly high prevalence within Bivalvia (Additional file 1: Fig. S9).

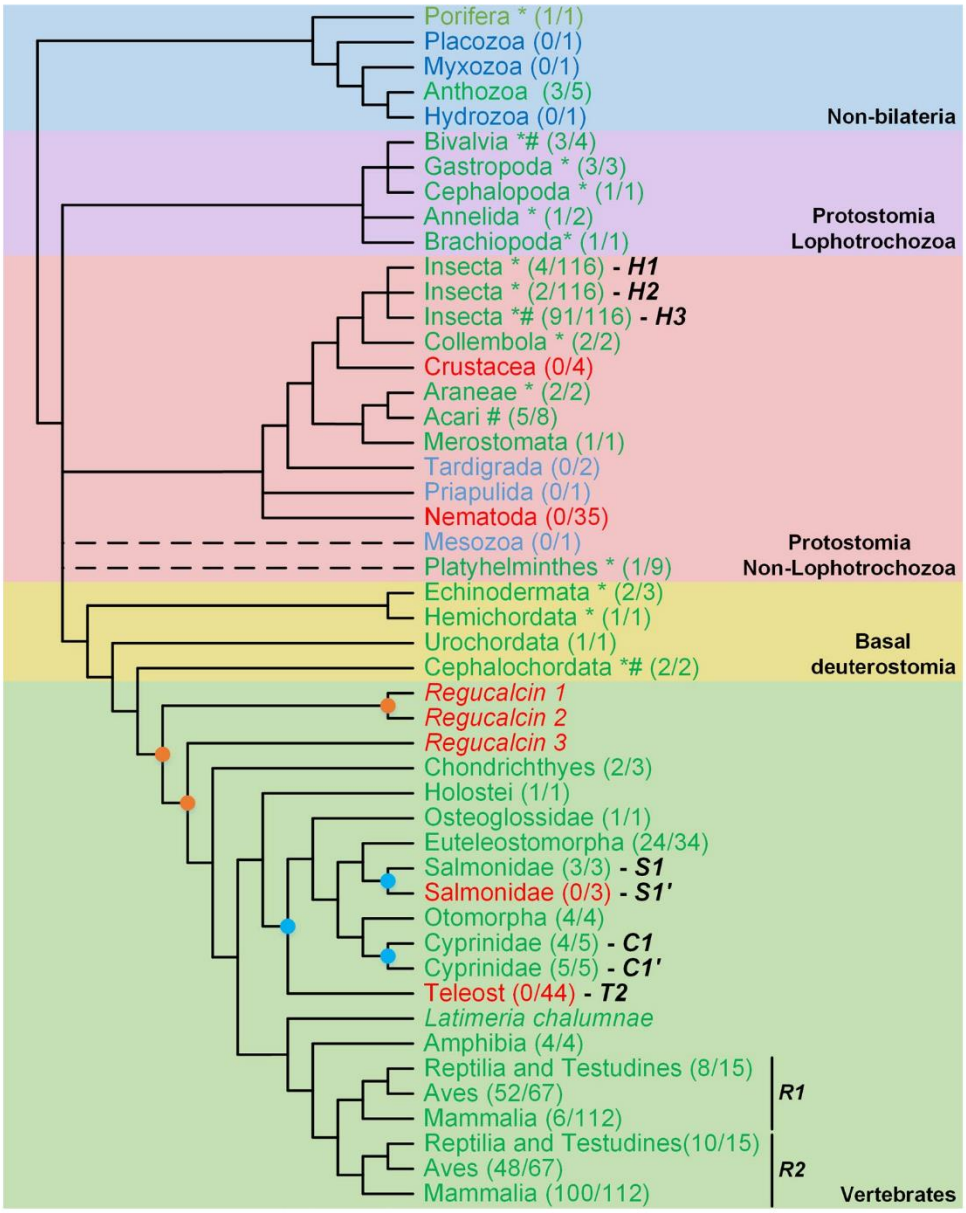


Fig. 2 *Regucalcin* evolution across the animal kingdom. *Regucalcin* presence is highlighted in green, absence in red and uncertainty regarding gene loss/presence in light blue. Gene duplications affecting a single species of a genus are marked with a *, while those which affected two or more species from the same genus are marked with a #. The number of species present in the final dataset and total number of species analysed for each taxonomic group, respectively, are shown within brackets. The identified gene duplicates are marked with letters H, S, C, T and R. Taxonomic groups are represented in different colours and labelled next to the cladogram. The known vertebrate Whole Genome Duplication (WGD) events are represented by orange circles, and the Teleost, Salmonidae and Cyprinidae specific WGDs are represented by blue circles, on top of the branches. The dashed branches represent taxonomic groups with uncertain position in the cladogram. The cladogram topology is depicted as in the Tree of Life web project [87]

Within deuterostomians, the *Regucalcin* evolutionary history appears to be less complex than the general scenario observed for the non-bilaterian and protostomian organisms. *Regucalcin* is present in all of the major basal deuterostomian groups, with specific duplications in Echinodermata, Hemichordata and Cephalochordata

species (Fig. 2 and Additional file 1: Fig. S10). As observed for protostomians, the overall tendency for duplications observed within basal deuterostomians raises once more the issue of whether homologous genes perform similar functions or not. A two-round whole genome duplication (2R-WGD) event has likely occurred within the Craniata subphylum, after the separation of vertebrates from invertebrate chordates [47–49]. One version of this hypothesis considers that the first round of WGD affected the common ancestor of all vertebrates, while the second affected the common ancestor of jawed vertebrates, after the separation from jawless vertebrates [48, 50]. Additionally, WGD were also extrapolated at the base of teleost fish, and after within the salmonids and some cyprinids [51]. This also seems to be the case in the allotetraploid *Xenopus laevis* [52]. *Regucalcin* is present in all the vertebrate taxonomic groups analysed (Fig. 2 and Additional file 1: Fig. S11). Nevertheless, the results suggest that three out of the four putative copies that resulted from the 2R-WGD events were lost, as were some of the duplicates expected from the proposed teleost-specific WGDs. The *Regucalcin* duplicates observed in the Reptilia and Testudines, Aves and Mammalia groups (genes *R1* and *R2*) could be the result of the 2R-WGD events. Nevertheless, since these duplicates are not present in the remaining vertebrate taxonomic groups, the inference of a duplication event in the common ancestor of the Reptilia and Testudines, Aves and Mammalia groups is the most parsimonious explanation for this scenario. Interestingly, only a small fraction of mammalian species still has gene *R1*, suggesting that this duplicate likely did not acquire relevant functions and was therefore lost independently in many lineages.

Regucalcin and Dca are involved in calcium homeostasis and oxidative stress response

As already stated, *Regucalcin* has been implicated in calcium homeostasis and oxidative stress, among others, in deuterostomians, but not in protostomians. Since the conservation of orthologous interactors is usually indicative of similar protein functional properties in distinct species [53], the comparison of the *D. melanogaster*, *M. musculus* and *H. sapiens* Regucalcin interactomes could provide clues regarding putative functions of this protein in the former species. Unexpectedly, the number of inferred *D. melanogaster* interactors (95) is substantially higher than the one obtained for *M. musculus* (three) and *H. sapiens* (nine). A significant fraction of the *D. melanogaster*, *M. musculus* and *H. sapiens* interactomes remains undiscovered [54–56], and thus this could explain the observed discrepancy. Nevertheless, it is also possible that Regucalcin may have more roles/interactors in *D. melanogaster* in comparison to the other species. Despite the small number of Regucalcin interactors reported for *M. musculus* and *H. sapiens*, one out of three *M. musculus* interactors (gene ID 22627) and four out of nine *H. sapiens* interactors (gene IDs 3094, 3336, 6647 and 9948) are also interactors in *D. melanogaster* (Additional file 1: Table S2). The shared *M. musculus* interactor (the 14-3-3 protein epsilon; gene ID 22627) is a calcium channel regulator (https://www.uniprot.org/uniprot/P62259). Moreover, one of the shared *H. sapiens* interactors (Histidine triad nucleotide-binding protein; gene ID 3094), is involved in the positive regulation of calcium-mediated signaling (https://www.uniprot.org/uniprot/P49773). These observations suggest a role of the *D. melanogaster* Regucalcin in calcium homeostasis, in contrast to previous claims of no calcium related functions [31]. Another shared *H. sapiens* interactor (Superoxide dismutase 1; gene ID 6647), which has been correlated with oxidative stress response/signaling [57] and is known to have increased activity in the presence of Regucalcin [58, 59] further suggests that the *D. melanogaster* Regucalcin may also be an apoptosis regulator [60]. The dysregulation of calcium homeostasis/signaling and apoptosis is thus likely the cause of the pupae lethality observed in *Regucalcin* RNAi experiments (see above). However, no conclusions can be drawn regarding VC synthesis with this data, since no interactor with relevant function in the known pathway was identified in either *M. musculus* or *H. sapiens*. For Dca there is no protein-protein interaction data. Nevertheless, the high similarity between Regucalcin and Dca suggests that they might share interactors. The protein docking inferences of Regucalcin and Dca versus the 14-3-3 protein epsilon, Histidine triad nucleotide-binding protein 1 and Superoxide dismutase 1 here made show that this is a possibility, although the predicted Dca and Regucalcin interacting regions are different (Additional file 1: Fig. S12). Indeed, in two out of three cases, Dca has more interacting amino acids than Regucalcin (39, 52 and 37 for Dca versus 41, 45 and 34 for Regucalcin, respectively), although the interacting surface of *D. melanogaster* Regucalcin is the same as that observed in humans for the same three interactors (Additional file 1: Fig. S13). A putative role of *Dca* in calcium homeostasis could explain the involvement of this gene in wing size variation, since organ development depends on coordinated cell-cell communication that in turn requires signal integration among multiple pathways, relying on second messengers such as calcium ions [61].

Inferences on PSS

The location of positively selected sites (PSS) on predicted protein structures can give insight into which protein regions are most often the target of positively selected amino acid changes. While the presence of PSS

around the lid/active site of Regucalcin homologs suggests a possible effect on the modulation of its gluconolactonase activity, clusters of PSS located elsewhere suggest a possible effect on the modulation of the interaction strength between Regucalcin and its interactors.

In total, 33 PSS are here inferred at the protein surface of Regucalcin homologs, in 12 animal lineages (Fig. 3), one of which being the Sophophora Dca lineage where evidence for positive selection has been previously uncovered [31]. Four PSS are found in the lid region of the inferred protein models (Sophophora Dca S^{132}, Formicoidea Hy1 Q^{138}, Cyprinidae C1.2 S^{118} and Aves R2 H^{130}, labeled in dark blue in Fig. 3) and two surrounding the active site (Sophophora Dca I^{60} and Reptilia and Testudines R2 S^{35}, labeled in brown in Fig. 3). These observations suggest that the modulation of Regucalcin´s gluconolactonase activity might be important for a wide variety of species, and that the gluconolactonase activity is important for *D. melanogaster´*s Dca function.

Twenty-seven PSS can be seen spread across the remaining regions of the protein models (labeled in purple in Fig. 3; Additional file 1: Table S3). Since they are away from the lid region and the active site, it seems likely that they define interaction surfaces with Regucalcin interactors. This is a strong possibility since in *D. melanogaster* Regucalcin has, so far, been reported to interact with 95 other proteins (see EvoPPI 1.0; *D. melanogaster* same species query; [62]). Nevertheless, only two *D. melanogaster* Regucalcin PSS (I^8 and A^9) are contained in the Regucalcin interacting region here inferred (Fig. 4 A and Additional file 1: Fig. S12). Surprisingly, the other PSS are clustered around the predicted interacting surface for Dca (Fig. 4 A and Additional file 1: Fig. S12). As for *D. melanogaster* Dca, two out of the three PSS not found in the lid region of the inferred protein models or surrounding the active site, are found in the interface region here predicted for Regucalcin (S^{157} and S^{197}; Fig. 4B). It is thus possible that there is more than one interface region in both Regucalcin and Dca. PSS away from the lid and the active site may be involved in the destabilization/strengthening of ancestral PPIs, or even be involved in the establishment of novel PPIs after the duplication event.

SVCT evolution in animals

The observation that *B. mori* is unable to synthesize enough VC during larval stages 3-LE up to 5-LE for its needs [20], implies the presence of a functional VC transporter in the gut tissues for the dietary uptake of this nutrient. In vertebrate species, this is the primary role of the SVCT1 protein [36]. A VC transporter with a function similar to SVCT1 protein should thus be present in *B. mori*, as well as in the many other protostomian species that are believed to be unable to synthesize VC. It should, however, be noted that in vertebrates not all SVCT are able to transport VC [36]. Therefore, the presence of a *SVCT*-like gene by itself cannot be taken as evidence for VC transport. The phylogenetic inferences here performed are also an opportunity to elucidate the origin and evolution of *SVCT* genes.

The more basal taxonomic groups depict an overall complex evolutionary history. The non-bilaterian *SVCT* (*SVCTNB*) gene appears independently duplicated in the represented taxonomic groups, with the exception of the Hydrozoa phylum. (Fig. 5 and Additional file 1: Fig. S14).

The Non-Lophotrochozoa *SVCT-protostomian* gene (*SVCTP NL*) might have been duplicated in the ancestral of the Pancrustacea, since two genes can be inferred for the Insecta and Crustacea groups (Fig. 5 and Additional file 1: Fig. S15). Nevertheless, the vast majority of insect species only have one copy (Gene *I1*). Therefore, it is likely that the single *SVCTP NL* gene found in *B. mori* is the VC transporter in this species. Curiously, the Coleoptera species seem to have lost the *SVCTP NL* gene. In the Nematoda taxonomic group, the *SVCTP NL* gene is also missing in 29 out of the 35 species analysed (Fig. 5 and Additional file 1: Fig. S15), including *C. elegans*, where VC synthesis has been described [18], and all the Trichinellida species. The loss of a SVCTP transporter in the coleopterans and many nematodes suggests that such species do not rely on an external VC source, implying that they must be able to synthesize their VC. Nevertheless, the possibility that a different VC transporter is used cannot be completely ruled out. Moreover, it suggests that in these lineages SVCTP has a dedicated function, such as VC transport, rather than a generalized function, since the loss of a gene performing multiple functions is likely lethal. The Platyhleminthes also suffered a notable *SVCTP NL* loss, since this gene is only present in one out

(See figure on next page.)

Fig. 3 Distribution of PSS on predicted I-TASSER protein structures. The representative protein sequences used for the inferences were *D. melanogaster* AAN09306.2, *D. melanogaster* AGB95961.1, *Bombus terrestris* XP_020722941.1, *Acromyrmex echinatior* XP_011068254.1, *Bombyx mori* XP_012549788.1, *Poecilia mexicana* XP_014859494.1, *Danio rerio* NP_991309.1, *Gekko japonicus* XP_015266456.1, *G. japonicus* XP_015266450.1, *Coturnix japonica* XP_015706795.1, *C. japonica* XP_015706720.1 and *H.* sapiens NP_690608.1, respectively from the top left to the bottom right. The overall predicted structures are highlighted in cyan, while the identified PSS are marked in dark blue, brown and purple according to their proximity to the active site (close to distant, respectively) and tagged with the corresponding amino acid symbol and position in the reference sequence

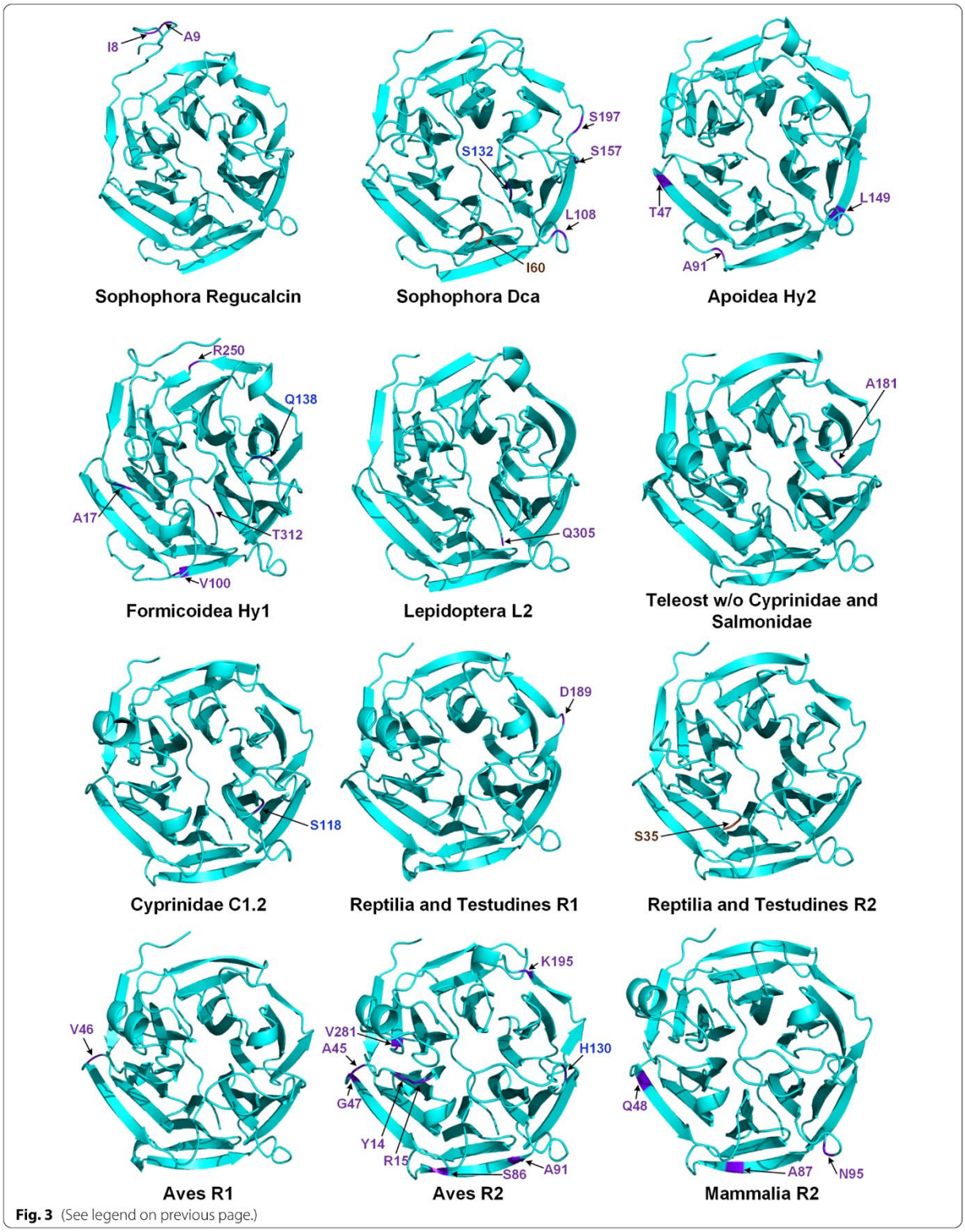


Fig. 3 (See legend on previous page.)

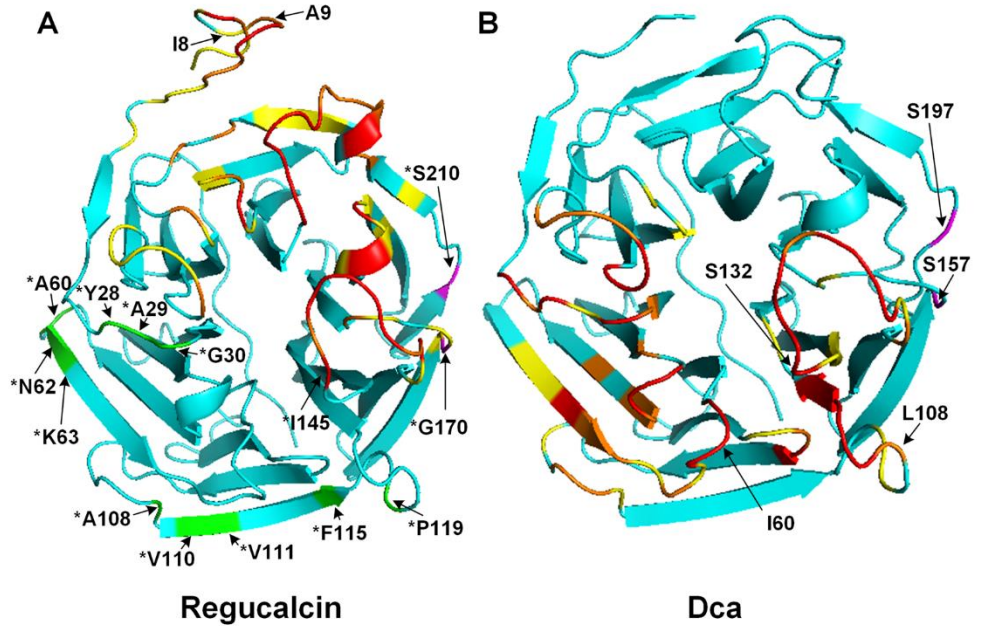


Fig. 4 Distribution of predicted interacting residues on the *D. melanogaster* Regucalcin AAN09306.2 (**A**) and Dca AGB95961.1 (**B**) models. The overall predicted structures are highlighted in cyan, while the interacting residues are marked in yellow (one predicted interaction), orange (two predicted interactions) and red (three predicted interactions). The previously inferred PSS positions are represented by the corresponding amino acid symbol and sequence position in both proteins. The I^{145},G^{170} and S^{210} Regucalcin positions (homologous to the Dca S^{132},S^{157} and S^{197}) are also displayed, as well as the homolog amino acids relative to the remaining datasets PSSs, with a "*" prefix. The Dca PSSs that were not correlated with the interaction regions (S^{157} and S^{197}), as well as their homologues in Regucalcin (G^{170} and S^{210}), are highlighted in magenta, while the remaining datasets PSSs homologs are shown in green. Although the distribution of interacting surfaces is distinct, the lid and active site regions of Regucalcin and Dca appear to be conserved

(See figure on next page.)

Fig. 5 Proposed *SVCT* gene evolution in Non-Bilateria (*SVCTNB*), protostomian Non-Lophotrochozoa (*SVCTP NL*), protostomian Lophotrochozoa (*SVCTP L*) and basal deuterostomian (*SVCT BD*) species. Gene presence is highlighted in green, absence in red and uncertainty regarding gene loss/presence in light blue. Gene duplications affecting a single species of a genus are marked with a *, while those which affected two or more species from a genus within a defined taxonomic group are marked with a #. The number of species present in the final dataset and total number of species analysed for each taxonomic group can be seen within parentheses, respectively. The inferred gene duplicates are represented next to the taxonomic group names. The cladogram topology is depicted as in the Tree of Life web project [87]

of nine species analysed (Fig. 5 and Additional file 1: Fig. S15).

The *SVCTP* evolutionary history in the Lophotrochozoa (*SVCT L*) has a remarkably difficult interpretation, mainly due to the numerous Bivalvia duplicates present (Fig. 5 and Additional file 1: Fig. S16). Indeed, the distribution pattern of the Bivalvia sequences implies up to 10 distinct genes (*Lop1* to *Lop10*). It is unclear whether each of the multiple genes found in Lophotrochozoa is dedicated to a specialized function as observed in vertebrates [36].

The results regarding the basal deuterostomians *SVCT* (*SVCT BD*) should be interpreted with caution, since the phylogeny obtained depicts considerable branch politomy (Additional file 1: Fig. S17). Nevertheless, one possible hypothesis implies the occurrence of a *SVCT BD* duplication in the common ancestor of Echinodermata and Hemichordata, and independent duplications in the ancestors of the Urochordata and Cephalochordata, respectively (Fig. 5 and Additional file 1: Fig. S17). If so, the duplicates themselves have been further duplicated in specific lineages/species.

Concerning the vertebrate species, the *SVCT1* gene is present in all the main lineages analysed (Fig. 6 and Additional file 1: Fig. S18). This gene is duplicated in salmonids and cyprinids, likely due to the two independent WGD events known to have affected the salmonid and some cyprinid species [51]. Although this gene is

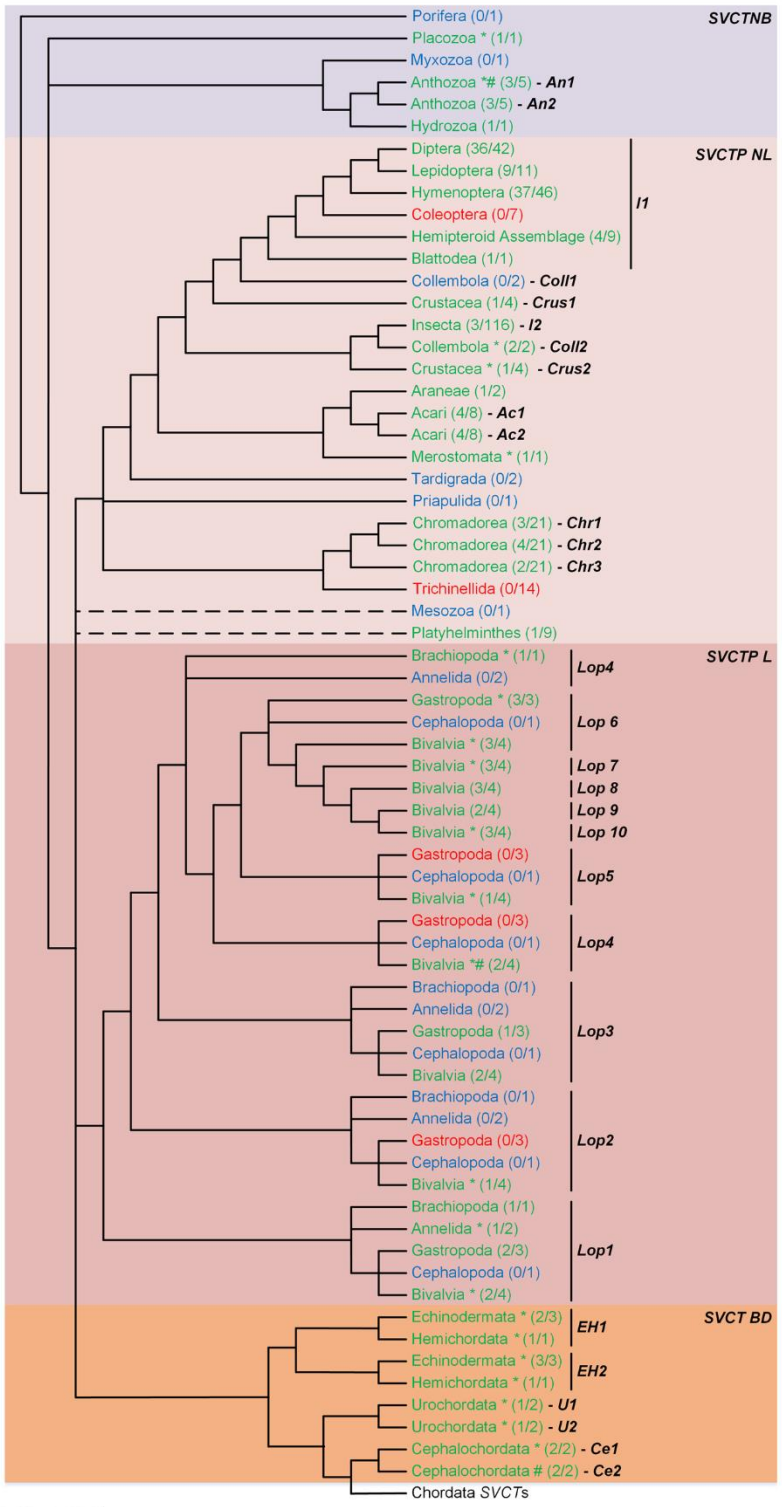


Fig. 5 (See legend on previous page.)

present in all the main vertebrate groups, it is only present in 17 out of 67 species within the Aves class, suggesting a considerable number of loss events in this lineage. Since SVCT1 is involved in VC uptake, this observation suggests that many Aves species do not rely on an external VC source. This is surprising, since several birds are known to have lost *GULO* [13, 21], and are theoretically incapable of synthesizing VC through the known animal pathway. From our data, 13 out of 67 species spread across nine Aves orders exhibit a lack of both *GULO* and *SVCT1*. It is, however, possible that in these 13 species *GULO* may be badly annotated or present in unplaced or non-sequenced regions of the genome.

The *SVCT2* gene appears to be present in all the main vertebrate lineages, apart from the Monotremata, Xenarthra and Chondrichthyes (Fig. 6 and Additional file 1: Fig. S19). Evidence for *SVCT2* gene duplications can be observed within the Actinopteri, possibly explained by the WGD events that occurred in the Salmonidae and Cyprinidae lineages [51], and an independent specific gene duplication in the Osteoglossidae family. Regarding the Amphibia, the extrapolated WGD that affected *X. laevis* [52] can explain the single duplication observed. Furthermore, the *SVCT2* gene presence within the Aves class is remarkably more accentuated when compared to that observed for *SVCT1*, as it is detected in 47 out of 67 species analysed.

The *SVCT3* gene does not seem to be closely related with either *SVCT1*, *SVCT2* and *SVCT4*. The somewhat distant phylogenetic relation with *SVCT1* and *SVCT2* was not surprising, as it was previously hypothesized that *SVCT3* diverged from these genes at an early evolutionary stage [36]. Nevertheless, the distant phylogenetic relation of this gene relative to *SVCT4* was not obvious, given that the SVCT3 transporter is proposed as the functional replacement of SVCT4 in species where *SVCT4* is pseudogenized [36]. However, a previous report revealed that the *SVCT3* sequences have unstable phylogenetic positions, even when using several prediction models [41]. An accelerated rate of nucleotide substitutions could be due, for instance, to a period of neutral evolution, before neo-functionalization. Although a conclusive explanation for this phenomenon cannot be extrapolated from this data, this hypothesis is compatible with the idea that *SVCT3* and *SVCT4* may derive from a single ancestral lineage affected by an event of WGD. Curiously, the *SVCT3* gene does not seem to be duplicated in any taxonomic group (Fig. 6 and Additional file 1: Fig. S20). In addition, the majority of Actinopteri (37 of 45) and almost half of the Aves species (31 of 67) seem to have lost the *SVCT3* gene. Although the loss of *SVCT3* in a large number of teleost fish was previously identified [41], to our knowledge it was not yet reported concerning the Aves class. In mammals, SVCT3 is known to be mainly present in the renal proximal straight tubule segments of the kidney, and is likely involved in molecule reabsorption processes [36], and thus *SVCT3* gene losses may be correlated with differences in the excretion mechanism of the species being compared.

SVCT4 is present within most of the vertebrate taxonomic groups (Fig. 7 A and B, and Additional file 1: Fig. S21). This gene is duplicated in 23 Acanthomorpha species, likely because of the teleost-specific WGD that is proposed to have taken place in the common ancestor of all teleosts [51, 63]. This hypothesis is compatible with the absence of duplications in *L. oculatus*, which suggests that the duplication event did not affect species from the Teleost sister group, Holostei. In addition, the *SVCT4* gene present in several Salmonidae and Cyprinidae species seem to have been affected by lineage-specific WGD events. A previous report stated that as in humans, several higher primate species could have lost the *SVCT4* gene [44], a finding that is here replicated. Nevertheless, the *SVCT4* gene is still found in three primate species, namely *Carlito syrichta* (Haplorrhini), *Propithecus coquereli* and *Microcebus murinus* (both Strepsirrhini). It is important to note that although *C. syrichta* belongs to the Haplorrhini suborder (where *SVCT4* is mostly loss or pseudogenized), it does not belong to the Simiiformes group (higher primates), but to the Tarsiiformes infraorder [64]. Given this evidence, it is likely that the *SVCT4* gene may have been lost in the Simiiformes, after the split of the Tarsiiformes. The initial *SVCT4* phylogenetic inference revealed an isolated cluster of 27 Actinopteri and three Amphibia coding sequences relative to the identified *SVCT4* sequences, leading to the creation of the *SVCT5* phylogeny (Figure S22). This phylogeny points to two possible scenarios: (i) a putative *SVCT4* duplication event that preceded the split of the Actinopteri and Amphibia taxonomic groups that originates the *SVCT5* gene (Fig. 7A), or (ii) two specific and independent *SVCT4* duplication events at the base of the Actinopteri and Amphibia taxonomic groups that originated two distinct genes similar to *SVCT4* (arbitrarily designated

(See figure on next page.)

Fig. 6 Proposed *SVCT1*, *SVCT2* and *SVCT3* gene evolution in vertebrate species. Gene presence is marked in green, absence in red and uncertainty regarding gene loss/presence in blue. The number of species present in the final dataset and total number of species analysed for each taxonomic group can be seen within parentheses, respectively. The identified gene duplicates are represented next to the cladogram. The known vertebrate WGD events are represented by orange circles in the branches, the Teleost, Salmonidae and Cyprinidae specific WGDs are represented by blue circles and the *X. laevis* WGD by green circles. The cladogram topology is depicted as in the Tree of Life web project [87]

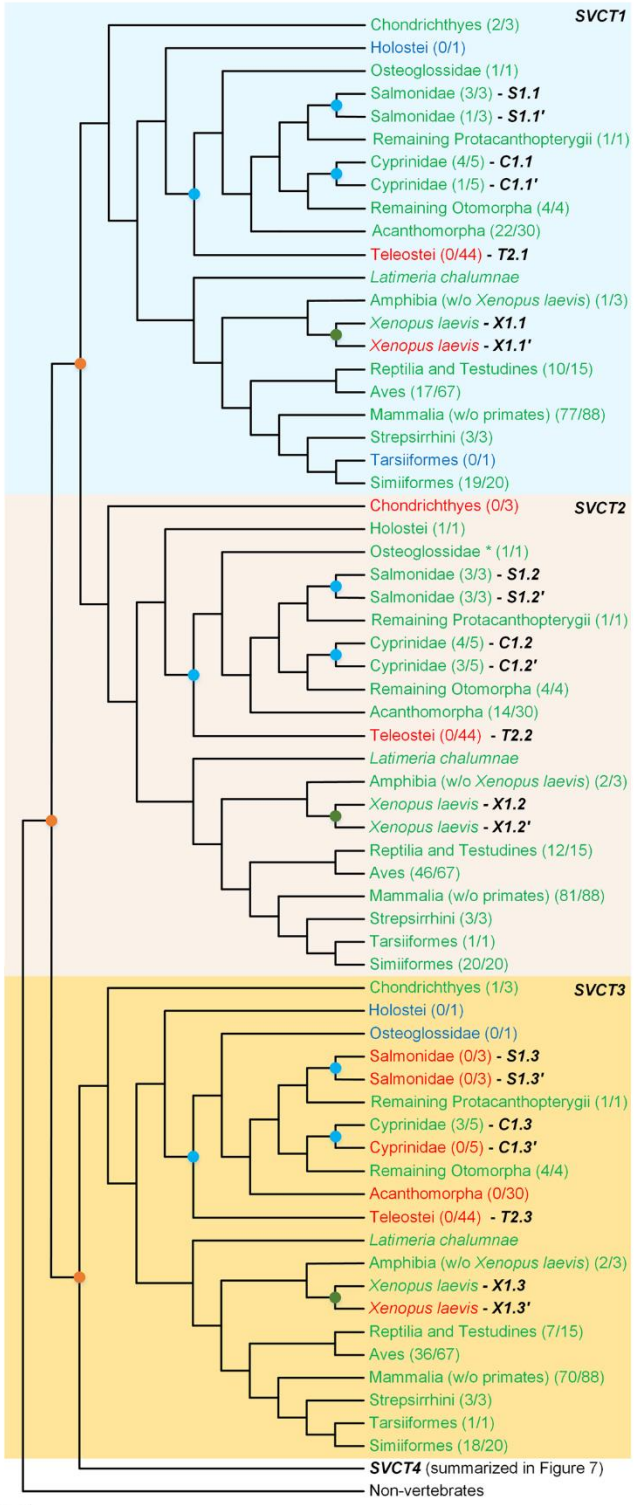


Fig. 6 (See legend on previous page.)

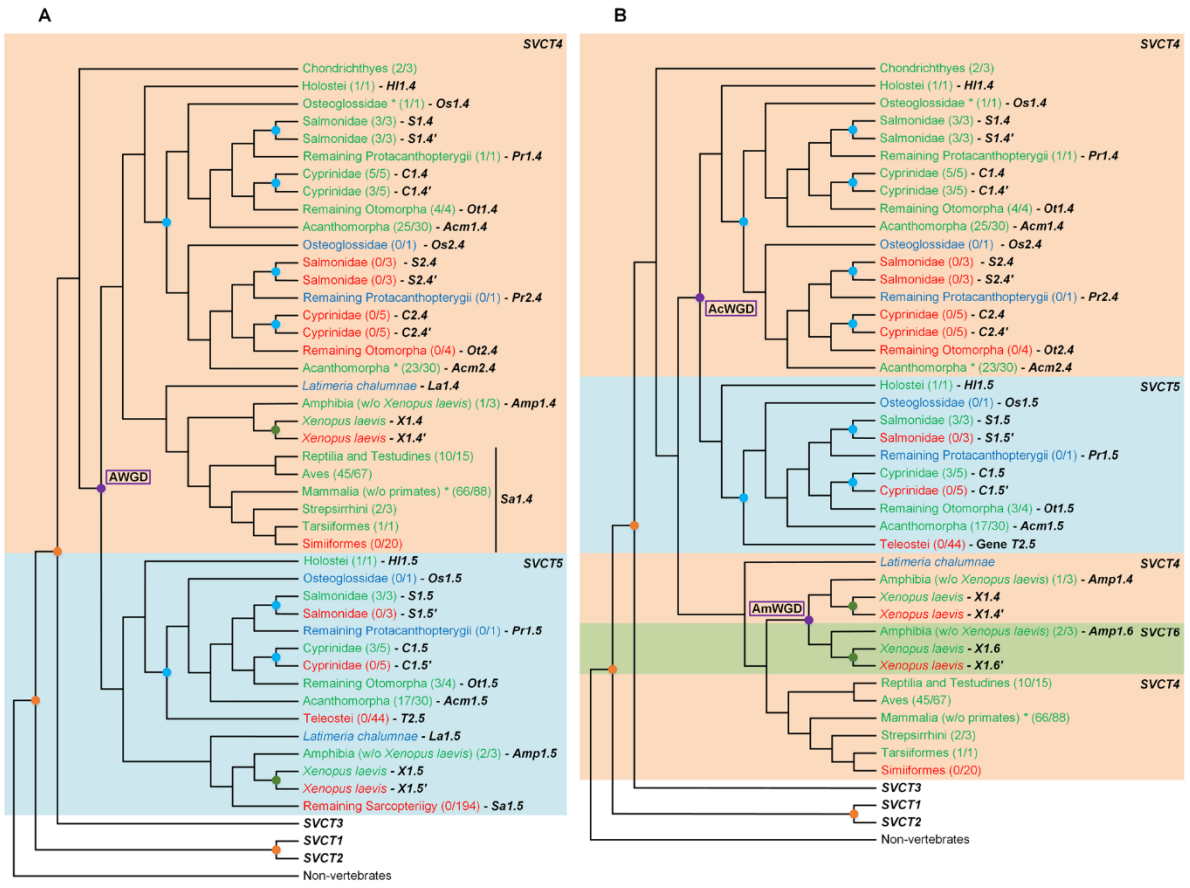


Fig. 7 Proposed *SVCT4* gene evolution in vertebrate species. Hypothesis (**A**) considers a possible ancestral duplication event (AWGD) that preceded all vertebrates after the separation from the Chondrichthyes, originating a putative *SVCT5*. Hypothesis (**B**) considers that two independent duplication events might have affected the Actinopteri and Amphibia groups (AcWGD and AmWGD, respectively), originating the *SVCT5* and *SVCT6* genes. *SVCT4, SVCT5* and *SVCT6* presence is highlighted in green, absence in red and uncertainty regarding gene loss/presence in light blue. The number of species present in the final dataset and total number of species analysed for each taxonomic group can be seen within parentheses, respectively. The identified gene duplicates are represented next to the cladogram. The known vertebrate WGD events are represented by orange circles in the branches, the Teleost, Salmonidae and Cyprinidae specific WGDs are represented by blue circles and the *X. laevis* WGD by green circles. The hypothesized duplication events are highlighted in purple. The cladogram topology is depicted as in the Tree of Life web project [87]

as *SVCT5* and *SVCT6*, respectively; Fig. 7B). Regardless of the evolutionary origin, these sequences appear to be clearly distinguishable from the known vertebrate *SVCT* genes and may represent novel nucleobase-ascorbate transporters yet uncharacterized.

SVCT functions in non-deuterostomian species

The inferences made on the evolutionary history of animal SVCTs (see above), in combination with the functional *B. mori* studies [20], suggests that SVCTP is able to perform VC transport. Therefore, as performed for Regucalcin, the comparison of the *D. melanogaster* SVCTP interactome with that of vertebrate SVCTs could provide evidence to support this hypothesis. It should be noted that although 25, 11, and 2 interactors have been reported for *H. sapiens* SVCT1, SVCT2 and SVCT3, there is no overlap between these and the *D. melanogaster* SVCTP paralog interactors (Additional file 1: Table S4). Moreover, no SVCT interactors have been reported for *M. musculus*. Therefore, at present, a given function cannot be correlated with the presence of a given interactor.

Amino acid motifs may also give information on substrate specificity, such as the SSSP motif reported in the literature [41]. This amino acid motif is only found in SVCT1 and SVCT2 proteins, the two SVCTs so far reported to be able to transport VC. None of the basal deuterostomian, protostomian and non-bilaterian SVCT sequences have this motif. It should be noted that in the

SSSP motif, the most important amino acids for substrate affinity (the second serine and the proline; [41]) are not simultaneously found in any sequence from non-deuterostomian and basal deuterostomian species. Non-bilaterian species from the Placozoa, Hydrozoa and Anthozoa phyla, as well as the majority of the basal deuterostomian species, have, however, one amino acid motif that is present in the SVCT4 transporter (SYSE), suggesting that these SVCTs are able to transport nucleobases. Insect species present the consensus motif TFGE, which is remarkably similar to the *E. coli* Uracil permease motif (TYGE; [41]). In *Drosophila*, uracil secreted by gut pathogens is crucial to activate the intestinal dual oxidase immune response at the midgut level [65]. This implies that a transporter of this nucleobase at the gut level might play an important role in host/pathogen interaction, in addition to the inferred uracil homeostasis functions. Although the *D. melanogaster SVCTP* has low expression in the midgut, it has moderate expression in the hindgut and high expression on the Malphigian tubules (https://flybase.org/reports/FBgn0037807), and therefore might intervene in such process in these specific tissues. The models presented in [65] consider only the recognition of uracil by surface receptors in enterocytes, but the transport of uracil and subsequent intracellular recognition of this nucleobase is also a possibility, since immune responses can also be mediated by cytoplasmic receptors [66]. Although the motif analyses suggest that SVCTNB and SVCTP are able to perform the transport of nucleobases only, they must also be able to transport VC given that *B. mori* is unable to synthesize enough VC during larval stages 3-LE up to 5-LE for its needs, and no other VC transporter is known [20].

Discussion

Regucalcin, *GULO* and *SVCT*: a history of neofunctionalization, pseudogenization, and subfunctionalization

A summary of the *Regucalcin*, *GULO* (data from [19, 21]) and *SVCT* presence/absence along the animal tree is shown in Fig. 8. For all lineages where there is enough data to be confident that a given gene is missing, when *GULO* is present, there is at least one *Regucalcin* homolog also present at a similar or higher frequency, with the exception of the Gastropoda. Therefore, species where a *GULO* gene is found and that belong to these lineages are inferred to synthesize VC through the described animal pathway. Gastropoda may also use the described animal pathway to synthesize VC, although in this case we have to assume that different Gastropoda species use different *Regucalcin*-like genes that belong to different gene lineages. In Monotremata, Priapulida, Myxozoa and Placozoa, a *GULO* gene has been identified but not a *Regucalcin* gene. Nevertheless, given that a single species is being analysed in each case, and given the possibility that genomes may not have been fully sequenced, at this point, we still consider that it is likely that these species synthesize VC through the described animal pathway.

As suggested before, *GULO* is likely only involved in the last step of the VC synthesis [13]. Despite the 2R-WGD event that likely occurred within the Craniata subphylum, after the separation of vertebrates from invertebrate chordates [47–49], the WGD that is proposed to have taken place in the common ancestor of all teleosts [51, 63], the two independent WGD that affected Salmonidae and Cyprinidae [51], and the WGD that affected *X. laevis* [52], *GULO* is almost always found as a single copy gene. As such, it seems that there is no advantage in having high L-gulonolactone oxidase activity levels. *GULO* is prone to pseudogenization in the lineages where VC is acquired through a VC-rich diet, which is likely the case of Haplorrhini primates, Passeriformes, Teleostei fish and some bats (included in the Laurasiatheria group). This

(See figure on next page.)

Fig. 8 Patterns of *Regucalcin*, *GULO* and *SVCT* gene presence/absence within the animal kingdom. NB, PRO and DEU represent the non-bilaterian, protostomian and deuterostomian species, respectively. Circles are representative of the *Regucalcin* gene, while squares and triangles represent the *GULO* and *SVCT* genes, respectively. The figures' fill is indicative of the percentage of species in which the given gene was identified relative to the total number of analysed species, while grey figures represent complete gene absence. The estimated GULO presence/absence frequency is based on previous reports [19, 21] where species for which no gene annotation is available, and thus where a manual gene annotation was performed, were also considered. The number of figures for each gene indicates the identified ancestral duplicates within each group/species. Numbers highlighted in the figures represent the inferred specific gene duplications found within the lineages. The purple colour outline within the *Regucalcin* Diptera section of the image represents the Sophophora *Dca* gene, while the dark blue, blue and orange represent the Drosophila *Regucalcin*, the Tephritidae *Regucalcin-like* and Culicidae *Regucalcin-like* genes, respectively. The four known *SVCT* genes within the vertebrates lineage are represented in red (*SVCT1*), orange (*SVCT2*), dark blue (*SVCT3*) and blue (*SVCT4*), while the putative *SVCT5* is highlighted in light blue. The *SVCTNB*, *SVCTP NL*, *SVCTP L* and *SVCT BD* genes are represented in burnt yellow. The blue branches in the cladogram highlight the taxonomic groups represented by less than three species in the analyses. The last two columns summarize the current hypothesis regarding VC synthesis (via GULO pathway) and uptake from diet, respectively. Within these, the information highlighted in light pink represents possible correlations between loss of *SVCT1* and *GULO* conservation, and vice-versa. The cladogram topology is depicted as in the Tree of Life web project [87]. The Euarchontoglires and Haplorrhini groups in the cladogram do not consider primates and *H. sapiens*, respectively, due to the subsequent taxa expansions

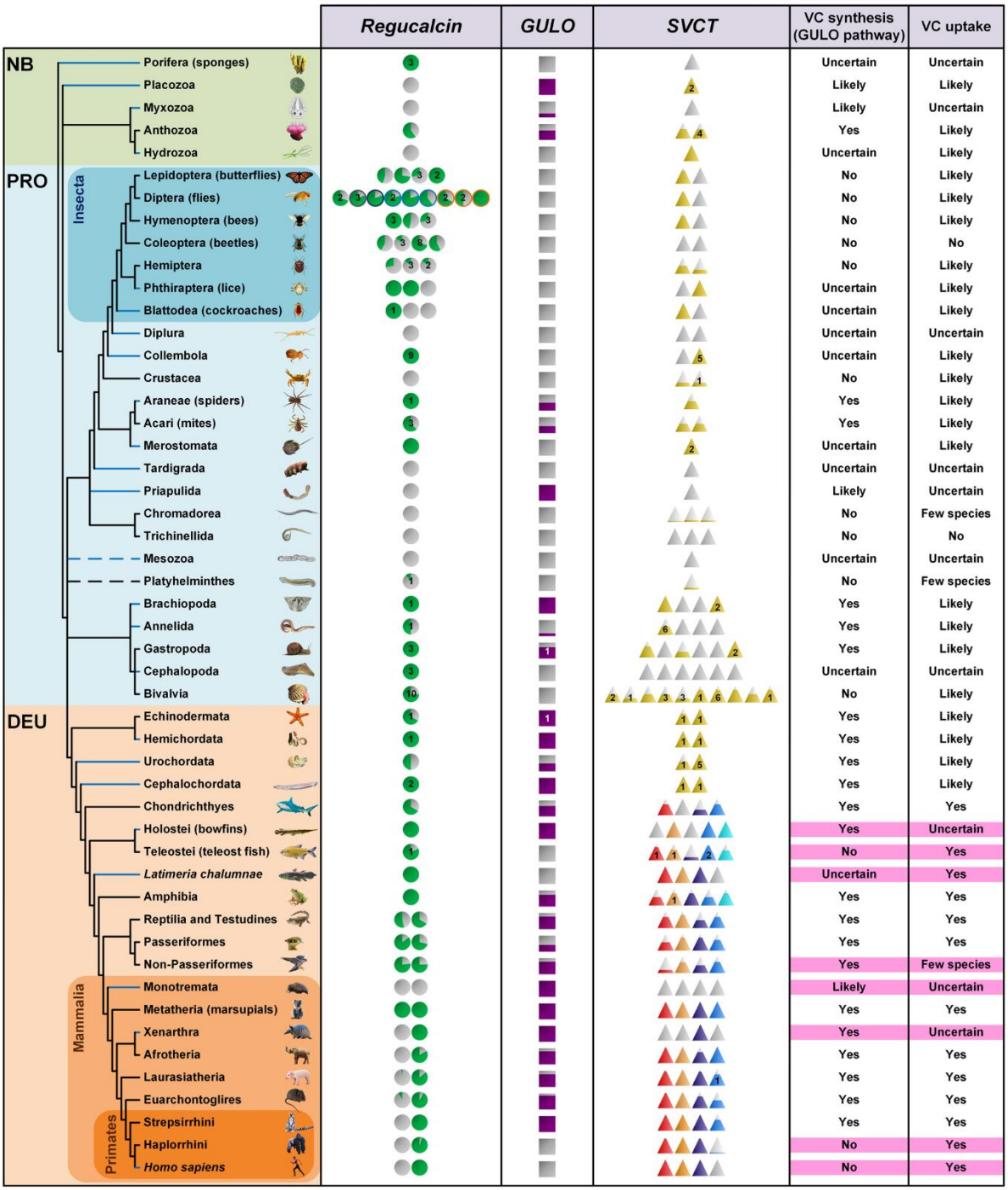


Fig. 8 (See legend on previous page.)

could also be the case of *L. chalumnae*, although in this case we cannot be certain that *GULO* was lost in this lineage since a single species was analysed. Nevertheless, if true, given that Teleostei species are more closely related to Holostei species than to *L. chalumnae*, this would imply another independent loss of the ability to

synthesize VC. This is a possibility since multiple independent *GULO* losses are common even at short evolutionary scales, as observed in bats [67].

Species from the Haplorrhini and Teleostei groups are known to develop scurvy in the absence of a VC rich diet [9, 68]. *SVCT1* is the only vertebrate gene that has been implicated in specific VC uptake from the diet [36, 42], and thus is an essential gene in species that lost the ability to synthesize VC. As expected, in Haplorrhini primates, Passeriformes and Teleostei fish, a *SVCT1* gene is almost always found. *SVCT1* can only be found in a few non-Passeriformes species, which implies that many of the analysed non-Passeriformes are unable to uptake VC from the diet. As such, they must synthesize their own VC, which is in agreement with the identification of a *GULO* gene in most non-Passeriformes species (Fig. 8). The loss of the *SVCT1* gene means that *GULO* is now an essential gene in this lineage, which can no longer be lost by pseudogenization. It is known that the basally branching actinopterygian fishes can synthesize VC in their kidneys [69], as amphibians, reptiles and the basally branching birds do [13], which is in agreement with the identification of a single copy of *GULO* and *Regucalcin* in Holostei fish (Fig. 8). Although a single species is being considered, and as such it is impossible to infer with certainty any gene loss in the Holostei lineage, the absence of a *SVCT1* gene, as here observed, would explain why *GULO* cannot be lost in this lineage.

Although a *GULO* gene is present in echinoderms, sea cucumbers have a daily VC requirement in their diet similar to that found in Teleostei fish [70]. Therefore, at least one of the two *SVCT*s found in these basal deuterostomian species must be able to transport VC. This observation suggests that SVCTs were already able to transport VC before the two-round whole genome duplication (2R-WGD) event that likely occurred within the Craniata subphylum, after the separation of vertebrates from invertebrate chordates [47–49]. This is further supported by the observation that the insect *B. mori* is unable to synthesize enough VC during larval stages 3-LE up to 5-LE for its needs [20], implying that the only SVCTP present in this species must be able to transport VC. Moreover, in bees, the dietary supplementation of VC increases resistance to infection, showing that VC must have been transported [71]. Therefore, despite the inferences here made based on amino acid patterns that SVCTPs are involved in the transport of nucleobases only, they must also transport VC. As such, the 2R-WGD event likely resulted in subfunctionalization and not in neofunctionalization. In some fish and amphibian species yet undescribed *SVCT* genes are here found, which might represent further subfunctionalization or neofunctionalization processes.

In almost all protostomian and non-bilaterian groups represented by more than one species and where no *GULO* gene was found, one *SVCT* gene is always found. The clear exceptions concerns the nematodes and coleopterans, since there is no *SVCTP* in *C. elegans* and the analysed species from the Trichinellida and Coleoptera orders. Therefore, these species must exclusively rely on the synthesis of VC to meet their nutritional needs. In agreement with this observation, *C. elegans* has been shown to synthesize VC [18]. Since there is no *GULO* and *Regucalcin* in these nematode species, VC must be synthesized through a novel pathway quite different from the one so far known for animals. Coleopterans have multiple *Regucalcin*-like genes, and as such, should theoretically be capable of synthesizing VC through one of the two hypothesized insect pathways. Since the coleopterans have a homolog of the *B. mori Gulo-like* gene while the dipterans such as *D. melanogaster* do not (data not shown), it is more parsimonious to assume that the coleopterans likely rely on the *B. mori* pathway to obtain this nutrient.

Besides VC synthesis, *Regucalcin* is also likely involved in calcium homeostasis, and oxidative stress response in Deuterostomes, Protostomes and non-Bilateria animals. Whether *Regucalcin's* gluconolactonase activity is essential for calcium homeostasis is unclear, since this role may be played through the inferred binding surface of Regucalcin-like proteins. In animals, with the clear exceptions of Insects and the Lophotrochozoa group, *Regucalcin* tends to be a single copy gene. Considering the aforementioned WGD events that likely occurred in vertebrates relative to the results here obtained, it can be concluded that the most likely fate of a *Regucalcin* duplicate is to be lost. The *Regucalcin* gene duplication inferred at the base of the reptiles/birds/mammals lineage also implies multiple independent gene losses of the *Regucalcin* duplicate. The most parsimonious explanation for these observations is that, generally speaking, despite its many roles, *Regucalcin* rarely evolves a new function. Insects are a clear exception, since *Regucalcin* homologues may have acquired a new function (neofunctionalization) related to the receptor-mediated uptake of hexamerin storage proteins from insect haemolymph by fat body cells, a feature that is only found in this group of species [72]. While in insects it is unclear whether *Regucalcin* plays any role in VC synthesis (with the possible exception of *B. mori*; [20]), this gene is likely involved in this process in the Brachiopoda, Annelida and Gastropoda, since *GULO* is present in some species from these groups, and as such, no *Regucalcin* neofunctionalization is expected. In Bivalvia, however, where no *GULO* gene has been found, it is possible that *Regucalcin* evolved new functions.

Conclusions

In animals, VC is an essential nutrient that can be synthesized/acquired from the diet. The gene involved in the last step of the only described animal synthesis pathway (*GULO*) has been lost multiple times during evolution in non-bilateria, protostomes and deuterostomes. Species that lack *GULO* are dependent on the VC present in the diet, which implies the presence of a functional VC transporter. It is here argued that this is mainly the case, although there are clear exceptions such as *C. elegans*, where a new VC synthesis pathway must be present. Here, it is also shown why in some basal animal lineages the ability to synthesize VC cannot be lost, even in the presence of VC rich diets, due to the loss of the VC transporter. Homologs of the gene involved in the penultimate step of VC synthesis (*Regucalcin*) are shown to be likely involved in VC synthesis, calcium homeostasis and the oxidative stress response in all animals, and to have acquired new roles during evolution in some taxonomic groups such as insects. Despite its multiple functions, there is no *Regucalcin* gene in *C. elegans*. Therefore, the presence of a *Regucalcin* gene alone cannot be taken as an indication of a partially conserved VC synthesis pathway in species where there is no *GULO* but VC synthesis occurs, such as insects.

Methods

Regucalcin and *SVCT* phylogenetic analyses

An adaptation of the protocol described in [21] was used to obtain the sequence sets for the phylogenetic inferences. CDS were downloaded from the NCBI RefSeq database, and sequence operations performed using the SEquence DAtaset builder's (SEDA) software [73]. The protein queries for the initial tblastn were *H. sapiens* Regucalcin (accession number NP_690608.1) and three distinct SVCT sequences, namely *H. sapiens* SVCT1 (NP_689898.2; XP_011542067.1), SVCT2 (CAB58120.1; NP_005107.4) and SVCT3 (NP_001138362.1). To obtain the final consensus *SVCT* dataset, the three distinct *SVCT1*, *SVCT2* and *SVCT3* output datasets were processed separately, subsequently merged, and then two additional operations, namely, "remove redundant sequences" and "remove isoforms", performed. Two *Manacus vitelinus* CDS with the same header and accession number (XP_008924532.1) but different nucleotide sequences (99% identical) caused issues with the isoform removal step when processing the *SVCT* datasets. This problem was solved by the random removal of one of the sequences. For both *Regucalcin* and *SVCT*, the complete list of sequences that have been removed is presented in Additional file 1: Table S5. Bayesian phylogenetic inferences were performed using ADOPS [74], using 1,000,000 generations for *Regucalcin* and 5,000,000 for *SVCT*s. The first 2500 *Regucalcin* and 12,500 *SVCT* samples were discarded (burn-in).

The phylogeny obtained using all *Regucalcin* sequences, had poor branch support values, due to the use of very divergent sequences (Additional file 1: Table S5). Therefore, the dataset was also divided into smaller datasets in order to analyze sequences from Non-Bilateria, Hemipteroid Assemblage and Blattodea, Coleoptera, Hymenoptera, Diptera, Lepidoptera, Non-Lophotrochozoa (without Insects), Lophotrochozoa, Basal Deuterostomia and Vertebrates species, independently of sequences from the other groups. Moreover, sequences associated with very long branches were also removed from the datasets. The resulting datasets were then analysed as described above. Convergence was achieved in all cases. The *H. sapiens Regucalcin* sequence (NP_690608.1) was used as root outgroup for the Non-Bilateria, protostomian and basal deuterostomian datasets, while the urochordate *C. intestinalis* sequence (XP_002120764.1) was used for the deuterostomian dataset.

The *SVCT* phylogeny did not converge for all the model parameters, likely due to the inherent difficulty in the alignment of sequences of distinct genes. To overcome this technical issue, sequences that caused abnormal branch lengths in the tree were removed (Additional file 1: Table S5), and the dataset was subdivided into smaller datasets representative of all *SVCT* genes here found *(SVCT1*, *SVCT2*, *SVCT3*, *SVCT4* and *SVCT5*), as well as four subsets, representative of non-bilaterians (*SVCTNB*), protostomian Lophotrochozoa (*SVCTP L*), protostomian Non-Lophotrochozoa (*SVCTP NL*) and basal deuterostomians (*SVCT BD*). Outgroup sequences were added to each dataset, from the following sequence pool: *D. melanogaster SVCTP* (AAF54519.1), *H. sapiens SVCT1* (XP_011542067.1), *H. sapiens SVCT2* (NP_005107.4), *H. sapiens SVCT3* (NP_001138362.1), *M. musculus SVCT4* (XP_006506197.1), *Trichoplax adhaerens SVCTNB* (EDV23955.1), *Aplysia californica SVCTP* (XP_012935012.1) and *Acanthaster planci SVCT* (XP_022103910.1). The resulting datasets were then analysed as described above, using 1,000,000 generations and a burn-in of 2500. Convergence was achieved in all cases.

The phylogenies obtained for *Regucalcin* and *SVCT* were rooted using MEGA X [75], after conversion of the files from Nexus to Newick format using the application available at http://phylogeny.lirmm.fr/phylo_cgi/data_converter.cgi.

Positively selected amino acid sites inference

Positively selected amino acid sites (PSS) were inferred using both codeML [76] as implemented in ADOPS [74] and FUBAR [77] as implemented in DataMonkey [78].

Gene sequence sets were defined based on the *Regucalcin* Bayesian phylogenies. Only sets with a minimum of five species were considered. The considered FASTA files represented the Sophophora *Regucalcin*, Sophophora *Dca*, Culicidae *Cu3*, Apoidea *Hy2*, Formicoidea *Hy1*, Lepidoptera *L2*, *L3* and *L4*, Ciprinidae *C1.2*, Teleost_w/o_Cyprinidae_Salmonidae and the Reptilia and Testudines, Aves and Mammalia *R1* and *R2* genes. codeML did not produce a valid output for the Aves *R1*, Mammalia *R2* and Teleost_w/o_Cyprinidae_Salmonidae datasets due to an unknown internal error that persisted after several trials. The results of the remaining datasets were deposited in the B + database (bpositive.i3s.up.pt; "The evolution of regucalcin-like genes" (BP2018000005)), which allows for the visualization and download of the project by any user [79, 80]. Inferred PSS with a probability equal or higher than 90% identified by both approaches, were considered as positive. In the Aves *R1*, Mammalia *R2* and Teleost_w/o_Cyprinidae_Salmonidae datasets, only the FUBAR results were considered. For each dataset, the PSS were marked on a 3D model structure prediction of a representative protein sequence, obtained using I-TASSER [81–83].

D. melanogaster Regucalcin and *SVCT* interactome analyses

Protein-protein interactions (PPI) were obtained from EvoPPI 1.0 web tool [62], a platform that compiles all PPI reported in 12 databases. Searches were performed by selecting the appropriate species under the "Same species" query, and by using all available interactome databases. Only direct interactions (Interaction level 1) were considered. The queried gene IDs are for *D. melanogaster*: 32165 (*Regucalcin*), 41786 (*Dca*; there are no reported interactions), and 41259 (*SVCTP*); for *M. musculus*: 19733 (*Regucalcin*), 20522 (*SVCT1)*, 54338 (*SVCT2*) and 22626 (*SVCT3*); for *H. sapiens*: 9104 (*Regucalcin*), 9963 (*SVCT1*), 9962 (*SVCT2*) and 151295 (*SVCT3*).

The gene IDs of the *D. melanogaster* interactors were processed using the DRSC Integrative Ortholog Prediction Tool [84] in order to determine orthologous genes in *M. musculus* and *H. sapiens*. Genes with low confidence score or without any hit were analysed using a BLASTP reciprocal best hit approach [85] using the NCBI BLAST web interface (https://blast.ncbi.nlm.nih.gov/Blast.cgi#). The selected word size was three and low complexity regions were excluded.

Regucalcin and Dca protein docking inferences

Protein docking inferences were performed using an in silico protocol available in the literature [86]. The *D. melanogaster* protein sequences were downloaded from NCBI as FASTA formatted files (accession numbers AAN09306.2 (Regucalcin), AGB95961.1 (Dca), NP_608711.3 (Histidine triad nucleotide binding protein 1), NP_476735.1 (Superoxide dismutase 1), and NP_732311.1 (14-3-3epsilon)).

D. melanogaster RNAi experiments

The *Actin5C* GAL4 driver stock (25374) was obtained from the Bloomington Drosophila Stock Center (https://bdsc.indiana.edu), while the RNAi stocks for *Regucalcin* (105509) and *Dca* (103377) were supplied by the Vienna Drosophila Resource Center (https://stockcenter.vdrc.at). Fly stocks were kept at environmental chambers with a constant temperature of 25°C and 12 h day/night cycles. Flies were reared on cornmeal food supplemented with yeast extract and depleted of VC. The driver/RNAi crosses were performed in both directions (♂ driver x ♀ RNAi; ♀ driver x ♂ RNAi) using 6 individuals of each stock per cross, using cornmeal food vials kept at 25°C and 12 h day/night cycles. The progenitors were subsequently transferred to new vials every two days to maximize sample acquisition, until six vials were gathered. At the day of birth, the progeny was sorted according to the phenotype and gender.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12862-022-02040-7.

Additional file 1. *Regucalcin and Dca D.* melanogaster RNAi crosses (**Table S1**), Regucalcin interactome information (**Table S2**), Regucalcin-like positively selected amino acid sites (**Table S3**), SVCT interactome information (**Table S4**), details on sequence data set preparation for phylogenetic analyses (**Table S5**), Regucalcin phylogenetic analyses (**Figures S1–S11**), distribution of Regucalcin predicted interacting residues (**Figures S12–S13**), and SVCT-like (**Figures S14–S22**) phylogenetic analyses.

Acknowledgements

Not applicable.

Author contributions

PD, CPV and JV designed the study. All authors carried out the experiments and analyses. All authors interpreted the data, and wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was financed by the National Funds through FCT—Fundação para a Ciência e a Tecnologia, I.P., under the project UIDB/04293/2020. PD is supported by a PhD fellowship from Fundação para a Ciência e Tecnologia (SFRH/BD/145515/2019), cofinanced by the European Social Fund through the Norte Portugal Regional Operational Programme (NORTE 2020).

Availability of data and materials

The FASTA files used for phylogenetic analyses, the PDB files containing the inferred structures and the Newick tree files are provided in Zenodo (https://doi.org/10.5281/zenodo.6518249). Positively selected amino acid inferences are publicly available at B + database (http://bpositive.i3s.up.pt) under project BP2018000005.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Rua de Jorge Viterbo Ferreira, 228, 4050-313 Porto, Portugal. [2]Faculdade de Ciências da Universidade do Porto (FCUP), Rua do Campo Alegre, s/n, 4169-007 Porto, Portugal. [3]Instituto de Investigação e Inovação em Saúde (I3S), Universidade do Porto, Rua Alfredo Allen, 208, 4200-135 Porto, Portugal. [4]Instituto de Biologia Molecular e Celular (IBMC), Rua Alfredo Allen, 208, 4200-135 Porto, Portugal.

Received: 8 October 2021 Accepted: 17 June 2022
Published online: 25 June 2022

References

1. Amatore C, Arbault S, Ferreira DCM, Tapsoba I, Verchier Y. Vitamin C stimulates or attenuates reactive oxygen and nitrogen species (ROS, RNS) production depending on cell state: quantitative amperometric measurements of oxidative bursts at PLB-985 and RAW 264.7 cells at the single cell level. J Electroanal Chem. 2008;615(1):34–44.
2. Radi AM, Mohammed ET, Abushouk AI, Aleya L, Abdel-Daim MM. The effects of abamectin on oxidative stress and gene expression in rat liver and brain tissues: modulation by sesame oil and ascorbic acid. Sci Total Environ. 2020;701:134882.
3. Figueroa-Méndez R, Rivas-Arancibia S, Vitamin C. Health and disease: its role in the metabolism of cells and redox state in the brain. Front Physiol. 2015;6:397.
4. Li X, Huang J, May JM. Ascorbic acid spares alpha-tocopherol and decreases lipid peroxidation in neuronal cells. Biochem Biophys Res Commun. 2003;305(3):656–61.
5. Szarka A, Lőrincz T. The role of ascorbate in protein folding. Protoplasma. 2014;251(3):489–97.
6. Du J, Cullen JJ, Buettner GR. Ascorbic acid: chemistry, biology and the treatment of cancer. Biochim Biophys Acta. 2012;1826(2):443–57.
7. Seo M-Y, Lee S-M. Protective effect of low dose of ascorbic acid on hepatobiliary function in hepatic ischemia/reperfusion in rats. J Hepatol. 2002;36(1):72–7.
8. Nam T-G. Lipid peroxidation and its toxicological implications. Toxicol Res. 2011;27(1):1–6.
9. Hirschmann JV, Raugi GJ. Adult scurvy. J Am Acad Dermatol. 1999;41(6):895–906 quiz 7–10.
10. Chang YL, Rossetti M, Vlamakis H, Casero D, Sunga G, Harre N, et al. A screen of Crohn's disease-associated microbial metabolites identifies ascorbate as a novel metabolic inhibitor of activated human T cells. Mucosal Immunol. 2019;12(2):457–67.
11. Wolucka BA, Communi D. Mycobacterium tuberculosis possesses a functional enzyme for the synthesis of vitamin C, L-gulono-1,4-lactone dehydrogenase. FEBS J. 2006;273(19):4435–45.
12. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. Nature. 2011;473(7346):174–80.
13. Drouin G, Godin J-R, Pagé B. The genetics of vitamin C loss in vertebrates. Curr Genomics. 2011;12(5):371–8.
14. Smirnoff N. Ascorbic acid metabolism and functions: a comparison of plants and mammals. Free Radic Biol Med. 2018;122:116–29.
15. Wheeler G, Ishikawa T, Pornsaksit V, Smirnoff N. Evolution of alternative biosynthetic pathways for vitamin C following plastid acquisition in photosynthetic eukaryotes. Elife. 2015;4:e06369.
16. Linster CL, Van Schaftingen E, Vitamin C. Biosynthesis, recycling and degradation in mammals. FEBS J. 2007;274(1):1–22.
17. Hollmann S. CHAPTER VI: Metabolism of glucuronic acid and ascorbic acid. In: Hollmann S, editor. Non-glycolytic pathways of metabolism of glucose. Cambridge: Academic Press; 1964. p. 83–114.
18. Patananan AN, Budenholzer LM, Pedraza ME, Torres ER, Adler LN, Clarke SG. The invertebrate Caenorhabditis elegans biosynthesizes ascorbate. Arch Biochem Biophys. 2015;569:32–44.
19. Henriques SF, Duque P, López-Fernández H, Vázquez N, Fdez-Riverola F, Reboiro-Jato M, et al. Multiple independent L-gulonolactone oxidase (GULO) gene losses and vitamin C synthesis reacquisition events in non-Deuterostomian animal species. BMC Evol Biol. 2019;19(1):126.
20. Hou M, Dai TM, Liang XY, Zhang SX, Cui WZ, Qiu JF, et al. Bombyx mori can synthesize ascorbic acid through the l-gulose pathway to varying degrees depending on developmental stage. Arch Insect Biochem Physiol. 2021;106(4):e21783.
21. López-Fernández H, Duque P, Henriques S, Vázquez N, Fdez-Riverola F, Vieira CP, et al. A Bioinformatics Protocol for Quickly Creating Large-Scale Phylogenetic Trees. In: Fdez-Riverola F, Mohamad MS, Rocha M, De Paz JF, González P, editors., et al., Practical Applications of Computational Biology and Bioinformatics, 12th International Conference. Cham: Springer International Publishing; 2019. p. 88–96.
22. Ewing TA, Fraaije MW, Mattevi A, van Berkel WJH. The VAO/PCMH flavoprotein family. Arch Biochem Biophys. 2017;632:104–17.
23. Yamaguchi M, Murata T. Involvement of regucalcin in lipid metabolism and diabetes. Metabolism. 2013;62(8):1045–51.
24. Ishigami A, Fujita T, Handa S, Shirasawa T, Koseki H, Kitamura T, et al. Senescence marker protein-30 knockout mouse liver is highly susceptible to tumor necrosis factor-alpha- and Fas-mediated apoptosis. Am J Pathol. 2002;161(4):1273–81.
25. Lee SF, Chen Y, Varan AK, Wee CW, Rako L, Axford JK, et al. Molecular basis of adaptive shift in body size in Drosophila melanogaster: functional and sequence analyses of the Dca gene. Mol Biol Evol. 2011;28(8):2393–402.
26. Ishigami A, Kondo Y, Nanba R, Ohsawa T, Handa S, Kubo S, et al. SMP30 deficiency in mice causes an accumulation of neutral lipids and phospholipids in the liver and shortens the life span. Biochem Biophys Res Commun. 2004;315(3):575–80.
27. Yumura W, Imasawa T, Suganuma S, Ishigami A, Handa S, Kubo S, et al. Accelerated tubular cell senescence in SMP30 knockout mice. Histol Histopathol. 2006;21(11):1151–6.
28. Nakajima Y, Natori S. Identification and characterization of an anterior fat body protein in an insect. J Biochem. 2000;127(5):901–8.
29. Hansen IA, Meyer SR, Schäfer I, Scheller K. Interaction of the anterior fat body protein with the hexamerin receptor in the blowfly Calliphora vicina. Eur J Biochem. 2002;269(3):954–60.
30. Vierstraete E, Cerstiaens A, Baggerman G, Van den Bergh G, De Loof A, Schoofs L. Proteomics in Drosophila melanogaster: first 2D database of larval hemolymph proteins. Biochem Biophys Res Commun. 2003;304(4):831–8.
31. Arboleda-Bustos CE, Segarra C. The Dca gene involved in cold adaptation in Drosophila melanogaster arose by duplication of the ancestral regucalcin gene. Mol Biol Evol. 2011;28(8):2185–95.
32. Goto SG. Expression of Drosophila homologue of senescence marker protein-30 during cold acclimation. J Insect Physiol. 2000;46(7):1111–20.
33. McKechnie SW, Blacket MJ, Song SV, Rako L, Carroll X, Johnson TK, et al. A clinally varying promoter polymorphism associated with adaptive variation in wing size in Drosophila. Mol Ecol. 2010;19(4):775–84.
34. Casas-Vila N, Bluhm A, Sayols S, Dinges N, Dejung M, Altenhein T, et al. The developmental proteome of Drosophila melanogaster. Genome Res. 2017;27(7):1273–85.
35. Brown JB, Boley N, Eisman R, May GE, Stoiber MH, Duff MO, et al. Diversity and dynamics of the Drosophila transcriptome. Nature. 2014;512(7515):393–9.
36. Burzle M, Suzuki Y, Ackermann D, Miyazaki H, Maeda N, Clemencon B, et al. The sodium-dependent ascorbic acid transporter family SLC23. Mol Aspects Med. 2013;34(2–3):436–54.
37. Savini I, Rossi A, Pierro C, Avigliano L, Catani MV. SVCT1 and SVCT2: key proteins for vitamin C uptake. Amino Acids. 2008;34(3):347–55.
38. Corpe CP, Tu H, Eck P, Wang J, Faulhaber-Walter R, Schnermann J, et al. Vitamin C transporter Slc23a1 links renal reabsorption, vitamin

C tissue accumulation, and perinatal survival in mice. J Clin Investig. 2010;120(4):1069–83.
39. Tsukaguchi H, Tokui T, Mackenzie B, Berger UV, Chen XZ, Wang Y, et al. A family of mammalian Na+-dependent L-ascorbic acid transporters. Nature. 1999;399(6731):70–5.
40. May JM. The SLC23 family of ascorbate transporters: ensuring that you get and keep your daily dose of vitamin C. Br J Pharmacol. 2011;164(7):1793–801.
41. Kourkoulou A, Pittis AA, Diallinas G. Evolution of substrate specificity in the nucleobase-ascorbate transporter (NAT) protein family. Microbial Cell. 2018;5(6):280–92.
42. Takanaga H, Mackenzie B, Hediger MA. Sodium-dependent ascorbic acid transporter family SLC23. Pflügers Archiv. 2004;447(5):677–82.
43. Duell EJ, Lujan-Barroso L, Llivina C, Muñoz X, Jenab M, Boutron-Ruault M-C, et al. Vitamin C transporter gene (SLC23A1 and SLC23A2) polymorphisms, plasma vitamin C levels, and gastric cancer risk in the EPIC cohort. Genes Nutr. 2013;8(6):549–60.
44. Yamamoto S, Inoue K, Murata T, Kamigaso S, Yasujima T, Maeda JY, et al. Identification and functional characterization of the first nucleobase transporter in mammals: implication in the species difference in the intestinal absorption mechanism of nucleobases and their analogs between higher primates and other mammals. J Biol Chem. 2010;285(9):6522–31.
45. Smith JE 3, Cronmiller C. The Drosophila daughterless gene autoregulates and is controlled by both positive and negative cis regulation. Development. 2001;128(23):4705–14.
46. Duffy JB. GAL4 system in drosophila: A fly geneticist's swiss army knife. Genesis. 2002;34(1–2):1–15.
47. Dehal P, Boore JL. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. 2005;3(10):e314.
48. Kasahara M. Impact of whole-genome duplication on vertebrate development and evolution. Semin Cell Dev Biol. 2013;24(2):81–2.
49. Van de Peer Y, Maere S, Meyer A. The evolutionary significance of ancient genome duplications. Nat Rev Genet. 2009;10(10):725–32.
50. Kasahara M. The 2R hypothesis: an update. Curr Opin Immunol. 2007;19(5):547–52.
51. Glasauer SM, Neuhauss SC. Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol Genet Genomics MGG. 2014;289(6):1045–60.
52. Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, et al. Genome evolution in the allotetraploid frog Xenopus laevis. Nature. 2016;538(7625):336–43.
53. Matthews LR, Vaglio P, Reboul J, Ge H, Davis BP, Garrels J, et al. Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs" Genome Res. 2001;11(12):2120–6.
54. Ding X-B, Jin J, Tao Y-T, Guo W-P, Ruan L, Yang Q-l, et al. Predicted Drosophila Interactome Resource and web tool for functional interpretation of differentially expressed genes. Database. 2020;2020.
55. Hart GT, Ramani AK, Marcotte EM. How complete are current yeast and human protein-interaction networks? Genome Biol. 2006;7(11):120.
56. Li X, Cai H, Xu J, Ying S, Zhang Y. A mouse protein interactome through combined literature mining with multiple sources of interaction evidence. Amino Acids. 2010;38(4):1237–52.
57. Pias EK, Ekshyyan OY, Rhoads CA, Fuseler J, Harrison L, Aw TY. Differential effects of superoxide dismutase isoform expression on hydroperoxide-induced apoptosis in PC-12 cells. J Biol Chem. 2003;278(15):13294–301.
58. Fukaya Y, Yamaguchi M. Regucalcin increases superoxide dismutase activity in rat liver cytosol. Biol Pharm Bull. 2004;27(9):1444–6.
59. Ichikawa E, Yamaguchi M. Regucalcin increases superoxide dismutase activity in the heart cytosol of normal and regucalcin transgenic rats. Int J Mol Med. 2004;14(4):691–6.
60. Yamaguchi M. The anti-apoptotic effect of regucalcin is mediated through multisignaling pathways. Apoptosis. 2013;18(10):1145–53.
61. Brodskiy PA, Wu Q, Soundarrajan DK, Huizar FJ, Chen J, Liang P, et al. Decoding calcium signaling dynamics during Drosophila wing disc development. Biophys J. 2019;116(4):725–40.
62. Vázquez N, Rocha S, López-Fernández H, Torres A, Camacho R, Fdez-Riverola F, et al. EvoPPI 1.0: a web platform for within- and between-species multiple interactome comparisons and application to nine PolyQ proteins determining neurodegenerative diseases. Interdiscipl Sci Comput Life Sci. 2019;11(1):45–56.
63. Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y. Genome duplication, a trait shared by 22,000 species of ray-finned fish. Genome Res. 2003;13(3):382–90.
64. Purvis A. A composite estimate of primate phylogeny. Philos Trans R Soc Lond Ser B Biol Sci. 1995;348(1326):405–21.
65. Lee K-A, Kim B, You H, Lee W-J. Uracil-induced signaling pathways for DUOX-dependent gut immunity. Fly (Austin). 2015;9(3):115–20.
66. Liwinski T, Zheng D, Elinav E. The microbiome and cytosolic innate immune receptors. Immunol Rev. 2020;297(1):207–24.
67. Cui J, Pan YH, Zhang Y, Jones G, Zhang S. Progressive pseudogenization: vitamin C synthesis and its loss in bats. Mol Biol Evol. 2011;28(2):1025–31.
68. Blom JH, Dabrowski K. Continuous of "pulse-and-withdraw" supply of ascorbic acid in the diet: a new approach to altering the bioavailability of ascorbic acid, using teleost fish as a scurvy-prone model. Int J Vitam Nutr Res. 1998;68(2):88–93.
69. Dabrowski K. Primitive actimoterigian fishes can synthesize ascorbic acid. Experientia. 1994;50(8):745–8.
70. Okorie OE, Ko SH, Go S, Lee S, Bae JY, Han K, et al. Preliminary Study of the optimum dietary ascorbic acid level in Sea Cucumber, Apostichopus japonicus (Selenka). J World Aquacult Soc. 2008;39(6):758–65.
71. Farjan M, Łopieńska-Biernat E, Lipiński Z, Dmitryjuk M, Żółtowska K. Supplementing with vitamin C the diet of honeybees (Apis mellifera carnica) parasitized with Varroa destructor: effects on antioxidative status. Parasitology. 2014;141(6):770–6.
72. Hansen IA, Gutsmann V, Meyer SR, Scheller K. Functional dissection of the hexamerin receptor and its ligand arylphorin in the blowfly Calliphora vicina. Insect Mol Biol. 2003;12(5):427–32.
73. López-Fernández H, Duque P, Vazquez N, Fdez-Riverola F, Reboiro-Jato M, Vieira CP, et al. SEDA: a Desktop Tool Suite for FASTA Files Processing. IEEE/ACM Trans Comput Biol Bioinform. 2020.
74. Reboiro-Jato D, Reboiro-Jato M, Fdez-Riverola F, Vieira CP, Fonseca NA, Vieira J. ADOPS–automatic detection of positively selected sites. J Integr Bioinform. 2012;9(3):200.
75. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547–9.
76. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.
77. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, et al. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol Biol Evol. 2013;30(5):1196–205.
78. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Mol Biol Evol. 2018;35(3):773–7.
79. Vázquez N, Vieira CP, Amorim BSR, Torres A, López-Fernández H, Fdez-Riverola F, et al. Large scale analyses and visualization of adaptive amino acid changes projects. Interdiscipl Sci Comput Life Sci. 2018;10(1):24–32.
80. Vázquez N, Vieira CP, Amorim BSR, Torres A, López-Fernández H, Fdez-Riverola F, et al. Automated Collection and Sharing of Adaptive Amino Acid Changes Data. In: Fdez-Riverola F, Mohamad MS, Rocha M, De Paz JF, Pinto T, editors., et al., 11th international conference on practical applications of computational biology and bioinformatics. Cham: Springer International Publishing; 2017. p. 18–25.
81. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. 2010;5(4):725–38.
82. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nat Methods. 2015;12(1):7–8.
83. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res. 2015;43(W1):W174–81.
84. Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, Perrimon N, et al. An integrative approach to ortholog prediction for disease-focused and other functional studies. BMC Bioinform. 2011;12(1):357.
85. Moreno-Hagelsieb G, Latimer K. Choosing BLAST options for better detection of orthologs as reciprocal best hits. Bioinformatics (Oxford, England). 2007;24(3):319–24.
86. Rocha S, Vieira J, Vázquez N, López-Fernández H, Fdez-Riverola F, Reboiro-Jato M, et al. ATXN1 N-terminal region explains the binding differences of wild-type and expanded forms. BMC Med Genomics. 2019;12(1):145.

87. Maddison DR, Schulz K-S, Maddison WP. The Tree of Life Web Project*. Zootaxa. 2007;1668(1):19–40.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC

2.1.1. Supplementary material

The evolution of vitamin C biosynthesis and transport in animals

Pedro Duque, Cristina P. Vieira, Bárbara Bastos, Jorge Vieira

Additional tables

Additional file 1: Table S1 – *Regucalcin* and *Dca D. melanogaster* RNAi crosses progeny count.
Additional file 1: Table S2 – Conserved *Regucalcin* interactors between *D. melanogaster* and *H. sapiens*/*M. musculus*.
Additional file 1: Table S3 – PSS located outside of the active site and lid regions relative to the dataset's reference protein structures.
Additional file 1: Table S4 – Retrieved *D. melanogaster* SVCT interactome and corresponding orthologs in *H. sapiens*, as well as the current SVCT1, SVCT2 and SVCT3 *H. sapiens* interactomes.
Additional file 1: Table S5 – *Regucalcin* and *SVCT* sequences removed during dataset processing.

Additional figures

Additional file 1: Fig. S1 – Non-Bilaterian *Regucalcin* original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The outgroup sequence is not highlighted and can be seen at the bottom of the phylogeny. The proposed gene clusters after manual curation can be observed in B.
Additional file 1: Fig. S2 – Hemipteroid Assemblage and Blattodea *Regucalcin* original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The general outgroup sequence is not highlighted and can be seen at the bottom of the phylogeny, while the Remaining Insecta representative sequence is marked in light purple. The proposed gene clusters after manual curation can be observed in B.
Additional file 1: Fig. S3 – Coleoptera *Regucalcin* original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The outgroup sequence is not highlighted and can be seen at the bottom of the phylogeny The proposed gene clusters after manual curation can be observed in B.
Additional file 1: Fig. S4 – Hymenoptera *Regucalcin* original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The outgroup sequence is not highlighted and can be seen at the bottom of the phylogeny The proposed gene clusters after manual curation can be observed in B.
Additional file 1: Fig. S5 – Diptera *Regucalcin* original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The outgroup sequence is not highlighted and can be seen at the bottom of the phylogeny The proposed gene clusters after manual curation can be observed in B.
Additional file 1: Fig. S6 – Lepidoptera *Regucalcin* original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The outgroup sequence is not highlighted and can be seen at the bottom of the phylogeny The proposed gene clusters after manual curation can be observed in B.
Additional file 1: Fig. S7 – Proposed *Regucalcin* evolutionary scenario across the Insecta. *Regucalcin* presence is highlighted in green, absence in red and uncertainty regarding gene loss/presence in light blue. Gene duplications affecting a single species of a genus are marked with a *, while those which affected two or more species from the same genus are marked with a #. The number of species present in the final dataset and total number of species analysed for each taxonomic group can be seen within parentheses, respectively. Taxonomic relationships are depicted as seen in literature [1-10].
Additional file 1: Fig. S8 – Non-Lophotrochozoa *Regucalcin* original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The general outgroup sequence is not highlighted and can be seen at the bottom of the phylogeny, while the Insecta representative sequence is marked in light pink. The proposed gene clusters after manual curation can be observed in B.
Additional file 1: Fig. S9 – Lophotrochozoa *Regucalcin* original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The outgroup sequence is not highlighted and can be seen at the bottom of the phylogeny The proposed gene clusters after manual curation can be observed in B.

Additional file 1: Fig. S10 – Basal deuterostomians *Regucalcin* original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The outgroup sequence is not highlighted and can be seen at the bottom of the phylogeny The proposed gene clusters after manual curation can be observed in B.

Additional file 1: Fig. S11 – Deuterostomians *Regucalcin* original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The outgroup sequence is not highlighted and can be seen at the top of the phylogeny The proposed gene clusters after manual curation can be observed in B.

Additional file 1: Fig. S12 – Distribution of predicted interacting residues on the *D. melanogaster* Regucalcin AAN09306.2 (A) and Dca AGB95961.1 (B) relative to the 14-3-3 protein epsilon (NP_033562.3), Histidine triad nucleotide-binding protein 1 (NP_005331.1) and Superoxide dismutase 1 (NP_000445.1) models. The Regucalcin and Dca predicted structutes are highlighted in cyan, while the remaining protein models are represented in green. The interacting residues can be seen in red. The represented interaction models are the ones with the highest Regucalcin and Dca interacting residues within the lowest Z-score cluster provided by HADDOCK. For comparison purposes, all the Regucalcin and Dca structures have the lid region on the top portion of the panels. This figure allows the visualization of the distinct overall interaction surfaces concerning both proteins.

Additional file 1: Fig. S13 – Distribution of predicted interacting residues on the human Regucalcin (RCSB PDB accession number 3G4E) relative to the human 14-3-3 protein epsilon (3UAL), Histidine triad nucleotide-binding protein 1 (6B42) and Superoxide dismutase 1 (6FLH) crystal structures. The Regucalcin structute is highlighted in cyan, while the remaining proteins are represented in green. The interacting residues can be seen in red. The represented interaction models are the ones with the highest Regucalcin interacting residues within the lowest Z-score cluster provided by HADDOCK. For comparison purposes, the Regucalcin structure has the lid region on the top portion of the panels. This figure allows the visualization of an interaction surface overall very similiar to the one observed in the *D. melanogaster* Regucalcin (AAN09306.2), indicating that the distinct interacting surface that emerged in the *D. melanogaster* Dca (AGB95961.1) is likely the consequence of gene neofunctionalization processes after duplication.

Additional file 1: Fig. S14 – Non-bilaterian *Sodium-dependent Vitamin C transporter (SVCTNB)* original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The three outgroup sequences are not highlighted and can be seen at the top of the phylogeny. The proposed gene clusters after manual curation can be observed in B.

Additional file 1: Fig. S15 – Non-Lophotrochozoa *Sodium-dependent Vitamin C transporter (SVCTP NL)* original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The three outgroup sequences are not highlighted and can be seen at the bottom of the phylogeny. The proposed gene clusters after manual curation can be observed in B.

Additional file 1: Fig. S16 – Lophotrochozoa *Sodium-dependent Vitamin C transporter (SVCTP L)* original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The three outgroup sequences are not highlighted and can be seen at the bottom of the phylogeny. The proposed gene clusters after manual curation can be observed in B.

Additional file 1: Fig. S17 – Basal deuterostomian *Sodium-dependent Vitamin C transporter (SVCT BD)* original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The seven outgroup sequences are not highlighted and can be seen spread across the phylogeny. The proposed gene clusters after manual curation can be observed in B.

Additional file 1: Fig. S18 – *Sodium-dependent Vitamin C transporter 1* (*SVCT1*) original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The four outgroup sequences are not highlighted and can be seen at the top of the phylogeny. The proposed gene clusters after manual curation can be observed in B.

Additional file 1: Fig. S19 – *Sodium-dependent Vitamin C transporter 2* (*SVCT2*) original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The four outgroup sequences are not highlighted, and can be seen at the top and bottom of the phylogeny. The proposed gene clusters after manual curation can be observed in B.

Additional file 1: Fig. S20 – *Sodium-dependent Vitamin C transporter 3* (*SVCT3*) original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The four outgroup sequences are not highlighted, and can be seen at the top and bottom of the phylogeny. The proposed gene clusters after manual curation can be observed in B.

Additional file 1: Fig. S21 – *Sodium-dependent Vitamin C transporter 4* (*SVCT4*) original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The four outgroup sequences are not highlighted, and can be seen at the top of the phylogeny.. The proposed gene clusters after manual curation can be observed in B.
Additional file 1: Fig. S22 – Putative *Sodium-dependent Vitamin C transporter 5* (*SVCT5*) original (A) and collapsed (B) Bayesian phylogenies. In A, the species relevant taxonomic groups are represented next to the corresponding sequences and the different taxonomic groups are highlighted with distinct colours, following the colour scheme presented below the consensus tree. The 16 outgroup sequences are not highlighted and can be seen at the top of the phylogeny. The proposed gene clusters after manual curation can be observed in B.

References

1. Maddison DR, Schulz K-S, Maddison WP. The Tree of Life Web Project*. Zootaxa. 2007;1668(1):19-40.
2. van der Linde K, Houle D, Spicer GS, Steppan SJ. A supermatrix-based molecular phylogeny of the family Drosophilidae. Genet Res (Camb). 2010;92(1):25-38.
3. Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim J-W, Lambkin C, et al. Episodic radiations in the fly tree of life. Proceedings of the National Academy of Sciences. 2011;108(14):5690.
4. Song N, Liang A-P, Bu C-P. A Molecular Phylogeny of Hemiptera Inferred from Mitochondrial Genome Sequences. PloS one. 2012;7(11):e48778.
5. Song F, Li H, Jiang P, Zhou X, Liu J, Sun C, et al. Capturing the Phylogeny of Holometabola with Mitochondrial Genome Data and Bayesian Site-Heterogeneous Mixture Models. Genome Biol Evol. 2016;8(5):1411-26.
6. Li H, Leavengood JM, Jr., Chapman EG, Burkhardt D, Song F, Jiang P, et al. Mitochondrial phylogenomics of Hemiptera reveals adaptive innovations driving the diversification of true bugs. Proceedings Biological sciences. 2017;284(1862):20171223.
7. Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, et al. Evolutionary History of the Hymenoptera. Current biology : CB. 2017;27(7):1013-8.
8. Johnson Kevin P, Dietrich Christopher H, Friedrich F, Beutel Rolf G, Wipfler B, Peters Ralph S, et al. Phylogenomics and the evolution of hemipteroid insects. Proceedings of the National Academy of Sciences. 2018;115(50):12775-80.
9. O'Grady PM, DeSalle R. Phylogeny of the Genus Drosophila. Genetics. 2018;209(1):1-25.
10. Zhang S-Q, Che L-H, Li Y, Dan L, Pang H, Ślipiński A, et al. Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. Nature Communications. 2018;9(1):205.

Additional file 1: Table S1

Crossing	Straight ♀	Curly ♀	Straight ♂	Curly ♂	X^2 ♀	X^2 ♂
RGN (#105509) ♂ x *Act5C/GAL4* (#25374) ♀	40	445	11	454	338,19 (P<0,001)	422,04 (P<0,001)
RGN (#105509) ♀ x *Act5C/GAL4* (#25374) ♂	81	407	3	355	217,78 (P<0,001)	346,10 (P<0,001)
Dca (#103377) ♂ x *Act5C/GAL4* (#25374) ♀	70	324	16	303	163,65 (P<0,001)	258,21 (P<0,001)
Dca (#103377) ♀ x *Act5C/GAL4* (#25374) ♂	56	194	25	159	76,176 (P<0,001)	97,59 (P<0,001)

Additional file 1: Table S2

D. melanogaster **Regucalcin interactors**	**Interactor homologs (*H. sapiens*)**	**Interactor homologs (*M. musculus*)**
33471	3094	n.a.
39251	6647	n.a.
39476	3336	n.a.
39505	9948	n.a.
42186	n.a.	22627

Additional file 1: Table S3

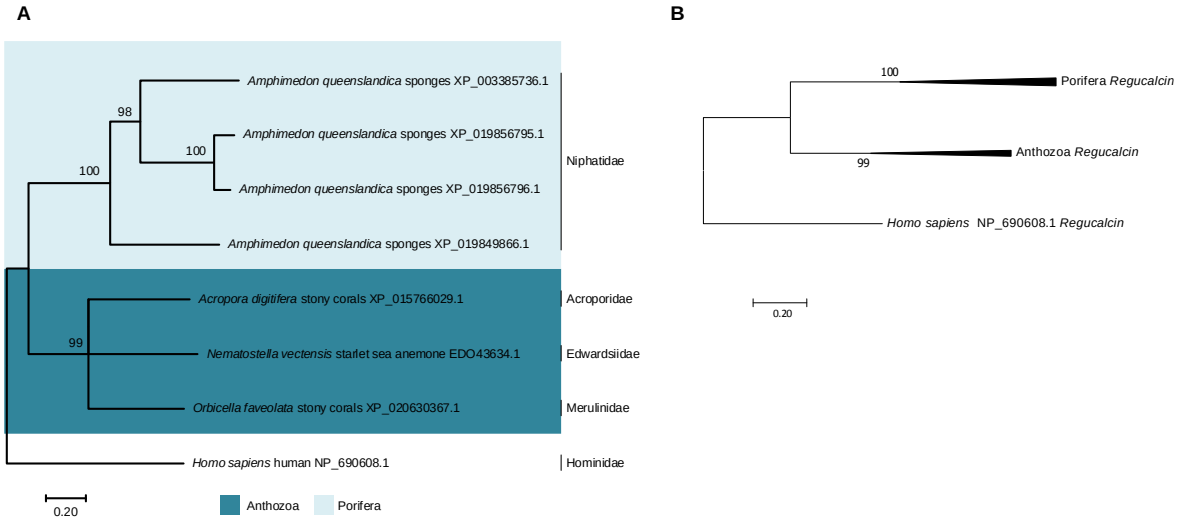
Dataset	Protein reference PSS
Sophophora Regucalcin	I^{8}, A^{9}
Sophophora Dca	L^{108}, S^{157}, S^{197}
Apoidea Hy2	T^{47}, A^{91}, L^{149}
Formicoidea Hy1	A^{17}, V^{100}, R^{250}, T^{312}
Lepidoptera L2	Q^{305}
Teleost without Salmonidae and Cyprinidae	A^{181}
Reptilia R1	D^{189}
Aves R1	V^{46}
Aves R2	Y^{14}, R^{15}, A^{45}, G^{47}, S^{86}, A^{91}, K^{195}, V^{281}
Mammalia R2	Q^{48}, A^{87}, N^{95}

Additional file 1: Table S4

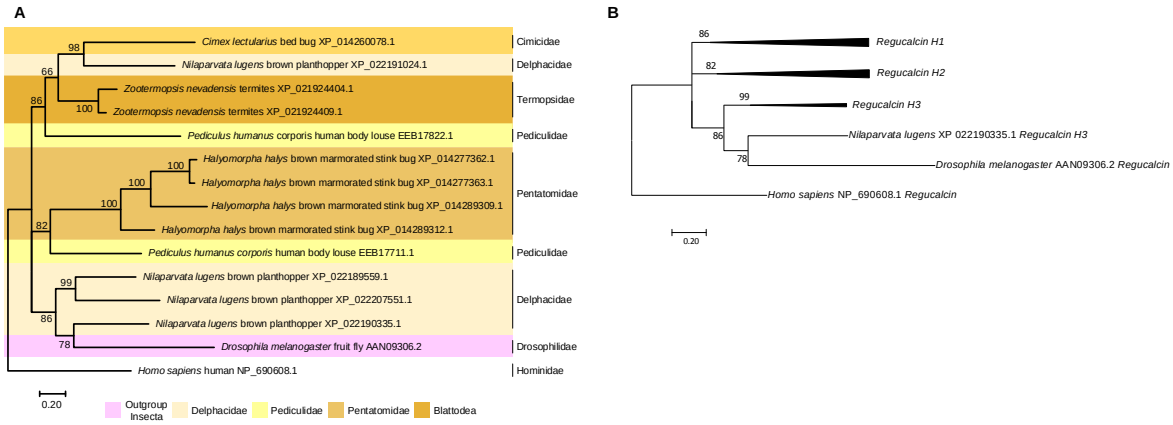
D. melanogaster SVCT Interactome	*H. sapiens* orthologs	SVCT1 Interactome	SVCT2 Interactome	SVCT3 Interactome
32179	No ortholog retrived (BLAST + DIOPT)	2885	4734	3480
32630	9972	3198	7316	5625
32851	129401	3881	10555	
34292	Low confidence hits	4223	27065	
35226	7274	4690	55238	
35305	6741	5295	55744	
35664	161424	5335	79669	
36491	55755	10768	79956	
37637	54940	23281	84522	
38045	27044	30008	85025	
39382	65095	54507	254428	
40233	8755	81851		
40532	No ortholog retrived (BLAST + DIOPT)	83900		
40806	152007	85291		
41270	56915	125115		
41360	1743	375791		
42453	1017	386675		
43057	No ortholog retrived (BLAST + DIOPT)	386676		
2768852	4140	386677		
		386680		
		386681		
		386682		
		386683		
		388677		
		653240		

Additional file 1: Table S5

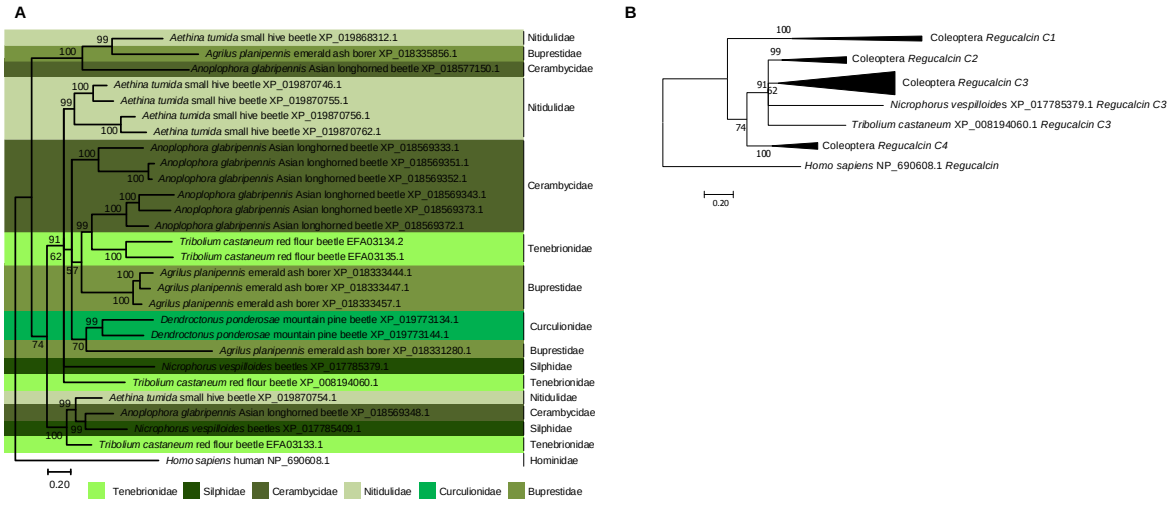
Isoforms removed		Redundant sequences from different species		Problematic Sequences Removed	
Repeatkin	**SVCT**	**Repeatkin**	**Repeatkin**	**Repeatkin**	**SVCT**
Pogona_vitticeps_XP_020552492.1	Acanthaster_planci_(crown-of-thorns_starfish)_Acanthasteridae_Asteroidea_XP_022093025.1	Haliaeetus_albicilla_(white-tailed_eagle)_Accipitridae_XP_009929557.1 replaces the following sequences:	Bactrocera_latifrons_XP_018802379.1	Acanthisitta_chloris_chloris_(Rifleman)_Acanthisittidae_Aves_XP_009079854.1	
Mesocricetus_auratus_(golden_hamster)_Cricetidae_Mammalia_XP_005072214.2	Actinotyga_jabusa_(cheetah)_Felidae_Mammalia_XP_014028547.1/XP_014028548.1	Haliaeetus_leucocephalus_(bald_eagle)_Accipitridae_XP_010578757.1	Bemisia_tabaci_XP_018902746.1	Aedes_aegypti_yellow_fever_mosquito_Culicidae_Insecta_EJY57974.1	
Tupaia_chinensis_ELV14210.1	Aldrovanda_melanoleuca_(giant_panda)_Ursidae_Mammalia_XP_011233547.1/XP_019661063.1	Haliaeetus_albicilla_(white-tailed_eagle)_Accipitridae_KFQ05724.1	Branchiostoma_floridae_EEN58249.1	Aedes_albopictus_Asian_tiger_mosquito_Culicidae_Insecta_XP_019543343.1	
Myotis_brandtii_XP_014398972.1	Alligator_mississippiensis_(American_alligator)_Alligatoridae_KYO34297.1/XP_019337984.1/XP_019346970.1		Gallus_gallus_XP_015156358.1	Anolis_carolinensis_green_anole_Dactyloidae_XP_008108293.1	
Fukomys_damilis_XP_021956085.1	Anolis_carolinensis_(green_anole)_Dactyloidae_XP_008109011.1	Corvus_brachyrhynchos_(American_crow)_Corvidae_XP_008627501.1 replaces the following sequences:	Nilaparvata_lugens_XP_022187325.1	Aplysia_californica_California_sea_hare_Aplysiidae_Gastropoda_XP_005098904.1	
Alligator_sinensis_(Chinese_alligator)_Alligatoridae_XP_006032553.1	Aotus_nancymaae_(Ma's_night_monkey)_Aotidae_Mammalia_XP_012301315.1/XP_012325553.1	Corvus_brachyrhynchos_(American_crow)_Corvidae_XP_017584065.1	Nilaparvata_lugens_XP_022203265.1	Aplysia_californica_California_sea_hare_Aplysiidae_Gastropoda_XP_005098905.2	
Cricetulus_griseus_XP_007633712.1/XP_007629332.1	Balaenoptera_acutorostrata_scammoni_(minke_whale)_Balaenopteridae_Mammalia_XP_007178911.1/XP_007188812.1/XP_007191768.1	Corvus_cornix_cornix_(hooded_crow)_Corvidae_XP_019138371.1	Nilaparvata_lugens_XP_022203265.1	Apostichopus_japonicus_Japanese_sea_cucumber_Stichopodidae_Holothuroidea_FBK43178.1	
Phascolarctos_cinereus_XP_020849583.1	Bison_bison_bison_(American_bison)_Bovidae_Mammalia_XP_010848922.1/XP_010859518.1/XP_010859519.1	Corvus_cornix_cornix_(hooded_crow)_Corvidae_XP_019138372.1	Odontopleura_muticulana_luteib_XP_015844817.1	Atta_colombica_ants_Formicidae_Insecta_KYM76322.1	
Charadrius_vociferus_XP_009880409.1	Bos_indicus_(zebu_cattle)_Bovidae_Mammalia_XP_019820577.1/XP_019820578.1	Corvus_cornix_cornix_(hooded_crow)_Corvidae_XP_019138373.1	Pogona_vitticeps_XP_020552480.1	Balearica_regulorum_gibbericeps_East_African_grey_crowned_crane_Gruidae_Aves_XP_010292977.1	
Opisthocomus_hoazin_XP_009929456.1/XP_009929435.1	Bos_mutus_(wild_yak)_Bovidae_Mammalia_XP_005897741.1	Corvus_brachyrhynchos_(American_crow)_Corvidae_KFO35830.1	Trichonephila_clavipes_mRNA_ONT_ONT_ONT



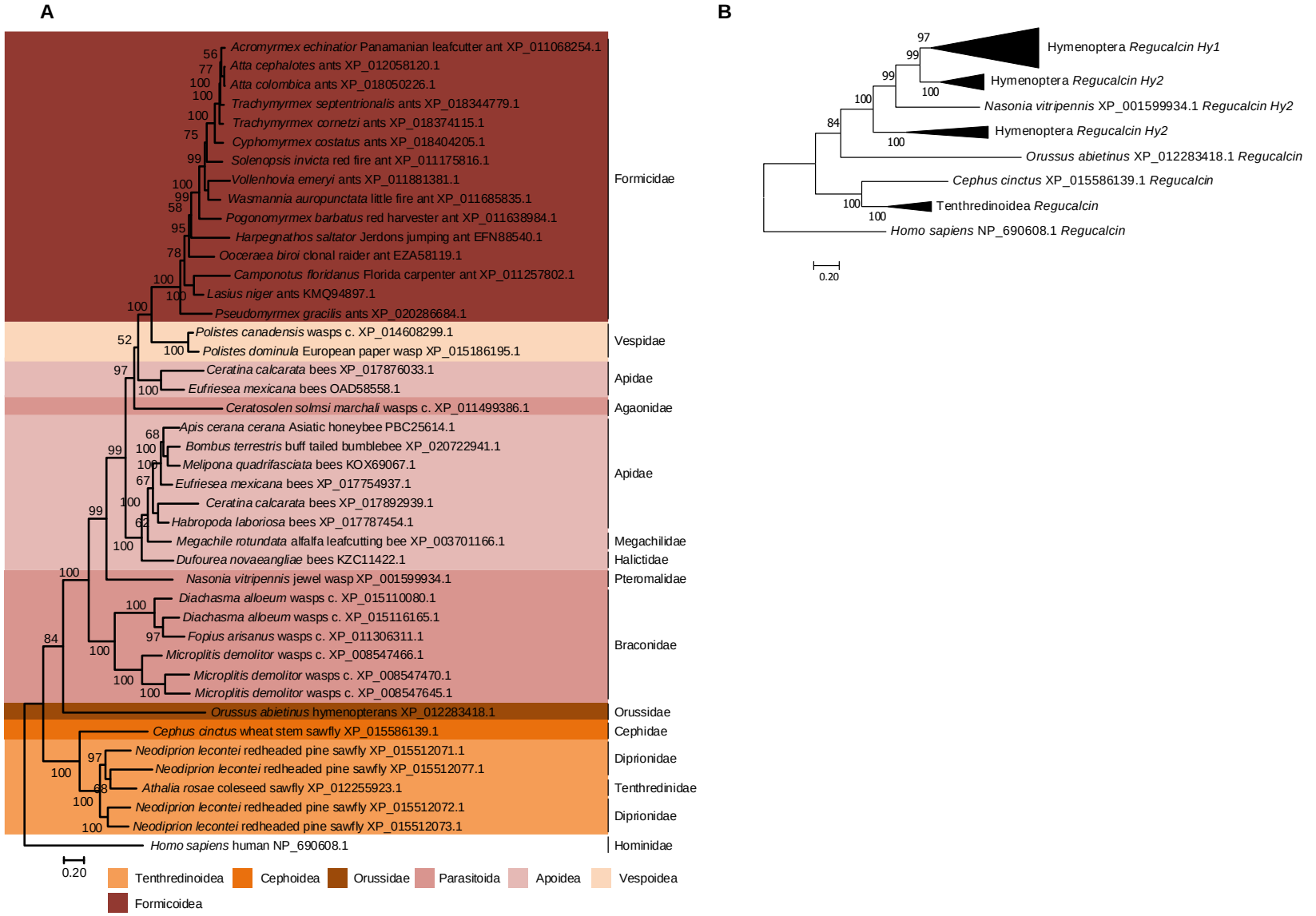
Additional file 1: Fig. S1



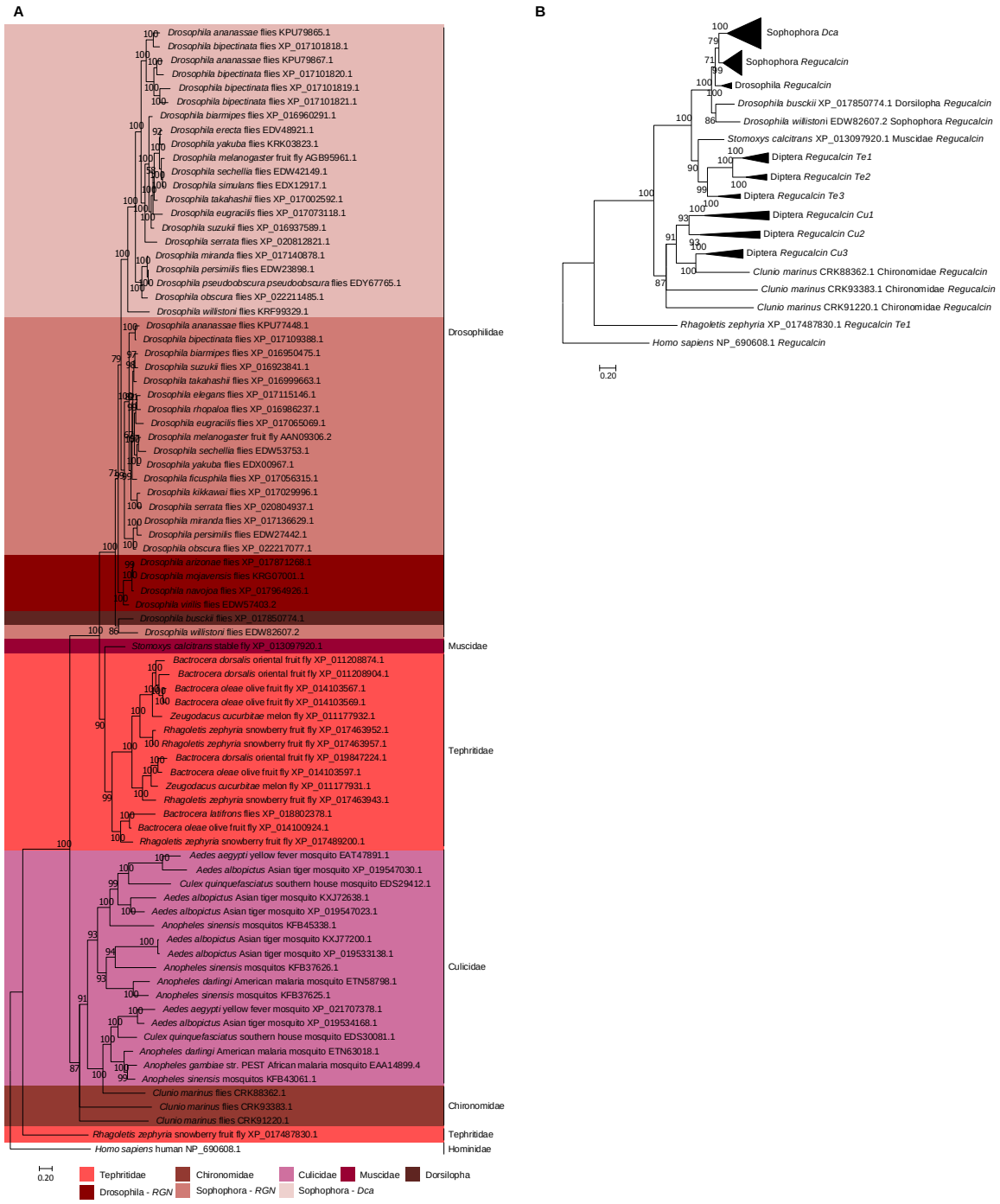
Additional file 1: Fig. S2



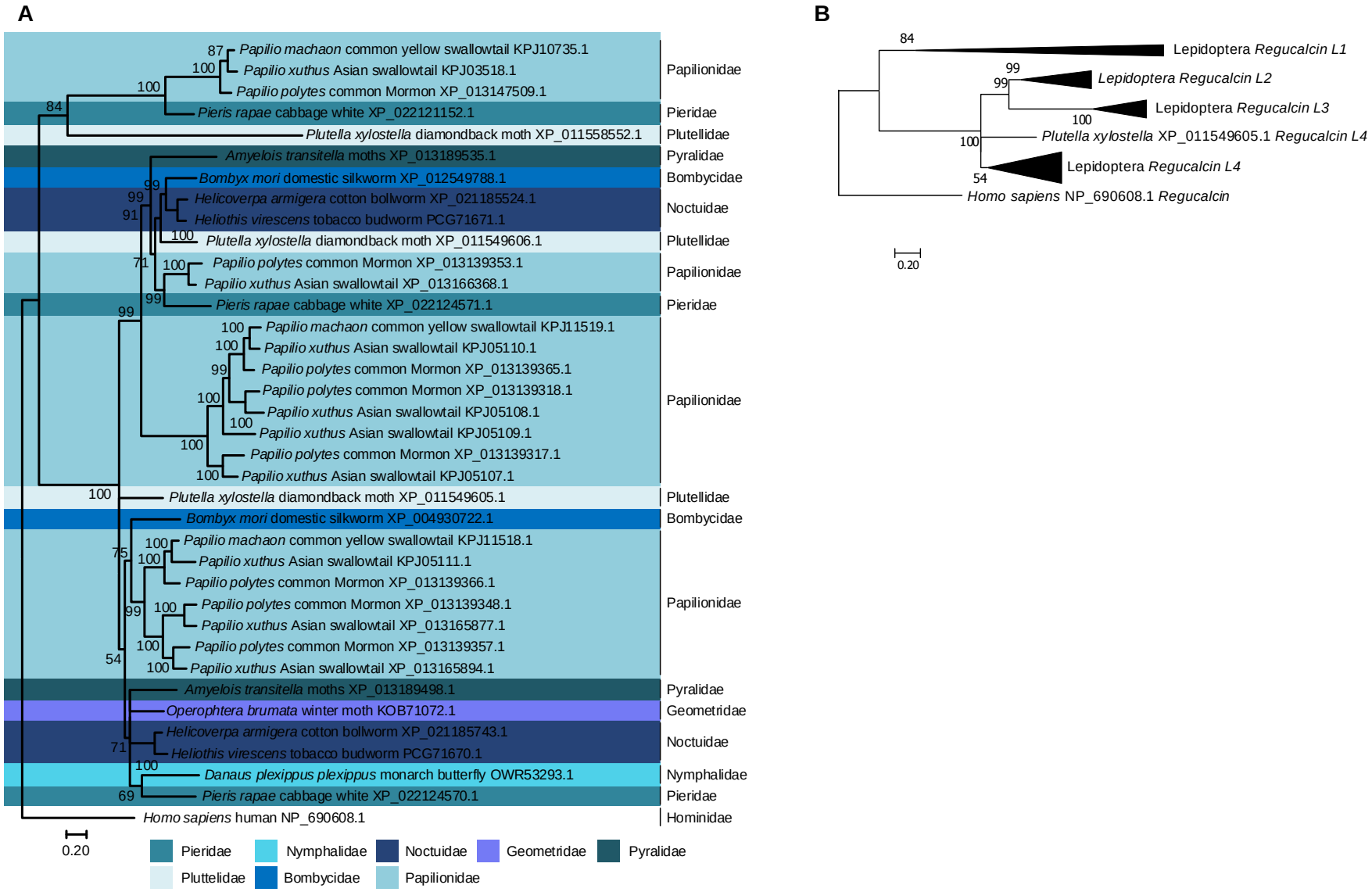
Additional file 1: Fig. S3



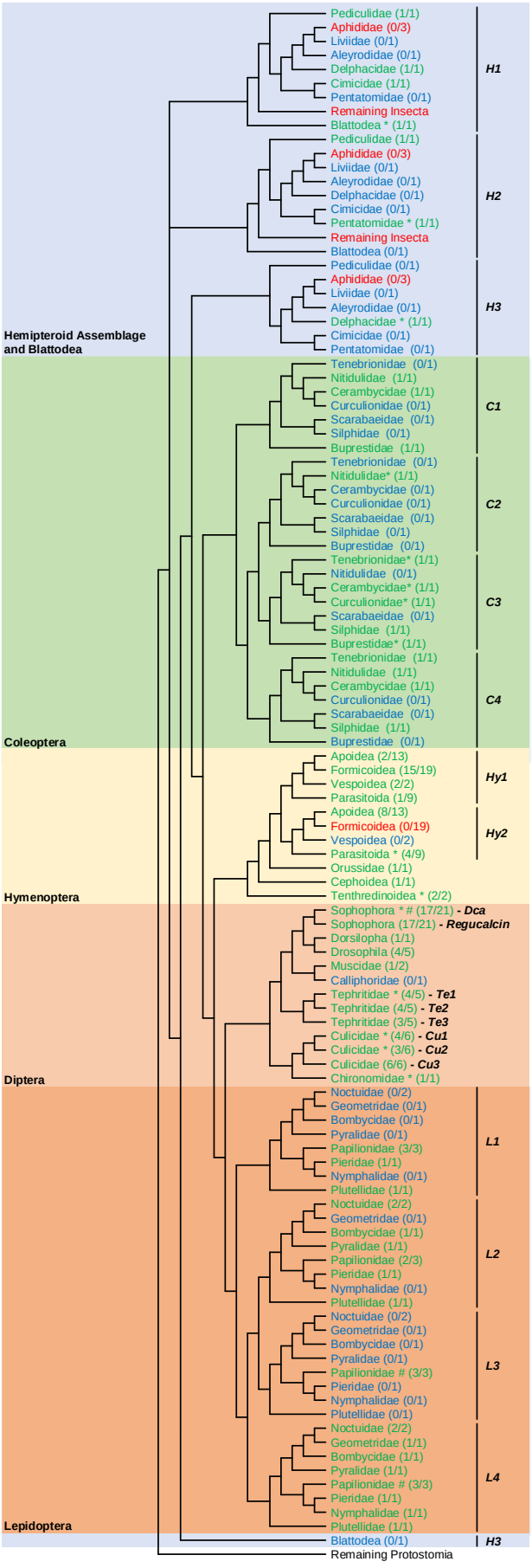
Additional file 1: Fig. S4



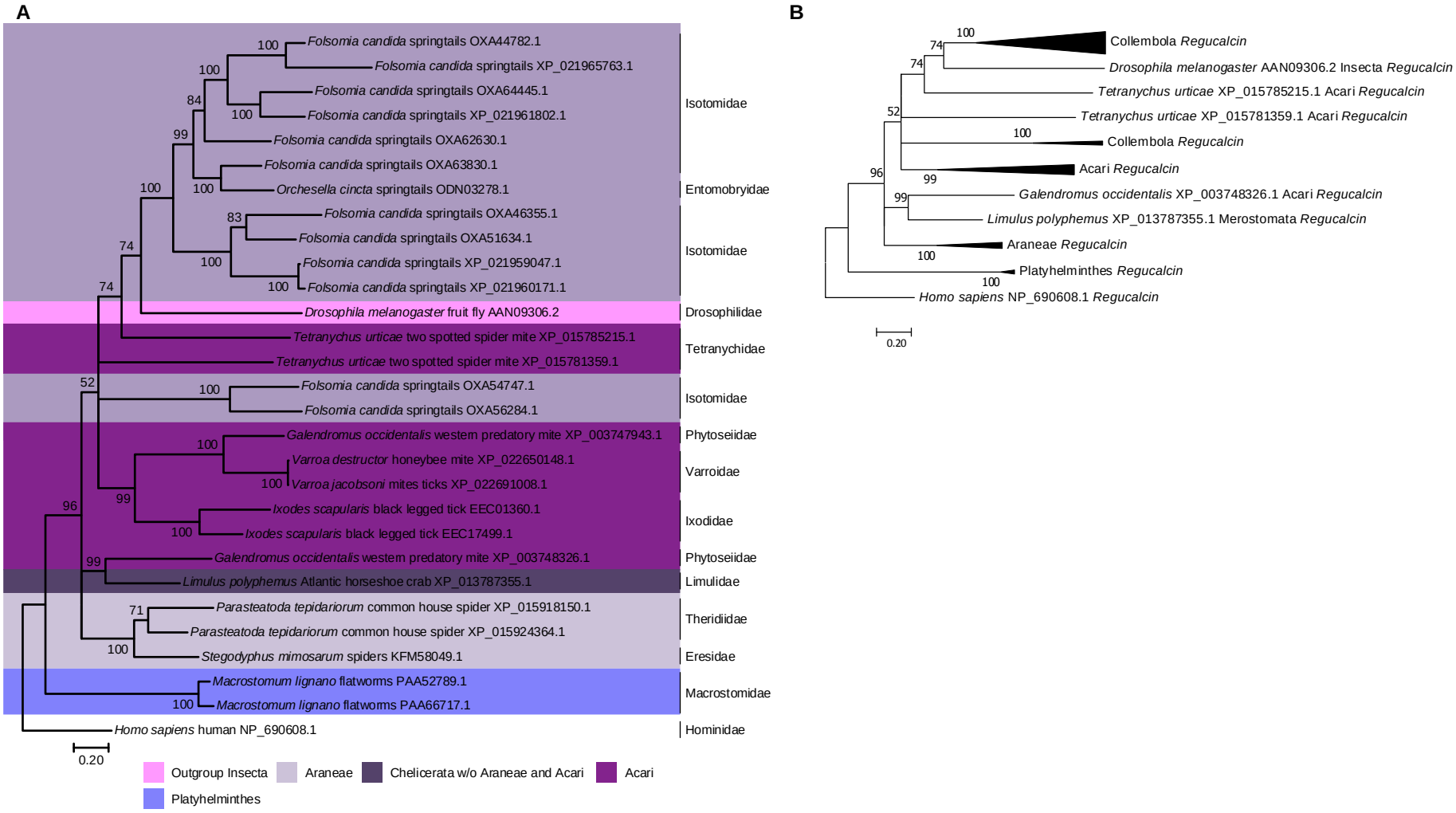
Additional file 1: Fig. S5



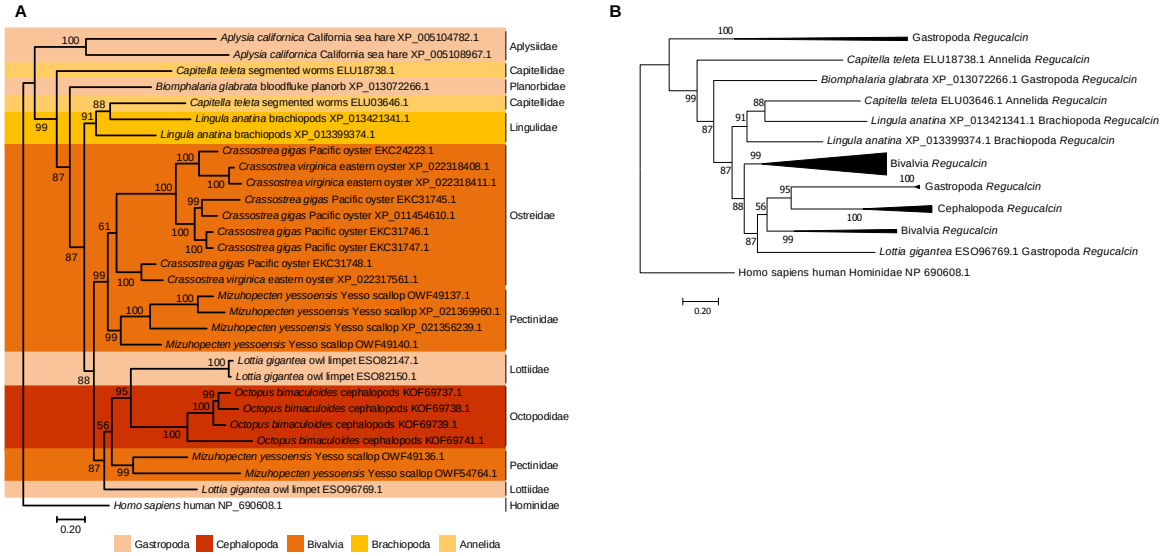
Additional file 1: Fig. S6



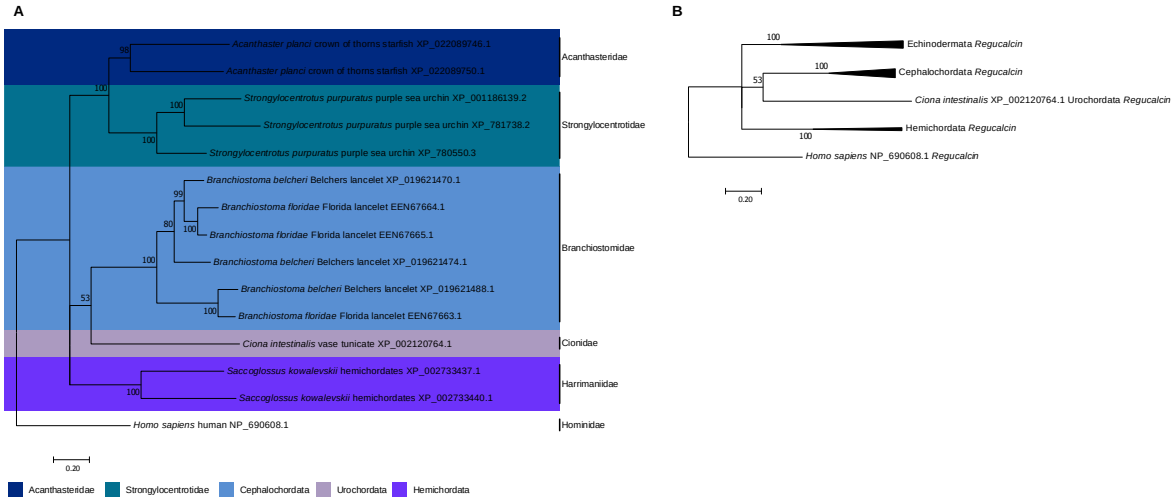
Additional file 1: Fig. S7



Additional file 1: Fig. S8



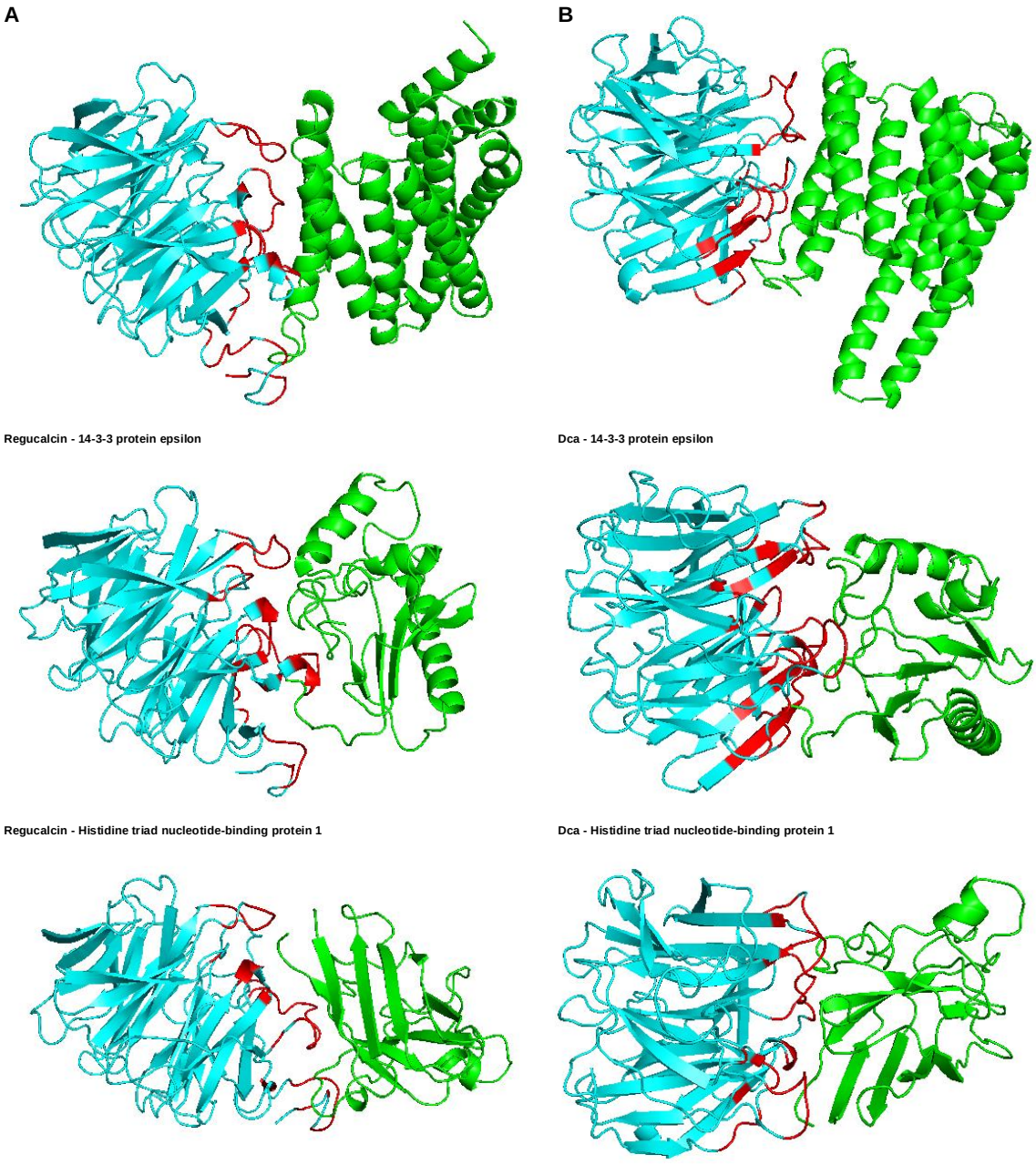
Additional file 1: Fig. S9



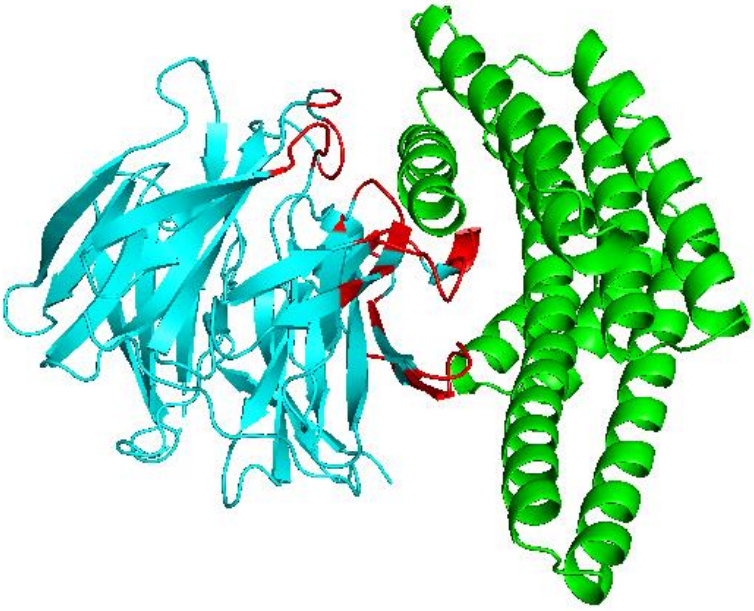
Additional file 1: Fig. S10



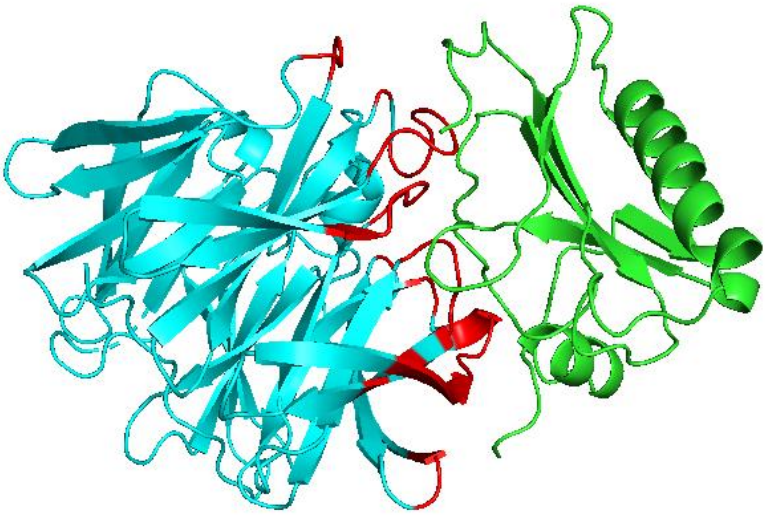
Additional file 1: Fig. S11



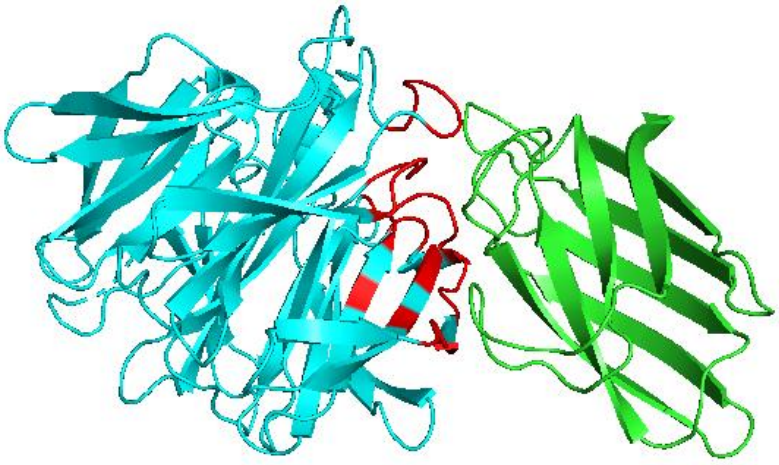
Additional file 1: Fig. S12



Regucalcin - 14-3-3 protein epsilon (Human)

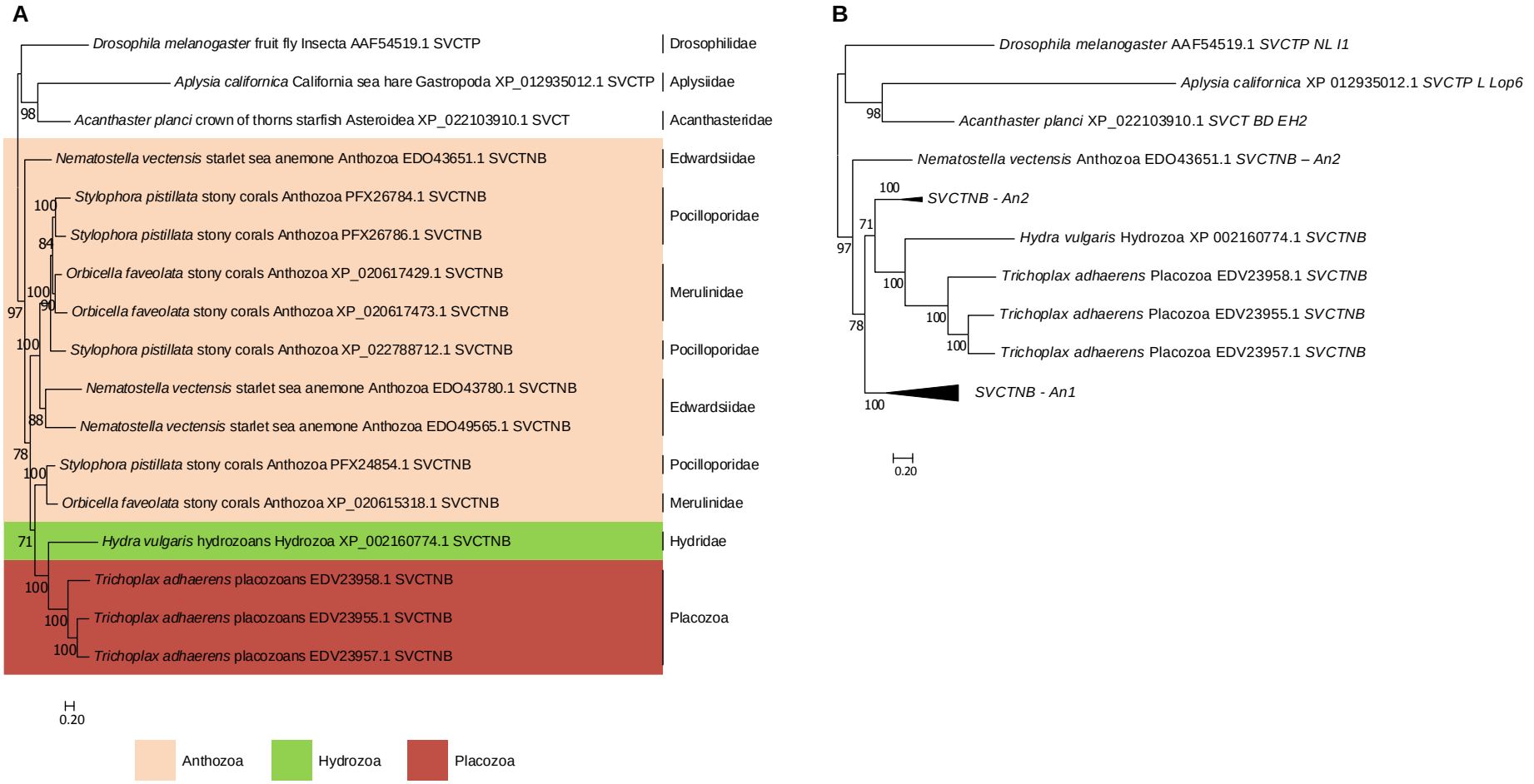


Regucalcin - Histidine triad nucleotide-binding protein 1 (Human)

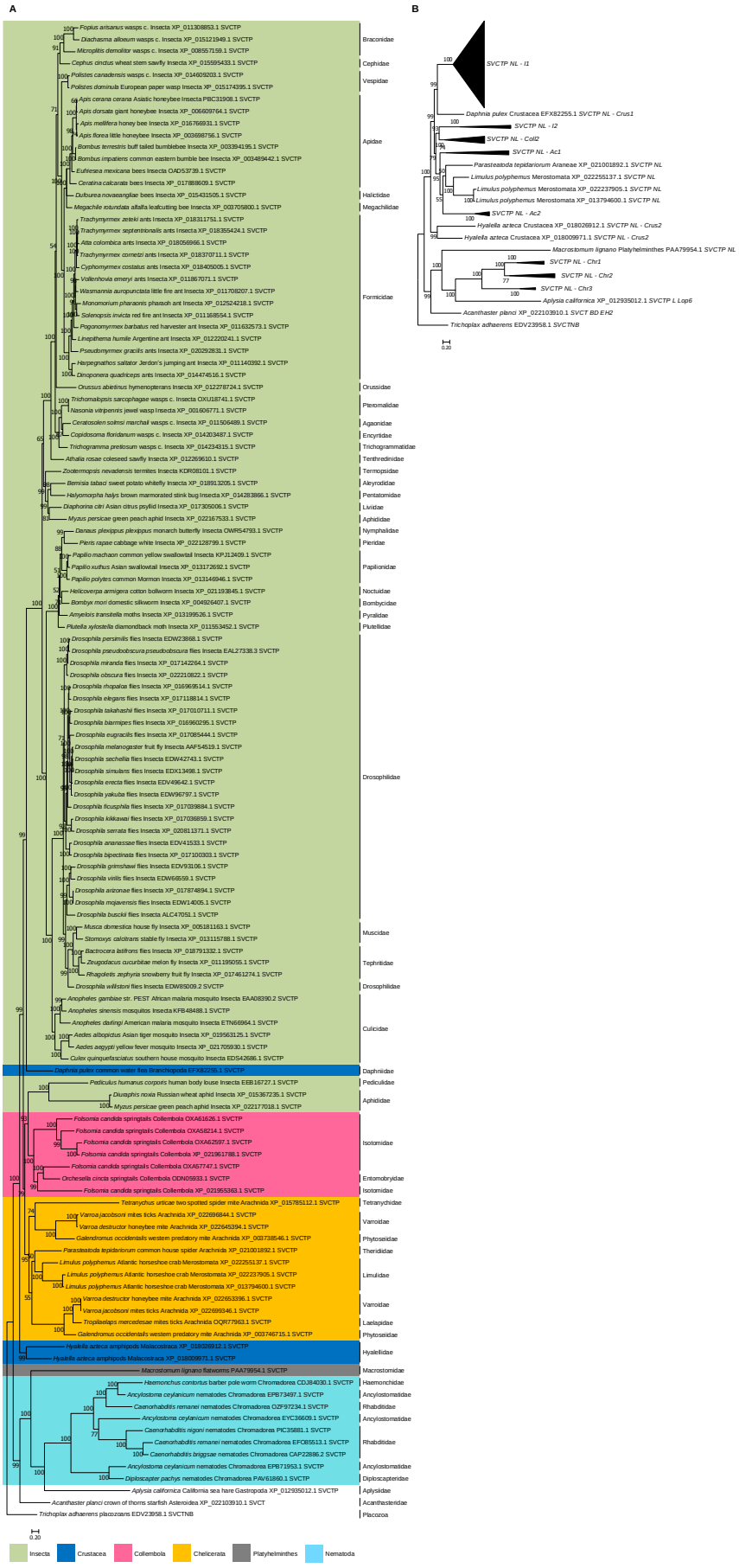


Regucalcin - Superoxide dismutase 1 (Human)

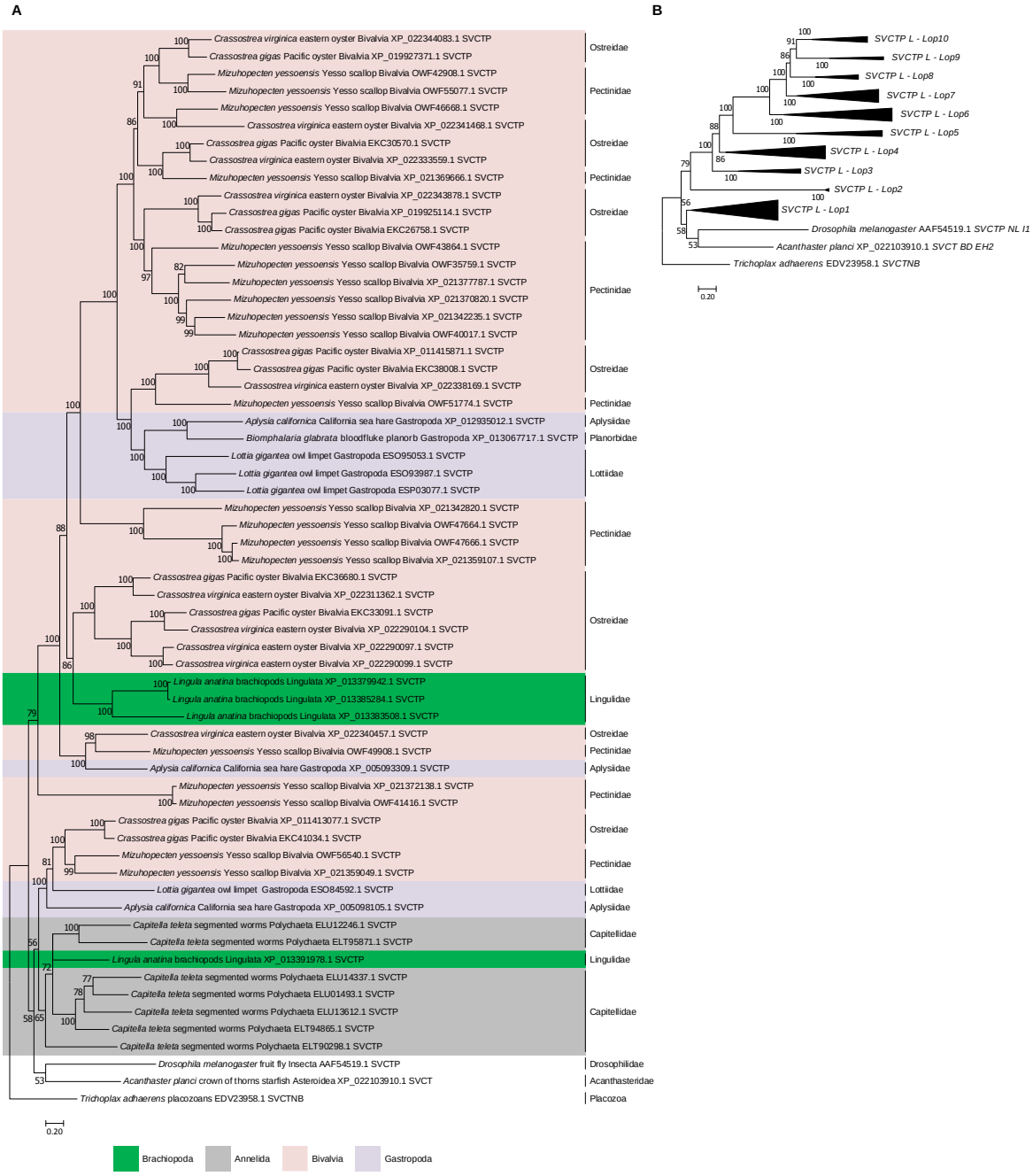
Additional file 1: Fig. S13



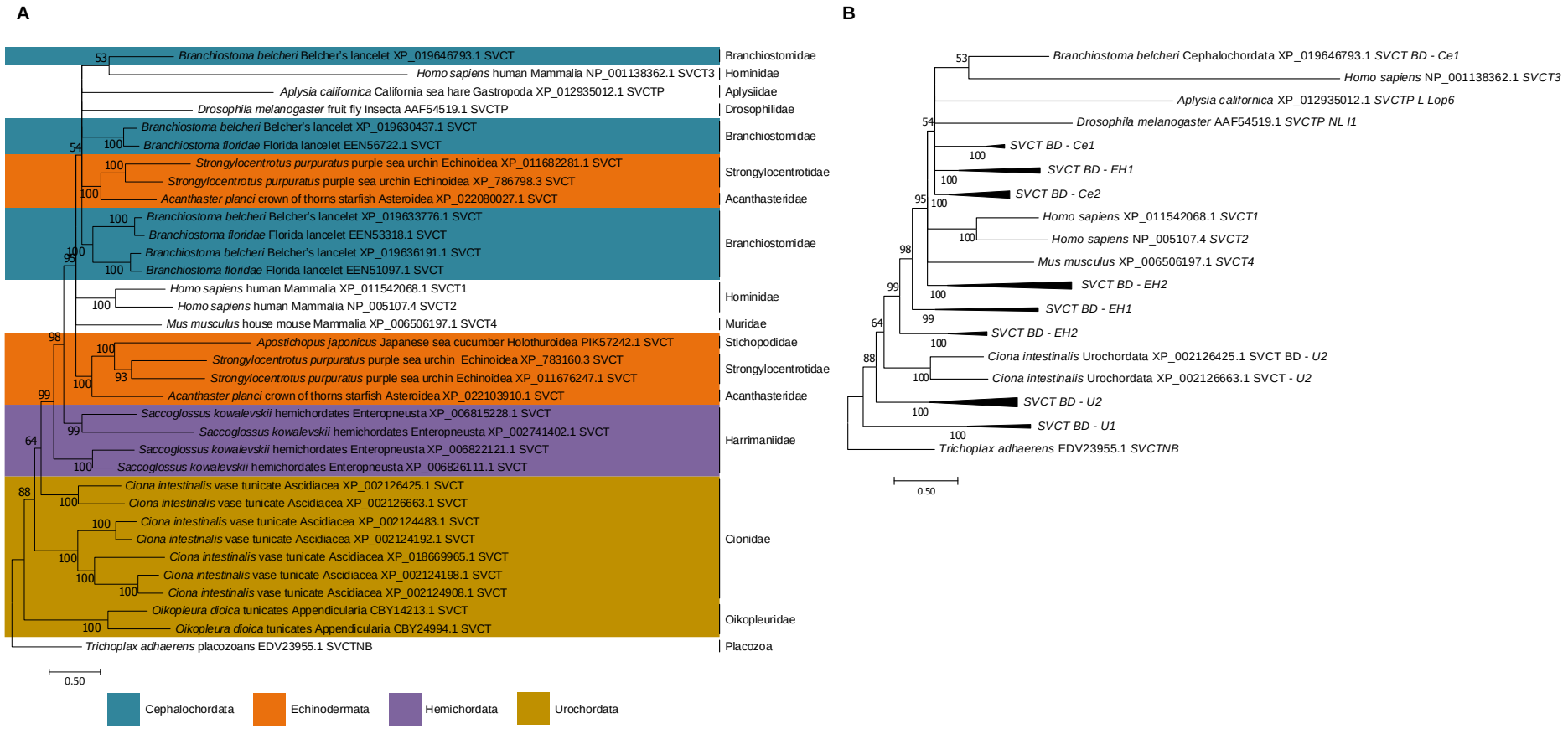
Additional file 1: Fig. S14



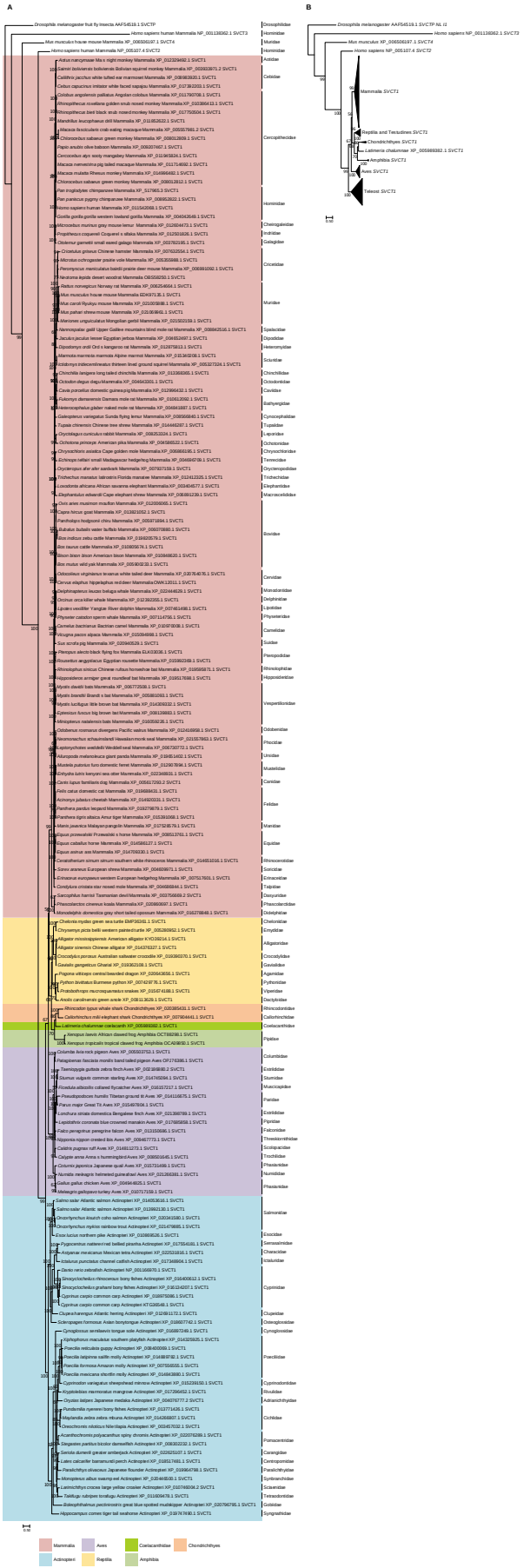
Additional file 1: Fig. S15



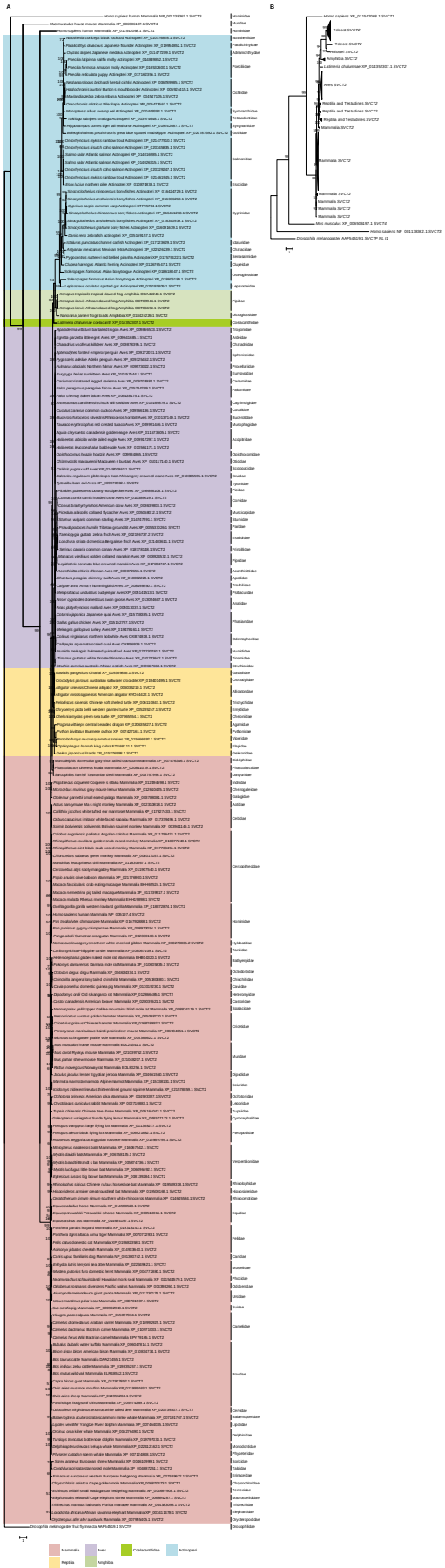
Additional file 1: Fig. S16



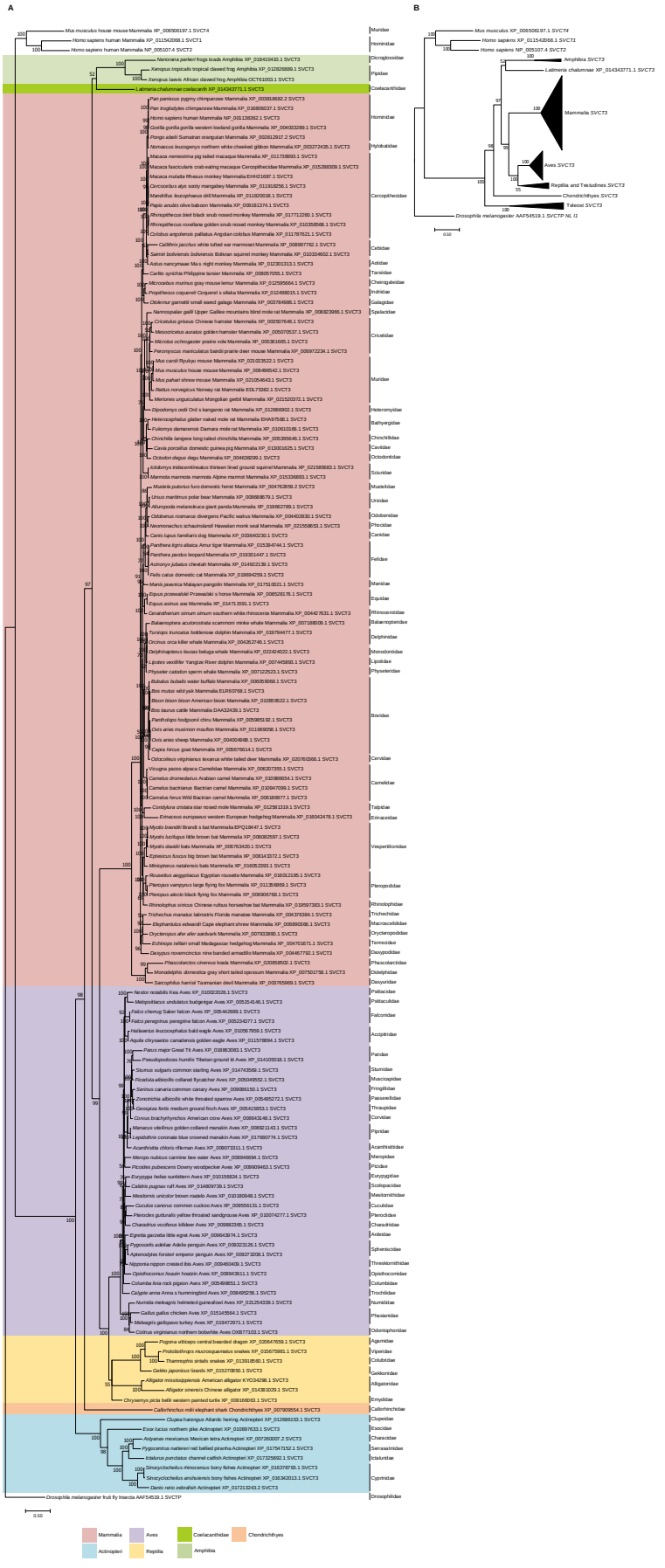
Additional file 1: Fig. S17



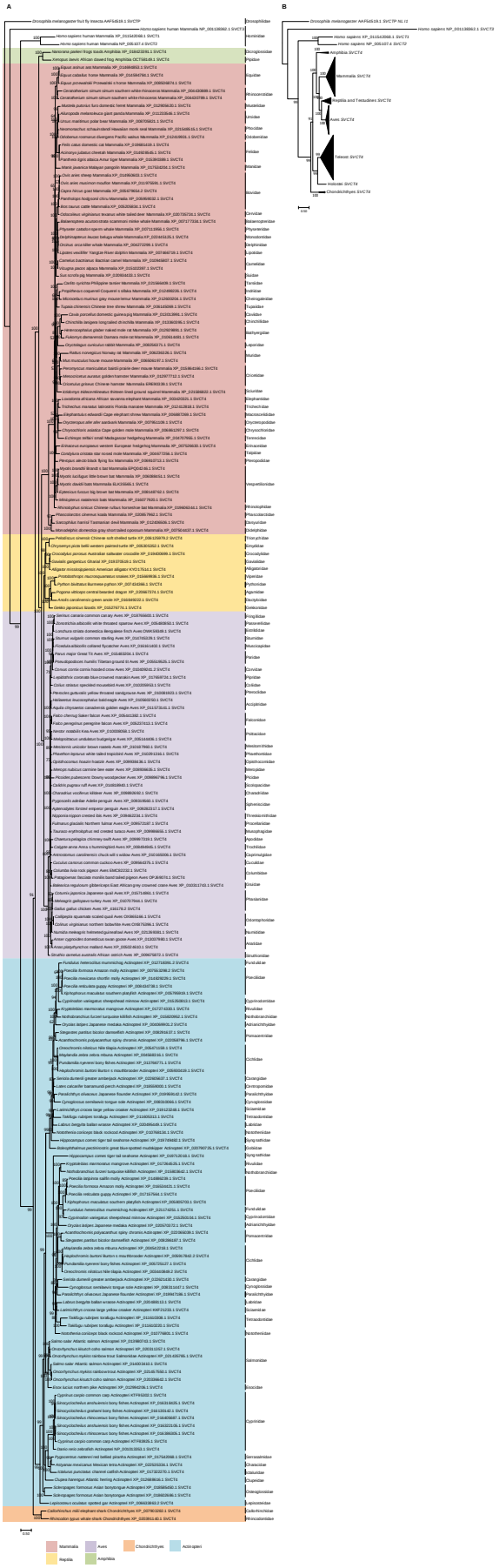
Additional file 1: Fig. S18



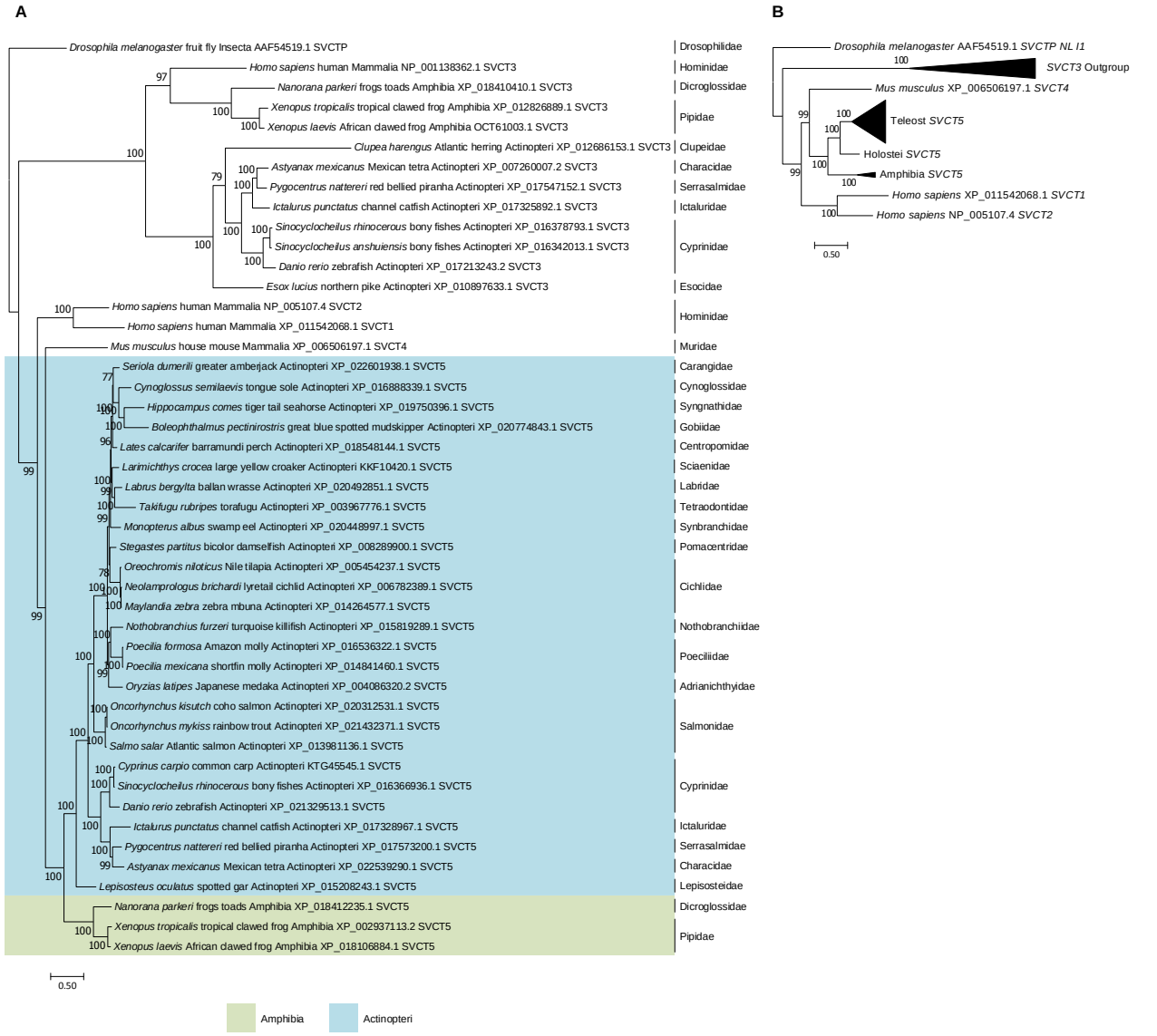
Additional file 1: Fig. S19



Additional file 1: Fig. S20



Additional file 1: Fig. S21



Additional file 1: Fig. S22

2.2. Addendum: Further experimental work and considerations in the context of "The evolution of vitamin C biosynthesis and transport in animals" article

2.2.1. Article synopsis

The inference of positively selected sites within several *Regucalcin* gene lineages indicated that *Regucalcin* and its duplicates were subjected to selective pressures throughout evolution in animals (Duque *et al.* 2022). Insects were highly affected by *Regucalcin* duplication events, and the conservation of the resulting gene duplicates was interpreted as an indication of their substantial role in biological adaptation processes within species from this taxonomic group (Duque *et al.* 2022). Bioinformatics analyses suggested that *Regucalcin* homologs are possibly linked to calcium homeostasis and oxidative stress response roles in *D. melanogaster*, and that those functions likely extend to other protostomians (Duque *et al.* 2022). Phylogenetic inferences further hinted that *Regucalcin* homologs could also be related to VC synthesis in these organisms, evidenced by the presence of two *Regucalcin* duplicates in *B. mori* (Duque *et al.* 2022). Nevertheless, the functional aspects of the latter biological role could not be adequately addressed in *D. melanogaster* due to the high pupal lethality associated with *Regucalcin* and *Dca* RNAi-mediated suppression (Duque *et al.* 2022). Thus, further research is still required to verify the roles of Regucalcin and Dca in this model organism, particularly concerning the characterization of the novel VC synthesis pathway. The absence of *Regucalcin* and *GULO* homologs in *C. elegans* despite its capacity for endogenous VC production highlighted the possibility of multi-enzymatic adaptations in the emergence of novel VC synthesis pathways in this species (Duque *et al.* 2022). Nevertheless, how similar evolutionary scenarios could have affected other basal animal lineages remains to be determined.

The lack of sufficient protein-protein interaction data for vertebrate SVCTs limited the efforts to implicate the single SVCT found in *D. melanogaster* in active VC transport (Duque *et al.* 2022). However, the reconciliation of past research with novel phylogenetic data suggested that SVCTs present in basal animals are capable of generalized VC and nucleobase transport, while the vertebrate SVCTs likely specialized in either function after duplications that derived from the inferred 2R-WGD (Duque *et al.* 2022). The absence of GULO-derived VC synthesis in species from basal taxonomic groups seemed strongly linked to the requirement of SVCT-mediated dietary VC uptake mechanisms (Duque *et al.* 2022). Nevertheless, SVCTs appeared to be not functionally important in many nematode species, including *C. elegans*, and coleopterans, suggesting that species from these taxonomic lineages are crucially reliant on endogenous VC synthesis for optimal metabolic function (Duque *et al.* 2022).

2.2.2. VC supplementation of *Regucalcin* and *Dca* RNAi crossings

D. melanogaster Regucalcin and *Dca* RNAi knockdowns have been previously performed to verify if the presence of *Regucalcin* homologs could be associated with VC synthesis in Sophophora species (Duque *et al.* 2022). The employed methodology was ultimately unable to validate this possibility, since the suppression of *Regucalcin* and *Dca* expression led to a lethal phenotype (Duque *et al.* 2022). However, if the lethality was associated with lack of sufficient VC for adequate pupal development, dietary VC supplementation could allow the reversion of this phenotype. Additional crosses were performed in cornmeal food vials with L-Ascorbic acid (50mM) (Sigma-Aldrich, product number A5960) to test this hypothesis. Nonetheless, the viability of flies under *Regucalcin* and *Dca* RNAi-mediated suppression could not be rescued using this experimental setup.

2.2.3. Additional insights regarding the VC synthesis capabilities of *C. elegans*

The evolutionary history of *Regucalcin* depicted in Duque *et al.* (Duque *et al.* 2022) was in accordance with previous literature concerning the loss of this gene in Nematoda species (López-Fernández *et al.* 2019, Patananan *et al.* 2015, Wheeler *et al.* 2015). The presented phylogenetic results raised doubts about the reported *Regucalcin* homolog (*E03H4.3*) previously identified in *C. elegans* by Yabuta *et al.* (Yabuta *et al.* 2020), although this was not discussed in the research paper at the time. Upon additional inspection, the E03H4.3 protein sequence (accession number NP_493146.2) did not show considerable homology to Regucalcin using the *M. musculus* or *R. norvegicus* non-redundant protein sequence databases available at National Center for Biotechnology Information (NCBI) blastp (https://blast.ncbi.nlm.nih.gov/Blast.cgi), opposed to what was stated in Yabuta *et al.* (Yabuta *et al.* 2020). Instead, E03H4.3 had considerable homology with a glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1 for both species. The absence of Regucalcin from the blastp hit table goes in agreement with the information available at the Ensembl database for *E03H4.3* (https://www.ensembl.org/; gene WBGene00008471), since *Regucalcin* is not referred to as a *E03H4.3* vertebrate homolog. Considering this information, it is then possible that VC synthesis in *C. elegans* occurs through an functional adaptation of E03H4.3 to replace the gluconolactonase role of Regucalcin, but E03H4.3 does not appear to be a Regucalcin homolog. This is also likely in respect to the proposed *C. elegans GULO* homolog *F54D5.12* (Yabuta *et al.* 2020), as both blastp results (using the *C. elegans* F54D5.12 query, accession number NP_496465.1) and Ensembl information (gene WBGene00010055) indicate that this gene encodes in fact for a D-2-hydroxyglutarate dehydrogenase. Interestingly, this protein is a α-hydroxy acid dehydrogenase that belongs to the vanillyl alcohol oxidase/para-cresol methylhydroxylase (VAO/PCMH) family, a

subgroup that is distantly related to other flavoproteins such as GULO and the *B. mori* DHCR24 (Ewing *et al.* 2017). Thus, it appears that distinct VAO/PCMH enzymes can be functionally recruited to assume GULO function depending on the considered animal lineage. Ultimately, this evidence does not alter any of the conclusions of the article by Duque *et al.* (Duque *et al.* 2022) and supports the hypothesis that VC synthesis in protostomians may have evolved through multiple enzymatic adaptations.

2.2.4. References

Duque, P., C. P. Vieira, B. Bastos and J. Vieira. **2022.** The evolution of vitamin C biosynthesis and transport in animals. *BMC Ecology and Evolution*. 22:84.

Ewing, T. A., M. W. Fraaije, A. Mattevi and W. J. van Berkel. **2017.** The VAO/PCMH flavoprotein family. *Archives of Biochemistry and Biophysics*. 632:104-117.

López-Fernández, H., P. Duque, S. Henriques, N. Vázquez, F. Fdez-Riverola, C. P. Vieira, M. Reboiro-Jato and J. Vieira. **2019.** A Bioinformatics Protocol for Quickly Creating Large-Scale Phylogenetic Trees. In: F. Fdez-Riverola, M. S. Mohamad, M. Rocha, J. F. De Paz and P. González, editors. *Practical Applications of Computational Biology and Bioinformatics, 12th International Conference.* Cham. Springer. p. 88-96.

Patananan, A. N., L. M. Budenholzer, M. E. Pedraza, E. R. Torres, L. N. Adler and S. G. Clarke. **2015.** The invertebrate *Caenorhabditis elegans* biosynthesizes ascorbate. *Archives of Biochemistry and Biophysics*. 569:32-44.

Wheeler, G., T. Ishikawa, V. Pornsaksit and N. Smirnoff. **2015.** Evolution of alternative biosynthetic pathways for vitamin C following plastid acquisition in photosynthetic eukaryotes. *eLife*. 4:e06369.

Yabuta, Y., R. Nagata, Y. Aoki, A. Kariya, K. Wada, A. Yanagimoto, H. Hara, T. Bito, N. Okamoto, S. Yoshida, A. Ishihara and F. Watanabe. **2020.** L-Ascorbate Biosynthesis Involves Carbon Skeleton Rearrangement in the Nematode *Caenorhabditis elegans*. *Metabolites*. 10:334.

Chapter 3. Vitamin C synthesis in *D. melanogaster*: bioinformatics and molecular biology approaches for the identification of novel pathways

This chapter contains phylogenetic data from the following submitted article:

Glez-Peña, D., López-Fernández, H., **Duque, P.**, Vieira, C. P., & Vieira, J. (2024). Inferences on the evolution of the ascorbic acid synthesis pathway in insects using Phylogenetic Tree Collapser (PTC), a tool for the automated collapsing of phylogenetic trees using taxonomic information. *Journal of Integrative Bioinformatics*. Under second round of revision.

3.1. Introduction

Previous literature has revealed that VC can be synthesized in *D. melanogaster*, *B. mori* and *C. elegans* despite the absence of *GULO* (Henriques *et al.* 2019, Hou *et al.* 2021, Massie *et al.* 1991, Patananan *et al.* 2015, Yabuta *et al.* 2020). In *B. mori*, this is likely achieved through the almost complete conservation of the characterized animal VC synthesis pathway and functional adaptation of DHCR24 to fulfil the role of GULO (Duque *et al.* 2022b, Hou *et al.* 2021). The evolutionary scenario is more complex regarding *C. elegans*, as previous phylogenetic analyses suggest that GULO and the gluconolactonase that precedes it in the pathway, Regucalcin, are lost in nematodes (Duque *et al.* 2022a), and are in part functionally substituted by E03H4.3 and F54D5.12, respectively (Yabuta *et al.* 2020). When both cases are considered, it is clear that divergent VC synthesis mechanisms have emerged within protostomians. Nevertheless, with the current knowledge, it is unknown how such distinct adaptive processes relate to the VC synthesis pathway of *D. melanogaster*.

B. mori and *D. melanogaster* are insects, and thus it is intuitive to assume that the substitution of GULO with a GULO-like enzyme would be the more parsimonious evolutionary explanation for the endogenous VC synthesis capacity of both organisms. In this context, the remaining enzymatic constituents of the characterized animal VC synthesis pathway should have conserved orthologs in *D. melanogaster*. It is known that vertebrate VC production relies on the initial enzymatic conversion of D-Glucose-6P to D-Glucose-1P (Caspi *et al.* 2020, Wheeler *et al.* 2015) (Figure 3.1). This reaction can be catalysed by Phosphoglucomutases 1 and 2 (PGM1 and PGM2), but PGM1 is mainly responsible for the total phosphoglucomutase activity in vertebrates, while PGM2 acts primarily as a phosphoribomutase (Beck 1979, Maliekal *et al.* 2007, Muenks *et al.* 2017). Furthermore, *PGM2* has been shown to have very reduced expression in kidney and liver relative to *PGM1* (Muenks *et al.* 2017). Thus, considering that VC synthesis in vertebrates occurs in the liver and/or kidney (Drouin *et al.* 2011, Dutta Gupta *et al.* 1973), PGM2 likely does not play a considerable role in VC production. D-Glucose-1P is after subsequently converted to D-Glucuronate through the successive intervention of the UDP-Glucose pyrophosphorylase 2 (UGP2), the UDP-Glucose 6-dehydrogenase (UGDH), and a putative UDP-Glucuronidase (UGUR) (Caspi *et al.* 2020, Linster and Van Schaftingen 2007, Wheeler *et al.* 2015) (Figure 3.1). The A1 and B1 members of the aldo-keto reductase family 1 (AKR1A1/AKR1B1) then transform D-Glucuronate into L-Gulonate (Caspi *et al.* 2020, Gabbay *et al.* 2010, Linster and Van Schaftingen 2007, Wheeler *et al.* 2015) (Figure 3.1). Although in vertebrates the AKR1A1 appears to be responsible for approximately 85% of this substrate processing (Gabbay *et al.* 2010), this protein has no considerable homolog in *D. melanogaster*

(information available at https://www.ensembl.org; gene ENSG00000117448) (Duque *et al.* 2022b). Nevertheless, the AKR1B1 appears to be an AKR1A1 synergist in vertebrates, being responsible for the remaining 15% of the D-Glucuronate conversion (Gabbay *et al.* 2010). Thus, if present, putative AKR1B1 homologs should contribute at least in part to VC synthesis in *D. melanogaster*. Finally, L-Gulonate is converted to the substrate of GULO, L-Gulono-1,4-lactone, by Regucalcin (Caspi *et al.* 2020, Gabbay *et al.* 2010, Linster and Van Schaftingen 2007, Wheeler *et al.* 2015) (Figure 3.1). Two *Regucalcin* homologs, *Regucalcin* and *Dca*, were already identified in *D. melanogaster* (Arboleda-Bustos and Segarra 2011, Duque *et al.* 2022a). However, the presence of the remaining pathway constituents remains to be addressed under the VC synthesis scope in this species. Currently, there is no candidate for the functional replacement of GULO in *D. melanogaster*, as previous sequence homology searches indicated that *DHCR24* homologs were not found in this species (Duque *et al.* 2022b). Nevertheless, detailed phylogenetic analyses regarding the distribution patterns of DHCR24 homologs within insects might reveal other taxonomic lineages capable of endogenous VC synthesis (Duque *et al.* 2022b).

Alternatively, as evidenced in *C. elegans* (Yabuta *et al.* 2020), it is possible that VC synthesis in *D. melanogaster* derives from multiple enzymatic adaptations. If so, then the identification of novel pathway constituents is hard, particularly when considering that previous literature regarding *Regucalcin* (e.g. (Duque *et al.* 2022a, Patananan *et al.* 2015, Wheeler *et al.* 2015)) has revealed that comprehensive searches for CDS homology cannot always help determine gene candidates susceptible to functional changes. In an extreme case, it is even possible that dramatically different novel VC synthesis pathways are present in protostomians, further complicating the characterization process. This is not unreasonable, since E03H4.3 and F54D5.12 seem to account for 18% and 19%, respectively, of total VC production in *C. elegans* (Yabuta *et al.* 2020), and thus are more likely associated with an accessory than a main VC synthesis pathway.

Moreover, although disregarded in the literature, myo-inositol metabolism may also contribute to the endogenous synthesis of this nutrient in animals. The animal MIOX catalyses the conversion of myo-inositol to D-Glucuronate (Brown *et al.* 2006, Thorsell *et al.* 2008), which is a key molecule in the glucuronate-xylulose pathway (Chang *et al.* 2015, Lu and Agarwal 2022) and the Regucalcin substrate in the vertebrate VC synthesis process (Caspi *et al.* 2020, Gabbay *et al.* 2010, Linster and Van Schaftingen 2007, Wheeler *et al.* 2015) (Figure 3.1). This gene has been previously implicated in VC synthesis in the kidney of avian species (Hänninen *et al.* 1971), and therefore, it has been suggested that it can also contribute to the endogenous production of this nutrient in protostomians such as *D. melanogaster* (Duque *et al.* 2022b). Interestingly, data available in the KEGG Pathway

(https://www.genome.jp/kegg/pathway.html; (Kanehisa and Goto 2000)) and MetaCyc databases (https://metacyc.org/; (Caspi *et al.* 2020)) indicates that myo-inositol can also be derived from D-Glucose-6P through two conversion steps catalysed by the inositol-3-phosphate synthase 1 (ISYNA1) and Inositol monophosphatases 1 and 2 (IMPA1 and IMPA2) in vertebrates, respectively. Therefore, the possibility of a synergistic interaction between the mainly conserved characterized animal pathway and an uncharacterized myo-inositol metabolism pathway for the purpose of VC synthesis in some animal lineages cannot be excluded (Figure 3.1).

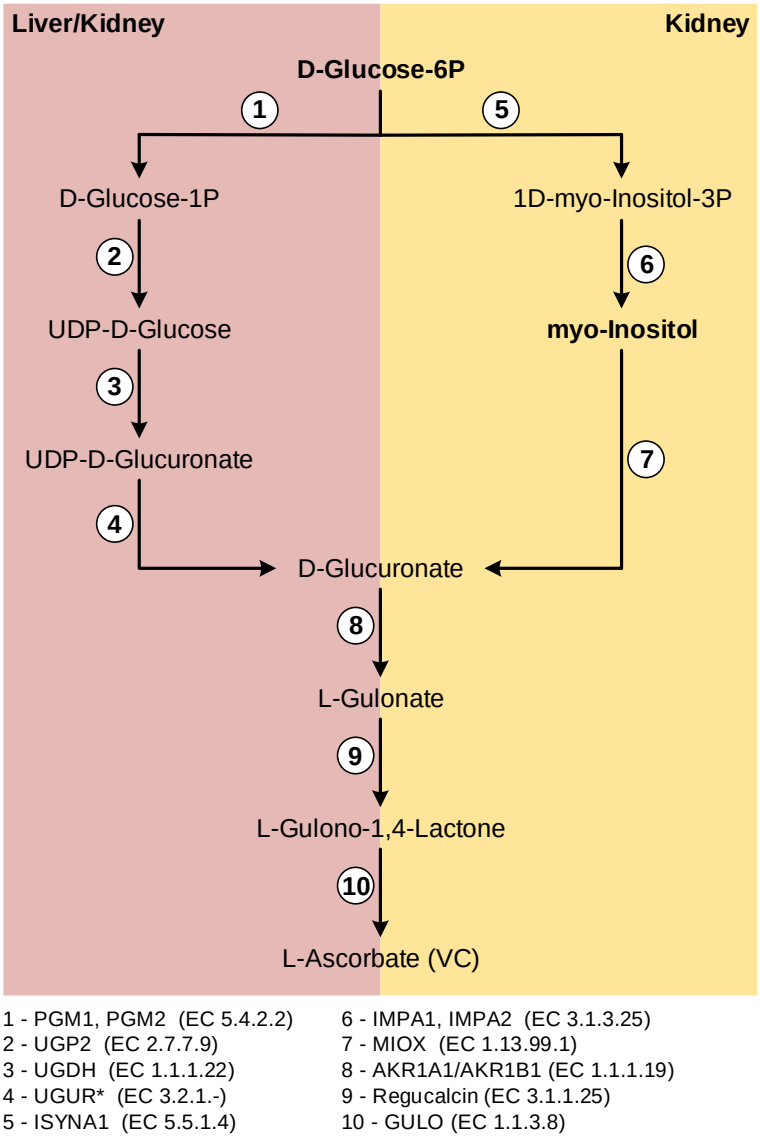


Figure 3.1. Graphical representation of possible VC synthesis pathways in vertebrates. The numbered circles represent the enzymes that catalyse each pathway step, according to the number legend presented at the bottom of the figure. UGUR is highlighted with "*" since remains to be identified and characterized. The light red background indicates pathway steps that are known to occur in the liver and kidney of vertebrates, while the yellow background indicates the putative myo-inositol pathway that is likely involved in kidney VC synthesis in galliform species.

Considering the presented background, none of the hypotheses regarding possible VC synthesis mechanisms in *D. melanogaster* should be ruled out from further studies. Fortunately, despite the distinct experimental challenges they presented, all hypotheses could be addressed simultaneously using a combination of bioinformatics and genomic screenings. As a first methodological approach, phylogenetic analyses of genes related to the characterized animal VC synthesis pathway in insects were performed to provide information regarding the presence of candidate homologs in *D. melanogaster*. Such information proved invaluable in order to reconstruct the possible VC synthesis pathway in this organism assuming the pathway conservation hypothesis, as well as a myo-inositol derived enzymatic mechanism. Moreover, these data was also useful in the context of further functional screenings, as specific RNAi-mediated suppression of identified homologs, and the use of chromosomal deletion stocks where they are partially deleted, hinted to their role in VC synthesis in this species. Coincidentally, chromosomal deletions constitute a prominent genetic tool for the mapping and identification of genes involved in closely related metabolic processes in *D. melanogaster* (Cook *et al.* 2012), and were thus additionally used for the identification of putative alternative candidate enzymes outside the scope of pathway conservation.

The performed phylogenetic analyses revealed that *D. melanogaster* can possibly synthesize VC using an almost completely conserved animal VC synthesis pathway and a myo-inositol pathway. The use of chromosomal deletions further allowed the inference of candidate genes likely involved in VC production in *D. melanogaster*, highlighting that *Pgm1* and *IMPA* homologs (*CG17026* to *CG17029*) can be involved in the process. The RNAi methodology utilized to verify this hypothesis presented technical limitations and was ultimately unable to confirm or dismiss a possible role of *Pgm1* and *CG17029* in *D. melanogaster* endogenous VC synthesis. Nevertheless, it showed that Miox-derived myo-inositol metabolism contributes to total VC levels in adult flies. Further research is still required to verify if the characterized animal VC synthesis pathway is acting synergistically with the myo-inositol pathway for the purpose of VC production in *D. melanogaster*.

3.2. Results and discussion

3.2.1. Identification of gene homologs related to described VC synthesis pathways in *D. melanogaster*

In order to determine if genes involved in the characterized vertebrate VC synthesis pathway, the adapted *B. mori* pathway or the myo-inositol metabolism pathway were conserved in *D. melanogaster*, phylogenetic analyses for the *PGM1*, *UGP2*, *UGDH*, *AKR1B1*, *DHCR24*, *MIOX*, *ISYNA1*, *IMPA1* and *IMPA2* were performed using annotated

Hexapoda genomes to provide evolutionary context. In these analyses, gene presence or absence were inferred using the restrictive criteria of previous research (Duque *et al.* 2022a).

The phylogenetic data indicates that the *PGM1* gene is present in all the major taxonomic groups analysed as a single copy, and thus displays a linear evolutionary history within insects (Figures 3.2 and S3.1). In this context, the single *D. melanogaster* sequence found (*Phosphoglucose mutase 1* (*Pgm1*), accession number NP_524675.1) is the likely vertebrate *PGM1* homolog.

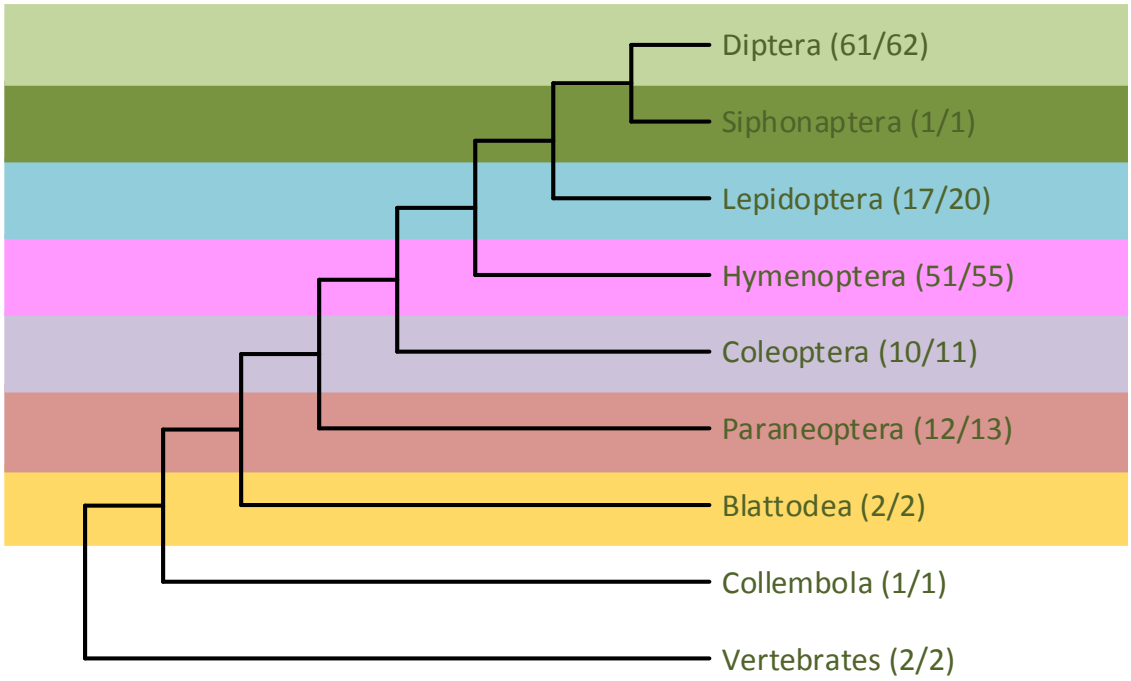


Figure 3.2. Proposed *PGM1* evolutionary scenario across the Hexapoda subphylum. The higher rank insect taxonomic groups are marked with distinct colours. Gene presence is highlighted in green. The number of species present in the final dataset and total number of species analysed for each taxonomic group can be seen within parentheses, respectively. Taxonomic relationships are depicted according to the Tree of Life web project (Maddison *et al.* 2007).

Despite the absence of *UGP2* in the Siphonaptera, the *UGP2* gene has an overall ubiquitous distribution across all the major taxonomic groups analysed (Figures 3.3 and S3.2). Nevertheless, since this group is only represented by a single species (*Ctenocephalides felis*), it is not possible to infer complete gene loss in this lineage. In general, the *UGP2* gene appears to follow a conserved single copy state in all Hexapoda groups, although it can be found locally duplicated in the dipteran *Rhagoletis* genus and in the Paraneoptera *Acyrthosiphon pisum* and *Thrips palmi* species. In *D. melanogaster*, a single sequence (*UDP-glucose pyrophosphorylase* (*UGP*), accession number NP_001163399.1) is the likely vertebrate *UGP2* homolog.

Due to the inherent variation of GC content within the *UGDH* dataset, two individual Bayesian trees were created and manually integrated to obtain a global phylogenetic perspective. By compiling the information of both phylogenies, it is possible to observe that the *UGDH* gene is represented in all the analysed higher taxonomic groups (Figures 3.4 and S3.3). Although this gene is vastly present as a single copy, it is possible to observe local gene duplications within the Diptera, Lepidoptera, Coleoptera and Paraneoptera taxonomic groups. After this analysis, a single *D. melanogaster* sequence (*sugarless* (*sgl*), accession number NP_476980.1) is the likely vertebrate *UGDH* homolog.

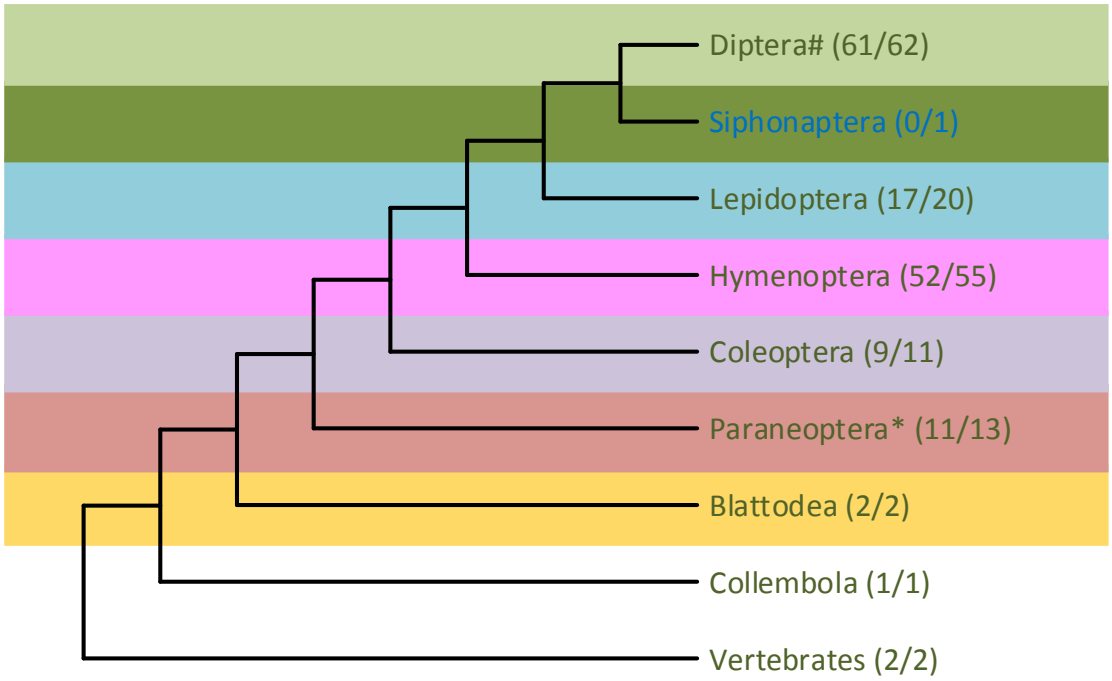


Figure 3.3. Proposed *UGP2* evolutionary scenario across the Hexapoda subphylum. The higher rank insect taxonomic groups are marked with distinct colours. Gene presence is highlighted in green and uncertainty regarding gene loss or presence in light blue. Gene duplications affecting a single species of a genus are marked with a *, while those which affected two or more species from the same genus are marked with a #. The number of species present in the final dataset and total number of species analysed for each taxonomic group can be seen within parentheses, respectively. Taxonomic relationships are depicted according to the Tree of Life web project (Maddison *et al.* 2007).

The *AKR1B1* dataset was particularly difficult to analyse since it not only had a very high number of sequences, but also high sequence divergence due to the presence of many members of the AKR1B protein family after the initial homolog search. Due to the impact of these limitations in the running time and final convergence of a Bayesian phylogenetic inference method, the FastTree 2 software (Price *et al.* 2010) was chosen for this process. Nevertheless, although this method is much faster for the processing of a large amount of sequences, it usually sacrifices accuracy in the determination of the phylogenetic relationship between them (Yang and Warnow 2011). This is evidenced for example by the

presence of tree branches with a zero-support value, which indicate that the branch location in some regions of the phylogeny is random and should not be interpreted in a deterministic way. As such, the summary of these results represents an approximated evolutionary hypothesis based on the global phylogeny, which required personal curation regarding current taxonomic context and understanding of duplication or loss events.

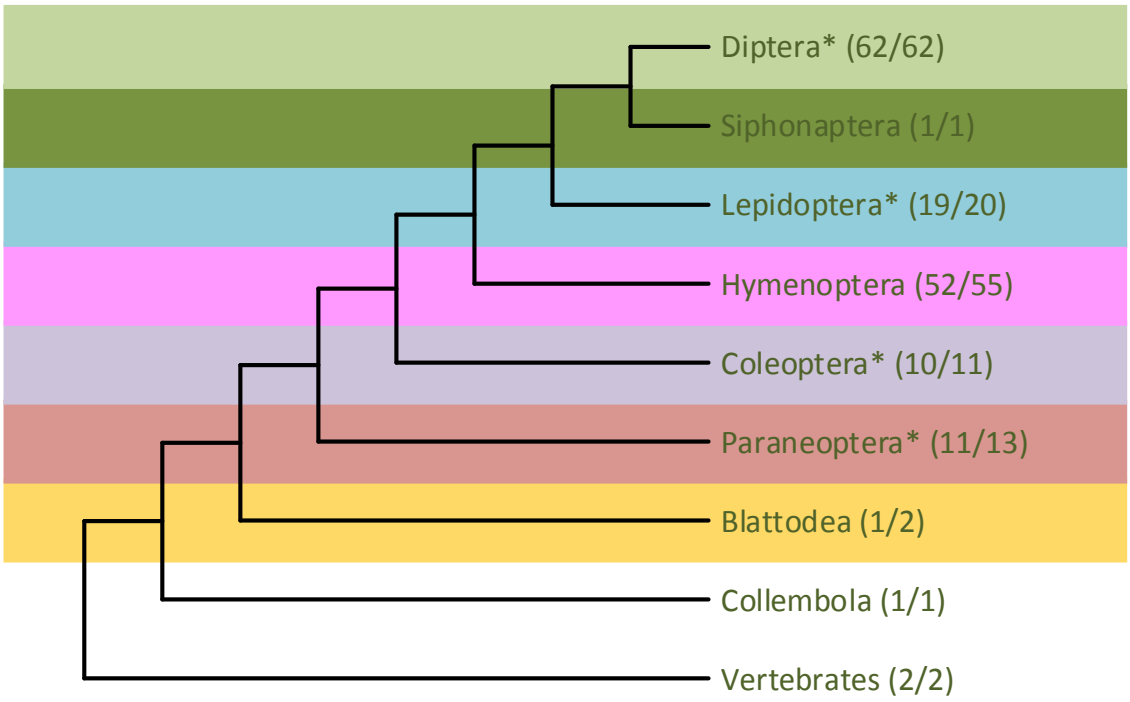


Figure 3.4. Proposed *UGDH* evolutionary scenario across the Hexapoda subphylum. The higher rank insect taxonomic groups are marked with distinct colours. Gene presence is highlighted in green. Small gene duplication events within specific taxonomic groups are marked with *. The number of species present in the final dataset and total number of species analysed for each taxonomic group can be seen within parentheses, respectively. Taxonomic relationships are depicted according to the Tree of Life web project (Maddison *et al.* 2007).

Overall, it was possible to infer three main gene duplicate clusters within the *AKR1B1* phylogeny. The *AKR1B1 Gene 1* is present in all the major taxonomic groups, and presents a high prevalence of conserved duplicates (Figures 3.5 and S3.4). Within the Diptera, the most evident duplication affected the common ancestor of the Drosophilidae and Tephritidae families, originating two distinct duplicates (*DT 1* and *DT 2*). In addition, both *DT 1* and *DT 2* can be found duplicated once in the Tephritidae family, originating four *AKR1B1* gene 1 homologs in these species. In a likely independent event, the *AKR1B1 Gene 1* can also be found duplicated within the Culicidae group. Contrary to the Siphonaptera group represented by a single copy gene, the Lepidoptera lineage also appears to have been affected by duplication events, since three distinct gene duplicates can be inferred at the base of the taxonomic group (*LEP 1* to *LEP 3*). Although not all

species appear to have the three genes, gene loss can only be inferred within the Bombycoidea group for gene *LEP 1*.

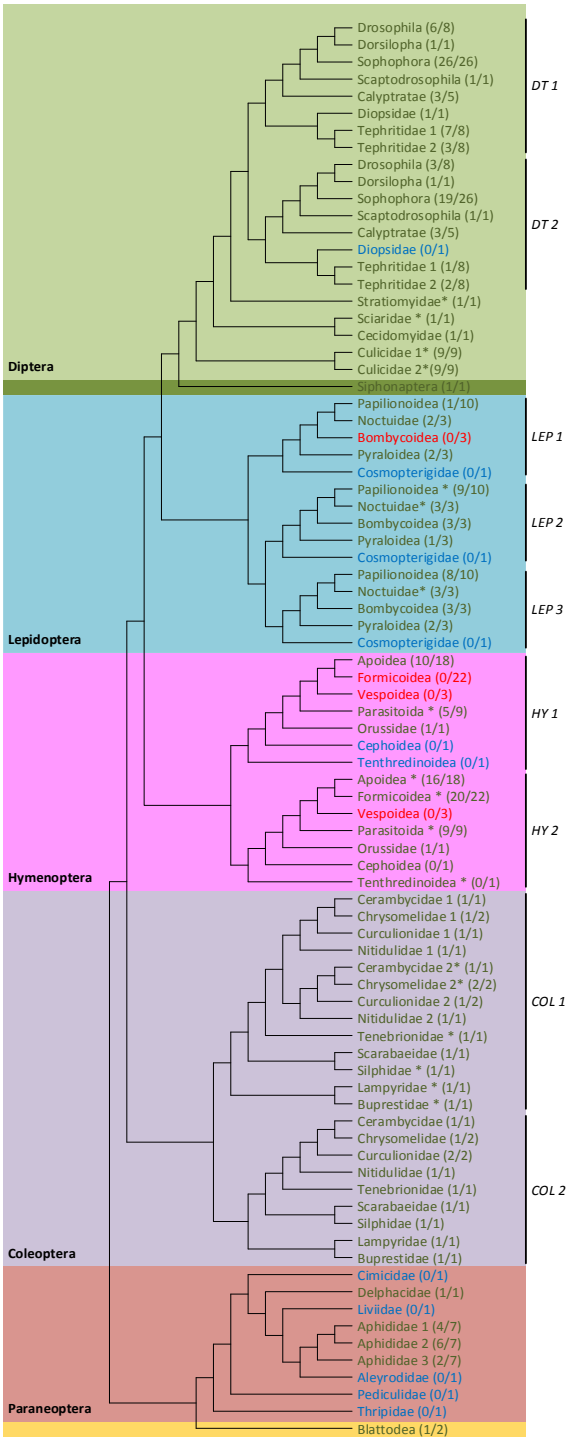


Figure 3.5. Proposed *AKR1B1 Gene 1* evolutionary scenario across the Insecta. The higher rank taxonomic groups are marked with distinct colours. Gene presence is highlighted in green, absence in red and uncertainty regarding gene loss or presence in light blue. Relevant gene duplicates are represented by tags present in the right side of the cladogram. Small gene duplication events within specific taxonomic groups are marked with *. The number of species present in the final dataset and total number of species analysed for each taxonomic group can be seen within parentheses, respectively. (continues next page)

Taxonomic relationships are depicted according to the Tree of Life web project (Maddison *et al.* 2007) and published literature (O'Grady and DeSalle 2018, Peters *et al.* 2017, Van Der Linde *et al.* 2010, Wiegmann *et al.* 2011, Zhang *et al.* 2018).

The Hymenoptera lineage is represented by two *AKR1B1 Gene 1* copies (*HY 1* and *HY 2*) overall spread across the analyzed taxonomic groups, with the exception of the Vespoidea (lost both genes) and Formicoidea (lost *HY 1*). From the global perspective within the Coleoptera, it is possible to observe two ancestral *AKR1B1 Gene 1* duplicates present in all taxonomic groups (*Col 1* and *Col 2*). Furthermore, *Col 1* appears to have been additionally duplicated in the common ancestor of the Cerambycidae, Chrysomelidae, Curculionidae and Nitidulidae. Although there is a high prevalence of undefined gene presence/loss scenarios within the Paraneoptera, *AKR1B1 Gene 1* appears to be predominantly single copy, with the exception of the Aphididae group, where a total of three genes can be inferred (*Aphididae 1* to *Aphididae 3*). In the most ancestral insect group (Blattodea), the *AKR1B1 Gene 1* can be found as a single copy. Overall, there is also a tendency for local duplication events within the higher categorized taxonomic groups, with the exception of the Siphonaptera, Paraneoptera and Blattodea.

The *AKR1B1 Gene 2* is present in all the major taxonomic groups and appears to have been subjected to various independent duplication events (Figures 3.6 and S3.4). The Diptera lineage appears to have been affected by duplication events since three distinct gene duplicates can be inferred at the base of the taxonomic group (*DT 1* to *DT 3*). Although not all species appear to have the three genes, gene loss can only be inferred within the Tephritidae (*DT 1* and *DT 2*), Culicidae (*DT 1* and *DT 3*) and Calyptratae (*DT 2*) groups. In addition, *DT 1* can be found duplicated in the Sophophora subgenus and *DT 2* in the Sophophora subgenus and Culicidae family. The Siphonaptera and Lepidoptera lineages appear to be represented by a single-copy gene, and almost all lepidopteran species seem to have lost it. The Hymenoptera lineage is represented by two *AKR1B1 Gene 2* copies (*HY 1* and *HY 2*) overall spread across the analysed taxonomic groups. Within these duplicates lineages, gene loss events can be inferred within the Megachilidae (*HY 1*), Halictidae (*HY 1*), Vespoidea (*HY 1*) and Parasitoida (*HY 2*) groups. The Coleoptera lineage is represent by two *AKR1B1 Gene 2* copies (*COL 1* and *COL 2*) which appears to be spread across the analysed taxonomic groups. Despite the high prevalence of undefined gene presence/loss representations within the Paraneoptera, the *AKR1B1 Gene 2* can be found in the Thripidae family as a single copy and seems to be absent in the Aphididae group. The *AKR1B1 Gene 2* can be found in the most ancestral insect group (Blattodea). Overall, there is a tendency for local duplication events within the higher categorized taxonomic groups, excluding only the Siphonaptera, Lepidoptera and Paraneoptera groups.

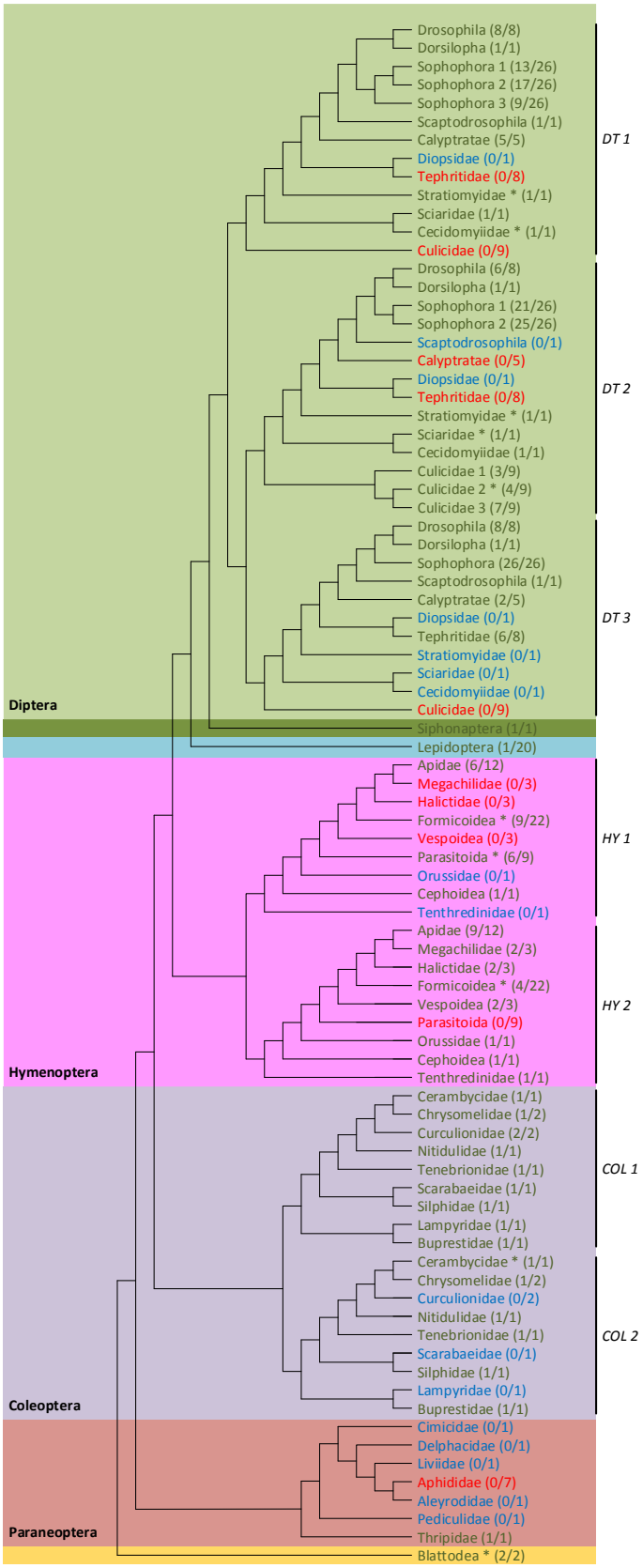


Figure 3.6. Proposed *AKR1B1 Gene 2* evolutionary scenario across the Insecta. The higher rank taxonomic groups are marked with distinct colours. (Continues next page)

Gene presence is highlighted in green, absence in red and uncertainty regarding gene loss or presence in light blue. Relevant gene duplicates are represented by tags present in the right side of the cladogram. Small gene duplication events within specific taxonomic groups are marked with *.The number of species present in the final dataset and total number of species analysed for each taxonomic group can be seen within parentheses, respectively. Taxonomic relationships are depicted according to the Tree of Life web project (Maddison *et al.* 2007) and published literature (O'Grady and DeSalle 2018, Peters *et al.* 2017, Van Der Linde *et al.* 2010, Wiegmann *et al.* 2011, Zhang *et al.* 2018).

The *AKR1B1 Gene 3* is present in all the major taxonomic groups, but unlike the previous two *AKR1B1* genes, it seems to have a less complex duplication pattern (Figures 3.7 and S3.4). The Culicidae group appears to have been affected by two ancestral duplication events, a scenario that is not observed in the remaining dipterans. Two independent gene loss events can be inferred concerning the Drosophila subgenus and Tephritidae family. Even though *AKR1B1 Gene 3* is present in the Sophophora group, many species also appear to have lost this gene. While the Siphonaptera group is represented by a single copy gene, the Lepidoptera lineage appears to have been affected by duplication events, since two distinct gene duplicates (*LEP 1* and *LEP 2*) can be inferred at the base of the taxonomic group. Various gene loss events can be inferred within the two duplicate lineages, namely the Papilionidae, Nymphalidae and Noctuidae for gene *LEP 1*, and Bombycoidea for both *LEP 1* and *LEP 2* genes. The Hymenoptera lineage is represented by a single copy gene, as well as the Coleoptera group. Similarly to *AKR1B1 Gene 1* and *AKR1B1 Gene 2*, *AKR1B1 Gene 3* presence appears to be restricted to a few Paraneoptera groups. Nevertheless, the *AKR1B1 Gene 3* gene is locally duplicated in the Cimicidae lineage and ancestrally duplicated in the Aphididae group, where a total of two genes can be inferred. It is not possible to assume gene presence or loss in the most ancestral insect group (Blattodea) with the available data. Overall, there is a tendency for local duplication events within the higher categorized taxonomic groups, excluding only the Siphonaptera and Hymenoptera groups.

When considering the full phylogenetic context, the three *AKR1B1* gene lineages appear to have derived from two subsequent duplication events of a common ancestral gene (Figure 3.8). Several vertebrate *AKR1B1* homologs (*Aldo-keto reductase 1B* (*Akr1B*; NP_001261717.1), *CG6083* (NP_648485.1), *CG12766* (NP_647839.1), *CG10863* (NP_647840.1), *CG10638* (NP_996068.1 and NP_729808.1) and *CG9436* (NP_610235.1)) are found in *D. melanogaster*. Unlike the other homologs, *CG10638* is represented by two isoform sequences with high sequence divergence. Whether this result is derived from a low quality annotation of one sequence or real post-transcriptional processing variability between both sequences remains to be determined.

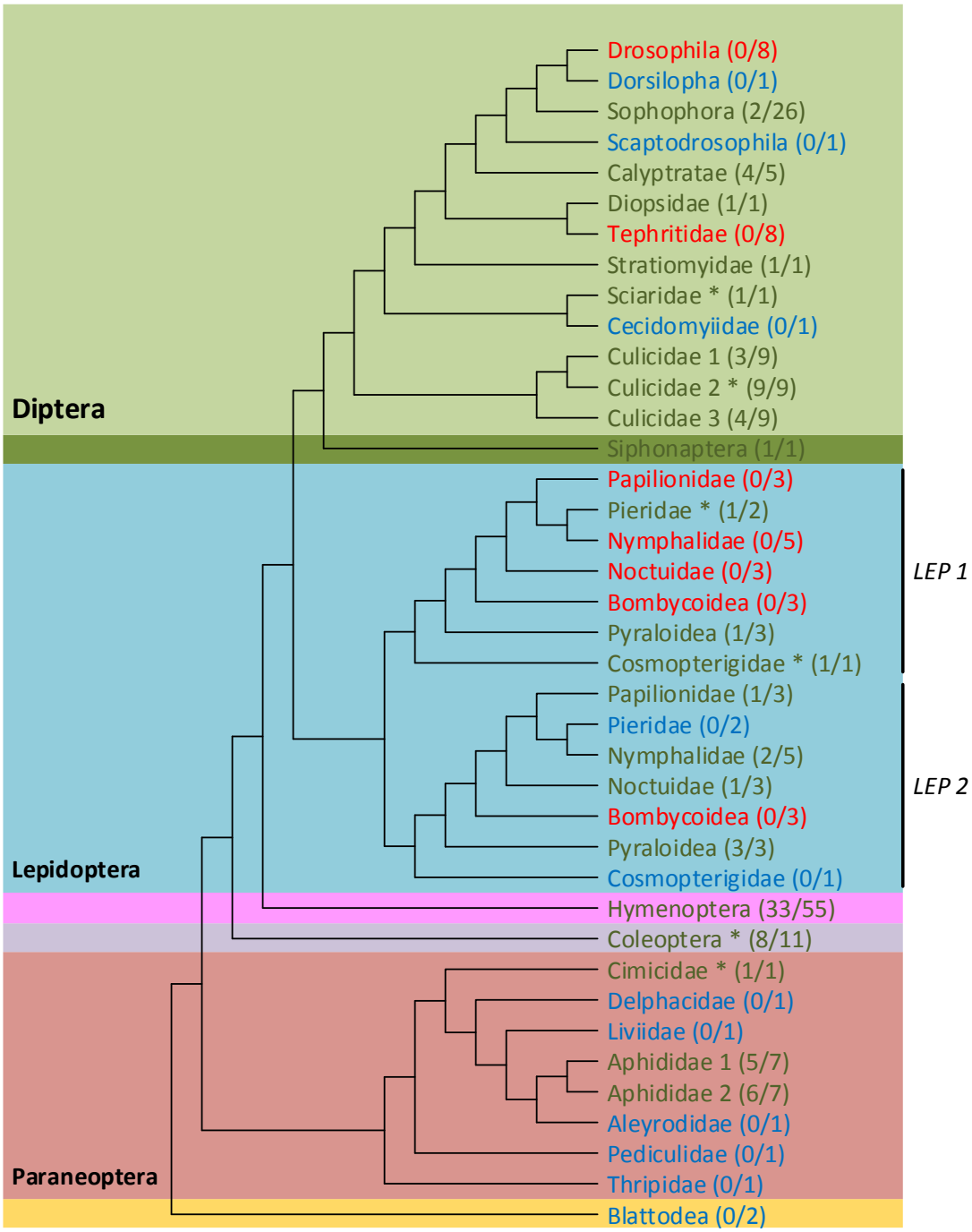


Figure 3.7. Proposed *AKR1B1 Gene 3* evolutionary scenario across the Insecta. The higher rank taxonomic groups are marked with distinct colours. Gene presence is highlighted in green, absence in red and uncertainty regarding gene loss or presence in light blue. Relevant gene duplicates are represented by tags present in the right side of the cladogram. Small gene duplication events within specific taxonomic groups are marked with *. The number of species present in the final dataset and total number of species analysed for each taxonomic group can be seen within parentheses, respectively. Taxonomic relationships are depicted according to the Tree of Life web project (Maddison *et al.* 2007) and published literature (O'Grady and DeSalle 2018, Peters *et al.* 2017, Van Der Linde *et al.* 2010, Wiegmann *et al.* 2011).

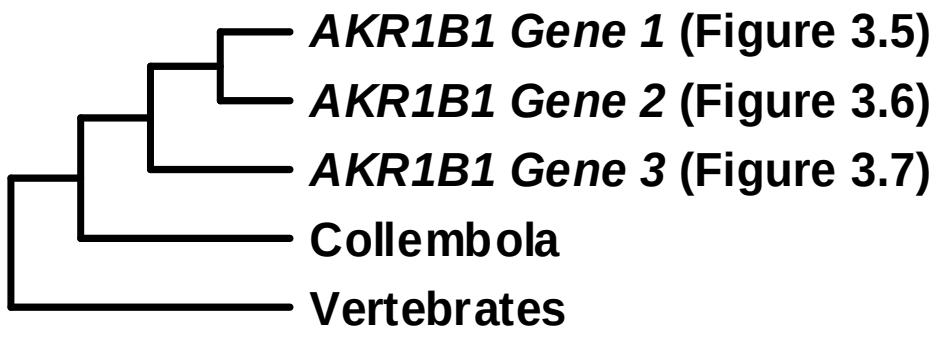


Figure 3.8. Summary cladogram of the proposed evolutionary history of the three *AKR1B1* genes within the Insecta, represented relative to the selected outgroups.

The gene identified in *B. mori* by Hou *et al.* as having GULO-like activity (Hou *et al.* 2021), *DHCR24*, is present as a single copy in the vertebrate species used as phylogenetic root (Figures 3.9 and S3.5). However, four distinct gene lineages are present in hexapods (*DHCR24 1* to *DHCR24 4*). The reported *B. mori DHCR24* belongs to the *DHCR24 1* lineage. This gene lineage was lost in the majority of the taxonomic groups analysed, being only present in the Lepidoptera and Paraneoptera orders. *DHCR24 2* is present in a reduced number of Hymenoptera and basal Diptera families, as well as the Coleoptera, Blattodea and Collembola. *DHCR24 3* gene is present in the Diptera, Lepidoptera, Coleoptera, Paraneoptera and Collembola. This gene appears to have high functional relevance within the Paraneoptera Aphididae family members, as two duplicates of the ancestral *DHCR24 3* lineage can be observed in these species. *DHCR24 4* is only present in Collembola, and appears to have been lost in the Insecta lineage. There are no homologs of any *DHCR24* lineage within the dipteran Brachycera suborder, where *D. melanogaster* is included. As previously stated (Duque *et al.* 2022b), this species is a cholesterol auxotroph (Jing and Behmer 2020, Lavrynenko *et al.* 2015) that has lost the ability to convert desmosterol to cholesterol (Svoboda and Feldlaufer 1991), a metabolic function that is likely related to the loss of *DHCR24*. Thus, as expected, VC synthesis in *D. melanogaster* does not derive from the DHCR24 evolutionary adaptation that is observed in *B. mori*.

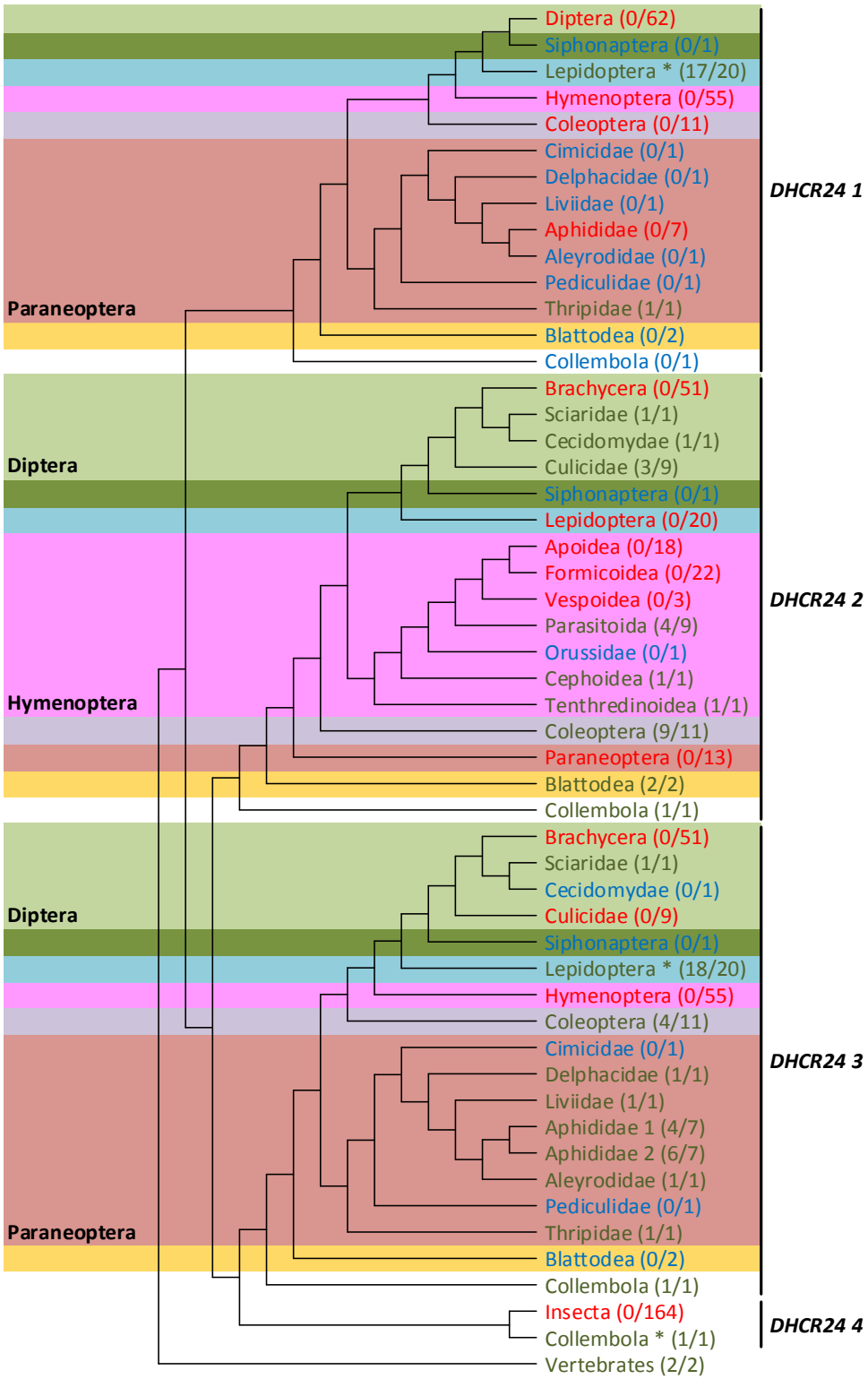


Figure 3.9. Proposed *DHCR24* evolutionary scenario across the Hexapoda subphylum. The higher rank insect taxonomic groups are marked with distinct colours. Gene presence is highlighted in green, absence in red and uncertainty regarding gene loss or presence in light blue. The inferred gene duplicates are tagged with a number suffix in the relevant taxonomic groups. Small gene duplication events within specific taxonomic groups are marked with *. The number of species present in the final dataset and total number of species analysed for each taxonomic group can be seen within parentheses, respectively. (Continues next page)

Taxonomic relationships are depicted according to the Tree of Life web project (Maddison *et al.* 2007) and published literature (O'Grady and DeSalle 2018, Peters *et al.* 2017, Van Der Linde *et al.* 2010, Wiegmann *et al.* 2011).

MIOX homologs are present in all of the higher taxonomic groups (Figures 3.10 and S3.6). According to the phylogeny, a specific gene duplication event has led to the emergence of two *MIOX* duplicates within lepidopterans (*LEP 1* and *LEP 2*). A single *MIOX* homolog is present in *D. melanogaster* (*Miox* or *CG6910*; NP_648556.3), in accordance with previous observations (Duque *et al.* 2022b).

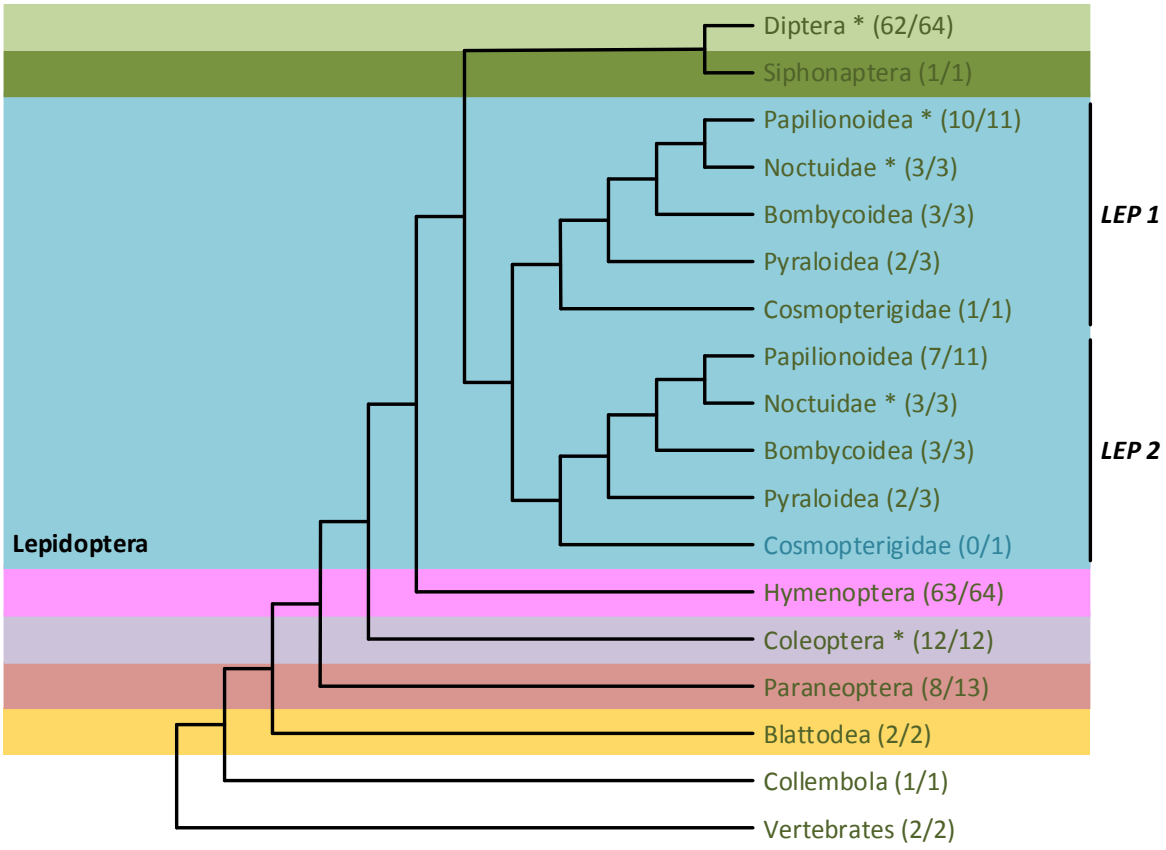


Figure 3.10. Proposed *MIOX* evolutionary scenario across the Hexapoda subphylum. The higher rank insect taxonomic groups are marked with distinct colours. Gene presence is highlighted in green and uncertainty regarding gene loss or presence in light blue. Relevant gene duplicates are represented by tags present in the right side of the cladogram. Small gene duplication events within specific taxonomic groups are marked with *. The number of species present in the final dataset and total number of species analysed for each taxonomic group can be seen within parentheses, respectively. Taxonomic relationships are depicted according to the Tree of Life web project (Maddison *et al.* 2007) and published literature (O'Grady and DeSalle 2018, Peters *et al.* 2017, Van Der Linde *et al.* 2010, Wiegmann *et al.* 2011).

The phylogenetic data suggests that the evolutionary history of *ISYNA1* in hexapods is arguably simple, as this gene is generally present as a single copy in species from all

major insect taxonomic groups (Figures 3.11 and S3.7). A single *ISYNA1* homolog exists in *D. melanogaster*, namely *myo-inositol-1-phosphate synthase* (*Inos*; NP_477405.1).

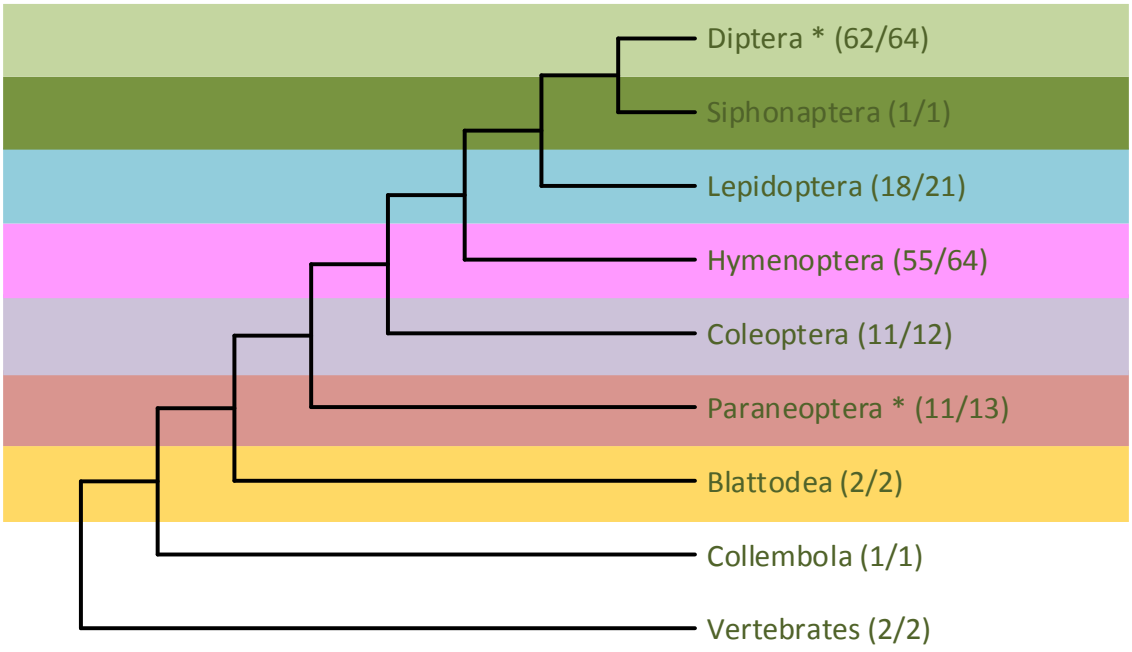


Figure 3.11. Proposed *ISYNA1* evolutionary scenario across the Hexapoda subphylum. The higher rank insect taxonomic groups are marked with distinct colours. Gene presence is highlighted in green. Small gene duplication events within specific taxonomic groups are marked with *. The number of species present in the final dataset and total number of species analysed for each taxonomic group can be seen within parentheses, respectively. The number of species present in the final dataset and total number of species analysed for each taxonomic group can be seen within parentheses, respectively. Taxonomic relationships are depicted according to the Tree of Life web project (Maddison *et al.* 2007).

Curiously, *IMPA* homologs in Hexapoda species form two separate gene clusters (*HEX 1* and *HEX 2*), in a matter likely independent from the duplication event that led to the emergence of *IMPA1* and *IMPA2* in vertebrates (Figures 3.12 and S3.8). The *HEX 1* gene lineage appears to have been duplicated once within dipterans after divergence from the Stratiomyidae group, originating *DT 1* and *DT 2*. Nevertheless, *DT 2* was lost in the Sophophora, Calyptratae and Tephritidae. This gene lineage also appears to have been affected by two additional duplication events in the common ancestor of Drosophila and Dorsilopha, although one of the three originated duplicates seems to have been lost in the Drosophila subgenus. Excluding specific duplications in Lepidoptera, Coleoptera and Paraneoptera species, *HEX 1* is present as a single copy in the non-Diptera hexapods. The *HEX 2* gene has been ancestrally duplicated within dipterans (*DIP 1* and *DIP 2*), similarly to what was observed for *HEX 1*. Both *DIP 1* and *DIP 2* where duplicated once, originating a total of four *HEX 2 IMPA* homologs in many dipteran species.

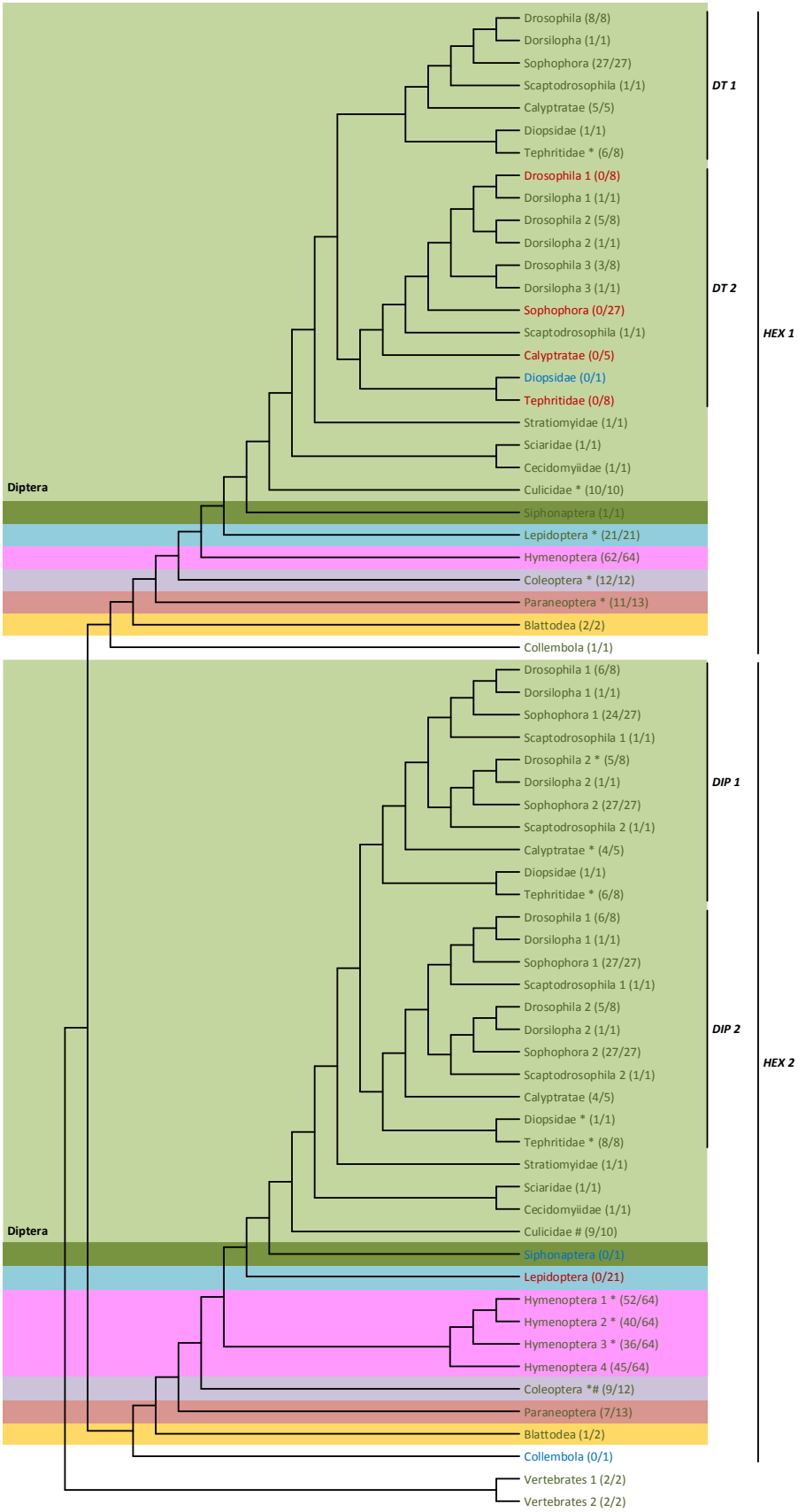


Figure 3.12. Proposed *IMPA* evolutionary scenario across the Hexapoda subphylum. The higher rank taxonomic groups are marked with distinct colours. (Continues next page)

Gene presence is highlighted in green, absence in red and uncertainty regarding gene loss or presence in light blue. Relevant gene duplicates are represented by tags present in the right side of the cladogram. Gene duplications affecting a single species of a genus are marked with a *, while those which affected two or more species from the same genus are marked with a #. The number of species present in the final dataset and total number of species analysed for each taxonomic group can be seen within parentheses, respectively. Taxonomic relationships are depicted according to the Tree of Life web project (Maddison *et al.* 2007) and published literature (O'Grady and DeSalle 2018, Peters *et al.* 2017, Van Der Linde *et al.* 2010, Wiegmann *et al.* 2011).

Four total *HEX 2 IMPA* homologs are also present in the majority of Hymenoptera species, although they seem to have emerged independently of the duplication events that led to the distinct dipteran *HEX 2* gene lineages. This gene lineage appears to have been duplicated in a small fraction of Curculionidae coleopterans, and is present as a single copy in Paraneoptera and Blattodea. No *HEX 2* was found in the analysed lepidopterans, Siphonaptera and Collembola. One *HEX 1* (*CG9391*; NP_730610.3) and four *HEX 2* (*CG17026* (NP_648820.1), *CG17027* (NP_001261909.1), *CG17028* (NP_001189114.1) and *CG17029* (NP_001261908.1)) *IMPA* homologs were identified in *D. melanogaster*.

When contextualizing the obtained data with a previous phylogenetic analysis of the *Regucalcin* gene (Duque *et al.* 2022a), it is evident that, excluding *GULO* (López-Fernández *et al.* 2019), genes related to the characterized animal VC synthesis pathway are generally conserved in insect species (Figure 3.13). The evaluation of *AKR1A1* distribution in hexapods could provide further insights on the efficiency of VC synthesis in insects capable of doing so, since AKR1A1 appears responsible for approximately 85% of the D-Glucuronate conversion to L-Gulonate in vertebrates (Gabbay *et al.* 2010). In this context, insects with AKR1A1 orthologs can likely possess optimized D-Glucuronate metabolism that leads to increased rates of VC production. Nevertheless, given the reported absence of considerable *AKR1A1* homologs in *D. melanogaster*, such analysis was outside of the scope of this work. Interestingly, the emergence of three insect *AKR1B1* gene lineages from a common ancestral lineage observed in Collembola species indicates that possible *AKR1B1* subfunctionalization or neofunctionalization processes originated advantageous phenotypes in insects. Therefore, it is possible that evolutionary constraints led to the adaptation and subsequent efficiency increase of AKR1B1 homologs in the context of VC synthesis. Interestingly, species from the Sophophora subgenus, such as *D. melanogaster*, are among those that have a particularly high prevalence of *AKR1B1* duplications. Thus, the ability to synthesize VC in this model organism might be correlated with conservation of certain *AKR1B1* duplicates. The adaptation of DHCR24 as functional replacement of GULO for VC synthesis purposes in *B. mori* is likely an isolated evolutionary event, although the same *DHCR24 1* gene lineage can be observed in Thripidae (Paraneoptera) species. Such

evolutionary scenario is, however, excluded within Brachycera species due to the loss of all *DHCR24* gene lineages. The phylogenetic data also hints that myo-inositol metabolism is essential in hexapods, given that *ISYNA1*, *IMPA* and *MIOX* homologs are found in all taxonomic groups. As observed for *AKR1B1*, *IMPA* has been duplicated many times within dipterans, hence suggesting that *IMPA* subfunctionalization or neofunctionalization processes had a considerable evolutionary impact in the biological performance of these species. Moreover, although *MIOX* duplications are seemingly rare in hexapods, it is interesting to observe that *MIOX* has only been duplicated within lepidopterans, one of which known to be able to synthesize VC (*B. mori*; (Hou *et al.* 2021)). Nevertheless, how myo-inositol relates to VC synthesis in hexapods remains to be addressed. Ultimately, the phylogenetic data suggests that VC synthesis in *D. melanogaster* can occur through both the conservation of the majority of the characterized animal VC synthesis pathway and the presence of a myo-inositol metabolic pathway. However, this organism is not capable of VC synthesis due to DHCR24 functional adaptation.

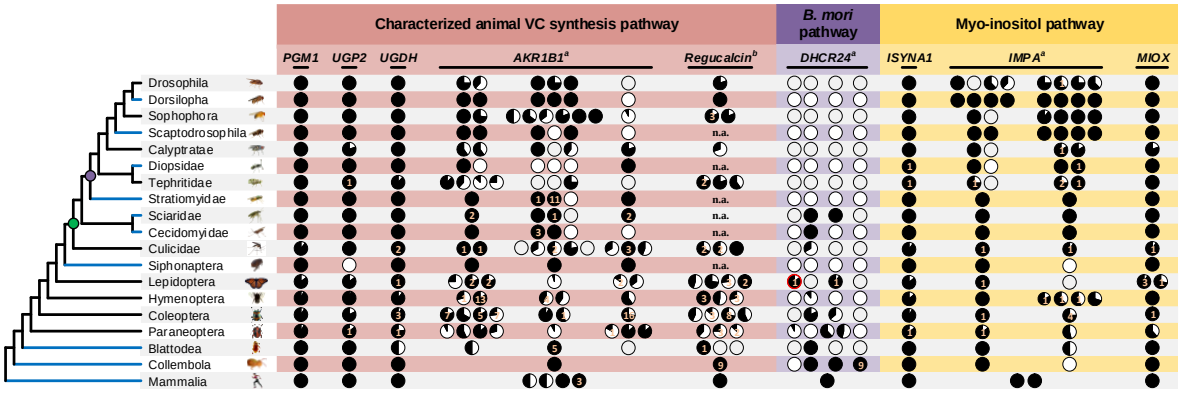


Figure 3.13. Phylogenetic distribution of the analysed genes and *Regucalcin* within Hexapoda species. Genes involved in distinct pathways are highlighted with different background colours. The circles represent the total number of inferred homologs for each taxonomic category in the cladogram. The filling within the circles indicates the number of species in which the given gene was identified relative to the total number of species analysed. Orange numbers superimposing the circles represent small gene duplication events within a specific taxonomic group. Blue branches in the cladogram highlight the taxonomic groups represented by less than three species. The green and purple circles included in the cladogram indicate the start of the Diptera order and Brachycera suborder, respectively. Taxonomic relationships are depicted as seen in the Tree of life web project (Maddison *et al.* 2007) and the literature (O'Grady and DeSalle 2018, Van Der Linde *et al.* 2010, Wiegmann *et al.* 2011). The gene sections marked with "a" represent genes with inferred ancestral duplicates separated by distinct columns from gene 1 (left side) up to gene 4 (right side). The red color outline within the *DHCR24* Lepidoptera section of the image represents the *B. mori DHCR24* gene lineage. (Continues next page)

The information present in "[b]" was adapted from Duque *et al.* (Duque *et al.* 2022a). CDS from *Clunio marinus* (Chironomidae) present in previous analyses (Duque *et al.* 2022a) were not available in the RefSeq datasets used for the newly performed phylogenetic inferences, and as such, the results regarding this species are excluded from the figure. Genome annotations for some species analysed in this work were not available in Duque *et al.* (Duque *et al.* 2022a), and thus information on the presence of *Regucalcin* homologs within certain taxonomic groups is represented as "not available" (n.a.).

3.2.2. VC colorimetric assays of *D. melanogaster* chromosomal deletion and RNAi stocks

3.2.2.1. *D. melanogaster* chromosomal deletion stocks screening

Chromosomal deletion stocks allow the identification of genes involved in closely related metabolic processes in *D. melanogaster* (Cook *et al.* 2012). Given this experimental potential, their use in conjugation with colorimetric VC assays could show a correlation between putative gene expression reductions associated with decreased gene dosage and low VC levels in this model organism. Bloomington Drosophila Stock Center (BDSC) deficiency kit stocks with cytologically defined chromosomal deletions (https://bdsc.indiana.edu/stocks/df/dfkit_retired_info.html) and VC colorimetric assays (e.g. EnzyChrom™; https://www.universalbiologicals.com/enzychromtm-ascorbic-acid-assay-kit-easc-100) were already commercially available, and thus a genomic screening could be conveniently performed in order to determine if VC synthesis in *D. melanogaster* was derived from a mostly conserved animal VC synthesis pathway or a yet unidentified pathway. Since these stocks can have heterogeneous backgrounds that may directly affect gene expression levels (Jin *et al.* 2001, Meiklejohn *et al.* 2003, Ryder *et al.* 2004), the interpretation of the results required caution. In an effort to reduce the impact of biologic variability in the analysis of colorimetric data, as a conservative approach, only stocks with VC absorbance values equal or below 50% of the Oregon-R control strain VC absorbance value (0,09 under 570nm) were initially considered as potential candidates for further experimentation. This criterion was applied to theoretically eliminate false positives from the final dataset, allowing a more accurate deficiency stock selection process.

Although the original objective was a wide genomic screening using 271 BDSC deficiency kit stocks, the persistent arrival of compromised stocks due to unknown disease or fungi growth limited the experimental execution. Nevertheless, a total of 185 chromosomal deletions stocks were analysed (Table S3.1), covering approximately 62% of the *D. melanogaster* genome (Figure 3.14A). Overall, the colorimetric assays revealed that over 80% of these stocks had lower VC levels than the Oregon-R strain used as control (Figure 3.14B).

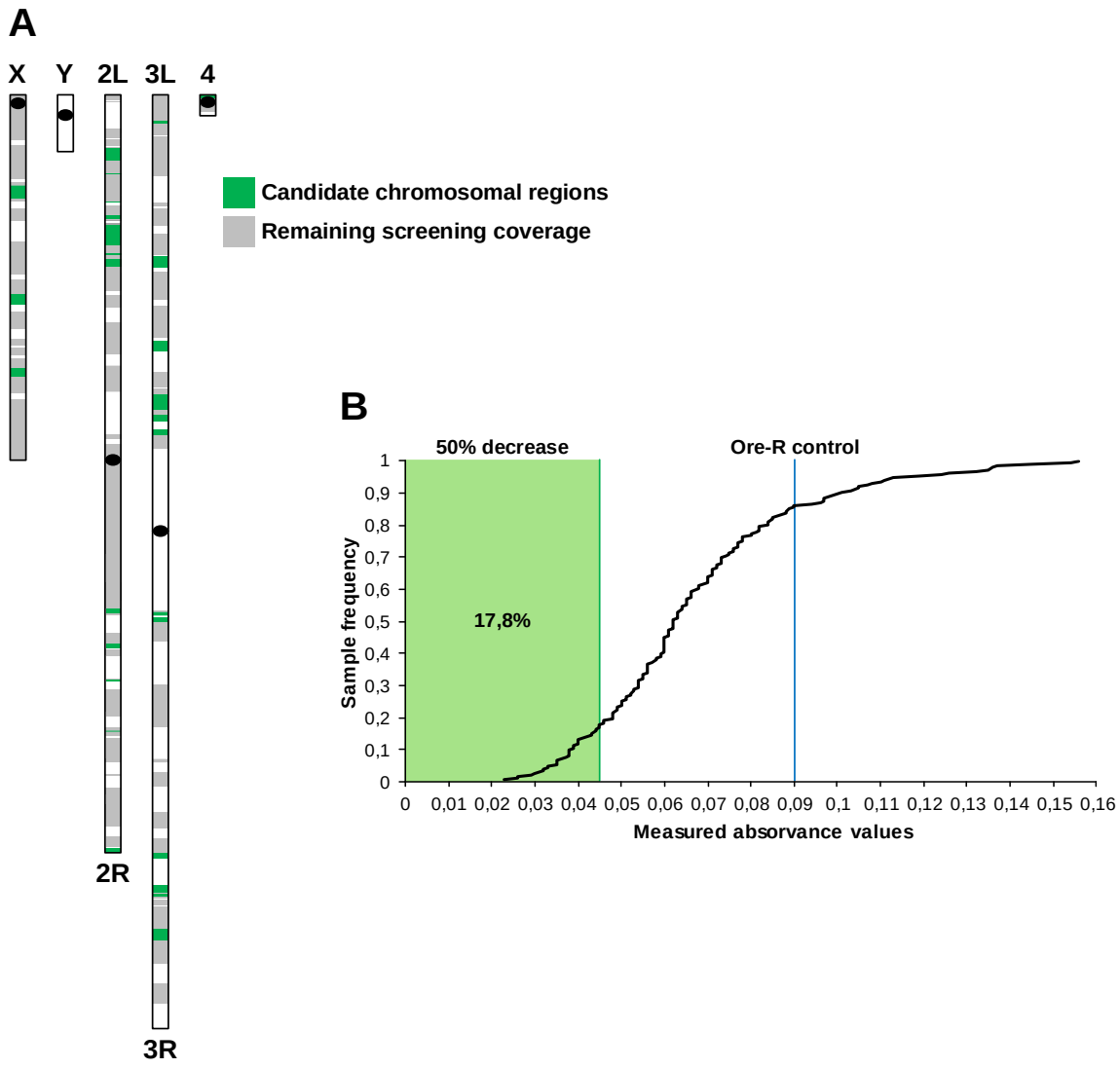


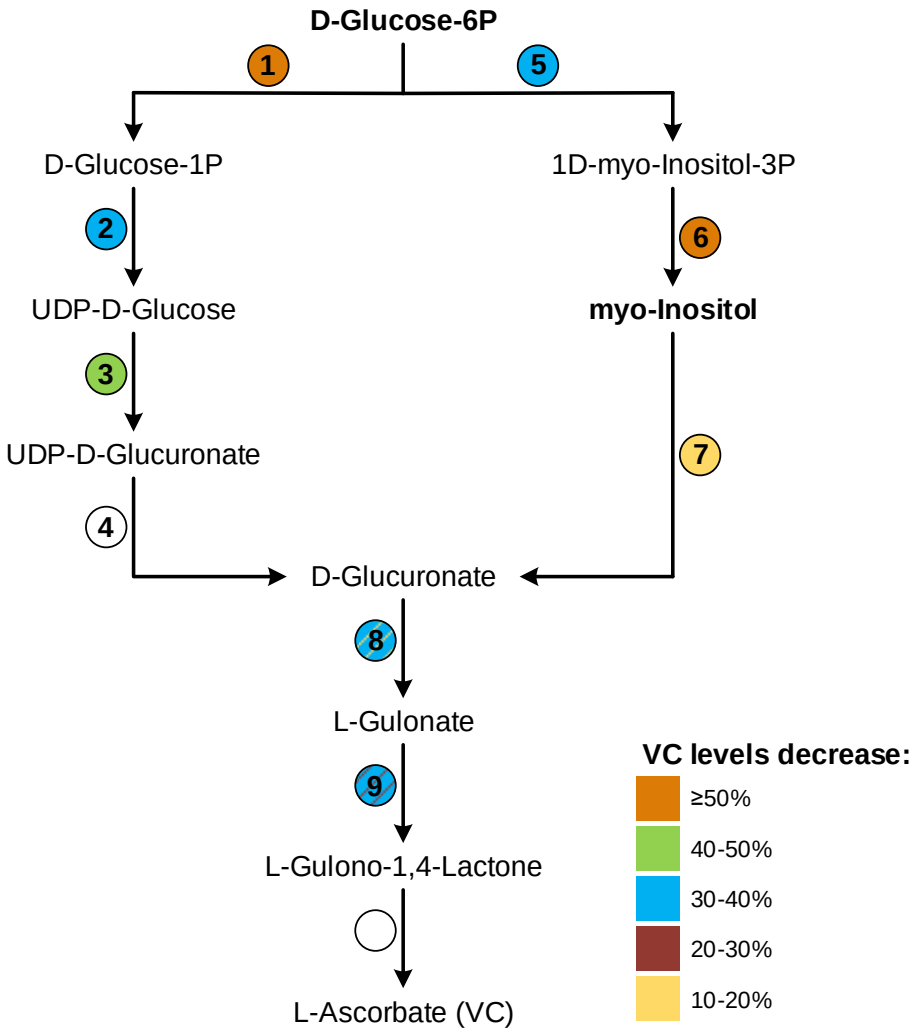
Figure 3.14. Distribution of chromosomal deletions and VC absorbance values of the analysed deficiency stocks. In panel A, the genomic coordinates of the analysed chromosomal deletions were mapped relative to the *D. melanogaster* karyotype. Chromosomal deletions associated with a 50% or higher decrease in total VC absorbance values are presented in green, while the remaining assayed genomic regions are highlighted in grey. The frequency of stocks that show lower, equal or higher VC absorbance values relative to the Oregon-R control strain (highlighted by the blue line) can be seen in panel B. The 50% and below cutoff in VC absorbance values (0,045) is shown in green.

After excluding the non-coding ribonucleic acids and pseudogenes contained within each analysed chromosomal deletion, the total number of genes subjected to experimental screening was 9193 (Table S3.2). From those, 1480 genes were partially deleted in 33 deficiency stocks that had VC absorbance values equal to or below 50% of the 0,09 from the Oregon-R strain (approximately 17,8% of the total number of stocks; Figure 3.14; Tables S3.1 and S3.3). This was a surprising result, since even with a conservative approach, the incidence of deletion stocks with considerably reduced VC levels was higher than the one

which would allow feasible gene candidate characterization within the time constraints of the delineated PhD project. However, it is highly unlikely that these findings reflect true correlation between reduced gene expression and directly reduced VC synthesis in all cases. That would imply that a minimum of 33 genes (one from each candidate deletion stock) was required for the production of this nutrient in *D. melanogaster*, hinting to the existence of a very long VC synthesis pathway or several considerably long synergistic VC synthesis pathways in this species. Alternatively, it was more parsimonious to assume that, in addition to the distinct genomic backgrounds of the utilized deficiency stocks, the extensive chromosomal deletions likely induced various degrees of pleiotropic effects in total VC levels. Independently of either conjecture, these unexpected results highlight that new methodological approaches and considerable data refinement are still required in order to conceivably assess the possible emergence of novel VC synthesis mechanisms in *D. melanogaster*.

Despite this limitation, the available data still provided interesting findings regarding the mostly conserved animal VC synthesis pathway and myo-inositol pathway hypotheses (Figure 3.15). With the exception of the *HEX 1 IMPA* homolog *CG9391*, the genes previously identified as candidates to play a role in VC synthesis in *D. melanogaster* were included in analysed chromosomal deletion stocks (Table 3.1). Interestingly, four stocks (1888, 2612, 2993 and 3686) even included more than one of these homologs, and were thus theoretically more susceptible to a drastic decline in total VC levels.

Overall, the colorimetric data suggested that all the candidate gene homologs analysed were contained within deficiency stocks with lower VC levels than the Oregon-R control strain. In this context, none should be excluded from further functional characterization efforts under the VC synthesis context. This was very relevant in the case of *Regucalcin* and *Dca*, as previous analyses could not determine if they played a role in this process (Duque *et al.* 2022a), but they could now be putatively related with observable reductions in total VC levels. From the 12 deficiency stocks considered, only 2993 and 3640 had a considerable decrease in VC levels under the conservative 50% reduction criterion. Curiously, the 2993 stock included a partial deletion of both *Pgm1* and four *HEX 2* IMPA homologs, *CG17026* to *CG17029*, while the 3640 stock only included the partial deletion of *CG17026* to *CG17029*. Assuming that the reduced dosage of these genes was directly responsible for the observed decrease in VC levels, some critical questions arose from these results.



1 - Pgm1 (EC 5.4.2.2)
2 - UGP (EC 2.7.7.9)
3 - sgl (EC 1.1.1.22)
4 - UGUR* (EC 3.2.1.-)
5 - Inos (EC 5.5.1.4)
6 - CG9391#, CG17026, CG17027, CG17028, CG17029 (EC 3.1.3.25)
7 - Miox (EC 1.13.99.1)
8 - Akr1B, CG6083, CG9436, CG10638, CG10863, CG12766 (EC 1.1.1.19)
9 - Regucalcin, Dca (EC 3.1.1.25)

Figure 3.15. Graphical representation of possible VC synthesis pathways in *D. melanogaster*. The numbered circles represent enzymes that may catalyse each pathway step, according to the number legend presented at the bottom of the figure. Each enzymatic step is associated with a VC level decrease according to the colour ranking presented. The stripped patterns exhibit the maximum and minimum decrease obtained when considering absorbance results for multiple gene homologs related to a given enzymatic step. UGUR is highlighted with "*" since remains to be identified and characterized. CG9391 is highlighted with "#" since no VC absorbance value was obtained in regards to the *G9391* gene. Given the lack of data regarding the identification of GULO substitutes in *D. melanogaster*, the last pathway step assuming the current hypotheses is represented by an empty circle.

Since the 2993 and 3640 stocks displayed similar decreases in total VC levels (58,9% and 50%, respectively), then the partial deletion of *CG17026*, *CG17027*, *CG17028* and

CG17029 could be responsible for the majority of the decrease in VC levels, thus suggesting that *Pgm1* had no significant contribution for the VC synthesis process.

Table 3.1. VC absorbance values of deletion stocks that include gene homologs related to possible VC synthesis pathways in *D. melanogaster*.

Vertebrate gene	*D. melanogaster* homologs	BDSC Deletion stock	VC absorbance value (570nm)	Absorbance variation (vs Oregon-R)
PGM1	*Pgm1*	2993	0,037	-58,9%
UGP2	*UGP*	997	0,063	-30,0%
		7079	0,06	-33,3%
UGDH	*sgl*	6867	0,052	-42,2%
AKR1B1	*Akr1B*	2612	0,078	-13,3%
	CG6083	2612	0,078	-13,3%
	CG9436	1007	0,072	-20,0%
		1888	0,056	-37,8%
	CG10638	5492	0,063	-30,0%
	CG10863	3686	0,065	-27,8%
	CG12766	3686	0,065	-27,8%
Regucalcin	*Regucalcin*	962	0,066	-26,7%
	Dca	29026	0,058	-35,6%
ISYNA1	*Inos*	1888	0,056	-37,8%
IMPA	*CG9391*	n.a.	n.a.	n.a.
	CG17026			
	CG17027	2993	0,037	-58,9%
	CG17028	3640	0,045	-50,0%
	CG17029			
MIOX	*Miox*	2612	0,078	-13,3%

This would imply that VC production derived from myo-inositol metabolism was the primary source of this nutrient in *D. melanogaster*. If so, then the modest VC absorbance reduction (13,3%) obtained for the deficiency stock that includes a partial deletion of *Miox*, 2612, was not concordant with this hypothesis. Moreover, this would also not explain why the deficiency stocks that include a partial deletion of *UGP* and *sgl* were associated with relatively high reductions in VC levels (over 30% in both cases). Nevertheless, as previously stated, deficiency stocks possess some experimental limitations, as they may include partial

deletions of other genes that can as well be contributing directly (e.g. have an unknown role in VC synthesis) or indirectly (e.g. have a role in oxidative stress response mechanisms) to the observed phenotype, consequently limiting the drawn conclusions. Thus, although this genomic screening was ultimately informative, further direct gene expression abolishment was required to address the role of *Pgm1*, *CG17026*, *CG17027*, *CG17028* and *CG17029* in *D. melanogaster*. For this purpose, a combination of GAL4-UAS RNAi gene suppression, Real-Time Polymerase Chain Reaction (RT-PCR) and VC colorimetric assays would likely provide important information.

3.2.2.2. *D. melanogaster* RNAi crossings screening

The RT-PCR primer design and availability of adequate RNAi stocks for *Pgm1* did not pose any problem regarding the chosen methodology. However, despite efforts to perform analyses on the four *HEX 2* IMPA homologs, several technical limitations occurred. Due to the high sequence identity between the *CG17026*, *CG17027*, *CG17028* and *CG17029* genes, the design of efficient primers following adequate thermodynamic stability and GC clamp to assure viability and specificity under the used RT-PCR conditions was only possible for *CG17029*. Moreover, even if efficient primers were designed for all genes, the VDRC RNAi stocks available for the *CG17026*, *CG17027* and *CG17028* genes present recognized gene suppression off-targets which would invalidate the direct correlation between gene suppression and any derived phenotypic effects (consult stock numbers 32813, 32819, 49199, 50075, 100100 and 103270 at https://stockcenter.vdrc.at). To surpass this technical issue, the RNAi suppression of *Miox* was selected as an alternative to directly address the contribution of myo-inositol metabolism for VC synthesis in *D. melanogaster*. Furthermore, given the role of D-Glucuronate as an intermediate substrate in both the characterized D-Glucose-6P and the possible myo-inositol VC synthesis pathways, the functional characterization of enzymes likely related with its conversion to L-Gulonate in *D. melanogaster* could indirectly highlight possible roles of Regucalcin and Dca in the production of VC. This would circumvent previous issues with the functional characterization of both proteins in this context (Duque *et al.* 2022a), and ultimately imply that at least a partial conservation of the characterized animal VC synthesis pathway is required for the process to happen in *D. melanogaster*. From the six *AKR1B1* homologs identified, so far only *CG10863* was found to encode for an enzyme with catalytic similarities to AKR1A1 and AKR1B1 (Di Luccio *et al.* 2006), and previously hypothesized to intervene in *D. melanogaster* VC synthesis (Duque *et al.* 2022b).

Unfortunately, *CG10863* RNAi gene suppression was completely lethal using this experimental setup (see “3.4. Materials and methods” for details). The associated lethality

was more accentuated than the one observed for *Regucalcin* and *Dca* knockdown (Duque *et al.* 2022a), as no progeny displaying the RNAi suppression phenotype was observed in 800 flies. Furthermore, similarly to what was observed for *Regucalcin* and *Dca*, *CG10863*-associated lethality was not reverted when maintaining the parental generation in food vials 50mM VC supplementation. Therefore, no further experiments were performed for this gene.

Under standard growth conditions, the RT-PCR data indicated that *Pgm1* and *Miox* RNAi-derived suppression was successfully achieved (*Pgm1*_7D and *Miox*_7D in Figure 3.16). Nevertheless, *Miox* knockdown was considerably higher (75,4%) than the knockdown of *Pgm1* (54,8%).

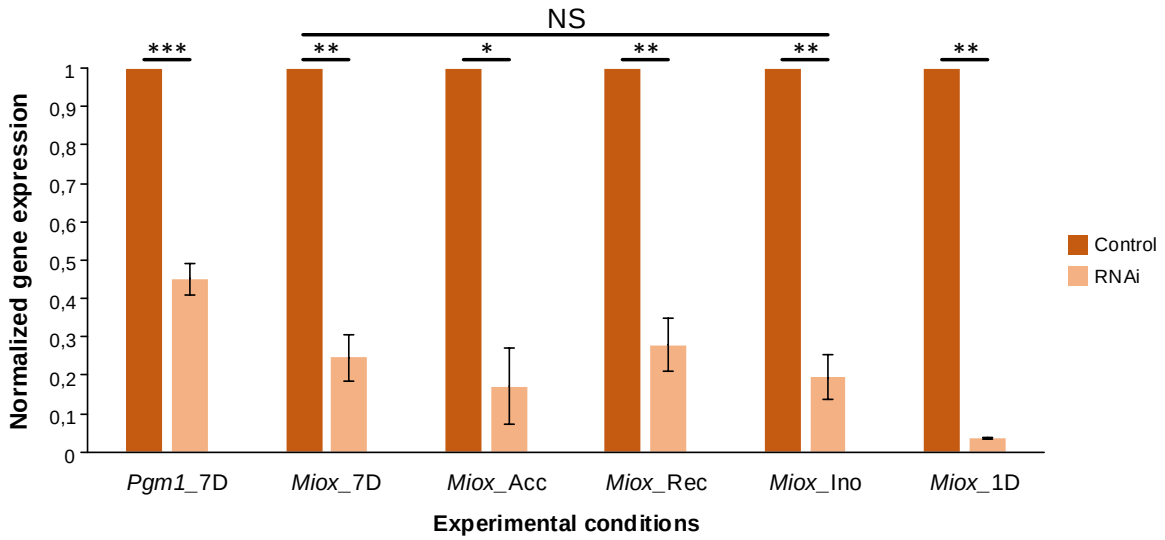


Figure 3.16. Results of the RT-PCR experiments performed for the *Pgm1* and *Miox* control and RNAi samples. Samples labelled as "Control" are representative of flies without RNAi expression, while the "RNAi" label concerns flies with RNAi-derived gene suppression. The experimental conditions acronyms represent fly samples from the following experimental setups: *Pgm1*_7D and *Miox*_7D: flies grown seven days under standard conditions (25ºC); *Miox*_Acc: flies grown seven days under standard conditions (25ºC) and after exposed to cold acclimation conditions (15ºC) for one day; *Miox*_Rec: *Miox*_Acc flies exposed to one day of recovery at 25ºC; *Miox*_Ino: flies grown to seven days under 25mM myo-inositol supplementation at 25ºC; *Miox*_1D: flies grown to 0-3h after birth under standard conditions (25ºC). The gene expression of control samples was normalized to 1 and included in the graph to facilitate the observation of gene expression change in RNAi samples. Gene suppression was achieved with statistical significance in all cases (*, ** and *** indicates p-values <0,05, <0,01 and <0,001 under a heteroscedastic two-tailed t-test, respectively).

This likely results from the differences in full-body expression of both genes, as available transcription data indicates that *Pgm1* is more expressed than *Miox* in adult wild-type flies (information available at https://flybase.org/reports/FBgn0003076.htm (*Pgm1*) and

https://flybase.org/reports/FBgn0036262.htm (*Miox*)). When comparing the RT-PCR information with data from VC colorimetric assays, it was possible to observe that only *Miox* supression was correlated with a modest but statistical significant 21% decrease in *D. melanogaster* VC levels (Figure 3.17). Although these results suggested that *Pgm1* does not contribute to VC synthesis in *D. melanogaster*, it is important to consider that the RNAi suppression of *Pgm1* was arguably low even when using a strong constitutively expressed *Actin 5C*-GAL4 driver, particularly considering that phenotypic effects are often expected at 70% or higher knockdown efficiency in *D. melanogaster* (Xia *et al.* 2021). Therefore, the unaffected VC levels are likely the result of insufficient *Pgm1* suppression, and as such, do not indicate the lack of contribution of this gene for VC synthesis in this organism. If so, then the results obtained from the chromosomal deletion stocks were not surprising, as the partial deletion of *Pgm1* would not likely affect total VC levels under normal experimental conditions. Ultimately, it was not possible to verify the role of *Pgm1* in VC synthesis under both experimental conditions.

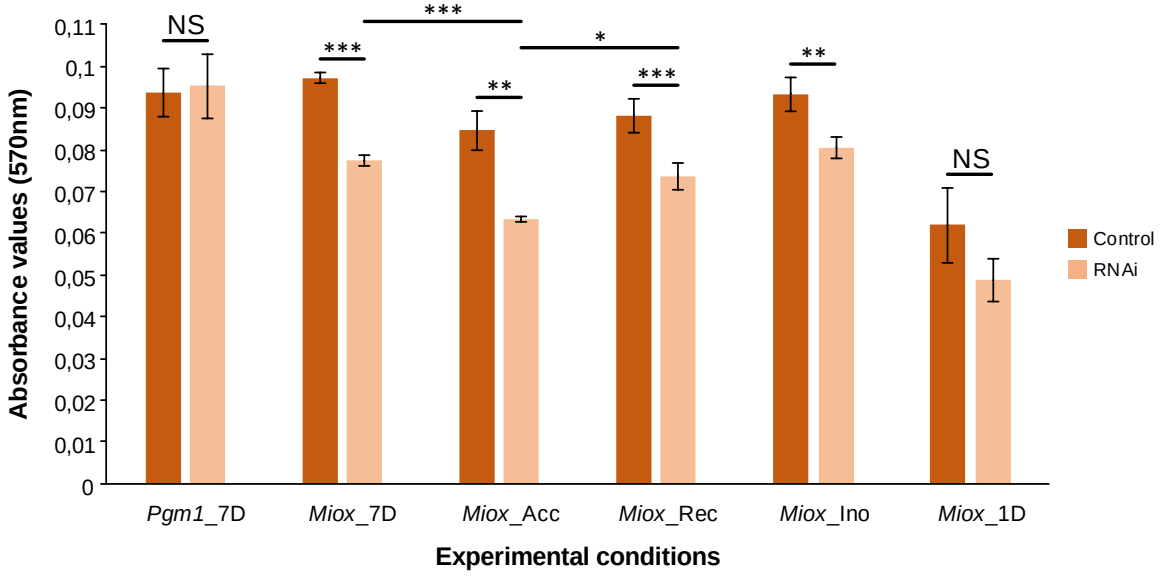


Figure 3.17. Results of colorimetric assays performed for the *Pgm1* and *Miox* control and RNAi samples. Samples labelled as "Control" are representative of flies without RNAi expression, while the "RNAi" label concerns flies with RNAi-derived gene suppression. The experimental conditions acronyms represent fly samples from the following experimental setups: *Pgm1*_7D and *Miox*_7D: flies grown seven days under standard conditions (25ºC); *Miox*_Acc: flies grown seven days under standard conditions (25ºC) and after exposed to cold acclimation conditions (15ºC) for one day; *Miox*_Rec: *Miox*_Acc flies exposed to one day of recovery at 25ºC; *Miox*_Ino: flies grown to seven days under 25mM myo-inositol supplementation at 25ºC; *Miox*_1D: flies grown to 0-3h after birth under standard conditions (25ºC). Statistical significance of VC level differences was only achieved in experimental conditions related to myo-inositol metabolism (** and *** indicates p-values <0,01 and <0,001 under a heteroscedastic two-tailed t-test, respectively).

The same, however, could not be said regarding *Miox*. Following on the results obtained for *Miox* RNAi suppression under standard growth conditions, additional experiments were performed to fully evaluate the contribution of this gene to the VC production process. Given that VC synthesis in *D. melanogaster* was inferred under recovery after cold acclimation exposure (Henriques *et al.* 2019), and the expression of *Miox* is reduced under cold acclimation in this species (MacMillan *et al.* 2016), an additional assay was performed to verify if *Miox* knockdown affected total VC levels under such experimental setup. Overall, the gap in VC levels between control and RNAi flies increased to 25% under acclimation conditions (*Miox*_Acc; Figure 3.17), and decreased to 17% after one day of recovery at 25ºC (*Miox*_Rec; Figure 3.17). Accordingly, *Miox* suppression varied from 82,9% to 72,1% between acclimation and recovery conditions, respectively, suggesting that, although not statistically different, the effectiveness of *Miox* knockdown could have contributed to the discrepancy in total VC values (Figure 3.16). These findings further highlighted that VC synthesis can be achieved through myo-inositol metabolism, and hinted that reduced *Miox* expression under cold acclimation conditions can possibly affect total VC levels in *D. melanogaster*. Two additional experiments were performed in order to verify if *Miox* overexpression led to increased total VC levels in flies, and if the 21% decrease in VC levels associated with *Miox* knockdown under standard growth conditions was influenced by VC accumulation during seven days, or alternatively, was effectively the contribution of myo-inositol for VC synthesis in *D. melanogaster*. Since *Miox* expression is known to be induced by myo-inositol supplementation (Contreras *et al.* 2023), increased Miox availability could lead to subsequent increases in VC levels. Nevertheless, the obtained RT-PCR data revealed that myo-inositol supplementation did not lead to *Miox* overexpression under these experimental conditions (*Miox*_Ino; Figure 3.16), and thus no conclusions could be drawn from the results. Interestingly, this condition was still the one that showed the smallest decrease in VC levels between control and *Miox* knockdown flies (14%; Figure 3.17). Considering that the half-life of VC can extend to 20 days in animals (Naidu 2003) and the supression of *Miox* was not complete in either of the conditions previously tested, it was possible that basal Miox-derived VC synthesis was leading to an tissue accumulation of VC throughout the lifetime of flies, before the performed colorimetric assays. If so, then total VC values in seven days or more flies would be biased and mask the effective role of *Miox* in the process. By analysing flies with 0 to 3h from birth (*Miox*_1D), it was possible to verify that this was not the case. In accordance with the reduced expression of *Miox* during early development of *D. melanogaster* relative to older life stages (data available at https://flybase.org/reports/FBgn0036262.htm), young flies exposed to RNAi displayed accentuated *Miox* knockdown (96,5%; Figure 3.16) and did not display statistically significant changes in VC levels relative to control individuals (Figure 3.17).

Interestingly, this experiment also showed that although myo-inositol metabolism contributes to VC synthesis in *D. melanogaster*, it likely only contributes to this process in adult stages.

Further analyses on *CG17029* RNAi-mediated suppression indicated that lower *CG17029* expression (27,3% relative to control; Figure S3.9A) does not affect total VC levels in *D. melanogaster* (Figure S3.9B), likely due to existing functional redundancy between this gene and the remaining three *HEX 2* IMPA homologs in this species.

3.3. Conclusion

The performed phylogenetic analyses revealed that not only *D. melanogaster*, but many hexapods should theoretically be capable of endogenous VC production using an almost completely conserved animal VC synthesis pathway. In addition, it was possible to observe that myo-inositol metabolism could also contribute to VC synthesis in these species. Nevertheless, DHCR24 functional adaptations seems to only be involved in the process in lepidopterans such as *B. mori*.

The use of chromosomal deletion stocks to address hypotheses regarding the preservation of the majority of the currently characterized animal VC synthesis pathway, or alternatively, the emergence of novel pathways in *D. melanogaster*, proved informative. Despite technical limitations, this approach highlighted *Pgm1* and *HEX 2 IMPA* homologs (*CG17026* to *CG17029*) as candidates for further functional characterization in this context.

RNAi-mediated suppression of *Pgm1* and *CG17029* and subsequent colorimetric assays could not implicate or exclude either gene from a role in VC synthesis. Nevertheless, this methodological approach allowed to verify that *Miox* was involved in the process, indicating that myo-inositol metabolism contributes to total VC levels in adult flies. However, further research is still required to clarify if the characterized D-Glucose-6P pathway is acting synergistically with the myo-inositol pathway for the purpose of VC synthesis in *D. melanogaster*. If not, then the gene candidate data obtained from the deficiency stocks can prove useful as a starting point in the determination of novel VC synthesis pathways in this organism.

3.4. Materials and methods

3.4.1. Phylogenetic analyses

An adaptation of the protocol described in the literature (López-Fernández *et al.* 2019) was used to obtain the sequence sets for the phylogenetic inferences. CDS were downloaded from the NCBI RefSeq database, and sequence operations performed using

the SEquence DAtaset builder (SEDA) software (López-Fernández *et al.* 2022). Due to the time lapse in the execution of the protocols, and subsequent cumulative deposit of new genome annotations in the NCBI RefSeq database, the phylogenetic analyses performed for the *PGM1*, *UGP2*, *UGDH* and *AKR1B1* genes were performed with 165 Hexapoda genomes, those performed for ISYNA1, *IMPA* and *MIOX* utilized 178 Hexapoda genomes, and the dataset for *DHCR24* contained 180 Hexapoda genomes.

The queries used for the *tblastn* searches were the human protein sequences for PGM1 (NP_002624.2), UGP2 (NP_006750.3), UGDH (NP_003350.1), AKR1B1 (NP_001619.1), ISYNA1 (NP_057452.1), IMPA1 (NP_005527.1), IMPA2 (NP_055029.1) and MIOX (NP_060054.4), as well as the *B. mori* DHCR24 sequence (XP_004926865.1). The BLAST algorithm version used was 2.7.1+ and the *tblastn* parameters selected included a 0.05 expectation value and a limitless number of BLAST hits to retrieve. The output files were further processed using SEDA's "NCBI rename" and “Merge” options to prefix the header of each of the retrieved sequences with the name of the species, common name, family and class and to include all results in a single file, respectively. Using the "Rename header” option, the sequence headers were altered to keep only the species name, common name, family, class and protein accession number. The file was refined using the “Pattern filtering” option and sequences with ambiguous nucleotides ([NRMSHDVYKWB]) were removed. Using the “Remove redundant sequences” and “Filtering” options, identical nucleotide sequences, sequences that are non-multiple of three, that do not have a valid start codon (ATG), and that have in frame stop codons were also removed. The sequences representative of the initial queries were reallocated to the first position in each of the files individually using the “Reallocate reference sequences” option, and then the “Filtering” operation was used to remove sequences with a size difference larger than 10% relative to these references to avoid badly annotated gene sequences. In the *AKR1B1* dataset this value was adapted to 2% to maintain only the sequences from the AKR1B protein family, since the *tblastn* output produced a high number of sequence hits that also represented distinct Aldo Keto Reductases with a lower degree of similarity to the query. Redundant sequence isoforms were after removed using adapted “Minimum word length” parameters according to the query sequence reference size in the “Remove isoforms” operation. Given the high overlap of *IMPA1* and *IMPA2 tblastn* hits, the datasets for both genes were merged and processed as one, labeled as *IMPA*.

The phylogenetic trees for the *PGM1*, *UGP2*, *UGDH*, *DHCR24, ISYNA1* and *MIOX* datasets were inferred using the Automatic Detection Of Positively Selected sites (ADOPS) (Reboiro-Jato *et al.* 2012) Docker image available at the Bioinformatics Docker Images Project (https://pegi3s.github.io/dockerfiles/), as well as Docker (https://www.docker.com/)

within a Linux virtualization in VirtualBox (https://www.virtualbox.org/). The phylogeny for the *AKR1B1* dataset was produced using the FastTree 2 software (Price *et al.* 2010) after a Clustal Omega (Sievers *et al.* 2011) alignment, using their respective Docker images (available at https://pegi3s.github.io/dockerfiles/). In the case of IMPA, the CDS were initially translated to protein sequences using SEDA and after aligned using the Clustal Omega Docker image. After, the protein alignment was converted back to nucleotide using the TranslatorX (Abascal *et al.* 2010) Docker image (https://hub.docker.com/r/pegi3s/translatorx/). This alignment file was converted to NEXUS format using the ALignment Transformation EnviRonment (Glez-Pena *et al.* 2010) Docker image (https://hub.docker.com/r/pegi3s/alter/) and analyzed using MrBayes (Ronquist *et al.* 2012) independently of the ADOPS pipeline (https://hub.docker.com/r/pegi3s/mrbayes/), with 2,000,000 requested tree generations and a defined burn-in of 25%. These datasets were processed using different methodologies due to computational issues with the high number of retrieved sequences from *tblastn* that turned further Bayesian analyses unfeasible (*AKR1B1*), and due to sequence alignment issues (*IMPA*).

When using the ADOPS pipeline, nucleotide sequences are first translated and aligned using the amino-acid alignment as a guide. MUSCLE (Edgar 2004) was used as the alignment algorithm as implemented in T-Coffee (Notredame *et al.* 2000). Only codons with a support value above two were used for phylogenetic reconstruction. MrBayes 3.1.2 (Ronquist *et al.* 2012) was used as the phylogenetic inference method, as implemented in the ADOPS pipeline. The general time-reversible model (GTR) of sequence evolution was implemented in the analysis, allowing for among-site rate variation and a proportion of invariable sites. Third codon positions were allowed to have a gamma distribution shape parameter different from that of first and second codon positions. Two independent runs of 1,500,000 generations with four chains each (one cold and three heated chains) were performed for all datasets with the exception of *ISYNA1*, which required 3,000,000 generations to achieve adequate convergence. The average standard deviation of split frequencies was always about 0.01 and the potential scale reduction factor for every parameter about 1.00 showing that convergence has been achieved. Trees were sampled every 100th generation with a defined burn-in of 25% for the complete analysis. The undiscarded trees were used to compute the Bayesian posterior probability values of each clade of the consensus tree. To achieve complete convergence and facilitate accurate phylogenetic inference, the *PGM1*, *UGP2*, *UGDH, DHCR24*, *ISYNA1, IMPA* and *MIOX* datasets were processed to remove sequence contaminants, as well as aberrant sequences due to unexpected high/low GC content. The *IMPA* dataset had many contaminants due to an unexpectedly high *tblastn* homology with a bisphosphate 3-prime-

nucleotidase (*BPNT1*) in particular, a gene involved in myo-inositol metabolism (Spiegelberg *et al.* 1999) outside the scope of the proposed VC synthesis pathway presented in this work. The removed sequences are listed in Table S3.4. The *UGDH* dataset was subdivided in two smaller ones, since the variable GC content did not allow for a convergent analysis. The Nexus format Bayesian trees produced by the ADOPS pipeline were converted to Newick format using the Format Conversion Website (http://phylogeny.lirmm.fr/phylo_cgi/data_converter.cgi). This Newick formatted file was imported to MEGA X (Kumar *et al.* 2018) in order to root the consensus phylogenetic tree using the outgroup vertebrate sequences.

For the *AKR1B1* dataset, the Clustal Omega and FastTree 2 were run with standard parameters, and the details on the phylogeny inference steps can be consulted in the FastTree 2 software manual (http://www.microbesonline.org/fasttree/#How).

3.4.2. *D. melanogaster* stock maintenance and crossings

The *Actin 5C-GAL4* driver stock (#25374) and the chromosomal deletion stocks (Table S3.1) were obtained from the BDSC. The RNAi stocks for *Pgm1* (#34954), *CG17029* (#49565), *Miox* (#103766) and *CG10863* (#103433) were supplied by the VDRC. Fly stocks were kept at environmental chambers with a constant temperature of 25ºC and 12h day/night cycles. Flies were reared on cornmeal food supplemented with yeast extract and depleted of VC.

The majority of standard driver/RNAi crosses were performed in the ♀ driver x ♂ RNAi direction with 8 individuals of each stock per cross, using cornmeal food vials kept at 25ºC and 12h day/night cycles. To verify if the chosen crossing direction was correlated with the fully lethal phenotype associated with *CG10863* suppression, crosses were also performed in the ♂ driver x ♀ RNAi direction for this gene. The progenitors were subsequently transferred to new vials every two days to maximize sample acquisition in all cases, until four vials were gathered. At the day of birth, the progeny was sorted according to the phenotype and gender. Most flies were kept under standard conditions until 7 days of age, before being frozen in liquid nitrogen and stored at -80ºC. In the case of one-day flies, samples were collected and preserved using the same methodology, within a 3h period after birth. In the cold acclimation and recovery experimental conditions, seven-day flies were exposed to 15ºC and 25ºC for 24h, respectively, while maintained in cornmeal food vials during 12h day/night cycles. In myo-inositol supplementation conditions, the ♀ driver x ♂ RNAi crossings were performed on cornmeal food vials with myo-inositol (25mM) (SERVA Electrophoresis GmbH, product number 26310.01). After sorting, the progeny was kept in new myo-inositol supplemented food vials during seven days, and subsequently

frozen in liquid nitrogen and stored at -80ºC. In order to verify if VC supplementation could reverse the lethal phenotype associated with *CG10863* suppression, additional crosses were performed in cornmeal food vials with L-Ascorbic acid (50mM) (Sigma-Aldrich, product number A5960).

Flies from the chromosomal deletion stocks were recovered from successive expansions of the original stocks using cornmeal food vials kept at 25ºC and 12h day/night cycles. A controlled parental population size (20-40 flies) was used to avoid progeny overcrowding in each case. The progenitors were transferred to new vials every two days until four vials were obtained. At the day of birth, the progeny was sorted according to gender. After, flies were kept under standard conditions until 7 days of age, before being frozen in liquid nitrogen and stored at -80ºC.

3.4.3. VC colorimetric assays

Samples of 25 female flies from the chromosomal deletion stocks and ♀ GAL4 driver x ♂ RNAi crossings were used for VC measurements along with the EnzyChrom™ Ascorbic Acid Assay Kit (https://www.universalbiologicals.com/enzychromtm-ascorbic-acid-assay-kit-easc-100). Samples were homogenized in 1ml of PBS (1X, pH 7.4) buffer in ice and transferred to 1.5ml tubes that were centrifuged for 5 minutes at 13300 rotations per minute (RPM) (4ºC) using a VWR® Micro Star 17R centrifuge. 700µl of the resultant supernatants were then transferred to new 1.5ml tubes and centrifuged 15 minutes at 13300 RPM (4ºC). 20µL of the resulting supernatants were transferred to a Thermo Scientific™ Nunc™ MicroWell™ 96-Well Microplate and mixed with 80µL of the EnzyChrom™ Ascorbic Acid Assay Kit reaction mix (constituted by 85µL of assay buffer, 1µL of enzyme mix and 1µL of dye reagent per sample). The mixtures were incubated for 10 minutes at 25ºC with occasional agitation, and were then placed in the BioTek Synergy™ 2 Multi-Mode Microplate Reader for absorbance measurements at 570nm wavelength. To validate the measurements, each assay was performed using 20µL of two VC standards (1000µM and 500µM) prepared using the 10mM standard provided by the kit and type 2 distilled water, mixed with 80µL of the reaction mix. The final VC absorbance values were obtained by subtracting the absorbance of white microplate wells (80µL of reaction mix plus 20µL PBS (1X, pH 7.4) without sample) from total absorbance values.

3.4.4. Gene expression analyses

Pgm1, Miox and *CG17029* expression levels were determined using three biological samples of three female flies displaying the curly (control) and straight (RNAi) wing phenotypes for all reported experimental conditions, respectively. Total RNA was isolated using the homogenization of biological samples in 1ml TRIzol™ Reagent (Invitrogen™) in

1.5ml tubes. The homogenates were initially agitated for five minutes at room temperature (23-25ºC), and after centrifuged at 5000 RPM (4ºC) for 10 minutes using a VWR® Micro Star 17R centrifuge. The supernatants were transferred to new 1.5ml tubes and mixed with 200µl of chloroform, and after agitated for three minutes at room temperature (23-25ºC) and centrifuged at 5000 RPM (4ºC) for 15 minutes. This step was performed an additional time with the obtained supernatants. The resulting supernatants were after mixed with 200µl of chloroform in new 1.5 ml tubes, agitated for three minutes at room temperature (23-25ºC), and the mixtures were transferred to 2ml 5PRIME Phase Lock Gel™ tubes to be centrifuged at 5000 RPM (4ºC) for 15 minutes. The supernatants were transferred to new 1.5ml tubes, mixed with 5µl of glycogen, agitated for three minutes at room temperature (23-25ºC), mixed with 500µl of isopropanol, agitated for ten minutes at room temperature (23-25ºC), and centrifuged at 7000 RPM (4ºC) for 10 minutes. The supernatants were removed from the 1.5ml tubes, and the resulting pellets were washed two times with 1ml of ethanol (75% V/V, prepared with nuclease-free water (Invitrogen™)) using two centrifugations at 5000 RPM (4ºC) for 5 minutes, left to dry during 10 minutes, and resuspended in 21,5µl of nuclease-free water (Invitrogen™). The samples were treated with the TURBO DNA-free™ kit (Invitrogen™) according to the manufacturer's instructions (see manual in https://www.thermofisher.com/order/catalog/product/AM1907) to remove possible genomic DNA contaminations. RNA concentration and integrity were measured using an Experion™ StdSens (Bio-Rad) protocol. cDNA was synthesized by reverse transcription using the SuperScript® III First-Strand Synthesis System for RT-PCR (Invitrogen™) according to the manufacturer's indications https://tools.thermofisher.com/content/sfs/manuals/superscriptIIIfirststrand_pps.pdf), using negative template and reverse transcriptase controls to validate the cDNA synthesis process. Efficient RT-PCR primers were designed for *Pgm1* (Forward - 5' TCCACACCACTCCACTGCCCG 3'; Reverse - 5' TTACGAAGAACGCCTTGCTGCC 3') and *CG17029* (Forward - 5' CAAACTTCATTCACCGCATCCCG 3'; Reverse - 5' CCATTCAGATAGGCTCCGTGTCCC 3'), using the OLIGO 4.0 software (https://www.oligo.net/index.html). Primers for *Miox* and the reference gene *ribosomal protein L32* (*RpL32*) were used as reported in the literature (Contreras *et al.* 2023, Reis *et al.* 2011). RT-PCR was performed with the iQ™ SYBR® Green Supermix (Bio-Rad) according to the manufacturer's instructions (https://www.bio-rad.com/webroot/web/pdf/lsr/literature/10016680.pdf) on a CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad) with the following program: 3min at 95°C; 40 cycles of 30s at 94°C, 30s at 56°C and 30s at 72°C followed by a standard melt curve. Change in gene expression was calculated using the $2^{-\Delta\Delta CT}$ method (Schmittgen and Livak 2008).

3.5. References

Abascal, F., R. Zardoya and M. J. Telford. **2010.** TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*. 38:W7-W13.

Arboleda-Bustos, C. E. and C. Segarra. **2011.** The *Dca* Gene Involved in Cold Adaptation in *Drosophila melanogaster* Arose by Duplication of the Ancestral *regucalcin* Gene. *Molecular Biology and Evolution*. 28:2185-2195.

Beck, W. **1979.** Differential function of the phosphoglucomutase isozymes PGM1 and PGM2. *Human Genetics*. 50:93-100.

Brown, P. M., T. T. Caradoc-Davies, J. M. Dickson, G. J. Cooper, K. M. Loomes and E. N. Baker. **2006.** Crystal structure of a substrate complex of myo-inositol oxygenase, a di-iron oxygenase with a key role in inositol metabolism. *Proceedings of the National Academy of Sciences*. 103:15032-15037.

Caspi, R., R. Billington, I. M. Keseler, A. Kothari, M. Krummenacker, P. E. Midford, W. K. Ong, S. Paley, P. Subhraveti and P. D. Karp. **2020.** The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Research*. 48:D445-d453.

Chang, H.-H., H.-N. Chao, C. S. Walker, S.-Y. Choong, A. Phillips and K. M. Loomes. **2015.** Renal depletion of myo-inositol is associated with its increased degradation in animal models of metabolic disease. *American Journal of Physiology-Renal Physiology*. 309:F755-F763.

Contreras, A., M. K. Jones, E. D. Eldon and L. S. Klig. **2023.** Inositol in Disease and Development: Roles of Catabolism via myo-Inositol Oxygenase in *Drosophila melanogaster*. *International Journal of Molecular Sciences*. 24:4185.

Cook, R. K., S. J. Christensen, J. A. Deal, R. A. Coburn, M. E. Deal, J. M. Gresens, T. C. Kaufman and K. R. Cook. **2012.** The generation of chromosomal deletions to provide extensive coverage and subdivision of the *Drosophila melanogaster* genome. *Genome Biology*. 13:R21.

Di Luccio, E., R. A. Elling and D. K. Wilson. **2006.** Identification of a novel NADH-specific aldo-keto reductase using sequence and structural homologies. *Biochemical Journal*. 400:105-114.

Drouin, G., J.-R. Godin and B. Pagé. **2011.** The Genetics of Vitamin C Loss in Vertebrates. *Current Genomics*. 12:371-378.

Duque, P., C. P. Vieira, B. Bastos and J. Vieira. **2022a.** The evolution of vitamin C biosynthesis and transport in animals. *BMC Ecology and Evolution*. 22:84.

Duque, P., C. P. Vieira and J. Vieira. **2022b.** Advances in Novel Animal Vitamin C Biosynthesis Pathways and the Role of Prokaryote-Based Inferences to Understand Their Origin. *Genes*. 13:1917.

Dutta Gupta, S., P. K. Choudhury and I. B. Chatterjee. **1973.** Synthesis of l-ascorbic acid from d-glucurono-1,4-lactone conjugates by different species of animals. *International Journal of Biochemistry*. 4:309-314.

Edgar, R. C. **2004.** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 32:1792-1797.

Gabbay, K. H., K. M. Bohren, R. Morello, T. Bertin, J. Liu and P. Vogel. **2010.** Ascorbate synthesis pathway: dual role of ascorbate in bone homeostasis. *Journal of Biological Chemistry*. 285:19510-19520.

Glez-Pena, D., D. Gómez-Blanco, M. Reboiro-Jato, F. Fdez-Riverola and D. Posada. **2010.** ALTER: program-oriented conversion of DNA and protein alignments. *Nucleic Acids Research*. 38:W14-W18.

Hänninen, O., R. Raunio and J. Marniemi. **1971.** Biosynthesis of L-ascorbic acid from myo-inositol. *Carbohydrate Research*. 16:343-351.

Henriques, S. F., P. Duque, H. López-Fernández, N. Vázquez, F. Fdez-Riverola, M. Reboiro-Jato, C. P. Vieira and J. Vieira. **2019.** Multiple independent *L-gulonolactone oxidase* (*GULO*) gene losses and vitamin C synthesis reacquisition events in non-Deuterostomian animal species. *BMC Evolutionary Biology*. 19:126.

Hou, M., T. M. Dai, X. Y. Liang, S. X. Zhang, W. Z. Cui, J. F. Qiu, Y. H. Sima, W. Z. Cui and S. Q. Xu. **2021.** *Bombyx mori* can synthesize ascorbic acid through the l-gulose pathway to varying degrees depending on developmental stage. *Archives of Insect Biochemistry and Physiology*. 106:e21783.

Jin, W., R. M. Riley, R. D. Wolfinger, K. P. White, G. Passador-Gurgel and G. Gibson. **2001.** The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nature Genetics*. 29:389-395.

Jing, X. and S. T. Behmer. **2020.** Insect sterol nutrition: physiological mechanisms, ecology, and applications. *Annual Review of Entomology*. 65:251-271.

Kanehisa, M. and S. Goto. **2000.** KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*. 28:27-30.

Kumar, S., G. Stecher, M. Li, C. Knyaz and K. Tamura. **2018.** MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*. 35:1547-1549.

Lavrynenko, O., J. Rodenfels, M. Carvalho, N. A. Dye, R. Lafont, S. Eaton and A. Shevchenko. **2015.** The ecdysteroidome of *Drosophila*: influence of diet and development. *Development*. 142:3758-3768.

Linster, C. L. and E. Van Schaftingen. **2007.** Vitamin C. Biosynthesis, recycling and degradation in mammals. *The FEBS Journal*. 274:1-22.

López-Fernández, H., P. Duque, S. Henriques, N. Vázquez, F. Fdez-Riverola, C. P. Vieira, M. Reboiro-Jato and J. Vieira. **2019.** A Bioinformatics Protocol for Quickly Creating Large-Scale Phylogenetic Trees. In: F. Fdez-Riverola, M. S. Mohamad, M. Rocha, J. F. De Paz and P. González, editors. *Practical Applications of Computational Biology and Bioinformatics, 12th International Conference.* Cham. Springer. p. 88-96.

López-Fernández, H., P. Duque, N. Vázquez, F. Fdez-Riverola, M. Reboiro-Jato, C. P. Vieira and J. Vieira. **2022.** SEDA: A Desktop Tool Suite for FASTA Files Processing. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 19:1850-1860.

Lu, Y. and A. Agarwal. **2022.** Myo-inositol oxygenase in cadmium-induced kidney injury. *American Journal of Physiology Renal Physiology*. 322:F470-F472.

MacMillan, H. A., J. M. Knee, A. B. Dennis, H. Udaka, K. E. Marshall, T. J. S. Merritt and B. J. Sinclair. **2016.** Cold acclimation wholly reorganizes the *Drosophila melanogaster* transcriptome and metabolome. *Scientific Reports*. 6:28999.

Maddison, D. R., K.-S. Schulz and W. P. Maddison. **2007.** The tree of life web project. *Zootaxa*. 1668:19-40.

Maliekal, P., T. Sokolova, D. Vertommen, M. Veiga-da-Cunha and E. Van Schaftingen. **2007.** Molecular Identification of Mammalian Phosphopentomutase and Glucose-1,6-bisphosphate Synthase, Two Members of the α-D-Phosphohexomutase Family. *Journal of Biological Chemistry*. 282:31844-31851.

Massie, H. R., M. E. Shumway, S. J. P. Whitney, S. M. Sternick and V. R. Aiello. **1991.** Ascorbic acid in *Drosophila* and changes during aging. *Experimental Gerontology*. 26:487-494.

Meiklejohn, C. D., J. Parsch, J. M. Ranz and D. L. Hartl. **2003.** Rapid evolution of male-biased gene expression in *Drosophila*. *Proceedings of the National Academy of Sciences*. 100:9894-9899.

Muenks, A. G., K. M. Stiers and L. J. Beamer. **2017.** Sequence-structure relationships, expression profiles, and disease-associated mutations in the paralogs of phosphoglucomutase 1. *PLoS One*. 12:e0183563.

Naidu, K. A. **2003.** Vitamin C in human health and disease is still a mystery? An overview. *Nutrition Journal*. 2:7.

Notredame, C., D. G. Higgins and J. Heringa. **2000.** T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*. 302:205-217.

O'Grady, P. M. and R. DeSalle. **2018.** Phylogeny of the Genus *Drosophila*. *Genetics*. 209:1-25.

Patananan, A. N., L. M. Budenholzer, M. E. Pedraza, E. R. Torres, L. N. Adler and S. G. Clarke. **2015.** The invertebrate *Caenorhabditis elegans* biosynthesizes ascorbate. *Archives of Biochemistry and Biophysics*. 569:32-44.

Peters, R. S., L. Krogmann, C. Mayer, A. Donath, S. Gunkel, K. Meusemann, A. Kozlov, L. Podsiadlowski, M. Petersen, R. Lanfear, P. A. Diez, J. Heraty, K. M. Kjer, S. Klopfstein, R. Meier, C. Polidori, T. Schmitt, S. Liu, X. Zhou, T. Wappler, J. Rust, B. Misof and O. Niehuis. **2017.** Evolutionary History of the Hymenoptera. *Current Biology*. 27:1013-1018.

Price, M. N., P. S. Dehal and A. P. Arkin. **2010.** FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One*. 5:e9490.

Reboiro-Jato, D., M. Reboiro-Jato, F. Fdez-Riverola, C. P. Vieira, N. A. Fonseca and J. Vieira. **2012.** ADOPS-Automatic detection of positively selected sites. *Journal of Integrative Bioinformatics*. 9:18-32.

Reis, M., C. P. Vieira, R. Morales-Hojas, B. Aguiar, H. Rocha, C. Schlötterer and J. Vieira. **2011.** A Comparative Study of the Short Term Cold Resistance Response in Distantly Related *Drosophila* Species: The Role of *regucalcin* and *Frost*. *PLOS ONE*. 6:e25520.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard and J. P. Huelsenbeck. **2012.** MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*. 61:539-542.

Ryder, E., F. Blows, M. Ashburner, R. Bautista-Llacer, D. Coulson, J. Drummond, J. Webster, D. Gubb, N. Gunton, G. Johnson, C. J. O'Kane, D. Huen, P. Sharma, Z. Asztalos, H. Baisch, J. Schulze, M. Kube, K. Kittlaus, G. Reuter, P. Maroy, J. Szidonya, A. s. Rasmuson-Lestander, K. Ekström, B. Dickson, C. Hugentobler, H. Stocker, E. Hafen, J. A. Lepesant, G. Pflugfelder, M. Heisenberg, B. Mechler, F. Serras, M. Corominas, S. Schneuwly, T. Preat, J. Roote and S. Russell. **2004.** The DrosDel Collection: A Set of P-Element Insertions for Generating Custom Chromosomal Aberrations in *Drosophila melanogaster*. *Genetics*. 167:797-813.

Schmittgen, T. D. and K. J. Livak. **2008.** Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*. 3:1101-1108.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson and D. G. Higgins. **2011.** Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*. 7:539.

Spiegelberg, B. D., J. P. Xiong, J. J. Smith, R. F. Gu and J. D. York. **1999.** Cloning and characterization of a mammalian lithium-sensitive bisphosphate 3'-nucleotidase inhibited by inositol 1,4-bisphosphate. *The Journal of Biological Chemistry*. 274:13619-13628.

Svoboda, J. A. and M. F. Feldlaufer. **1991.** Neutral sterol metabolism in insects. *Lipids*. 26:614-618.

Thorsell, A.-G., C. Persson, N. Voevodskaya, R. D. Busam, M. Hammarström, S. Gräslund, A. Gräslund and B. M. Hallberg. **2008.** Structural and biophysical characterization of human myo-inositol oxygenase. *Journal of Biological Chemistry*. 283:15209-15216.

Van Der Linde, K. I. M., D. Houle, G. S. Spicer and S. J. Steppan. **2010.** A supermatrix-based molecular phylogeny of the family Drosophilidae. *Genetics Research*. 92:25-38.

Wheeler, G., T. Ishikawa, V. Pornsaksit and N. Smirnoff. **2015.** Evolution of alternative biosynthetic pathways for vitamin C following plastid acquisition in photosynthetic eukaryotes. *eLife*. 4:e06369.

Wiegmann, B. M., M. D. Trautwein, I. S. Winkler, N. B. Barr, J.-W. Kim, C. Lambkin, M. A. Bertone, B. K. Cassel, K. M. Bayless, A. M. Heimberg, B. M. Wheeler, K. J. Peterson, T. Pape, B. J. Sinclair, J. H. Skevington, V. Blagoderov, J. Caravas, S. N. Kutty, U. Schmidt-Ott, G. E. Kampmeier, F. C. Thompson, D. A. Grimaldi, A. T. Beckenbach, G. W. Courtney, M. Friedrich, R. Meier and D. K. Yeates. **2011.** Episodic radiations in the fly tree of life. *Proceedings of the National Academy of Sciences*. 108:5690.

Xia, S., N. W. VanKuren, C. Chen, L. Zhang, C. Kemkemer, Y. Shao, H. Jia, U. Lee, A. S. Advani, A. Gschwend, M. D. Vibranovski, S. Chen, Y. E. Zhang and M. Long. **2021.** Genomic analyses of new genes and their phenotypic effects reveal rapid evolution of essential functions in *Drosophila* development. *PLOS Genetics*. 17:e1009654.

Yabuta, Y., R. Nagata, Y. Aoki, A. Kariya, K. Wada, A. Yanagimoto, H. Hara, T. Bito, N. Okamoto, S. Yoshida, A. Ishihara and F. Watanabe. **2020.** L-Ascorbate Biosynthesis Involves Carbon Skeleton Rearrangement in the Nematode *Caenorhabditis elegans*. *Metabolites*. 10:334.

Yang, J. and T. Warnow. **2011.** Fast and accurate methods for phylogenomic analyses. *BMC Bioinformatics*. 12:S4.

Zhang, S. Q., L. H. Che, Y. Li, L. Dan, H. Pang, A. Ślipiński and P. Zhang. **2018.** Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. *Nature Communications*. 9:205.

3.6. Supplementary material

Table S3.1. Utilized BDSC deletion stocks experimental information.

BDSC Deletion stock	VC absorbance value (570nm)	Deletion cytology	Genome coordinates
125	0,137	14B8:14C1	X:16363428-16363994
140	0,097	28E1:28E5	2L:8071609;2L:8110561
190	0,113	47D3:48B2	2R:11155185;11644257
198	0,078	43F:44D3-8	2R:7861514..8660244
201	0,07	44D1-4:44F12	2R:8479476;8993811
282	0,082	58D1-2:59A	2R:22193486;22790099
420	0,066	36C2-4:37B9-C1	2L:17436830;2L:19123869
439	0,108	61C5-8:62A8	3L:563756..1622626
463	0,089	63E6-9;64A8-9	3L:3651164..4245410
490	0,077	25F3-26A1:26D3-11	2L:5682522..6514155
600	0,062	62A10-B1:62D2-5	3L:1651548..2310495
669	0,053	99A1-2:99B6-11	3R:29184632..29775722
693	0,084	24C2-8;25C8-9	3L:3726432..5164742
739	0,077	41A:41A	2R:1..4582213
749	0,088	h42-h43;42A2-3	2R:1512810..5929207
756	0,078	88F9-89A1:89B9-10	3R:11598111..16276804
757	0,056	55E2-4:56C1-11	2R:18664328..19187938
781	0,073	25D2-4:26B2-5	2L:5240240..6110195
823	0,055	97E3:98A5	3R:27093906;3R:27532761
936	0,076	3E1-2;3C2	X:2173103..2907914
939	0,06	3C11:3E4	X:3255980..X:3676945
940	0,056	3D6-E1;4F5	X:3554315..5384034
945	0,054	5A8-9;5C5-6	X:5613782..5878554
946	0,084	5C2:5D5-6	X:5766013..6181605
949	0,088	7D1:7D5-6	X:7901331;X:8115848
952	0,051	8E:9C-D	X:9495619..10460894
954	0,06	9B1-2;10A1-2	X:10055611..11002008
962	0,066	10F7;11D1	X:11920399..12911046
967	0,04	11D-E;12A1-2	X:12836467..13497993
970	0,082	17A1:18A2	X:18168380..19079978
971	0,103	18A5;18D	X:19110894..19204912
972	0,054	18E1-2;20	X:19676127..23108331
977	0,094	19F1-2;20E-F	X:20939503;X:23542271
997	0,063	67A2;67D11-13	3L:9190597..10316395
998	0,065	12D2-E1;13A2-5	X:13990704..15059747
1007	0,072	42A1-2;42E6-F1	2R:5798246;2R:7009943
1045	0,111	30D-30F;31F	2L:9795948..10557961
1145	0,075	48A3-4;48C6-8	2R:11580032..11909032
1329	0,136	1A1;2A	X:103615..1475668
1420	0,081	65F3:66B10	3L:7356061;3L:8141044
1491	0,074	35D1;36A6-7	2L:15271254..16485982
1541	0,097	66B8-9;66C9-10	3L:8135912..8395475
1546	0,049	1B2-1A3;A3	X:1;X:2480289
1547	0,073	55A:55F	2R:17890760..18782915
1567	0,069	23C1-2;23E1-2	2L:2875793..3298921
1743	0,077	46A:46C	2R:9637539..9988605
1785	0,061	101F1;102F8	4:49043..1098782
1884	0,055	84A1-2;84B1-2	3R:6625559..7089239
1888	0,056	42B3-5;43E1S-18	2R:6651936..7861513
1910	0,048	97A:98A1-2	3R:26559054;3R:27480200
1920	0,07	88B5;89C1-7	3R:16145093;3R:16480326
1990	0,062	83C1-2;84B1-2	3R:5551168;3R:7086599
2052	0,057	77A1:77D1	3L:20168063..20567591
2363	0,054	95F7:96A7-18	3R:24253317;3R:24761404
2393	0,099	83E1-2:84A5	3R:6192284..6850227
2400	0,11	62B4-7:62D5-E5	3L:1861374;3L:2537148
2577	0,054	61A:61D3	3L:10001..890468
2604	0,08	60C5-6;60D9-10	2R:24268358;2R:24612892
2608	0,09	75A6-7:75C1-2	3L:17865330..18262608
2612	0,078	68C8-11:69B4-5	3L:11431881..12421235
2892	0,035	29C1-2;30C8-9	2L:8369332;2L:9806483
2990	0,073	75B8:75F1	3L:18040889;3L:18841891
2993	0,037	72C1-D1;73A3-4	3L:15957888..16539625
3003	0,055	86E2-4:87C6-7	3R:11246530;3R:12675277
3011	0,082	90F1-F4:91F5	3R:18118268..19177589
3079	0,086	32F1-3;33F1-2	2L:11437223;2L:12626238
3124	0,076	70C1-2:70D4-5	3L:13592111..14381025
3126	0,068	70D2-3:71E4-5	3L:14074501..15674313
3127	0,085	77B-C:77F-78A	3L:20376442..20920517
3128	0,065	86C1:87B1-5	3R:10873773;3R:12350415
3217	0,054	14B13;15A9	X:16328437..16806047
3221	0,132	7B2-4;7C3-4	X:7337288..7933660
3347	0,063	19F1:14B1	X:15795007..16175032
3467	0,06	56F9-17;57D11-12	2R:21089480;2R:21496204
3468	0,145	96A2-7:98D2-4	3R:24420761..25411217
3518	0,154	51D3-8;52F5-9	2R:14796263..16211184
3520	0,101	52F5-9;52F10-53A1	2R:16059113;2R:16339585
3547	0,066	99B5-6:99F1	3R:29468941;3R:30495521
3560	0,096	9F:10C3-5	X:10781791..11559989
3588	0,055	35B4-6:35F1-7	2L:14648532..16125490
3591	0,124	44F10:45D9-E1	2R:8861340;2R:9524270
3617	0,08	76B1-2:76D5	3L:19353707..19938587
3638	0,058	21A1:21B7-8	2L:7525..324485
3640	0,045	71F1-4:72D1-10	3L:15860743..16099816
3649	0,105	63C2:63F7	3L:3303480..3935599
3650	0,119	63F:63D	3L:2674593;3L:3338860
3686	0,065	63F6-7:64C13-15	3L:3910034;3L:5235796
3714	0,1	20A:20F	X:21377456..23108331
4741	0,06	15D1:16A4-6	X:17020532..17469753
4787	0,053	82F3-4:82F10-11	3R:5168562..5364155
4953	0,156	16A2:16C7-10	X:17198399..17776521
4956	0,071	27E2:28D1	2L:7117729..7826302
4959	0,065	h35;h38L	2L:22536884;2L:23513712
4960	0,098	48E:49A	2R:12200815;2R:12530888
4961	0,038	60F1:60F5	2R:25052331;2R:25286936
4962	0,062	92B3;92F13	3R:19815663..20794570
4966	0,068	45A6-7:45E2-3	2R:9160438..9524769
5126	0,043	76B4;77B	3L:19486521..20297553
5246	0,071	57D2-8:58D1	2R:21214877..22212886
5281	0,039	5C3-10;5C3-12	X:5848082;X:6674387
5330	0,044	24A2:24D4	2L:3606756;2L:4198404
5420	0,045	27C2-3;27C4-5	2L:6880270..6899661
5492	0,063	69A4-5;69D4-6	3L:12265641..12736889
5574	0,049	54B16;54B16	2R:17448011..17481670
5601	0,035	96F1;97B1	3R:25712017..26496050
5680	0,035	54B17-C4;54C1-4	2R:17456472;2R:17457129
5694	0,038	82F8-10;83A1-3	3R:5260969..5452637
5869	0,062	32D1:32F1-3	2L:11021966..11650043
5878	0,042	77E2-4:78A2-4	3L:20644926..21030467
5879	0,084	48E12-F4;49A1-B6	2R:12232476..12582293
5951	0,048	79D3-E1;79F3-6	3L:22261962..22718905
6217	0,041	16C:16F	X:17678211..18188379
6299	0,05	26B1-2;26D1-2	2L:5980165..6448047
6338	0,107	26D3-E1;26F4-7	2L:6411355;2L:6664150
6374	0,066	26D10-E1;27C1	2L:6499667..6880269
6456	0,105	69D4-5;69F5-7	3L:12707218..13083087
6457	0,06	69F6-70A1;70A1-2	3L:13044488;3L:13227765
6471	0,04	67E3-7;68A2-6	3L:10413428..11106298
6478	0,063	30C3-5;30F1	2L:9706076;2L:9967229
6507	0,033	23F3-4;24A1-2	2L:3466844;2L:3659853
6608	0,062	21C3-4;21C6-8	2L:456210..464653
6646	0,045	76A7-B1;76B4-5	3L:19346700..19512438
6648	0,059	22E4-F2:22F3-23A1	2L:2459823;2L:2768590
6649	0,073	79E5-F1;80A2-3	3L:22492296..23811039
6754	0,048	75F10-11;76A1-5	3L:19145365..19302094
6779	0,062	54D1-2;54E5-7	2R:17616779;2R:17739916
6866	0,126	56C4:58D6-10	2R:19220990..19468915
6867	0,052	65D4-5:65E4-6	3L:6942786;3L:7156003
6917	0,048	45D3-4:45F2-6	2R:9397820..9613601
6965	0,059	23E5;23F4-5	2L:3344928..3486273
6999	0,056	34A3;34B7-9	2L:12942603..13393105
7079	0,06	66F1-2:67B2-3	3L:8961731..9397796
7080	0,063	85F1-2:86C7-8	3R:9907979..10917238
7142	0,038	32A1-2:32C5-D1	2L:10557962..11067015
7143	0,061	32D1:33D4-E1	2L:11021966..11218807
7144	0,087	22D2-3;22F1-2	2L:2175620;2L:2450829
7145	0,038	48C5-D1:48D5-E1	2R:11848883..12104508
7147	0,051	28A4-B1:28D3-9	2L:7465903..8048798
7273	0,068	59B:59D8-E1	2R:22766695..23562485
7339	0,05	9D5-E1:9E7-8 + 13B5-6:13E1-2	X:10556555;X:10768893
7412	0,072	98B1-2:98B3-5	3R:27670955..27915954
7413	0,023	92F7-93A1:93B3-6	3R:20762299;3R:21112186
7414	0,06	54B1-2:54B7-10	2R:17227748..17438696
7441	0,059	54C8-D1:54E2-7	2R:17561145..17749868
7443	0,044	83B7-C1;83C6-D1	3R:5823351;3R:5861298
7497	0,048	25C8;25D5	2L:5147258;2L:5305646
7566	0,09	62A2:62A6	3L:1478674;3L:1586881
7674	0,06	95A4:95B1	3R:23414436;3R:23723902
7676	0,075	95D8:95E1	3R:24031427;3R:24141369
7681	0,051	96D1:96D1	3R:25202997;3R:25515673
7682	0,071	96E2:96E6	3R:25515898;3R:25637876
7721	0,046	96E2:96E6	X:19663016;X:19788713
7754	0,05	18A2:18A2	X:19028718;X:19153633
7975	0,061	50D4:50E4	2R:14073080;2R:14212783
7896	0,064	56F11:56F16	2R:20245186;2R:20313635
7990	0,039	94E9:94E13	3R:23279758;3R:23346387
7992	0,07	95B1:95D1	3R:23773121;3R:23943004
8067	0,032	102A1:102A6	4:69710;4:193485
8082	0,064	75F2:76A1	3L:18995894;3L:19170702
8101	0,026	78D5:79A2	3L:21533807;3L:21880685
8103	0,032	83B4:83B6	3R:5600629;3R:5624095
8491	0,043	94D2-10;94E1-6	3R:22859679..23220975
8583	0,071	94E1-2:94F1-2	3R:23114713..23422676
8674	0,036	25C4:25C8	2L:5073453;2L:5145500
8836	0,046	28F5:29B1	2L:8240266;2L:8362842
9069	0,056	60C8:60E8	2R:24402666;2R:24942839
9410	0,033	45F6:46B4	2R:9594814-9594924;2R:9860827
9496	0,04	50E1:50E6	2R:14176098;2R:14265801
9497	0,026	94F1:95A4	3R:23369716;3R:23605672
9500	0,052	96F1:96F10	3R:25735650;3R:26055872-26055884
9502	0,029	28C3;28D3	2L:7774037;2L:8012787
9503	0,03	31B1;31D9	2L:10209408;2L:10333704
9505	0,085	32C1;32C1	2L:10667405;2L:11001966
9506	0,066	34C1;34C6	2L:13445419;2L:13665417
9510	0,07	40A5:40E5	2L:21828581;2L:22139023
9529	0,072	97F1-2:98A	3R:27264827..27670954
9594	0,056	34B4:34C4	2L:13290762;2L:13536086
9596	0,135	54B2:54B17	2R:17304783;2R:17494828
9625	0,06	6C12:6D8	X:6759293;X:6862948
9693	0,039	62A11:62B7	3L:1688724;3L:1841694
9700	0,061	79A3:79B3	3L:21916420;3L:22085436
23149	0,064	79B2:79D1	3L:22076093;3L:22207852
23668	0,049	67C7;67D5	3L:9760499;3L:10055226
25078	0,077	53C1:53C6	2R:16336781;2R:16494732
29026	0,058	88D2:88D6	3R:14731527..14885474

Table S3.2. List of partially deleted genes within the analysed BDSC deletion stocks.



Table S3.3. List of partially deleted genes within the BDSC deletion stocks associated with an over 50% decrease in total VC levels.



Table S3.4. Divergent sequences removed from CDS datasets prior to the final phylogenetic analyses.

PGM1	UGP2	UGDH	DHCR24	DDX46	ABPA	MIOX
Bombyx_mori_(domestic_silkworm)_Insecta_Bombycidae_XP_004926437.1	Aethina_tumida_(small_hive_beetle)_Insecta_XP_019892389.1	Rhagoletis_zephyria_(snowberry_fruit_fly)_Insecta_Tephritidae_XP_031477995.1	Amyelois_transitella_(moth)_Lepidoptera_Pyralidae_XP_013195185.1	Vespula_pensylvanica_(western_yellowjacket)_Hymenoptera_Vespidae_XP_043964611.1	Bactrocera_oleae_(olive_fruit_fly)_Diptera_Tephritidae_XP_029214465.1	Dioscorea_tenuipes_(Chinese_yam)_Tracheophyta_Dioscoreales_XP_039127999.1
Bombyx_mandarina_(wild_silkworm)_Insecta_Bombycidae_XP_028036827.1	Papilio_xuthus_(Asian_swallowtail)_Insecta_Papilionidae_XP_013142709.1	Rhagoletis_zephyria_(snowberry_fruit_fly)_Insecta_Tephritidae_XP_031483009.1	Galleria_mellonella_(greater_wax_moth)_Lepidoptera_Pyralidae_XP_026753343.1	Nylanderia_fulva_(tropical_crazy_ant)_Hymenoptera_Formicidae_XP_029035058.1	Vespa_mandarinia_(Asian_giant_hornet)_Hymenoptera_Vespidae_XP_029718386.1	Arctia_agaricis_(brown_wyandot_caterpillar)_Lepidoptera_Lycaenidae_XP_043772760.1
Manduca_sexta_(tobacco_hornworm)_Insecta_Sphingidae_XP_037502995.1	Papilio_xuthus_(Asian_swallowtail)_Insecta_Papilionidae_XP_013142341.1		Apis_(laborious_(Himalayan_honeybee)_Hymenoptera_Apidae_XP_026302555.1	Drosophila_Anopheles_(Bero)_Diptera_Drosophilidae_XP_016964620.1	Lucilia_cuprina_(Australian_sheep_blowfly)_Diptera_Calliphoridae_XP_023297261.1	
Danaus_plexippus_plexippus_(monarch_butterfly)_Insecta_Nymphalidae_XP_032519260.1	Galleria_mellonella_(greater_wax_moth)_Lepidoptera_Pyralidae_XP_026753343.1		Chrysopelia_paradisi_(snake)_XP_054755705.1		Lucilia_cuprina_(Australian_sheep_blowfly)_Diptera_Calliphoridae_XP_023297975.1	
Vanessa_tameamea_(Nafferton)_Insecta_Nymphalidae_XP_026485675.1	Papilio_xuthus_(Asian_swallowtail)_Insecta_XP_013350501.1		Oncaus_abietinus_(hymenopteran)_Hymenoptera_Oncaeidae_XP_012301185.1		Bradysia_coprophila_(Bero)_Diptera_Sciaridae_XP_037542708.1	
Hirtisomatus_armigera_(cotton_bollworm)_Insecta_Noctuidae_XP_031943535.1	Trichoplusia_ni_(cabbage_looper)_Insecta_Trichoplusiinae_XP_026437997.1		Amphylaeus_gibbiventris_(Asian_bogbumee_beetle)_Coleoptera_Cerambycidae_XP_028849923.1		Danaus_plexippus_plexippus_monarch_butterfly_Lepidoptera_XP_032519260.1	
Papilio_polytes_(common_Mormon)_Insecta_Papilionidae_XP_013150186.1	Sitophilus_oryzae_(rice_weevil)_Insecta_Curculionidae_XP_030737353.1		Apis_(eastern_(Himalayan_honeybee)_Hymenoptera_Apidae_XP_016921389.1		Cyphomyrmex_costatus_ants_Hymenoptera_Formicidae_XP_029031602.1	
Hyposmocoma_kahamanoa_moths_(Hyposmocoma_kahamanoa_moth)_Insecta_Cosmopterigidae_XP_026316870.1	Nylanderia_fulva_(tropical_crazy_ant)_Hymenoptera_Formicidae_XP_029035058.1		Cryptotermes_secundus_(termites)_Blattodea_Kalotermitidae_XP_023763582.1		Melipona_quadrifasciata_ants_bees_Hymenoptera_Apidae_XP_015894182.1	
Pieris_rapae_(XP_022123934.1)	Monomorium_(ingle)_Insecta_XP_012838966.1		Orthaltiraphaga_(cactus_(beetles)_Coleoptera_Scraptiidae_XP_022953302.1		Camponotus_ants_bees_Hymenoptera_Formicidae_XP_024475582.1	
Tetrapus_(abbreviated_(Bero)_Insecta_Diopsidae_XP_037929721.1	Amphylaeus_cubaunti_(mosquitos)_Insecta_Culicidae_XP_028235626.1		Cryptotermes_secundus_(termites)_Blattodea_Kalotermitidae_XP_023723618.1		Paraage_angelina_(speckled_wood_butterfly_Lepidoptera_XP_039765736.1	
Tetrapus_(abbreviated_(Bero)_Insecta_Diopsidae_XP_037929477.1	Acyrthosiphon_pisum_(pea_aphid)_Insecta_Aphididae_XP_001943680.1		Zootermopsis_nevadensis_(termites)_Blattodea_Termopsidae_XP_021937079.1		Leptothorax_(formicidae)_(Colorado_potato_beetle)_Coleoptera_Chrysomelidae_XP_023524231.1	
Cimex_lectularius_(bed_bug)_Insecta_Cimicidae_XP_014255767.1	Agrilus_planipennis_(emerald_ash_borer)_Insecta_Buprestidae_XP_032852990.1				Plautus_pyralis_common_western_firefly_Coleoptera_Lampyridae_XP_031787730.1	
Bombycus_(silkworm_(bubble_fly)_Insecta_Bombycidae_XP_033534990.1	Amphylaeus_gibbiventris_(Asian_bogbumee_beetle)_Insecta_Cerambycidae_XP_028179897.1				Tetrapus_(abbreviated_(Bero_Diptera_Diopsidae_XP_037955145.1	
Lucilia_cuprina_(Australian_sheep_blowfly)_Insecta_Calliphoridae_XP_023297598.1	Amphylaeus_gibbiventris_(Asian_bogbumee_beetle)_Insecta_Cerambycidae_XP_028353065.1				Tetrapus_(abbreviated_(Bero_Diptera_Diopsidae_XP_037955307.1	
Lucilia_sericata_(common_green_bottle_fly)_Insecta_Calliphoridae_XP_037819286.1	Amphylaeus_gibbiventris_(Asian_bogbumee_beetle)_Insecta_XP_028179898.1				Rhagoletis_zephyria_(snowberry_fruit_fly)_Diptera_Tephritidae_XP_031748731.1	
Nicrophorus_vespilloides_(beetles)_Insecta_Silphidae_XP_017780560.1	Apis_cerana_(Eastern_honeybee)_Insecta_Apidae_XP_016899252.2				Drosophila_serrata_Bero_Diptera_Drosophilidae_XP_032756585.1	
Rhagoletis_zephyria_(snowberry_fruit_fly)_Insecta_Tephritidae_XP_031477995.1	Apis_cerana_(Eastern_honeybee)_Insecta_Apidae_XP_039437391.1				Drosophila_ficusphila_Bero_Diptera_Drosophilidae_XP_017044178.1	
Drosophila_eugracilis_(Bero)_Insecta_Drosophilidae_XP_017035918.1	Metaphycita_luschnati_(aphids)_Insecta_Aphididae_XP_031783936.1				Lucilia_cuprina_(Australian_sheep_blowfly)_Diptera_Calliphoridae_XP_023296597.1	
Homo_sapiens_human_Mammalia_Hominidae_NP_068802.2	Nicrophorus_vespilloides_(beetles)_Insecta_Silphidae_XP_017780985.1				Drosophila_biarmipes_Bero_Diptera_Drosophilidae_XP_016961930.2	
Mus_musculus_house_mouse_Mammalia_Muridae_NP_776178.3	Bradysia_coprophila_(Bero)_Insecta_Sciaridae_XP_037537706.1				Homo_sapiens_human_Primates_Hominidae_NP_056970.4	
	Anopheles_stephensi_(Asian_malaria_mosquito)_Insecta_Culicidae_XP_035903578.1				Drosophila_miranda_Bero_Diptera_Drosophilidae_XP_017527754.1	
	Copidosoma_floridanum_(wasps_ants_and_bees)_Insecta_Encyrtidae_XP_014514939.1				Drosophila_takahashii_Bero_Diptera_Drosophilidae_XP_017552775.1	
	Eufriesea_mexicana_(Bero)_Insecta_Apidae_XP_017760797.1				Drosophila_ananassae_Bero_Diptera_Drosophilidae_XP_032955255.1	
	Danaus_plexippus_plexippus_(monarch_butterfly)_Insecta_XP_032519260.2				Drosophila_kikkawai_Bero_Diptera_Drosophilidae_XP_017010900.1	
	Danaus_plexippus_plexippus_(monarch_butterfly)_Insecta_XP_032532926.1				Drosophila_subobscura_Bero_Diptera_Drosophilidae_XP_034001915.1	
	Chelonus_insularis_(wasps_ants_and_bees)_Insecta_Braconidae_XP_024859422.1				Stomoxys_calcitrans_(mites_fly_Diptera_Muscidae_XP_013127481.1	
	Rhopalosiphum_maidis_(corn_leaf_aphid)_Insecta_Aphididae_XP_026021545.1				Drosophila_pseudoobscura_Bero_Diptera_Drosophilidae_XP_001358979.3	
	Bombus_(Mexico_(Bero)_Insecta_Apidae_XP_033151227.1				Drosophila_persimilis_Bero_Diptera_Drosophilidae_XP_002015216.1	
	Nylanderia_(fulva_(ants)_Insecta_Formicidae_XP_029034835.1				Drosophila_obscura_Bero_Diptera_Drosophilidae_XP_032721400.1	
	Bombus_(mastrucs_(common_western_bumble_bee)_Insecta_Apidae_XP_003397979.1				Drosophila_bipectinata_Bero_Diptera_Drosophilidae_XP_017038067.2	
	Bicyclus_(eyranes_(squirting_bush_brown)_Insecta_XP_029969852.1				Mus_musculus_house_mouse_Rodentia_Muridae_NP_035969.2	
	Leptocybranus_(dendrobranchia_(Columbia_(asian_beetle)_Insecta_Chrysomelidae_XP_023523960.1				Drosophila_kikkawai_Bero_Diptera_Drosophilidae_XP_017034583.1	
	Leptocybranus_(dendrobranchia_(Columbia_(asian_beetle)_Insecta_Chrysomelidae_XP_023523979.1				Drosophila_(Bero_(Diptera_Drosophilidae_XP_017011706.1	
	Dendroctonus_ponderosae_(mountain_pine_beetle)_Insecta_Curculionidae_XP_019760272.1				Drosophila_willistoni_Bero_Diptera_Drosophilidae_XP_002070350.1	
	Dendroctonus_ponderosae_(mountain_pine_beetle)_Insecta_Curculionidae_XP_019760722.1				Drosophila_guanche_Bero_Diptera_Drosophilidae_XP_035023725.1	
	Anopheles_(eubrachis_(mosquitoes)_Insecta_Culicidae_XP_035018588.1				Drosophila_hydei_Bero_Diptera_Drosophilidae_XP_023164588.1	
	Centaurea_maculosa_(beetle_midgid_Insecta_Ceratopogonidae_XP_031437397.1				Drosophila_(ebulla_Bero_Diptera_Drosophilidae_XP_032958836.1	
	Bombus_(terrestris_(bufftailed_bumblebee)_Insecta_Apidae_XP_003399507.1				Drosophila_subpulchrella_Bero_Diptera_Drosophilidae_XP_037737324.1	
	Aphis_gossypii_(cotton_aphid)_Insecta_Aphididae_XP_027637011.1				Drosophila_melanogaster_fruit_fly_Diptera_Drosophilidae_XP_002732.1	
	Microplitis_demolitor_(wasps_ants_and_bees)_Insecta_Braconidae_XP_008556265.1				Drosophila_(willistella_Bero_Diptera_Drosophilidae_XP_034337351.1	
	Photinus_pyralis_(common_western_firefly)_Insecta_Lampyridae_XP_031319960.1				Drosophila_(anamensis(ca_Bero_Diptera_Drosophilidae_XP_035965600.1	
	Cotesia_(glomerata_(cabbage_white_butterfly)_Insecta_Braconidae_XP_034183202.1				Drosophila_(suzukii_(spotted_wing_drosophila)_Insecta_Drosophilidae_XP_032925842.1	
	Folsomia_(candida_(springtail)_(Collembola_Insecta_XP_035707282.1				Rhagoletis_(zephyria_(snowberry_fruit_fly_Diptera_Tephritidae_XP_031748284.1	
	Oncaus_abietinus_(hymenopteran)_Insecta_Oncaeidae_XP_012298632.1				Drosophila_(melgas_Bero_Diptera_Drosophilidae_XP_017957645.1	
	Apis_(florea_(little_honeybee)_Insecta_Apidae_XP_003691870.1				Drosophila_(miranda_Bero_Diptera_Drosophilidae_XP_035974701.1	
	Diachasma_alloeum_(wasps_ants_and_bees)_Insecta_Braconidae_XP_015123685.1				Drosophila_(mojavensis_Bero_Diptera_Drosophilidae_XP_032958451.1	
	Nilaparvata_(lugens_(brown_planthopper)_(Delphacidae_Insecta_XP_039019651.1				Drosophila_(navojoa_Bero_Diptera_Drosophilidae_XP_017035788.1	
	Culex_(pipiens_(pallens_(northern_house_mosquito)_Insecta_Culicidae_XP_039439892.1				Drosophila_(subobscura_Bero_Diptera_Drosophilidae_XP_034003927.1	
	Rhopalosiphum_(maidis_(corn_leaf_aphid)_Insecta_Aphididae_XP_026021545.1				Zeugodacus_(cucurbitae_(melon_fly_Diptera_Tephritidae_XP_011183866.1	
	Myzus_(persicae_(green_peach_aphid)_Insecta_Aphididae_XP_022180705.1				Drosophila_(arizonae_Bero_Diptera_Drosophilidae_XP_017974183.1	
	Bombus_(latreil_(lucent_pocket_umbrella)_Insecta_Hymenoptera_XP_033952555.1				Drosophila_(ambi_Bero_Diptera_Drosophilidae_XP_032953825.1	
	Aethina_tumida_(small_hive_beetle)_Insecta_XP_019878989.1				Anopheles_(coluzzii_(mosquitoes_Diptera_Culicidae_XP_040327958.1	
	Nasonia_(vitripennis_(jewel_wasp)_Insecta_Pteromalidae_XP_016619025.1				Anopheles_(albimanus_mosquitos_Diptera_Culicidae_XP_035762256.1	
	Sipha_(flava_(yellow_sugarcane_aphid)_Insecta_Aphididae_XP_025629466.1				Anopheles_(merus_mosquitos_Diptera_Culicidae_XP_041976670.1	
	Drosophila_(grimshawi_(Bero)_Insecta_Drosophilidae_XP_001992139.2				Bradysia_coprophila_(Bero_Diptera_Sciaridae_XP_037554067.1	
	Trifolium_(subterran_(leaf_bee_Hymenoptera_Halictidae_XP_039437391.1				Aedes_(aegypti_(yellow_fever_mosquito_Diptera_Culicidae_XP_035992712.1	
	Trifolium_(subterran_(leaf_bee_Hymenoptera_Halictidae_XP_039437391.1				Aedes_(aegypti_(yellow_fever_mosquito_Diptera_Culicidae_XP_035992559.2	
	Anopheles_(stephensi_(Asian_malaria_mosquito_Diptera_Culicidae_XP_035903578.1				Anopheles_(arabiensis_mosquitos_Diptera_Culicidae_XP_040172268.1	
	Copidosoma_(floridanum_(wasps_ants_and_bees)_Insecta_Encyrtidae_XP_014514939.1				Centaurea_(maculosa_(beetle_midgid_Diptera_Ceratopogonidae_XP_031437397.1	
	Eufriesea_(mexicana_(Bero_Insecta_Apidae_XP_017760797.1				Aedes_(albopictus_(Asian_tiger_mosquito_Diptera_Culicidae_XP_029504867.2	
	Danaus_(plexippus_(plexippus_(monarch_butterfly_Insecta_XP_032519260.2				Culex_(quinquefasciatus_(southern_house_mosquito_Diptera_Culicidae_XP_001845423.2	
	Danaus_(plexippus_(plexippus_(monarch_butterfly_Insecta_XP_032532926.1				Manduca_(sexta_(tobacco_hornworm_Lepidoptera_Sphingidae_XP_030236923.2	
	Chelonus_(insularis_(wasps_ants_and_bees)_Insecta_Braconidae_XP_024859422.1				Papilio_(xuthus_(Asian_swallowtail_Lepidoptera_Papilionidae_XP_013157620.1	
	Rhopalosiphum_(maidis_(corn_leaf_aphid)_Insecta_Aphididae_XP_026021545.1				Spodoptera_(frugiperda_(fall_armyworm_Lepidoptera_XP_035819055.1	
	Bombus_(Mexico_(Bero)_Insecta_Apidae_XP_033151227.1				Papilio_(xuthus_(Asian_swallowtail_Lepidoptera_XP_013142341.1	
	Nylanderia_(fulva_(ants)_Insecta_Formicidae_XP_029034835.1				Acyrthosiphon_(pisum_(pea_aphid)_Hemiptera_Aphididae_XP_016656301.1	
	Bombus_(mastrucs_(common_western_bumble_bee)_Insecta_Apidae_XP_003397979.1				Hyposmocoma_kahamanoa_moths_Lepidoptera_Cosmopterigidae_XP_026316870.1	
	Bicyclus_(eyranes_(squirting_bush_brown)_Insecta_XP_029969852.1				Melanaphis_(sacchari_(aphids_Hemiptera_Aphididae_XP_025637088.1	
	Leptocybranus_(dendrobranchia_(Columbia_(asian_beetle)_Insecta_Chrysomelidae_XP_023523960.1				Acica_agaricis_(brown_wyandot_caterpillar)_Lepidoptera_Lycaenidae_XP_043966668.1	
	Leptocybranus_(dendrobranchia_(Columbia_(asian_beetle)_Insecta_Chrysomelidae_XP_023523979.1				Bombyx_(mori_(domestic_silkworm_Lepidoptera_Bombycoidae_XP_004925439.1	
	Dendroctonus_ponderosae_(mountain_pine_beetle)_Insecta_Curculionidae_XP_019760272.1				Danaus_(plexippus_(plexippus_monarch_butterfly_Lepidoptera_XP_032519260.2	
	Dendroctonus_ponderosae_(mountain_pine_beetle)_Insecta_Curculionidae_XP_019760722.1				Danaus_(plexippus_(plexippus_monarch_butterfly_Lepidoptera_XP_032532926.1	
	Anopheles_(eubrachis_(mosquitoes)_Insecta_Culicidae_XP_035018588.1				Amyelois_(transitella_(moths_Lepidoptera_Pyralidae_XP_013195185.1	
	Centaurea_(maculosa_(beetle_midgid_Insecta_Ceratopogonidae_XP_031437397.1				Bicyclus_(eyranes_(squirting_bush_brown_Lepidoptera_XP_029969852.1	
	Bombus_(terrestris_(bufftailed_bumblebee)_Insecta_Apidae_XP_003399507.1				Galleria_(mellonella_(greater_wax_moth_Lepidoptera_Pyralidae_XP_026775387.1	
	Aphis_(gossypii_(cotton_aphid)_Insecta_Aphididae_XP_027637011.1				Manduca_(sexta_(tobacco_hornworm_Lepidoptera_Sphingidae_XP_030236923.2	
	Microplitis_(demolitor_(wasps_ants_and_bees)_Insecta_Braconidae_XP_008556265.1				Ostrinia_(furnacalis_(Asian_corn_borer_Lepidoptera_Crambidae_XP_028160168.1	
	Photinus_(pyralis_(common_western_firefly)_Insecta_Lampyridae_XP_031319960.1				Vanessa_(tameamea_(Nafferton_Lepidoptera_XP_026485837.1	
	Cotesia_(glomerata_(cabbage_white_butterfly)_Insecta_Braconidae_XP_034183202.1				Aphis_(gossypii_(cotton_aphid_Hemiptera_Aphididae_XP_027637011.1	
	Folsomia_(candida_(springtail)_(Collembola_Insecta_XP_035707282.1				Jeremy_(ornatus_(Agiltus_butterfly_Lepidoptera_Pieridae_XP_030054560.1	
	Oncaus_abietinus_(hymenopteran)_Insecta_Oncaeidae_XP_012298632.1				Papilio_(xuthus_(Asian_swallowtail_Lepidoptera_XP_013142341.1	
	Apis_(florea_(little_honeybee)_Insecta_Apidae_XP_003691870.1				Helicoverpa_(armigera_(cotton_bollworm_Lepidoptera_XP_021181392.1	
	Diachasma_alloeum_(wasps_ants_and_bees)_Insecta_Braconidae_XP_015123685.1				Myzus_(persicae_(green_peach_aphid_Hemiptera_Aphididae_XP_022180705.1	
	Nilaparvata_(lugens_(brown_planthopper)_(Delphacidae_Insecta_XP_039019651.1				Sipha_(flava_(yellow_sugarcane_aphid_Hemiptera_Aphididae_XP_025629466.1	
	Culex_(pipiens_(pallens_(northern_house_mosquito)_Insecta_Culicidae_XP_039439892.1				Papilio_(polytes_(common_Mormon_Lepidoptera_Papilionidae_XP_013139375.1	
	Rhopalosiphum_(maidis_(corn_leaf_aphid)_Insecta_Aphididae_XP_026021545.1				Bicyclus_(martelliana_(wild_silkworm_Lepidoptera_Bombycoidae_XP_028978383.1	
	Myzus_(persicae_(green_peach_aphid)_Insecta_Aphididae_XP_022180705.1				Paraage_(segesta_(speckled_wood_butterfly_Lepidoptera_XP_039757635.1	
	Bombus_(latreil_(lucent_pocket_umbrella)_Insecta_Hymenoptera_XP_033952555.1				Trichoplusia_(ni_(cabbage_looper_Lepidoptera_Trichoplusiinae_XP_026437997.1	
	Aethina_tumida_(small_hive_beetle)_Insecta_XP_019878989.1				Cyphomyrmex_(costatus_(ants_Hymenoptera_Formicidae_XP_029031602.1	
	Nasonia_(vitripennis_(jewel_wasp)_Insecta_Pteromalidae_XP_016619025.1				Vespula_(mandarinia_(Asian_giant_hornet_Hymenoptera_Vespidae_XP_029718372.1	
	Sipha_(flava_(yellow_sugarcane_aphid)_Insecta_Aphididae_XP_025629466.1				Atta_(cephalotes_(ants_Hymenoptera_Formicidae_XP_012307205.1	
	Drosophila_(grimshawi_(Bero)_Insecta_Drosophilidae_XP_001992139.2				Vollenhovia_(emeryi_(ants_Hymenoptera_Formicidae_XP_012370298.1	
	Trifolium_(subterran_(leaf_bee_Hymenoptera_Halictidae_XP_039437391.1				Atta_(colombica_(ants_Hymenoptera_Formicidae_XP_040362203.1	
	Trifolium_(subterran_(leaf_bee_Hymenoptera_Halictidae_XP_039437391.1				Trachymyrmex_(cornetzi_(ants_Hymenoptera_Formicidae_XP_018372891.1	
	Anopheles_(gambiae_(vc_PEST_(African_malaria_mosquito_Insecta_Culicidae_XP_321500.4				Diachasma_(alloeum_(wasps_ants_and_bees_Hymenoptera_Braconidae_XP_015123685.2	
	Aedes_(albopictus_(Asian_tiger_mosquito_Insecta_Culicidae_XP_029504867.2				Pogonomyrmex_(barbatus_(red_harvester_ant_Hymenoptera_Formicidae_XP_039378088.1	
	Paraage_(segesta_(speckled_wood_butterfly_Insecta_XP_039757635.1				Cyphomyrmex_(costatus_(ants_Hymenoptera_Formicidae_XP_029031602.1	
	Danaus_(plexippus_(plexippus_monarch_butterfly_Insecta_XP_032519260.2				Formica_(exsecta_(ants_Hymenoptera_Formicidae_XP_029971580.1	
	Danaus_(plexippus_(plexippus_monarch_butterfly_Insecta_XP_032532926.1				Trachymyrmex_(septentrionalis_ants_Hymenoptera_Formicidae_XP_018372312.1	
	Chelonus_(insularis_(wasps_ants_and_bees)_Insecta_Braconidae_XP_024859422.1				Harpegnathos_(saltator_(jerdon's_jumping_ant_Hymenoptera_Formicidae_XP_011143264.1	
	Rhopalosiphum_(maidis_(corn_leaf_aphid)_Insecta_Aphididae_XP_026021545.1				Monomorium_(pharaonis_(pharaoh_ant_Hymenoptera_Formicidae_XP_036928547.1	
	Bombus_(Mexico_(Bero)_Insecta_Apidae_XP_033151227.1				Camponotus_floridanus_(Florida_carpenter_ant_Hymenoptera_Formicidae_XP_011264087.2	
	Nylanderia_(fulva_(ants)_Insecta_Formicidae_XP_029034835.1				Vespula_pensylvanica_(western_yellowjacket_Hymenoptera_Vespidae_XP_043983211.1	
	Bombus_(mastrucs_(common_western_bumble_bee)_Insecta_Apidae_XP_003397979.1				Solenopsis_(invicta_(red_fire_ant_Hymenoptera_Formicidae_XP_011165075.1	
	Bicyclus_(eyranes_(squirting_bush_brown)_Insecta_XP_029969852.1				Dinoponera_(quadriceps_(ants_Hymenoptera_Formicidae_XP_014272899.1	
	Leptocybranus_(dendrobranchia_(Columbia_(asian_beetle)_Insecta_Chrysomelidae_XP_023523960.1				Linepithema_(humile_(Argentine_ant_Hymenoptera_Formicidae_XP_012230708.1	
	Leptocybranus_(dendrobranchia_(Columbia_(asian_beetle)_Insecta_Chrysomelidae_XP_023523979.1				Pogonomyrmex_(barbatus_(red_harvester_ant_Hymenoptera_Formicidae_XP_011272816.1	
	Dendroctonus_ponderosae_(mountain_pine_beetle)_Insecta_Curculionidae_XP_019760272.1				Acromyrmex_(echinatior_(Panamanian_leafcutter_ant_Hymenoptera_Formicidae_XP_011259738.1	
	Dendroctonus_ponderosae_(mountain_pine_beetle)_Insecta_Curculionidae_XP_019760722.1				Odontomachus_(brunneus_(ants_Hymenoptera_Formicidae_XP_032685308.1	
	Anopheles_(eubrachis_(mosquitoes)_Insecta_Culicidae_XP_035018588.1				Apis_(dorsata_(giant_honeybee_Hymenoptera_Apidae_XP_006625426.1	
	Centaurea_(maculosa_(beetle_midgid_Insecta_Ceratopogonidae_XP_031437397.1				Cullomia_(gigas_(ants_Hymenoptera_Formicidae_XP_012255563.1	
	Bombus_(terrestris_(bufftailed_bumblebee)_Insecta_Apidae_XP_003399507.1				Apis_(mellifera_(honey_bee_Hymenoptera_Apidae_XP_026937950.1	
	Aphis_(gossypii_(cotton_aphid)_Insecta_Aphididae_XP_027637011.1				Apis_(mellifera_(Western_honeybee_Hymenoptera_Apidae_XP_016770679.1	
	Microplitis_(demolitor_(wasps_ants_and_bees)_Insecta_Braconidae_XP_008556265.1				Apis_(cerana_(Indian_honeybee_Hymenoptera_Apidae_XP_016899252.1	
	Oncaus_(abietinus_(hymenopteran)_Insecta_Oncaeidae_XP_012298632.1				Ceratina_(calcarata_(bees_Hymenoptera_Apidae_XP_017780556.1	
	Diachasma_(alloeum_(wasps_ants_and_bees_Hymenoptera_Braconidae_XP_015123685.1				Nasonia_(vitripennis_(jewel_wasp_Hymenoptera_Halictidae_XP_031942336.1	
	Leptopilina_(heterotoma_(wasps_ants_and_bees_Hymenoptera_Figitidae_XP_023180227.1				Osmia_(bicornis_(bicornis_red_mason_bee_Hymenoptera_Megachilidae_XP_029036077.1	
	Vanessa_(cardui_(painted_lady_butterfly_Lepidoptera_Nymphalidae_XP_023944775.1				Eufriesea_(mexicana_(bees_Hymenoptera_Apidae_XP_017757389.1	
	Diabrotica_(virgifera_(virgifera_western_corn_rootworm_Coleoptera_Chrysomelidae_XP_028140782.1				Bombus_(terrestris_(bees_Hymenoptera_Apidae_XP_033151430.1	
	Trichoplusia_(ni_(cabbage_looper_Lepidoptera_Trichoplusiinae_XP_026437997.1				Megabombus_(pensilis_(bees_Hymenoptera_Halictidae_XP_035350751.1	
	Trichoplusia_(ni_(cabbage_looper_Lepidoptera_Trichoplusiinae_XP_026437997.1				Osmia_(ribaria_(orchard_mason_bee_Hymenoptera_Megachilidae_XP_030240561.1	
	Cimex_(lectularius_(bed_bug_Hemiptera_Cimicidae_XP_014255767.1				Habropoda_(laboriosa_(southeastern_blueberry_bee_Hymenoptera_Apidae_XP_035489188.1	
	Cimex_(lectularius_(bed_bug_Insecta_Cimicidae_XP_014255767.1				Bombus_(impatiens_(bees_Hymenoptera_Apidae_XP_033167990.1	
	Sipunculus_(nudus_(Common_peanutworm_Sipuncula_Sipunculidae_XP_012802933.1				Dufourea_(novaeangliae_(bees_Hymenoptera_Halictidae_XP_034208821.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Melipona_(quadrifasciata_(bees_Hymenoptera_Apidae_XP_015894182.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Frieseomelitta_(varia_(bees_Hymenoptera_Apidae_XP_019704564.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Bombus_(terrestris_(bees_Hymenoptera_Apidae_XP_033167990.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Apis_(laboriosa_(Himalayan_honeybee_Hymenoptera_Apidae_XP_026302555.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Pollenia_(rudis_(bees_Hymenoptera_Apidae_XP_015770529.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Pollenia_(amadensis_(wasps_ants_and_bees_Hymenoptera_Vespidae_XP_014019279.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Apis_(florea_(little_honeybee_Hymenoptera_Apidae_XP_012348592.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Bombus_(terrestris_(Hymenoptera_Apidae_XP_033168990.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Bombus_(terrestris_(Hymenoptera_Apidae_XP_033168990.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Habropoda_(laboriosa_(bees_Hymenoptera_Apidae_XP_017785628.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Apidius_(gillanicki_(wasps_ants_and_bees_Hymenoptera_Braconidae_XP_016011165.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Papias_(indianus_(wasps_ants_and_bees_Hymenoptera_Braconidae_XP_013159965.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Cryptotermes_secundus_(termites_Blattodea_Kalotermitidae_XP_023734575.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Zootermopsis_nevadensis_(termites_Blattodea_Termopsidae_XP_021937222.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Copidosoma_floridanum_(wasps_ants_and_bees_Hymenoptera_Encyrtidae_XP_014235787.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Leptopilina_(heterotoma_(wasps_ants_and_bees_Hymenoptera_Figitidae_XP_023180227.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Athalia_rosae_(cabbage_(sawfly_Hymenoptera_Tenthredinidae_XP_028357438.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Microplitis_(demolitor_(wasps_ants_and_bees_Hymenoptera_Braconidae_XP_008556265.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Oncaus_(abietinus_(hymenopteran)_Hymenoptera_Oncaeidae_XP_012298632.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Diachasma_(alloeum_(wasps_ants_and_bees_Hymenoptera_Braconidae_XP_015123685.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Leptopilina_(heterotoma_(wasps_ants_and_bees_Hymenoptera_Figitidae_XP_023180227.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Vanessa_(cardui_(painted_lady_butterfly_Lepidoptera_Nymphalidae_XP_023944775.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Diabrotica_(virgifera_(virgifera_western_corn_rootworm_Coleoptera_Chrysomelidae_XP_028140782.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Trichoplusia_(ni_(cabbage_looper_Lepidoptera_Trichoplusiinae_XP_026437997.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Trichoplusia_(ni_(cabbage_looper_Lepidoptera_Trichoplusiinae_XP_026437997.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Cimex_(lectularius_(bed_bug_Hemiptera_Cimicidae_XP_014255767.1	
	Ixodes_(scapularis_(black-legged_tick_Arthropoda_Ixodidae_XP_002402161.1				Cimex_(lectularius_(bed_bug_Insecta_Cimicidae_X

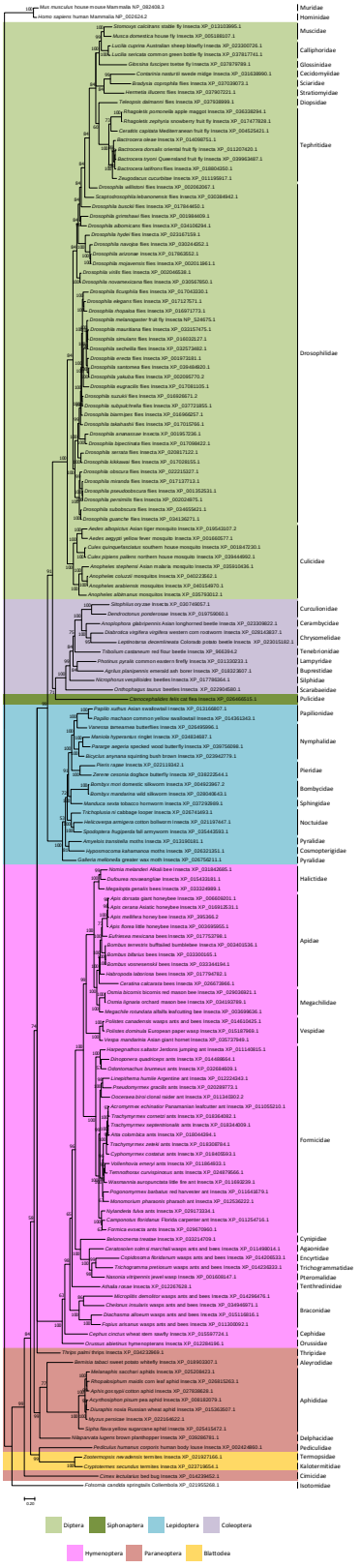


Figure S3.1. *PGM1* Bayesian Phylogeny. Taxonomic family names are represented to the right of tree branches. The relevant Insecta higher taxonomic groups are highlighted with distinct colours according to the legend below the phylogeny. The vertebrate and Collembola outgroup sequences are not highlighted.

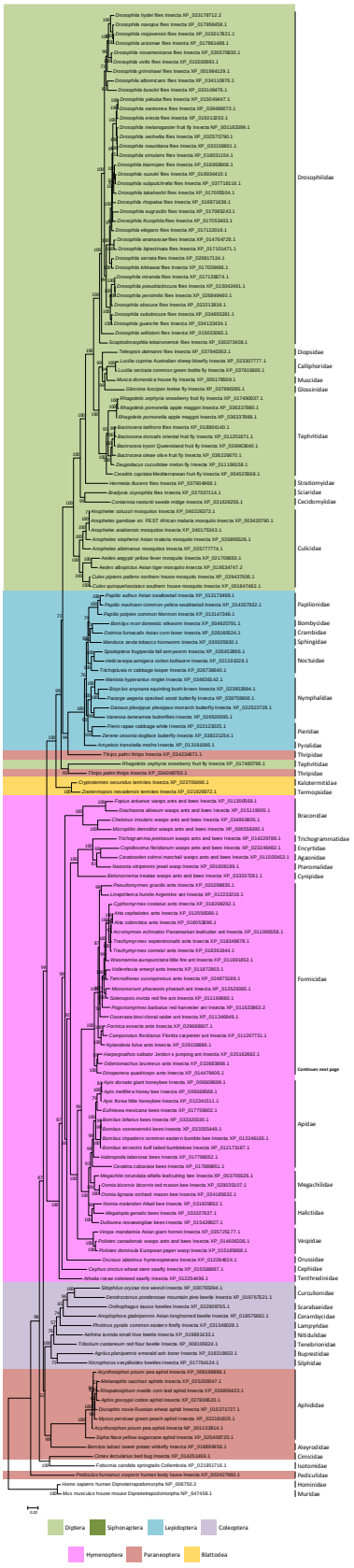


Figure S3.2. *UGP2* Bayesian Phylogeny. Taxonomic family names are represented to the right of tree branches. The relevant Insecta higher taxonomic groups are highlighted with distinct colours according to the legend below the phylogeny. The vertebrate and Collembola outgroup sequences are not highlighted.

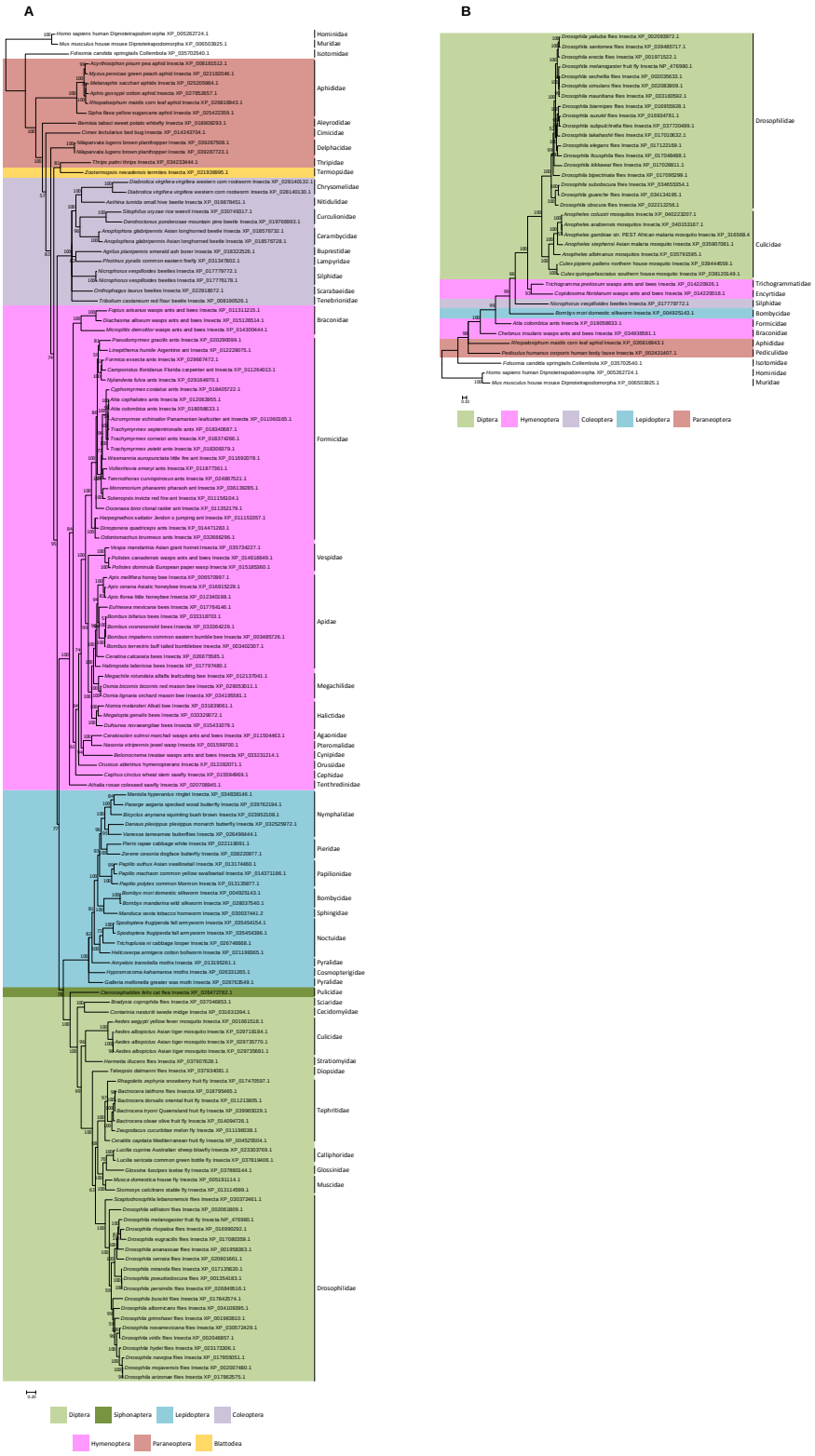


Figure S3.3. *UGDH* Bayesian Phylogenies. The phylogeny containing the main UGDH dataset CDS is presented in panel A. The complementary phylogeny constituted by the most divergent *UGDH* CDS is presented in panel B. Taxonomic family names are represented to the right of tree branches. The relevant Insecta higher taxonomic groups are highlighted with distinct colours according to the legend below the phylogeny. The vertebrate and Collembola outgroup sequences are not highlighted.



Figure S3.4. *AKR1B1* FastTree Phylogeny. Taxonomic family names are represented to the right of tree branches. The relevant Insecta higher taxonomic groups are highlighted with distinct colours according to the legend below the phylogeny. The vertebrate and Collembola outgroup sequences are not highlighted.

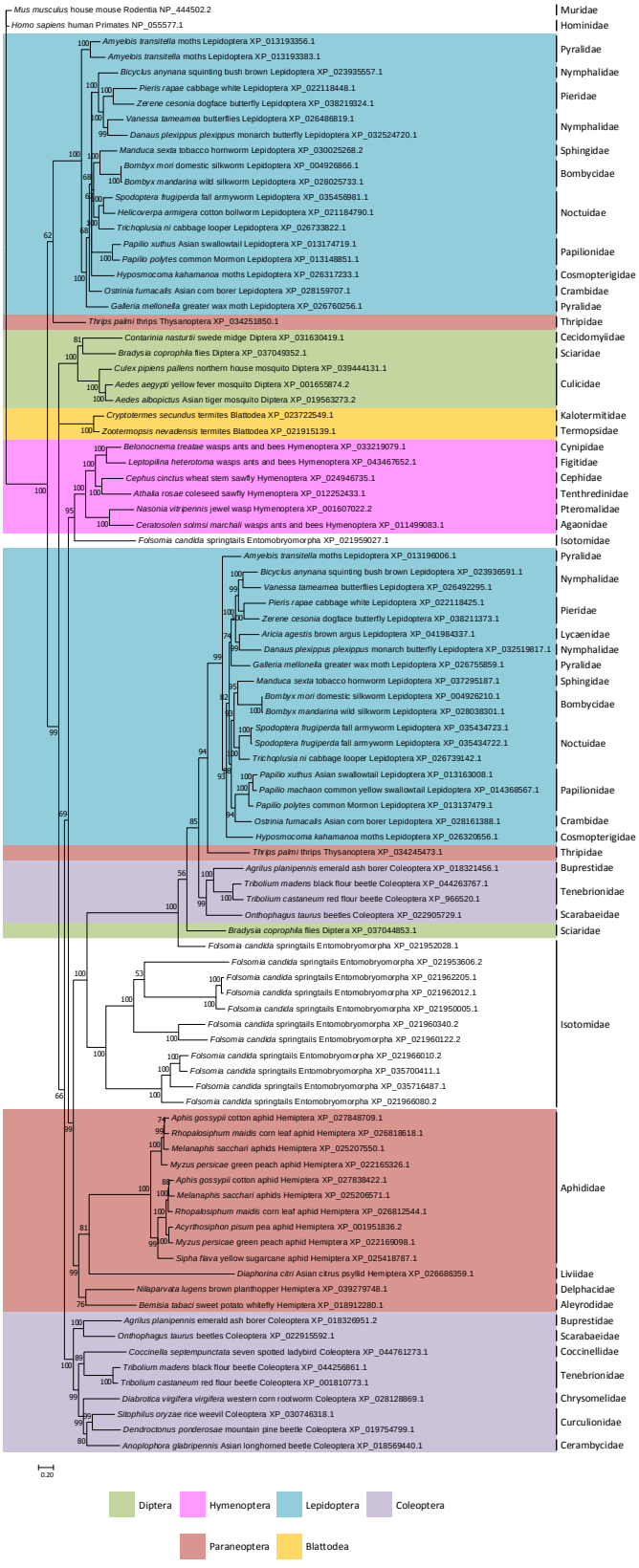


Figure S3.5. *DHCR24* Bayesian Phylogeny. Taxonomic family names are represented to the right of tree branches. The relevant Insecta higher taxonomic groups are highlighted with distinct colours according to the legend below the phylogeny. The vertebrate and Collembola outgroup sequences are not highlighted.

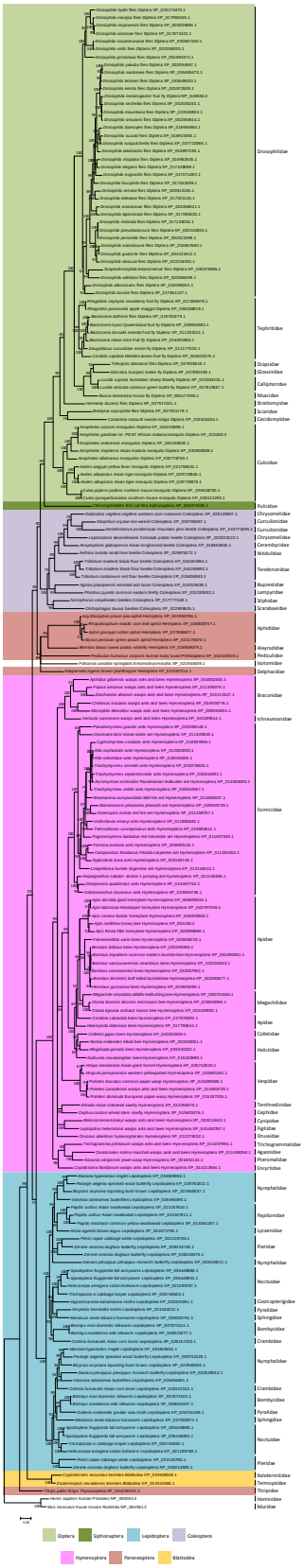


Figure S3.6. *MIOX* Bayesian Phylogeny. Taxonomic family names are represented to the right of tree branches. The relevant Insecta higher taxonomic groups are highlighted with distinct colours according to the legend below the phylogeny. The vertebrate and Collembola outgroup sequences are not highlighted.

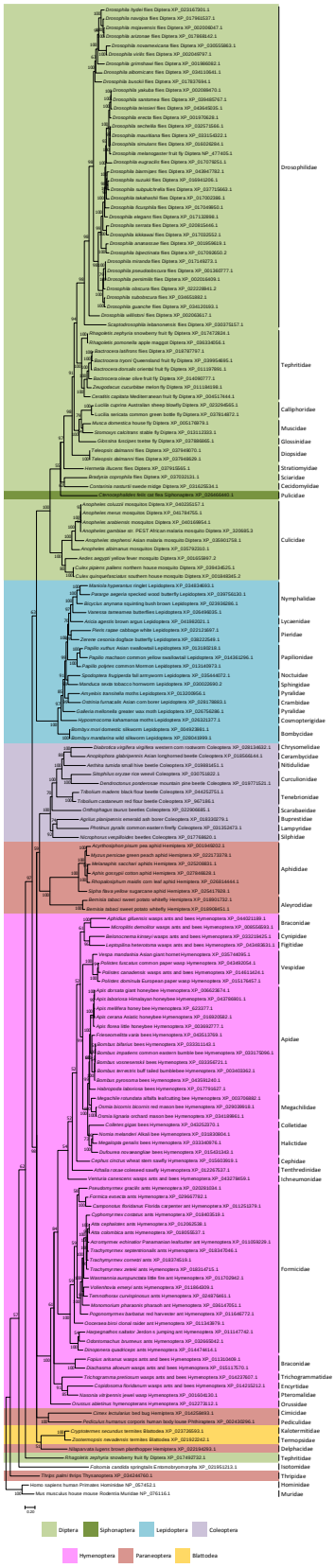


Figure S3.7. *ISYNA1* Bayesian Phylogeny. Taxonomic family names are represented to the right of tree branches. The relevant Insecta higher taxonomic groups are highlighted with distinct colours according to the legend below the phylogeny. The vertebrate and Collembola outgroup sequences are not highlighted.



Figure S3.8. *IMPA* Bayesian Phylogeny. Taxonomic family names are represented to the right of tree branches. The relevant Insecta higher taxonomic groups are highlighted with distinct colours according to the legend below the phylogeny. The vertebrate and Collembola outgroup sequences are not highlighted.

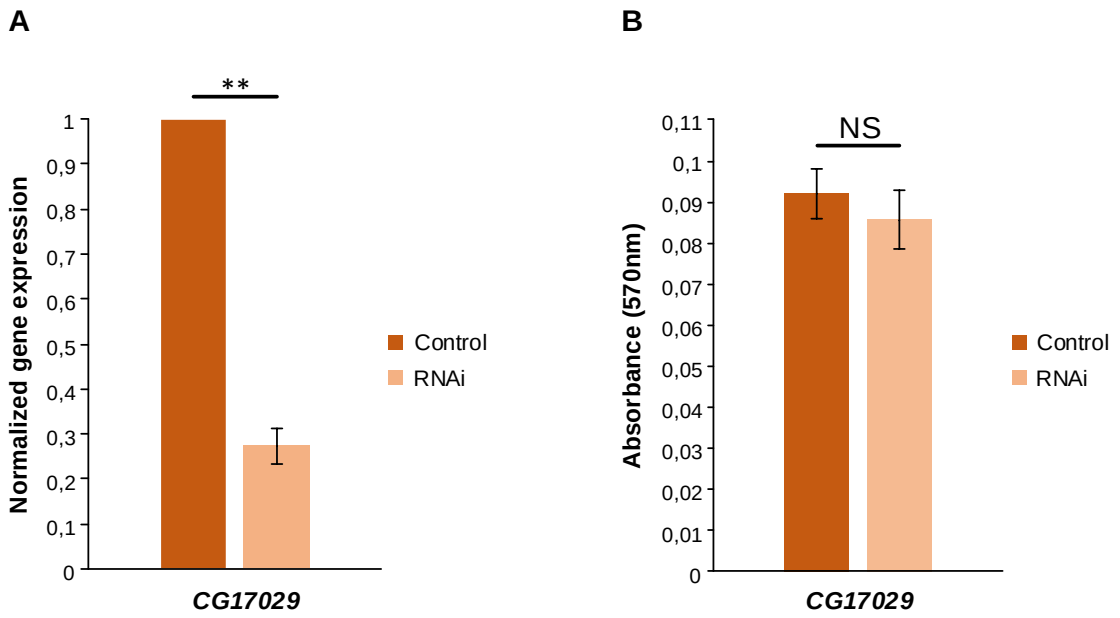


Figure S3.9. Results of the RT-PCR (A) and colorimetric (B) experiments performed for the *CG17029* control and RNAi samples. Samples labelled as "Control" are representative of flies without RNAi expression, while the "RNAi" label concerns flies with RNAi-derived gene suppression. The gene expression of control samples was normalized to 1 and included in the graph to facilitate the observation of gene expression change in RNAi samples. Gene suppression was achieved with statistical significance (** indicates a p-value <0,01 under a heteroscedastic two-tailed t-test).

Chapter 4. Phylogenetic inference of AOs presence within bacteria using auto-phylo

This chapter contains the following conference paper:

López-Fernández, H., Pinto, M., Vieira, C.P., **Duque, P.**, Reboiro-Jato, M., & Vieira, J. (2023). Auto-phylo: A Pipeline Maker for Phylogenetic Studies. In: Rocha, M., Fdez-Riverola, F., Mohamad, M.S., Gil-González, A.B. (eds) *Practical Applications of Computational Biology and Bioinformatics, 17th International Conference (PACBB 2023)*. PACBB 2023. Lecture Notes in Networks and Systems, vol 743. Springer, Cham. https://doi.org/10.1007/978-3-031-38079-2_3

This article is here reproduced with permission from Springer Nature, according to the following clause available at https://www.springernature.com/gp/partners/rights-permissions-third-party-distribution:

“Springer Nature Book and Journal Authors have the right to reuse the Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution. Authors must properly cite the published work in their thesis according to current citation standards and include the following acknowledgement: 'Reproduced with permission from Springer Nature'.”



Auto-phylo: A Pipeline Maker for Phylogenetic Studies

Hugo López-Fenández[1,2], Miguel Pinto[3,4], Cristina P. Vieira[3,5], Pedro Duque[3,4,5,6], Miguel Reboiro-Jato[1,2], and Jorge Vieira[3,5(✉)]

[1] CINBIO, Department of Computer Science, ESEI—Escuela Superior de Ingeniería Informática, Universidade de Vigo, 32004 Ourense, Spain

[2] SING Research Group, Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, 36213 Vigo, Spain

[3] Instituto de Investigação e Inovação em Saúde (I3S), Universidade do Porto, Rua Alfredo Allen, 208, 4200-135 Porto, Portugal

jbvieira@i3s.up.pt

[4] Faculdade de Ciências da Universidade do Porto (FCUP), Rua do Campo Alegre, S/N, 4169-007 Porto, Portugal

[5] Instituto de Biologia Molecular e Celular (IBMC), Rua Alfredo Allen, 208, 4200-135 Porto, Portugal

[6] School of Medicine and Biomedical Sciences (ICBAS), Porto University, Rua de Jorge Viterbo Ferreira, 228, 4050-313 Porto, Portugal

Abstract. Inferences on the evolutionary history of a gene can provide insight into whether the findings made for a given gene in a given species can be extrapolated to other species, including humans, help explain morphological evolution or give an explanation for unexpected findings regarding gene expression suppression experiments, among others. The large amount of sequence data that is already available, and that is predicted to dramatically increase in the next few years, means that life science researchers need efficient automated ways of analyzing such data. Moreover, especially when dealing with divergent sequences, inferences can be affected by the chosen alignment and tree building algorithms, and thus the same dataset should be analyzed in different ways, reinforcing the need for the availability of efficient automated ways of analyzing the sequencing data. Therefore, here, we present auto-phylo, a simple pipeline maker for phylogenetic studies, and provide two examples of its utility: one involving a small already formatted sequenced dataset (41 CDS) to determine the impact of the use of different alignment and tree building algorithms in an automated way, and another one involving the automated identification and processing of the sequences of interest, starting from 16550 bacterial CDS FASTA files downloaded from the NCBI Assembly RefSeq database, and subsequent alignment and tree building inferences.

Keywords: pipeline maker · file processing · phylogenetics · docker

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
M. Rocha et al. (Eds.): PACBB 2023, LNNS 743, pp. 24–33, 2023.
https://doi.org/10.1007/978-3-031-38079-2_3

1 Introduction

Inferences on the evolutionary history of a gene or gene family can offer insight into the function of homologous genes in different organisms, and thus provide insight into whether the findings made for a given species can be extrapolated to other species, including humans, by highlighting instances of gene duplication leading to hypofunctionalization, subfunctionalization, or neofunctionalization (see [1] for a review). They also help explain morphological evolution (see for instance, [2]), and can provide an explanation for why the suppression of the expression of a given gene has no apparent consequences (see for instance, [3]), among others.

The large amount of data that is already available, and that is predicted to drastically increase in the next seven years, means that life science researchers need efficient automated ways of analyzing such data. Indeed, the Earth Biogenome Project aims at obtaining high-quality reference genome sequences for all 1.8 million named living eukaryote species until 2030 [4, 5], giving an opportunity to understand the evolution of any eukaryote gene with unprecedented detail. Nevertheless, parsing a large number of FASTA files in order to have them with the desired format for the phylogenetic analyses is a challenge per se, especially for users without an informatics background. In this context, SEDA (SEquence DAtaset builder; [6]), an open-source, multiplatform application, with an easy-to-use GUI, that allows performing a large number of operations on FASTA files containing DNA and protein sequences, is useful. The GUI, however, cannot be incorporated into automated pipelines. For this reason, in SEDA v1.6.0, which is still under development, a new command-line interface (CLI) that allows using SEDA's operations in the context of automated pipelines will be introduced. Almost all operations (29) have been already migrated to the CLI, with the only exception of the PfamScan protein annotation operation, some BLAST operations (3; "BLAST" is already implemented), and some gene annotation operations (4; "CGA" is already implemented). The CLI development is expected to be completed by the end of 2023 along with a generic way of creating SEDA-CLI pipelines based in Compi [7]. A Docker image is, however, already available for the current SEDA-CLI development version (tagged as "pegi3s/seda:1.6.0-v2304"), at the pegi3s Bioinformatics Docker Images Project [8] with the already implemented operations. We have taken advantage of the pegi3s/seda:1.6.0-v2304 Docker image to develop several auto-phylo modules that are useful in a variety of situations.

When dealing with distantly related sequences, as it is often the case when analyses are performed at the kingdom level, it is important to show that the inferred evolutionary history does not depend on the alignment or phylogenetic method used, which usually implies the installation of different software packages. Having a single software application, that easily allows the creation of automated pipelines, that can be downloaded when needed and erased when no longer needed, and where the installed scientific software is ready to run, would facilitate such analyses. The latter two conditions can be achieved by using Docker images, such as those available at the pegi3s/Bioinformatics Docker Images Project [8].

The availability of a SEDA-CLI version and other relevant images available at the pegi3s/Bioinformatics Docker Images Project allowed the development of the auto-phylo application, here presented, that is made available as a Docker image. This application allows users to easily create automated pipelines, by only specifying the auto-phylo

module/s to be used, as well as the input and output directories for each module. While researchers with an informatics background may easily combine different informatic tools, life science researchers without an informatics background do not have such knowledge, but can easily use auto-phylo. At present, auto-phylo includes modules for (i) tblastx, ii) parsing of CDS FASTA files downloaded from the NCBI Assembly database, (iii) checking of unwanted genomes, (iv) sequence disambiguation, (v) header prefixing, (vi) adding taxonomy information to sequence headers, (vii) removal of stop codons, (viii) merging FASTA files, (ix) two alignment methods (Clustal Omega [9] and T-Coffee [10]; in the case of CDS data, the alignment is made at the amino acid level, and the corresponding nucleotide alignment is then obtained), (x) seven phylogenetic inference methods (FastTree [11], MrBayes [12], Maximum Likelihood, Neighbor-Joining, Minimum Evolution, Maximum Parsimony, and UPGMA as implemented in MEGA-CC [13]) with the possibility of rooting the trees using RootDigger [14], (xi) JModel testing (to check the appropriateness of the models implemented in FastTree and MrBayes; [15]), and (xii) Phylogenetic Tree Collapser (that allows the automated collapsing of taxonomically related groups of sequences; available at the pegi3s Bioinformatics Docker Images Project).

2 Material and Methods

The auto-phylo Docker image here described can be used, to address the impact of choosing different alignment and phylogenetic inferences methods when using the same set of sequences. For instance, when addressing the evolution of vitamin C synthesis, Duque et al. [16] could not fully resolve the relationship of the different types of aldonolactone oxidoreductases (AOs), using a set of sequences aligned with MUSCLE ([17], as implemented in T-Coffee [10]), and Bayesian phylogenetic inferences [12], and the use of different alignment and tree inference methods could help resolve this issue. Therefore, the coding sequences (CDS), associated with the accession numbers NP_216287.1, MBI0379321.1, NP_254014.1, MCO4763851.1, MBS1617249.1, HIA12326.1, NEO81287.1, BAY85903.1, MCP8715902.1, PSP48796.1, BAO94255.2, XP_844937.1, XP_001682394.1, NP_848862.1, EDL85374.1, OCT81467.1, XP_002122023.1, XP_005443605.1, XP_024182616.1, XP_024165881.1, XP_008240731.1, XP_008220057.1, XP_004291192.1, XP_004303609.1, NP_190376.1, NP_182199.1, NP_564393.1, NP_182198.2, NP_182197.1, NP_200460.1, XP_023896433.1, XP_023896434.1, XP_023900169.1, XP_023914113.1, XP_043131130.1, XP_002422445.1, NP_013624.1, XP_003651626.1, XP_049145240.1, NP_055577.1, and XP_004926865.1, were retrieved from the NCBI nucleotide database. As alignment algorithms, both the `Clustal_Omega_codons` and the `T-coffee_codons` auto-phylo modules were used. As tree inference methods, the `Fasttree`, `MrBayes`, `me_tree`, `ml_tree`, and `nj_tree` auto-phylo modules were used. Therefore, in total, we obtained 10 different trees. In order to show that the GTR + I + G model used in the `Fasttree` and `MrBayes` modules is adequate, the `JModel_test` auto-phylo module was used. In order to achieve this, the `pipeline` and `config` files shown in Fig. 1 were used. In the `config` file, we specified the SEDA version to be used (seda:1.6.0-v2304), 2,000,000

tree generations and a burn-in of 25% for `MrBayes`, "Complete deletion" and 1000 bootstraps for the Minimum Evolution (`me_tree`), Maximum likelihood (`ml_tree`), and Neighbor-Joining (`nj_tree`) auto-phylo modules, and the option of not rooting the trees (two outgroup sequences are present in the dataset).

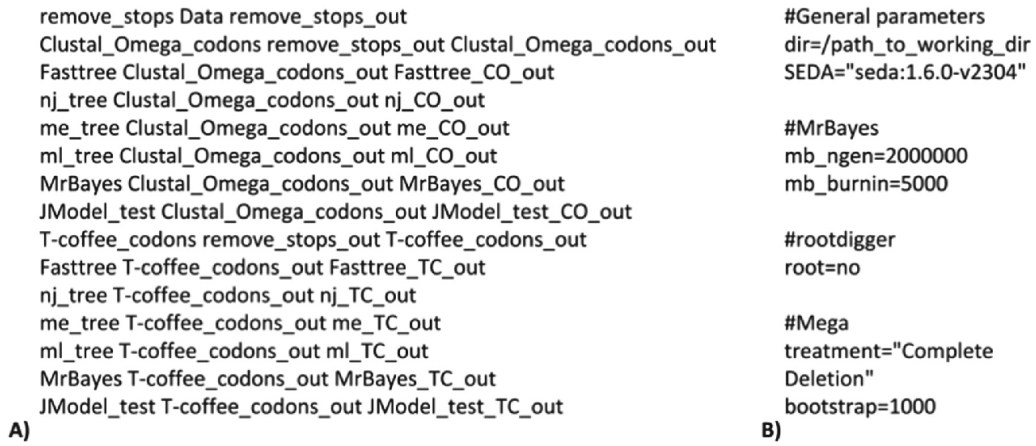


Fig. 1. The auto-phylo `pipeline` (A) and `config` (B) files used to generate 10 gene trees (see Results, Sect. 3.3), using two alignment methods (Clustal Omega and T-Coffee) and five tree inference methods (FastTree, MrBayes, and Maximum Likelihood, Neighbor-Joining, and Minimum Evolution as implemented in MEGA-CC), as well as the `JModel_test` output.

In order to identify putative AO sequences of bacterial origin, that may share similar biological functions with the last enzyme involved in the animal VC synthesis pathway, namely, L-gulonolactone oxidase (GULO) [16], the auto-phylo Docker image here described was also used, with the `pipeline` and `config` files shown in Fig. 2, to perform large scale analyses. Since this protocol has three operations that require an internet connection, namely the `check_contamination`, `add_taxonomy`, and the `tree_collapser` modules, and such connection may sometimes fail, the data processing was divided into six different steps. The query file contains the following sequences: *Mycobacterium tuberculosis* NP_216287, *Pseudomonas aeruginosa* NP_254014, *Streptomyces albiflaviniger* MBI0379321, *Halobacteriales archaeon* PSP48796, *Aspergillus chevalieri* XP_043131130, *Myxococcales bacterium* MCO4763851, *Bacteroidota bacterium* MBS1617249, *Flavobacteriales bacterium* HIA12326, *Mus musculus* NP_848862, *Rattus norvegicus* EDL85374, *Falco cherrug* XP_005443605, *Xenopus laevis* OCT81467, *Ciona intestinalis* XP_002122023, *Moorena sp.* SIO4G3 NEO81287, *Calothrix parasitica* NIES-267 BAY85903, *Trypanosoma brucei brucei* TREU927 XP_844937, and *Leishmania major* strain Friedlin XP_001682394 sequences, and the reference_file the *Mycobacterium tuberculosis* NP_216287 sequence. By using a value of 100000 for the `isoform_min_word_length` parameter, the remove isoforms step of the `GCF_and_GCA_CDS_processing` module is effectively skipped (bacterial genes do not usually code for isoforms), as no CDS is that long.

3 Results

3.1 Auto-phylo Modules

The 23 modules that are at present available for auto-phylo are based on the pegi3s Bioinformatics Docker Images Project [8] images, as well as on the CLI recently made available as an early access for the SEDA v1.6.0 [6] software. Detailed information on each module can be found at https://hub.docker.com/r/pegi3s/auto-phylo. It should be noted that the input/output type defines the compatibility between the different modules. For instance, the tree inference modules accept as input a single aligned FASTA file, and will fail if given multiple files or non-aligned FASTA files. Therefore, it is important to check the compatibility of the different modules before declaring a pipeline. Moreover, in order to use some modules (such as the `CGF_and_CGA_CDS_processing`), the values of some variables must be declared in the `config` file. It should be noted that depending on the value that is declared, the corresponding operations could have no effect. For instance, when using the `CGF_and_CGA_CDS_processing` module, if the size variation (in percentage) relative to the reference sequence is a very large number (for instance, 100000), the specified pattern is ".", or the isoform minimum word length is greater than the sequence sizes, these operations will have no effect, since all sequences will always match such criteria. When processing a large number of files, in order to avoid out of memory errors, it is advisable to use the `split` option. This is a special command, that is invoked after a regular command, and that takes as input a single argument, namely, the number of groups to consider (see Fig. 2 for an example). For instance, the instruction: "`CGF_and_CGA_CDS_processing my_data_dir out_dir split 20`" will split the files that are in the `my_data_dir` directory into 20 equal size subfolders. The data on each subfolder will be processed independently, thus avoiding out of memory errors. The output of all independent analyses will still be saved in the same `out_dir` directory.

While it is foreseeable that, in the future, further modules will be provided for other commonly used alignment and tree inference methods, the way FASTA files should be parsed to have the format desired by the user depends on the source database as well as in the user needs. The script here provided for parsing CDS files downloaded from the NCBI Assembly database likely accommodates most user needs. Nevertheless, the structure of a basic auto-phylo script is relatively simple (http://evolution6.i3s.up.pt/static/auto-phylo/docs/), and thus, even a researcher with very basic knowledge on Bash scripting should be able to write a simple module to parse FASTA files, using SEDA-CLI operations. Assuming that such module is named `my_module` and that it is located in the working directory (`/your/data/dir`), it can be copied into the Docker image, and then be invoked as any other auto-phylo module, with the following command:

```
docker run --rm \
  -v /your/data/dir:/data \
  -v /var/run/docker.sock:/var/run/docker.sock \
  pegi3s/auto-phylo \
  bash -c "cp /data/my_module /opt && /opt/main"
```

3.2 Setting up an Auto-phylo Pipeline

In order to set up a pipeline, using the available auto-phylo modules, the user only needs to edit two text files, named `pipeline` and `config` that must be present in the working directory. In the `pipeline` file, the user declares the order by which the operations should be invoked and the name of the respective input and output directories, one instruction per line. Examples of `pipeline` and `config` files are given in Figs. 1 and 2. The output directory of the previous operation is usually the input directory of the next one, but it is also possible to declare branched pipelines, such as the one in Fig. 1. Intermediate files that are produced during the processing of the data may contain information that is potentially relevant, and thus, a prefix made of a block identifier, command identifier, and step is added to a general name describing the operation, so that the user can easily determine the order of the operations. A block is defined as a set of consecutive commands where the output directory of the previous command is the input directory of the next one. When this condition no longer holds, the next set of instructions is defined as a new block and the block identifier increases by one unit. A file named `blocks_and_commands` indicating where the intermediary files have been saved is also produced. It should be noted that, although intended to be used as a pipeline maker, auto-phylo can also be used to execute a single command at a time. For instance, when processing thousands of files, the operations that are dependent on an internet connection may sometimes fail due to a broken connection, meaning that the user must repeat the operation for the files for which the operation failed. For convenience, such files are identified and saved on a separate folder, in the same place where the corresponding intermediate files are located.

A)
```
check_contamination my_data no_contamination_out
check_contamination files_that_failed no_contamination_out
tblastx no_contamination_out tblastx_out
```

B)
```
CGF_and_CGA_CDS_processing tblastx_out SEDA_processing_out split 20
add_taxonomy SEDA_processing_out taxonomy_out
```

C)
```
add_taxonomy files_that_failed taxonomy_out
```

D)
```
merge  taxonomy_out merged_out
prefix  merged_out prefixed_out
tblastx prefixed_out second_tblastx_out
prefix_out second_tblastx_out prefix_out
Clustal_Omega_codons prefix_out CO_out
Fasttree  CO_out tree_out
tree_collapse tree_out collapsed_tree_out
```

E)
```
tree_collapse tree_out collapsed_tree_out
```

F)
```
#General parameters
dir=/path_to_working_dir
SEDA="seda:1.6.0-v2304"

#tree_collapser
tc_taxonomy=taxonomy
tc_stsm=sequence_to_species_mapping
```

G)
```
#General parameters
dir=/path_to_working_dir
SEDA="seda:1.6.0-v2304"

#Blast related parameters
query=query
expect=0.05

#SEDA-cli
taxonomy=superkingdom
category=Bacteria
isoform_min_word_length=100000
isoform_ref_size=1300
start_codon="ATG GTG TTG CTG ATT ATC"
reference_file=reference_file
max_size_difference=10
pattern="HW[AG]K"
codon_table=1
taxonomy_header="Class Family"

#Fasttree
root=yes
mode=exhaustive
```

H)

Fig. 2. The six (A-F) auto-phylo `pipeline` files and two `config` files (G–H) used. The `config` file shown in G) was used with the `pipeline` files shown in A)–E), while the `config` file shown in H) was used with the `pipeline` file shown in F).

If parameters need to be passed to the scripts being used, they must be declared in the `config` file, one per line. Examples of such files are given in Figs. 1 and 2. It should be noted that a file named `config` must always be given, since at least the working directory location where the folder with the input files is located and where results will be saved, as well as the SEDA Docker image to be used, must be always declared.

3.3 Bacterial AOs May Have a Function Similar to Animal GULOs

When performing Bayesian phylogenetic inferences, using a set of sequences aligned using MUSCLE [17], Duque et al. [16] could not fully resolve the relationship of the different types of AOs. Here, by using a simple branched `pipeline` file (see Material and Methods), we explored the consequences of using different alignment and tree building methods. Although in Fig. 3 we show that the chosen alignment and tree building method can have an impact on the conclusions that are taken, overall, the gene trees do support the hypothesis that there is a group of sequences present in bacteria, especially in Cyanobacteria, that may have a function similar to animal GULOs. Given the small number of sequences used, it is nevertheless unclear whether these are the result of gene transfer events or if there are large bacterial groups where the majority of species shows such sequences.

3.4 Identification of Bacterial Species Groups that Have AOs Closely Related to Animal GULOs

Given the result obtained in Sect. 3.3 we now search for all bacterial species groups that have putative AOs that may play functions similar to GULO [16]. Such characterization can shed light not only on the evolutionary origin of eukaryote *GULO*, but also can open new perspectives for industrial applications [16].

The NCBI Assembly RefSeq database was queried for representative bacterial genomes with an annotation, in order to download one genome per species (16550 on the fourth of March 2023). For eight species a consistent “500 Server Error” was obtained, and thus those files were not used. Of the remaining species, 15910 species had at least one tblastx hit with one of the 17 nucleotide sequences present in the query file (see Material and Methods). After removing sequences showing ambiguous nucleotides, 15908 species remain. The removal of sequences with a length not multiple of 3, that do not have a valid start codon, or show in frame stop codons, reduces this number to 15766 species, and the removal of sequences that are 10% longer or shorter than the reference sequence (the *Mycobacterium tuberculosis* NP_216287 sequence), further reduces this number to 13215 species. After removing sequences that do not show the typical AO "HW[AG]K" pattern, 3888 species remain. In these files, there are 5384 sequences in total. An attempt to align these 5384 sequences revealed several highly divergent sequence groups. Therefore, a second tblastx, using the same query file as before, was performed. Because the size of the blast database is now much smaller, now only 2143 sequences show a significant hit. These were aligned using Clustal Omega [9], a tree obtained using FastTree [11], and collapsed using the Phylogenetic Tree Collapse program (available at the pegi3s Bioinformatics Docker Images Project [8]).



Fig. 3. Inferences on AO gene evolution using different alignment (CO – Clustal Omega; TC – T-Coffee) and tree building algorithms (FT – FastTree; MB – MrBayes; ME – Minimum Evolution; ML – Maximum Likelihood; NJ – Neighbor-Joining). The number of species in each group is indicated within brackets. Group names as in [16].

Due to the large number of sequences involved, JModel test could not be run. The 2143 sequences represent 2073 species, since 70 species contribute with two sequences (32 *Streptomyces*, 7 *Streptacidiphilus*, 5 *Nocardioides*, 4 *Mycobacterium*, 3 *Parvibaculum* and *Amycolatopsis*, 2 *Burkholderia*, and one *Actinomadura*, *Gordonia*, *Gottfriedia*, *Hydrocarboniphaga*, *Kitasatospora*, *Nocardia*, *Phenylobacterium*, *Pseudomaricurvus*, *Skermania*, *Solimonas*, *Stenotrophobium*, *Streptomonospora*, *Tepidicaulis* and *Terricaulis* species). The results show that while some taxa are well represented in the resulting phylogeny, others are not and are likely the result of horizontal gene transfer events, as it is the case for all taxa that are represented by less than 10 species in Fig. 4A. Nevertheless, even taxa that are well represented, such as the *Glycomycetales* (Fig. 4B), can

show evidence for horizontal gene transfer events (in this case 2 events). The obtained pattern is thus complex, and needs further validation that is beyond the scope of this work.

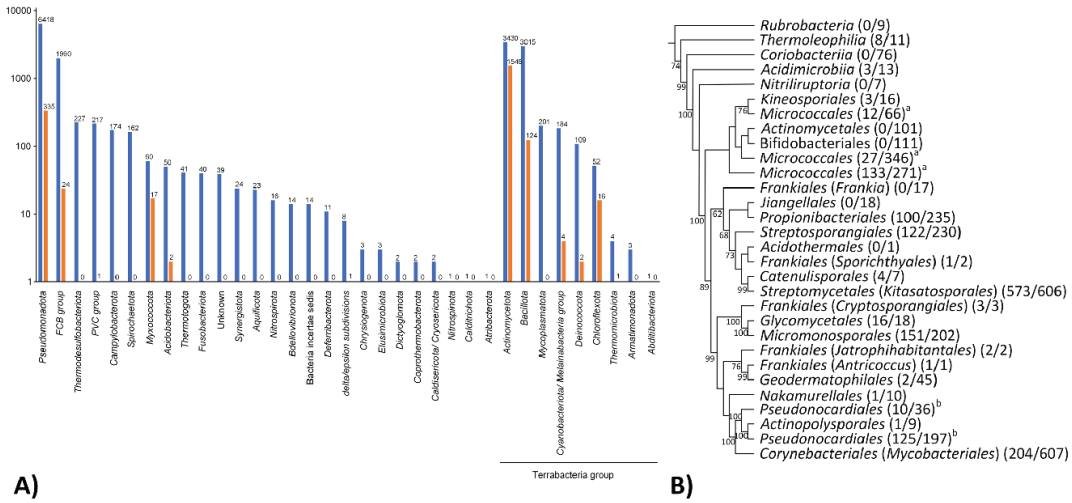


Fig. 4. Bacterial taxa showing sequences with homology to AOs. A) Blue and orange bars represent the total number of species analyzed and the number of species showing AOs, respectively. B) For *Actinomycetota* the number of species showing AOs and the total number of species analyzed is shown on a cladogram, based on the tree presented by [18]. [a] 37 (in the phylogeny) and 129 species (total) could not be placed in the cladogram. [b] 7 (in the phylogeny) and 12 species (total) could not be placed in the cladogram. For *Motilibacterales*, *Nanopelagicales*, one unclassified *Actinomycetes* species, and one unclassified *Actinomycetota*, 0 and 5, 0 and 9, 0 and 1, and 0 and 1 (in the phylogeny and in total respectively) are not shown, respectively.

4 Conclusion

The auto-phylo Docker image allows the creation of complex automated pipelines with little effort, since the user just needs to declare in the `pipeline` file the order by which the operations should be performed and the name of the respective input and output directories, one instruction per line. In the `config` file the user specifies the mandatory parameters (path to working directory and SEDA version to be used), as well as other parameters that may be required by the auto-phylo modules. Two biological examples on the evolution of AOs show the usefulness of auto-phylo.

Acknowledgments. This research was financed by the National Funds through FCT—Fundação para a Ciência e a Tecnologia, I.P., under the project UIDB/04293/2020, and by the Consellería de Educación, Universidades e Formación Profesional (Xunta de Galicia), under the scope of the strategic funding ED431C2018/55-GRC and ED431C 2022/03-GRC Competitive Reference Group. PD is supported by a PhD fellowship from Fundação para a Ciência e Tecnologia (SFRH/BD/145515/2019), co-financed by the European Social Fund through the Norte Portugal Regional Operational Programme (NORTE 2020).

References

1. Birchler, J.A., Yang, H.: The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. Plant Cell **34**(7), 2466–2474 (2022)
2. Merabet, S., Carnesecchi, J.: Hox dosage and morphological diversification during development and evolution. In: Seminars in Cell and Developmental Biology. Elsevier (2022)
3. e Silva, R.S., et al.: The Josephin domain (JD) containing proteins are predicted to bind to the same interactors: implications for spinocerebellar ataxia type 3 (SCA3) studies using Drosophila melanogaster mutants. Front. Mol. Neurosci. **16** (2023)
4. Gupta, P.K.: Earth Biogenome project: present status and future plans. Trends Genet. (2022)
5. Lewin, H.A., et al.: Earth BioGenome project: sequencing life for the future of life. Proc. Natl. Acad. Sci. **115**(17), 4325–4333 (2018)
6. López-Fernández, H., et al.: SEDA: a desktop tool suite for FASTA files processing. IEEE/ACM Trans. Comput. Biol. Bioinf. **19**(3), 1850–1860 (2020)
7. López-Fernández, H., et al.: Compi: a framework for portable and reproducible pipelines. PeerJ Comput. Sci. **7**, e593 (2021)
8. López-Fernández, H., Ferreira, P., Reboiro-Jato, M., Vieira, C.P., Vieira, Jorge: The pegi3s bioinformatics docker images project. In: Rocha, M., Fdez-Riverola, F., Mohamad, M.S., Casado-Vara, R. (eds.) PACBB 2021. LNNS, vol. 325, pp. 31–40. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-86258-9_4
9. Sievers, F., Higgins, D.G.: Clustal omega. Current Protocols Bioinform. **48**(1), 3.13. 1–3.13. 16 (2014)
10. Notredame, C., Higgins, D.G., Heringa, J.: T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. **302**(1), 205–217 (2000)
11. Price, M.N., Dehal, P.S., Arkin, A.P.: FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. **26**(7), 1641–1650 (2009)
12. Huelsenbeck, J.P., Ronquist, F.: MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17**(8), 754–755 (2001)
13. Kumar, S., et al.: MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. Bioinformatics **28**(20), 2685–2686 (2012)
14. Bettisworth, B., Stamatakis, A.: Root digger: a root placement program for phylogenetic trees. BMC Bioinform. **22**(1), 1–20 (2021)
15. Darriba, D., et al.: jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods **9**(8), 772 (2012)
16. Duque, P., Vieira, C.P., Vieira, J.: Advances in novel animal vitamin c biosynthesis pathways and the role of prokaryote-based inferences to understand their origin. Genes **13**(10), 1917 (2022)
17. Edgar, R.C.: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**(5), 1792–1797 (2004)
18. Nouioui, I., et al.: Genome-based taxonomic classification of the phylum Actinobacteria. Front. Microbiol. **9**, 2007 (2018)

Chapter 5. General discussion

5.1. Initial background and objectives synopsis

The loss of *GULO,* the gene that encodes for the enzyme that catalyses the final step of the characterized animal VC synthesis pathway, was formerly associated with VC auxotrophy in species such as humans and teleosts (Drouin *et al.* 2011). Many protostomian lineages have also lost *GULO*, suggesting that many basal animal species should not be able to synthesize VC (Henriques *et al.* 2019, López-Fernández *et al.* 2019). Nevertheless, although not commonly discussed, the preconceived notion of *GULO*-derived VC auxotrophy in protostomians was previously challenged in a small number of reports. In those, evidence for endogenous VC synthesis in insects, namely *D. melanogaster* (Henriques *et al.* 2019, Massie *et al.* 1991) and *B. mori* (Hou *et al.* 2021), and the nematode *C. elegans* (Patananan *et al.* 2015, Yabuta *et al.* 2020), was presented. However, the molecular mechanisms behind the capacity for VC production in *D. melanogaster* remained unknown. Considering that the mechanistic role of VC in important cellular processes related to human health and development is mostly undiscovered, it was therefore argued that further characterization of VC synthesis in models organisms could provide an experimental advantage in future efforts to describe novel VC functional properties.

Independently of the source of VC, the homeostasis of this nutrient is imperative for optimal cellular function in animal species (Burzle *et al.* 2013, Rivas *et al.* 2008, Savini *et al.* 2008). In vertebrates, this stability is mainly maintained by active VC transport mediated by SVCT1 and SVCT2 (Savini *et al.* 2008), while two other orphan transporters, SVCT3 and SVCT4, are not involved in the process (Burzle *et al.* 2013, Yamamoto *et al.* 2010). Although *SVCT* homologs similar to the vertebrate *SVCT4* were previously found in protostomians, none similar to the vertebrate *SVCT1* or *SVCT2* were identified in these species (Kourkoulou *et al.* 2018), raising questions regarding the existence of dedicated VC transport mechanisms in basal animal lineages. In this context, considering that SVCT4 is capable of nucleobase transport (Yamamoto *et al.* 2010), it was hypothesized that vertebrate *SVCT*s could be derived from an ancestral *SVCT* that encoded a protein with capacity for active VC and nucleobase transport, which underwent subfunctionalization after duplication in vertebrates. Under such an evolutionary model, the existence of protostomian SVCTs could ultimately translate to the development of novel experimental setups that are adequate to verify the functional relevance of VC, using model organisms such as *D. melanogaster* or *C. elegans*.

The divergence of AOs from prokaryotic sources has recently emerged as a possible evolutionary scenario that could explain the existence of the distinct eukaryote VC synthesis pathways (Duque *et al.* 2022b). A small phylogenetic screening further hinted to this

possibility, revealing that putative AOs are present in bacteria and archaea species (Duque *et al.* 2022b). Nevertheless, this hypothesis was never explored in a large-scale genomic context, despite the possibility of leading to future improvements of the costly two-step fermentation process applied in the industrial production of VC. Therefore, the phylogenetic analysis of a large bacterial CDS dataset could uncover the general distribution of AOs in these prokaryotes, ultimately dismissing or validating such theoretical framework.

Following on this background and the expertise available at the host lab, the work performed in this thesis focused on two main objectives: i) the characterization of the *D. melanogaster* VC synthesis pathway and ii) the identification of possible VC transporters present in basal animal species. Additionally, the distribution pattern of AOs in bacterial lineages was also addressed.

5.2. The identification of genes related to VC synthesis in *D. melanogaster*

5.2.1. *Regucalcin* and *Dca* characterization efforts

The available literature suggested that the ability to synthesize VC without the intervention of GULO could have arisen multiple times in protostomians. Considering that the insect *B. mori* relies on the functional adaptation of DHCR24 to replace GULO (Hou *et al.* 2021), it was therefore possible that the ability for VC synthesis in other basal animal species could derive from the conservation of the remaining animal VC synthesis pathway in association with a GULO-like enzyme replacement. Thus, the initial methodological efforts concerned the identification of homologs of the gene that encodes for the enzyme that precedes GULO in the animal VC synthesis pathway, Regucalcin, in a generalized animal phylogenetic context. The phylogenetic analysis revealed for the first time that, when present, *Regucalcin* tends to be a single copy gene in most animal lineages (Duque *et al.* 2022a). This result indicated that the most likely fate of any *Regucalcin* duplicates in vertebrates was to be lost, evidenced by multiple independent gene losses after a duplication event that occurred in the common ancestor of reptiles, birds and mammals (Duque *et al.* 2022a). Therefore, despite playing crucial roles in VC synthesis, calcium homeostasis, and oxidative stress response in deuterostomes, *Regucalcin* rarely evolves a new function in these species (Duque *et al.* 2022a). Nevertheless, the particularly high prevalence of *Regucalcin* homologs in insects and lophotrochozoans hinted to the possible emergence of alternative functions of this gene lineage in some basal animals (Duque *et al.* 2022a). This hypothesis correlated well with previous reports that suggested novel roles for *Regucalcin* homologs in receptor-mediated uptake of hexamerin storage proteins (Hansen *et al.* 2002), response to cold exposure (Arboleda-Bustos and Segarra 2011, Goto 2000), and wing size variation (Lee *et al.* 2011, McKechnie *et al.* 2010) in insects. Moreover,

such interpretation was also supported by the inference of PSS in several protostomian *Regucalcin* lineages (Duque *et al.* 2022a). Nevertheless, this phylogenetic information was insufficient to highlight potential events of *Regucalcin* subfunctionalization after duplication and the specific functional roles of *Regucalcin* homologs. Such knowledge was particularly relevant in relation to the VC synthesis pathway of *D. melanogaster*, as *Regucalcin* is duplicated in this model organism, and its duplicate, Dca, could be related to VC synthesis after cold exposure conditions (Duque *et al.* 2022b). The *Actin 5C*-mediated constitutive suppression of *Regucalcin* or *Dca* was thus performed to verify the role of both genes in the VC synthesis context. However, despite highlighting that both genes are required for *D. melanogaster* viability independently of the presence of exogenous VC sources (Duque *et al.* 2022a), this methodology was not informative for further VC synthesis pathway characterization efforts in this organism. An additional in-silico approach reliant on the comparison of reported interactomes allowed to overcome some difficulties in the inference of the functional roles of Regucalcin and Dca in *D. melanogaster*, suggesting for the first time that both are involved in calcium homeostasis and oxidative stress response (Duque *et al.* 2022a). Nonetheless, the available interactome information was insufficient to implicate either gene in the VC synthesis process in this species at the time (Duque *et al.* 2022a).

5.2.2. Extended search for other gene candidates: The identification of *Miox* as a contributor to VC synthesis in *D. melanogaster*

The observation that the distantly related *E03H4.3* can functionally replace *Regucalcin* in *C. elegans* hinted that VC synthesis could alternatively derive from multiple enzymatic adaptations in protostomes. Furthermore, given that Miox-derived myo-inositol metabolism contributes to the endogenous synthesis of this nutrient in vertebrates (Hänninen *et al.* 1971), it was also possible that this substrate was related to the process in these basal animals. Thus, following on the lack of informative results regarding the functional characterization attempts of *Regucalcin* and *Dca,* the possibility of alternative VC synthesis mechanisms in *D. melanogaster* was additionally taken in consideration to extend the scope of pathway identification efforts. The combined utilization of phylogenetic analyses, colorimetric assays, chromosomal deletion stocks and RNAi-mediated suppression of relevant genes proved efficient for this purpose. The presented phylogenetic data has shown for the first time that genes related to the characterized animal VC synthesis pathway, and the myo-inositol pathway, are generally conserved in insect species. Nevertheless, *AKR1A1* homologs are absent in *D. melanogaster,* and thus *AKR1A1* was likely replaced by *AKR1B1*-like genes in this species (Duque *et al.* 2022b). This is a parsimonious explanation, since species from the Sophophora subgenus, such as *D.*

melanogaster, have a particularly high prevalence of *AKR1B1* duplications. Such observation also suggests that *AKR1B1* subfunctionalization or neofunctionalization processes could have originated advantageous phenotypes in these species. If evolutionary constraints led to the adaptation and subsequent efficiency increase of AKR1B1-like enzymes in the context of VC synthesis remains, however, to be determined. The adaptation of DHCR24 as functional replacement of GULO for VC synthesis purposes was excluded within Brachycera species due to the loss of all *DHCR24* homologs, and thus this gene lineage does not contribute to the VC synthesis pathway in *D. melanogaster*.

Complementary colorimetric data gathered from the BDSC deficiency stocks showed that 1480 partially deleted genes were associated with VC absorbance values equal to or below 50% of the 0,09 value obtained for the control *D. melanogaster* Oregon-R strain. This unexpected result was not revealing in regards to the presence of possibly divergent VC synthesis mechanisms in *D. melanogaster*, and hinted to possible technical limitations imposed on the applied methodology by the distinct genomic backgrounds of the utilized deficiency stocks, as well as the presence of extensive chromosomal deletions. Despite this limitation, the data still highlighted that *Pgm1* and *HEX 2 IMPA* homologs *CG17026* to *CG17029* could be related to VC synthesis in *D. melanogaster*, ultimately suggesting that either D-Glucose-6-P or myo-inositol metabolism were essential for the process. The subsequent RNAi-mediated suppression of *Pgm1* did not allow to disclose the role of this gene in VC synthesis under the utilized experimental conditions, likely due to insufficient expression knockdown in order to produce a phenotypic effect in total VC levels. However, the RNAi-mediated suppression of *Miox* used as a technical alternative to overcome issues with the knockdown screening of *HEX 2 IMPA* homologs implicated this gene in the VC synthesis process of *D. melanogaster*. This result has shown for the first time that, as in vertebrates capable of VC synthesis in the kidney, myo-inositol can be a VC precursor in basal animal species. Furthermore, it also indicated that the conversion of this substrate to D-Glucuronate catalysed by Miox intervenes in the synthesis pathway, thus highlighting the evolutionary conservation of this enzymatic step between deuterostomes and protostomes. This finding suggests that, although commonly disregarded, VC synthesis derived from myo-inositol is a likely ancestral trait in animal species. Given that the metabolic contribution of Miox across the five experimental conditions tested always accounted for under 26% of total VC levels in *D. melanogaster*, it is clear that other VC synthesis mechanisms are synergistically occurring in this organism. In this sense, further research is required to clarify if any of the characterized animal D-Glucose-6P pathway homologs or other unrelated alternative enzymatic mechanisms are responsible for the majority of VC synthesis in *D. melanogaster*.

5.2.3. Future perspectives on the characterization of the *D. melanogaster* VC synthesis pathways

To improve upon the VC biosynthetic pathway identification efforts in *D. melanogaster*, the RNAi experimental methodology presented in this thesis should be applied to other characterized animal VC synthesis pathway and myo-inositol pathway homologs such as *UGP*, *sgl* and *Inos*, as these genes have no duplicates that could functionally compensate high levels of suppression. The pathway steps possibly catalyzed by more than one protein, as for example the case of the D-Glucuronate conversion to L-Gulonate by AKR1B1 homologs, can also been addressed using the simultaneously targeting of multiple genes using an RNAi system based on the pNP vector (Qiao *et al.* 2018, Wang *et al.* 2019). Moreover, the possibility of a synergistic interaction between the characterized animal VC synthesis pathway and the myo-inositol pathway for the production of VC could also be explored using such approach, as the simultaneous double knockdown of genes related to these pathways (e.g. *Pgm1* and *Miox*) could accentuate the observed decrease in total VC levels in *D. melanogaster*. However, as previously shown and discussed in previous chapters, RNAi-derived approaches are prone to possible technical issues such as progeny lethality or low gene expression knockdown. Therefore, a new combination of distinct GAL4 drivers should also be considered in order to overcome such limitations and even refine the current results. In this perspective, tissue-specific gene suppression might be useful to determine not only the genes involved in the VC synthesis process, but the tissues where this nutrient is synthesized in *D. melanogaster*. VC synthesis in vertebrates capable of doing so is known to occur in the kidney, the liver or both organs depending on the considered species (Drouin *et al.* 2011, Dutta Gupta *et al.* 1973). Coincidently, gene expression information available at the FlyBase database (https://flybase.org/) indicates that the majority of the genes related to the characterized animal VC synthesis pathway and the myo-inositol pathway are considerably expressed in the malpighian tubule system and the fat body in *D. melanogaster*, tissues considered analogous to the vertebrate kidney (Nocelli *et al.* 2016) and liver (Musselman *et al.* 2013), respectively. Thus, it is parsimonious to assume that VC synthesis might be taking place within either or both of these tissues in *D. melanogaster*. Given that GAL4 drivers capable of specific RNAi expression within these tissues are already available (Weaver *et al.* 2020), the role of all relevant genes should also be evaluated considering these hypotheses. If all the presented methodologies are unable to provide a discernable description of the VC synthesis mechanisms in *D. melanogaster*, then the further refinement of the list of 1480 genes obtained from the colorimetric screening of deficiency stocks emerges as a potential starting point in order to identify additional pathway possibilities.

5.3. Identification of SVCT homologs as possible VC transporters in protostomian species

The phylogenetic inferences regarding the evolutionary history of the *SVCT* gene lineage have demonstrated for the first time that vertebrate SVCTs can derive from the known 2R-WGD events that affected species from this taxonomic group (Duque *et al.* 2022a). This finding supported the hypothesis that the biological roles of SVCT1, SVCT2, SVCT3 and SVCT4 are likely the result of subfunctionalization of a single ancestral SVCT after duplication (Duque *et al.* 2022a). A further homology screening within the aligned region of the conserved amino acid motif associated with VC specificity in SVCT1 and SVCT2 (SSSP; (Kourkoulou *et al.* 2018)) has revealed that none of the basal deuterostomian, protostomian and non-bilaterian SVCT sequences contain this motif (Duque *et al.* 2022a). Moreover, the second serine and the proline residues, which are regarded as the most important for substrate affinity within this motif (Kourkoulou *et al.* 2018), were also not simultaneously found in the analysed non-deuterostomian and basal deuterostomian SVCTs (Duque *et al.* 2022a). The amino acid motif present in the vertebrate SVCT4 transporter (SYSE; (Kourkoulou *et al.* 2018)) was, however, found in non-bilaterians and the majority of the basal deuterostomian species, suggesting that these SVCTs are able to transport nucleobases (Duque *et al.* 2022a). Curiously, insect SVCTs display the consensus amino acid motif TFGE, which is similar to the *E. coli* UraA uracil permease motif (TYGE; (Kourkoulou *et al.* 2018)). This finding was interesting when considering that even though the role of UraA in VC transport was seemingly never addressed, VC is transported into *E. coli* cells through the ulaABC system, but also by a still unidentified transporter (Amábile-Cuevas 2023, Campos *et al.* 2007). Since previous research revealed that some NAT transporters with uracil affinity (Goudela *et al.* 2005) can also bind to VC (Kourkoulou *et al.* 2018), UraA should arguably represent a valid contributor to VC transport in *E. coli*. If so, then insect SVCTs could also represent transporters with capacity for nucleobase and VC transport, a hypothesis that correlates well with the dietary VC requirements of *B. mori* (Hou *et al.* 2021) and the presence of a single *SVCT* gene in this species (Duque *et al.* 2022a). Ultimately, this would imply that *CG6293* is the main VC transporter in *D. melanogaster*, but further functional assessment is still required to validate this hypothesis.

5.3.1. Future perspectives on the characterization of the *D. melanogaster* CG6293

Although *CG6293* has a seemingly generalized distribution in *D. melanogaster*'s tissues, it is primarily expressed in the malphigian tubule system and ovaries (https://flybase.org/reports/FBgn0037807). This information correlates in part with the possibility of VC synthesis within malphigian tubules in *D. melanogaster*, as the high

availability of VC transporters in this tissue would facilitate the whole-body distribution of this nutrient through the haemolymph in this species. Alternatively, the high CG6293 presence in this tissue might rather be related to the reabsorption of this nutrient. In this sense, it is known that fly females have approximately three times more total body VC than males, an observation that could not be exclusively explained by the different bodyweight between genders in previous research (Henriques *et al.* 2019). Although not as drastic, female mice also possess more VC than their male counterpart due, in part, to distinct renal VC excretion rates between genders, as females are able to retain more VC from urine than males (Kuo *et al.* 2004). Coincidentally, *CG6293* is expressed approximately three times more in female malphigian tubules, supporting this extrapolation in *D. melanogaster*. If the high expression of *CG6293* in the ovaries is related to VC accumulation requirements is unknown. It is, however, unlikely to concern possible VC synthesis, as genes related to D-Glucose-6P and myo-inositol metabolism in the context of this process in *D. melanogaster* have low expression levels in this tissue.

Considering the available information, the RNAi-mediated suppression of *CG6293* using a specific GAL4 driver expressed in malphigian tubules (e.g. BDSC #91415 (*Uro-GAL4*)) is likely to provide valuable insights on VC homeostasis in *D. melanogaster*. The excision of this tissue in fly individuals with *CG6293* knockdown will be a requirement to simultaneously evaluate the role of this gene in the VC transport context and validate if malphigian tubules are responsible for VC synthesis or reabsorption in this organism.

5.4. Bacterial AOs as indication of possible endogenous VC synthesis in prokaryotes: the first large-scale step and future perspectives

Following on previous efforts regarding the understanding of the evolutionary origins of animal VC synthesis mechanisms (Duque *et al.* 2022b), a search for putative AOs that may play functions similar to GULO in 16542 bacterial species was performed (López-Fernández *et al.* 2023). Interestingly, despite previous literature hinting to the general VC synthesis inability of prokaryotes and the extensive absence of *GULO* homologs in these species (Wheeler *et al.* 2015), the phylogenetic results presented in this work highlight that putative AOs can be found in 2073 bacteria (López-Fernández *et al.* 2023). Although the phylogenetic inference did not allow to disclose the evolutionary origins of the identified AOs, it was clear that the presence of these putative enzymes is particularly high in Actinomycetota species (1545 out of the total 2073 bacteria species) (López-Fernández *et al.* 2023). This suggests that this taxonomic group could be the evolutionary source of AOs in the majority of prokaryotes due to processes of, for example, horizontal gene transfer. Nevertheless, this hypothesis still requires further validation. Interestingly, within the

Actinomycetota, 573 out of 606 analysed Streptomycetales have a putative AO (López-Fernández *et al.* 2023), possibly indicating that a completely functional VC synthesis pathway is present in species from this taxonomic group, and that this nutrient is required for the optimal function of these organisms. This extrapolation is concordant with a previous report that inferred the presence of such enzymes in *Streptomyces coelicolor* and *Streptomyces avermitilis* (Wolucka and Communi 2006), and also correlates well with the notorious capacity of *Streptomyces* species for the production of bioactive metabolites (Bérdy 2005, Jagannathan *et al.* 2021).

The presented phylogenetic data depicted the first large-scale analysis performed for the identification of potential AOs homologs in prokaryotes, and highlighted for the first time that AOs possibly related to VC synthesis are present in a considerable number of bacteria species. Such methodology can be further applied using homology searches for genes associated with the remaining characterized animal, plant and photosynthethic protists VC synthesis pathways, hopefully allowing the phylogenetic reconstruction of the complete bacterial VC biosynthetic mechanisms. Further refinement of such analyses will likely reveal the evolutionary scenario behind the emergence of the divergent VC synthesis pathways in eukaryotes.

5.5. References

Amábile-Cuevas, C. F. **2023.** Ascorbate and Antibiotics, at Concentrations Attainable in Urine, Can Inhibit the Growth of Resistant Strains of *Escherichia coli* Cultured in Synthetic Human Urine. *Antibiotics*. 12:985.

Arboleda-Bustos, C. E. and C. Segarra. **2011.** The *Dca* Gene Involved in Cold Adaptation in *Drosophila melanogaster* Arose by Duplication of the Ancestral *regucalcin* Gene. *Molecular Biology and Evolution*. 28:2185-2195.

Bérdy, J. **2005.** Bioactive Microbial Metabolites. *The Journal of Antibiotics*. 58:1-26.

Burzle, M., Y. Suzuki, D. Ackermann, H. Miyazaki, N. Maeda, B. Clemencon, R. Burrier and M. A. Hediger. **2013.** The sodium-dependent ascorbic acid transporter family SLC23. *Molecular Aspects of Medicine*. 34:436-454.

Campos, E., C. Montella, F. Garces, L. Baldoma, J. Aguilar and J. Badia. **2007.** Aerobic l-ascorbate metabolism and associated oxidative stress in *Escherichia coli*. *Microbiology*. 153:3399-3408.

Drouin, G., J.-R. Godin and B. Pagé. **2011.** The Genetics of Vitamin C Loss in Vertebrates. *Current Genomics*. 12:371-378.

Duque, P., C. P. Vieira, B. Bastos and J. Vieira. **2022a.** The evolution of vitamin C biosynthesis and transport in animals. *BMC Ecology and Evolution*. 22:84.

Duque, P., C. P. Vieira and J. Vieira. **2022b.** Advances in Novel Animal Vitamin C Biosynthesis Pathways and the Role of Prokaryote-Based Inferences to Understand Their Origin. *Genes*. 13:1917.

Dutta Gupta, S., P. K. Choudhury and I. B. Chatterjee. **1973.** Synthesis of l-ascorbic acid from d-glucurono-1,4-lactone conjugates by different species of animals. *International Journal of Biochemistry*. 4:309-314.

Goto, S. G. **2000.** Expression of *Drosophila* homologue of s*enescence marker protein-30* during cold acclimation. *Journal of Insect Physiology*. 46:1111-1120.

Goudela, S., P. Karatza, M. Koukaki, S. Frillingos and G. Diallinas. **2005.** Comparative substrate recognition by bacterial and fungal purine transporters of the NAT/NCS2 family. *Molecular Membrane Biology*. 22:263-275.

Hänninen, O., R. Raunio and J. Marniemi. **1971.** Biosynthesis of L-ascorbic acid from myo-inositol. *Carbohydrate Research*. 16:343-351.

Hansen, I. A., S. R. Meyer, I. Schäfer and K. Scheller. **2002.** Interaction of the anterior fat body protein with the hexamerin receptor in the blowfly *Calliphora vicina*. *European Journal of Biochemistry*. 269:954-960.

Henriques, S. F., P. Duque, H. López-Fernández, N. Vázquez, F. Fdez-Riverola, M. Reboiro-Jato, C. P. Vieira and J. Vieira. **2019.** Multiple independent *L-gulonolactone oxidase* (*GULO*) gene losses and vitamin C synthesis reacquisition events in non-Deuterostomian animal species. *BMC Evolutionary Biology*. 19:126.

Hou, M., T. M. Dai, X. Y. Liang, S. X. Zhang, W. Z. Cui, J. F. Qiu, Y. H. Sima, W. Z. Cui and S. Q. Xu. **2021.** *Bombyx mori* can synthesize ascorbic acid through the l-gulose pathway to varying degrees depending on developmental stage. *Archives of Insect Biochemistry and Physiology*. 106:e21783.

Jagannathan, S. V., E. M. Manemann, S. E. Rowe, M. C. Callender and W. Soto. **2021.** Marine Actinomycetes, New Sources of Biotechnological Products. *Marine Drugs*. 19.

Kourkoulou, A., A. A. Pittis and G. Diallinas. **2018.** Evolution of substrate specificity in the Nucleobase-Ascorbate Transporter (NAT) protein family. *Microbial Cell*. 5:280-292.

Kuo, S.-M., M. E. MacLean, K. McCormick and J. X. Wilson. **2004.** Gender and sodium-ascorbate transporter isoforms determine ascorbate concentrations in mice. *The Journal of nutrition*. 134:2216-2221.

Lee, S. F., Y. Chen, A. K. Varan, C. W. Wee, L. Rako, J. K. Axford, R. T. Good, M. J. Blacket, C. Reuter and L. Partridge. **2011.** Molecular basis of adaptive shift in body size in *Drosophila melanogaster*: functional and sequence analyses of the *Dca* gene. *Molecular Biology and Evolution*. 28:2393-2402.

López-Fernández, H., P. Duque, S. Henriques, N. Vázquez, F. Fdez-Riverola, C. P. Vieira, M. Reboiro-Jato and J. Vieira. **2019.** A Bioinformatics Protocol for Quickly Creating Large-

Scale Phylogenetic Trees. In: F. Fdez-Riverola, M. S. Mohamad, M. Rocha, J. F. De Paz and P. González, editors. *Practical Applications of Computational Biology and Bioinformatics, 12th International Conference.* Cham. Springer. p. 88-96.

López-Fernández, H., M. Pinto, C. P. Vieira, P. Duque, M. Reboiro-Jato and J. Vieira. **2023.** Auto-phylo: A Pipeline Maker for Phylogenetic Studies. In: M. Rocha, F. Fdez-Riverola, M. S. Mohamad and A. B. Gil-González, editors. *Practical Applications of Computational Biology and Bioinformatics, 17th International Conference (PACBB 2023)*. Cham. Springer Nature Switzerland. p. 24-33.

Massie, H. R., M. E. Shumway, S. J. P. Whitney, S. M. Sternick and V. R. Aiello. **1991.** Ascorbic acid in *Drosophila* and changes during aging. *Experimental Gerontology*. 26:487-494.

McKechnie, S. W., M. Blacket, S. Song, L. Rako, X. Carroll, T. Johnson, L. Jensen, S. Lee, C. Wee and A. A. Hoffmann. **2010.** A clinally varying promoter polymorphism associated with adaptive variation in wing size in *Drosophila*. *Molecular ecology*. 19:775-784.

Musselman, L. P., J. L. Fink, P. V. Ramachandran, B. W. Patterson, A. L. Okunade, E. Maier, M. R. Brent, J. Turk and T. J. Baranski. **2013.** Role of Fat Body Lipogenesis in Protection against the Effects of Caloric Overload in *Drosophila*. *Journal of Biological Chemistry*. 288:8028-8042.

Nocelli, R. C. F., P. Cintra-Socolowski, T. C. Roat, E. C. M. Silva-Zacarin and O. Malaspina. **2016.** Comparative physiology of Malpighian tubules: form and function. *Open Access Insect Physiology*. 6:13-23.

Patananan, A. N., L. M. Budenholzer, M. E. Pedraza, E. R. Torres, L. N. Adler and S. G. Clarke. **2015.** The invertebrate *Caenorhabditis elegans* biosynthesizes ascorbate. *Archives of Biochemistry and Biophysics*. 569:32-44.

Qiao, H.-H., F. Wang, R.-G. Xu, J. Sun, R. Zhu, D. Mao, X. Ren, X. Wang, Y. Jia, P. Peng, D. Shen, L.-P. Liu, Z. Chang, G. Wang, S. Li, J.-Y. Ji, Q. Liu and J.-Q. Ni. **2018.** An efficient and multiple target transgenic RNAi technique with low toxicity in *Drosophila*. *Nature Communications*. 9:4160.

Rivas, C. I., F. A. Zúñiga, A. Salas-Burgos, L. Mardones, V. Ormazabal and J. C. Vera. **2008.** Vitamin C transporters. *Journal of Physiology and Biochemistry*. 64:357-375.

Savini, I., A. Rossi, C. Pierro, L. Avigliano and M. V. Catani. **2008.** SVCT1 and SVCT2: key proteins for vitamin C uptake. *Amino Acids*. 34:347-355.

Wang, F., H. H. Qiao, R. G. Xu, J. Sun, R. Zhu, D. Mao and J. Q. Ni. **2019.** pNP Transgenic RNAi System Manual in *Drosophila*. *Bio Protoc*. 9:e3158.

Weaver, L. N., T. Ma and D. Drummond-Barbosa. **2020.** Analysis of Gal4 Expression Patterns in Adult *Drosophila* Females. *G3 Genes|Genomes|Genetics*. 10:4147-4158.

Wheeler, G., T. Ishikawa, V. Pornsaksit and N. Smirnoff. **2015.** Evolution of alternative biosynthetic pathways for vitamin C following plastid acquisition in photosynthetic eukaryotes. *eLife*. 4:e06369.

Wolucka, B. A. and D. Communi. **2006.** *Mycobacterium tuberculosis* possesses a functional enzyme for the synthesis of vitamin C, L-gulono-1,4-lactone dehydrogenase. *Febs j*. 273:4435-4445.

Yabuta, Y., R. Nagata, Y. Aoki, A. Kariya, K. Wada, A. Yanagimoto, H. Hara, T. Bito, N. Okamoto, S. Yoshida, A. Ishihara and F. Watanabe. **2020.** L-Ascorbate Biosynthesis Involves Carbon Skeleton Rearrangement in the Nematode *Caenorhabditis elegans*. *Metabolites*. 10:334.

Yamamoto, S., K. Inoue, T. Murata, S. Kamigaso, T. Yasujima, J. Y. Maeda, Y. Yoshida, K. Y. Ohta and H. Yuasa. **2010.** Identification and functional characterization of the first nucleobase transporter in mammals: implication in the species difference in the intestinal absorption mechanism of nucleobases and their analogs between higher primates and other mammals. *Journal of Biological Chemistry*. 285:6522-6531.