Evolution of the vitamin C biosynthetic pathway and transport mechanism: from the global perspective to a *Drosophila melanogaster* case study

Pedro Miguel Barros Duque
Cell and Molecular Biology Master's Degree
Science Department, FCUP
2018

Supervisor
Jorge Manuel de Sousa Basto Vieira
Main researcher at IBMC
jbvieira@ibmc.up.pt

Co-supervisor
Cristina Alexandra Gonçalves Paula Vieira
Auxiliary researcher at IBMC
cgvieira@ibmc.up.pt

Institute Address
IBMC – Instituto de Biologia Molecular e Celular
I3s building – Rua Alfredo Allen nº 208
4200-135 Porto, Portugal

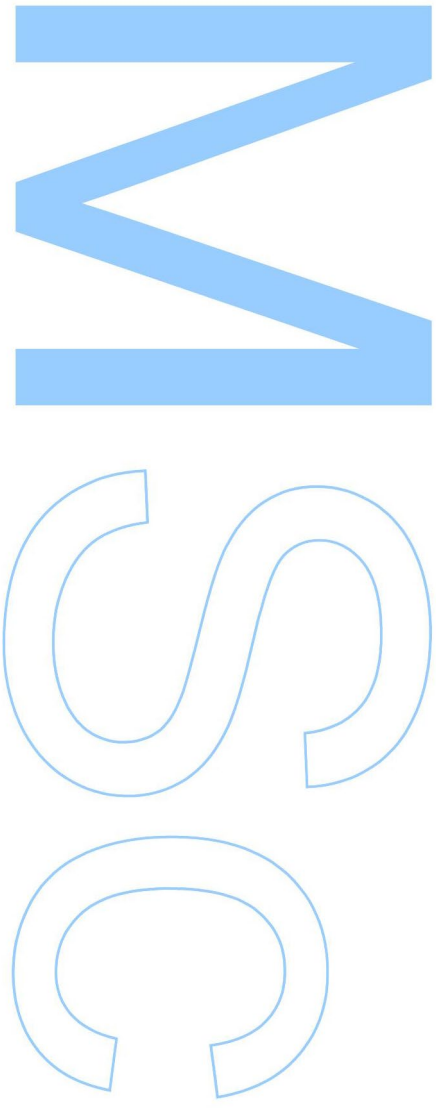
U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

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

O Presidente do Júri,

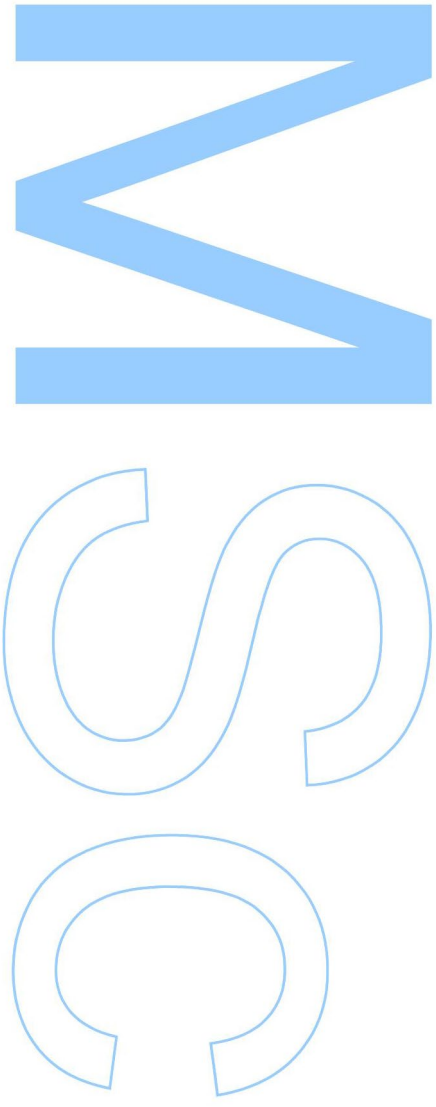
Porto, _____/_____/_____



Declaração de não-plágio

Eu, Pedro Miguel Barros Duque, aluno com o número 201004984, inscrito no mestrado de Biologia Celular e Molecular no presente ano letivo de 2017/2018, declaro por minha honra que sou o autor da totalidade do texto apresentado, não apresento texto plagiado, e tomei conhecimento das consequências de uma situação de plágio.

Porto, 30 de Setembro de 2018



Acknowledgements

This work was performed in the Phenotypic Evolution group at the I3s facilities, as part of the dissertation thesis of the Cellular and Molecular Biology Master´s degree that belongs to the Faculty of Sciences of the University of Porto. My sincere and great thanks to all the colleagues I had the chance to work with, but especially to Sílvia, Sara, Hugo, Pedro, José, Joel, Rodrigo and Vanessa for the knowledge and friendship provided. Also my thanks to FCUP, I3s and IBMC for allowing the opportunity to perform this work.

I also would like to say a special thank you to Jorge and Cristina Vieira for their patience, great advisement and teaching along this proficuous experience.

Furthermore, I want to thank the support of all my friends, but even more the sacrifice of my family, especially my parents, for allowing me to come back to university while always motivating me to become resilient and hard-working without questioning my capabilities. A thank you is not nearly enough, but I hope in time I get a chance to return their vote of confidence in a similar fashion.

Abstract

Ascorbic acid (vitamin C) is a crucial co-factor for several enzymatic reactions and an indispensable antioxidant agent, playing an important role in normal function and development of eukaryotic cells. Many deuterostomians are capable of synthesizing ascorbic acid, however some species such as Haplorrhini primates, teleost fish and *Cavia porcellus* (guinea pig) lost this ability due to the loss of the *L-gulonolactone oxidase* (*GULO*) gene. Protostomians are often regarded as not having *GULO*, although there are hints that this may not be the case. We used all available genomic information in GenBank and RefSeq for animal species to clarify this issue. We show that *GULO* was not lost at the split between the protostomians and deuterostomians, and added supporting evidence that this gene could in fact be present in the ancestral of animals and Fungi. It is consensual that *GULO* was lost in the Insecta taxonomic group, and our results go in accordance with this finding. Nevertheless, ascorbic acid levels can be detected in one representative species from this group, the model organism *Drosophila melanogaster* (fly), even in the absence of a dietary source of this vitamin. Given this evidence, it was possible that the fly microbiome could be responsible for the supply of ascorbic acid. Still, we were able to determine that the microbiome is not responsible for ascorbic acid synthesis. Furthermore, we observed that after cold acclimation conditions, *D. melanogaster* is able to replenish a break in ascorbic acid levels after one day of recovery, which is strong evidence of putative synthesis.

In deuterostomians, specifically in the vertebrates group, ascorbic acid homeostasis is facilitated by the presence of Sodium-dependent Vitamin C Transporters (SVCTs). Four transporters (SVCT1 to 4) have been identified, although only SVCT1 and SVCT2 are known to be correlated with ascorbic acid cellular transport. The evolutionary history of these four transporters remains undiscovered, as well as the ancestral subtract specificity traits, but it is suspected that *SVCT1* and *SVCT2* genes are derived from a common ancestor and that *SVCT3* and *SVCT4* are "orphan genes". Furthermore, one uncharacterized SVCT transporter is also detected in *D. melanogaster*. Using all available animal genome annotations, we sought out to understand the evolutionary history of the vertebrate transporters and its phylogenetic relationship to the *SVCT* transporters observed in protostomian species. We uncovered that within vertebrates, the general presence of four transporters is likely the result of two rounds of whole genome duplication that are already reported in literature, from a single ancestral gene. The protostomian *SVCT* seems to be duplicated independently several times in many distinct taxonomic groups, nevertheless our results also indicate the presence of a single ancestral gene at the base of this taxonomic group. Nevertheless, we were unable to imply the protostomian SVCTs in ascorbic acid transport.

Keywords: *GULO*, *SVCT*, ascorbic acid, *D. melanogaster*, microbiome, synthesis, transport

Resumo

O ácido ascórbico (vitamina C) é um cofator crucial em várias reações enzimáticas e um agente antioxidante indispensável, contribuindo de forma notória para o normal funcionamento e desenvolvimento de células eucarióticas. Muitos deuterostómios são capazes de sintetizar ácido ascórbico, no entanto algumas espécies como os primatas Haplorrhini, os peixes teleósteos e *Cavia porcellus* (porquinho da Índia) não possuem esta capacidade devido há perda do gene *L-gulonolactona oxidase* (*GULO*). É considerado que os protostómios não possuem *GULO*, mas existem pistas que sugerem o contrário. Usámos toda a informação genómica disponível para espécies animais nas bases de dados GenBank e RefSeq para clarificar este assunto. Demonstrámos que o gene *GULO* não foi perdido na divergência entre protostómios e deuterostómios, e acrescentámos evidência da possível presença deste gene no ancentral dos animais e Fungi. É consensual que o gene *GULO* foi perdido no grupo taxonómico Insecta, e os nossos resultados estão de acordo com esta premissa. No entanto, níveis de ácido ascórbico podem ser detetados numa espécie representativa deste grupo taxonómico, *Drosophila melanogaster*, mesmo na ausência de uma fonte desta vitamina na dieta. Dada esta evidência, seria possível que o microbioma fosse responsável pela suplementação de ácido ascórbico. Contudo, fomos capazes de determinar que o microbioma não desempenha esse papel. Observamos ainda que em condições de aclimatação ao frio, *D. melanogaster* é capaz de normalizar uma descida dos níveis de ácido ascórbico após um dia de recuperação, o que é uma forte evidência de produção putativa.

Em deuterostómios, especificamente no grupo dos vertebrados, a homeostasia do ácido ascórbico é facilitada pela presença de Transportadores de Vitamina C dependentes de Sódio (SVCTs). Quatro transportadores (SVCT1 a 4) foram identificados, apesar de apenas os transportadores SVTC1 e SVCT2 estarem implicados no transporte de ácido ascórbico. A história evolutiva destes transportadores permanece desconhecida, bem como as propriedades funcionais ancestrais dos mesmos, mas é extrapolado que os genes *SVCT1* e *SVCT2* derivem de um ancestral comum enquanto os genes *SVCT3* e *SVCT4* sejam "órfãos". Um transportador SVCT não caracterizado foi também descoberto em *D. melanogaster*. Usando todas as anotações de genoma disponíveis para espécies animais, procurámos compreender a história evolutiva dos transportadores identificados em vertebrados e a sua relação filogenética com os transportadores observados em protostómios. Descobrimos que dentro dos vertebrados, a presença de quatro transportadores pode ser correlacionada com duas duplicações totais de genoma reportadas na literatura, a partir de um único gene ancestral. O gene *SVCT* parece estar duplicado independentemente em várias linhagens de protostómios, no entanto os nossos resultados indicam que o mesmo é derivado de uma cópia ancestral única. Todavia, não fomos capazes de implicar os transportadores SVCT de protostómios na possível capacidade de transporte de ácido ascórbico.

Palavras chave: *GULO*, *SVCT*, ácido ascórbico, *D. melanogaster*, microbioma, produção, transporte

Index

Index of Figures

List of abreviations

2R-WGD – Two rounds of whole genome duplication

ADOPS - Automatic Detection Of Positively Selected Sites

ALO - D-arabino-1,4-lactone oxidase

CDS - Coding nucleotide sequences

DHA - Dehydro-L-ascorbic acid

DTPA - Diethylenetriaminepentaacetic acid

FAD - Flavin adenine dinucleotide

GlcUAR - D-glucuronate reductase

GLDH - L-Galactono-lactone dehydrogenase

GLUT - Sodium-independent facilitative glucose transporters

GTR - General time-reversible model

GULO - L-gulonolactone oxidase

HEPES - Hydroxyethyl piperazineethanesulfonic acid

HPLC - High-Performance Liquid Chromatography

L-galDH - L-galactose dehydrogenase

MDHA - Monodehydroascorbic acid

mRNA - Messenger ribonucleic acid

NAT - Nucleobase-ascorbate transporter

NCBI - National Center for Biotechnology Information database

ORF - Open reading frames

ROS – Reactive Oxigen species

SEDA - Sequence Dataset builder

SVCT- Sodium-dependent Vitamin C Transporter

UDP - Uridine diphosphate glucose

UTR - Untranslated regions

VTC4 - L-galactose-1-phosphate phosphatase

WGD - Whole genome duplication

I. Introduction

L-ascorbic acid, also known as vitamin C ($C_6H_8O_6$), is a water-soluble vitamin that can be found in solution in a reduced ionizable form (L-ascorbic acid), in a one electron oxidized form (monodehydroascorbic acid (MDHA)) or an oxidized nonionic form (dehydro-L-ascorbic acid (DHA)) (Muñoz *et al.* 2015, Smirnoff 2018). It is known that this micronutrient is necessary for normal cell function, growth and development (Subramanian *et al.* 2017), acting as an important cellular antioxidant capable of detoxifying exogenous radical species present in the cell or those that have arisen due to excess superoxide generation by mitochondrial metabolism (May and Harrison 2013; Zhang *et al.* 2014). The fully oxidized version of ascorbic acid originated by the reduction of radical species, namely DHA, is usually reduced back to ascorbic acid by glutathione (GSH) directly (Winkler *et al.* 1994) or by glutaredoxin (Wells *et al.* 1990), whereas the ascorbic acid radical form, MDHA, is reduced back to ascorbic acid by the action of for example the NADH-cytochrome b5 reductase (Borgese *et al.* 1987) or thioredoxin reductase (Du *et al.* 2012). This antioxidant ability is known to be correlated with protection against degenerative diseases and cancer (Figueroa-Méndez and Rivas-Arancibia 2015). Nevertheless, this vitamin may also display pro-oxidative behavior in higher doses inside the cell, when in the presence of catalytic metal ions (Buettner and Jurkiewicz 1996, Podmore *et al.* 1998, Halliwell 1999, Bahadorani *et al.* 2008, Frei and Lawson 2008). In these conditions, for example, ascorbic acid reduces ferric (Fe^{3+}) to ferrous (Fe^{2+}) iron, becoming an oxidized radical, while the resultant Fe^{2+} easily reacts with O_2, originating a superoxide radical. This radical can later dismute into H_2O_2 and O_2, and between the interaction of H_2O_2 and the Fe^{2+} ion that can be recycled by the presence of ascorbic acid, several radical oxygen species can be produced via Fenton reaction (Du *et al.* 2012). Among various other key roles, ascorbic acid is also essential in collagen biosynthesis, serving as a cofactor for collagen stabilization enzymes, namely prolysyl and lysyl hydroxylase, but also by stimulating lipid peroxidation (Podmore *et al.* 1998, Traikovich 1999, Szarka and Lőrincz 2014). Furthermore, the presence of this vitamin is required for proper brain development, as several reports show that the inadequate levels of ascorbic acid lead to ineffective neuromodulation, which in turn results in impared cognitive function or even death (Gale *et al.* 1996, Tveden-Nyborg and Lykkesfeldt 2009, Tveden-Nyborg *et al.* 2009, Hansen *et al.* 2014). Additionally, He *et al.* (2015) has shown that ascorbic acid can be used as cofactor in enzymes involved in DNA or histones demethylation such as *TET1* and *JMJD3*, playing a role as modulator in epigenetic modifications.

Regarding the synthesis of ascorbic acid, three main biosynthetic pathways are currently described in the literature: the mammals pathway, the plant pathway and the photosynthetic protists pathway (Wheeler *et al.* 2015, Smirnoff 2018). The mammals pathway is characterized

by the use of glucose as an initial precursor that is ultimately converted to L-gulonolactone by the action of several enzymes such as D-glucuronate reductase (GlcUAR) and SMP30/Regucalcin. This molecule, through an oxidation process catalized by L-gulonolactone oxidase (GULO), leads to the synthesis of 2-keto-L-gulonolactone, which is spontaneously converted to ascorbic acid (Linster and Van Schaftingen 2007, Wheeler *et al.* 2015, Aumailley *et al.* 2016). The plant pathway is somewhat different as it uses fructose as the initial precursor. Fructose is gradually converted through the action of several enzymes, such as L-galactose-1-phosphate phosphatase (VTC4) and L-galactose dehydrogenase (L-galDH), to L-galactono-lactone, which is oxidized to ascorbic acid by the enzyme L-Galactono-lactone dehydrogenase (GLDH) (Wheeler *et al.* 2015). As for the photosynthetic protists pathway, glucose is again used as a precursor molecule as seen in the mammals pathway. Nevertheless, in this case, the glucose is converted to L-galactono-lactone, and this molecule is oxidized by the action of GLDH to ascorbic in the final step of the metabolic pathway, as observed in plants (Wheeler *et al.* 2015, Jiang *et al.* 2018, Smirnoff 2018). In Fungi, several species are able to synthesize ascorbic acid analogues due to the action of D-arabino-1,4-lactone oxidase (ALO), such as D-erythroascorbate (Loewus 1999), using the conversion of D-arabinose as the initial substrate through a pathway considered similar to the one found in plants (Wheeler *et al.* 2015). The known ascorbic acid and D-erythroascorbate synthesis pathways can be observed in Figure 1.

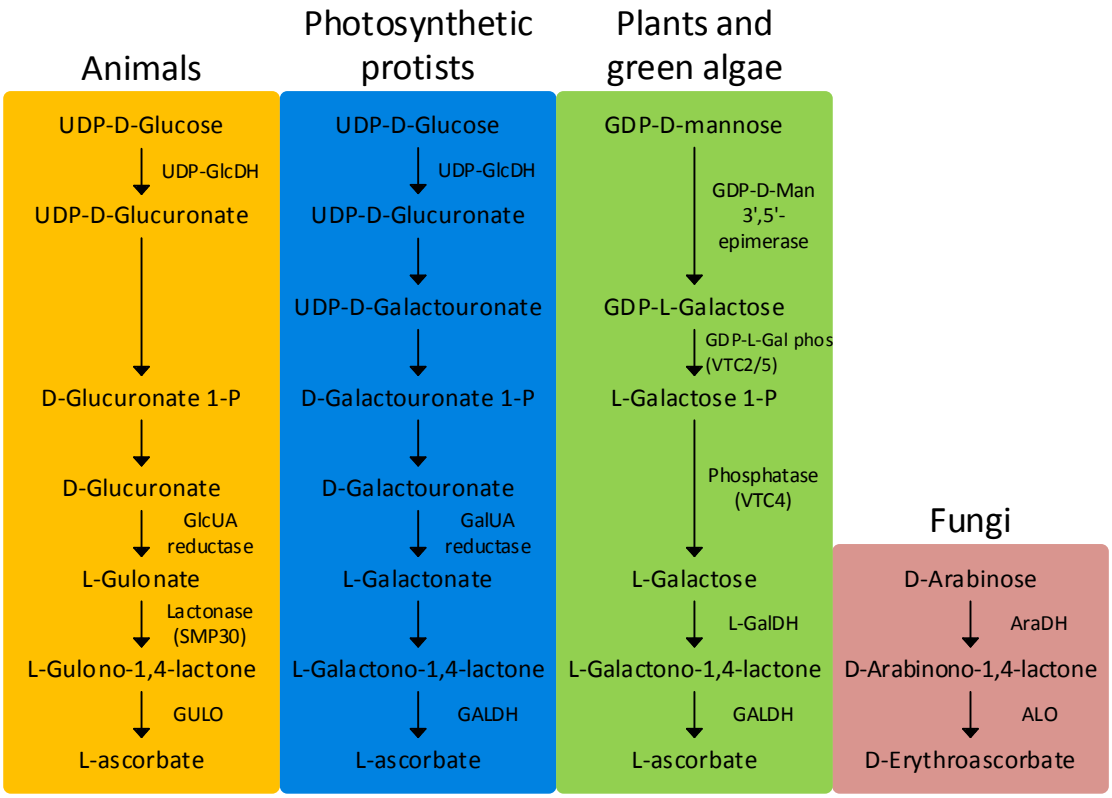


Figure 1 – Graphic display of the currently known ascorbic acid synthetic pathways. The final oxidation step of the distinct aldono-1,4-lactones to ascorbate is performed by an FAD-linked oxidase or dehydrogenase (GULO, GALDH or ALO). Photosynthetic protists

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 (Wheeler *et al.* 2015). The figure here presented and the corresponding description were adapted from Smirnoff (2018).

In summary, the three major pathways of ascorbic acid biosynthesis use different routes and initial substrates to synthetize an aldonolactone precursor (L-gulono-lactone or L-galactono-lactone), which is converted to ascorbic acid by either GULO (animal pathway) or GLDH (plant and photosynthetic protists pathways) (Shigeoka *et al.* 1979, Wheeler *et al.* 1998, Loewus 1999, Wheeler *et al.* 2015, Smirnoff 2018). It is interesting to note that both these aldonolactone oxidoreductases possess a well-characterized conserved HWXK motif, known to be involved in FAD-binding at the catalytic domain (Fraaije *et al.* 1999, Logan *et al.* 2007, Aboobucker and Lorence 2016). It is not known how or why the distinct pathways of ascorbic acid biosynthesis arose in animals, plants and algae. In this context, two evolutionary scenarios should be considered (Wheeler *et al.* 2015). In the first scenario, a gene duplication event could have occurred in the last common ancestral of these taxonomic groups, followed by a differential loss of either gene in the different lineages. In the second, a lateral gene transfer event of a novel gene was followed by functional replacement of the ancestral gene (Keeling and Inagaki 2004). In the currently proposed model, ancestral eukaryotes synthesized ascorbic acid via GULO, and GLDH appeared later in the Archaeplastida (land plants, green algae, red algae, and glaucophytes) lineage following endosymbiosis of a cyanobacterium, after the divergence of the glaucophytes (Wheeler *et al.* 2015). Nevertheless, it is known that within Fungi, several species are able to synthesize ascorbic acid analogues (Loewus 1999, Wheeler *et al.* 2015). Moreover, this protein also contains the conserved HWXK amino acid motif known to exist in the animal GULO and plant GLDH (Fraaije *et al.* 1999, Logan *et al.* 2007, Aboobucker and Lorence 2016). This evidence adds further questions regarding the ancestral pathway of ascorbic acid synthesis within eukaryote species, and as such, further analysis regarding the molecular evolution of the enzymes participating in this pathway are needed to uncover the probable evolutionary history.

The ability to produce endogenous ascorbic acid is not ubiquitous to all eukaryotic organisms. In humans (*Homo sapiens*), vitamin C deficiency caused by a daily lack of ingestion of this vitamin is often correlated with absence of collagen hydroxylation, leading in extreme cases to scurvy (Lux-Battistelli and Battistelli 2017). Humans are unable to synthesize ascorbic acid, making them auxotrophs regarding this molecule (Davey *et al.* 2002, Montel-Hagen *et al.* 2008). Like humans, non-human primates, the teleost fishes, some birds, *Cavia porcellus* (guinea pig) and various bats have lost the ability to synthesize ascorbic acid, due to the complete or partial loss of the *GULO* gene (Drouin *et al.* 2011). Within the Euteleostomi (bony vertebrates), it is known that some ancestral actinopterygian fish species, like cartilaginous and non-teleost (Holostei) bony fishes, are able to synthesize ascorbic acid, placing the probable *GULO* loss event in teleost fishes around 200 to 210 million years ago (Dabrowski 1994, Moreau and Dabrowski

1998, Moreau and Dabrowski 2005, Cho *et al.* 2007). Given the time scale of *GULO* loss in teleosts, there is no any identifiable remnant gene sequence in these species genomes (Lachapelle and Drouin 2011). However, in Haplorrhini primates and *C. porcellus* (guinea pig), evidence for partial *GULO* gene sequences has been found (Nishikimi *et al.* 1992, Nishikimi *et al.* 1994, Ohta and Nishikimi 1999). These findings are evidence of much more recent independent *GULO* gene loss events, which were calculated to have happened around 61 million years ago in Haplorrhini primates and 14 million years ago in *C. porcellus* (guinea pig) (Lachapelle and Drouin 2011). Nevertheless, in teleosts, Haplorrhini primates and *C. porcellus*, no gene reactivation events occurred since the loss of *GULO*, while in some bats (Cui *et al.* 2011, Drouin *et al.* 2011) and passeriform birds (Drouin *et al.* 2011) that phenomenon seems to have happened many times independently. Despite many exhaustive studies regarding the genetics behind the loss of ascorbic acid production, in several species, the reasons behind the loss of function regarding the *GULO* gene across several taxonomic groups are not yet fully understood (Drouin *et al.* 2011, Fernie and Tohge 2015, Wheeler *et al.* 2015, Smirnoff 2018). Several authors believe that the *GULO* gene is "predisposed" to pseudogenization when faced against other genes belonging to the animal ascorbic acid biosynthetic pathway. These authors argue that GULO is only implicated in the production of ascorbic acid, a compound unnecessary for other metabolic pathways (Linster *et al.* 2007), whereas proteins encoded by other genes of the pathway, such as Regucalcin, affect many metabolic traits when absent (in the example given, caprolactam degradation, and the pentose phosphate pathway, among others) (Moreno *et al.* 2017). Given that the synthesis of ascorbic acid catalyzed by GULO is attached with the production of hydrogen peroxide (H_2O_2), alternatively, some authors suggest that the loss of *GULO* is related with the evasion of H_2O_2-induced oxidative stress by diminishing the concentration of this molecule in the cell (Smirnoff 2018). Other hypothesis relies on the evidence that, in humans, the loss of the enzyme responsible for the oxidation of uric acid, uricase, may be related with the loss of *GULO*. Ames *et al.* (1981) showed that uric acid can act almost with the same efficacy as an antioxidant, when compared to ascorbic acid. Furthermore, they showed that uric acid plasma levels in human cells were notably higher than the ascorbic acid levels, a possible indication of the importance this molecule has in the oxidative stress response. Given these reports, it is possible that with the loss of uricase, the availability of uric acid in the cells rose, and that phenomenon led to the facultative use of ascorbic acid as an antioxidant, which ultimately resulted in the loss of GULO (Ames *et al.* 1981, Smirnoff 2018). Yet another hypothesis relies on the need to detoxify the cellular environment when it is exposed to the prejudicial substances. The UDP-glucuronate that results from the activity of the UDP-D-glucose dehydrogenase enzyme in the third step of the animal ascorbic acid biosynthesis pathway (Wheeler *et al.* 2015), can be used to remove xenobiotics or endobiotics from the cell through a process of glucuronidation (Ritter 2000). Given that 30% of the available UDP-glucuronate is used for ascorbic acid synthesis in rat (*Rattus norvegicus*) liver (Linster and Van

Schaftingen 2007), it is possible that with the loss of ability to synthesize ascorbic acid, the increased UDP-glucuronate levels in the cell allow for more effective detoxification processes (Linster and Van Schaftingen 2007, Smirnoff 2018).

While *GULO* presence or loss in vertebrates have been thoroughly scrutinized (Drouin *et al.* 2011, Yang 2013), it has not been carefully attended in the remaining animal taxonomic groups, especially in the more basal protostomian and non-bilaterian groups. Within the non-bilaterians little is known, with a single report from Wheeler *et al.* (2015) indicating that the *GULO* gene is present in the poriferan *Amphimedon queenslandica*, the placozoan *Trichoplax adhaerens* and the cnidarian *Nematostella vectensis*. As for the protostomians, Wheeler *et al.* (2015) also reports the detection of *GULO* in the genome of the annelid *Capitella teleta*, the acari tick *Ixodes scapularis*, the centipede *Strigamia maritima* and the gastropods *Aplysia californica*, *Haliotis discus hannai* and *Lottia gigantea*. Nevertheless, no details were provided on how the data was gathered and the analysis performed, and apart from the conclusion that the ability to synthesize ascorbic acid via *GULO* is an ancestral trait of the Ophistokonta (animal and fungi) taxonomic group that was probably lost in some lineages, there is no discussion on these findings. Nevertheless, an interesting report with much informative context is available regarding *Caenorhabditis elegans*, an invertebrate nematode, where a novel ascorbic acid synthesis pathway may have been found. Although the final enzyme of either the animal or plant ascorbic acid synthesis pathway is not present in *C. elegans*, Patananan *et al.* (2015) demonstrated the incorporation of ^{13}C into *C. elegans* ascorbic acid pool, using ^{13}C-labeled *Escherichia coli* food. This information is highly curious as it suggests that ascorbic acid could be synthetized by a pathway not reliant on either GULO or GLDH to catalyze the final oxidation reaction. However, in this work no other enzyme was proposed as a potential candidate to do the biological function of both GULO and GLDH. Furthermore, ascorbic acid levels were detected in several marine invertebrates without an identifiable *GULO* gene, but the general consensus regarding these results implies that these species obtain this vitamin through their diet, and not because of *de novo* synthesis (Carr and Neff 1980, Carr *et al.* 1983, Dabrowski and Hinterleitner 1989, López-Fernández *et al.* 2018). It is well described that *GULO* was lost in the insects lineage (Wheeler *et al.* 2015, López-Fernández *et al.* 2018). Nevertheless, ascorbic acid levels can also be measured in several species from this taxonomic group, since high levels of ascorbic acid possibly associated with enhanced tolerance against the ROS-inducing agent tannin, were found in the moth *Orgyia leucostigma* (Barbehenn *et al.* 2001). Moreover, ascorbic acid levels are thought to aid the enzymatic antioxidant systems in the *Callosobrochus maulatus* beetle, since the concentration of this vitamin decreases in a dose-dependent manner in response to the presence of different ROS-inducing insecticides in this species (Kolawole *et al.* 2014). However, the initial ascorbic acid levels detected in these species could again be explained by dietary

supplementation, as seen for marine invertebrates. Still, one interesting report was published concerning *Drosophila melanogaster* Oregon-R, where not only it is shown that *D. melanogaster* adult flies reared on ascorbic acid free food maintain detectable levels of ascorbic acid, but also that these levels increase when flies are exposed to 4ºC for 10 min (cold shock conditions) (Massie *et al.* 1991). Since *D. melanogaster* does not have the GULO protein that allows for the synthesis of ascorbic acid and does not obtain this vitamin from the food source, the most logical explanation for the increased levels detected seems to be endogenous synthesis through an undescribed pathway as seen in *C. elegans*. However, although many authors still consider that prokaryotes do not synthetize or depend on ascorbic acid, several symbiotic bacteria living in metazoan hosts that are able to produce this vitamin have been identified, namely *Mycobacterium tuberculosis* and a particular strain of *Pseudomonas aeruginosa* (Wolucka and Communi 2006, Chang *et al.* 2018). Furthermore, it is known that the human gut commensal bacteria can synthesize and supply vitamins to the host (LeBlanc *et al.* 2013), and that Vitamin B1 can be synthetized by *D. melanogaster's* microbiota in sufficient amounts to support the viability of its offspring (Sannino *et al.* 2018). With this information, we are able to extrapolate that in addition to a possible alternative pathway of ascorbic acid synthesis, it is reasonable to consider the hypothesis that the *D. melanogaster* microbiome may be synthesizing ascorbic acid.

Independently from the means of obtaining ascorbic acid (synthesis, microbiome or diet), the ascorbic acid levels in tissues need to be in homeostasis for the optimal function of the organism (Bürzle *et al.* 2013). In deuterostomians, one important regulatory mechanism shown to be involved in ascorbic acid homeostasis relies on transporter proteins to control the accumulation of this vitamin in several tissues (Savini *et al.* 2008, Du *et al.* 2012, Bürzle *et al.* 2013, Lindblad *et al.* 2013). Two known classes of transporter proteins with this specific function are already identified, namely the Sodium-dependent vitamin C transporters (SVCTs) that are related to the absorption and distribution of ascorbic acid through cells, and the Sodium-independent facilitative glucose transporters (GLUTs), responsible for the absorption of DHA (Diliberto *et al.* 1983, Welch *et al.* 1993, Vera *et al.* 1995, Welch *et al.* 1995, Savini *et al.* 2008, Du *et al.* 2012). However, studies revealed that although GLUTs contribute to DHA absorption, the contribution of these transporters for ascorbic acid concentrations in the cell is rather small, and thus, SVCTs have been shown to be the main regulator of ascorbic acid uptake (Tsukaguchi *et al.* 1999, Corpe *et al.* 2010).

SVCTs are surface glycoproteins that belong in the nucleobase-ascorbate transporter (NAT) protein family (Bürzle *et al.* 2013). The proteins included in this family are assorted into three distinct groups given their corresponding substrate specificity: i) xanthine and uric acid, ii) uracil or iii) ascorbic acid (Bürzle *et al.* 2013). The ascorbic acid group proteins are known to be exclusive to vertebrate species and are designated as SVCT1, SVCT2, SVCT3 and SVCT4 (de

Koning and Diallinas 2000, Yamamoto *et al.* 2010). Curiously, from these four proteins, only SVCT1 and SVCT2, the translated product of the *SLC23A1* and *SLC23A2* genes (Muñoz *et al.* 2015), respectively, are involved in ascorbic acid uptake and share a unique and characteristic conserved amino acid motif (SSSP) (Wang *et al.* 2000, Corpe *et al.* 2005, Wilson 2005, Biondi *et al.* 2007, Godoy *et al.* 2007, Luo *et al.* 2008, Mackenzie *et al.* 2008, Nualart *et al.* 2014, Kourkoulou *et al.* 2018). In humans, the SVCT1 transporter is mainly expressed in the epithelial tissues of several organs, such as the intestine, kidney and liver, while SVCT2 is expressed ubiquitously throughout the body (Rajan *et al.* 1999, Tsukaguchi *et al.* 1999, Wang *et al.* 2000, Clark *et al.* 2002, Lee *et al.* 2006). Furthermore, it is known that the SVCT1 transporter contributes mainly to ascorbic acid uptake and therefore whole-body ascorbic acid level regulation, whereas the SVCT2 transporter is linked with specific responses to oxidative stress in the cells (Bürzle *et al.* 2013). Moreover, Kuo *et al.* (2004) showed that the SVCT1 and SVCT2 transporters seem to function and be expressed independently in mice (*Mus musculus*), since a lower expression of SVCT2 in heterozygous SVCT2 knockout individuals did not affect the expression of SVCT1 in the kidney and liver, condition that allowed for normal ascorbic acid levels in these organs. In addition, Kuo *et al.* (2004) also showed that the ascorbic acid levels in SVCT2-predominat organs, such as the brain or spleen, were lower in these mutant mice, possible evidence that this transporter is essential for the maintenance of ascorbic acid levels in tissues without notable presence of SVCT1. Further support for this hypothesis is given by an independent study performed by Sotiriou *et al.* (2002), in which similar results are obtained.

No function has yet been attributed to the SVCT3 transporter, encoded by the *SLC23A3* gene. Nevertheless, this transporter was shown to be mainly expressed in the kidney in both human and mouse, and in the later organism, likely present in the S3 segment of renal proximal tubules (Bürzle *et al.* 2013). This result led Bürzle *et al.* (2013) to conclude that the SVCT3 transporter might be responsible for the reabsorption of substrates that would otherwise be excreted in the kidney, although no likely substrate was found. In fact, Bürzle *et al.* (2013) showed that SVCT3 does not transport either ascorbic acid or nucleobases. Coincidently, several phylogenetic analyses indicate that SVCT3 may have diverged early in evolution from the SVCT1 and SVCT2 transporters, and it is probable that this event led to the loss of ascorbic acid transport capacity, while allowing for specialization in the regulation of the absorption of other substrates (Bürzle *et al.* 2013, Kourkoulou *et al.* 2018). The fact that this transporter has a similar conserved amino acid motif [SS(FIV)(PAS)] to the one characteristic of the known ascorbic acid transporters SVCT1 and SVCT2 (SSSP), further supports the hypothesis of a common ancestor between them (Kourkoulou *et al.* 2018).

The SVCT4 transporter was initially described in *M. musculus* (mouse) and is encoded by the *SLC23A4* gene, which was found to be a pseudogene (*SLC23A4P*) in *H. sapiens* (humans)

(Yamamoto *et al.* 2010). In *M. musculus* (mouse), this transporter is known to transport various nucleobases, such as xanthine, hypoxanthine, guanine, thymine and uracil, but not ascorbic acid (Yamamoto *et al.* 2010). SVCT4 appears to be mostly expressed in the apical membrane of the mouse small intestine, and is likely to have an important role in uracil uptake from the diet (Yamamoto *et al.* 2010). Although unconfirmed, it is proposed that the SVCT3 may perform the functions of this transporter in species were SVCT4 was lost, for example *H. sapiens* (Yamamoto *et al.* 2010, Bürzle *et al.* 2013). It is possible that evolutionary pressure to suppress the absorption of nucleobases may have contributed to the loss of function of SVCT4 gene in humans (Yamamoto *et al.* 2010). It is also interesting to note that SVCT4 loss in some species appears to have a correlation with inexistent ascorbic acid synthesis (Kourkoulou *et al.* 2018).

In terms of molecular evolution, it is accepted that the *SVCT1* and *SVCT2* genes probably arose from a duplication of a common ancestral gene about 450 million years ago, before the divergence of bony fish (Osteichthyes) and tetrapods (Savini *et al.* 2008, Kourkoulou *et al.* 2018). Many evidences support this hypothesis, such as for example, relatively similar-sized open reading frames (ORFs) between the two genes, the identical exon-intron borders positions with the exception of the 5' and 3' untranslated regions (UTRs) and highly homologous mRNAs across several species (Savini *et al.* 2008). Furthermore, it has been shown that the neighboring genes of *SVCT1* and *SVCT2* are highly conserved in *H. sapiens* (humans) and *M. musculus* (mouse) (Savini *et al.* 2008). Nevertheless, both *SVCT3* and *SVCT4* genes are considered orphan genes (Bürzle *et al.* 2013, Nualart *et al.* 2014) and their evolutionary origin is rather undefined at the moment. It is known that orphan genes can be the result of duplication events with subsequent fast divergence (Tautz and Domazet-Lošo 2011). In fact, gene duplications, along with genome rearrangement events, are very relevant in the origin of new genes and phenotypes, and are thought to have had a crucial role in the diversification of vertebrates (Tautz and Domazet-Lošo 2011, Cañestro *et al.* 2013). Duplicated genes can undergo a process of subfunctionalization, in which the ancestral functions of the ancestral gene are subdivided between the daughter genes, without consequences regarding loss of function (Wolfe 2001, Glasauer and Neuhauss 2014). Alternatively, they can also suffer a process of neofunctionalization, in which one of the duplicated genes acquires mutations that eventually confer a novel function while the other copy retains the ancestral function without any mutational event (Wolfe 2001, Glasauer and Neuhauss 2014). However, the most common scenario is the process of non-functionalization, in which one of the duplicated genes simply accumulates mutations and is eventually lost (Wolfe 2001, Glasauer and Neuhauss 2014). One acknowledged phenomenon that induces higher genome complexity is known as whole genome duplication (WGD). This event is extremely important regarding the adaptation of several species to new environmental conditions and usually results in notable genome diversification. (Kasahara 2013, Moriyama and Koshiba-Takeuchi 2018). Currently, it is

proposed that vertebrates underwent two rounds of whole genome duplication, and several evidences for this hypothesis where gathered. For example, it has been shown that approximately 25% of the *H. sapiens* (human) genome is covered by four sets of paralogous regions, and that by comparing this genome to one belonging to the invertebrate *Branchiostoma floridae* (amphioxus), there is an evident occurrence of quadruple conserved synteny (gene order on chromosomes) (Putnam *et al.* 2008, Kasahara 2013). Moreover, several sets of paralogous genes thought to have emerged by the two rounds of whole genome duplication (2R-WGD) at the stem of the vertebrates lineage, are present in cartilaginous fish but not in invertebrate chordates (Putnam *et al.* 2008). Given these evidences, it is proposed that 2R-WGD occurred in vertebrates after the separation from invertebrate chordates. Furthermore, it is considered that the first round of WGD affected the common ancestor of all vertebrates, while the second concerns the common ancestor of jawed vertebrates, after the separation from jawless vertebrates (as lampreys and hagfish) (Dehal and Boore 2005, Kasahara 2013).

Nevertheless, it is thought that a teleost-specific whole genome duplication has taken place in the common ancestor of all teleosts (Taylor *et al.* 2003, Glasauer and Neuhauss 2014). It has long been known that several genes from the tetrapods (four-legged vertebrate) lineage can be seen duplicated in teleosts (Taylor *et al.* 2001, Glasauer and Neuhauss 2014). However, only when the four tetrapod *Hox* genes clusters (homeotic genes found in all animal groups and thus conserved genes, essential for the development of organisms) were found duplicated within the most basal teleost groups Elopomorpha (Guo *et al.* 2010, Henkel *et al.* 2012) and Osteoglossomorpha (Chambers *et al.* 2009), did this hypothesis gain strength (Glasauer and Neuhauss 2014). Given that these are the most ancient representatives of the bony fish lineage, it is proposed that the WGD occurred at the base of the teleost origin, but before these species radiation (Glasauer and Neuhauss 2014). In fact, this WGD event appears to have taken place around 320-350 million years ago, according to the analysis performed by Vandepoele *et al.* (2004) using the well-established divergence time between bony fish and tetrapods (450 million years ago) as a reference point for a molecular clock approach. Curiously, within teleosts, additional lineage-specific whole genome duplications appear to have occurred in salmonids (Johnson *et al.* 1987, Alexandrou *et al.* 2013, Glasauer and Neuhauss 2014) and some cyprinids (Uyeno and Smith 1972, Ferris and Whitt 1977, David 2003, Wang *et al.* 2012, Zhang *et al.* 2013, Glasauer and Neuhauss 2014).

Within the tetrapod Amphibia taxonomic group, a WGD event that may have affected specifically *Xenopus laevis* (african clawed frog) around 21 to 54 million years ago, but not species of the same genera (as *Xenopus tropicalis*) has also been proposed (Evans *et al.* 2005, Chain and Evans 2006, Pollet and Mazabraud 2006, Sémon and Wolfe 2008). Recently, this hypothesis has been proven to be correct, since Session *et al.* (2016) was able to demonstrate that

the tetraploid *X. laevis* has two partitioned subgenomes which likely belong to two distinct diploid progenitor species.

Although this much knowledge is available for deuterostomian, and most particularly, to the vertebrates lineage, the knowledge of WGD regarding protostomian lineages is still limited (Li *et al.* 2018). Coincidently, although molecular evolution analyses were performed for the distinct *SVCT* genes identified in vertebrates, very few reports mention studies in protostomian species. In fact, a SVCT protein can be found in *D. melanogaster* (fly) (AAF54519.1), although so far it has been not characterized regarding ascorbic acid transport ability. Knowing that *D. melanogaster* (fly) may be producing ascorbic acid, the presence of a transporter protein with the capacity to maintain ascorbic acid homeostasis would certainly be indicative of a putative important role of this vitamin in this species and ultimately, in other protostomians.

In this work, we seek to describe the evolutionary history of the *GULO* gene within all animal lineages with available genomes, focusing our attentions on the scarcely analyzed Protostomia and Non-Bilateria taxonomic groups. Furthermore, we want to evaluate if species thought to be unable to synthesize ascorbic acid due to the loss of *GULO*, such as *D. melanogaster*, can eventually be using either an alternative biosynthetic pathway, an alternative GULO-like protein or the ascorbic acid supplied by the microbiome to maintain homeostatic levels of this vitamin in the organism. Related with ascorbic acid homeostasis, we desire to elucidate the molecular evolution of the SVCT transporters in non-bilaterian and protostomian lineages, and uncover if the function traits of the ancestral protein could include the ability to transport ascorbic acid.

II. Materials and methods

II.1. Animal *GULO* and *SVCT* CDS phylogenies

Coding sequences (CDS) files were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/assembly/) by typing "Animals" under the "Assembly" search option. Given the incomplete overlap in the CDS annotations between the GenBank and RefSeq databases, we downloaded all of the available data in FASTA format from both, seeking to obtain the maximum information possible. Next, using the SEDA (http://sing-group.org/seda/) software "NCBI Rename" option, we added a prefix to each file name with information on the species name, common name, and kingdom to which the species belongs to. This step allowed us to identify contaminations with badly classified species in the downloaded files, as were the case of *Escherichia coli* (bacteria) and *Bovine orthopneumovirus* (virus). These FASTA files were then removed from the dataset. Due to the cheer size of the animal complete CDS FASTA files, namely 6.3 and 26.4 GB for GenBank and Refseq, respectively, we then proceeded to narrow the information for our genes of interest, namely *GULO*, *SVCT1*, *SVCT2*, *SVCT3* and *SVCT4*.

Regarding the *GULO* gene, a *tblastn* search was performed using the SEDA software. The *M. musculus* GULO protein available at NCBI (NP_848862.1) was used as query against the GenBank and RefSeq CDS files, separately. 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. These output files were further processed using SEDA's "NCBI rename" option, to prefix the header of each of the retrieved sequences with the name of the species, common name, and the family name to which the species belongs to. For both GenBank and RefSeq data, we used SEDA´s "Merge" option so the files would be merged into a single file. The GenBank and RefSeq files were then processed for the removal of sequence line breaks using the "Reformat file" option. Using the "Rename header" option, we altered the sequence headers and kept only the species name, common name, family name, and protein accession number. Again using the "Merge" option, the GenBank and RefSeq files were merged into a single file. This file was refined using the "Pattern filtering" option and sequences with ambiguous nucleotides ([NRMSHDVYKWB]), as well as those not showing the typical amino acid HWXK motif, were removed. Next, using the "Remove redundant sequences" option, identical nucleotide sequences were removed and a list of merged sequence headers was produced. This list was exported to check if different species had identical nucleotide sequences. None of those cases were found. After, using the SEDA "Filtering" option, sequences that are non-multiple of three, that do not have a valid start codon (ATG), and that have in frame stop codons were removed. The *M. musculus GULO* CDS was then reallocated to the first sequence position of the file using the "Reallocate reference sequences" option, and then the "Filtering" option was used to remove

sequences with a size difference larger than 10% relative to the *M. musculus* sequence. This 10% size difference limitation was imposed to eliminate badly annotated *GULO* sequences. Using the MEGA7 software (https://www.megasoftware.net/) we then aligned the sequences in the processed FASTA file using the "MUSCLE (Codons)" option. Still in MEGA7 and using this aligned sequence file, we were able to obtain a neighbor-joining phylogeny using the standard parameters. This phylogeny was used to identify possible CDS isoforms in need of removal. The identified isoforms were confirmed by protein sequence comparison, using the "Align two or more sequences" option in a standard protein BLAST available at NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins). The identification of isoforms was performed following these criteria: 98% or more similarity between sequences, sequences with less than 98% similarity but with obvious annotation errors (such as wrong intron locations) or sequences already identified with an "Isoform" tag by the NCBI database. In the case of 100% similarity between isoforms, we chose the isoform to remove randomly. In the remaining cases, isoforms chosen for removal were the least similar to *M. musculus* GULO between the compared lot, in terms of size and/or identity. The resulting FASTA file was used to produce the alignment after adding five Fungi *ALO* (D-arabinono-1,4-lactone oxidase) CDSs available at NCBI database (*Saccharomyces cerevisiae* S288C (NP_013624.1), *Sugiyamaella lignohabitans* (XP_018736459.1), *Fusarium oxysporum f. sp. lycopersici* 4287 (XP_018236955.1), *Pochonia chlamydosporia* 170 (XP_018144218.1) and *Metarhizium majus* ARSEF 297 (XP_014580409.1)). These sequences were used as an outgroup to facilitate the future rooting of the final Bayesian phylogenetic tree.

The *GULO* Bayesian phylogenetic tree was produced by the analysis of the final FASTA format file (Animals *GULO* CDS plus five Fungi *ALO* CDS) using the ADOPS (Automatic Detection Of Positively Selected Sites) pipeline (Reboiro-Jato *et al.* 2012). In this pipeline, nucleotide sequences are first translated and aligned using the amino-acid alignment as a guide. We used the MUSCLE alignment algorithm as implemented in T-Coffee (Notredame *et al.* 2000). Only codons with a support value above two were used for phylogenetic reconstruction. We used MrBayes 3.1.2 (Ronquist *et al.* 2012) 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 1,000,000 generations with four chains each (one cold and three heated chains) were performed. 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 (first 2500 samples were discarded). The undiscarded

trees were used to compute the Bayesian posterior probability values of each clade of the consensus tree.

The Nexus format Bayesian trees produced as output 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 formated file was imported to MEGA7 in order to root the consensus phylogenetic tree using the five Fungi *ALO* CDSs. This protocol is based on that described in López-Fernández *et al.* (2018).

Regarding the *SVCT* genes, the initial *tblastn* search was performed three times with different reference protein sequences, namely *H. sapiens* SVCT1 (NP_689898.2; XP_011542067.1), SVCT2 (CAB58120.1; NP_005107.4) and SVCT3 (NP_001138362.1). The BLAST parameters used were the same as for the *GULO* protocol. The three resulting files were separately processed using the protocol already described for *GULO*, although without a specific amino acidic pattern filtering and the addition of a Fungi outgroup to the processed final FASTA files. In the size difference step, the chosen reallocated reference sequence for each file was respectively the corresponding initial *tblastn H. sapiens* reference sequence. An alignment file was then produced for each individual *SVCT1*, *SVCT2* and *SVCT3* datasets using the Clustal Omega software (https://www.ebi.ac.uk/Tools/msa/clustalo/), posteriorly used to obtain a Neighbor-Joining tree in MEGA7 from which isoforms could be detected and analyzed for future removal. Due to the presence of two sequences from *Manacus vitelinus* with the same header and accession number (XP_008924532.1) but different nucleotide sequences (99% identical), the initial alignments failed for the datasets. For this reason, we performed the removal of one of the problematic sequences since it constitutes redundant information for posterior analysis. The Neighbor-joining trees for *SVCT1*, *SVCT2* and *SVCT3* were rooted using a branch belonging to a basal animal taxonomic group, namely the Placozoa (represented by *T. adhaerens*). After the removal of all isoforms from the datasets, we merged the final files for each *SVCT* gene into a single one, using the SEDA software. After, also using SEDA, we performed the removal of redundant sequences since some species *SVCT* sequences could be represented in more than one of the merged datasets. After verifying the presence of any remnant isoforms in this dataset using the methodology already mentioned, we used the finalized dataset to produce a Bayesian phylogeny.

The Bayesian tree was obtained using MrBayes 3.1.2 as implemented in the ADOPS pipeline. The general time-reversible model (GTR) of sequence evolution was implemented in the analyses, 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 5,000,000 generations with four

chains each (one cold and three heated chains) were performed. 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 100^{th} generation with a defined burn-in of 25% for the complete analysis (first 12500 samples were discarded). The remaining trees were used to compute the Bayesian posterior probability values of each clade of the consensus tree.

Unfortunately, the output phylogeny did not converge in all the model parameters and as such could not be utilized as a valid representation of phylogenetic relationship between the *SVCT* genes. Seeking to overcome this technical limitation, we tried to further refine the dataset by manually observing the produced MUSCLE alignment file and excluded sequences that originated several alignment gaps, improving the amount of information gathered by MrBayes for the inference of phylogenetic relationship and hopefully allowing for the convergence of a new consensus tree. However, this approach led to same result obtained for the original file. Since the sequences present in the dataset represent greatly divergent species and belong to different genes that can have distinct rates of sequence evolution, perhaps our methodology simply did not allow for highly defined results. As such, we decided to subdivide our dataset into smaller files representative of all *SVCT* genes we could detect in our merged file, and create individual ADOPS runs for each one. For this purpose, we used the refined but unfinished Nexus format Bayesian tree obtained before as a draft representation to manually identify clusters of sequences that could belong to different *SVCT* genes. This tree was converted to Newick format using the Format Conversion Website and later imported to MEGA7 for the placement of a root (sequences belonging to cnidarian species from the Anthozoa group). By observing the tree, we were able to group sequences based on the position they had relative to our initial *H. sapiens SVCT* reference sequences. For instance, the sequence of *M. musculus* present in the same branch of the *H. sapiens SVCT1*, was considered an ortholog of *SVCT1*. Using this approach, we were able to detect *SVCT1*, *SVCT2*, *SVCT3*, *SVCT4* and *SVCT* sequences belonging to protostomian species, which we designated *SVCTP*, and created individual files for each group of sequences. Notably, the *SVCT4* CDS group was validated by the absence of sequences from any primate species, in which this gene is known to be lost (Yamamoto *et al.* 2010). In addition, several Echinodermata, hemichordate, urochordate and cephalochordate sequences, as well as non-bilaterian ones, did not group with any of the *SVCT* gene sequences identified. Given the context of our analysis and knowing that the first round of WGD likely occurred after the divergence of vertebrate and invertebrate chordates (Kasahara 2013), these sequences were included in the *SVCTP* file. Using the SEDA software, we added a tag suffix in the header of the sequences based on the representative file in which they were included (like "_SVCT1" for the sequences identified as *SVCT1*). Five sequences were selected to represent the outgroups in further analysis for the newly

created files, namely *D. melanogaster* SVCTP (AAF54519.1), *H. sapiens SVCT1* (XP_011542067.1), *H. sapiens SVCT2* (NP_005107.4), *H. sapiens SVCT3* (NP_001138362.1) and *M. musculus SVCT4* (XP_006506197.1), each properly placed in the corresponding files.

The *SVCT1*, *SVCT2*, *SVCT3*, *SVCT4* and *SVCTP* files were then analyzed using the ADOPS pipeline as described above. Convergence was achieved in all cases. After converting the Nexus format trees to images and placing the appropriate roots, we verified that a group of sequences in the *SVCT4* phylogeny did not actually appear phylogenetically close to any of the five *SVCT* genes identified. As such, a new file with the *SVCT5* tag was created with these sequences, the five outgroup sequences already mentioned, nine Actinopteri *SVCT3* sequences and three Amphibia *SVCT3* sequences. This file was analyzed using the already described parameters for the ADOPS pipeline and the resulting Nexus formatted tree treated as previously described.

II.2. *GULO* CDS annotations

The non-annotated species representative genomes were obtained from NCBI by querying for various non-bilaterian and protostomian taxonomic groups under the "Assembly" option. Only genomes represented in the GenBank database were downloaded in FASTA format, since many of the species of interest did not have a representative genome in the RefSeq database. Using the SEDA software, a tblastn was performed using *M. musculus* GULO sequence available at NCBI (NP_848862.1) as a protein query against the species genomes previously obtained. The BLAST algorithm version used was 2.7.1+ and the tblastn parameters selected included a 0.05 expectation value, but also a "extract only hit regions" option with a window of 5000 flanking nucleotides. The tblastn results obtained for each species genome (FASTA format) were further processed using the "Grow sequences" option included in the SEDA software, with a selected minimum overlap of 2500 nucleotides. This step was important because the species representative genomes were available mainly as contigs. As such, this option allowed several contigs and even scaffolds to be expanded into a larger representative sequence, later simplifying the annotation process. Next, still using the SEDA software, the possible redundant sequences (originated by overlapping contigs for instance) were removed using the "Remove redundant sequences option". The headers of the redundant sequences removed were merged.

Using the NCBI BLAST website (https://blast.ncbi.nlm.nih.gov), a tblastn was performed using the *M. musculus* GULO protein sequence (NP_848862.1) as query against individual processed sequence datasets (FASTA format) originated using SEDA for the various species of interest. The default BLAST parameters were altered regarding the word size (from 3 to 2) and the low complexity regions filter was removed. This allowed a better alignment between sequences representing greatly divergent species. Using the tblastn results, it was possible to

identify nucleotide region coordinates of putative introns and exons along some of the previously processed nucleotide sequences, and consequently, to annotate the putative *GULO* gene coding sequences in several non-annotated species genomes.

All of the performed CDS annotations were included in a FASTA format file along with the *GULO* CDS of *M. musculus* (NP_848862.1), *X. laevis* (OCT81467.1), *Priapulus caudatus* (XP_014666894.1), *Gallus gallus* (XP_015140704.1), *Alligator mississippiensis* (KYO43973.1), *Lepisosteus oculatus* (XP_015207781.1), *Branchiostoma belcheri* (XP_019645195.1), *N. vectensis* (EDO44935.1), *S. cerevisiae* S288C (NP_013624.1) and *M. majus* ARSEF 297 (XP_014580409.1), available at NCBI. This file was processed using the ADOPS pipeline. In this pipeline, nucleotide sequences are first translated and aligned using the amino-acid alignment as a guide. We used the MUSCLE alignment algorithm as implemented in T-Coffee. Only codons with a support value above two are used for phylogenetic reconstruction when using this pipeline.

Bayesian trees were obtained using MrBayes 3.1.2 as implemented in the ADOPS pipeline. The general time-reversible model (GTR) of sequence evolution was implemented in the analyses, 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 5,000,000 generations with four chains each (one cold and three heated chains) were performed. 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 100^{th} generation with a defined burn-in of 25% for the complete analysis (first 12500 samples were discarded). The remaining trees were used to compute the Bayesian posterior probability values of each clade of the consensus tree.

The Nexus format Bayesian trees produced as output by the ADOPS software were converted to Newick format using the Format Conversion Website. Then, the Newick format files were imported to MEGA7. The root of the consensus tree was placed at the split of the Fungi and the remaining species.

II.3. *Drosophila melanogaster* Oregon-R maintenance

The fly strain used in all experiments is the Oregon-R strain that was obtained from Drosophila Stock Centre (http://blogs.cornell.edu/drosophila/).

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.

II.4. Control and cold exposure experimental conditions

Six male and six female virgin Oregon-R flies were collected from the stocks to new vials and reared at 25ºC with 12h day/night cycles. These flies were transferred to new vials every one to two days up to five times and then discarded. Using this transfer strategy, we were able to ensure that newborn flies would inhabit the distinct transfer vials during successive days in a rather large number, optimizing the sampling needed for future experiments. At the day of their birth, newborn flies were separated according to their gender into different vials, and were then kept at 25ºC for seven days with a 12h day/night cycle.

As controls, male and female flies were collected separately and snap frozen in liquid nitrogen in sets of 25 individuals after this seven-day period.

Using seven days male flies, different cold exposure experimental conditions were tested. In the case of cold acclimation, flies were transferred to 15ºC for one day with a 12h day/night cycle and then back to 25ºC for one day for recovery. Biological samples with 25 individuals were snap frozen with liquid nitrogen at two different time points, namely immediately after cold exposure and after one day of recovery.

Regarding cold shock, flies were transferred from 25ºC to ice containers and kept at 4ºC for four hours. After the exposure to 4ºC, flies were transferred back to 25ºC and collected after a recovery time of 30 minutes, two hours and two days, being then snap frozen with liquid nitrogen in sets of 25 individuals. The frozen samples for all the experimental conditions were after kept at -80ºC until further analysis.

We did not select flies younger than seven days to avoid possible problems caused by differences regarding gene expression and gut microbiome population between same age flies, which may be significant at an early stage (Carlson *et al.* 2015, Odamaki *et al.* 2016). These differences could lead to intrinsic ascorbic acid concentration differences between biological samples and easily mislead our interpretation of the results.

II.5. Generation of axenic *D. melanogaster*

To obtain axenic flies, Oregon-R flies moved from a stock vial into a fly trap were allowed to lay eggs for two hours in a plate containing yeast extract media. Next, the media was resuspended in water to allow the eggs retrieval resorting to a sieve, and the consequent transferal to a 1.5 ml tube. The eggs were then centrifuged twice at 350G for 2.5 minutes in a sodium hypochlorite solution (2.7% v/v), seeking to remove the chorion from the samples and, as such, any microbial population present (as described in Aboobucker and Lorence 2016). From this step forward, the protocol was performed under sterile conditions. The obtained dechorinated eggs were after washed three times using a sterile saline triton solution (300 $\mu l.l^{-1}$ Triton X-100, 4 $g.l^{-}$

[1] NaCl,) and posteriorly placed in UV-sterilized food vials. To verify if the flies microbiome was effectively removed, three axenic flies and three control flies were separately collected, washed with sodium hypochlorite solution (2.7% v/v) and homogenized in 200 µl of a sterile 0.9% (w/v) NaCl solution. One hundred microliters of these homogenates were cultured in plates containing Luria-Bertoli (LB) solid medium at 25 °C for two days. The homogenates belonging to the axenic flies did not display bacterial growth.

II.6. High Performance Liquid Chromatography (HPLC) analysis

An extraction buffer containing 1 ml of SIGMA® HEPES sodium salt (Ref.H3784) at 250 mM concentration with pH 7.2 and 1 ml ALDRICH® Diethylenetriaminepentaacetic acid (DTPA; Ref.D1133) at 2.5 mM in 23 ml tri-distilled water was prepared. Samples of 25 frozen flies were homogenized by maceration in 500 µl of extraction buffer and then transferred to a 1.5 ml tube. To avoid loss of sample, the material used for maceration was washed with another 500 µl of extraction buffer that was then added to the 1.5 ml tube of the homogenized sample. Every step was performed in ice. The homogenate was then subjected to a centrifugation of five minutes at 9000G inside a 4ºC chamber, and 850 µl of the supernatant were transferred to another 1.5 ml tube present in ice. Next and again at 4ºC, a second centrifugation of 20 minutes at 16000G was performed for that tube and three aliquots (250 µl) of the supernatant were transferred to three separate 1.5 ml tubes placed in ice (technical replicas). Finally, at 4ºC, a third centrifugation of 15 minutes at 16000G was applied to the three technical replicas. All the centrifugations were performed using a Eppendorf® 5415D centrifuge and they allowed the deprivation of any suspended particles in solution that could influence the HPLC ascorbic acid quantification of the samples. In parallel, three 1 ml ascorbic acid standard controls with different concentrations were prepared for each HPLC run for both male (25 µM, 10 µM and 5 µM) and female (100 µM, 25 µM and 10 µM) flies samples, using a stock of SIGMA® L-Ascorbic acid BioXtra (Ref.A5960) and extraction buffer as a solvent. The ascorbic acid controls were then preserved in ice and deprived of direct light to avoid degradation. Low temperature centrifugations were used to avoid possible ascorbic acid degradation along the protocol steps, as well as the conservation of intermediary samples in ice.

Two hundred microliters of each processed technical replicas and ascorbic acid controls were loaded into a HPLC 96-well plate, making a total of six samples per run for each condition tested. A single biological sample (divided by three technical replicas) was used in each HPLC run to prevent any kind of ascorbic acid degradation that could arise due to the increased waiting time between sample measurements in a fully loaded plate, in which case a single HPLC run could take up to days. The buffers used in the HPLC analysis were buffer A, consisting of 20 mM SIGMA-ALDRICH® Triethylammonium acetate buffer (Ref.69372) in tri-distilled water at pH 6, and buffer B, 20 mM SIGMA-ALDRICH® Triethylammonium acetate buffer (Ref.69372) in

40% Acetonitrile, LiChrosolv® Reag. Ph. Eur. (Ref.1.00030.1000) at pH 6. The HPLC run was performed at 25ºC with an injection volume of 90 µl and a flow rate of 1 ml/min. The six samples were separated using a LiChrospher® 100 RP-18 (5 µm) LiChroCART® 250-4 reversed-phase column using the following buffer gradient: constant flow of 100% buffer A for 7 minutes, a 1 min linear gradient from 100% buffer A to 0% buffer A, 5 min at 0% buffer A, a 1 min linear gradient from 0% buffer A to 100% buffer A, and constant flow for 10 min at 100% buffer A. Ascorbic acid was detected at a wavelength of 265 nm and eluted between 3 and 3.6 minutes. This HPLC methodology was adapted from Patananan *et al.* (2015).

We also used a single biological sample of 25 female control flies (seven days) to determine the elution time and chromatogram peak region of ascorbic acid during a HPLC run. The experimental protocol was very similar to the one used for test samples, with an additional step before the third centrifugation. In this step, we added 10 units (10 µl) of SIGMA® Ascorbate Oxidase from *Cucurbita* sp. (Ref.A0157) to one of three 250 µl technical replicas, and incubated all three replicas at 25ºC for 25 minutes. After that step, we resumed the standard protocol already described. Using this approach, we sought to use the two unaltered technical replicas as controls for the detection of ascorbic acid, while using the oxidase treated replica as an ascorbic acid deprived control, due to the oxidase activity. This essay allowed us to validate our ascorbic acid peak location in the test samples for the various experimental conditions, avoiding possible false positives in the measurements.

The ascorbic acid levels in the samples were determined by using an ascorbic acid calibration curve and normalized by the number of flies in the individual samples (25 in every case). Differences between treated samples and the appropriate control samples was tested using an t-test statistic after checking if data points are normally distributed.

II.7. Determination of L-ascorbate in microbiome cultures expanded ex-vivo

The possible microbiome contribution for the production of ascorbic acid was assayed using microbiome cultures expanded ex-vivo in De Man, Rogosa and Sharpe (MRS) media. For this purpose, 25 seven day flies were initially collected and washed with sodium hypochlorite solution (2.7 % v/v), followed by three washes with sterile water. After, the flies were homogenized in a 0.9 % (w/v) NaCl solution for a total volume suspension of 1 ml, with this suspension being posteriorly used to inoculate distinct flasks with 100 ml of MRS medium supplemented, or not, with 2% w/v glucose. These flasks were then incubated at 25 and 30 °C, and the microbial growth was monitored using a spectrometric approach (600 nm wavelength). Samples were collected one, two and three days after the inoculation, from both the supernatant and the pellet of bacterial cells. The measurement of ascorbic acid levels in these samples was

performed using a HPLC approach. For the supernatant, 900 µl of the clarified medium were combined with 100 µl of 10x HEPES/DTPA extraction buffer prior to injection in the HPLC. As for the pellet assay, 1 ml of 1 % SDS (Sodium dodecyl sulfate), 0.2 M NaCl solution was used to resuspend the cells, and the obtained homogenate was vortexed for 1 min. This homogenate was then centrifuged at 18000G for 5 min at 4 °C, and 800 µl of the acquired supernatant were mixed with 200 µl of 10x HEPES/DTPA extraction buffer before the HPLC injection.

III. Results and discussion

III.1. *GULO* gene CDS Bayesian phylogeny

The phylogenetic relationship of the 118 animal *GULO* sequences that we have identified after the *GULO* CDS dataset filtering are represented in Supplementary figure 1.

Regarding the presence or loss of the *GULO* gene, the identification of a GULO protein with all expected features (similarity with reference sequence, presence of the typical amino acid pattern, expected size, and expected position in the phylogenetic tree), allows us to extrapolate that the *GULO* gene has not been lost in the lineage of the species in which it was detected. However, we cannot assume that the lack of a detectable GULO protein in a single species means there was loss of the *GULO* gene in that species lineage. This is due to technical issues such as incompleteness of the genome sequence for that particular species, or even failure in the gene annotation. As such, a minimum of three species of a given lineage in which *GULO* CDS with all expected features is not present is needed to sustain the hypothesis that the *GULO* gene was lost. Using this approach, we can achieve higher confidence in our conclusions because it is unlikely that technical issues affect three different species genomes in the same way. Moreover, *GULO* CDSs removed from the final phylogenetic analysis due to size difference or in-frame stop codons can be the result of miss-annotation of a functional *GULO* gene or a *GULO* pseudogene, respectively. Using these criteria, we were able to summarize all of the findings regarding *GULO* presence/absence in animal lineages on a cladogram, seen in Figure 2.

The *A. queenslandica* sequence (XP_003389075.2) that represents the Porifera taxonomic group in the consensus phylogeny appears misplaced in the taxonomic context, being an outgroup to the five Fungi *ALO* CDSs (Supplementary figure 1). In addition, the amino acidic motif present in the protein encoded by this sequence is HWGK, rather than the HWAK motif seen in the majority of the remaining sequences. As such, this species does not likely have a GULO, but another FAD domain protein. Nevertheless, since a single species was present in the analysis, we cannot assume that *GULO* was lost in the Porifera group.

The *GULO* gene seems lost in the non-bilaterian species from the Placozoa group. Still, since a single species (*T. adhaerens*) was analyzed concerning this taxonomic group, we cannot formulate any hypothesis on the subject (Figure 2). Within the non-bilaterian Cnidaria group, we identified a putative *GULO* gene sequence in three out of five species of the Anthozoa class. Although no likely functional *GULO* gene sequence was found for both Myxozoa and Hydrozoa classes, a single species of each (*Thelohanellus kitauei* and *Hydra vulgaris*, respectively) was analyzed, and thus no conclusions were taken (Figure 2). This result goes in accordance with putative *GULO* gene identification by Wheeler *et al.* (2015) in the Cnidaria group, although the results for the Placozoa and Porifera do not.

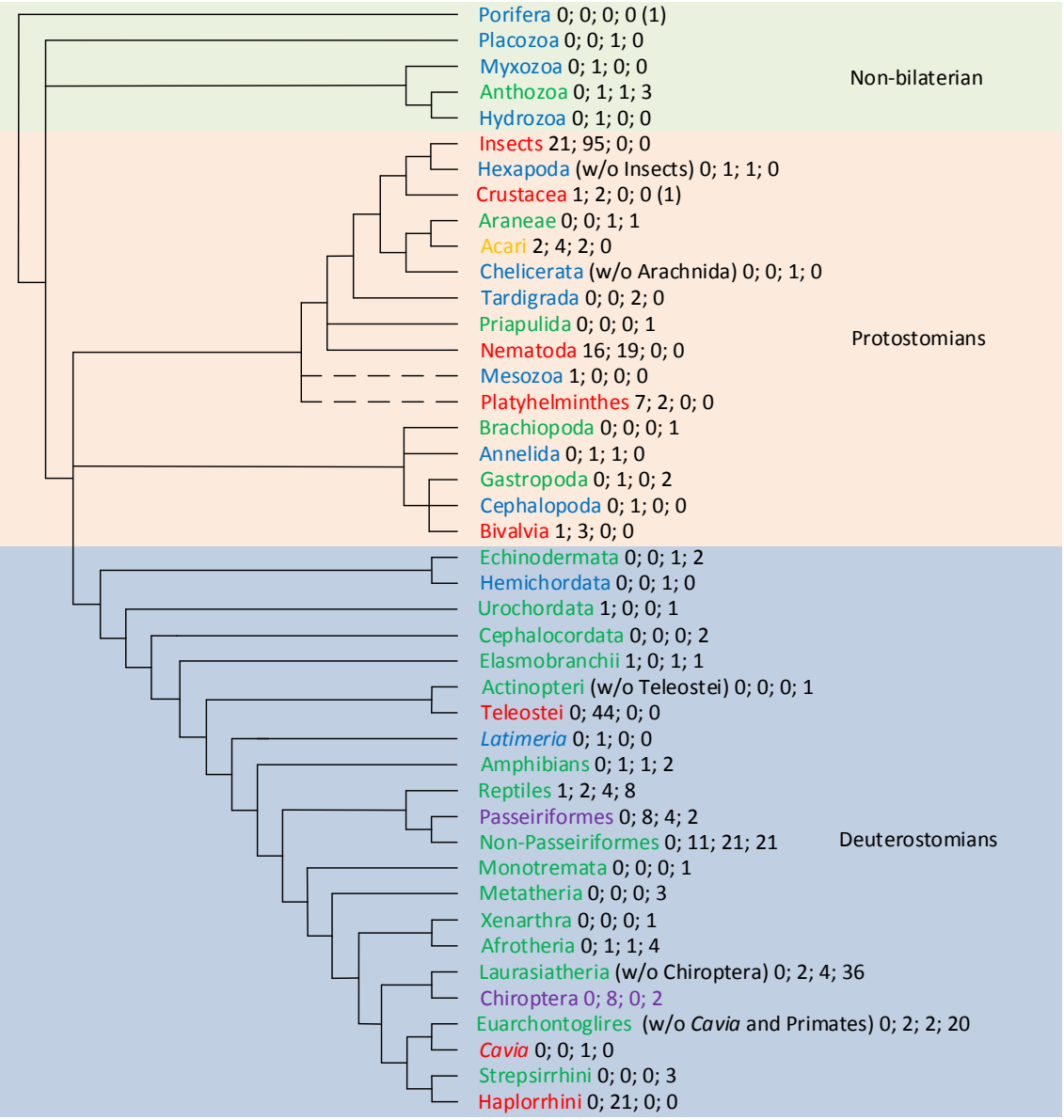


Figure 2 - Summary of the findings regarding the presence of putative functional *GULO* genes in non-bilaterian, protostomian and deuterostomian lineages. The lineages where a possibly functional *GULO* has been detected are represented in green, while lineages where the *GULO* gene has not been detected are presented in red. Additionally, lineages for which there is insufficient information to extrapolate a conclusion are represented in blue, with lineages showing a notable number of species with a functional and non-functional *GULO* gene represented in violet. The specific Acari lineage case, in which some species were excluded from the final phylogeny but may potentially have a functional *GULO* gene, is represented in orange. The first three numbers next to the represented lineages indicate species excluded from the dataset because: i) no sequence with significant homology was found in the initial BLAST; ii) the sequences did not possess the typical GULO amino acid pattern or showed ambiguous nucleotide positions; iii) the sequences do not present an ATG start codon, are non-multiple of three, have in frame stop codons, or have a size difference larger than 10% relative to the reference *M. musculus GULO* sequence. The last number indicates the number of species from each lineage present in the final tree. Numbers in parentheses indicate species that were found taxonomically misplaced in the final phylogeny, and that do not likely have a *GULO* gene. Broken lines show uncertain relationships. Taxonomic relationships are depicted as in the Tree of life web project (http://www.tolweb.org/tree/) and in Helgen (2011). This cladogram is depicted an in López-Fernández *et al.* (2018).

The bilaterian group is comprised of the Protostomia and Deuterostomia taxonomic groups. The deuterostomes can be divided into two large taxonomic branches: one leading to the Echinodermata and Hemichordata sister groups, and the other leading to the Chordata group, in

which the vertebrates are included. In the Echinodermata group, we were able to identify putative *GULO* genes for two species, namely *Acanthaster planci* and *Strongylocentrotus purpuratus*, with the later species result going in accordance with the findings by Wheeler *et al.* (2015). As for the Hemichordata, we could not find any probable *GULO* gene, although only one species was analyzed (*Saccoglossus kowalevskii*). Nevertheless, this was also the expected result (Wheeler *et al.* 2015).

Inside the Chordata group, we found putative *GULO* genes in one Urochordata species (*Ciona intestinalis*), already referenced as a *GULO* containing species by Wheeler *et al.* (2015). Still regarding the Chordata group, we detected a potentially functional *GULO* in two species belonging to the Cephalochordata subphylum, namely *B. belcheri* and *B. floridae*. All of the remaning deuterostomian species analyzed belong to the Gnathostomata taxonomic group, included in the Craniata subphylum.

We observed the lack of a putative *GULO* gene in the teleost (bony) fishes group, in which 44 species were analyzed. Nevertheless, this was the expected result since reports state that no teleost fish species is able to synthesize ascorbic acid due to the loss of the *GULO* gene (Drouin *et al.* 2011). Interestingly, we detected a putative *GULO* gene in a species (*L. oculatus*) belonging to a Teleostei sister group inside the Neopterygii subclass, Holostei, represented as "Actinopteri (w/o Teleostei)" in Figure 2. Furthermore, we could not identify a *GULO* gene in *Latimeria chalumnae* (coelacanth), a fish species more closely related to tetrapods (four-limbed vertebrates) than to teleost fish. Moreover, we detected a putative functional *GULO* gene sequence in one Elasmobranch species (*Rhincodon typus*).

We also observed the presence of a potentially functional *GULO* gene in two amphibian species belonging to the Anura order, namely *Nanorana parkeri* and *X. laevis*.

Regarding the birds taxonomic group, we identified a putative *GULO* in 23 species representing 15 different orders. We also detected that 25 species were excluded from the analysis due to either a size difference larger than 10% relative to the reference CDS, absence of a start codon, sequences non-multiple of three or in frame stop codons, suggesting the presence of a *GULO* pseudogene. Furthermore, we found that in the Galliformes, Gruiformes, Pelecaniformes, and Passeriformes orders there were at least three species removed in this step. As such, for each of these groups, we performed a tblastn search using the GULO mouse sequence as query (NP_848862.1) against the genomes of the species excluded, and found that the *GULO* gene is not well annotated in the Galliformes and Pelecaniformes groups. In these cases, there were no in-frame stop codons and all exons could be found, so it is likely that these genomes could harbor a functional *GULO* gene. In the case of Gruiformes, the *GULO* sequences were only missing the first methionine. However, we observed that the second codon was preceded by the nucleotides AG, which could indicate a canonical GT-AG splice site at this location, and therefore, the presence of an intron after the exon containing the first methionine. Given this hypothesis, in the

absence of mRNA data, it is impossible to accurately find the first exon of this putative *GULO* gene. It is known that, within Passeriformes, several species do not present a functional *GULO* gene. Also, there is evidence of provable *GULO* loss and reacquisition along the evolutionary history. We found that four passerine species (*Acanthisitta chloris*, *Lonchura striata*, *Ficedula albicollis* and *M. vitellinus*) do not have a well annotated *GULO* gene since, like in the Galliformes and Pelecaniformes groups, there were no in-frame stop codons detected and all exons could be found. Given the passerine species complex evolutionary history regarding the *GULO* gene, as they belong to a lineage were containing functional and non-functional *GULO* is rather random, we decided to represent this order in a featured position in Figure 2. All of the other bird species analyzed are represented with the "Non-Passeiriformes" tag.

Now regarding the Reptiles taxonomic group, we detected a putative functional *GULO* gene in four species of the Crocodylia order (*A. mississippiensis*, *Alligator sinensis*, *Crocodylus porosus* and *Gavialis gangeticus*), in one species of the Testudines order (*Chrysemys picta bellii*) and in three species included in the Squamata order (*Gekko japonicus*, *Ophiophagus hannah* and *Python bivittatus*). We noticed that four species were removed from the phylogenetic analysis in the size difference step of the protocol. Utilizing the tblastn approach mentioned in the birds case, we found that the *GULO* gene is not well annotated in three squamata species (*Protobothrops mucrosquamatus*, *Pogona vitticeps* and *Anolis carolinensis*) and one testudine species (*Chelonia mydas*).

A putative *GULO* gene was detected in the one species (*Ornithorhynchus anatinus*) of the Monotremata taxonomic group that was analyzed. Moreover, a potentially functional GULO gene was identified in three species of the Metatheria group (*Monodelphis domestica*, *Phascolarctos cinereus* and *Sarcophilus harrisii*), one species of the Xenarthra group (*Dasypus novemcinctus*) and four species of the Afrotheria group (*Echinops telfairi*, *Loxodonta africana*, *Trichechus manatus* and *Orycteropus afer afer*).

The Chiroptera (bats) taxonomic group belongs to the Laurasiatheria superorder. For several years, it was considered that bats were unable to synthesize ascorbic acid, due to the loss of *GULO*. However, later reports have refuted that hypothesis, since ascorbic acid synthesis was observed in *Rousettus leschenaultia* and *Hipposideros armiger* (Cui *et al.* 2011). Additionally, a functional *GULO* gene was detected in *Rhinolophus ferrumequinum* and *Rousettus aegyptiacus*, amongst other bat species (Cui *et al.* 2011). We identified a putative *GULO* gene in only two (*R. aegyptiacus* and *Rhinolophus sinicus*) out of ten bat species analyzed, a result that goes in accordance with Cui *et al.* (2011) results. Given the particular phylogenetic characteristics of the Chiroptera group regarding the *GULO* gene, we gave a featured position to this taxonomic group in Figure 2. All of the other Laurasiatheria species were included in the "Laurasiatheria (w/o Chiroptera)" category. Thirty-six out of 42 of these remaining Laurasiatheria species had a putative functional *GULO* gene present.

The Primates order and the species *C. porcellus* (guinea pig) are contained in the Euarchontoglires superorder. The Primates order is comprised by the Haplorhini suborder (which includes the anthropoid primates) and the Strepsirrhini (prosimians) suborder. Several reports show that anthropoid primates and guinea pigs have lost the capacity to synthesize ascorbic acid (Drouin *et al.* 2011). Nevertheless, their genomes still contain trace sequences similar to *GULO* gene sequences (Drouin *et al.* 2011). In agreement with these reports, although all 21 Haplorhini genomes showed at least one hit when performing the tblastn step of the protocol, none had the typical HWAK GULO amino acid motif. We were also able to identify a potentially functional *GULO* gene in three species *(Microcebus murinus*, *Propithecus coquereli* and *Otolemur garnettii*) belonging to the Haplorhini sister group, Strepsirrhini. Furthermore, although a *GULO* gene is annotated in the guinea pig genome, the CDS is not present in the phylogenetic tree shown in Supplementary Figure 1. This putative functional gene encodes a protein that is half the size of the ones seen in other species, and thus it is considered a pseudogene. This is an example relative to the advantages of using a sequence size removal step, showing how useful this option can be to avoid misleading information available in many databases, such as bad annotations. The Haplorhini and Strepsirrhini suborders, such as the guinea pig, were discriminated in Figure 2. All of the remaining Euarchontoglires species analyzed were included in the "Euarchontoglires (w/o *Cavia* and Primates)" category. Regarding this category, we were able to identify a putative *GULO* gene in 20 out of the 24 species available in the protocol.

Analyzing the Prostotomian species, we verified that there was a *GULO* gene loss in the Pancrustacea taxonomic group. We could not identify any putative *GULO* in 116 insect species. The insects are included in the Hexapoda subphylum, and although we could not find a putative *GULO* gene in any insect species, the number of remaining Hexapoda species analyzed was not sufficient to conclude that the gene was lost in this complete lineage. We also could not identify any potentially functional *GULO* gene in the analyzed species of the Crustacea subphylum. Although one crustacean species (*Hyalella azteca*) *GULO* CDS is present in the final phylogeny, the sequence appears misplaced in the tree. In addition, the amino acidic motif of the protein encoded by this sequence is HWGK, rather than the typically seen HWAK pattern. With this information, we did not consider this sequence to be representative of the *GULO* gene.

The Arachnida taxonomic group contains the Araneae order and the Acari subclass, being integrated the Chelicerata subphylum. We failed to detect a potentially functional *GULO* gene in eight species of the Acari subclass. Still, using a tblastn against the two genomes of the species removed by size (*Euroglyphus maynei* and *I. scapularis*), we found that both of them could contain a functional *GULO* sequence (HWAK pattern and great alignment coverage), a likely occurrence for at least *I. scapularis* given the report by Wheeler *et al.* (2015). Their exclusion from the analysis could therefore be the result of wrong gene annotation, as seen various times in deuterostomes. Regarding the Araneae order, a putative functional GULO sequence was

identified in *Parasteatoda tepidariorum*, while none was found in *Stegodyphus mimosarum*. We also could not find a putative *GULO* gene in the remaining Chelicerata species, namely *Limulus polyphemus*. Similarly to the Hexapoda case, we cannot assume that the *GULO* gene is lost in the whole Chelicerata (w/o Arachnida) lineage, since only one species was analyzed.

Two Tardigrades species (*Hypsibius dujardini* and *Ramazzottius varieornatus*) were analyzed and no putative *GULO* gene was found, even after a tblastn search against the species genomes. Given the number of species analyzed, we cannot assume that *GULO* gene is lost in this lineage.

We detected a potentially functional *GULO* gene in *Priapulus caudatus*, a member of the Priapulida phylum. In contrast, when observing the 35 analyzed species belonging to the Nematoda phylum, we failed to detect a putative *GULO* gene. We obtained equal results when observing the nine analyzed species included in the Platyhelminthes phylum, strongly indicating the loss of *GULO* in this lineage. In the Mesozoa subphylum, only *Intoshia linei* was analyzed, being removed at the initial tblasnt step. Since a single species was observed, we could not draw any conclusions regarding this lineage. *GULO* appears present in the Brachiopoda phylum, as we detected a putative gene in *Lingula anatina*. We could not formulate any hypothesis regarding the Annelida phylum, since we could not identify a possibly functional *GULO* in *Helobdella robusta* and *C. teleta*.

The Bivalvia, Cephalopoda and Gastropoda classes are included in the Mollusca phylum. Regarding the four analyzed species representative of the Bivalvia class (*Mytilus galloprovincialis, Crassostrea gigas*, *Crassostrea virginica* and *Mizuhopecten yessoensis*), we could not detect any putative *GULO* gene, and thus, it is likely the gene was lost in this lineage. Moving to the Gastropoda class, we were able to detect a potentially functional *GULO* in two species, namely, *Aplysia californica* and *L. gigantea*, in accordance with the reports by Wheeler *et al.* (2015). This result indicates that the *GULO* gene was probably maintained in this lineage. We could not take any conclusions concerning the Cephalopoda class, since we were not able to identify a putative *GULO* gene in the only species analyzed (*Octopus bimaculoides*).

The remaining sequences detected that were not placed in the expected phylogenetic position, were not analyzed. We observed that these *GULO* CDSs encoded for proteins that had the amino acidic motif HWGK (as the Porifera and Crustacea sequences), while all the GULO sequences had the motif HWAK. This information suggests that the GULO protein always has the HWAK motif across all animal species, and that proteins with the HWGK motif are not truly GULO, but rather FAD domain proteins coded by a gene other than *GULO*. Given this hypothesis, we propose that the real GULO amino acidic motif is HWAK, rather than the referenced HWXK. Interestingly, the five outgroup Fungi *ALO* sequences also seem to encode for proteins that contain the conserved HWAK motif.

Our results suggests that *GULO* origin predates at least the split of the Cnidaria and Bilateria taxonomic groups, around 642 million years ago. This is due to findings regarding the non-bilaterian Anthozoa class, in which we found a putative *GULO* gene, and to observations concerning the presence of *GULO* in 21 bilaterian taxonomic groups. This is a conservative overview when compared to the hypothesis by Wheeler *et al.* (2015). Nevertheless, following our criteria and considering the shortage of information regarding the basal Porifera group, we cannot assume that *GULO* was present or absent in the ancestor species from which all of the Metazoa (animal) taxonomic groups descended. Furthermore, although Fungi ALO synthetizes D-erythroascorbic acid instead of L-ascorbic acid, it shares several characteristics with animal GULO and is known to have some substrate specificity to L-Gulonolactone (Smirnoff *et al.* 2001). Given this evidence, *GULO* may be in fact much older than we can estipulate in our results, possibly being present in the ancestral of the Metazoa and Fungi taxonomic groups, as proposed by Wheeler *et al.* (2015).

The *GULO* gene appears to have been lost independently in many lineages, both in the Protostomia and Deuterostomia groups (five and three times, respectively). Inside the protostomians, regarding the Arthropoda phylum, we cannot draw conclusions concerning the *GULO* gene in the Hexapoda (w/o Insects) category, given that only one species was analyzed. Nevertheless, we suspect that this gene was lost after the split between the Pancrustacea and the Chelicerata taxonomic groups in the former lineage. Failure to detect a putative functional *GULO* in the Crustacea subphylum and Insecta class while confirming the presence of *GULO* in the Araneae order strongly supports this hypothesis. Regarding the Nematoda, there seems to be a clear loss of a putative *GULO* gene. Given the presence of a potentially functional *GULO* in the Priapulida and Arthropoda phyla and the absence of this gene in the Nematoda phylum, it seems evident that the gene was lost independently in the latter. These three phyla belong to the Ecdysozoa group and, as such, should share a common ancestor carrying a *GULO* gene. We also hypothesize that the Lophotrochozoa (comprised of Mollusca, Annelida and Brachiopoda phyla) ancestral had a potentially functional *GULO* gene, and that this gene was lost in the Bivalvia after the radiation of the Conchifera subphylum (which includes the Bivalvia, Cephalopoda and Gastropoda classes), while it was maintained in the Gastropoda class. The Platyhelminthes phylum has an uncertain taxonomic position in the Protostomia group. We do know that this phylum is not included within the analyzed Ecdysozoa and Lophotrochozoa, and thus it is possible that the *GULO* gene was also lost independently in this unplaced group. Observing the wide context regarding the detection of *GULO* sequences in Protostomian species, it is surprising that although we failed to detect *GULO* in three large Protostomian data sets (117 Insects, 25 Nematoda, and 10 Platyhelminthes genomes), we could identify a putative *GULO* in five rather small datasets (five Anthozoa, two Araneae, one Priapulida, one Brachipoda and three Gastropoda genomes). Furthermore, these putative *GULO* sequences do not seem the result of contamination,

since they are present in the expected position in the tree. Moreover, they are not *GULO* gene paralogous sequences, since they are closely related to the deuterostomian *GULO* sequences than to five Fungi *ALO* sequences used as outgroup, but also because do not appear in the consensus tree grouped with the nine sequences with a HWGK pattern. Even so, many non-bilaterian and protostomian taxonomic groups remain uncharacterized regarding the possible presence or absence of a putative *GULO* gene, and the evolutionary history of this gene is rather difficult to evaluate given the results obtained.

It seems clear that the deuterostomian ancestor species possessed a putative functional *GULO*, given the wide spectrum of Deuterostomia taxonomic groups in which we could detect this gene, as seen in Figure 2. Nevertheless, there were some apparent gene loss events detected. One noticeable loss event of the *GULO* gene is observed in the Teleostei infraclass. This observation, in combination with the detection of a putative *GULO* gene in the Holostei taxonomic group, suggest that the gene was present in the Actinopteri ancestor species and was lost independently in the Teleostei lineage after the radiation of the Neopterygii subclass. Other important gene loss events are observed in the Euarchontoglires superorder, namely in *C. porcellus* and in all 21 analyzed species representing the Haplorhini suborder. *C. porcellus* (guinea pig) belongs to the Caviidae family, included in the Hystricomorpha suborder of the Rodentia order. We cannot observe a putative *GULO* in this species, but we can identify *GULO* in several species included in other Hystricomorpha families, such as *Octodon degus* (Octodontidae), *Heterocephalus glaber* and *Fukomys damarensis* (both included in Bathyergidae). Given this information, we suggest that the *GULO* gene was lost independently in the Caviidae family, after the Hystricomorpha order radiation. Regarding the Primates order, our interpretation suggests that, due to the clear absence of a putative *GULO* gene in the Haplorhini suborder, and presence of a potentially functional gene in the Strepsirrhini suborder, the loss event should have occurred after the radiation of the Primates into these two sister groups. As already stated, the Passeriformes and the Chiroptera orders represent particular cases, in which there is a mix of species with a functional and non-functional *GULO* gene. Concerning the Passeriformes, we detected a putative *GULO* in two species, namely, *Corvus brachyrhynchos* and *Pseudopodoces humilis*. They belong to the Corvoidea and Sylvioidea superfamilies, respectively, which are included within the Passeri (songbirds) suborder. The songbirds comprise almost half of all identified avian species and represent the biggest observed radiation event of the birds taxonomic group. This radiation event culminated in a worldwide species dispersion with an overwhelming ecological and behavioral diversity (Keith Barker *et al.* 2004). Given this immense diversity, it is rather difficult tracking a gene evolutionary path within Passeri. Even within the *Corvus* genus, a rather specific taxonomic group, we obtained results in which a putative *GULO* gene is detected (*C. brachyrhynchos*), but also results in which a GULO protein pattern was not present (*Corvus cornix cornix*). To widen the analysis in an effort to understand

the *GULO* gene history within Passeriformes, even when considering the four species in which we detected a badly annotated *GULO* as effectively having one, the result is the same. The Passeriformes order is comprised of the Oscines (which includes the Passeri suborder) and Suboscines clades, as well as an Acanthisittidae family sister-group. With this approach, we consider that *GULO* is present in a single species belonging to the Acanthisittidae family (*A. chloris*) and a single species included in the Suboscines group (*M. vitellinus*). Furthermore, the two remaining species with a possible putative *GULO*, namely *L. striata* and *F. albicollis*, are included in two distinct superfamilies of Passeri, Passeroidea and Muscicapoidea, respectively. Given the information acquired from this dataset, different hypothesis are possible: i) considering only the confirmed putative *GULO* observations, the gene could have been lost at the base of the Passeriformes group, and eventually reacquired by *C. brachyrhynchos* and *P. humilis* or; ii) considering those two species but also the four with poor annotation, the gene was present at the base of all Passeriformes but lost independently in many species afterwards. We cannot discriminate either one with the available *GULO* CDSs available at the time. Regarding the Chiroptera case, we observe that the vast majority of the Laurasiatheria species analyzed, excluding the bats, have an identifiable potentially functional *GULO*. Nevertheless, when observing the bats specifically, the majority of the species sequences analyzed were excluded in the pattern-filtering step. The bat species in which we can identify a putative *GULO*, *Rousettus aegyptiacus* and *Rhinolophus sinicus*, belong, respectively, to the Megachiroptera and Microchiroptera suborders. Interestingly, when overlapping the excluded and analyzed sequences taxonomic data, we found that the Pteropodinae subfamily (Megachiroptera) was represented in both cases: *R. aegyptiacus* has a putative *GULO*, while *Pteropus Alecto* and *Pteropus vampyrus* do not. Adding the results of Cui *et al.* (2011) regarding this family, we also know that a functional GULO with 440 amino acid residues and a HWAK pattern, sharing 90% identity with *M. musculus* GULO (NP_848862.1), is present in *Rousettus leschenaultia* (GenBank accession number ADP88813.1). Furthermore, we observe that of the 10 bat species analyzed, seven are included in the Microchiroptera suborder, spread across three families: Hipposideridae (*H. armiger*), Vespertilionidae (*Eptesicus fuscus*, *Miniopterus natalensis*, *Myotis davidii*, *Myotis brandtii* and *Myotis lucifugus*) and Rhinolophinae (*Rhinolophus sinicus*). We report the detection of a putative *GULO* in *R. sinicus*, however, we know that these results are incomplete. Cui *et al.* (2011) was able to show that *Hipposideros armiger* has a functional GULO (GenBank accession number ADP88814.1) with 440 amino acid residues and a HWAK pattern, sharing 91% identity with *M. musculus* GULO (NP_848862.1). Tracking how we could have missed this information along the dataset processing, we verified that when downloading all the animal annotated CDS, the GenBank files regarding species with valid *GULO* CDS accession numbers such as *R. leschenaultia* and *H. armiger* were rather incomplete. Somehow, even a model species like *H. sapiens* (human) only has 13 gene CDS annotated according to GenBank. Moreover, although

four putative *GULO* CDSs are identified in the initial tblastn regarding the RefSeq file for *H. armiger*, only one is a predicted as being a real GULO (XP_019491656.1). Neverthless, even this one is badly annotated and does not even contain the HWXK amino acid pattern. This is a peculiar case in which, even when overlapping the GenBank and RefSeq data to obtain the most informative FASTA file possible, the file ends up incomplete and thus misleading when analyzed. This does not represent a methodological problem, but rather a limitation of the databases. Even so, this is a great example that illustrates the need for caution when formulating hypotheses based on obtained phylogenetic results. Returning to the results, we observe that *Galeopterus variegatus*, the only species belonging to the Dermoptera order in this dataset, is excluded in the GULO HWXK pattern filtering step. When only one species is analyzed, our criteria indicates that we cannot make conclusions regarding a possible loss of the *GULO* gene in that given lineage. In spite of that, it is curious that the Chiroptera and Dermoptera orders are sister groups, although they belong to distinct superorders, Laurasiatheria and Euarchontoglires, respectively. Also within the Euarchontoglires, a putative *GULO* is observed within the Primates order, but also in the Scadentia sister group (represented by *Tupaia chinensis*). Collecting all this evidence, we suggest two possible hypotheses regarding the *GULO* gene in bats. The first implies that *GULO* could have been lost after the Boreoeutheria group (comprised of Laurasiatheria and Euarchontoglires) radiation, in the lineage that leads to the Chiroptera/Dermoptera phylogenetic branch, while it was maintained in the Primate and Scadentia lineage. As such, the current taxonomic knowledge regarding the placement of the Chiroptera and Dermoptera orders in distinct superorders does not correctly apply in this specific case. The second hypothesis suggests an independent gene loss event in both Chiroptera and Dermoptera orders that occurred after the Laurasiatheria and Euarchontoglires radiations, respectively. The second option seems more likely with the information we have, while the first needs some complementary work concerning the number of species analyzed in the Dermoptera order. Nevertheless, in both alternatives there is a common assumption: given the bat phylogeny presented by Teeling *et al.* (2005) already adapted by Cui *et al.* (2011) and also our results, it seems plausible that *GULO* was lost at the base of all Chiroptera, and that the inactive gene was reactivated during evolution in some species. Even so, and as Cui *et al.* (2011) states, we cannot assume that *GULO* may not evolve to become a pseudogene eventually, since it is clear that a large number of bat species do not require this functional gene.

The consensus phylogeny obtained (Supplementary figure 1) also shows that *GULO* is a single copy gene in all animal species analyzed, with the exception of *A. planci*. Therefore, it seems that *GULO* duplicates were not retained after the two rounds of closely spaced auto-tetraploidization events (commonly known as 1R and 2R) that occurred early in vertebrate evolution (Dehal and Boore 2005; Putnam *et al.* 2008). Some reports suggest that they may have occurred during chordate evolution, after the split of the urochordate and cephalochordate

lineages but before the radiation of gnathostomes. It is possible that *GULO* had already been lost in the common ancestor of all Teleosti fish, and thus the whole-genome duplication event that occurred at that time (Taylor 2003) is not relevant regarding the gene evolutionary history within this group. Moreover, a single *GULO* gene is found even in a recent allotetraploid species such as *Xenopus leavis* (Session *et al.* 2016).

Our results imply that *GULO* was not lost a single time in the Protostomia group, but multiple times independently, after the split of the Protostomia and the Deuterostomia lineages, which complement the hypothesis formulated by Wheeler *et al.* (2015).

III.2. Non-Bilateria/Protostomia *GULO* gene annotations and phylogenetic analysis

III.2.1. *GULO* sequence annotation process

The results concerning the phylogenetic analysis of available annotated animal *GULO* CDS implied the unreported presence of a putative functional *GULO* gene in three Protostomia lineages, namely the Araneae family and the Priapulida and Brachiopoda phyla. Nevertheless, many protostomian and non-bilaterian lineages remained uncategorized regarding the presence or absence of a potential *GULO* gene. The main reason for these inconclusive results was the number of species analyzed with an already annotated genome, since we could not assert the absence of *GULO* with any less than three representative species CDS within each taxa, and in many cases, only one or two species were available in the final processed dataset. To overcome this lack of information, we performed annotations for the representative GenBank genomes of all available species included in the Protostomia and non-bilaterian groups highlighted in Figure 2.

Within the Non-Bilateria, we analyzed two species included in the Porifera phylum, namely *A. queenslandica* and *Aplysina aerophoba*. In *A. queenslandica*, the tblastn results obtained showed a high alignment coverage between *M. musculus* GULO and the used genome, dispersed across several scaffolds. Moreover, we detected a putative exonic sequence that encoded for the HWGK amino acid pattern. We cannot assume that the scaffolds in which we detected potential *GULO* exons are successional portions of the genome, since they are not assembled and can even be present in different chromosomes. As such, we failed to annotate the *GULO* gene in this species. As for *A. aerophoba*, we could not detect a probable *GULO* exon that encodes for the conserved HWXK GULO amino acid motif, and so the annotation was not performed. We were able to perform an annotation for *T. adhaerens*, the only available species belonging to the Placozoa phylum. According to our annotation, we were able to identify 13 introns for this potential gene, which encodes a protein with a HWAK amino acid motif. Regarding the Myxozoa phylum, we tried to perform annotations using the genomes of four species, namely *T. kitauei*, *Kudoa iwatai*, *Sphaeromyxa zaharoni* and *Enteromyxum leei*, but were only able to annotate a putative *GULO* in *K. iwatai*. Our annotation suggests the presence of 13 intronic regions within this *K. iwatai* potential gene, which may code for a putative protein with a HWAK amino acid motif. *T. kitauei* tblastn result presented a complete sequence coverage with a possibly encoded HWGK amino acid motif in the alignment. Even so, we verified the presence of in-frame stop codons within the alignment, and as such, could not annotate a putative gene. For *S. zaharoni* and *E. leei*, we observed a partial alignment coverage in the tblastn results, in both cases with exons that may code for a HWAK amino acid motif. Given the alignment gaps observed in these results, we were not able to perform an annotation for these species. We did not analyze any Anthozoa species since we considered that the results obtained for our *GULO* CDS

dataset and the reports by Wheeler *et al.* (2015) represented strong enough evidence that the gene may be present within this taxonomic group. As for the Hydrozoa phylum, we analyzed only one genome belonging to *Hydra magnipapillata*. In this species, we could not identify a single genomic sequence with homology with the *M. musculus* GULO that could code for the conserved HWXK amino acid pattern, and so an annotation was not performed.

Concerning the Protostomia, we analyzed four Hexapoda (excluding insects) species genomes, namely *Catajapyx aquilonaris*, *Folsomia candida*, *H. duospinosa* and *Orchesella cincta*. We observed a low coverage sequence alignment with no putative encoded HWXK amino acid motif for *C. aquilonaris*, and therefore did not perform an annotation for this species. In *O. cincta*, we were able to annotate a putative *GULO* gene with three introns, which may code for a protein with an HWGK amino acid pattern. Furthermore, we identified a possibly duplicated *GULO* gene in *F. candida* and *H. duospinosa*. Regarding *F. candida*, two gene annotations were performed, which we designated as *Folsomia candida* 1 and 2, respectively. The first annotated putative *GULO* sequence had four introns, whereas the second had two. Nevertheless, both sequences may code for proteins with an HWGK amino acid pattern. In *H. duospinosa*, the two performed sequence annotations were entitled *Holacanthella duospinosa* 1 and 2, and both presented a single intron. In addition, we were able to observe that both potential *GULO* sequences could possibly code for a protein with an HWGK amino acid pattern. Given the results obtained for the Insecta class using the previously described *GULO* CDS dataset approach, in which a putative *GULO* was not detected in 116 species, we considered that further analyzing this taxonomic group would lead to the same conclusion, and did not try to perform annotations in this case. As for the Crustacea, we felt that the previous results, where we did not detect a *GULO* in four species, were on the brink of our three species criteria for hypothesizing a gene loss within a lineage. Although we were confident in our results, to strengthen our conclusion, we further analyzed this taxonomic group by performing a direct tblastn in the NCBI database, using as query *M. musculus* GULO against only the Crustacea class species representatives. Using this assay, we were able to determine that only two species had a tblastn alignment hit, namely *Eurytemora affinis* and *Hyalella azteca*. Furthermore, by observing the sequence alignment results, we uncovered that the only species which could code for a protein with a conserved HWXK pattern was *H. azteca*. Given this result, we proceeded to download this species genome for a more detailed analysis. Using the *H. azteca* genome, we were able to annotate two distinct putative *GULO* sequences with a single intron each, both probably encoding for a potential HWGK pattern containing protein. These sequences were labeled as *Hyalella azteca* 1 and 2.

Concerning the Araneae family, included in the Chelicerata subphylum, we were able to refine the annotation already available for *GULO* in *Parasteatoda tepidariorum* (XP_015913395.1) that was displayed in the *GULO* CDS protocol consensus phylogeny. Since

we already had an annotation for this species, we did not perform a new annotation using the complete genome. Instead, we performed a blastp to identify the regions of homology between the *P. tepidariorum* and the *M. musculus* GULO. Verifying the blastp results, we observed that the GULO protein of *P. tepidariorum* was 26 amino acids bigger in the N-terminus region when compared with *M. musculus* GULO, given that the first methionine of *M. musculus* GULO aligned with the methionine at the amino acidic position 27 of *P. tepidariorum* GULO. Given this evidence, we assumed an alternative putative ATG start for the available CDS and shortened it in 78 nucleotides at the beginning of the sequence, leaving the rest of the sequence unchanged. For the remaining five Araneae species analyzed, we tried to perform a *de novo* annotation using the available genomes information. We failed to annotate a potential *GULO* gene in three of these species, namely *L. reclusa*, *S. mimosarum* and *Acanthoscurria geniculata*. In *L. reclusa*, the tblastn genomic hit region obtained contained many ambiguous nucleotide sites, which partially overlapped putative exons and did not allow a complete annotation. As for *S. mimosarum* and *A. geniculata*, we obtained a partial sequence alignment coverage spread across several scaffolds in the tblastn results, and therefore, an annotation was not performed. Nevertheless, in both species, genomic regions that potentially code for the GULO HWAK conserved amino acid motif were identified. We were able to perform an annotation for the two remaining Araneae species analyzed, *Latrodectus hesperus* and *Nephila clavipes*. In *L. hesperus* we detected a possible *GULO* gene with nine putative introns, while in *N. clavipes* we observed a potential *GULO* with 11 introns. Furthermore, we found that both annotations possibly code for proteins with an HWAK amino acid pattern. Still within the Chelicerata, and now regarding the Acari subclass, we used 16 species genomes to execute our annotation analysis. No putative *GULO* sequence could be annotated in eight of these species genomes, belonging to *Galendromus occidentalis*, *Varroa destructor*, *Varroa jacobsoni*, *Sarcoptes scabiei*, *Tropilaelaps mercedesae*, *Tetranychus urticae*, *Ixodes ricinus* and *Dermatophagoides pteronyssinus*. In the first five mentioned species, the results obtained from the tblastn showed no similarities between any genomic scaffold and the *M. musculus* GULO used as query. In *T. urticae*, we were not able to identify any genomic region in the tblastn alignment that could code for a conserved HWXK amino acid motif, and thus it is not likely that the sequences identified in the tblastn results belong to *GULO* gene. As for *I. ricinus*, the nucleotide sequences that share identity with *M. musculus* GULO are scattered across different scaffolds, which may or may not be successive in the genomic context. Although we did not carry out an annotation for a potential *GULO* in this species, we verified that one of the sequences present in the alignment could code for a protein region with a HWAK amino acid pattern. This conserved amino acid pattern could also be encoded by a sequence retrieved from the *D. pteronyssinus* genome after the tblastn search. Even so, an annotation was not performed for this species due to the presence of a single in-frame stop codon preceding the third potential exon of a putative *GULO* gene. We performed partial gene annotations for *I. scapularis*,

Steganacarus magnus and *E. maynei*. In *I. scapularis*, we failed to annotate the putative *GULO* last exon, due to the presence of ambiguous nucleotide positions in this genome region. For *S. magnus*, we annotated a truncated putative *GULO* sequence that shared identity with the *M. musculus* GULO only after amino acidic position 109. The premature end of the scaffold that we used to perform the annotation at the 5' extremity did not allow a further annotation. As for *E. maynei*, we annotated a truncated putative *GULO* sequence at the 5' end, but also missing the final putative exon. This sequence shared identity with the *M. musculus* GULO after amino acidic positions 134 until 398, and could not be further annotated due to the premature end of the analyzed genomic scaffold at both extremities. Although these three sequences are all partial annotations, they have high Blast alignment scores (>=200) and can potentially code for a protein with a HWAK conserved motif. We were able to annotate a potential *GULO* in the remaining six species genomes analyzed, belonging to *Dermatophagoides farinae*, *Achipteria coleoptrata*, *Hypochthonius rufulus*, *Platynothrus peltifer* and *Rhipicephalus microplus*, respectively. These annotated sequences enabled us to infer five probable introns within the putative *GULO* gene for *D. farinae*, *A. coleoptrata* and *H. rufulus*, three introns for *P. peltifer* and one for *R. microplus*. Furthermore, all of these annotated sequences might code for a protein with a conserved HWAK amino acid motif. We also tried to perform annotations for three Chelicerata species that were not included in the Araneae and Acari taxonomic groups, namely *Limulus polyphemus*, *Mesobuthus martensii* and *Centruroides sculpturatus*. No annotation resulted from the analysis of the genomes for each of these species, since the sequences that had similarities with the *M.musculus* GULO obtained after the tblastn belonged to different and possibly non-successive genomic scaffolds. Nevertheless, all these species had a putative exonic sequence that could code for a protein region with a conserved HWAK amino acid motif. Regarding the Tardigrada phylum, we analyzed the available genomes for *R. varieornatus* and *H. dujardini*. We failed to annotate a probable *GULO* gene in these species, since the alignment coverage of the sequences retrieved from the tblastn initial results was rather incomplete in both cases. Even so, by observing these low coverage sequence alignments, we detected possible genomic sequences that could code for a protein with a HWGK amino acid pattern, for both species. The only representative genome available for the Priapulida species in the GenBank database belongs to *P. caudatus*. Since we already obtained a representative *GULO* sequence for this species in the previous animal *GULO* CDS protocol analysis (XP_014666894.1), we did not perform a new annotation using the genome information. Alternatively, we performed a blastp to identify the regions of homology between the *P. caudatus* and the *M. musculus* GULO to verify if the available annotation could be further refined, similarly to the case of *P. tepidariorum*. The blastp alignment had full coverage between the query and target protein sequences, except for the first four amino acids, with only one observable gap out of 437 amino acid positions analyzed. Moreover, the *P. caudatus GULO* CDS encodes for a protein that has only four more amino acids than the *M. musculus* GULO. These evidences are

indicative of a good annotation process, and as such, we did not perform additional adjustments to the *P. caudatus* available CDS. The GULO gene in *P. caudatus* has 11 potential introns and may code for a protein with a HWAK amino acid pattern. For the Nematoda and Platyhelminthes, the results obtained in the previous animal CDS analysis suggested the absence of a *GULO* gene in 35 and nine species, respectively. Even so, we sought out to strengthen these results using a tblastn in the NCBI database, using as query *M. musculus* GULO against the Nematoda and Platyhelminthes phyla species representatives. Using this approach, we were able to determine that only one Nematoda species (*Brugia timori*) had a tblastn alignment hit. However, the genomic sequence with similarities to *M. musculus* GULO did not code for a HWXK amino acid pattern and had poor alignment coverage with a expect value of 0.002. With these results, we considered unnecessary any attempt to perform a putative *GULO* gene annotation in the species belonging to these taxonomic groups, given the clear evidence for the lack of *GULO* seen within the nematodes and Platyhelminthes genomes analyzed so far. As for the Mesozoa, we analyzed the only available species genome, belonging to *I. linei*. We could not perform a gene annotation for these species since the tblastn output did not contain any sequences with homology with *M. musculus* GULO. Regarding the Brachiopoda group, we already detected a putative *GULO* sequence in *Lingula anatina* (XP_013393535.1), using the animal available CDS protocol. Like in *P. tepidariorum* and *P. caudatus*, we verified if this annotation could be further refined using a blastp approach. We verified that the obtained alignment between *L. anatina* GULO and the *M. musculus* GULO had full coverage, with the exception of the first three amino acids. Furthermore, the alignment had only two observable gaps out of 437 amino acid positions analyzed, and the GULO protein encoded by the available *L. anatina GULO* CDS had merely five more amino acids than the reference *M. musculus* GULO. By observing these parameters, we did not find evidence for a bad annotation process, and as such, did not perform additional adjustments to the *L. anatina* annotated *GULO* CDS. The GULO gene in *L. anatina* has 11 potential introns and may code for a protein with a HWAK amino acid pattern. Still regarding the Brachipoda phylum, we analyzed the available genome for *Phoronis australis*, but could not perform a gene annotation given that no nucleotide sequences with homology with *M. musculus* GULO were detected through the tblast search. We also tried to annotate a putative *GULO* gene in six Annelida species genomes, namely *Amynthas corticis* A, *Amynthas corticis* C, *C. teleta*, *Eisenia fetida*, *H. robusta* and *Hydroides elegans*. For *H. robusta* and *H. elegans*, we did not obtain sequences with homology to *M. musculus* GULO from the output of the tblastn, and thus no annotation was performed. Moreover, we were not able to annotate a probable *GULO* sequence in *A. corticis* A and *A. corticis* C. In *A. corticis* A, the nucleotide sequences that shared homology with *M. musculus* GULO were scattered across several genomic scaffolds. Even so, we observed that one of the sequences could code for a protein with a HWAK amino acid. In *A. corticis* C, we found no identifiable homologous nucleotide sequence that could code for a HWXK amino acid pattern. For *C. teleta*

and *E. fetida*, we were able to annotate potential *GULO* gene sequences that may code for proteins with a conserved HWAK and HWGK amino acid motifs, respectively. However, the obtained *E. fetida* sequence is missing nucleotide information that may code for the first 58 amino acid positons of a putative GULO protein, due to the precocious ending of the genomic scaffold we used to perform the annotation at the 5' terminus. We were able to infer 11 possible introns within the annotated *C. teleta GULO*, while for *E. fetida* we could not make an estimate.

Within the Gastropoda group, we evaluated if the available *GULO* annotations for *A. californica* (XP_005103891.1) and *Lottia gigantea* (ESO97787.1) could be further refined. With a blastp search using *M. musculus* GULO as query, we were able to determine that the obtained sequence alignments had full coverage for both species, with the exception of the first three amino acids. In addition, the alignments had only two observable gaps out of 437 amino acid positions analyzed for *A. californica*, and four distinguishable gaps out of 439 amino acid positions analyzed for *L. gigantea*. Moreover, the putative GULO protein encoded by the *A. californica* CDS had fifteen more amino acids than the reference *M. musculus* GULO, while the protein possibly encoded by the *L. gigantea* CDS had sixteen more. With this information, we considered the available CDSs for these species rather well annotated, and as such, did not perform additional annotation adjustments. The GULO gene in *A. californica* has 10 potential introns and may code for a protein with a HWAK amino acid pattern, as also seen in *L. gigantea*. We also tried to perform our own *GULO* gene annotations using the genomes of *Biomphalaria glabrata*, *Colubraria reticulata*, *Conus tribblei*, *Lymnaea stagnalis* and *Radix auricularia*. We failed to annotate a putative gene in *B. glabrata* and *C. tribblei*. In *B. glabrata*, the annotation process led to an incomplete CDS with in-frame stop codons. Furthermore, the only potential exonic sequence that could code for a protein region with a HWAK amino acid motif was in a different translation frame from the one seen in the rest of the putative gene. As for *C. tribblei,* the tblastn results revealed a poor sequence alignment coverage between the *M. musculus* GULO and the used genome, with the detected homologous sequences dispersed across several scaffolds. It was also possible to observe that, for this species, two distinct exonic sequences that could code for HWAK and HWGK amino acid patterns were identified. We were able to obtain annotations for *C. reticulata*, *L. stagnalis* and *R. auricularia*. The putative *GULO* annotations for *L. stagnalis* and *R. auricularia* allowed us to infer the presence of 11 possible introns in the gene sequence. As for *C. reticulata*, the tblastn results allowed us to annotate one full-coverage putative *GULO* sequence and another truncated at both extremities due to premature scaffold ending, which we labeled as *Colubraria reticulata* 1 and 2, respectively. We deduced the presence of one intron in the full-coverage annotation, while for the truncated sequence we were unable to estimate an intron number due to lack of information regarding the first 171 amino acidic positions in the tblastn alignment. Neverthless, all the annotated sequences for these three species may code for proteins

that contain a conserved HWAK amino acid pattern. Concerning the Bivalvia taxonomic group, we sought to annotate a putative GULO gene in ten species genomes, belonging to *Bankia setacea*, *Bathymodiolus platifrons*, *Modiolus philippinarum, M. galloprovincialis*, *C. virginica*, *Pinctada martensii*, *C. gigas*, *Corbicula fluminea, M. yessoensis* and *Dreissena polymorpha*. The tblastn search with *M. musculus* GULO as query only found homologous nucleotide sequences in the *B. setacea*, *B. platifrons*, *M. philippinarum* and *M. galloprovincialis* species genomes, while in the remaining genomes none was retrieved. However, the identified homologous sequences did not present any putative exonic regions that could code for the typical GULO HWXK amino acid pattern, and thus, an annotation was not performed. Regarding the Cephalopoda group, we tried to annotate the *GULO* gene in the only available representative genome for this class, belonging to *Octopus bimaculoides*. However, we could not detect any nucleotide sequence with homology to *M. musculus* GULO through the tblastn search, and as such, did not perform an annotation.

In all our *de novo* gene annotations, we were not able to detect the potential ATG start codon that encodes for the first translated protein amino acid, methionine. Trying to overcome this problem, we performed several tblastn searchs using *M. musculus* GULO as query against our annotations plus their extended 5' flanking genomic region, whenever possible. Using this approach, we hoped to detect a putative first exon for our gene annotations that could contain the ATG start of the coding sequence. Even so, this methodology did not work. Still, the passerine coding sequences analyzed in the previous dataset gave information regarding a potential ATG site preceded by the AG nucleotides, which could indicate a canonical GT-AG splice site at that location. Given this evidence, it is possible to speculate that our annotations may present the same features as these passerine sequences, and that without gene expression analysis, we cannot simply infer the ATG start codon using only a tblastn search. Even so, our annotations have high sequence alignment coverage with expect values remarkably close to zero when queried against the *M. musculus* GULO protein, while also not presenting any in-frame stop codons. Therefore, it is rather unlikely that these gene annotations could represent a pseudogene using only the absence of an identifiable ATG start codon as evidence, and as such, we are confident that we are identifying functional genes. As such, we adjusted the number of observed putative introns (by adding one intron) in our gene annotations, to follow the hypothesis in which the first exon containing the ATG start codon is not found, but expected to exist preceding our annotations in the analyzed genomes. The number of identified introns for each annotated putative *GULO* sequence presented until now already considers this correction.

III.2.2. *GULO* CDS annotations phylogenetic analysis

The consensus phylogeny obtained from the ADOPS pipeline output concerning our annotations is presented in Figure 3. Given the results obtained for the animal *GULO* CDS protocol, we suspected that sequences that code for a protein with an HWGK amino acid motif likely do not constitute a real *GULO*. By performing our own annotations, we were able to obtain eight sequences that may encode putative proteins with this HWGK pattern, namely for three Hexapoda species from the Collembola class (*O. cincta*, *H. duospinosa* and *F. candida*), for one species from the Crustacea group (*H. Azteca*) and for one Annelida species from the Clitellata group (*E. fetida)*. The Bayesian phylogeny reveals that these eight sequences are represented as an external tree branch to our selected Fungi *ALO* outgroup. This result further reinforces the evidence found in the previous protocol, and suggests that these annotations do not represent a real *GULO*, but rather another gene that could possibly code for a distinct FAD-domain protein. Interestingly, the annotations belonging to *K. iwatai* and *R. microplus* also appear misplaced in the phylogeny, even though they may code for proteins with an HWAK amino acid motif. To identify the possible cause for this unexpected result, we started by performing blastp searches using the translated *K. iwatai* and *R. microplus* annotations against the complete NCBI database to check for possible genome contamination or horizontal gene transfer events that could influence the phylogeny outcome. For *R. microplus*, we observed that the translated annotation is remarkably similar to a FAD-linked oxidoreductase from the Alphaproteobacteria *Sphingomonas* sp. 67-36 (OJV32873.1), sharing 88% identity across a 420 amino acid alignment with the mutual presence of a conserved HWAK pattern. Moreover, the alignment result had a 0.0 expect value, strongly suggesting that no protein sequences are expected to match as well or better than the hit we obtained, and so it is very unlikely that this result constitutes a false positive. Furthermore, the annotated sequence for *R. microplus* derived from a single continuous exon. Given these observations, it seems likely that a process of genome contamination with exogenous bacterial genetic material may have occurred, which in turn, can explain the abnormal position of the putative *R. microplus GULO* sequence in the consensus phylogeny. Coincidentally, this also seems to be the case for the *E. fetida* annotation grouped with the *R. microplus* sequence. Through a blastp search, we observe that the putative protein encoded by the clitellate *E. fetida* gene sequence is 99% identical to a FAD-binding protein from the *Candidatus lumbricidophila eiseniae* actinobacteria (PDQ34519.1) across the 355 amino acid alignment, with a 0.0 expect value. In addition, this annotation was obtained from a single continuous putative exon within the *E. fetida* genome, further suggesting a contamination process. Regarding the *K. iwatai* annotation, we did not find evidence for any of these events, and the relatively low 71% posterior credibility

value suggests that this sequence may in fact be placed in another position of the phylogeny with a considerable probability.

Within the green highlighted sequences that may represent a real *GULO* (Figure 3), we observe that the partial annotation for the parasitiform tick *Ixodes scapularis* is abnormaly placed in the phylogeny. Given that *I. scapularis* is a protostome included within the Acari subclass, theoretically, this sequence should be grouped with the Acari Sarcoptiformes taxonomic group comprised of the *A. coleoptrata*, *P. peltifer*, *H. rufulus*, *S. magnus*, *D. farinae* and *E. maynei* species, in the consensus tree. Instead, this sequence is grouped with five *GULO* sequences representative of the deuterostome Reptiles (*A. mississippiensis*), Aves (*G. gallus*), Euarchontoglires (*M. musculus*), Amphibia (*Xenopus laevis*) and Actinopteri (*L. oculatus*) taxonomic groups.

It is know that ticks such as *I. scapularis* are blood-feeding ectoparasites with a broad host range, ranging from small vertebrates to humans (Gulia-Nuss *et al.* 2016). Our annotation protocol for *I. scapularis* allowed us to identify 11 putative introns within a possible *GULO* gene, the same intron number identified in the *G. gallus*, *M. musculus*, *X. laevis* and *L. oculatus GULO*. However, annotations performed for the remaining Acari species from the Sarcoptiformes taxonomic group allowed us to account for only three (*P. peltifer*) to five (*A. coleoptrata, D. farinae* and *Hypochthonius rufulus*) introns within the putative gene. Given these considerations, it would seem plausible that the probable *GULO* annotated in *I. scapularis* resulted from technical issues, due to contamination of the sequenced genome with genetic material from some deuterostome host species. Nevertheless, this does not appear to be the case, since by analyzing our *I. scapularis* translated annotation using a blastp search against all annotated species NCBI database, we did not obtain hit results with values of identity over 53%, in alignments with 98 to 100% query coverage. Alternatively, as considered for *K. iwatai*, the 77% posterior credibility value given to the *I. scapularis* sequence branch suggests that the sequence can be placed somewhere else in the consensus tree. Still regarding the taxonomic context, observing the non-bilaterian *N. vectensis* species adjacent to the deuterostomian cephalochordate *B. belcheri* in a branch with 100% posterior credibility value is surprising. Moreover, the ancestral species belonging to the Placozoa (*T. adhaerens*) phyla is represented as a sister group of the Acari (Sarcoptiformes) and Araneae (Theridiidae and Nephilidae), in a branch with a 88% posterior

credibility value.

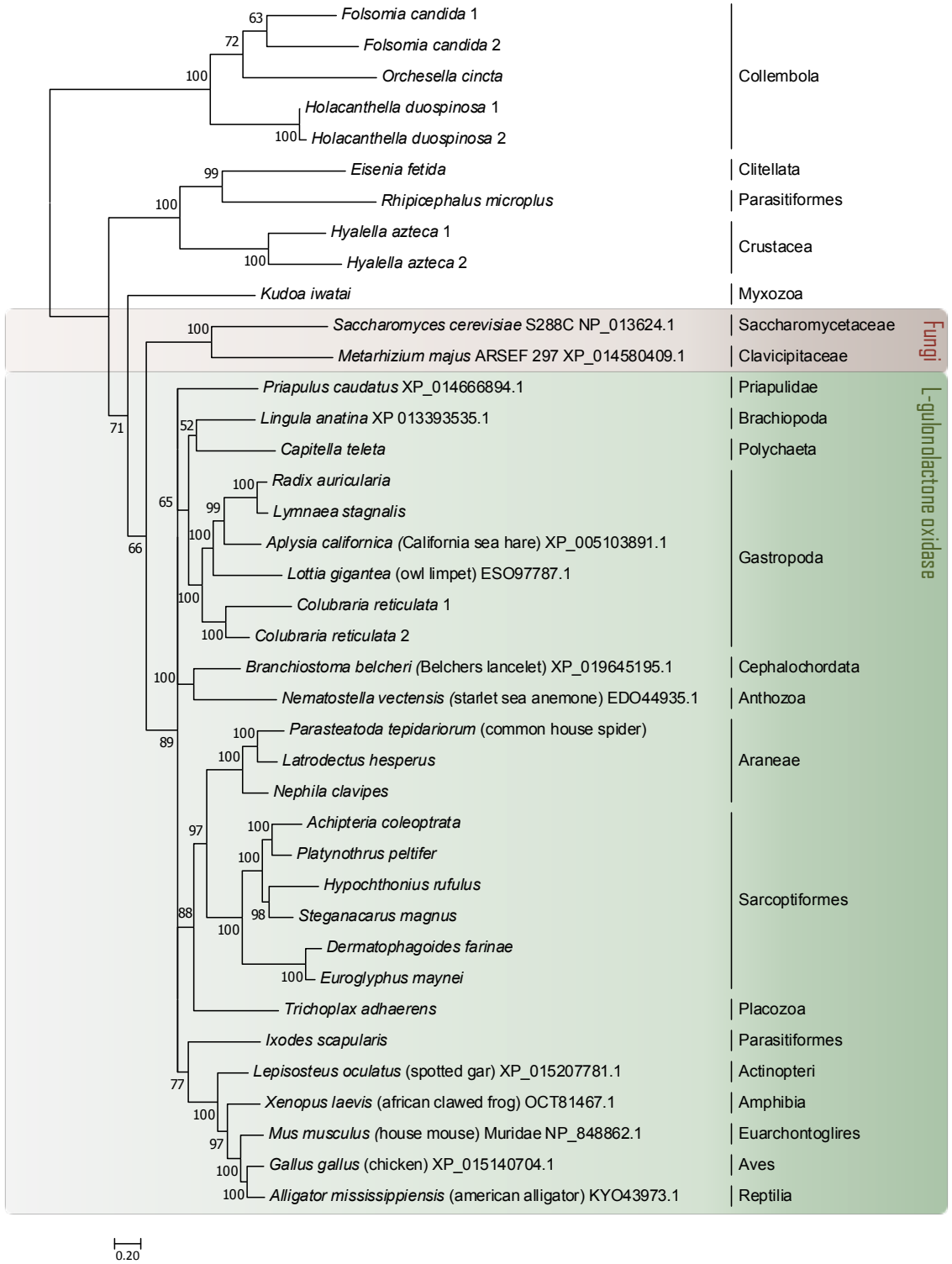


Figure 3 – Protostomes and non-bilaterians putative *L-gulonolactone oxidase* (*GULO*) annotations phylogeny. Two *GULO* CDS from Fungi species were used to help rooting the tree. Six deuterostomian species (representative of the Actinopteri, Amphibia, Euarchontoglires, Aves, Reptilia and Cephalochordata groups) *GULO* CDS were used to facilitate the interpretation of the results. Relevant higher taxonomic classifications are shown next to the species name.

It is rather common to find conflicting information when analyzing real datasets and the events leading to these disaccords are well described. Possible horizontal gene transfer incidents, lack of synapomorphy due to multiple nucleotide substitutions or stochastic errors can lead to erratic phylogenies. Nucleotide substitutions may also randomly create sequence similarities that, in turn, may cluster distantly related lineages in a phylogeny (Wägele and Mayer 2007). It is known that substitution models can sometimes correct a few of these effects (Arenas 2015). However, the sequences displayed in the consensus phylogeny represent species with an evolutionary divergence time up to 656 million years (Peterson *et al.* 2004), which in turn originates a very complex dataset for phylogenetic analysis. Given this time scale, it was expected that even using the most likely and complete phylogenetic relations inference model for our data, some annotations would be misplaced in the phylogeny due to, for example, greatly distinct rates of sequence evolution. Nevertheless, the obtained phylogeny seems to follow a correct taxonomic context in general. Furthermore, apart from the *K. iwatai* annotation, even the particular sequence placements mentioned above are suggestive of *GULO*, since they share an internal position relative to the Fungi outgroup.

Hopping to generate a more defined phylogeny from which we could draw conclusions regarding the *K. iwatai* putative *GULO* sequence placement, we produced a new FASTA file without the annotations that may code for a protein with an amino acid HWGK conserved pattern and the *R. microplus* sequence we think derives from bacterial contamination. This file was analyzed with the same parameters as before using the ADOPS pipeline, and the phylogeny output can be observed in Figure 4. In the new consensus tree, it is possible to observe that all the Gastropoda species annotations are clustered with 100% posterior credibility value, with the Polychaeta and Brachiopoda groups (represented by the *C. teleta* and *L. anatina* annotations, respectively) placed as sister groups, as expected. Additionally, *GULO* appears to be duplicated in the gastropod *C. reticulata*. Contrary to the remaining Gastropoda annotations, these two putative gene sequences resulted from different single continuous exons. Even so, they are included in the correct taxonomic position in the phylogeny and should likely constitute a real *GULO*. The three Araneae species annotations are placed together with 100% posterior credibility value, seen as a sister group to the six Sarcoptiformes species putative *GULO* sequences branch. Furthermore, we can also see that the *K. iwatai* annotation is now placed internally relative to the Fungi outgroup. This result suggests that the *K. iwatai* sequence appears in fact closely related to the remaining annotated putative *GULO* sequences than to the Fungi *GULO*, strongly corroborating our hypothesis of technical issues regarding the previous phylogeny representation. No changes are seen regarding the non-bilaterian *N. vectensis* and cephalochordate *B. belcheri* species annotations, remaining grouped in a branch with 99% posterior credibility value derived from the polytomy. Additionally, the placement of the placozoan *T. adhaerens* annotation is still

adjacent to the Acari (Sarcoptiformes) and Araneae (Theridiidae and Nephilidae) sequences, in a branch with an even higher 99% posterior credibility value, when compared to the previous phylogeny. The *P. caudatus* sequence is now unexpectedly grouped with five deuterostomians (*A. mississippiensis*, *G. gallus*, *M. musculus*, *X. laevis* and *L. oculatus*) and the parasitiform *I. scapularis GULO* annotations. However, the very low 60% posterior credibility value suggests that this annotation can be placed almost randomly in the phylogeny, and could likely be inserted in the correct taxonomic context (near the Araneae and Sarcoptiformes annotations). The *I. scapularis* annotation maintains the same surprising placement in the phylogeny when compared to the previous consensus tree, with a higher posterior credibility value for the corresponding branch (rise from 77 to 82%).

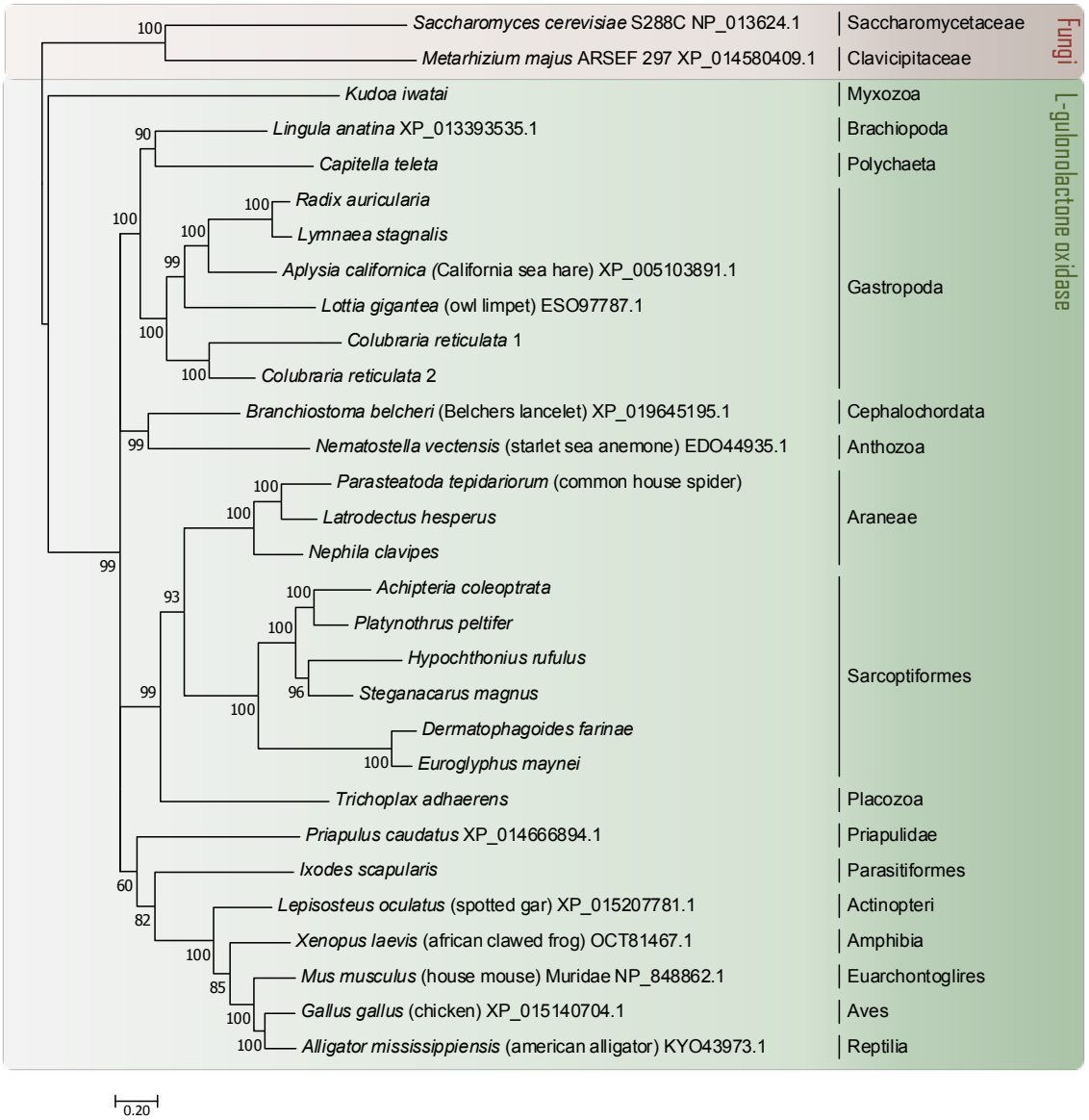


Figure 4 – Refined protostomes and non-bilaterians putative *GULO* annotations phylogeny. Two *GULO* CDS from Fungi species were used to help rooting the tree. Six deuterostomian species (representative of the Actinopteri, Amphibia, Euarchontoglires, Aves, Reptilia and Cephalochordata groups) *GULO* CDS were used to facilitate the interpretation of the results. Relevant higher taxonomic classifications are shown next to the species name.

In conclusion, we can summarize the findings concerning the Araneae, Acari and Mollusca taxonomic groups and hypothesize the *GULO* evolutionary history within these lineages. Starting with the Araneae order, the six species analyzed are included in two superorders: Araneomorphae (*L. hesperus*, *P. tepidariorum*, *N. clavipes*, *S. mimosarum* and *L. reclusa*) and Mygalomorphae (*A. geniculata*). Inside the Araneomorphae, we have one species representing the Haplogynae (*L. reclusa*) and four species representing the known sister group, Entelegynae (*L. hesperus*, *P. tepidariorum*, *N. clavipes* and *S. mimosarum*). Concerning the Entelegynae, we analyzed three species from the Orbiculariae group (*L. hesperus*, *P. tepidariorum* and *N. clavipes*) and one from the Eresoidea group (*S. mimosarum*). We were able to annotate a putative *GULO* in *L. hesperus* and *N. clavipes*, but also to refine the annotation already available for *P. tepidariorum*, coincidently the three representative species for the Orbiculariae group in this work. As such, we can assume that the *GULO* gene is maintained at least in the Orbiculariae species lineage, specifically in the Theridiidae (*L. hesperus* and *P. tepidariorum*) and Nephilidae (*N. clavipes*) families. Although we did not perform annotations for *S. mimosarum*, *L. reclusa* and *A. geniculata*, we cannot rule out the presence of a *GULO* gene in these species genomes. For *S. mimosarum* and *A. geniculata*, we were able to find potential exonic sequences dispersed across several scaffolds, including some that may code for a protein with the typical HWAK amino acid pattern, while for *L. reclusa* we could not obtain a reliable tblastn alignment due to the bad quality of the genome. According to these evidences, it is likely that with better genome sequencing and assembly we can in fact perform a complete putative *GULO* annotation for these species. The summarized results for the Araneae group analysis can be observed in Figure 5.

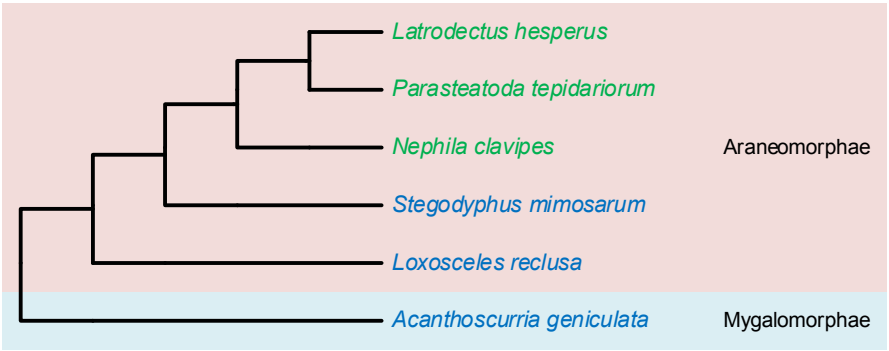


Figure 5 – Cladogram representation of the findings regarding the annotation of *GULO* in six Araneae species. The species in which *GULO* is likely present are highlighted in green, while the species where the presence or absence of *GULO* cannot be inferred are highlighted in blue. The pink and light blue regions differentiate the Araneae species analyzed into two superorders, respectively Araneomorphae and Mygalomorphae. The cladogram branches represent the taxonomic relation between the species analyzed, depicted as in the Tree of life web project (http://tolweb.org/tree/) and in Wheeler *et al.* (2015).

The Acari species analyzed are representative of two Acari superorders, namely Acariformes, and Parasitiformes. In the NCBI database, we were able to obtain genomes from seven Parasitiformes species (*I. scapularis*, *I. ricinus*, *R. microplus*, *G. occidentalis*, *V. destructor*,

V. jacobsoni and *T. mercedesae*) and nine Acariformes species (*A. coleoptrata*, *P. peltifer*, *S. magnus*, *H. rufulus*, *D. farinae*, *D. pteronyssinus*, *E. maynei*, *S. scabiei* and *T. urticae*). Within the Parasitiformes, we were able to annotate a putative *GULO* gene in two species belonging to the Ixodidae family included in the Ixodida order (*I. scapularis* and *R. microplus*), although with further phylogenetic analysis we uncovered that the *R. microplus* annotation obtained may in fact be the result of genome contamination. Curiously, the *I. scapularis* annotation also shows irregular placement in the obtained phylogenies for this dataset, when observing the taxonomic context. This is probably due to the complexity of our dataset regarding the greatly divergent species analyzed rather than a problem with our annotation, and as such, we consider *GULO* to be present in this species. This consideration is supported by the results of Wheeler *et al.* (2015), where *I. scapularis* was presented as a species with a identifiable *GULO* gene, although no details are given by these authors. Additionally, the *I. ricinus* species (also included in the Ixodida order) genome had features of a possible *GULO* gene presence, however the tblastn alignment was scattered across many genomic scaffolds. Using our criteria, we cannot infer the presence or absence of *GULO* in *I. ricinus*, but given the results obtained for *I. scapularis*, it is possible to extrapolate that a *GULO* gene may be present in the Ixodida order, specifically in the Ixodinae subfamily in which these two species are included. We could not annotate a putative *GULO* gene in the four remaining Parasitiformes species (*G. occidentalis*, *V. destructor*, *V. jacobsoni* and *T. mercedesae*) belonging to the Mesostigmata order (Gamasina infraorder), due to the lack of sequence homology in the performed tblastn search. This result presents strong evidence regarding the possible loss of *GULO* in the Mesostigmata lineage. Concerning the Acariformes, the nine analyzed species can be included in two sister orders, namely Trombidiformes (*T. urticae*) and Sarcoptiformes (*A. coleoptrata*, *P. peltifer*, *S. magnus*, *H. rufulus*, *D. farinae*, *D. pteronyssinus*, *E. maynei* and *S. scabiei*). Regarding the Trombidiformes, we did not annotate a putative *GULO* for *T. urticae* since no exonic sequence coded for a conserved HWXK amino acid motif. Nevertheless, since only one species was analyzed, we cannot exclude the presence of *GULO* in this lineage. As for the Sarcoptiformes, the eight species analyzed are representative of the Astigmata (*D. farinae*, *D. pteronyssinus*, *E. maynei* and *S. scabiei*) and Oribatida (*A. coleoptrata*, *P. peltifer*, *S. magnus* and *H. rufulus*) suborders. Within the Astigmata, we were able to annotate a putative *GULO* for *D. farinae* and *E. maynei*, species that belong to the Pyroglyphidae family. However, we were not able to obtain an annotation for another species of this family (*D. pteronyssinus*), due to the presence of a single in-frame stop codon in the expected CDS. It is surprising to find species of the same genus with opposite annotation results. Even so, several reports show that although *D. farinae* and *D. pteronyssinus* (house dust mites) have a worldwide geographical distribution, and sometimes even share the same microenvironment (*e.g.* carpets), they have distinct optimal growth conditions (Thomas 2010). It is rather common to see an alternation between the most predominant *Dermatophagoides* species present across several

geographical locations, and conventionally *D. farinae* is known as the American house dust mite and *D. pteronyssinus* as the European house dust mite due to the effective population size of each species in these regions (Thomas 2010, Liu *et al.* 2018). Perhaps the observed species regionalization pattern, which seems to derive from different growth conditions needs, can account for the results obtained. Since distinct selective pressures are present in the preferable environments of each of these species, we can extrapolate that they may condition the putative *GULO* gene evolutionary path by enhancing a purifying selection mechanism for *D. farinae* (leading to *GULO* conservation), while for *D. pteronyssinus* they do not exert an effect and allow a *GULO* pseudogeneization event. As for *S. scabiei*, the absence of a valid putative *GULO* annotation is surprising. We could not obtain any homologous sequence hit through the tblastn search against this species genome, even though we found potential *GULO* sequences in the remaining seven Sarcoptiformes species. This is unexpected since even if *GULO* was lost in the *S. scabiei* lineage, we should theoretically observe partial remnants of the gene undergoing pseudogenization using a tblastn approach, given that putative *GULO* sequences were easily found in closely related species and usually the accumulation of genomic mutations leading to loss-of-function is a rather slow process. As such, considering these observations, we contemplate that an annotation could not be performed for *S. scabiei* likely due to a technical issue that did not allow the genomic region in which *GULO* is contained to be sequenced. Alternatively, the gene could have been physically removed because of unequal crossing over during meiosis or even due to the mobilization of a transposable element. Given these results, we cannot extrapolate the presence or absence of *GULO* in this organism. Regarding the Oribatida group, we were able to perform annotations for the four analyzed species, which appear in the correct phylogenetic context in the consensus phylogeny. The summarized results for the Acari group analysis can be observed in Figure 6.

Based on the cladogram, we can infer that *GULO* is present in the Acariformes and Parasitiformes lineages. Within the Parasitiformes, *GULO* appears to have been lost in the Mesostigmata order and maintained in the Ixodida order, while in the Acariformes, *GULO* seems to be present in the Sarcoptiformes order and may be present or absent in the Trombidiformes order.

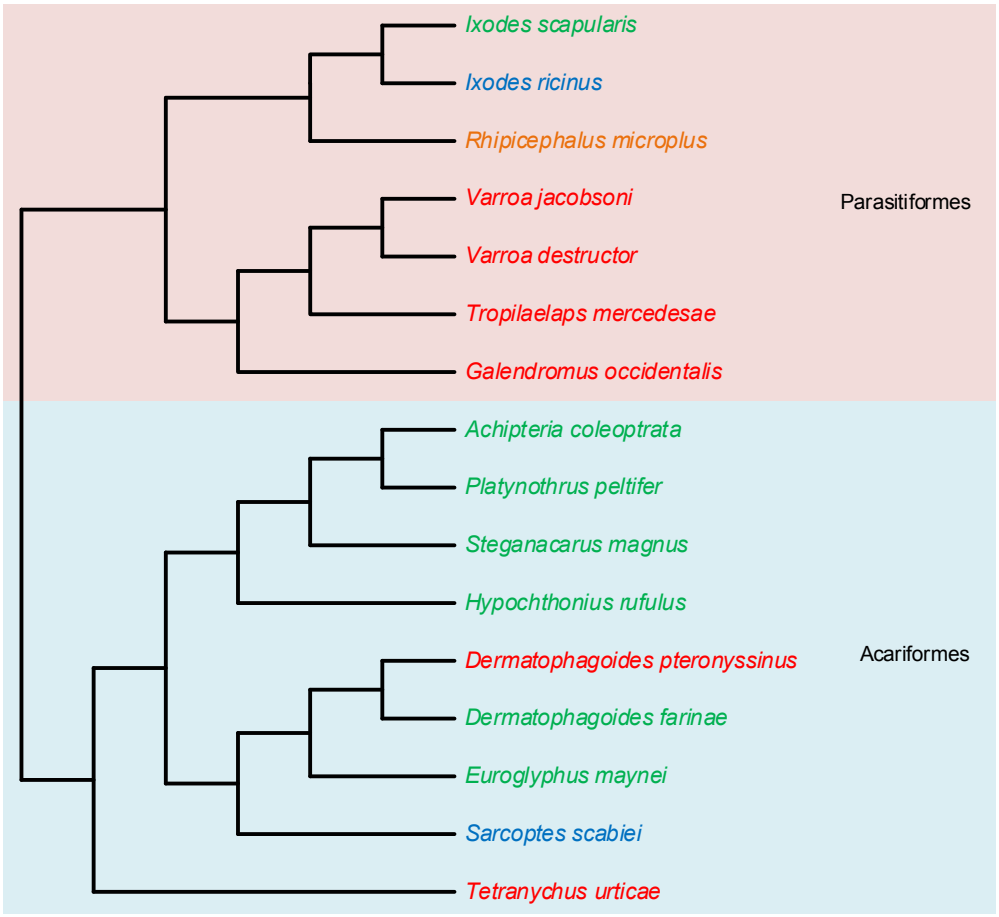


Figure 6 - Cladogram representation of the findings regarding the annotation of *GULO* in 16 Acari species. The species in which *GULO* is likely present are highlighted in green, the species where the presence or absence of *GULO* cannot be inferred are highlighted in blue, the species in which the annotations may be the result of genome contamination are highlighted in orange and the species where *GULO* is probably absent are highlighted in red. The pink and light blue regions differentiate the Acari species analyzed into two superorders, respectively Parasitiformes and Acariformes. The cladogram branches represent the taxonomic relationship between the species analyzed, depicted as in the Tree of life web project (http://tolweb.org/tree/), Black *et al.* (1997), Liana and Witaliński (2005), Domes *et al.* (2007) and Dowling and OConnor (2010).

Regarding the Mollusca phylum, we analyzed 18 species that represent three separate taxonomic classes, namely the Gastropoda, the Bivalvia and the Cephalopoda. Remarkably, we could not annotate a single putative *GULO* gene in any of the ten bivalve species genomes downloaded (*B. setacea*, *B. platifrons*, *M. philippinarum, M. galloprovincialis*, *C. virginica*, *P. martensii*, *C. gigas*, *C. fluminea, M. yessoensis* and *D. polymorpha*), due to lack of homology or identifiable HWXK amino acid motif encoded by the potential exons found. This is also the case for the only Cephalopoda species evaluated in this work (*O. bimaculoides*), since there were no homologous sequences to the *M. musculus* GULO found when performing the tblastn search. In addition, we were able to annotate a putative *GULO* in three Gastropoda species genomes (*C. reticulata*, *L. stagnalis* and *R. auricularia*) and confirm the already available *GULO* CDS for *A. californica* and *L. gigantea*. For the remaining two Gastropoda species, *C. tribblei* and *B. glabrata*, no annotations were performed due to the spread of putative exonic sequences across

several scaffolds or in-frame stop codons in the CDS, respectively. The results for the Molusca group are summarized in Figure 7.

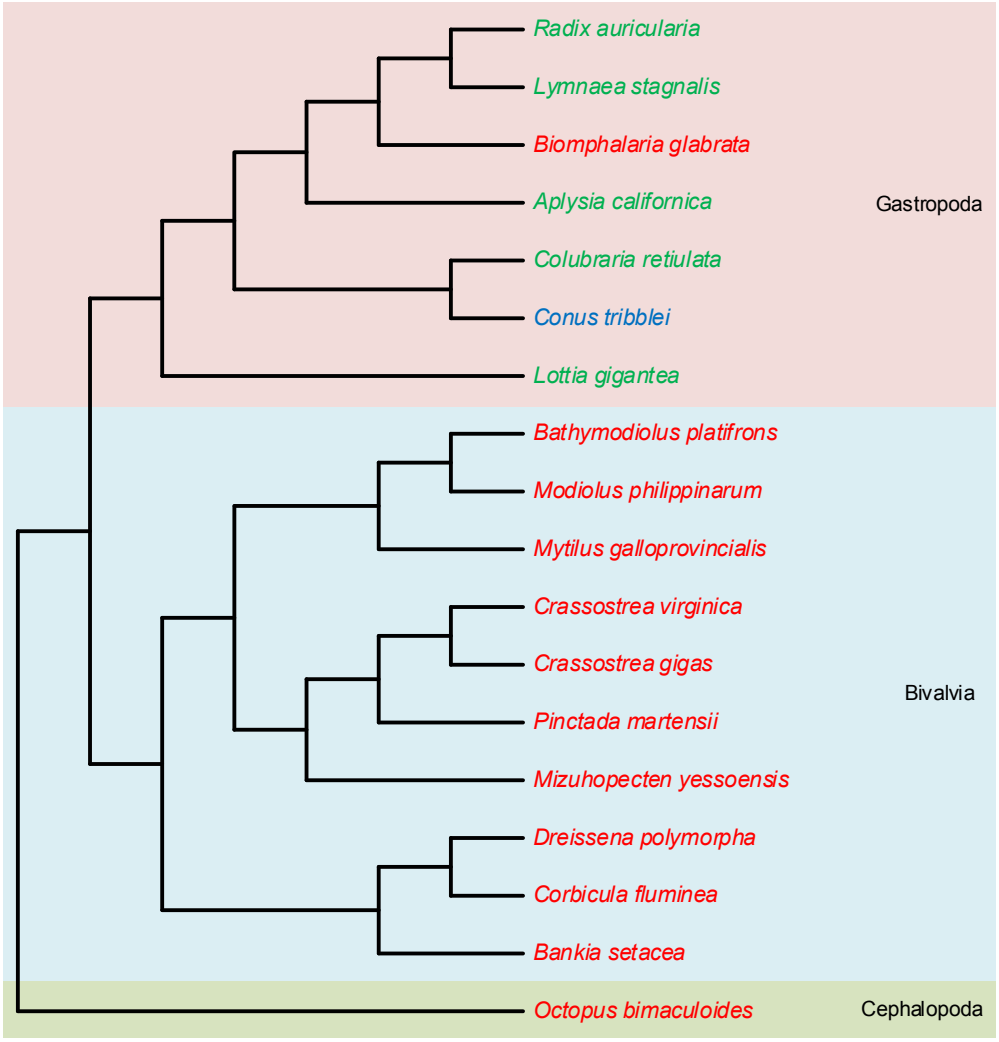


Figure 7 - Cladogram representation of the findings regarding the annotation of *GULO* in 18 Mollusca species. The species in which *GULO* is likely present are highlighted in green, the species where the presence or absence of *GULO* cannot be inferred are highlighted in blue and the species where *GULO* is probably absent are highlighted in red. The pink, light blue and olive green regions differentiate the Mollusca species analyzed into three classes, respectively Gastropoda, Bivalvia and Cephalopoda. The cladogram branches represent the taxonomic relationship between the species analyzed, depicted as in the Tree of life web project (http://tolweb.org/tree/), Taylor *et al.* (2007), Plazzi *et al.* (2011), Zapata *et al.* (2014) and Liu *et al.* (2018).

It is likely that *GULO* was lost in the Bivalvia class but was present in the common ancestral of both Bivalvia and Gastropoda taxonomic groups, given that this gene is maintained in the majority of the gastropods analyzed. Inside the Gastropoda group, *GULO* may have been recently lost in *B. glabrata*, a member of the Planorboidea superfamily included in one of the Gastropoda subclasses, Heterobranchia, since it appears present in two closely related species, *L. stagnalis* and *R. auricularia*. Regarding the Cephalopoda group, we cannot infer the presence or absence of *GULO* since only one representative genome was available for inquiry, belonging to *O. bimaculoides*. Nevertheless, *GULO* seems to not be present at least in this species.

Using non-annotated genomes, we were able to infer the presence of *GULO* in the Placozoa, Myxozoa, Acari and Annelida taxonomic groups and extrapolate the absence of *GULO* in the Hexapoda (w/o Insects) rank, going in accordance and complementing some results of Wheeler *et al.* (2015). Moreover, we also reinforced the evidence of *GULO* presence in the Gastropoda and Araneae groups. In general, it seems that the presence of a *GULO* gene is actually common across the protostomian and non-bilaterian species. Nevertheless, *GULO* appears to have been lost independently in four relatively large taxonomic groups, namely the Pancrustacea, Nematoda, Platyhelminthes and Bivalvia. The refined cladogram that considers the new findings regarding the analysis of non-annotated genomes is represented in Figure 8.

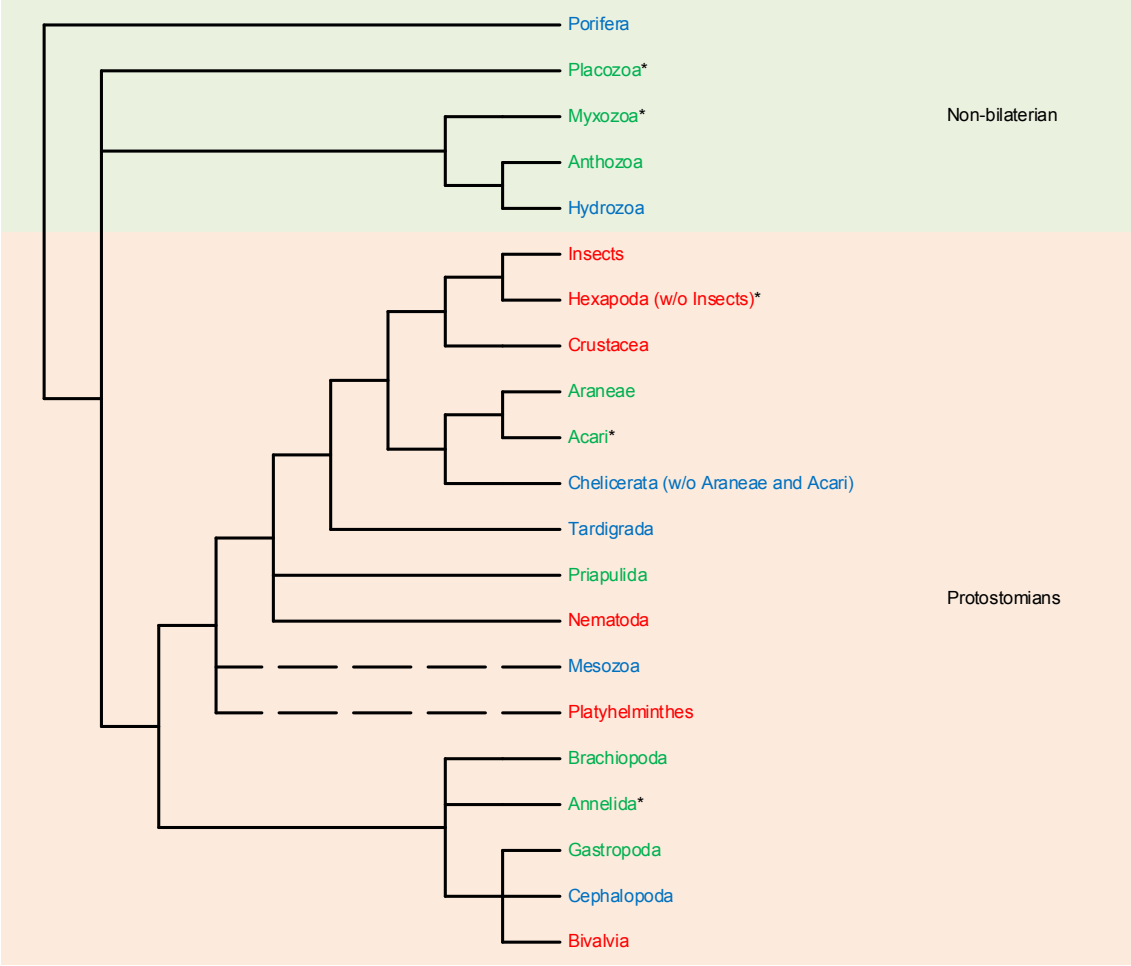


Figure 8 - Summary of the findings regarding the presence of putative functional *GULO* genes in protostomian and non-bilaterian animal lineages. In green, red and blue, are, respectively, the lineages where a likely functional *GULO* has been detected, lineages where a functional *GULO* gene has not been detected and lineages for which there is indecisive data and thus no firm conclusions can be made. The new findings concerning the analysis of non-annotated genomes are marked with an "*". Broken lines show uncertain relationships. Taxonomic relationships are depicted as in the Tree of life web project (http://tolweb.org/tree/).

III.3. Microbiome and HPLC assays

Ascorbic acid can be detected in a great number of insects, including *D. melanogaster*, and the current consensus regarding the source of this nutrient in these species lies on direct absorption from their available diet. This hypothesis goes in accordance with several results that suggest that *GULO* was lost in the insect lineage (Massie *et al.* 1991, Barbehenn *et al.* 2001, López-Fernández *et al.* 2018). Nevertheless, we know that *D. melanogaster* can be reared in an ascorbic acid depleted medium for many generations. Therefore, if *D. melanogaster* obtains ascorbic acid from the available diet, we should not be able to detect this vitamin in flies reared under standard conditions. However, our results demonstrate that this is not the case (Figures 9, 10, 11 and 12). Moreover, an interesting report suggests that the invertebrate *C. elegans* is able to synthesize ascorbic acid even in the absence of a functional *GULO* gene (Patananan *et al.* 2015). Given these evidences, we were able to formulate two distinct hypotheses: i) the *D. melanogaster*'s gut microbiome may be fully or at least partially responsible for the ascorbic acid synthesis, or ii) *D. melanogaster* may inherently be able to synthetize this vitamin either using a GULO-alternative final step enzyme or a completely alternative metabolic pathway.

To address the first hypothesis, we need to compare ascorbic acid levels in flies without gut microbiome (axenic) and control flies. However, contrary to control flies, male and female axenic flies could not be separated after hatch to ensure that they were kept under sterile conditions until collection. This represented a relevant technical issue, given that reports show that mated female flies commonly alter their diet regime toward specific nutrients, such as amino acids, and this could lead to biased results (Ribeiro and Dickson 2010, Uchizono *et al.* 2017). To overcome this problem, we also performed ascorbic acid measurements using control male and female flies kept together during a seven-day period seeking to verify if these conditions led to different results. Furthermore, to ensure that the axenic flies used were depleted of bacterial contamination from exogenous sources that could influence the results, flies were collected in parallel of the HPLC assay, homogenized and plated in standard LB (Luria-Bertani) solid medium. 7-day flies with microbiome were also used in the same conditions to ensure a positive growth control.

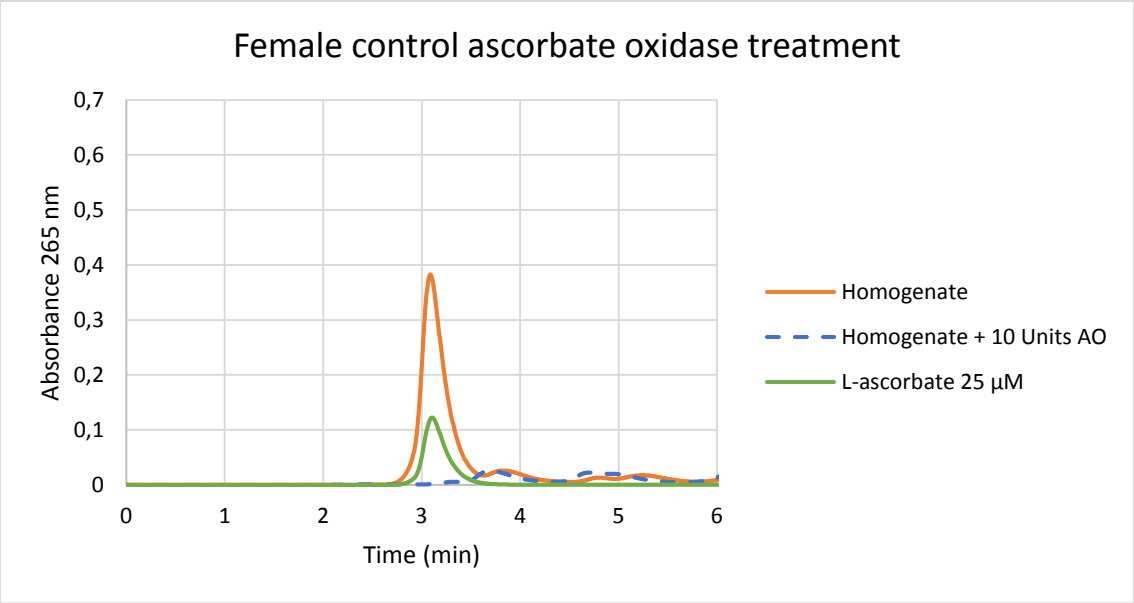


Figure 9 – Graphical portrayal of the data obtained from the HPLC regarding the female biological sample used as reference for the elution time and chromatogram peak region of ascorbic acid. The orange line represents a control homogenate technical replica, the blue dash line represents a control homogenate technical replica with ascorbate oxidase treatment and the green line represents a 25 µM ascorbic acid standard solution.

The measurement results of ascorbic acid levels in seven-day axenic flies (mated), seven-day control flies (virgin) and seven-day control flies (mated) are shown in Figure 10.

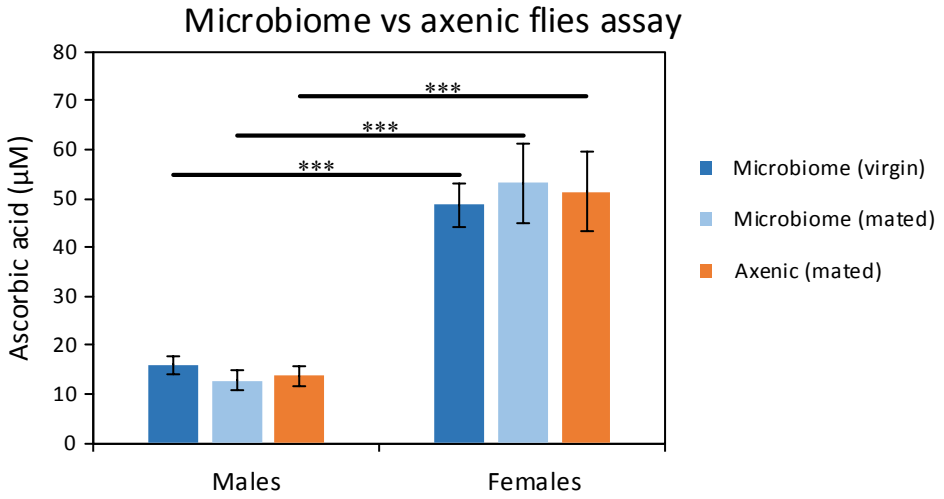


Figure 10 – Ascorbic acid levels (µM) of seven-day axenic and control flies. The graph displays separate columns for male and female individuals, with control virgin flies represented in blue, control mated flies in light blue and axenic mated flies in orange. No significant statistical differences were found (*** corresponds to P≤0.001) within the male and female flies groups for the three evaluated conditions, suggesting that the microbiome does not contribute for the synthesis of this nutrient. Nevertheless, there are significant statistical differences when comparing male and female values for the same experimental conditions, with females showing values around three fold higher than males.

No significantly different ascorbic acid levels were seen between matted, virgin or axenic individuals within the same gender, strongly suggesting that the microbiome has minimal or null influence in the synthesis of this vitamin. Moreover, we performed a microbiome ex-vivo assay,

using a homogenate obtained from 25 whole individuals without environmental microbial contaminants to inoculate MRS liquid medium. The MRS media is known to allow the propagation of several bacteria genera that are likely involved in ascorbic acid synthesis, such as *Gluconacetobacter, Gluconobacter* and *Acetobacter*, and as such was an ideal choice for the experimental approach (Bremus et al 2006, Newell *et al.* 2014, Simhadri *et al.* 2017). Nevertheless, no ascorbic acid was detected in the supernatant or the bacterial pellet after one, two and three days.

Another possible observation regarding the results shown in Figure 10 concerns the unexpected difference in ascorbic acid levels (μM) exhibited by male (16.0 ± 1.8) and female (48.6 ± 4.4) flies. It is known that female *D. melanogaster* individuals are normally bigger than their male counterparts, a consideration confirmed by our own measurements, in which we determined that the body weight (wet weight) of our female flies is approximately 50% higher than the males body weight. Nevertheless, that weight difference alone cannot explain the 3-fold higher ascorbic acid levels measured in female flies when compared to males. This evidence seems to reinforce the hypothesis that *D. melanogaster* might be able to synthesise ascorbic acid, since flies with the same microbiome and sharing the same environmental conditions should not likely produce this vitamin in such distinct amounts, suggesting possible gender-specific differences regarding molecular regulatory mechanisms or physiological processes. Interestingly, in *M. musculus* females, the plasma ascorbic acid levels were also found to be higher when compared to males. This difference was primarily attributed to discrepant ascorbate excretion rates due to sexual dimorphism of the renal tubule. As such, it seems plausible that the different ascorbic acid levels in male and female flies could for example be related with gender-specific traits regarding ascorbic acid excretion rates across the *D. melanogaster* Malpighian tubule system.

Seeking to reinforce the intrinsic *D. melanogaster* ascorbic acid synthesis hypothesis, we decided to follow up on the evidence found by Massie *et al.* (1991), in which ascorbic acid levels of male *D. melanogaster* Oregon-R flies were shown to increase after a brief cold shock of 10 min at 4 °C. For this purpose, we performed two distinct cold exposure protocols, namely cold shock and cold acclimation, to check if an equal response occurred. The cold acclimation protocol consisted in measuring ascorbic acid levels in 7-days-old male virgin flies (control) and 7-days-old male virgin flies subjected to 15 °C for one day with or without one day of post-recovery at 25ºC, and the obtained results are presented in Figure 11.

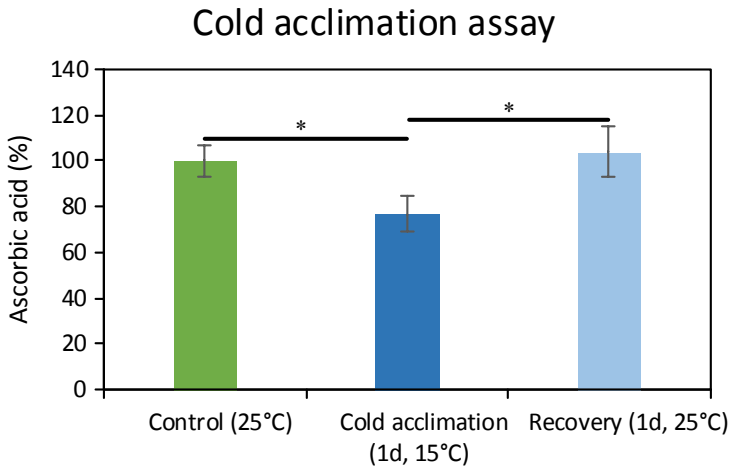


Figure 11 – Seven-day male *D. melanogaster* Oregon-R flies ascorbic acid levels in control (green), cold acclimation (blue) and recovery (light blue) conditions. The levels are represented in percentage relative to the control used (100%). After one day of cold acclimation, the flies ascorbic acid levels significantly decrease (* corresponds to $P \leq 0.05$), returning to control values after allowing one day of recovery at 25ºC.

In these conditions, the ascorbic acid levels decreased 23 % in flies exposed one day to 15ºC when compared to control flies, although the level recovers to control values after one day at 25ºC. These results allow us to infer putative ascorbic acid synthesis, or alternatively, the reduction of dehydroascorbic acid back to ascorbic acid due to enzymatic recycling pathways. Knowing that ascorbic acid levels can be easily detected in *D. melanogaster*, but also that the microbiome is not likely involved in ascorbic acid synthesis and that this vitamin is not obtained through the provided food source, the evidence seems to indicate probable *de novo* ascorbic acid production in the flies because no alternative source is evident. As such, it seems more parsimonious to assume that *D. melanogaster* is primarily producing ascorbic acid in the recovery phase after the cold acclimation stress.

In the cold shock protocol, 7-days-old male virgin flies were subjected to 4 °C for four hours and collected after 30 minutes, two hours, and 48 hours of recovery at 25ºC for ascorbic acid level measurements. The corresponding experimental results can be seen in Figure 12.

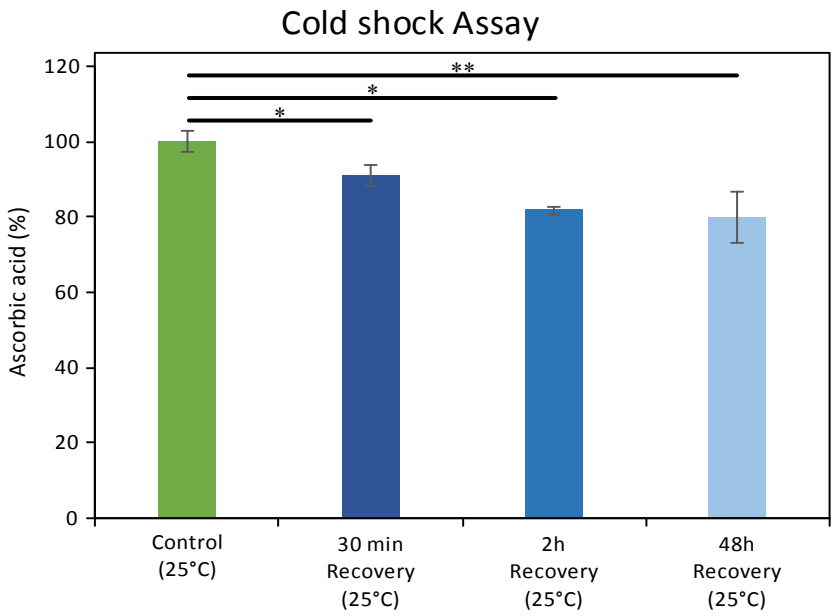


Figure 12 - Seven-day male *D. melanogaster* Oregon-R flies ascorbic acid levels in control (green) and after 30 minutes, two hours, or 48 hours of cold shock recovery conditions (ranging from darker to lighter blue, respectively). The levels are presented in percentage relative to the control used (100%). After exposure to cold shock, the flies ascorbic acid levels significantly decrease at all the measured recovery stages (* corresponds to P≤0.05 and ** to P≤0.01), when compared to control values.

In these conditions, the ascorbic acid levels decreased 9 % in flies with 30 min of recovery when compared to the control. When we extended the recovery period to two and 48 hours, the ascorbic acid levels decreased even more to values of 18 and 19 %, respectively, when compared to the control.

It is known that cold shock stress and cold acclimation stress conditions originate two distinct phenotypic reactions in the cell. The first condition is the most aggressive to the cell and usually leads to mechanical membrane damage due to ice crystals formation, accumulation of reactive oxygen species that contributes to protein/DNA denaturation, and drastic variations of osmotic potential, all possible inductors of cell death (Panoff *et al.* 1997, Amato and Christner 2009). Transcriptional and translational alterations that allow for the expression of isozymes and cold shock proteins, which in turn stabilize affected nucleic acids and proteins, commonly characterize the cold acclimation stress response (Phadtare *et al.* 1999, Hoyoux *et al.* 2004). It was previously observed that the activity of enzymes from the oxidative pathway was increased in response to cold acclimation in ectotherm species (Guderley 2004). Since mitochondrial activity is known to potentiate reactive oxygen species production, the antioxidant activity of ascorbic acid emerges as an important non-enzymatic scavenger mechanism of these molecules (Sastre *et al.* 2000). In fact, previous studies have correlated the reduced expression of the human ascorbic acid transporter SVCT2 with a reduced ability of neurons and platelets to cope with oxidative stress, implicating vitamin C in the direct protection against this stress (Qiu *et al.* 2007, Savini *et al.* 2007). With increased mitochondrial activity, we can deduce that the adenosine

triphosphate (ATP) levels should also rise in the cell. In fact, metabolomics analysis of cold-acclimated flies showed an increase of adenosine triphosphate (ATP) levels, and other studies reported that cold stress leads to the increase of ATP levels in cricket muscles and beetle larvae (Colinet 2011, MacMillan *et al.* 2012, MacMillan *et al.* 2016). Given these reports, it was suggested that insects could increase ATP levels during cold acclimation as a defence mechanism to handle a possible forthcoming more acute stress, since the extra ATP produced could later be allocated to the recovery process. From our point of view, this mechanism can explain the results obtained for our cold acclimated flies. Following this interpretation and knowing that cold shock and cold acclimation correspond to different stress stimuli, the contrasting results for our cold shock experiment can be related to very distinct physiological responses of the cell. Coincidently, studies performed in the flesh fly (*Sarcophaga crassipalpis*) showed that ATP levels of flies maintained at 0 °C for 20 days are significantly lower than the ATP levels detected in flies that received a 24 hour warming pulse (15ºC or 20ºC) at the 10th day of cold exposure (Dollo *et al.* 2010). This report suggests that the regeneration of ATP reserves needed for the reduction of chilling injuries and subsequent promotion of organism survival is allowed through brief warming pulses during the cold exposure (Dollo *et al.* 2010). This evidence suggests that, contrary to the cold acclimation conditions, the cold shock exposure does not inherently allow for the creation of ATP reserves needed for an effective recovery within our defined 30 min, two hour and two day time periods, which could be related to the absence of ascorbic acid replenishment detected.

III.4. *SVCT* phylogenies

In the previous chapters, we show possible synthesis of ascorbic acid in *D. melanogaster* even in the absence of *GULO*. Knowing that in deuterostomians ascorbic acid homeostasis is maintained by transporter proteins designated as SVCTs (Lindblad *et al.* 2013), we sought out to understand if the SVCT transporter known to exist in protostomians could eventually play the same role. Our phylogenetic approach regarding these transporters sought out to uncover any correlation between the evolutionary paths of the *SVCT1*, *SVCT2*, *SVCT3*, *SVCT4* genes and the protostomian *SVCT*, which we designated as *SVCTP*. With this information, we expected to deduce substrate specificity characteristics of these transporters and extrapolate if ascorbic acid transport could be an ancestral trait or a product of gene neofunctionalization after events of WGD in vertebrates.

The *SVCT1* consensus tree (Supplementary figure 2) shows that the *H. sapiens SVCT2* sequence (NP_005107.4) is phylogenetically closer to the animal *SVCT1* sequences out of all the used outgroups. Regarding the Chondrichthyes class, we are able to detect a *SVCT1* gene in both species analyzed, belonging to the Elasmobranchii (*R. typus*) and Holocephali (*Callorhinchus milii*) subclasses. Concerning the Actinopteri class, the *SVCT1* gene was not found in *Fundulus heteroclitus*, *Labrus bergylta*, *L. oculatus*, *Nothobranchius furzeri* and *Notothenia coriiceps*. *F. heteroclitus*, *L. bergylta*, *N. furzeri* and *N. coriiceps* are the only representatives of the Fundulidae, Labridae, Nothobranchiidae and Nototheniidae families (all Teleostei), while *L. oculatus* is the only representative of the Lepisosteidae family (Holostei), respectively. Since these five families are represented by only one species each, we cannot assume that the *SVCT1* gene was lost in these lineages, because there is a high chance that incomplete genome sequencing/annotation or wrong annotation of the gene could lead to the result we observed. This conclusion is further reinforced by the detection of the *SVCT1* gene in the remaining 40 Actinopteri species studied, since it is highly unlikely that five independent gene losses occurred when the vast majority of species still maintain a putative functional transporter. Interestingly, we also observed that *SVCT1* is duplicated in *Salmo salar* and *Cyprinus carpio*, belonging to the Salmonidae and Cyprinidae families, respectively. This duplication event can be correlated with the WGD events known to affect specifically the salmonid and some cyprinid lineages (Glasauer and Neuhauss 2014). We observed a *SVCT1* copy in *L. chalumnae*. Now regarding the four Amphibia species analyzed, we were able to detect the *SVCT1* gene in two members of the Pipidae family (*X. laevis* and *X. tropicalis*), but were unable to find this gene in the single representatives of the Dicroglossidae and Ranidae familes (*N. parkeri* and *Rana catesbeiana*, respectively). Within the Reptilia group, we were able to detect the *SVCT1* gene in 11 species out of 15, but not in *Thamnophis sirtalis*, *O. hannah*, *G. japonicus* and *Pelodiscus sinensis*. These species are the single representatives of four reptile families, respectively Colubridae, Elapidae, Gekkonidae and Trionychidae, and as such,

we cannot extrapolate if *SVCT1* was lost in these lineages, similarly to the Actinopteri cases mentioned. When observing the species representative of the Aves class, we verify that the *SVCT1* gene was only detected in 17 out of 67 genomes. From the 50 species in which we do not detect the *SVCT1* gene, only in *Aquila chrysaetos canadensis*, *Haliaeetus albicilla* and *Haliaeetus leucocephalus* we found enough evidence of possible gene loss, since they belong to the same taxonomic family (Accipitridae) and it is highly unlikely that a genomic information gap could be present consistently in the three genomes. Concerning the Mammalia class, we were able to detect the *SVCT1* gene in 99 species out of the 112 analyzed. Of the 13 excluded species, only four were solo representatives of their corresponding family, namely *Balaenoptera acutorostrata scammoni* (Balaenopteridae), *Castor canadensis* (Castoridae), *Nomascus leucogeny* (Hylobatidae) and *Carlito syrichta* (Tarssiidae). As such, as seen for the Actinopteri and Reptilia, although we were not able to find a *SVCT1* gene in these species, we cannot infer if the gene was lost in their lineages with the limited information gathered. Nevertheless, given the overwhelming number of species in which we detected a putative functional *SVCT1*, this appears to be the result of bad gene annotation or incomplete genome sequencing and not a gene loss event.

The *SVCT2* consensus tree (Supplementary figure 3) shows that the *H. sapiens SVCT1* sequence (XP_011542068.1) is phylogenetically closer to the animal *SVCT2* sequences out of all the used outgroups, as expected from the previous *SVCT1* results. No *SVCT2* gene was found in the two analyzed species from the Chondrichthyes class, namely *R. typus* and *C. milii*, but since they are the only representatives of the Elasmobranchii and Holocephali subclasses, respectively, no conclusions regarding possible gene loss were performed. Within Actinopteri, eleven species that were shown to contain the *SVCT1* gene do not seem to possess the *SVCT2* gene. They are almost exclusively single representatives of their corresponding taxonomic families, such as *Seriola dumerili* (Carangidae), *Lates calcarifer* (Centropomidae), *Cynoglossus semilaevis* (Cynoglossidae), *Cyprinodon variegatus* (Cyprinodontidae), *Kryptolebias marmoratus* (Rivulidae) and *Larimichthys crocea* (Sciaenidae). Inside Poeciliidae, two (*Xiphophorus maculatus* and *Poecilia mexicana*) out of the five species that contained the *SVCT1* gene, may have lost the *SVCT2* gene. The opposite scenario is observed for *Sinocyclocheilus anshuiensis* (Cyprinidae), *L. oculatus* (Lepisosteidae), *N. coriiceps* (Nototheniidae), *H. burtoni* and *N. brichardi* (both contained in Cichlidae), species in which no *SVCT1* gene was identified but that seem to contain *SVCT2*. Interestingly, *F. heteroclitus* (Fundulidae), *L. bergylta* (Labridae) and *N. furzeri* (Nothobranchiidae) do not seem to possess any of the *SVCT1* and *SVCT2* genes associated with the ability to transport ascorbic acid. Several evidences for *SVCT2* gene duplications can be observed within the Actinopteri, specifically in the Salmonidae (*Oncorhynchus kisutch*, *Oncorhynchus mykiss* and *S. salar*), Cyprinidae (*Sinocyclocheilus anshuiens* and *Sinocyclocheilus rhinocerous*) and Osteoglossidae (*Scleropages formosus*) families. The

duplications in the Salmonidae and Cyprinidae species can again be explained by the WGD events that occurred in these specific family lineages. There is also an identifiable *SVCT2* duplication in *S. formosus*, that seems to be an independent local duplication. We observed a *SVCT2* copy in *L. chalumnae*. Now regarding the Amphibia group, we were able to identify the *SVCT2* gene in *X. laevis*, *X. tropicalis a*nd *N. parkeri*, but not in *R. catesbeiana*. The first two species had the *SVCT1* gene, whereas *N. parkeri* and *R. catesbeiana* did not seem to possess it. A duplication event seems to have occurred in *X. laevis* but not in *X. tropicalis*, a result that can be explained by the known specific *X. laevis* WGD that did not affect *X. tropicalis* (Sémon and Wolfe 2008). Within the Reptilia, we could observe that three species in which we did not identify the *SVCT1* gene have the *SVCT2* gene, namely *O. hannah*, *G. japonicus* and *P. sinensis*. Contrary to these cases, although we could detect the *SVCT1* gene in *A. carolinensis*, we were not able to observe the *SVCT2* gene in this species. Interestingly, *T. sirtalis* does not appear to have both the *SVCT1* and *SVCT2* genes. Concerning the Aves taxonomic group, the *SVCT2* is detected in 47 out of 67 species, a remarkably higher value than the one seen for the *SVCT1* gene. It is interesting to note that the *SVCT2* gene can be found in the species representing the single family for which we found evidence of possible *SVCT1* loss, *A. chrysaetos canadensis*, *H. albicilla* and *H. leucocephalus* (Accipitridae). Moreover, there seems to be species that may not have both the *SVCT1* or *SVCT2* genes, namely *Colius striatus* (Coliidae), *Merops nubicus* (Meropidae), *Mesitornis unicolor* (Mesitornithidae), *Zonotrichia albicollis* (Passerellidae), *Phaethon lepturus* (Phaethontidae), *Nestor notabilis*, *Amazona aestiva* (both Psittacidae), *Pterocles gutturalis* (Pteroclidae) and *Geospiza fortis* (Thraupidae). We were able to detect the *SVCT2* gene in 105 out of the 112 Mammalia species analyzed, verifying that only two mammal families (Manidae and Dasypodidae) are missing due to the exclusion of their only representatives, *Manis javanica* and *Dasypus novemcinctus*, respectively. Although we cannot assert possible *SVCT2* absence or presence in these specific lineages, it is very likely that the gene may not be well annotated in these species, or alternatively, contained in a genome region that could not be sequenced or assembled, since the vast majority of mammal species have this gene.

The *SVCT3* consensus tree (Supplementary figure 4) shows that this gene does not appear to be closely related with the *SVCT1*, *SVCT2* and *SVCT4* genes. This somewhat distant phylogenetic relation with *SVCT1* and *SVCT2* was expected, since Bürzle *et al.* (2013) has shown that *SVCT3* appears to have diverged from these genes at an early evolutionary stage. Nevertheless, knowing that the SVCT3 transporter is proposed as the replacement of SVCT4 function in species were *SVCT4* is pseudogenized (Yamamoto *et al.* 2010, Bürzle *et al.* 2013), it was expected that these genes would be phylogenetically closer than what is shown in our results. Even so, in our consensus phylogeny, the *SVCT3* representative sequences have remarkably big branch lengths, which are indicative of great genetic divergence between the *SVCT3* gene and the

used outgroups. These results go in accordance with what is reported by Kourkoulou *et al.* (2018), where even using several phylogenetic prediction models, the *SVCT3* sequences appear to be unstable in the consensus tree. It is known that genes can have distinct rates of evolution correlated with, for example, different expression levels in the organism (Wolf *et al.* 2009). This can possibly explain the observed results for the *SVCT3* gene, since an accelerated rate of evolution by comparison to the other *SVCT* representatives could result in a greater accumulation of nucleotide substitutions, and therefore, higher divergence levels. Although we cannot extrapolate a reasonable explanation for this phenomenon, this hypothesis is compatible with the idea in which *SVCT3* and *SVCT4* derive from a single ancestral lineage, although they are considered to be orphan genes (Bürzle *et al.* 2013, Nualart *et al.* 2014). Contrary to the results observed for the *SVCT1* and *SVCT2* genes, the *SVCT3* gene does not seem to be duplicated in any taxonomic group. In the Chondrichthyes class, we are able to detect the *SVCT3* gene in *C. milii*, while for *R. typus* no sequence is represented. Remarkably, many Actinopteri species seem to have lost the *SVCT3* gene, since only eight species out of the 45 analyzed are present in the final consensus phylogeny. These results support the hypothesis of Kourkoulou *et al.* (2018), in which he extrapolates that a notable *SVCT3* loss event happened in the very disparate Acanthomorpha taxonomic group. The most representative cases of possible *SVCT3* loss are seen within the Cichlidae and Salmonidae families, since five Cichlidae species (*H. burtoni*, *N. brichardi*, *Maylandia zebra*, *Oreochromis niloticus* and *Pundamilia nyererei*) and three Salmonidae species (*O. kisutch*, *O. mykiss* and *S. salar*) do not seem to possess this gene. Furthermore, the *SVCT3* gene was not detected in *F. heteroclitus* (Fundulidae), *L. bergylta* (Labridae) and *N. furzeri* (Nothobranchiidae), species that already did not seem to possess *SVCT1* and *SVCT2*. We observed a *SVCT3* copy in *L. chalumnae.* Regarding the Amphibia class, we were able to detect the *SVCT3* gene in *X. laevis*, *X. tropicalis* and *N. parkeri*, while for *R. catesbeiana* no sequence is presented in the final dataset. Inside the Reptilia group, we observed that of the 15 species analyzed, *Chelonia mydas* (Cheloniidae), *Crocodylus porosus* (Crocodylidae), *Anolis carolinensis* (Dactyloidae), *O. hannah* (Elapidae), *Python bivittatus* (Pythonidae) and *P. sinensis* (Trionychidae) may not have the *SVCT3* gene. Notably, the *SVCT3* gene appears to be present in *T. sirtalis* (Colubridae), while the *SVCT1* and *SVCT2* are not. Within the Aves, we detected the *SVCT3* gene in 36 out of the analyzed 67 species. We can observe that the two representative species from the Estrildidae family (*Lonchura striata domestica* and *Taeniopygia guttata*) possess the *SVCT1* and *SVCT2* genes, but do not seem to have the *SVCT3* gene. In addition, six species without identifiable *SVCT1* and *SVCT2* genes appear to possess the *SVCT3* gene, namely *M. nubicus* (Meropidae), *M. unicolor* (Mesitornithidae), *Z. albicollis* (Passerellidae), *N. notabilis* (Psittacidae), *P. gutturalis* (Pteroclidae) and *G. fortis* (Thraupidae). Nevertheless, none of the *SVCT1*, *SVCT2* or *SVCT3* genes was detected in *C. striatus* (Coliidae), *P. lepturus* (Phaethontidae) and *A. aestiva* (Psittacidae). Concerning the Mammalia class, we were able to

detect the *SVCT3* gene in 92 of the total 112 species analyzed. Although they possess the *SVCT1* and *SVCT2* genes, *Chrysochloris asiatica* (Chrysochloridae), *Galeopterus variegatus* (Cynocephalidae), *Jaculus jaculus* (Dipodidae), *Loxodonta africana* (Elephantidae), *Hipposideros armiger* (Hipposideridae), *Oryctolagus cuniculus* (Leporidae), *Ochotona princeps* (Ochotonidae), *Sorex araneus* (Soricidae), *Sus scrofa* (Suidae) and *Tupaia chinensis* (Tupaiidae) do not have any representative sequence for the *SVCT3* gene in the consensus phylogeny. Coincidentally, *D. novemcinctus* (Dasypodidae) appears to have only the *SVCT3* gene.

The *SVCT4* consensus tree (Supplementary Figure 5) indicates that this gene is more closely related to *SVCT1* and *SVCT2* than to the remaining outgroup genes used. Furthermore, the *SVCT4* gene appears to have been duplicated several times, mainly within the Actinopteri group. Curiously, there is also evidence for duplication in the mammal *Ceratotherium simum simum* (Rhinocerotidae). This duplication is unexpected, since in the Mammalia group there are no other identifiable duplications across a wide number of species. However, it may be the result of independent local gene duplication in this species. Indeed, a report by Lund and Sherman (1998) demonstrates that the *LGHbeta* gene is duplicated in *C. simum simum*, although it was only believed to be duplicated in primates at the time. To our knowledge, no descriptive articles are available regarding the genome of *C. simum simum*, but with these findings, it is possible that further studies concerning this species will reveal evidence of duplication events undisclosed until now. In Chondrichthyes class, we are able to detect the *SVCT4* gene in *R. typus* and *C. milii*. Regarding the Actinopteri group, the majority of species analyzed appear to possess the *SVCT4* gene, since only the Synbranchidae family out of 28 is excluded from the final phylogeny, due to the possible absence of *SVCT4* in *Monopterus albus*. Moreover, the *SVCT4* gene does not seem duplicated in *P. mexicana* (Poeciliidae), *Haplochromis burtoni* (Cichlidae), *L. calcarifer* (Centropomidae), *Paralichthys olivaceus* (Paralichthyidae), *Boleophthalmus pectinirostris* (Gobiidae), *Poecilia latipinna* (Poeciliidae), *Esox lucius* (Esocidae), *Sinocyclocheilus grahami, Danio rerio* (both Cyprinidae), *Pygocentrus nattereri* (Serrasalmidae), *Astyanax mexicanus* (Characidae), *Ictalurus punctatus* (Ictaluridae), *Clupea harengus* (Clupeidae) and *L. oculatus* (Lepisosteidae). In addition, it seems duplicated two times in *Takifugu rubripes* (Tetraodontidae), while for the remaining presented species, only once. We did not observe a *SVCT4* copy in *L. chalumnae.* Within the Amphibia, the *SVCT4* is only present in *X. laevis* and *N. parkeri*, whereas in *X. tropicalis* and *R. catesbeiana* it seems to be missing. As for the Reptilia class, three single representative species do not seem to possess *SVCT4*, namely *C. mydas* (Cheloniidae), *T. sirtalis* (Colubridae) and *O. hannah* (Elapidae), whereas the remaining reptile families are represented by at least one species. Forty-five Aves species within 37 taxonomic families have the *SVCT4* gene, while the single species representatives of ten families, namely Acanthisittidae, Ardeidae, Bucerotidae, Cariamidae, Eurypygidae, Otididae, Thraupidae, Tinamidae, Trogonidae and

Tytonidae are not represented in the final phylogeny. Concerning the Mammalia class, as expected and in agreement with results from Yamamoto *et al.* (2010), the majority of primates have lost the *SVCT4* gene. However, *C. syrichta* (Tarsiidae), *Propithecus coquereli* (Indriidae) and *Microcebus murinus* (Cheirogaleidae) still maintain the *SVCT4* gene. Although *C. syrichta* is present in the same suborder as for example *H. sapiens* and *Gorilla gorilla*, where *SVCT4* is known to be pseudogenized, it belongs to a distinct infraorder from these excluded primates, namely Tarsii. Given this evidence, it is likely that the *SVCT4* gene may have been lost at the split of the Tarsiiformes from the other Haplorrhini infraorder, Simiiformes. Curiously, five non-primate species who displayed identifiable *SVCT1*, *SVCT2* and *SVCT3* genes, respectively *Canis lupus* (Canidae), *Dipodomys ordii* (Heteromyidae), *Octodon degus* (Octodontidae) and *Nannospalax galili* (Spalacidae), do not appear to possess the *SVCT4* gene.

Following the refinement of the *SVCT4* dataset, the *SVCT5* file was created with 30 Actinopteri/Amphibia CDSs that did not appear phylogenetically close to any of the *SVCT* genes identified. The Bayesian consensus phylogeny obtained for this dataset can be observed in Supplementary Figure 6. From the species represented in the *SVCT5* phylogeny, only two Actinopteri (*M. albus* and *Neolamprologus brichardi*) and one Amphibia (*X. tropicalis*) species did not appear to possess the *SVCT4* gene, but it is possible that the gene may be present but not correctly annotated, or alternatively, contained in a genome information gap. Considering the premise above, one possible origin for these putative *SVCT5* gene sequences could be related to an ancestral duplication event that preceded the split of the Actinopteri and Amphibia taxonomic groups. Alternatively, two local and independent *SVCT4* duplication events at the base of the Actinopteri and Amphibia taxonomic groups could originate two distinct genes similar to *SVCT4*, designated by us as *SVCT5* and *SVCT6*. Unfortunately, with the abcense of the only representative species that could provide evidence of one or another hypothesis given the intermediate taxonomic position relative to both groups (the coelacanth *L. chalumnae*), we cannot determine the most likely scenario. Nevertheless, regardless of the evolutionary origin, these sequences certainly do not represent any of the known *SVCT* genes so far characterized. Further studies are needed for the characterization of this putative gene or genes.

In the *SVCTP* consensus phylogeny (Supplementary Figure 7), we observed great branch polytomy, probably due to the high sequence divergence between the representative non-bilaterian, protostomian and basal deuterostomian species analyzed. Given the polytomy, the relationship between the *SVCTNB, SVCTP* and *SVCT* gene sequences and the corresponding *SVCT1*, *SVCT2*, *SVCT3* and *SVCT4* outgroups is not well defined, although it is possible to verify that the first, second and fourth outgroups are clustered together, while the *SVCT3* gene is represented alone with an abnormal branch length. Concerning the Non-Bilateria, we can observe that the *SVCTNB* gene (not refered as *SVCTP* since it does not follow the taxonomic

nomenclature) is not present in the only Porifera species analyzed (*A. queenslandica*) and is duplicated in *T. adhaerens* (Placozoa). As for the non-bilaterian Anthozoa, we were able to observe *SVCTNB* duplications in *Stylophora pistillata, Orbicella faveolata* and *N. vectensis*, while no evidence for *SVCTNB* was found in *Exaiptasia pallida* and *Acropora digitifera*. In *T. adhaerens*, the local duplications happened after speciation, while for the *S. pistillata*, *O. faveolata* and *N. vectensis*, there is evidence of one duplication event in the common ancestor but also of duplication after speciation. We were also able to observe that the only representative of the Hydrozoa phylum (*H. vulgaris*) appears to have a single *SVCTNB* copy and the only representative of the Myxozoa phylum (*T. kitauei*) is not represented in the final phylogeny. Within the Protostomia, specifically in the Insecta class, the *SVCTP* gene is present as a single copy in 90 species, including *D. melanogaster*. This is particularly interesting, since this gene appears to be frequently duplicated in many of the remaining protostomian taxonomic groups represented. Excluding the insect species, but within the Hexapoda phylum, we were able to detect a single *SVCTP* copy in *O. cincta* (springtail) and several duplications in *F. candida* (springtail), one putatively present in the ancestor of this species and several local duplications after speciation. In the Crustacea, one local *SVCTP* duplication seems to have occurred in the *H. azteca* genome, but not in *D. pulex* since only one sequence is represented. As for *D. magna*, no *SVCTP* sequence was observed. Regarding the Araneae, only one *SVCTP* sequence was detected in *P. tepidariorum*, while for *S. mimosarum* no sequence is represented in the final dataset. In the Acari, there is evidence of an ancient *SVCTP* duplication event before speciation for *V. destructor*, *V. jacobsoni* and *G. occidentalis*, while for *T. mercedesae* (bee mite) and *T. urticae* (two-spotted spider mite) only a single *SVCT* copy was identified. For *E. maynei* and *I. scapularis* no *SVCTP* sequences were identified. In the remaining Chelicerata species (*L. polyphemus*), there is evidence of two *SVCTP* duplications. Concerning the Tardigrada, none of the two species analyzed (*H. dujardini* and *R. varieornatus*) seem to contain the *SVCTP* gene. This result is also observed for the single Priapulida species represented in the initial dataset (*P. caudatus*) since no sequence is included in the consensus phylogeny. In the Nematoda taxonomic group, the *SVCTP* gene is duplicated in *Ancylostoma ceylanicum* and *Caenorhabditis remanei*, is a single copy gene in *Haemonchus contortus*, *Caenorhabditis nigoni*, *Caenorhabditis briggsae* and *Diploscapter pachys* and is not detected in the remaining 29 species analyzed. This gene appears to be lost in the Trichinellida order, given the twelve *Trichinella* species excluded from the final phylogeny. In the Mesozoa, we were not able to detect a *SVCTP* sequence in the only analyzed species *I. linei*. Out of the nine Platyhleminthes species analyzed, we were only able to identify one *SVCTP* sequence in *Macrostomum lignano*. Within the Brachiopoda, we were able to detect *SVCTP* duplications before and after speciation in *L. anatina*, while in the Annelida, we were able to detect several local duplications in *C. teleta* but no *SVCTP* sequence representative for *H. robusta*. Regarding the Mollusca phylum, we were able to find *SVCTP* duplications in two Gastropoda

species (*L. gigantea* and *A. californica*) and only one representative *SVCTP* sequence in the gastropod *B. glabrata*. In the Bivalvia class, the *SVCTP* appears to be duplicated a remarkable number of times compared to the remaining protostomians in *C. gigas*, *C. virginica* and *M. yessoensis*, since in the consensus phylogeny 20 *SVCTP* copies represent each of these species. Nevertheless, the remaining Bivalvia species analyzed (*M. galloprovincialis*) does not appear to have the *SVCTP* gene. As for the Cephalopoda, we were not able to detect the *SVCTP* gene in *O. bimaculoides*.

Within the deuterostome Echinodermata taxonomic group, the *SVCT* gene (not refered as *SVCTP* since it does not follow the taxonomic nomenclature) can be found duplicated one time in *A. planci*, two times in *S. purpuratus* and as a single copy in *Apostichopus japonicus*. Within the Hemichordata, the *SVCT* gene can be seen duplicated in *S. kowalevskii*, and in the Urochordata the same scenario is seen for *O. dioica* and *C. intestinalis*. Finally, regarding the Cephalochordata group, the *SVCT* gene is duplicated in *B. floridae* and *B. belcheri*. Seeing duplications in the four most basal deuterostomian groups implicates a possible ancestral duplication event at the base of all the deuterostomian species.

It is interesting to note that the vertebrate *SVCT3* gene with unknown function emerges as the one closely related to *SVCTP*. However, the phylogenetic relations displayed for the *SVCT3* gene relative to the other genes analyzed may be biased, since this gene in particular seems to present signs of accelerated rate of evolution, as can be seen by the particularly long branches representative of the *SVCT3* sequences in the consensus phylogenies (Supplementary figures 2-7). This is a known technical issue regarding the Bayesian phylogenetic relations inference (Wägele and Mayer 2007), and it may be the reason why this gene is not represented closely related to the *SVCT4* instead of the *SVCTP*. Considering this hypothesis, we were able to summarize our findings regarding the possible evolutionary history of the *SVCT* genes in cladogram format, represented in Figures 13 to 16B from the most basal animal taxonomic groups to the more recently diverged.

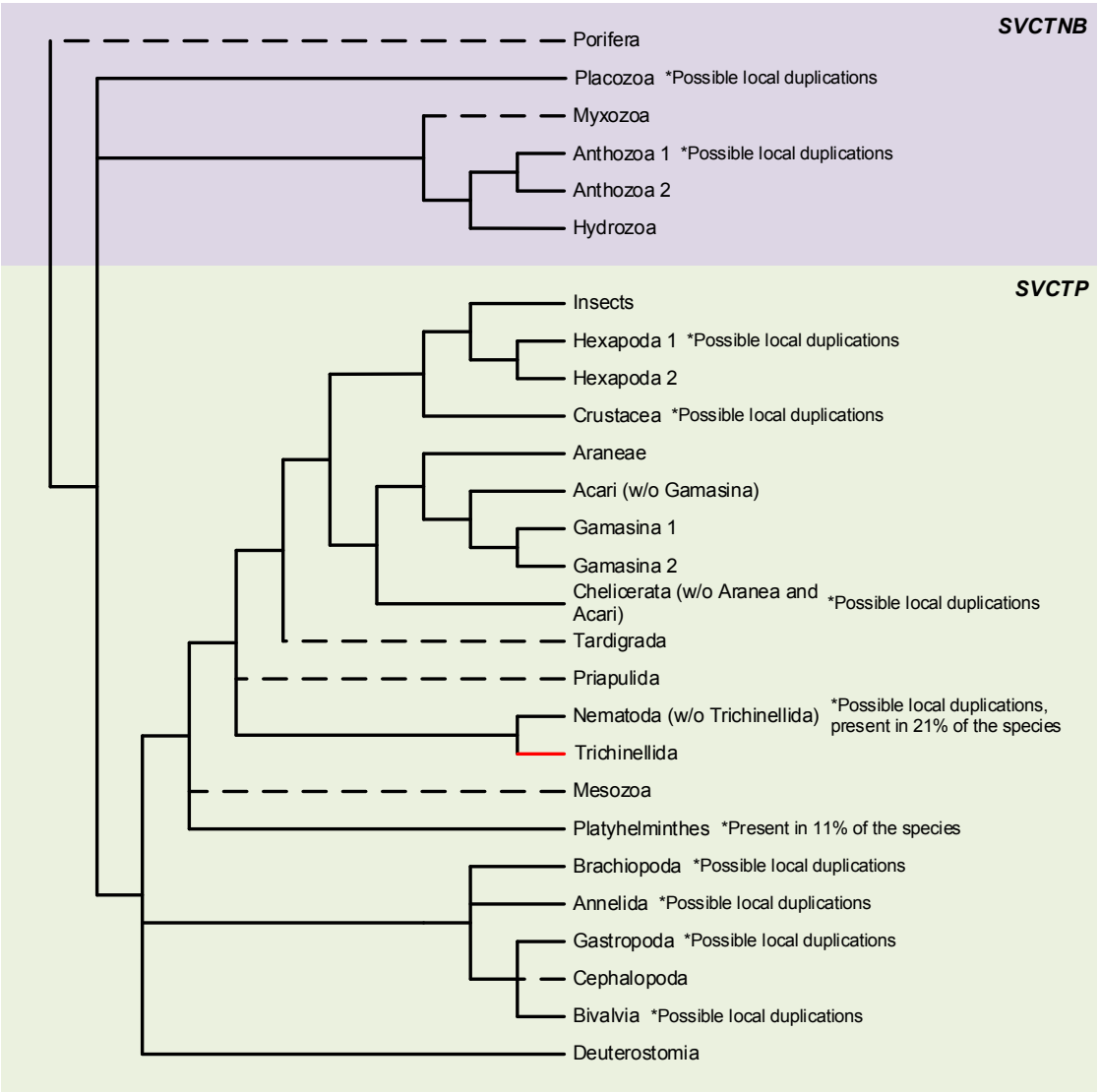


Figure 13 – *SVCTNB* and *SVCTP* evolutionary history cladogram. The non-bilaterian taxonomic groups are highlighted in purple, while the protostomian groups and deuterostomia split branch are highlighted in olive green. Lineages where a gene is extrapolated as lost have red branches, while dashed branches are representative of lineages where gene loss cannot be inferred with the available data. The "*" marks lineages with possible local duplications and where important gene loss events may have happened (see text for details). Taxonomic groups duplicated with "1" and "2" tags are affected by duplication events before speciation. Taxonomic relationships are depicted as in the Tree of life web project.

Figure 13 indicates the most likely scenario regarding the non-bilaterian and protostomian species. However, when considering the taxonomic transition to basal deuterostomians and subsequence deuterostomian radiation, there are two scenarios to consider: i) the *SVCT1*, *SVCT2*, *SVCT3* and *SVCT4* genes in chordates arose from an ancestral basal deuterostome duplicated *SVCT* while the other copy was lost (Figure 14A); or ii) the *SVCT1*, *SVCT2*, *SVCT3* and *SVCT4* genes in chordates arose from an ancestral basal deuterostome *SVCT* gene and the Echinodermata/ Hemichordata, Urochordata and Cephalochordata *SVCT* gene was independently duplicated three times (Figure 14B).

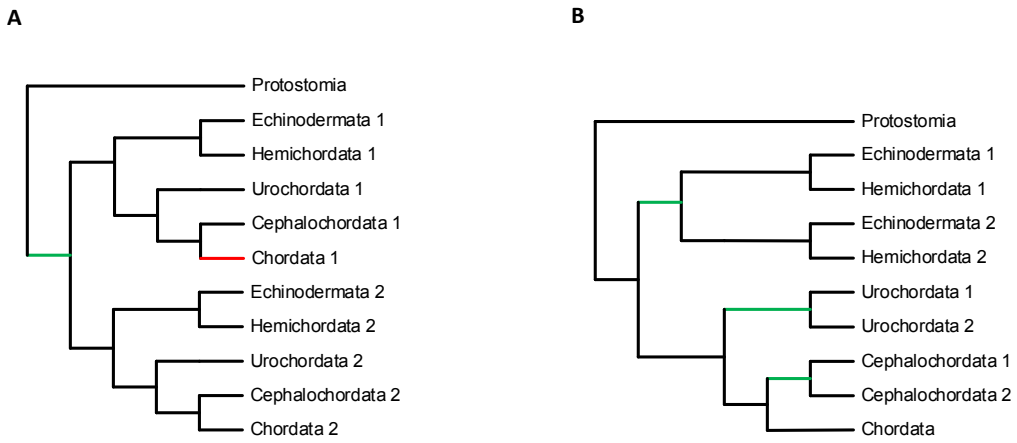
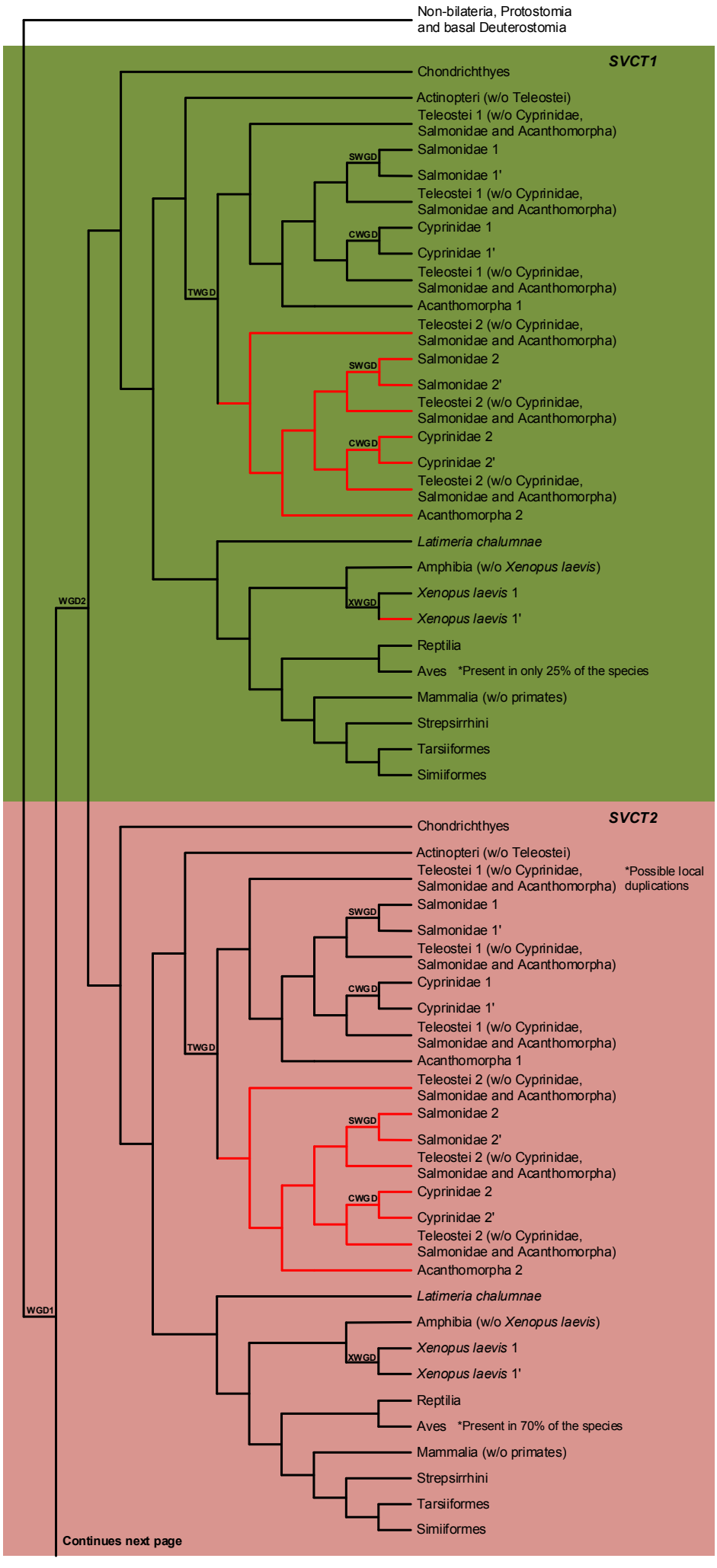


Figure 14 – Basal deuterostomian species inferred *SVCT* evolutionary histories. A) The green branch represents the ancestral duplication that may have affected all deuterostomian species. In this scenario, the Echinodermata, Hemichordata, Urochordata and Cephalochordata retained two *SVCT* copies in their genomes, while the remaining Chordata species *SVCT*s evolved from a single copy (Chordata 2) while the other was lost (Chordata 1, in the branch represented in red). B) In this scenario, the three green branches represent independent duplications that originated two copies of *SVCT* in the Echinodermata/Hemichordata, the Urochordata and the Cephalochordata groups, while the Chordata species *SVCT*s derived from the single ancestral *SVCT* copy. Taxonomic relationships are depicted as in the Tree of life web project.

Although we cannot extrapolate the most likely situation at the base of deuterostomians, we can infer the phylogenetic history of the *SVCT1*, *SVCT2*, *SVCT3* and *SVCT4* genes within the Chordata. Given the more complex history regarding the possible duplication of the *SVCT4* gene in Actinopteri and Amphibia species, we first summarize our results for the *SVCT1, SVCT2* and *SVCT3* genes in Figure 15.


Non-bilateria, Protostomia and basal Deuterostomia
SVCT1
Chondrichthyes
Actinopteri (w/o Teleostei)
Teleostei 1 (w/o Cyprinidae, Salmonidae and Acanthomorpha)
Salmonidae 1
SWGD
Salmonidae 1'
Teleostei 1 (w/o Cyprinidae, Salmonidae and Acanthomorpha)
Cyprinidae 1
CWGD
Cyprinidae 1'
Teleostei 1 (w/o Cyprinidae, Salmonidae and Acanthomorpha)
TWGD
Acanthomorpha 1
Teleostei 2 (w/o Cyprinidae, Salmonidae and Acanthomorpha)
Salmonidae 2
SWGD
Salmonidae 2'
Teleostei 2 (w/o Cyprinidae, Salmonidae and Acanthomorpha)
Cyprinidae 2
CWGD
Cyprinidae 2'
Teleostei 2 (w/o Cyprinidae, Salmonidae and Acanthomorpha)
Acanthomorpha 2
Latimeria chalumnae
Amphibia (w/o Xenopus laevis)
Xenopus laevis 1
XWGD
Xenopus laevis 1'
WGD2
Reptilia
Aves *Present in only 25% of the species
Mammalia (w/o primates)
Strepsirrhini
Tarsiiformes
Simiiformes
SVCT2
Chondrichthyes
Actinopteri (w/o Teleostei)
Teleostei 1 (w/o Cyprinidae, Salmonidae and Acanthomorpha)
*Possible local duplications
Salmonidae 1
SWGD
Salmonidae 1'
Teleostei 1 (w/o Cyprinidae, Salmonidae and Acanthomorpha)
Cyprinidae 1
CWGD
Cyprinidae 1'
Teleostei 1 (w/o Cyprinidae, Salmonidae and Acanthomorpha)
TWGD
Acanthomorpha 1
Teleostei 2 (w/o Cyprinidae, Salmonidae and Acanthomorpha)
Salmonidae 2
SWGD
Salmonidae 2'
Teleostei 2 (w/o Cyprinidae, Salmonidae and Acanthomorpha)
Cyprinidae 2
CWGD
Cyprinidae 2'
Teleostei 2 (w/o Cyprinidae, Salmonidae and Acanthomorpha)
Acanthomorpha 2
Latimeria chalumnae
WGD1
Amphibia (w/o Xenopus laevis)
Xenopus laevis 1
XWGD
Xenopus laevis 1'
Reptilia
Aves *Present in 70% of the species
Mammalia (w/o primates)
Strepsirrhini
Tarsiiformes
Simiiformes
Continues next page

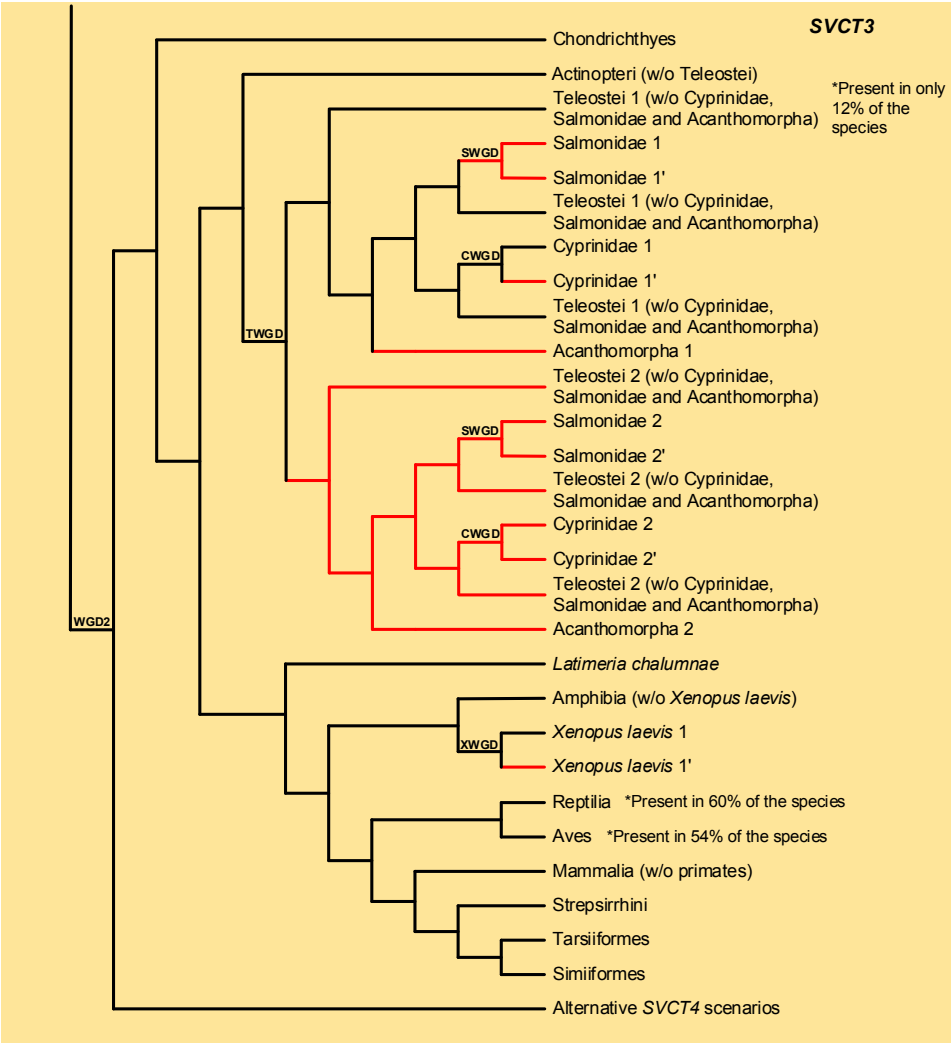


Figure 15 – Inferred evolutionary history for the *SVCT1*, *SVCT2* and *SVCT3* genes. *SVCT1* (green) is depicted as the phylogenetically closer to *SVCT2* (pink), while *SVCT3* (yellow) closer to *SVCT4* in a distinct lineage. The two rounds of whole genome duplication known to have occurred in vertebrates are represented respectively as WGD1 and WGD2. Further lineage-specific duplications can also be observed in teleosts (TWGD), Salmonidae (SWGD), Cyprinidae (CWGD) and *X. laevis* (XWGD). Lineages where either gene is extrapolated as lost have red branches. The "*" marks lineages with possible local duplications and where important gene loss events may have happened (see text for details). Taxonomic relationships are depicted as in the Tree of life web project and in Helgen (2011).

According to our results, the *SVCT4* gene appears to be duplicated in the Actinopteri and Amphibia lineages. Neverthess, this duplication can be the result of either a single local duplication on the common ancestral of these two taxonomic groups or, alternatively, two independent duplications at the base of Actinopteri and Amphibia. These two possible scenarios can be observed in figures 16A and 16B, and are complementary to Figure 15.

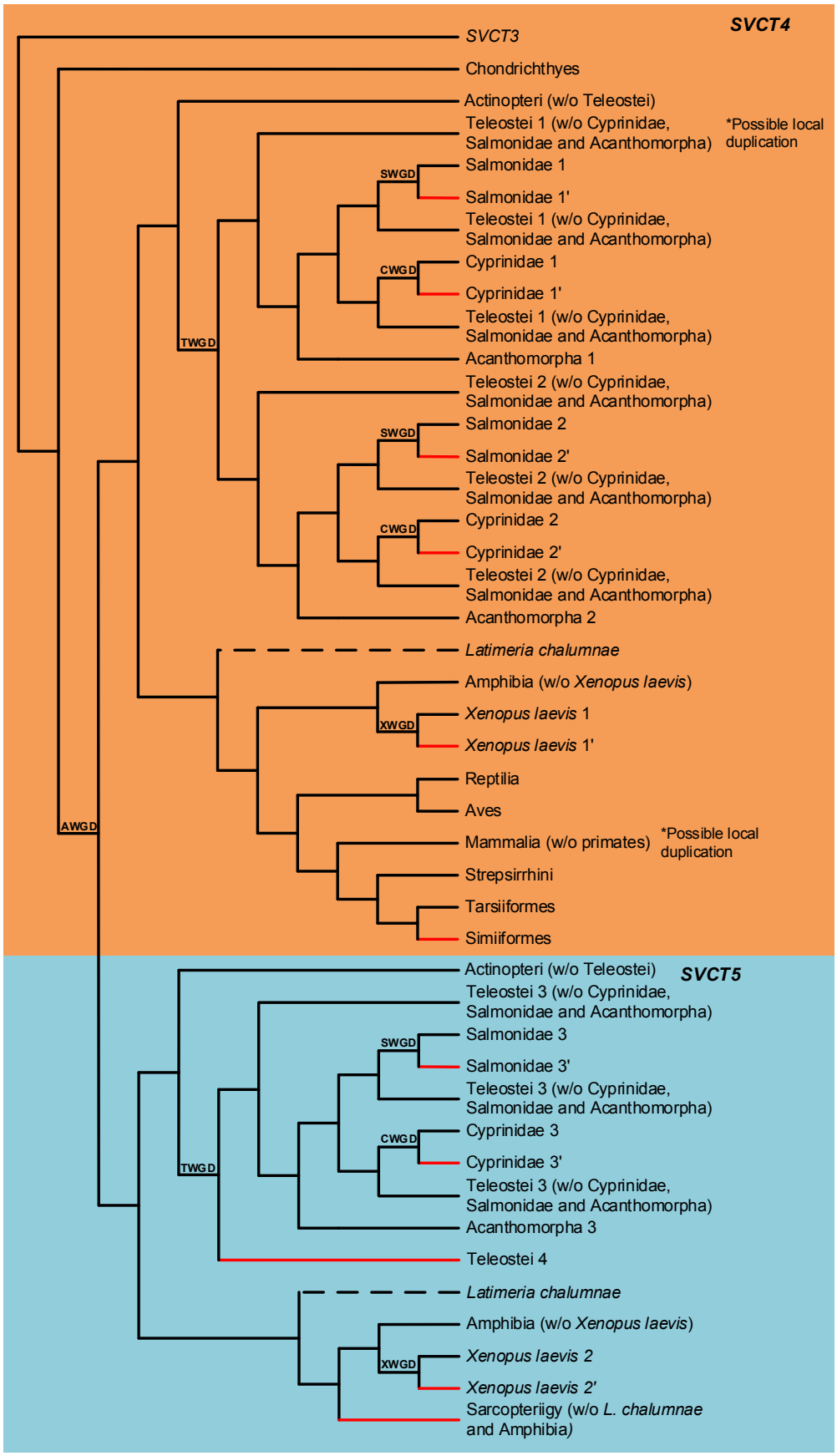


Figure 16A - Inferred evolutionary history for the *SVCT4* and *SVCT5* genes. *SVCT4* (orange) is depicted as phylogenetically closer to *SVCT5* (blue), and is linked to the *SVCT3* gene at the top of the cladogram. The putative ancestral duplication event at the base of

the Actinopteri and Amphibia is represented as AWGD. Further lineage-specific duplications can also be observed in teleosts (TWGD), Salmonidae (SWGD), Cyprinidae (CWGD) and *X. laevis* (XWGD). Lineages where either gene is extrapolated as lost have red branches, while dashed branches are representative of lineages where gene loss cannot be inferred with the available data. The "*" marks lineages with possible local duplications (see text for details). Taxonomic relationships are depicted as in the Tree of life web project and in Helgen (2011).

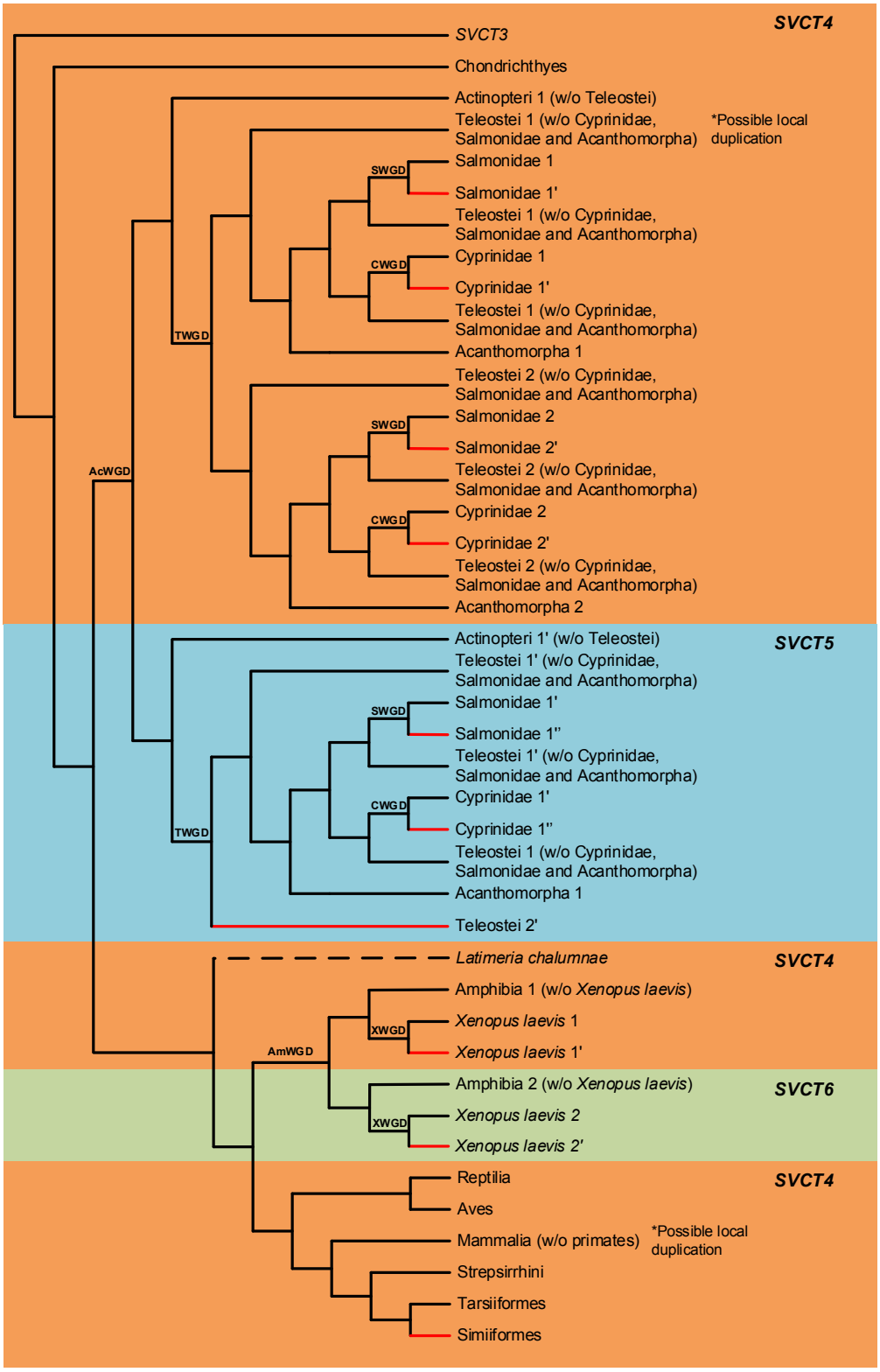


Figure 16B - Inferred evolutionary history for the *SVCT4*, *SVCT5* and *SVCT6* genes. The *SVCT4* gene (orange) is linked to the *SVCT3*

represented as AcWGD and AmWGD, respectively. Further lineage-specific duplications can also be observed in teleosts (TWGD), Salmonidae (SWGD), Cyprinidae (CWGD) and *X. laevis* (XWGD). *SVCT5* (blue) is the the prevalent gene from the independently duplicated *SVCT4* gene in Actinopteri, while *SVCT6* (green) is the remaining gene from the *SVCT4* duplication in the Amphibia. Lineages where genes are extrapolated as lost have red branches, while dashed branches are representative of lineages where gene presence or loss cannot be inferred with the available data. The "*" marks lineages with possible local duplications (see text for details). Taxonomic relationships are depicted as in the Tree of life web project and in Helgen (2011).

Given the intermediate taxonomic position of *L. chalumnae,* the presence of either one or two *SVCT4* copies in this species could provide support to the hypotheses of Figure 16A or Figure 16B, respectively. However, in the absence of this information, none of the scenarios seems more likely.

In summary, the *SVCT* genes are generally ubiquous in all animal taxonomic groups. Nevertheless, several considerable gene loss events may have happened in specific and rather large lineages. In protostomians, the *SVCTP* gene appears to have been lost in the Trichinellida group, as well as several Platyhelminthes lineages. In deuterostomians, the *SVCT1* gene may have been lost in many Aves lineages ancestrally, such as the *SVCT3*. Furthermore, the *SVCT3* gene also seems to be lost in many Actinopteri lineages, such as the Acanthomorpha. The other loss event already reported in the literature (Yamamoto *et al.* 2010) is also worth mentioning, regarding the *SVCT4* gene in the Simiiformes group. However, the *SVCT4* gene also seems duplicated specifically in the Actinopteri and Amphibia lineages, a rather unexpected result.

Although a possible *SVCT* molecular evolution history can be extrapolated, what happened many basal animal lineages is quite difficult to desbribe. The fact that, in humans, the *SVCT1* transporter is actively involved in ascorbic acid absorption through the gastrointestinal tract and the *SVCT2* transporter in direct ascorbic acid uptake from specific tissues such as the brain (May 2011, Savini *et al.* 2008), suggests that processes of neofunctionalization may have been important during *SVCT* evolution, contributing to gene fixation. Nevertheless, we cannot tell if the ancestor gene of the *SVCT1/SVCT2* and *SVCT3/SVCT4* lineages underwent a process of subfunctionalization after a duplication event in vertebrates, which led the *SVCT1/SVCT2* lineage to acquire the capacity to transport ascorbic acid, or alternatively, if neofunctionalization occurred after a duplication event. As such, it is also not possible to infer whether protostomians *SVCTP* is able to transport vitamin C. Therefore, in the absence of convincing phylogenetic results, we turned to conserved protein motifs to possibly find answers. By comparison of aligned protein sequences, we were able to find a mutual eight amino acid conserved protein motif for SVCT1 and SVCT2 (QHYLTCFS) for all the animal species present in the final datasets, with the exception of a single sequence belonging to *Aotus nancymaae* (XP_012329492.1). Furthermore, Kourkoulou *et al.* (2018) was also able to discover a conserved SSSP amino acid motif for these transporters. Since the ascorbic acid transport capacity is largely attributed to these proteins, it is possible that the found protein motifs can be related with this functional characteristic. Following

this hypothesis, if the conserved QHYLTCFS pattern was identified in the SVCTP proteins, that could be indicative of similar substrate specificity for ascorbic acid. Nevertheless, we uncovered that none of the conserved patterns are found in the SVCTP proteins, but also that both conserved patterns are unique of these transporters, and thus not found in the SVCT3, SVCT4 and SVCTP proteins. These results corroborate and add information to the evidences presented by Kourkoulou *et al.* (2018), but do not allow to assume that the ancestral transporter protein could in fact be transporting ascorbic acid.

IV. Conclusion

Using a molecular evolution approach, we were able to determine that the *GULO* gene is present in several lineages within Protostomia, adding support to the hints given by Wheeler *et al.* (2015). Furthermore, we were able to describe several *GULO* loss events within several animal lineages, and verify that in general, this gene is not duplicated in deuterostomians, protostomians and non-bilaterians. Furthermore, with the evidence provided by our phylogenetic analysis, we were able to trace a putative *GULO* to the ancestral of animal and Fungi taxonomic groups, in accordance with the hypothesis of Wheeler *et al.* (2015). Furthermore, we provide evidence that the conserved amino acid motif in the GULO protein may not be the HWXK described in the literature, but a more specific HWAK amino acid pattern found in animal GULO and Fungi ALO. Moreover, we were also able to extrapolate putative ascorbic acid synthesis in a model organism (*D. melanogaster*) without an identifiable *GULO*, given that our results show that the dietary source and microbiome do not influence the ascorbic acid levels detected in this species. In addition, to reinforce this idea, and following the reports by Massie *et al.* (1991), we were able to determine that after cold acclimation conditions, *D. melanogaster* is able to compensate loss in ascorbic acid levels after being allowed to recover for one day at 25ºC. Considering our results, we propose that either a novel pathway may exist for ascorbic acid synthesis in *D. melanogaster*, or alternatively, that a single enzyme may be performing the same function as GULO in this species. Either way, further studies are needed to uncover candidate genes that can be involved in ascorbic acid synthesis in GULO-absent species.

The four *SVCT* genes found in most vertebrates seem to follow two whole genome duplications from a single ancestral gene. Specifically, the *SVCT4* gene appears to have been duplicated in two specific lineages (Amphibia and Actinopteri), either by two independent local duplications or a single ancestral duplication that affected the ancestral of these two taxonomic groups. However, with the available data, it is impossible to infer the most likely scenario. We were also able to reinforce the evidence of *SVCT3* loss in many teleost lineages but also *SVCT4* loss in the Simiiformes, where *H. sapiens* is included. Moreover, the Aves taxonomic group appears to follow a complex gene loss scenario for the *SVCT1* and *SVCT3* genes. Although in the Chordata the duplications observed can be attributed to a single ancestral gene, duplications observed in Echinodermata, Hemichordata, Urochordata and Cephalochordata indicate that basal deuterostomians may have been affected by three independent genome duplications or an ancient duplication event. Within protostomians and non-bilaterians, the *SVCT* gene appears to be independently duplicated many times, and the more ancient duplications detected regard the Anthozoa, Hexapoda and Gamasina taxonomic groups. Nevertheless, notable gene loss appears to have affected the Nematoda lineages (essencially the Trichinellida) and the Platyhelminthes.

Curiously, the Insecta species have always only one *SVCT* copy. With the molecular evolution findings regarding the *SVCT* gene, we could not uncover putative ascorbic acid transport traits in the ancient transporter. Trying to overcome this issue, we used conserved protein motifs to establish a correlation between the ascorbic acid transporters SVCT1 and SVCT2 and the remaining proteins analyzed, but no conclusion that supports ancestral ascorbic acid transport traits was taken. Further analyses are therefore needed to further describe the SVCTs functional characteristics, and establish a correlation between the presence of SVCTs in protostomians and ascorbic acid homeostasis.

V. References

Aboobucker, S. I. and A. Lorence (2016). "Recent Progress on the Characterization of Aldonolactone Oxidoreductases." Plant physiology and biochemistry". PPB / Societe francaise de physiologie vegetale 98: 171-185.

Alexandrou, M. A., B. A. Swartz, N. J. Matzke and T. H. Oakley (2013). "Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae." Mol Phylogenet Evol 69(3): 514-523.

Amato, P. and B. C. Christner (2009). "Energy metabolism response to low-temperature and frozen conditions in Psychrobacter cryohalolentis." Appl Environ Microbiol 75(3): 711-718.

Ames, B. N., R. Cathcart, E. Schwiers and P. Hochstein (1981). "Uric acid provides an antioxidant defense in humans against oxidant- and radical-caused aging and cancer: a hypothesis." Proceedings of the National Academy of Sciences 78(11): 6858.

Arenas, M. (2015). "Trends in substitution models of molecular evolution." Frontiers in Genetics 6: 319.

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(3): 458-483.

Bahadorani, S., P. Bahadorani, J. P. Phillips and A. J. Hilliker (2008). "The Effects of Vitamin Supplementation on Drosophila Life Span Under Normoxia and Under Oxidative Stress." The Journals of Gerontology: Series A 63(1): 35-42.

Barbehenn, R. V., S. L. Bumgarner, E. F. Roosen and M. M. Martin (2001). "Antioxidant defenses in caterpillars: role of the ascorbate-recycling system in the midgut lumen." J Insect Physiol 47(4-5): 349-357.

Barker, F. K., A. Cibois, P. Schikler, J. Feinstein and J. Cracraft (2004). "Phylogeny and diversification of the largest avian radiation." Proceedings of the National Academy of Sciences of the United States of America 101(30): 11040-11045.

Biondi, C., B. Pavan, A. Dalpiaz, S. Medici, L. Lunghi and F. Vesce (2007). "Expression and characterization of vitamin C transporter in the human trophoblast cell line HTR-8/SVneo: effect of steroids, flavonoids and NSAIDs." Mol Hum Reprod 13(1): 77-83.

Black, W. C. t., J. S. Klompen and J. E. Keirans (1997). "Phylogenetic relationships among tick subfamilies (Ixodida: Ixodidae: Argasidae) based on the 18S nuclear rDNA gene." Mol Phylogenet Evol 7(1): 129-144.

Borgese, N., G. Pietrini and S. Gaetani (1987). "Concentration of NADH-cytochrome b5 reductase in erythrocytes of normal and methemoglobinemic individuals measured with a quantitative radioimmunoblotting assay." Journal of Clinical Investigation 80(5): 1296-1302.

Bremus, C., U. Herrmann, S. Bringer-Meyer and H. Sahm (2006). "The use of microorganisms in L-ascorbic acid production." J Biotechnol 124(1): 196-205.

Buettner, G. R. and B. A. Jurkiewicz (1996). "Catalytic metals, ascorbate and free radicals: combinations to avoid." Radiat Res 145(5): 532-541.

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." Mol Aspects Med 34(2-3): 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(2): 83-94.

Carlson, K. A., K. Gardner, A. Pashaj, D. J. Carlson, F. Yu, J. D. Eudy, C. Zhang and L. G. Harshman (2015). "Genome-Wide Gene Expression in relation to Age in Large Laboratory Cohorts of Drosophila melanogaster." Genetics Research International 2015: 19.

Carr, R. S., M. B. Bally, P. Thomas and J. M. Neff (1983). "Comparison of methods for determination of ascorbic acid in animal tissues." Anal Chem 55(8): 1229-1232.

Carr, R. S. and J. M. Neff (1980). "Determination of ascorbic acid in tissues of marine animals by liquid chromatography with electrochemical detection." Anal Chem 52(14): 2428-2430.

Chain, F. J. and B. J. Evans (2006). "Multiple mechanisms promote the retained expression of gene duplicates in the tetraploid frog Xenopus laevis." PLoS Genet 2(4): e56.

Chambers, K. E., R. McDaniell, J. D. Raincrow, M. Deshmukh, P. F. Stadler and C. H. Chiu (2009). "Hox cluster duplication in the basal teleost Hiodon alosoides (Osteoglossomorpha)." Theory Biosci 128(2): 109-120.

Chang, Y. L., M. Rossetti, H. Vlamakis, D. Casero, G. Sunga, N. Harre, S. Miller, R. Humphries, T. Stappenbeck, K. W. Simpson, R. B. Sartor, G. Wu, J. Lewis, F. Bushman, D. P. B. McGovern, N. Salzman, J. Borneman, R. Xavier, C. Huttenhower and J. Braun (2018). "A screen of Crohn's

disease-associated microbial metabolites identifies ascorbate as a novel metabolic inhibitor of activated human T cells." Mucosal Immunol.

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." Comp Biochem Physiol B Biochem Mol Biol 147(2): 178-190.

Clark, A. G., A. L. Rohrbaugh, I. Otterness and V. B. Kraus (2002). "The effects of ascorbic acid on cartilage metabolism in guinea pig articular cartilage explants." Matrix Biol 21(2): 175-184.

Colinet, H. (2011). "Disruption of ATP homeostasis during chronic cold stress and recovery in the chill susceptible beetle (Alphitobius diaperinus)." Comp Biochem Physiol A Mol Integr Physiol 160(1): 63-67.

Corpe, C. P., J. H. Lee, O. Kwon, P. Eck, J. Narayanan, K. L. Kirk and M. Levine (2005). "6-Bromo-6-deoxy-L-ascorbic acid: an ascorbate analog specific for Na+-dependent vitamin C transporter but not glucose transporter pathways." J Biol Chem 280(7): 5211-5220.

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." J Clin Invest 120(4): 1069-1083.

Cui, J., Y. H. Pan, Y. Zhang, G. Jones and S. Zhang (2011). "Progressive pseudogenization: vitamin C synthesis and its loss in bats." Mol Biol Evol 28(2): 1025-1031.

Dabrowski, K. (1994). "Primitive actimoterigian fishes can synthesize ascorbic acid." Experientia 50(8): 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(1): 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(7): 825-860.

David, L., S. Blum, M. W. Feldman, U. Lavi and J. Hillel (2003). "Recent duplication of the common carp (Cyprinus carpio L.) genome as revealed by analyses of microsatellite loci." Mol Biol Evol 20(9): 1425-1434.

de Koning, H. and G. Diallinas (2000). "Nucleobase transporters (review)." Mol Membr Biol 17(2): 75-94.

Dehal, P. and J. L. Boore (2005). "Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate." PLOS Biology 3(10): e314.

Diliberto, E. J., Jr., G. D. Heckman and A. J. Daniels (1983). "Characterization of ascorbic acid transport by adrenomedullary chromaffin cells. Evidence for Na+-dependent co-transport." J Biol Chem 258(21): 12886-12894.

Dollo, V. H., S. X. Yi and R. E. Lee, Jr. (2010). "High temperature pulses decrease indirect chilling injury and elevate ATP levels in the flesh fly, Sarcophaga crassipalpis." Cryobiology 60(3): 351-353.

Domes, K., M. Althammer, R. A. Norton, S. Scheu and M. Maraun (2007). "The phylogenetic relationship between Astigmata and Oribatida (Acari) as indicated by molecular markers." Exp Appl Acarol 42(3): 159-171.

Dowling, A. P. G. and B. M. Oconnor (2010). "Phylogenetic relationships within the suborder Dermanyssina (Acari: Parasitiformes) and a test of dermanyssoid monophyly." International Journal of Acarology 36(4): 299-312.

Drouin, G., J.-R. Godin and B. Pagé (2011). "The Genetics of Vitamin C Loss in Vertebrates." Current Genomics 12(5): 371-378.

Du, J., J. J. Cullen and G. R. Buettner (2012). "Ascorbic acid: Chemistry, biology and the treatment of cancer." Biochimica et biophysica acta 1826(2): 443-457.

Evans, B. J., D. B. Kelley, D. J. Melnick and D. C. Cannatella (2005). "Evolution of RAG-1 in polyploid clawed frogs." Mol Biol Evol 22(5): 1193-1207.

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

Ferris, S. D. and G. S. Whitt (1977). "Duplicate gene expression in diploid and tetraploid loaches (Cypriniformes, Cobitidae)." Biochem Genet 15(11-12): 1097-1112.

Figueroa-Méndez, R. and S. Rivas-Arancibia (2015). "Vitamin C in Health and Disease: Its Role in the Metabolism of Cells and Redox State in the Brain." Frontiers in Physiology 6: 397.

Fraaije, M. W., R. H. van den Heuvel, W. J. van Berkel and A. Mattevi (1999). "Covalent flavinylation is essential for efficient redox catalysis in vanillyl-alcohol oxidase." J Biol Chem 274(50): 35514-35520.

Frei, B. and S. Lawson (2008). "Vitamin C and cancer revisited." Proc Natl Acad Sci U S A 105(32): 11037-11038.

Gale, C. R., C. N. Martyn and C. Cooper (1996). "Cognitive impairment and mortality in a cohort of elderly people." Bmj 312(7031): 608-611.

Glasauer, S. M. and S. C. Neuhauss (2014). "Whole-genome duplication in teleost fishes and its evolutionary consequences." Mol Genet Genomics 289(6): 1045-1060.

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." J Biol Chem 282(1): 615-624.

Guderley, H. (2004). "Metabolic responses to low temperature in fish muscle." Biol Rev Camb Philos Soc 79(2): 409-427.

Gulia-Nuss, M., A. B. Nuss, J. M. Meyer, D. E. Sonenshine, R. M. Roe, R. M. Waterhouse, D. B. Sattelle, J. de la Fuente, J. M. Ribeiro, K. Megy, J. Thimmapuram, J. R. Miller, B. P. Walenz, S. Koren, J. B. Hostetler, M. Thiagarajan, V. S. Joardar, L. I. Hannick, S. Bidwell, M. P. Hammond, S. Young, Q. Zeng, J. L. Abrudan, F. C. Almeida, N. Ayllón, K. Bhide, B. W. Bissinger, E. Bonzon-Kulichenko, S. D. Buckingham, D. R. Caffrey, M. J. Caimano, V. Croset, T. Driscoll, D. Gilbert, J. J. Gillespie, G. I. Giraldo-Calderón, J. M. Grabowski, D. Jiang, S. M. S. Khalil, D. Kim, K. M. Kocan, J. Koči, R. J. Kuhn, T. J. Kurtti, K. Lees, E. G. Lang, R. C. Kennedy, H. Kwon, R. Perera, Y. Qi, J. D. Radolf, J. M. Sakamoto, A. Sánchez-Gracia, M. S. Severo, N. Silverman, L. Šimo, M. Tojo, C. Tornador, J. P. Van Zee, J. Vázquez, F. G. Vieira, M. Villar, A. R. Wespiser, Y. Yang, J. Zhu, P. Arensburger, P. V. Pietrantonio, S. C. Barker, R. Shao, E. M. Zdobnov, F. Hauser, C. J. P. Grimmelikhuijzen, Y. Park, J. Rozas, R. Benton, J. H. F. Pedra, D. R. Nelson, M. F. Unger, J. M. C. Tubio, Z. Tu, H. M. Robertson, M. Shumway, G. Sutton, J. R. Wortman, D. Lawson, S. K. Wikel, V. M. Nene, C. M. Fraser, F. H. Collins, B. Birren, K. E. Nelson, E. Caler and C. A. Hill (2016). "Genomic insights into the Ixodes scapularis tick vector of Lyme disease." Nature Communications 7: 10507.

Guo, B., X. Gan and S. He (2010). "Hox genes of the Japanese eel Anguilla japonica and Hox cluster evolution in teleosts." J Exp Zool B Mol Dev Evol 314(2): 135-147.

Halliwell, B. (1999). "Vitamin C: poison, prophylactic or panacea?" Trends Biochem Sci 24(7): 255-259.

Hansen, S. N., P. Tveden-Nyborg and J. Lykkesfeldt (2014). "Does Vitamin C Deficiency Affect Cognitive Development and Function?" Nutrients 6(9): 3818-3846.

He, X. B., M. Kim, S. Y. Kim, S. H. Yi, Y. H. Rhee, T. Kim, E. H. Lee, C. H. Park, S. Dixit, F. E. Harrison and S. H. Lee (2015). "Vitamin C facilitates dopamine neuron differentiation in fetal midbrain through TET1- and JMJD3-dependent epigenetic control manner." Stem Cells 33(4): 1320-1332.

Helgen, K. M. (2011). "The Mammal Family Tree." Science 334(6055): 458-459.

Henkel, C. V., E. Burgerhout, D. L. de Wijze, R. P. Dirks, Y. Minegishi, H. J. Jansen, H. P. Spaink, S. Dufour, F. A. Weltzien, K. Tsukamoto and G. E. van den Thillart (2012). "Primitive duplicate Hox clusters in the European eel's genome." PLoS One 7(2): e32231.

Hoyoux, A., V. Blaise, T. Collins, S. D'Amico, E. Gratia, A. L. Huston, J. C. Marx, G. Sonan, Y. Zeng, G. Feller and C. Gerday (2004). "Extreme catalysts from low-temperature environments." J Biosci Bioeng 98(5): 317-330.

Jiang, Z. Y., Y. Zhong, J. Zheng, M. Ali, G. D. Liu and X. L. Zheng (2018). "L-ascorbic acid metabolism in an ascorbate-rich kiwifruit (Actinidia. Eriantha Benth.) cv. 'White' during postharvest." Plant Physiol Biochem 124: 20-28.

Johnson, K. R., J. E. Wright, Jr. and B. May (1987). "Linkage relationships reflecting ancestral tetraploidy in salmonid fish." Genetics 116(4): 579-591.

Kasahara, M. (2013). "Impact of whole-genome duplication on vertebrate development and evolution." Semin Cell Dev Biol 24(2): 81-82.

Keeling, P. J. and Y. Inagaki (2004). "A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1α." Proceedings of the National Academy of Sciences of the United States of America 101(43): 15380-15385.

Kolawole, A. O., F. M. Olajuyigbe, J. O. Ajele and C. O. Adedire (2014). "Activity of the Antioxidant Defense System in a Typical Bioinsecticide-and Synthetic Insecticide-treated Cowpea Storage Beetle Callosobrochus maculatus F. (Coleoptera: Chrysomelidae)." International Journal of Insect Science 6: IJIS.S19434.

Kourkoulou, A., A. A. Pittis and G. Diallinas (2018). "Evolution of substrate specificity in the Nucleobase-Ascorbate Transporter (NAT) protein family." Microbial Cell 5(6): 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." J Nutr 134(9): 2216-2221.

Lachapelle, M. Y. and G. Drouin (2011). "Inactivation dates of the human and guinea pig vitamin C genes." Genetica 139(2): 199-207.

LeBlanc, J. G., C. Milani, G. S. de Giori, F. Sesma, D. van Sinderen and M. Ventura (2013). "Bacteria as vitamin suppliers to their host: a gut microbiota perspective." Curr Opin Biotechnol 24(2): 160-168.

Lee, J. H., C. S. Oh, G. H. Mun, J. H. Kim, Y. H. Chung, Y. I. Hwang, D. H. Shin and W. J. Lee (2006). "Immunohistochemical localization of sodium-dependent L-ascorbic acid transporter 1 protein in rat kidney." Histochem Cell Biol 126(4): 491-494.

Li, Z., G. P. Tiley, S. R. Galuska, C. R. Reardon, T. I. Kidder, R. J. Rundell and M. S. Barker (2018). "Multiple large-scale gene and genome duplications during the evolution of hexapods." Proc Natl Acad Sci U S A 115(18): 4713-4718.

Liana, M. and W. Witalinski (2005). "Sperm structure and phylogeny of Astigmata." J Morphol 265(3): 318-324.

Lindblad, M., P. Tveden-Nyborg and J. Lykkesfeldt (2013). "Regulation of Vitamin C Homeostasis during Deficiency." Nutrients 5(8): 2860-2879.

Linster, C. L. and E. Van Schaftingen (2007). "Vitamin C. Biosynthesis, recycling and degradation in mammals." Febs j 274(1): 1-22.

Liu, J., H. Liu and H. Zhang (2018). "Phylogeny and evolutionary radiation of the marine mussels (Bivalvia: Mytilidae) based on mitochondrial and nuclear genes." Mol Phylogenet Evol 126: 233-240.

Liu, X.-Y., K. Y. Yang, M.-Q. Wang, J. S.-L. Kwok, X. Zeng, Z. Yang, X.-J. Xiao, C. P.-Y. Lau, Y. Li, Z.-m. Huang, J.-g. Ba, A. K.-Y. Yim, C.-Y. Ouyang, S.-M. Ngai, T.-F. Chan, E. L.-H. Leung, L. Liu, Z.-G. Liu and S. K.-W. Tsui (2018). "High-quality assembly of Dermatophagoides pteronyssinus genome and transcriptome reveals a wide range of novel allergens." Journal of Allergy and Clinical Immunology 141(6): 2268-2271.e2268.

Loewus, F. A. (1999). "Biosynthesis and metabolism of ascorbic acid in plants and of analogs of ascorbic acid in fungi." Phytochemistry 52(2): 193-210.

Logan, F. J., M. C. Taylor, S. R. Wilkinson, H. Kaur and J. M. Kelly (2007). "The terminal step in vitamin C biosynthesis in Trypanosoma cruzi is mediated by a FMN-dependent galactonolactone oxidase." Biochem J 407(3): 419-426.

López-Fernández, H., P. Duque, S. Henriques, N. Vázquez, F. Fdez-Riverola, C. P. Vieira, M. Reboiro-Jato and J. Vieira (2018). A Bioinformatics Protocol for Quickly Creating Large-Scale

Phylogenetic Trees. Practical Applications of Computational Biology and Bioinformatics, 12th International Conference, Cham, Springer International Publishing.

Lund, L. A. and G. B. Sherman (1998). "Duplication of the southern white rhinoceros (Ceratotherium simum simum) luteinizing hormone beta subunit gene." J Mol Endocrinol 21(1): 19-30.

Luo, S., Z. Wang, V. Kansara, D. Pal and A. K. Mitra (2008). "Activity of a sodium-dependent vitamin C transporter (SVCT) in MDCK-MDR1 cells and mechanism of ascorbate uptake." International journal of pharmaceutics 358(1-2): 168-176.

Lux-Battistelli, C. and D. Battistelli (2017). "Latent scurvy with tiredness and leg pain in alcoholics: An underestimated disease three case reports." Medicine 96(47): e8861.

Mackenzie, B., A. C. Illing and M. A. Hediger (2008). "Transport model of the human Na+-coupled L-ascorbic acid (vitamin C) transporter SVCT1." Am J Physiol Cell Physiol 294(2): C451-459.

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.

Macmillan, H. A., C. M. Williams, J. F. Staples and B. J. Sinclair (2012). "Metabolism and energy supply below the critical thermal minimum of a chill-susceptible insect." J Exp Biol 215(Pt 8): 1366-1372.

Massie, H. R., M. E. Shumway, S. J. Whitney, S. M. Sternick and V. R. Aiello (1991). "Ascorbic acid in Drosophila and changes during aging." Exp Gerontol 26(5): 487-494.

May, J. M. (2011). "The SLC23 family of ascorbate transporters: ensuring that you get and keep your daily dose of vitamin C." Br J Pharmacol 164(7): 1793-1801.

May, J. M. and F. E. Harrison (2013). "Role of Vitamin C in the Function of the Vascular Endothelium." Antioxidants & Redox Signaling 19(17): 2068-2083.

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(6): 1039-1048.

Moreau, R. and K. Dabrowski (1998). "Body pool and synthesis of ascorbic acid in adult sea lamprey (Petromyzon marinus): An agnathan fish with gulonolactone oxidase activity." Proceedings of the National Academy of Sciences of the United States of America 95(17): 10279-10282.

Moreau, R. and K. Dabrowski (2005). "Biosynthesis of ascorbic acid by extant actinopterygians." Journal of Fish Biology 57(3): 733-745.

Moreno, K. X., C. E. Harrison, M. E. Merritt, Z. Kovacs, C. R. Malloy and A. D. Sherry (2017). "Hyperpolarized delta-[1-(13) C] gluconolactone as a probe of the pentose phosphate pathway." NMR Biomed 30(6).

Moriyama, Y. and K. Koshiba-Takeuchi (2018). "Significance of whole-genome duplications on the emergence of evolutionary novelties." Briefings in Functional Genomics: ely007-ely007.

Muñoz, A., M. Villagrán, P. Guzmán, C. Solíz, M. Gatica, C. Aylwin, K. Sweet, M. Maldonado, E. Escobar, A. M. Reyes, J. R. Toledo, O. Sánchez, S. A. Oñate, J. Carlos Vera and C. I. Rivas (2015). "Cis-regulatory elements involved in species-specific transcriptional regulation of the SVCT1 gene in rat and human hepatoma cells." Free radical biology & medicine 85: 183-196.

Newell, P. D., J. M. Chaston, Y. Wang, N. J. Winans, D. R. Sannino, A. C. Wong, A. J. Dobson, J. Kagle and A. E. Douglas (2014). "In vivo function and comparative genomic analyses of the Drosophila gut microbiota identify candidate symbiosis factors." Front Microbiol 5: 576.

Nishikimi, M., R. Fukuyama, S. Minoshima, N. Shimizu and K. Yagi (1994). "Cloning and chromosomal mapping of the human nonfunctional gene for L-gulono-gamma-lactone oxidase, the enzyme for L-ascorbic acid biosynthesis missing in man." J Biol Chem 269(18): 13685-13688.

Nishikimi, M., T. Kawai and K. Yagi (1992). "Guinea pigs possess a highly mutated gene for L-gulono-gamma-lactone oxidase, the key enzyme for L-ascorbic acid biosynthesis missing in this species." J Biol Chem 267(30): 21967-21972.

Notredame, C., D. G. Higgins and J. Heringa (2000). "T-Coffee: A novel method for fast and accurate multiple sequence alignment." J Mol Biol 302(1): 205-217.

Nualart, F., L. Mack, A. García, P. Cisternas, E. R. Bongarzone, M. Heitzer, N. Jara, F. Martínez, L. Ferrada, F. Espinoza, V. Baeza and K. Salazar (2014). "Vitamin C Transporters, Recycling and the Bystander Effect in the Nervous System: SVCT2 versus Gluts." Journal of stem cell research & therapy 4(5): 209.

Odamaki, T., K. Kato, H. Sugahara, N. Hashikura, S. Takahashi, J.-z. Xiao, F. Abe and R. Osawa (2016). "Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study." BMC Microbiology 16: 90.

Ohta, Y. and M. Nishikimi (1999). "Random nucleotide substitutions in primate nonfunctional gene for L-gulono-gamma-lactone oxidase, the missing enzyme in L-ascorbic acid biosynthesis." Biochim Biophys Acta 1472(1-2): 408-411.

Panoff, J. M., D. Corroler, B. Thammavongs and P. Boutibonnes (1997). "Differentiation between cold shock proteins and cold acclimation proteins in a mesophilic gram-positive bacterium, Enterococcus faecalis JH2-2." Journal of Bacteriology 179(13): 4451-4454.

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." Arch Biochem Biophys 569: 32-44.

Peterson, K. J., J. B. Lyons, K. S. Nowak, C. M. Takacs, M. J. Wargo and M. A. McPeek (2004). "Estimating metazoan divergence times with a molecular clock." Proceedings of the National Academy of Sciences of the United States of America 101(17): 6536.

Phadtare, S., J. Alsina and M. Inouye (1999). "Cold-shock response and cold-shock proteins." Curr Opin Microbiol 2(2): 175-180.

Plazzi, F., A. Ceregato, M. Taviani and M. Passamonti (2011). "A Molecular Phylogeny of Bivalve Mollusks: Ancient Radiations and Divergences as Revealed by Mitochondrial Genes." PLOS ONE 6(11): e27147.

Podmore, I. D., H. R. Griffiths, K. E. Herbert, N. Mistry, P. Mistry and J. Lunec (1998). "Vitamin C exhibits pro-oxidant properties." Nature 392: 559.

Pollet, N. and A. Mazabraud (2006). "Insights from Xenopus genomes." Genome Dyn 2: 138-153.

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(7198): 1064-1071.

Qiu, S., L. Li, E. J. Weeber and J. M. May (2007). "Ascorbate transport by primary cultured neurons and its role in neuronal function and protection against excitotoxicity." J Neurosci Res 85(5): 1046-1056.

Rajan, D. P., W. Huang, B. Dutta, L. D. Devoe, F. H. Leibach, V. Ganapathy and P. D. Prasad (1999). "Human placental sodium-dependent vitamin C transporter (SVCT2): molecular cloning and transport function." Biochem Biophys Res Commun 262(3): 762-768.

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." J Integr Bioinform 9(3): 200.

Ribeiro, C. and B. J. Dickson (2010). "Sex Peptide Receptor and Neuronal TOR/S6K Signaling Modulate Nutrient Balancing in Drosophila." Current Biology 20(11): 1000-1005.

Ritter, J. K. (2000). "Roles of glucuronidation and UDP-glucuronosyltransferases in xenobiotic bioactivation reactions." Chem Biol Interact 129(1-2): 171-193.

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(3): 539-542.

Sannino, D. R., A. J. Dobson, K. Edwards, E. R. Angert and N. Buchon (2018). "The Drosophila melanogaster Gut Microbiota Provisions Thiamine to Its Host." MBio 9(2).

Sastre, J., F. V. Pallardo, J. Garcia de la Asuncion and J. Vina (2000). "Mitochondria, oxidative stress and aging." Free Radic Res 32(3): 189-198.

Savini, I., M. V. Catani, R. Arnone, A. Rossi, G. Frega, D. Del Principe and L. Avigliano (2007). "Translational control of the ascorbic acid transporter SVCT2 in human platelets." Free Radic Biol Med 42(5): 608-616.

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(3): 347-355.

Sémon, M. and K. H. Wolfe (2008). "Preferential subfunctionalization of slow-evolving genes after allopolyploidization in Xenopus laevis." Proceedings of the National Academy of Sciences of the United States of America 105(24): 8333-8338.

Session, A. M., Y. Uno, T. Kwon, J. A. Chapman, A. Toyoda, S. Takahashi, A. Fukui, A. Hikosaka, A. Suzuki, M. Kondo, S. J. van Heeringen, I. Quigley, S. Heinz, H. Ogino, H. Ochi, U. Hellsten, J. B. Lyons, O. Simakov, N. Putnam, J. Stites, Y. Kuroki, T. Tanaka, T. Michiue, M. Watanabe, O. Bogdanovic, R. Lister, G. Georgiou, S. S. Paranjpe, I. van Kruijsbergen, S. Shu, J. Carlson, T. Kinoshita, Y. Ohta, S. Mawaribuchi, J. Jenkins, J. Grimwood, J. Schmutz, T. Mitros, S. Mozaffari, Y. Suzuki, Y. Haramoto, T. S. Yamamoto, C. Takagi, R. Heald, K. Miller, C. Haudenschild, J. Kitzman, T. Nakayama, Y. Izutsu, J. Robert, J. Fortriede, K. Burns, V. Lotay, K. Karimi, Y. Yasuoka, D. S. Dichmann, M. F. Flajnik, D. W. Houston, J. Shendure, L. DuPasquier, P. D. Vize, A. M. Zorn, M. Ito, E. Marcotte, J. B. Wallingford, Y. Ito, M. Asashima, N. Ueno, Y. Matsuda, G. J. C. Veenstra, A. Fujiyama, R. M. Harland, M. Taira and D. S. Rokhsar

(2016). "Genome evolution in the allotetraploid frog Xenopus laevis." Nature 538(7625): 336-343.

Shigeoka, S., Y. Nakano and S. Kitaoka (1979). "The biosynthetic pathway of L-ascorbic acid in Euglena gracilis." Journal of Nutritional Science and Vitaminology 25(4): 299-307.

Simhadri, R. K., E. M. Fast, R. Guo, M. J. Schultz, N. Vaisman, L. Ortiz, J. Bybee, B. E. Slatko and H. M. Frydman (2017). "The Gut Commensal Microbiome of Drosophila melanogaster Is Modified by the Endosymbiont Wolbachia" mSphere 2(5).

Smirnoff, N. (2001). "L-ascorbic acid biosynthesis." Vitam Horm 61: 241-266.

Smirnoff, N. (2018). "Ascorbic acid metabolism and functions: A comparison of plants and mammals." Free Radical Biology and Medicine.

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." Nat Med 8(5): 514-517.

Subramanian, V. S., P. Srinivasan, A. J. Wildman, J. S. Marchant and H. M. Said (2017). "Molecular mechanism(s) involved in differential expression of vitamin C transporters along the intestinal tract." Am J Physiol Gastrointest Liver Physiol 312(4): G340-g347.

Szarka, A. and T. Lőrincz (2014). "The role of ascorbate in protein folding." Protoplasma 251(3): 489-497.

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

Taylor, J. D., S. T. Williams, E. A. Glover and P. Dyal (2007). "A molecular phylogeny of heterodont bivalves (Mollusca: Bivalvia: Heterodonta): new analyses of 18S and 28S rRNA genes." Zoologica Scripta 36(6): 587-606.

Taylor, J. S., I. Braasch, T. Frickey, A. Meyer and Y. Van de Peer (2003). "Genome Duplication, a Trait Shared by 22,000 Species of Ray-Finned Fish." Genome Research 13(3): 382-390.

Taylor, J. S., Y. Van de Peer, I. Braasch and A. Meyer (2001). "Comparative genomics provides evidence for an ancient genome duplication event in fish." Philos Trans R Soc Lond B Biol Sci 356(1414): 1661-1679.

Teeling, E. C., M. S. Springer, O. Madsen, P. Bates, J. O'Brien S and W. J. Murphy (2005). "A molecular phylogeny for bats illuminates biogeography and the fossil record." Science 307(5709): 580-584.

Thomas, W. R. (2010). "Geography of house dust mite allergens." Asian Pac J Allergy Immunol 28(4): 211-224.

Traikovich, S. S. (1999). "Use of topical ascorbic acid and its effects on photodamaged skin topography." Archives of Otolaryngology–Head & Neck Surgery 125(10): 1091-1098.

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(6731): 70-75.

Tveden-Nyborg, P., L. K. Johansen, Z. Raida, C. K. Villumsen, J. O. Larsen and J. Lykkesfeldt (2009). "Vitamin C deficiency in early postnatal life impairs spatial memory and reduces the number of hippocampal neurons in guinea pigs." Am J Clin Nutr 90(3): 540-546.

Tveden-Nyborg, P. and J. Lykkesfeldt (2009). "Does vitamin C deficiency result in impaired brain development in infants?" Redox Rep 14(1): 2-6.

Uchizono, S., Y. Tabuki, N. Kawaguchi, T. Tanimura and T. Q. Itoh (2017). "Mated Drosophila melanogaster females consume more amino acids during the dark phase." PLOS ONE 12(2): e0172886.

Uyeno, T. and G. R. Smith (1972). "Tetraploid origin of the karyotype of catostomid fishes." Science 175(4022): 644-646.

Vandepoele, K., W. De Vos, J. S. Taylor, A. Meyer and Y. Van de Peer (2004). "Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates." Proc Natl Acad Sci U S A 101(6): 1638-1643.

Vera, J. C., C. I. Rivas, F. V. Velasquez, R. H. Zhang, Concha, II and D. W. Golde (1995). "Resolution of the facilitated transport of dehydroascorbic acid from its intracellular accumulation as ascorbic acid." J Biol Chem 270(40): 23706-23712.

Wägele, J. W. and C. Mayer (2007). "Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects." BMC Evolutionary Biology 7: 147-147.

Wang, J.-T., J.-T. Li, X.-F. Zhang and X.-W. Sun (2012). "Transcriptome analysis reveals the time of the fourth round of genome duplication in common carp (Cyprinus carpio)." BMC Genomics 13(1): 96.

Wang, Y., B. Mackenzie, H. Tsukaguchi, S. Weremowicz, C. C. Morton and M. A. Hediger (2000). "Human vitamin C (L-ascorbic acid) transporter SVCT1." Biochem Biophys Res Commun 267(2): 488-494.

Welch, R. W., P. Bergsten, J. D. Butler and M. Levine (1993). "Ascorbic acid accumulation and transport in human fibroblasts." Biochem J 294 (Pt 2): 505-510.

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." J Biol Chem 270(21): 12584-12592.

Wells, W. W., D. P. Xu, Y. F. Yang and P. A. Rocque (1990). "Mammalian thioltransferase (glutaredoxin) and protein disulfide isomerase have dehydroascorbate reductase activity." J Biol Chem 265(26): 15361-15364.

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.

Wilson, J. X. (2005). "Regulation of vitamin C transport." Annu Rev Nutr 25: 105-125.

Winkler, B. S., S. M. Orselli and T. S. Rex (1994). "The redox couple between glutathione and ascorbic acid: a chemical and physiological perspective." Free Radic Biol Med 17(4): 333-349.

Wolf, Y. I., P. S. Novichkov, G. P. Karev, E. V. Koonin and D. J. Lipman (2009). "The universal distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent ages." Proceedings of the National Academy of Sciences of the United States of America 106(18): 7273-7280.

Wolfe, K. H. (2001). "Yesterday's polyploids and the mystery of diploidization." Nat Rev Genet 2(5): 333-341.

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(19): 4435-4445.

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." J Biol Chem 285(9): 6522-6531.

Yang, H. (2013). "Conserved or Lost: Molecular Evolution of the Key Gene GULO in Vertebrate Vitamin C Biosynthesis." Biochemical Genetics 51(5): 413-425.

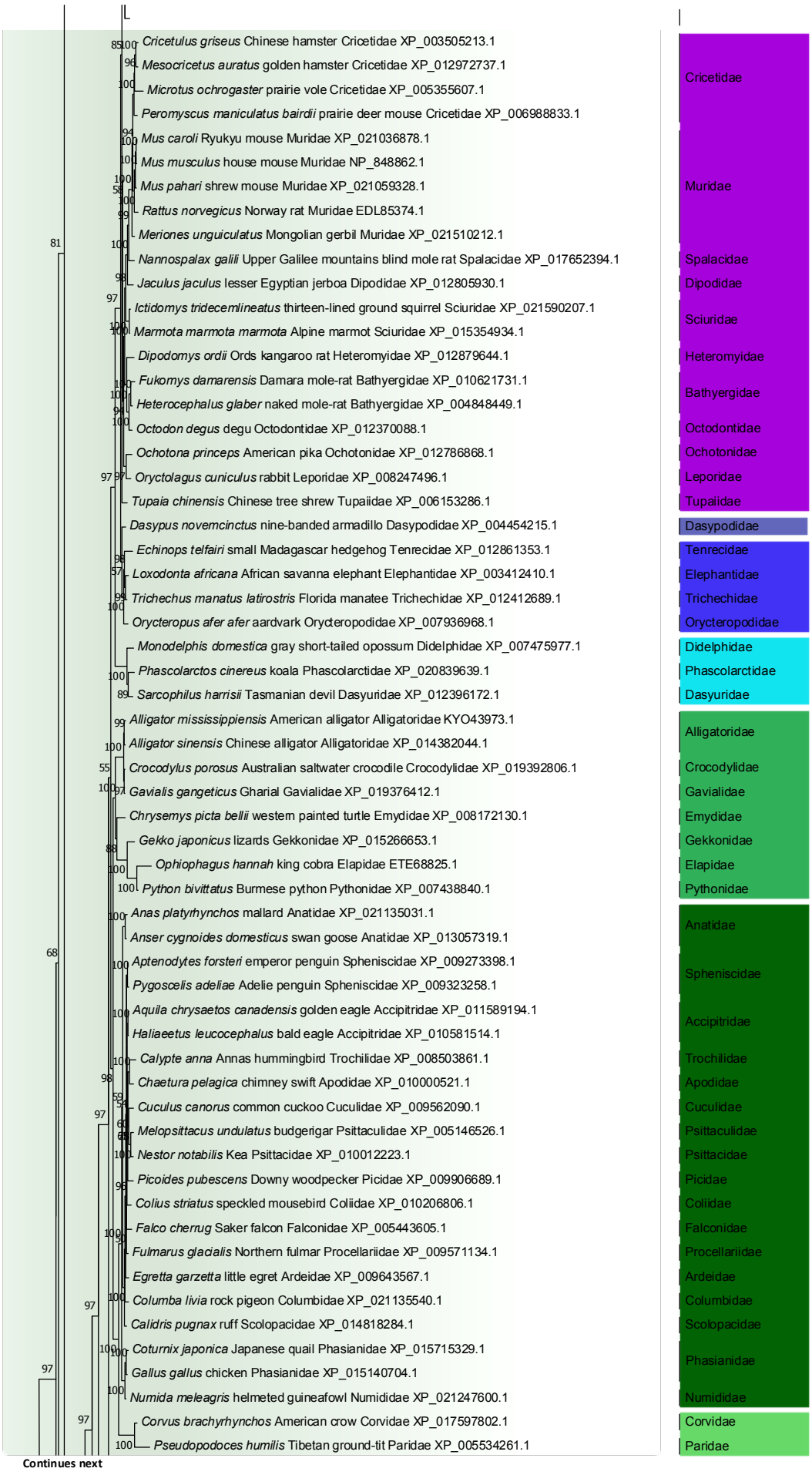
Zapata, F., N. G. Wilson, M. Howison, S. C. Andrade, K. M. Jorger, M. Schrodl, F. E. Goetz, G. Giribet and C. W. Dunn (2014). "Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda." Proc Biol Sci 281(1794): 20141739.

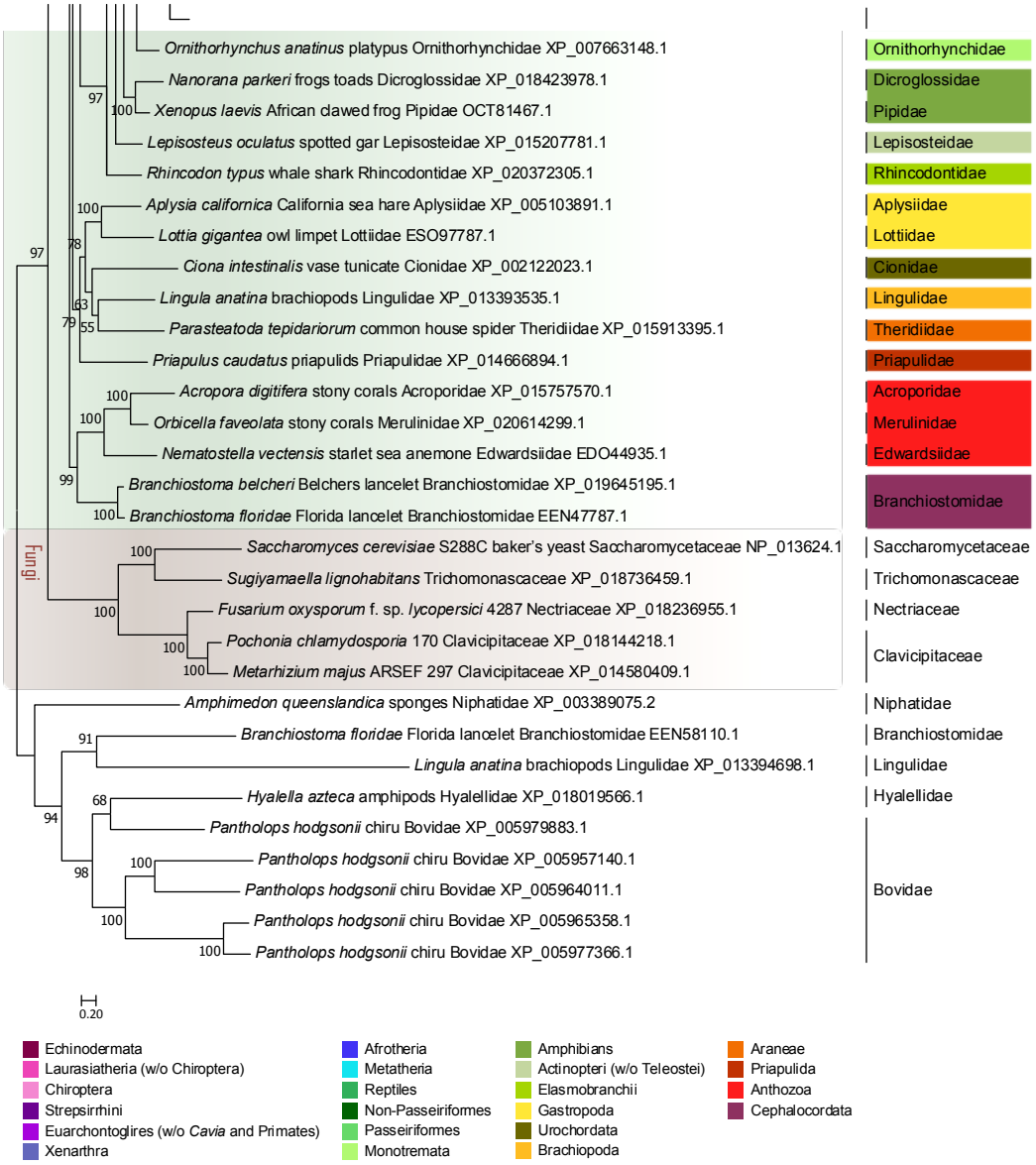
Zhang, P. Y., X. Xu and X. C. Li (2014). "Cardiovascular diseases: oxidative damage and antioxidant protection." Eur Rev Med Pharmacol Sci 18(20): 3091-3096.

Zhang, X., Y. Zhang, X. Zheng, Y. Kuang, Z. Zhao, L. Zhao, C. Li, L. Jiang, D. Cao, C. Lu, P. Xu and X. Sun (2013). "A consensus linkage map provides insights on genome character and evolution in common carp (Cyprinus carpio L.)." Mar Biotechnol (NY) 15(3): 275-312.

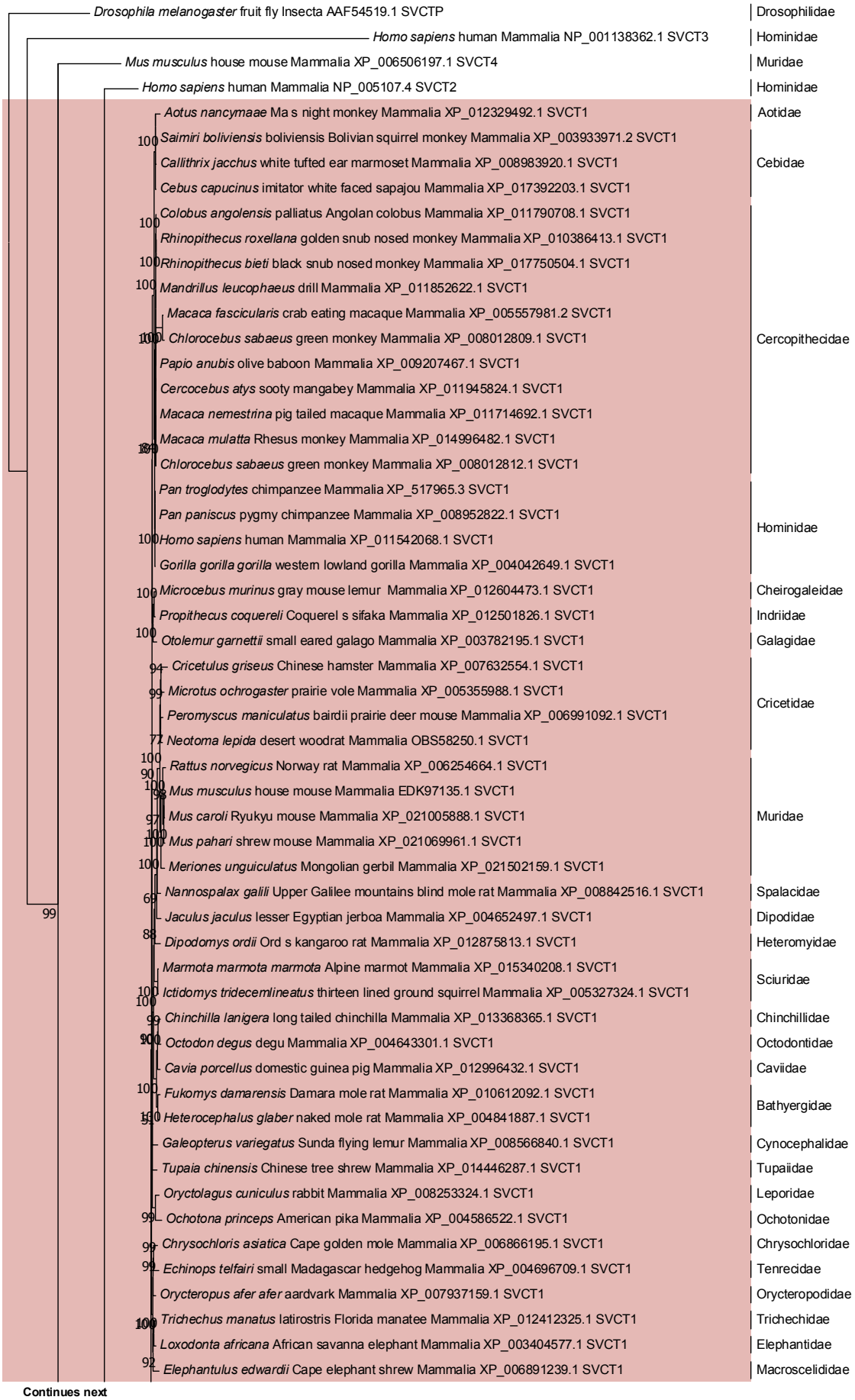
VI. Supplementary Data

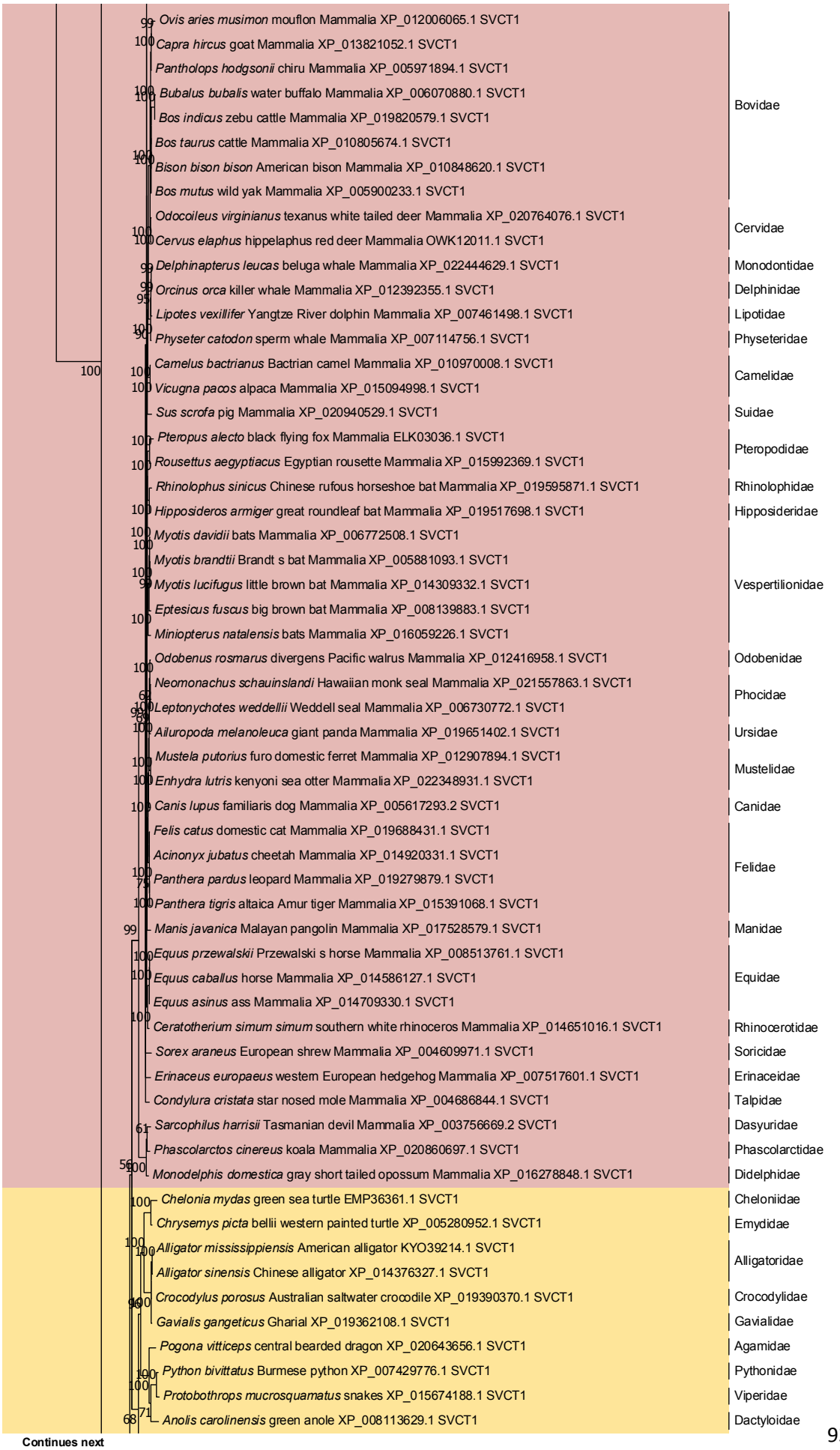
Continues next page


Cricetulus griseus Chinese hamster Cricetidae XP_003505213.1
Mesocricetus auratus golden hamster Cricetidae XP_012972737.1
Microtus ochrogaster prairie vole Cricetidae XP_005355607.1
Peromyscus maniculatus bairdii prairie deer mouse Cricetidae XP_006988833.1
Mus caroli Ryukyu mouse Muridae XP_021036878.1
Mus musculus house mouse Muridae NP_848862.1
Mus pahari shrew mouse Muridae XP_021059328.1
Rattus norvegicus Norway rat Muridae EDL85374.1
Meriones unguiculatus Mongolian gerbil Muridae XP_021510212.1
Nannospalax galili Upper Galilee mountains blind mole rat Spalacidae XP_017652394.1
Jaculus jaculus lesser Egyptian jerboa Dipodidae XP_012805930.1
Ictidomys tridecemlineatus thirteen-lined ground squirrel Sciuridae XP_021590207.1
Marmota marmota marmota Alpine marmot Sciuridae XP_015354934.1
Dipodomys ordii Ords kangaroo rat Heteromyidae XP_012879644.1
Fukomys damarensis Damara mole-rat Bathyergidae XP_010621731.1
Heterocephalus glaber naked mole-rat Bathyergidae XP_004848449.1
Octodon degus degu Octodontidae XP_012370088.1
Ochotona princeps American pika Ochotonidae XP_012786868.1
Oryctolagus cuniculus rabbit Leporidae XP_008247496.1
Tupaia chinensis Chinese tree shrew Tupaiidae XP_006153286.1
Dasypus novemcinctus nine-banded armadillo Dasypodidae XP_004454215.1
Echinops telfairi small Madagascar hedgehog Tenrecidae XP_012861353.1
Loxodonta africana African savanna elephant Elephantidae XP_003412410.1
Trichechus manatus latirostris Florida manatee Trichechidae XP_012412689.1
Orycteropus afer afer aardvark Orycteropodidae XP_007936968.1
Monodelphis domestica gray short-tailed opossum Didelphidae XP_007475977.1
Phascolarctos cinereus koala Phascolarctidae XP_020839639.1
Sarcophilus harrisii Tasmanian devil Dasyuridae XP_012396172.1
Alligator mississippiensis American alligator Alligatoridae KYO43973.1
Alligator sinensis Chinese alligator Alligatoridae XP_014382044.1
Crocodylus porosus Australian saltwater crocodile Crocodylidae XP_019392806.1
Gavialis gangeticus Gharial Gavialidae XP_019376412.1
Chrysemys picta bellii western painted turtle Emydidae XP_008172130.1
Gekko japonicus lizards Gekkonidae XP_015266653.1
Ophiophagus hannah king cobra Elapidae ETE68825.1
Python bivittatus Burmese python Pythonidae XP_007438840.1
Anas platyrhynchos mallard Anatidae XP_021135031.1
Anser cygnoides domesticus swan goose Anatidae XP_013057319.1
Aptenodytes forsteri emperor penguin Spheniscidae XP_009273398.1
Pygoscelis adeliae Adelie penguin Spheniscidae XP_009323258.1
Aquila chrysaetos canadensis golden eagle Accipitridae XP_011589194.1
Haliaeetus leucocephalus bald eagle Accipitridae XP_010581514.1
Calypte anna Annas hummingbird Trochilidae XP_008503861.1
Chaetura pelagica chimney swift Apodidae XP_010000521.1
Cuculus canorus common cuckoo Cuculidae XP_009562090.1
Melopsittacus undulatus budgerigar Psittaculidae XP_005146526.1
Nestor notabilis Kea Psittacidae XP_010012223.1
Picoides pubescens Downy woodpecker Picidae XP_009906689.1
Colius striatus speckled mousebird Coliidae XP_010206806.1
Falco cherrug Saker falcon Falconidae XP_005443605.1
Fulmarus glacialis Northern fulmar Procellariidae XP_009571134.1
Egretta garzetta little egret Ardeidae XP_009643567.1
Columba livia rock pigeon Columbidae XP_021135540.1
Calidris pugnax ruff Scolopacidae XP_014818284.1
Coturnix japonica Japanese quail Phasianidae XP_015715329.1
Gallus gallus chicken Phasianidae XP_015140704.1
Numida meleagris helmeted guineafowl Numididae XP_021247600.1
Corvus brachyrhynchos American crow Corvidae XP_017597802.1
Pseudopodoces humilis Tibetan ground-tit Paridae XP_005534261.1
Cricetidae
Muridae
Spalacidae
Dipodidae
Sciuridae
Heteromyidae
Bathyergidae
Octodontidae
Ochotonidae
Leporidae
Tupaiidae
Dasypodidae
Tenrecidae
Elephantidae
Trichechidae
Orycteropodidae
Didelphidae
Phascolarctidae
Dasyuridae
Alligatoridae
Crocodylidae
Gavialidae
Emydidae
Gekkonidae
Elapidae
Pythonidae
Anatidae
Spheniscidae
Accipitridae
Trochilidae
Apodidae
Cuculidae
Psittaculidae
Psittacidae
Picidae
Coliidae
Falconidae
Procellariidae
Ardeidae
Columbidae
Scolopacidae
Phasianidae
Numididae
Corvidae
Paridae
Continues next page



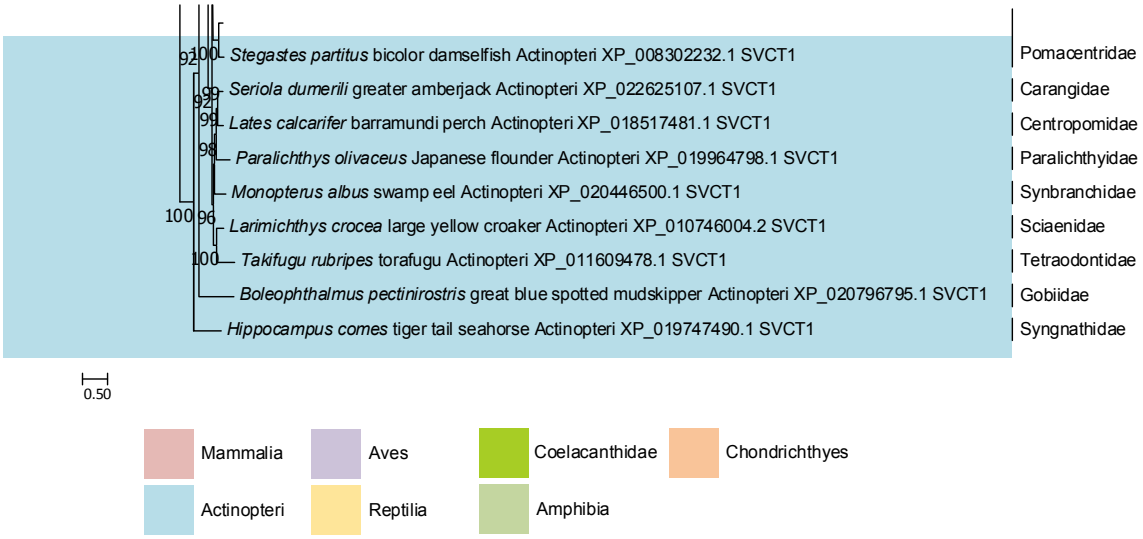
Supplementary Figure 1 - Animal L-gulonolactone oxidase (*GULO*) Bayesian phylogeny. Five *GULO* sequences from Fungi species were used to help rooting the tree. The family name is shown next to the species name. Families are coloured according to the relevant taxonomic groups analysed in this work, ranging from Deuterostomes (cold colors) to Protostomes and non-bilaterians (hot colors). Image and description are represented as seen in López-Fernández *et al.* (2018).


Drosophila melanogaster fruit fly Insecta AAF54519.1 SVCTP
Homo sapiens human Mammalia NP_001138362.1 SVCT3
Mus musculus house mouse Mammalia XP_006506197.1 SVCT4
Homo sapiens human Mammalia NP_005107.4 SVCT2
Aotus nancymaae Ma s night monkey Mammalia XP_012329492.1 SVCT1
Saimiri boliviensis boliviensis Bolivian squirrel monkey Mammalia XP_003933971.2 SVCT1
Callithrix jacchus white tufted ear marmoset Mammalia XP_008983920.1 SVCT1
Cebus capucinus imitator white faced sapajou Mammalia XP_017392203.1 SVCT1
Colobus angolensis palliatus Angolan colobus Mammalia XP_011790708.1 SVCT1
Rhinopithecus roxellana golden snub nosed monkey Mammalia XP_010386413.1 SVCT1
Rhinopithecus bieti black snub nosed monkey Mammalia XP_017750504.1 SVCT1
Mandrillus leucophaeus drill Mammalia XP_011852622.1 SVCT1
Macaca fascicularis crab eating macaque Mammalia XP_005557981.2 SVCT1
Chlorocebus sabaeus green monkey Mammalia XP_008012809.1 SVCT1
Papio anubis olive baboon Mammalia XP_009207467.1 SVCT1
Cercocebus atys sooty mangabey Mammalia XP_011945824.1 SVCT1
Macaca nemestrina pig tailed macaque Mammalia XP_011714692.1 SVCT1
Macaca mulatta Rhesus monkey Mammalia XP_014996482.1 SVCT1
Chlorocebus sabaeus green monkey Mammalia XP_008012812.1 SVCT1
Pan troglodytes chimpanzee Mammalia XP_517965.3 SVCT1
Pan paniscus pygmy chimpanzee Mammalia XP_008952822.1 SVCT1
Homo sapiens human Mammalia XP_011542068.1 SVCT1
Gorilla gorilla gorilla western lowland gorilla Mammalia XP_004042649.1 SVCT1
Microcebus murinus gray mouse lemur Mammalia XP_012604473.1 SVCT1
Propithecus coquereli Coquerel s sifaka Mammalia XP_012501826.1 SVCT1
Otolemur garnettii small eared galago Mammalia XP_003782195.1 SVCT1
Cricetulus griseus Chinese hamster Mammalia XP_007632554.1 SVCT1
Microtus ochrogaster prairie vole Mammalia XP_005355988.1 SVCT1
Peromyscus maniculatus bairdii prairie deer mouse Mammalia XP_006991092.1 SVCT1
Neotoma lepida desert woodrat Mammalia OBS58250.1 SVCT1
Rattus norvegicus Norway rat Mammalia XP_006254664.1 SVCT1
Mus musculus house mouse Mammalia EDK97135.1 SVCT1
Mus caroli Ryukyu mouse Mammalia XP_021005888.1 SVCT1
Mus pahari shrew mouse Mammalia XP_021069961.1 SVCT1
Meriones unguiculatus Mongolian gerbil Mammalia XP_021502159.1 SVCT1
Nannospalax galili Upper Galilee mountains blind mole rat Mammalia XP_008842516.1 SVCT1
Jaculus jaculus lesser Egyptian jerboa Mammalia XP_004652497.1 SVCT1
Dipodomys ordii Ord s kangaroo rat Mammalia XP_012875813.1 SVCT1
Marmota marmota marmota Alpine marmot Mammalia XP_015340208.1 SVCT1
Ictidomys tridecemlineatus thirteen lined ground squirrel Mammalia XP_005327324.1 SVCT1
Chinchilla lanigera long tailed chinchilla Mammalia XP_013368365.1 SVCT1
Octodon degus degu Mammalia XP_004643301.1 SVCT1
Cavia porcellus domestic guinea pig Mammalia XP_012996432.1 SVCT1
Fukomys damarensis Damara mole rat Mammalia XP_010612092.1 SVCT1
Heterocephalus glaber naked mole rat Mammalia XP_004841887.1 SVCT1
Galeopterus variegatus Sunda flying lemur Mammalia XP_008566840.1 SVCT1
Tupaia chinensis Chinese tree shrew Mammalia XP_014446287.1 SVCT1
Oryctolagus cuniculus rabbit Mammalia XP_008253324.1 SVCT1
Ochotona princeps American pika Mammalia XP_004586522.1 SVCT1
Chrysochloris asiatica Cape golden mole Mammalia XP_006866195.1 SVCT1
Echinops telfairi small Madagascar hedgehog Mammalia XP_004696709.1 SVCT1
Orycteropus afer afer aardvark Mammalia XP_007937159.1 SVCT1
Trichechus manatus latirostris Florida manatee Mammalia XP_012412325.1 SVCT1
Loxodonta africana African savanna elephant Mammalia XP_003404577.1 SVCT1
Elephantulus edwardii Cape elephant shrew Mammalia XP_006891239.1 SVCT1
Drosophilidae
Hominidae
Muridae
Hominidae
Aotidae
Cebidae
Cercopithecidae
Hominidae
Cheirogaleidae
Indriidae
Galagidae
Cricetidae
Muridae
Spalacidae
Dipodidae
Heteromyidae
Sciuridae
Chinchillidae
Octodontidae
Caviidae
Bathyergidae
Cynocephalidae
Tupaiidae
Leporidae
Ochotonidae
Chrysochloridae
Tenrecidae
Orycteropodidae
Trichechidae
Elephantidae
Macroscelididae
Continues next page

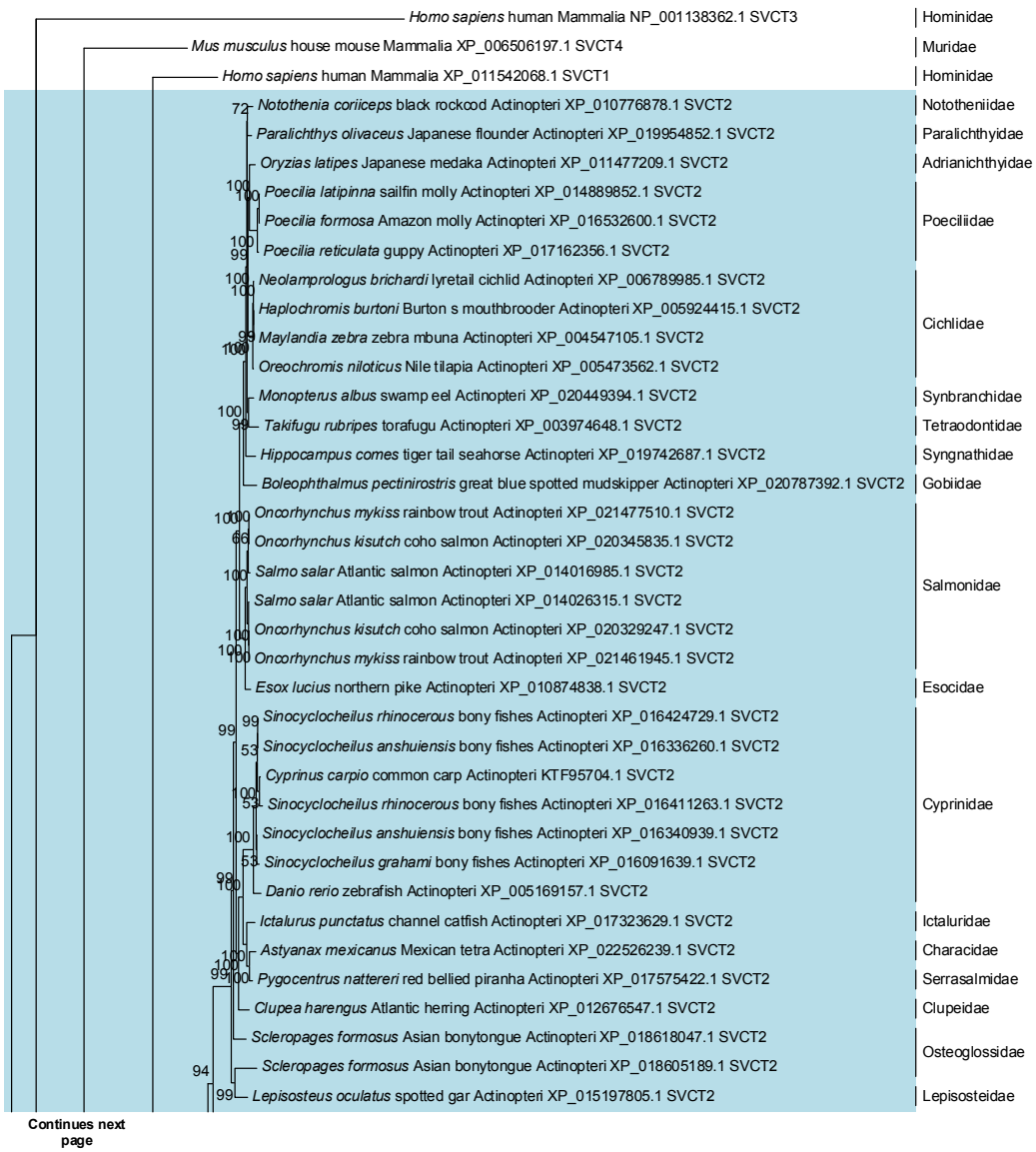


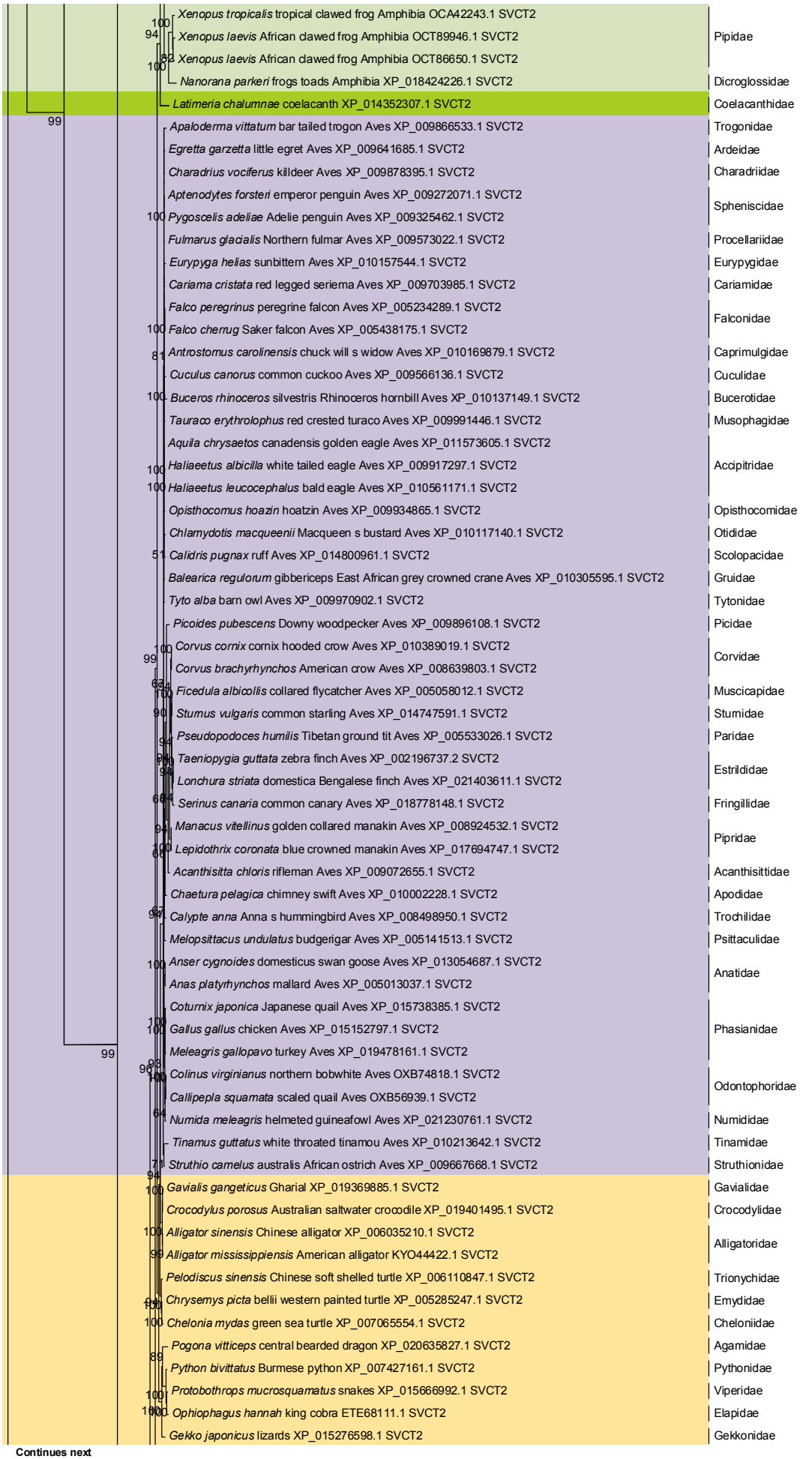
Continues next page


Rhincodon typus whale shark Chondrichthyes XP_020385431.1 SVCT1
Rhincodontidae
Callorhinchus milii elephant shark Chondrichthyes XP_007904441.1 SVCT1
Callorhinchidae
Latimeria chalumnae coelacanth XP_005989382.1 SVCT1
Coelacanthidae
Xenopus laevis African clawed frog Amphibia OCT88298.1 SVCT1
Xenopus tropicalis tropical clawed frog Amphibia OCA39850.1 SVCT1
Pipidae
Columba livia rock pigeon Aves XP_005503753.1 SVCT1
Patagioenas fasciata monilis band tailed pigeon Aves OPJ76386.1 SVCT1
Columbidae
Taeniopygia guttata zebra finch Aves XP_002186880.2 SVCT1
Estrildidae
Sturnus vulgaris common starling Aves XP_014745094.1 SVCT1
Sturnidae
Ficedula albicollis collared flycatcher Aves XP_016157217.1 SVCT1
Muscicapidae
Pseudopodoces humilis Tibetan ground tit Aves XP_014116675.1 SVCT1
Parus major Great Tit Aves XP_015497804.1 SVCT1
Paridae
Lonchura striata domestica Bengalese finch Aves XP_021398789.1 SVCT1
Estrildidae
Lepidothrix coronata blue crowned manakin Aves XP_017685858.1 SVCT1
Pipridae
Falco peregrinus peregrine falcon Aves XP_013150686.1 SVCT1
Falconidae
Nipponia nippon crested ibis Aves XP_009467773.1 SVCT1
Threskiornithidae
Calidris pugnax ruff Aves XP_014811273.1 SVCT1
Scolopacidae
Calypte anna Anna s hummingbird Aves XP_008501645.1 SVCT1
Trochilidae
Coturnix japonica Japanese quail Aves XP_015731499.1 SVCT1
Phasianidae
Numida meleagris helmeted guineafowl Aves XP_021266381.1 SVCT1
Numididae
Gallus gallus chicken Aves XP_004944825.1 SVCT1
Meleagris gallopavo turkey Aves XP_010717159.1 SVCT1
Phasianidae
Salmo salar Atlantic salmon Actinopteri XP_014053616.1 SVCT1
Salmo salar Atlantic salmon Actinopteri XP_013992130.1 SVCT1
Oncorhynchus kisutch coho salmon Actinopteri XP_020341580.1 SVCT1
Oncorhynchus mykiss rainbow trout Actinopteri XP_021479885.1 SVCT1
Salmonidae
Esox lucius northern pike Actinopteri XP_010869526.1 SVCT1
Esocidae
Pygocentrus nattereri red bellied piranha Actinopteri XP_017554181.1 SVCT1
Serrasalmidae
Astyanax mexicanus Mexican tetra Actinopteri XP_022531816.1 SVCT1
Characidae
Ictalurus punctatus channel catfish Actinopteri XP_017348904.1 SVCT1
Ictaluridae
Danio rerio zebrafish Actinopteri NP_001166970.1 SVCT1
Sinocyclocheilus rhinocerous bony fishes Actinopteri XP_016400612.1 SVCT1
Sinocyclocheilus grahami bony fishes Actinopteri XP_016134207.1 SVCT1
Cyprinus carpio common carp Actinopteri XP_018975086.1 SVCT1
Cyprinus carpio common carp Actinopteri KTG36548.1 SVCT1
Cyprinidae
Clupea harengus Atlantic herring Actinopteri XP_012691172.1 SVCT1
Clupeidae
Scleropages formosus Asian bonytongue Actinopteri XP_018607742.1 SVCT1
Osteoglossidae
Cynoglossus semilaevis tongue sole Actinopteri XP_016897249.1 SVCT1
Cynoglossidae
Xiphophorus maculatus southern platyfish Actinopteri XP_014325925.1 SVCT1
Poecilia reticulata guppy Actinopteri XP_008400069.1 SVCT1
Poecilia latipinna sailfin molly Actinopteri XP_014889782.1 SVCT1
Poecilia formosa Amazon molly Actinopteri XP_007556555.1 SVCT1
Poecilia mexicana shortfin molly Actinopteri XP_014843880.1 SVCT1
Poeciliidae
Cyprinodon variegatus sheepshead minnow Actinopteri XP_015239150.1 SVCT1
Cyprinodontidae
Kryptolebias marmoratus mangrove Actinopteri XP_017296452.1 SVCT1
Rivulidae
Oryzias latipes Japanese medaka Actinopteri XP_004076777.2 SVCT1
Adrianichthyidae
Pundamilia nyererei bony fishes Actinopteri XP_013771426.1 SVCT1
Maylandia zebra zebra mbuna Actinopteri XP_014266807.1 SVCT1
Oreochromis niloticus Nile tilapia Actinopteri XP_003457032.1 SVCT1
Cichlidae
Acanthochromis polyacanthus spiny chromis Actinopteri XP_022076289.1 SVCT1
Pomacentridae
Continues next page

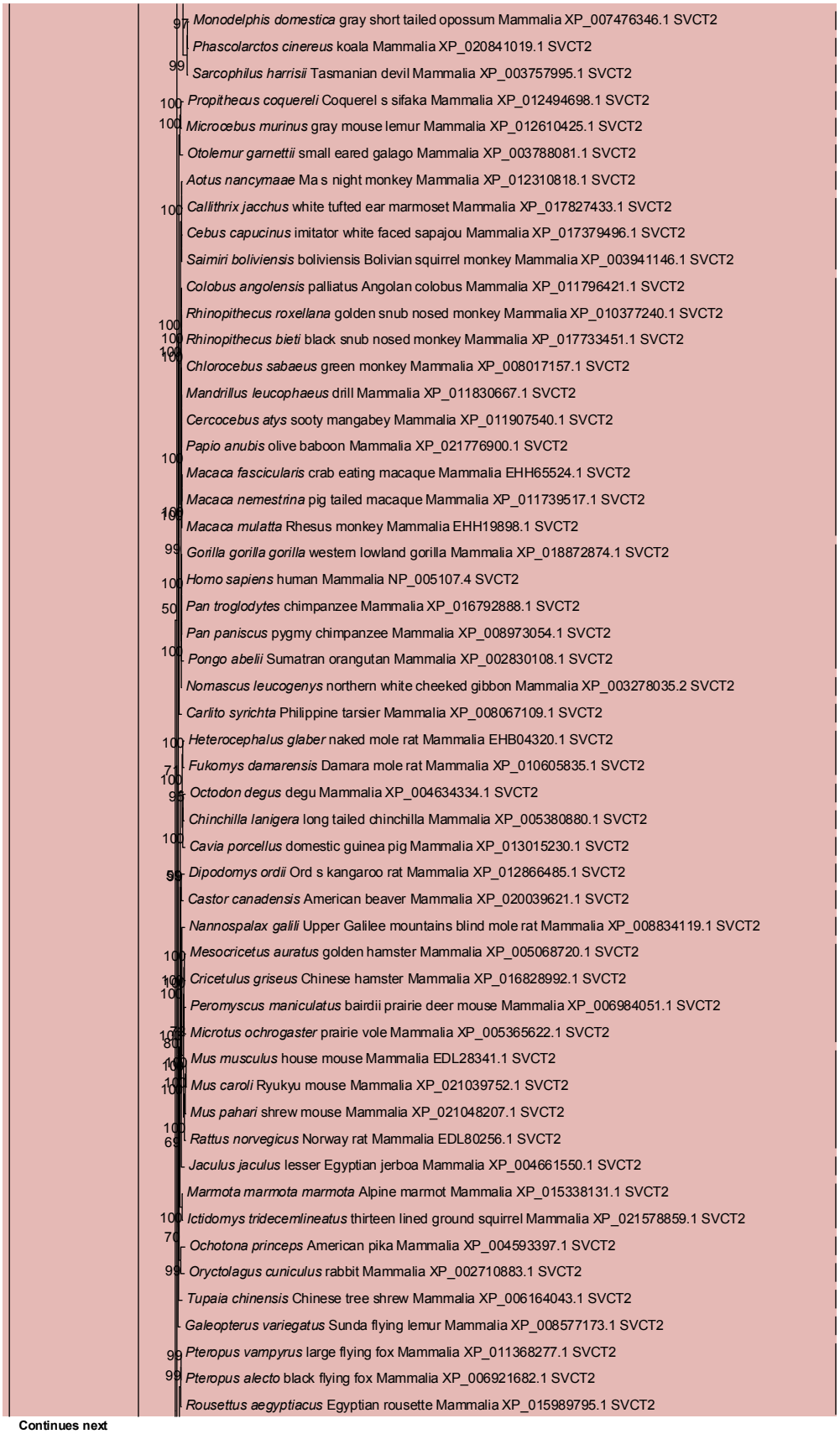


Supplementary Figure 2 - *Sodium-dependent Vitamin C transporter 1* (*SVCT1*) Bayesian phylogeny. The species family names are represented next to the corresponding sequences. Different taxonomic groups are highlighted with distinct colours: Mammalia in light red, Actinopteri in light blue, Aves in purple, Reptilia in yellow, Chondrichthyes in orange, Amphibia in light green and Coelacanthidae in green. The four outgroup sequences are not highlighted and can be seen at the top of the phylogeny.

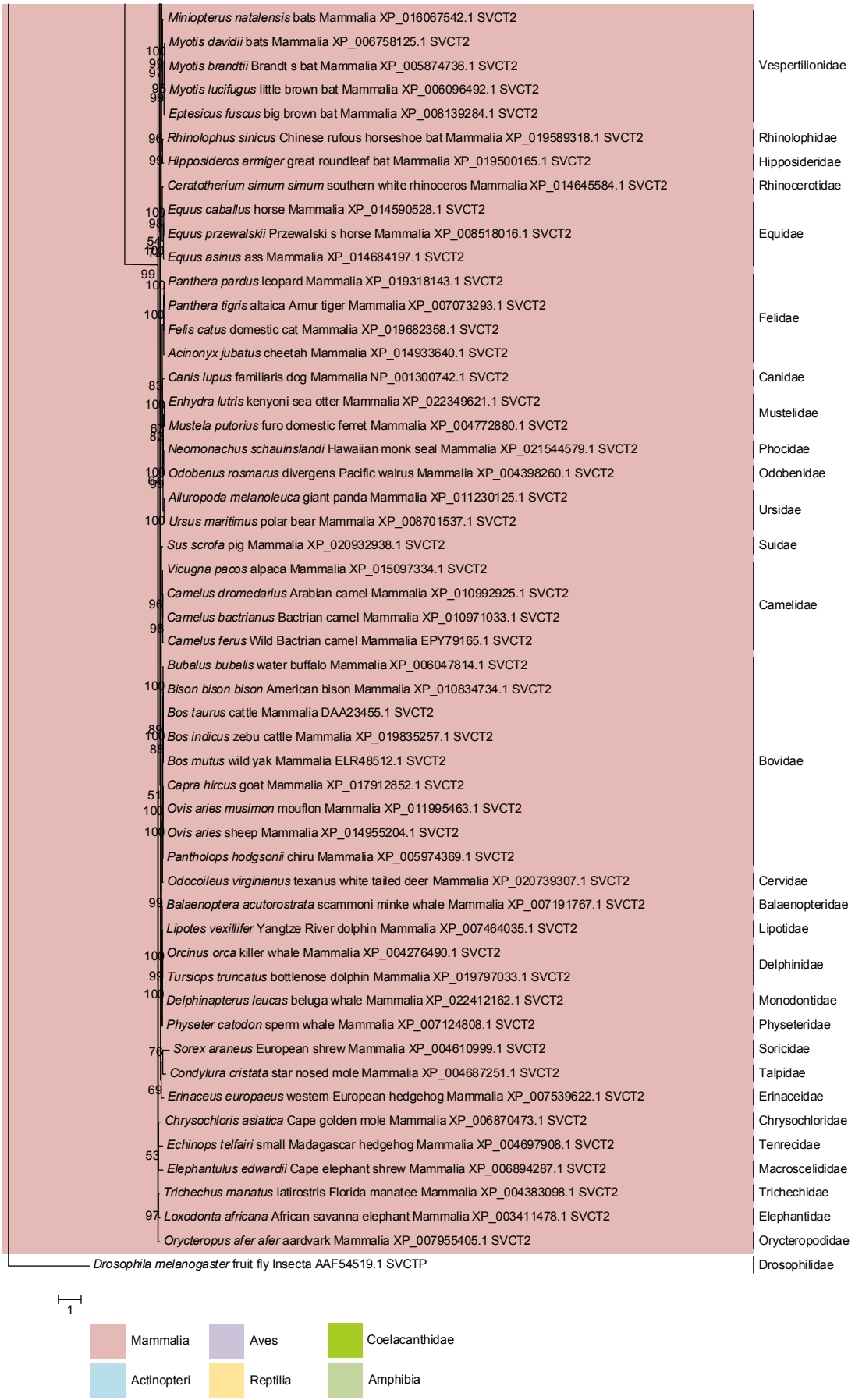




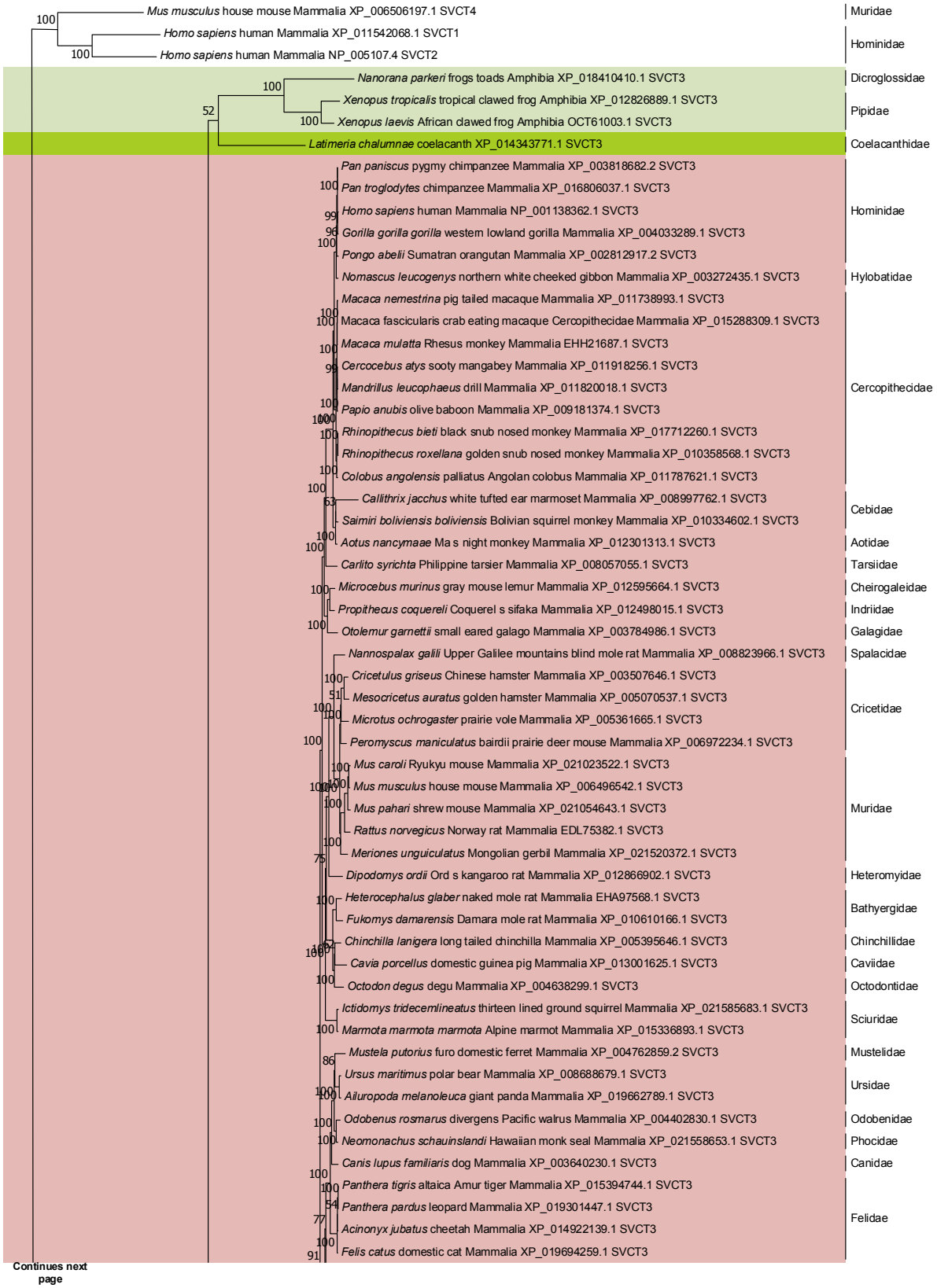
Continues next page

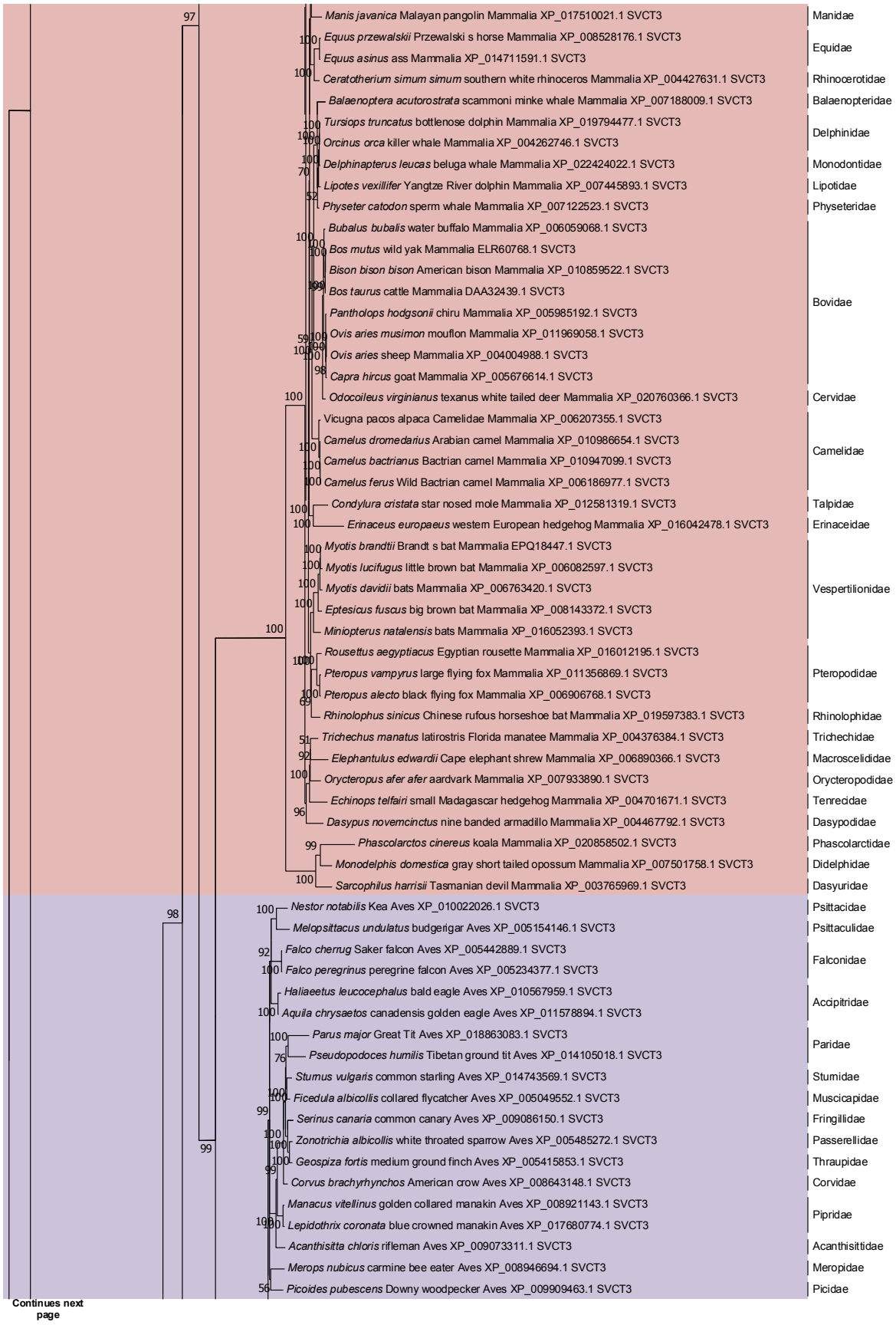


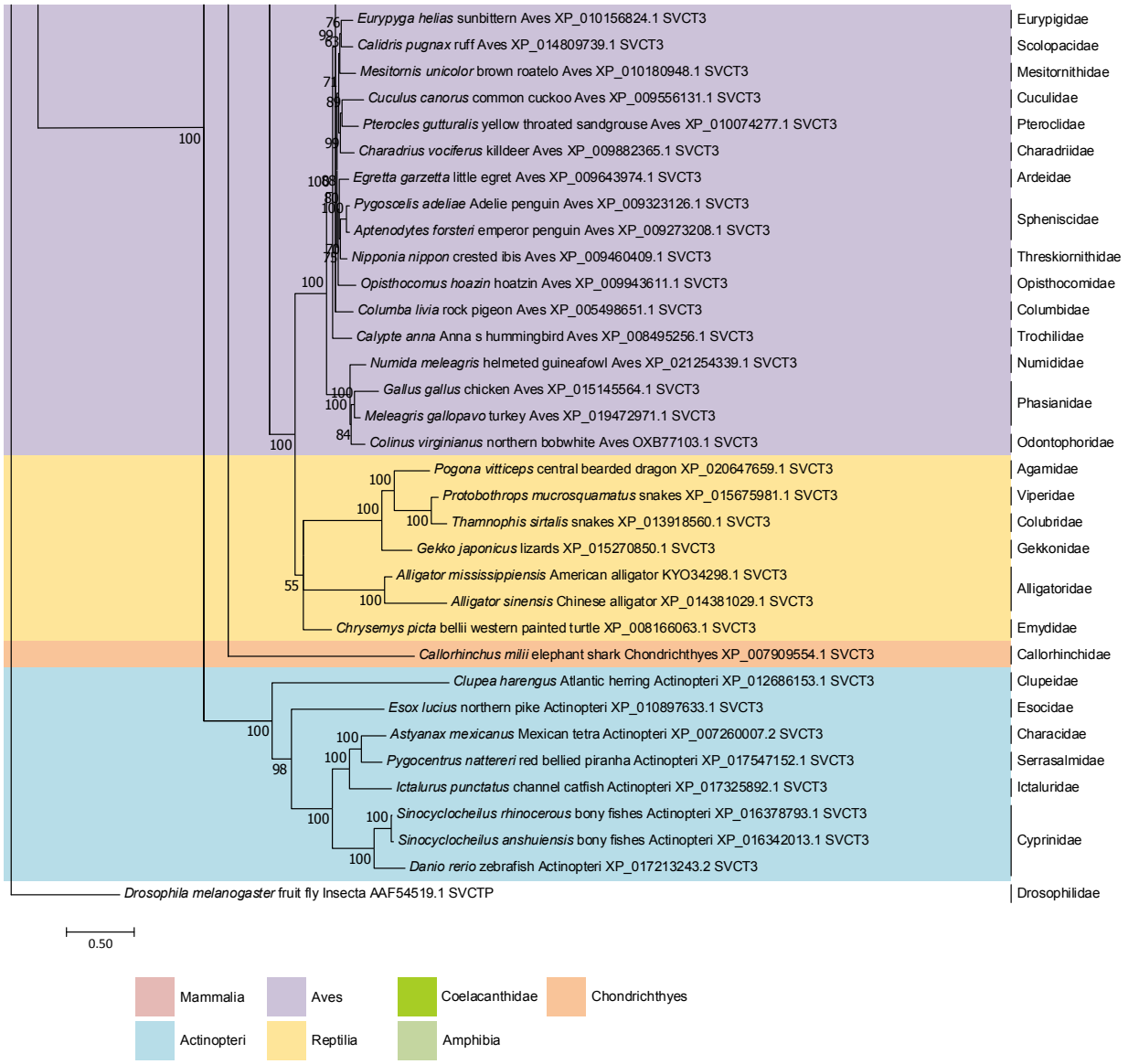
Continues next page


Miniopterus natalensis bats Mammalia XP_016067542.1 SVCT2
Myotis davidii bats Mammalia XP_006758125.1 SVCT2
Myotis brandtii Brandt s bat Mammalia XP_005874736.1 SVCT2
Myotis lucifugus little brown bat Mammalia XP_006096492.1 SVCT2
Eptesicus fuscus big brown bat Mammalia XP_008139284.1 SVCT2
Vespertilionidae
Rhinolophus sinicus Chinese rufous horseshoe bat Mammalia XP_019589318.1 SVCT2
Rhinolophidae
Hipposideros armiger great roundleaf bat Mammalia XP_019500165.1 SVCT2
Hipposideridae
Ceratotherium simum simum southern white rhinoceros Mammalia XP_014645584.1 SVCT2
Rhinocerotidae
Equus caballus horse Mammalia XP_014590528.1 SVCT2
Equus przewalskii Przewalski s horse Mammalia XP_008518016.1 SVCT2
Equus asinus ass Mammalia XP_014684197.1 SVCT2
Equidae
Panthera pardus leopard Mammalia XP_019318143.1 SVCT2
Panthera tigris altaica Amur tiger Mammalia XP_007073293.1 SVCT2
Felis catus domestic cat Mammalia XP_019682358.1 SVCT2
Acinonyx jubatus cheetah Mammalia XP_014933640.1 SVCT2
Felidae
Canis lupus familiaris dog Mammalia NP_001300742.1 SVCT2
Canidae
Enhydra lutris kenyoni sea otter Mammalia XP_022349621.1 SVCT2
Mustela putorius furo domestic ferret Mammalia XP_004772880.1 SVCT2
Mustelidae
Neomonachus schauinslandi Hawaiian monk seal Mammalia XP_021544579.1 SVCT2
Phocidae
Odobenus rosmarus divergens Pacific walrus Mammalia XP_004398260.1 SVCT2
Odobenidae
Ailuropoda melanoleuca giant panda Mammalia XP_011230125.1 SVCT2
Ursus maritimus polar bear Mammalia XP_008701537.1 SVCT2
Ursidae
Sus scrofa pig Mammalia XP_020932938.1 SVCT2
Suidae
Vicugna pacos alpaca Mammalia XP_015097334.1 SVCT2
Camelus dromedarius Arabian camel Mammalia XP_010992925.1 SVCT2
Camelus bactrianus Bactrian camel Mammalia XP_010971033.1 SVCT2
Camelus ferus Wild Bactrian camel Mammalia EPY79165.1 SVCT2
Camelidae
Bubalus bubalis water buffalo Mammalia XP_006047814.1 SVCT2
Bison bison bison American bison Mammalia XP_010834734.1 SVCT2
Bos taurus cattle Mammalia DAA23455.1 SVCT2
Bos indicus zebu cattle Mammalia XP_019835257.1 SVCT2
Bos mutus wild yak Mammalia ELR48512.1 SVCT2
Capra hircus goat Mammalia XP_017912852.1 SVCT2
Ovis aries musimon mouflon Mammalia XP_011995463.1 SVCT2
Ovis aries sheep Mammalia XP_014955204.1 SVCT2
Pantholops hodgsonii chiru Mammalia XP_005974369.1 SVCT2
Bovidae
Odocoileus virginianus texanus white tailed deer Mammalia XP_020739307.1 SVCT2
Cervidae
Balaenoptera acutorostrata scammoni minke whale Mammalia XP_007191767.1 SVCT2
Balaenopteridae
Lipotes vexillifer Yangtze River dolphin Mammalia XP_007464035.1 SVCT2
Lipotidae
Orcinus orca killer whale Mammalia XP_004276490.1 SVCT2
Tursiops truncatus bottlenose dolphin Mammalia XP_019797033.1 SVCT2
Delphinidae
Delphinapterus leucas beluga whale Mammalia XP_022412162.1 SVCT2
Monodontidae
Physeter catodon sperm whale Mammalia XP_007124808.1 SVCT2
Physeteridae
Sorex araneus European shrew Mammalia XP_004610999.1 SVCT2
Soricidae
Condylura cristata star nosed mole Mammalia XP_004687251.1 SVCT2
Talpidae
Erinaceus europaeus western European hedgehog Mammalia XP_007539622.1 SVCT2
Erinaceidae
Chrysochloris asiatica Cape golden mole Mammalia XP_006870473.1 SVCT2
Chrysochloridae
Echinops telfairi small Madagascar hedgehog Mammalia XP_004697908.1 SVCT2
Tenrecidae
Elephantulus edwardii Cape elephant shrew Mammalia XP_006894287.1 SVCT2
Macroscelididae
Trichechus manatus latirostris Florida manatee Mammalia XP_004383098.1 SVCT2
Trichechidae
Loxodonta africana African savanna elephant Mammalia XP_003411478.1 SVCT2
Elephantidae
Orycteropus afer afer aardvark Mammalia XP_007955405.1 SVCT2
Orycteropodidae
Drosophila melanogaster fruit fly Insecta AAF54519.1 SVCTP
Drosophilidae
1
Mammalia
Aves
Coelacanthidae
Actinopteri
Reptilia
Amphibia

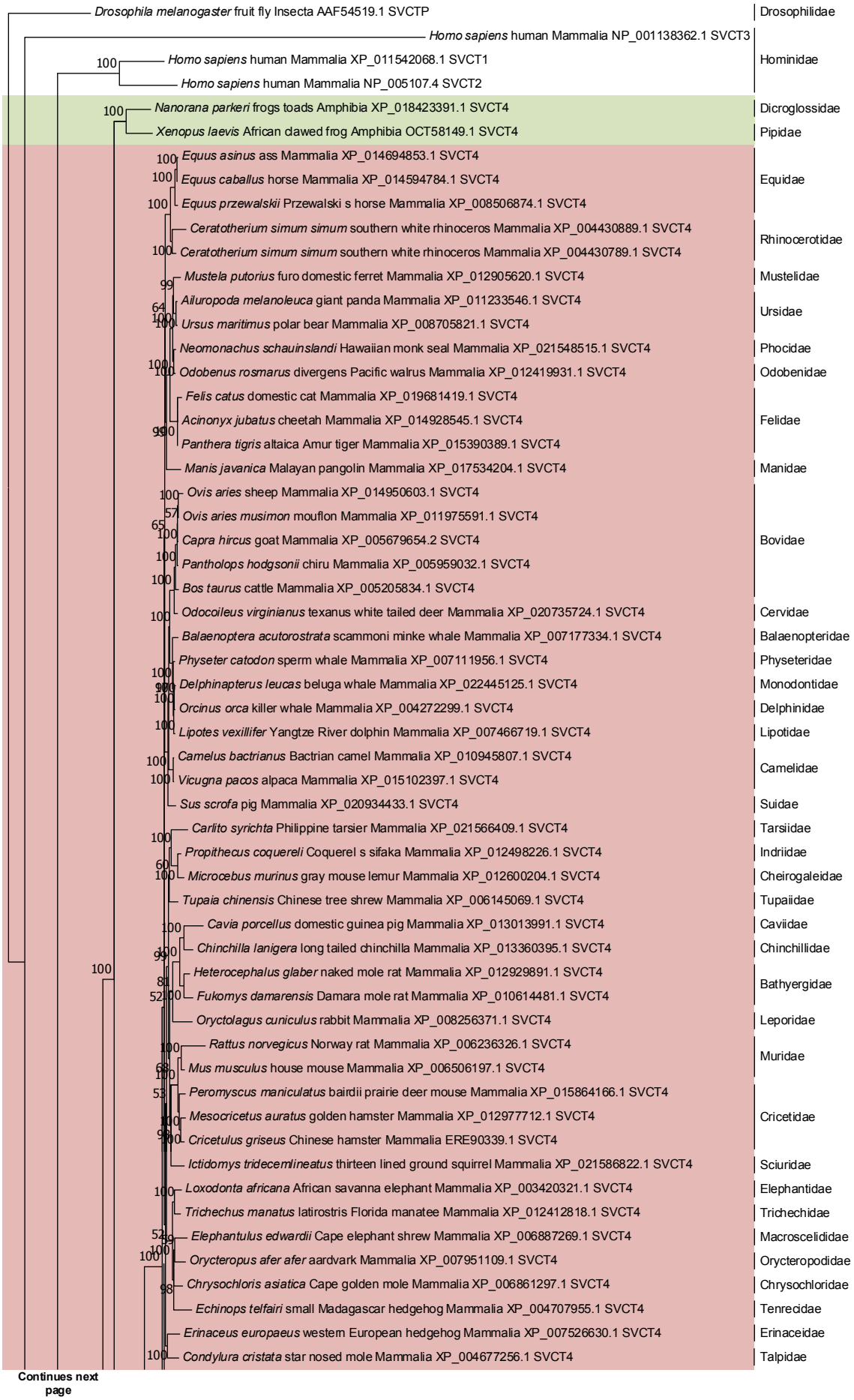
Supplementary Figure 3 - *Sodium-dependent Vitamin C transporter 2* (*SVCT2*) Bayesian phylogeny. The species family names are represented next to the corresponding sequences. Different taxonomic groups are highlighted with distinct colours: Mammalia in light red, Actinopteri in light blue, Aves in purple, Reptilia in yellow, Amphibia in light green and Coelacanthidae in green. The four outgroup sequences are not highlighted, and can be seen at the top (*H. sapiens SVCT1*, *H. sapiens SVCT3* and *M. musculus SVCT4*) and bottom of the phylogeny (*D. melanogaster SVCTP*).

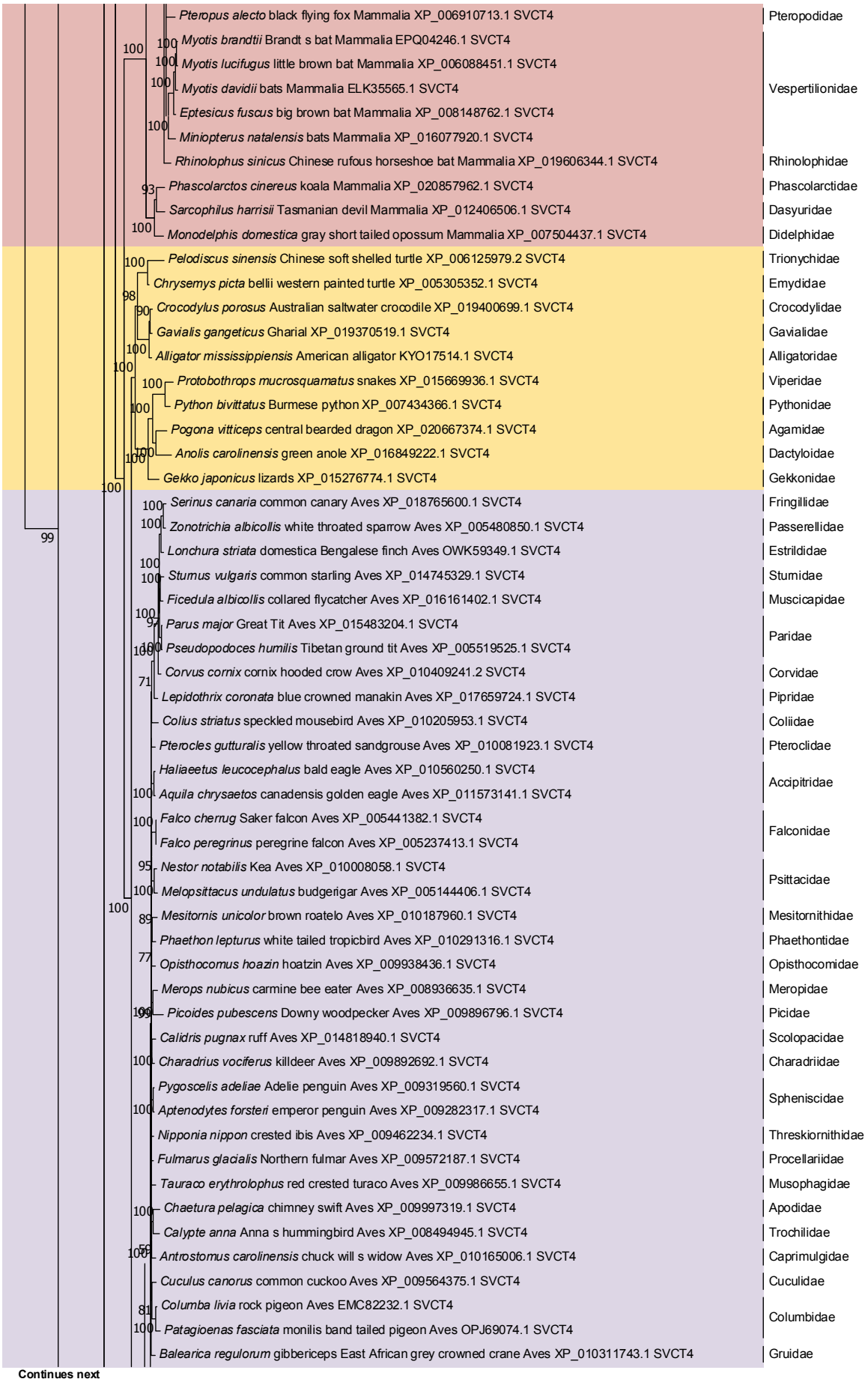


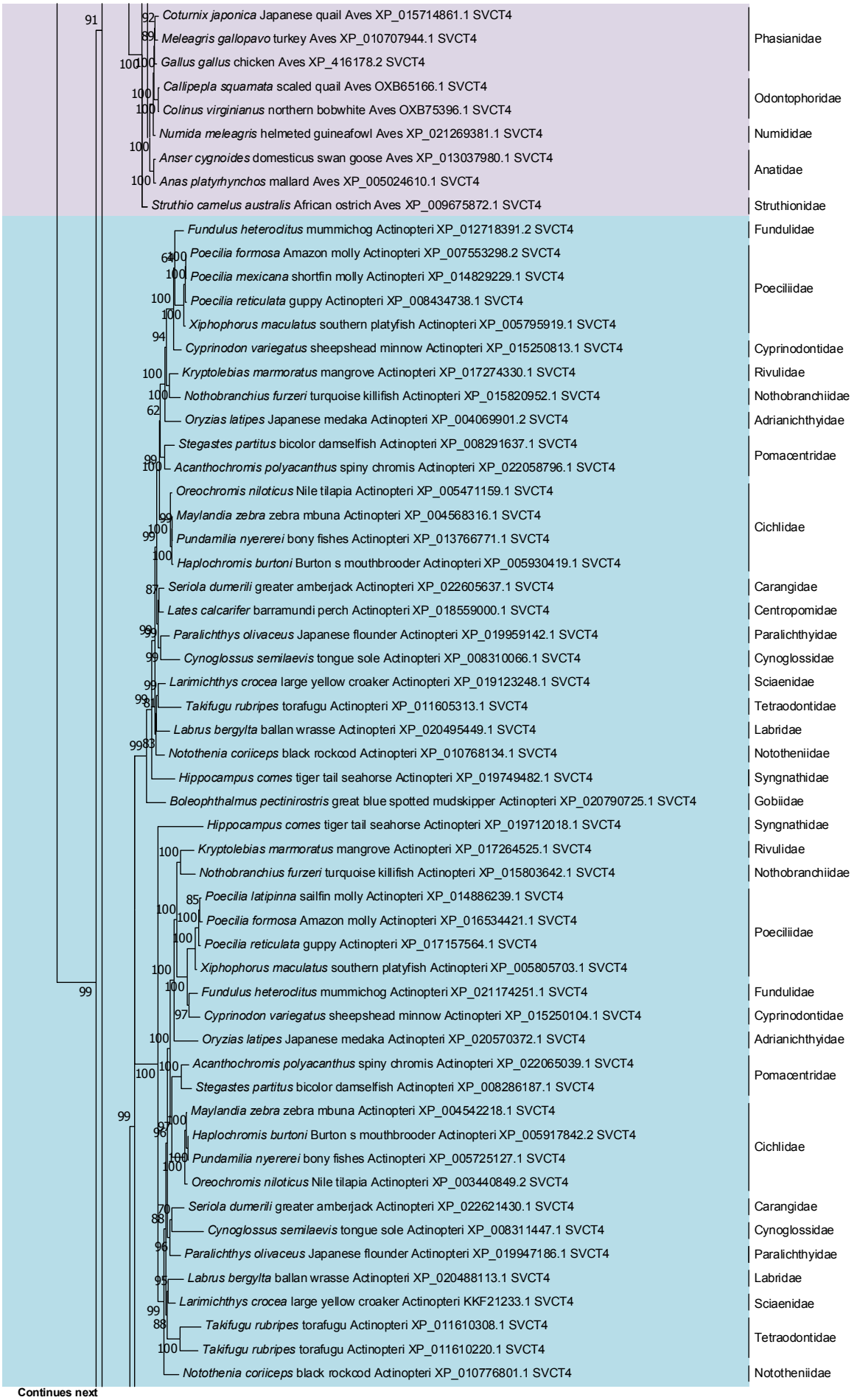

Manis javanica Malayan pangolin Mammalia XP_017510021.1 SVCT3 Manidae
Equus przewalskii Przewalski s horse Mammalia XP_008528176.1 SVCT3 Equidae
Equus asinus ass Mammalia XP_014711591.1 SVCT3
Ceratotherium simum simum southern white rhinoceros Mammalia XP_004427631.1 SVCT3 Rhinocerotidae
Balaenoptera acutorostrata scammoni minke whale Mammalia XP_007188009.1 SVCT3 Balaenopteridae
Tursiops truncatus bottlenose dolphin Mammalia XP_019794477.1 SVCT3 Delphinidae
Orcinus orca killer whale Mammalia XP_004262746.1 SVCT3
Delphinapterus leucas beluga whale Mammalia XP_022424022.1 SVCT3 Monodontidae
Lipotes vexillifer Yangtze River dolphin Mammalia XP_007445893.1 SVCT3 Lipotidae
Physeter catodon sperm whale Mammalia XP_007122523.1 SVCT3 Physeteridae
Bubalus bubalis water buffalo Mammalia XP_006059068.1 SVCT3 Bovidae
Bos mutus wild yak Mammalia ELR60768.1 SVCT3
Bison bison bison American bison Mammalia XP_010859522.1 SVCT3
Bos taurus cattle Mammalia DAA32439.1 SVCT3
Pantholops hodgsonii chiru Mammalia XP_005985192.1 SVCT3
Ovis aries musimon mouflon Mammalia XP_011969058.1 SVCT3
Ovis aries sheep Mammalia XP_004004988.1 SVCT3
Capra hircus goat Mammalia XP_005676614.1 SVCT3
Odocoileus virginianus texanus white tailed deer Mammalia XP_020760366.1 SVCT3 Cervidae
Vicugna pacos alpaca Camelidae Mammalia XP_006207355.1 SVCT3 Camelidae
Camelus dromedarius Arabian camel Mammalia XP_010986654.1 SVCT3
Camelus bactrianus Bactrian camel Mammalia XP_010947099.1 SVCT3
Camelus ferus Wild Bactrian camel Mammalia XP_006186977.1 SVCT3
Condylura cristata star nosed mole Mammalia XP_012581319.1 SVCT3 Talpidae
Erinaceus europaeus western European hedgehog Mammalia XP_016042478.1 SVCT3 Erinaceidae
Myotis brandtii Brandt s bat Mammalia EPQ18447.1 SVCT3 Vespertilionidae
Myotis lucifugus little brown bat Mammalia XP_006082597.1 SVCT3
Myotis davidii bats Mammalia XP_006763420.1 SVCT3
Eptesicus fuscus big brown bat Mammalia XP_008143372.1 SVCT3
Miniopterus natalensis bats Mammalia XP_016052393.1 SVCT3
Rousettus aegyptiacus Egyptian rousette Mammalia XP_016012195.1 SVCT3 Pteropodidae
Pteropus vampyrus large flying fox Mammalia XP_011356869.1 SVCT3
Pteropus alecto black flying fox Mammalia XP_006906768.1 SVCT3
Rhinolophus sinicus Chinese rufous horseshoe bat Mammalia XP_019597383.1 SVCT3 Rhinolophidae
Trichechus manatus latirostris Florida manatee Mammalia XP_004376384.1 SVCT3 Trichechidae
Elephantulus edwardii Cape elephant shrew Mammalia XP_006890366.1 SVCT3 Macroscelididae
Orycteropus afer afer aardvark Mammalia XP_007933890.1 SVCT3 Orycteropodidae
Echinops telfairi small Madagascar hedgehog Mammalia XP_004701671.1 SVCT3 Tenrecidae
Dasypus novemcinctus nine banded armadillo Mammalia XP_004467792.1 SVCT3 Dasypodidae
Phascolarctos cinereus koala Mammalia XP_020858502.1 SVCT3 Phascolarctidae
Monodelphis domestica gray short tailed opossum Mammalia XP_007501758.1 SVCT3 Didelphidae
Sarcophilus harrisii Tasmanian devil Mammalia XP_003765969.1 SVCT3 Dasyuridae
Nestor notabilis Kea Aves XP_010022026.1 SVCT3 Psittacidae
Melopsittacus undulatus budgerigar Aves XP_005154146.1 SVCT3 Psittaculidae
Falco cherrug Saker falcon Aves XP_005442889.1 SVCT3 Falconidae
Falco peregrinus peregrine falcon Aves XP_005234377.1 SVCT3
Haliaeetus leucocephalus bald eagle Aves XP_010567959.1 SVCT3 Accipitridae
Aquila chrysaetos canadensis golden eagle Aves XP_011578894.1 SVCT3
Parus major Great Tit Aves XP_018863083.1 SVCT3 Paridae
Pseudopodoces humilis Tibetan ground tit Aves XP_014105018.1 SVCT3
Sturnus vulgaris common starling Aves XP_014743569.1 SVCT3 Sturnidae
Ficedula albicollis collared flycatcher Aves XP_005049552.1 SVCT3 Muscicapidae
Serinus canaria common canary Aves XP_009086150.1 SVCT3 Fringillidae
Zonotrichia albicollis white throated sparrow Aves XP_005485272.1 SVCT3 Passerellidae
Geospiza fortis medium ground finch Aves XP_005415853.1 SVCT3 Thraupidae
Corvus brachyrhynchos American crow Aves XP_008643148.1 SVCT3 Corvidae
Manacus vitellinus golden collared manakin Aves XP_008921143.1 SVCT3 Pipridae
Lepidothrix coronata blue crowned manakin Aves XP_017680774.1 SVCT3
Acanthisitta chloris rifleman Aves XP_009073311.1 SVCT3 Acanthisittidae
Merops nubicus carmine bee eater Aves XP_008946694.1 SVCT3 Meropidae
Picoides pubescens Downy woodpecker Aves XP_009909463.1 SVCT3 Picidae
Continues next page



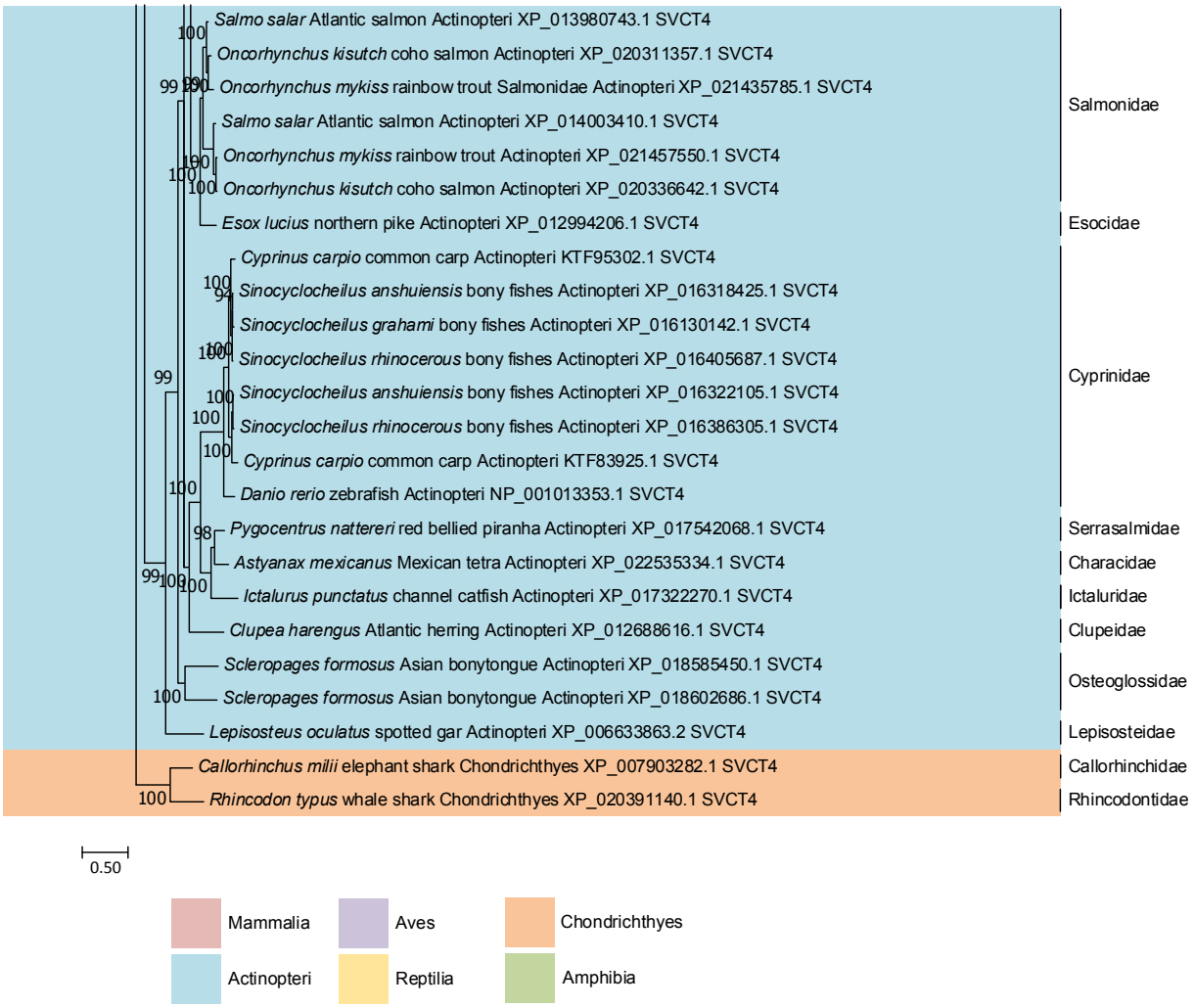
Supplementary Figure 4 - *Sodium-dependent Vitamin C transporter 3* (*SVCT3*) Bayesian phylogeny. The species family names are represented next to the corresponding sequences. Different taxonomic groups are highlighted with distinct colours: Mammalia in light red, Actinopteri in light blue, Aves in purple, Reptilia in yellow, Chondrichthyes in orange, Amphibia in light green and Coelacanthidae in green. The four outgroup sequences are not highlighted, and can be seen at the top (*H. sapiens SVCT1*, *H. sapiens SVCT2* and *M. musculus SVCT4*) and bottom of the phylogeny (*D. melanogaster SVCTP*).


Drosophila melanogaster fruit fly Insecta AAF54519.1 SVCTP
Drosophilidae
Homo sapiens human Mammalia NP_001138362.1 SVCT3
Homo sapiens human Mammalia XP_011542068.1 SVCT1
Hominidae
Homo sapiens human Mammalia NP_005107.4 SVCT2
Nanorana parkeri frogs toads Amphibia XP_018423391.1 SVCT4
Dicroglossidae
Xenopus laevis African clawed frog Amphibia OCT58149.1 SVCT4
Pipidae
Equus asinus ass Mammalia XP_014694853.1 SVCT4
Equus caballus horse Mammalia XP_014594784.1 SVCT4
Equus przewalskii Przewalski s horse Mammalia XP_008506874.1 SVCT4
Equidae
Ceratotherium simum simum southern white rhinoceros Mammalia XP_004430889.1 SVCT4
Ceratotherium simum simum southern white rhinoceros Mammalia XP_004430789.1 SVCT4
Rhinocerotidae
Mustela putorius furo domestic ferret Mammalia XP_012905620.1 SVCT4
Mustelidae
Ailuropoda melanoleuca giant panda Mammalia XP_011233546.1 SVCT4
Ursus maritimus polar bear Mammalia XP_008705821.1 SVCT4
Ursidae
Neomonachus schauinslandi Hawaiian monk seal Mammalia XP_021548515.1 SVCT4
Phocidae
Odobenus rosmarus divergens Pacific walrus Mammalia XP_012419931.1 SVCT4
Odobenidae
Felis catus domestic cat Mammalia XP_019681419.1 SVCT4
Acinonyx jubatus cheetah Mammalia XP_014928545.1 SVCT4
Panthera tigris altaica Amur tiger Mammalia XP_015390389.1 SVCT4
Felidae
Manis javanica Malayan pangolin Mammalia XP_017534204.1 SVCT4
Manidae
Ovis aries sheep Mammalia XP_014950603.1 SVCT4
Ovis aries musimon mouflon Mammalia XP_011975591.1 SVCT4
Capra hircus goat Mammalia XP_005679654.2 SVCT4
Pantholops hodgsonii chiru Mammalia XP_005959032.1 SVCT4
Bos taurus cattle Mammalia XP_005205834.1 SVCT4
Bovidae
Odocoileus virginianus texanus white tailed deer Mammalia XP_020735724.1 SVCT4
Cervidae
Balaenoptera acutorostrata scammoni minke whale Mammalia XP_007177334.1 SVCT4
Balaenopteridae
Physeter catodon sperm whale Mammalia XP_007111956.1 SVCT4
Physeteridae
Delphinapterus leucas beluga whale Mammalia XP_022445125.1 SVCT4
Monodontidae
Orcinus orca killer whale Mammalia XP_004272299.1 SVCT4
Delphinidae
Lipotes vexillifer Yangtze River dolphin Mammalia XP_007466719.1 SVCT4
Lipotidae
Camelus bactrianus Bactrian camel Mammalia XP_010945807.1 SVCT4
Vicugna pacos alpaca Mammalia XP_015102397.1 SVCT4
Camelidae
Sus scrofa pig Mammalia XP_020934433.1 SVCT4
Suidae
Carlito syrichta Philippine tarsier Mammalia XP_021566409.1 SVCT4
Tarsiidae
Propithecus coquereli Coquerel s sifaka Mammalia XP_012498226.1 SVCT4
Indriidae
Microcebus murinus gray mouse lemur Mammalia XP_012600204.1 SVCT4
Cheirogaleidae
Tupaia chinensis Chinese tree shrew Mammalia XP_006145069.1 SVCT4
Tupaiidae
Cavia porcellus domestic guinea pig Mammalia XP_013013991.1 SVCT4
Caviidae
Chinchilla lanigera long tailed chinchilla Mammalia XP_013360395.1 SVCT4
Chinchillidae
Heterocephalus glaber naked mole rat Mammalia XP_012929891.1 SVCT4
Fukomys damarensis Damara mole rat Mammalia XP_010614481.1 SVCT4
Bathyergidae
Oryctolagus cuniculus rabbit Mammalia XP_008256371.1 SVCT4
Leporidae
Rattus norvegicus Norway rat Mammalia XP_006236326.1 SVCT4
Mus musculus house mouse Mammalia XP_006506197.1 SVCT4
Muridae
Peromyscus maniculatus bairdii prairie deer mouse Mammalia XP_015864166.1 SVCT4
Mesocricetus auratus golden hamster Mammalia XP_012977712.1 SVCT4
Cricetulus griseus Chinese hamster Mammalia ERE90339.1 SVCT4
Cricetidae
Ictidomys tridecemlineatus thirteen lined ground squirrel Mammalia XP_021586822.1 SVCT4
Sciuridae
Loxodonta africana African savanna elephant Mammalia XP_003420321.1 SVCT4
Elephantidae
Trichechus manatus latirostris Florida manatee Mammalia XP_012412818.1 SVCT4
Trichechidae
Elephantulus edwardii Cape elephant shrew Mammalia XP_006887269.1 SVCT4
Macroscelididae
Orycteropus afer afer aardvark Mammalia XP_007951109.1 SVCT4
Orycteropodidae
Chrysochloris asiatica Cape golden mole Mammalia XP_006861297.1 SVCT4
Chrysochloridae
Echinops telfairi small Madagascar hedgehog Mammalia XP_004707955.1 SVCT4
Tenrecidae
Erinaceus europaeus western European hedgehog Mammalia XP_007526630.1 SVCT4
Erinaceidae
Condylura cristata star nosed mole Mammalia XP_004677256.1 SVCT4
Talpidae
Continues next page

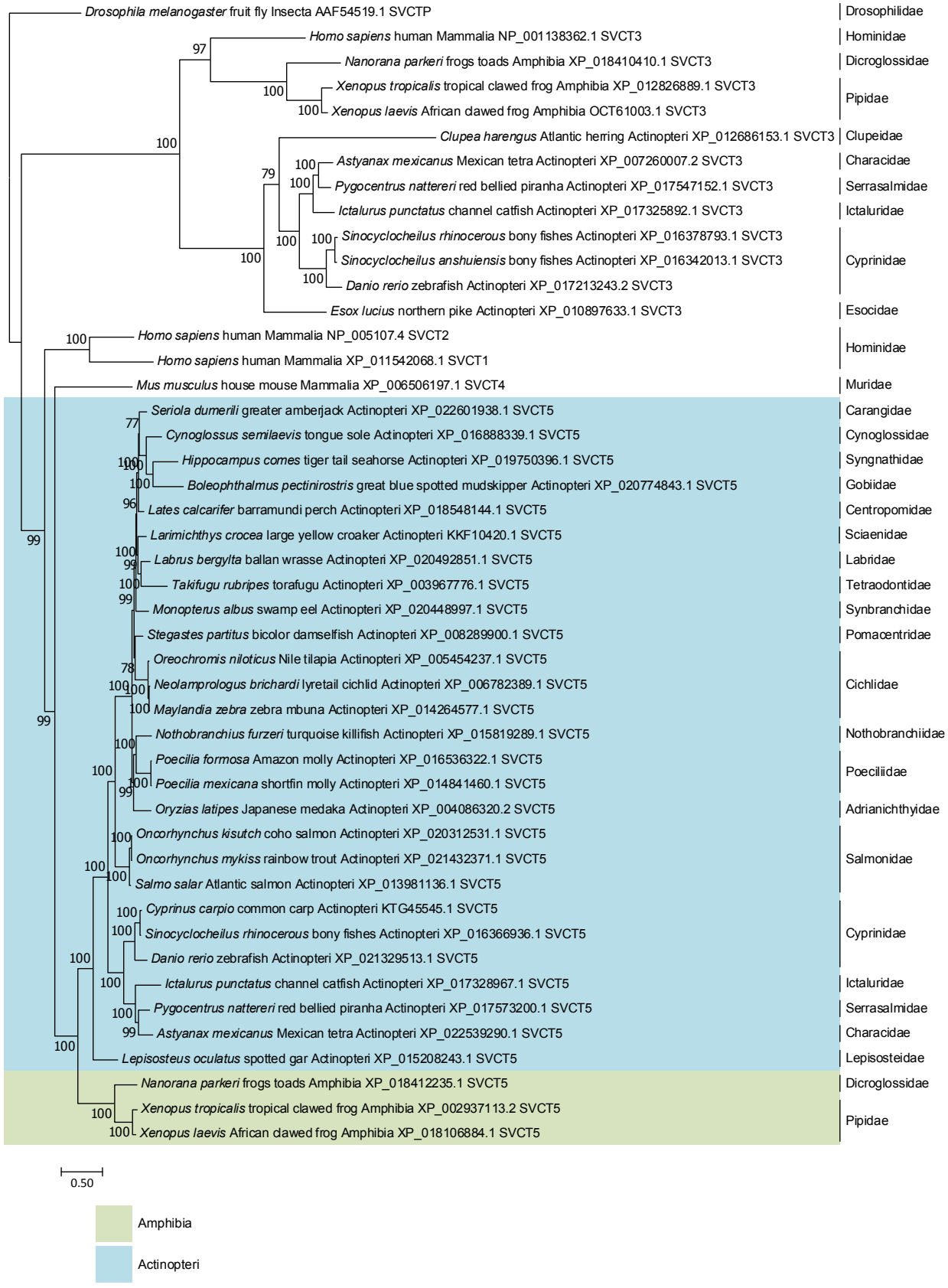

Pteropus alecto black flying fox Mammalia XP_006910713.1 SVCT4 | Pteropodidae
Myotis brandtii Brandt s bat Mammalia EPQ04246.1 SVCT4
Myotis lucifugus little brown bat Mammalia XP_006088451.1 SVCT4
Myotis davidii bats Mammalia ELK35565.1 SVCT4
Eptesicus fuscus big brown bat Mammalia XP_008148762.1 SVCT4
Miniopterus natalensis bats Mammalia XP_016077920.1 SVCT4 | Vespertilionidae
Rhinolophus sinicus Chinese rufous horseshoe bat Mammalia XP_019606344.1 SVCT4 | Rhinolophidae
Phascolarctos cinereus koala Mammalia XP_020857962.1 SVCT4 | Phascolarctidae
Sarcophilus harrisii Tasmanian devil Mammalia XP_012406506.1 SVCT4 | Dasyuridae
Monodelphis domestica gray short tailed opossum Mammalia XP_007504437.1 SVCT4 | Didelphidae
Pelodiscus sinensis Chinese soft shelled turtle XP_006125979.2 SVCT4 | Trionychidae
Chrysemys picta bellii western painted turtle XP_005305352.1 SVCT4 | Emydidae
Crocodylus porosus Australian saltwater crocodile XP_019400699.1 SVCT4 | Crocodylidae
Gavialis gangeticus Gharial XP_019370519.1 SVCT4 | Gavialidae
Alligator mississippiensis American alligator KYO17514.1 SVCT4 | Alligatoridae
Protobothrops mucrosquamatus snakes XP_015669936.1 SVCT4 | Viperidae
Python bivittatus Burmese python XP_007434366.1 SVCT4 | Pythonidae
Pogona vitticeps central bearded dragon XP_020667374.1 SVCT4 | Agamidae
Anolis carolinensis green anole XP_016849222.1 SVCT4 | Dactyloidae
Gekko japonicus lizards XP_015276774.1 SVCT4 | Gekkonidae
Serinus canaria common canary Aves XP_018765600.1 SVCT4 | Fringillidae
Zonotrichia albicollis white throated sparrow Aves XP_005480850.1 SVCT4 | Passerellidae
Lonchura striata domestica Bengalese finch Aves OWK59349.1 SVCT4 | Estrildidae
Sturnus vulgaris common starling Aves XP_014745329.1 SVCT4 | Sturnidae
Ficedula albicollis collared flycatcher Aves XP_016161402.1 SVCT4 | Muscicapidae
Parus major Great Tit Aves XP_015483204.1 SVCT4
Pseudopodoces humilis Tibetan ground tit Aves XP_005519525.1 SVCT4 | Paridae
Corvus cornix cornix hooded crow Aves XP_010409241.2 SVCT4 | Corvidae
Lepidothrix coronata blue crowned manakin Aves XP_017659724.1 SVCT4 | Pipridae
Colius striatus speckled mousebird Aves XP_010205953.1 SVCT4 | Coliidae
Pterocles gutturalis yellow throated sandgrouse Aves XP_010081923.1 SVCT4 | Pteroclidae
Haliaeetus leucocephalus bald eagle Aves XP_010560250.1 SVCT4
Aquila chrysaetos canadensis golden eagle Aves XP_011573141.1 SVCT4 | Accipitridae
Falco cherrug Saker falcon Aves XP_005441382.1 SVCT4
Falco peregrinus peregrine falcon Aves XP_005237413.1 SVCT4 | Falconidae
Nestor notabilis Kea Aves XP_010008058.1 SVCT4
Melopsittacus undulatus budgerigar Aves XP_005144406.1 SVCT4 | Psittacidae
Mesitornis unicolor brown roatelo Aves XP_010187960.1 SVCT4 | Mesitornithidae
Phaethon lepturus white tailed tropicbird Aves XP_010291316.1 SVCT4 | Phaethontidae
Opisthocomus hoazin hoatzin Aves XP_009938436.1 SVCT4 | Opisthocomidae
Merops nubicus carmine bee eater Aves XP_008936635.1 SVCT4 | Meropidae
Picoides pubescens Downy woodpecker Aves XP_009896796.1 SVCT4 | Picidae
Calidris pugnax ruff Aves XP_014818940.1 SVCT4 | Scolopacidae
Charadrius vociferus killdeer Aves XP_009892692.1 SVCT4 | Charadriidae
Pygoscelis adeliae Adelie penguin Aves XP_009319560.1 SVCT4
Aptenodytes forsteri emperor penguin Aves XP_009282317.1 SVCT4 | Spheniscidae
Nipponia nippon crested ibis Aves XP_009462234.1 SVCT4 | Threskiornithidae
Fulmarus glacialis Northern fulmar Aves XP_009572187.1 SVCT4 | Procellariidae
Tauraco erythrolophus red crested turaco Aves XP_009986655.1 SVCT4 | Musophagidae
Chaetura pelagica chimney swift Aves XP_009997319.1 SVCT4 | Apodidae
Calypte anna Anna s hummingbird Aves XP_008494945.1 SVCT4 | Trochilidae
Antrostomus carolinensis chuck will s widow Aves XP_010165006.1 SVCT4 | Caprimulgidae
Cuculus canorus common cuckoo Aves XP_009564375.1 SVCT4 | Cuculidae
Columba livia rock pigeon Aves EMC82232.1 SVCT4
Patagioenas fasciata monilis band tailed pigeon Aves OPJ69074.1 SVCT4 | Columbidae
Balearica regulorum gibbericeps East African grey crowned crane Aves XP_010311743.1 SVCT4 | Gruidae
Continues next page



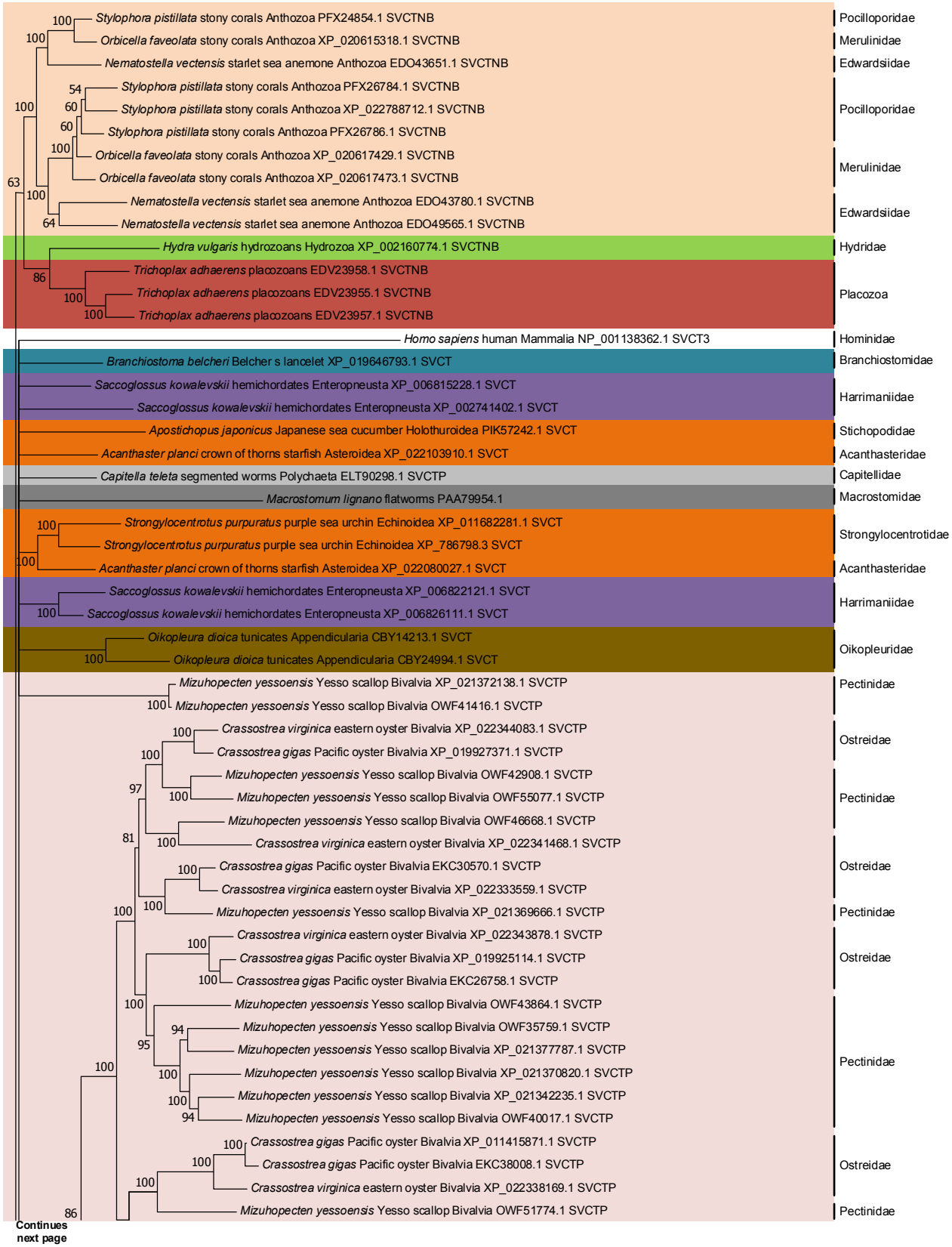
Continues next page

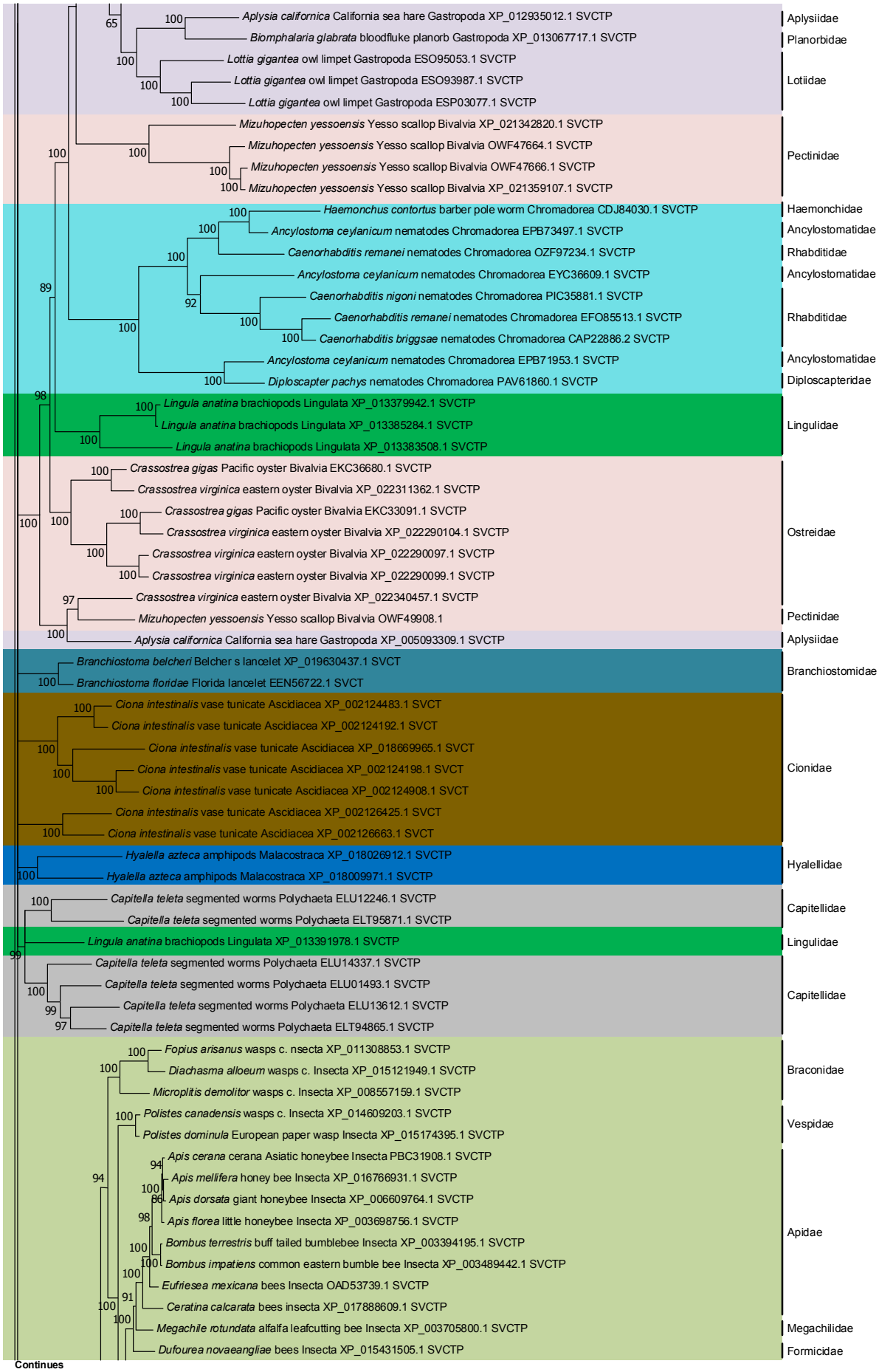


Supplementary Figure 5 - *Sodium-dependent Vitamin C transporter 4* (*SVCT4*) Bayesian phylogeny. The species family names are represented next to the corresponding sequences. Different taxonomic groups are highlighted with distinct colours: Mammalia in light red, Actinopteri in light blue, Aves in purple, Reptilia in yellow, Chondrichthyes in orange and Amphibia in light green. The four outgroup sequences are not highlighted and can be seen at the top of the phylogeny.

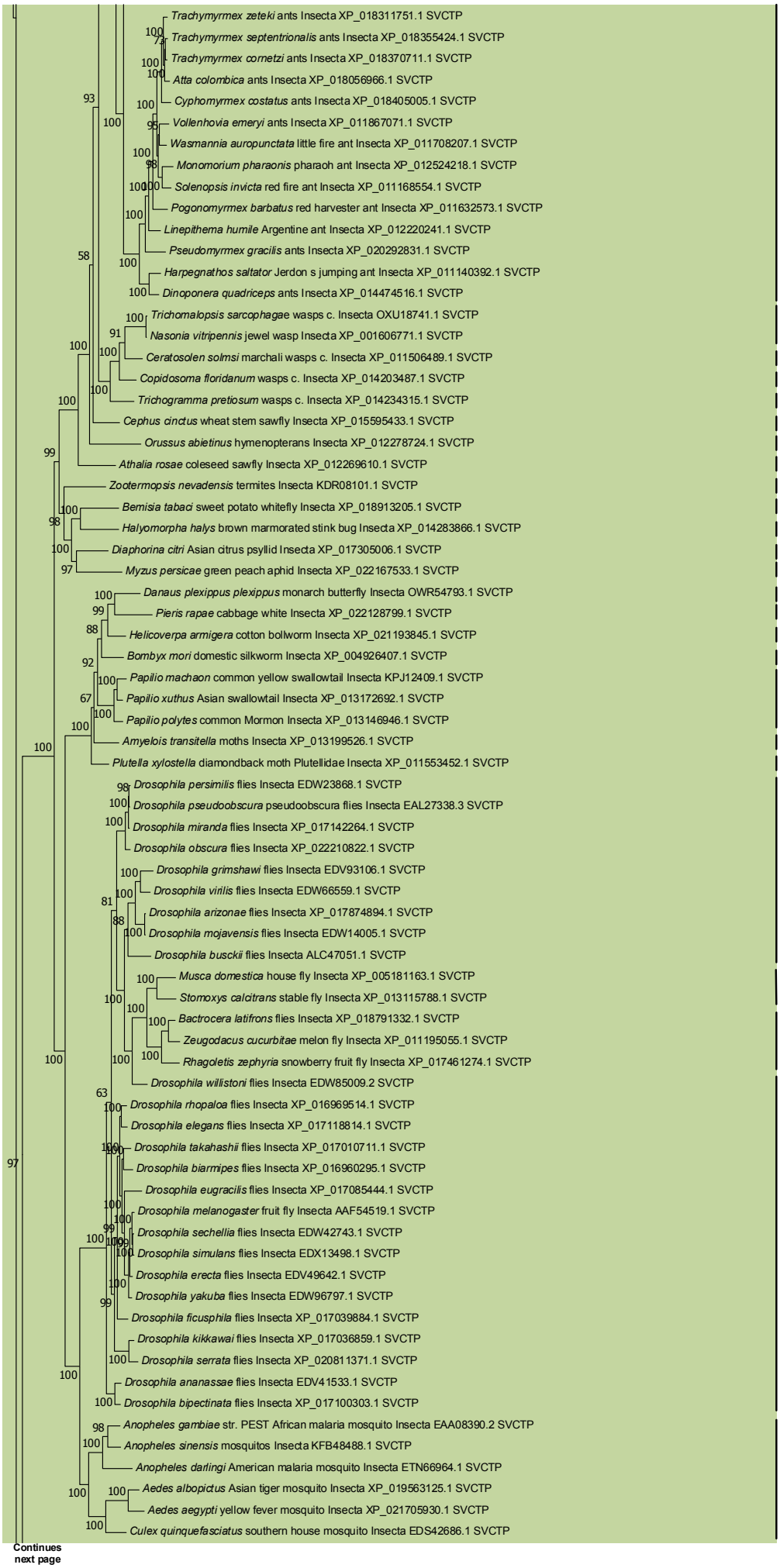


Supplementary Figure 6 - Putative *Sodium-dependent Vitamin C transporter 5* (*SVCT5*) Bayesian phylogeny. The species family names are represented next to the corresponding sequences. Different taxonomic groups are highlighted with distinct colours, namely Actinopteri in light blue and Amphibia in light green. The outgroup sequences used are not highlighted and can be seen at the top of the phylogeny.

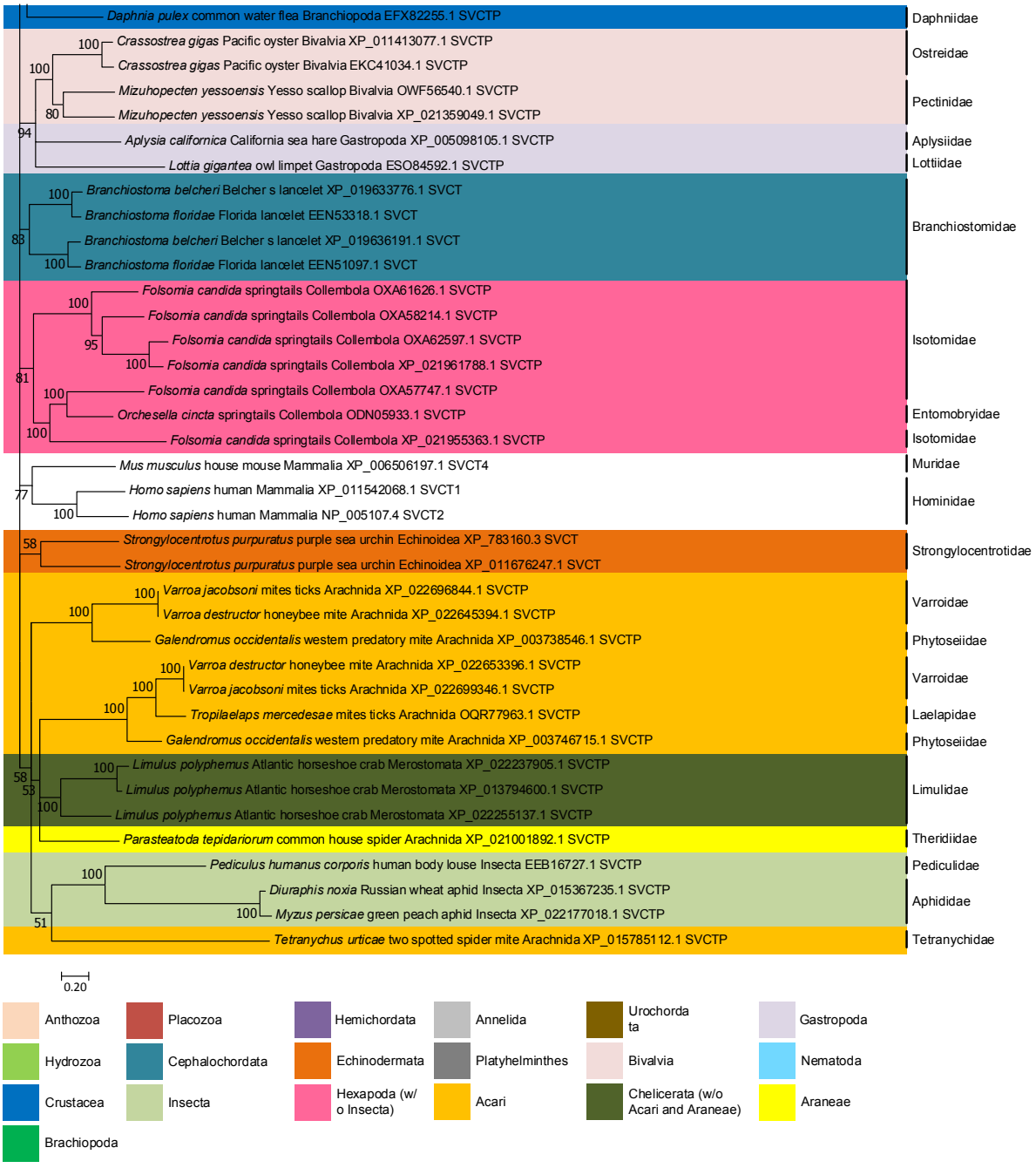

Stylophora pistillata stony corals Anthozoa PFX24854.1 SVCTNB
Orbicella faveolata stony corals Anthozoa XP_020615318.1 SVCTNB
Nematostella vectensis starlet sea anemone Anthozoa EDO43651.1 SVCTNB
Stylophora pistillata stony corals Anthozoa PFX26784.1 SVCTNB
Stylophora pistillata stony corals Anthozoa XP_022788712.1 SVCTNB
Stylophora pistillata stony corals Anthozoa PFX26786.1 SVCTNB
Orbicella faveolata stony corals Anthozoa XP_020617429.1 SVCTNB
Orbicella faveolata stony corals Anthozoa XP_020617473.1 SVCTNB
Nematostella vectensis starlet sea anemone Anthozoa EDO43780.1 SVCTNB
Nematostella vectensis starlet sea anemone Anthozoa EDO49565.1 SVCTNB
Hydra vulgaris hydrozoans Hydrozoa XP_002160774.1 SVCTNB
Trichoplax adhaerens placozoans EDV23958.1 SVCTNB
Trichoplax adhaerens placozoans EDV23955.1 SVCTNB
Trichoplax adhaerens placozoans EDV23957.1 SVCTNB
Homo sapiens human Mammalia NP_001138362.1 SVCT3
Branchiostoma belcheri Belcher s lancelet XP_019646793.1 SVCT
Saccoglossus kowalevskii hemichordates Enteropneusta XP_006815228.1 SVCT
Saccoglossus kowalevskii hemichordates Enteropneusta XP_002741402.1 SVCT
Apostichopus japonicus Japanese sea cucumber Holothuroidea PIK57242.1 SVCT
Acanthaster planci crown of thorns starfish Asteroidea XP_022103910.1 SVCT
Capitella teleta segmented worms Polychaeta ELT90298.1 SVCTP
Macrostomum lignano flatworms PAA79954.1
Strongylocentrotus purpuratus purple sea urchin Echinoidea XP_011682281.1 SVCT
Strongylocentrotus purpuratus purple sea urchin Echinoidea XP_786798.3 SVCT
Acanthaster planci crown of thorns starfish Asteroidea XP_022080027.1 SVCT
Saccoglossus kowalevskii hemichordates Enteropneusta XP_006822121.1 SVCT
Saccoglossus kowalevskii hemichordates Enteropneusta XP_006826111.1 SVCT
Oikopleura dioica tunicates Appendicularia CBY14213.1 SVCT
Oikopleura dioica tunicates Appendicularia CBY24994.1 SVCT
Mizuhopecten yessoensis Yesso scallop Bivalvia XP_021372138.1 SVCTP
Mizuhopecten yessoensis Yesso scallop Bivalvia OWF41416.1 SVCTP
Crassostrea virginica eastern oyster Bivalvia XP_022344083.1 SVCTP
Crassostrea gigas Pacific oyster Bivalvia XP_019927371.1 SVCTP
Mizuhopecten yessoensis Yesso scallop Bivalvia OWF42908.1 SVCTP
Mizuhopecten yessoensis Yesso scallop Bivalvia OWF55077.1 SVCTP
Mizuhopecten yessoensis Yesso scallop Bivalvia OWF46668.1 SVCTP
Crassostrea virginica eastern oyster Bivalvia XP_022341468.1 SVCTP
Crassostrea gigas Pacific oyster Bivalvia EKC30570.1 SVCTP
Crassostrea virginica eastern oyster Bivalvia XP_022333559.1 SVCTP
Mizuhopecten yessoensis Yesso scallop Bivalvia XP_021369666.1 SVCTP
Crassostrea virginica eastern oyster Bivalvia XP_022343878.1 SVCTP
Crassostrea gigas Pacific oyster Bivalvia XP_019925114.1 SVCTP
Crassostrea gigas Pacific oyster Bivalvia EKC26758.1 SVCTP
Mizuhopecten yessoensis Yesso scallop Bivalvia OWF43864.1 SVCTP
Mizuhopecten yessoensis Yesso scallop Bivalvia OWF35759.1 SVCTP
Mizuhopecten yessoensis Yesso scallop Bivalvia XP_021377787.1 SVCTP
Mizuhopecten yessoensis Yesso scallop Bivalvia XP_021370820.1 SVCTP
Mizuhopecten yessoensis Yesso scallop Bivalvia XP_021342235.1 SVCTP
Mizuhopecten yessoensis Yesso scallop Bivalvia OWF40017.1 SVCTP
Crassostrea gigas Pacific oyster Bivalvia XP_011415871.1 SVCTP
Crassostrea gigas Pacific oyster Bivalvia EKC38008.1 SVCTP
Crassostrea virginica eastern oyster Bivalvia XP_022338169.1 SVCTP
Mizuhopecten yessoensis Yesso scallop Bivalvia OWF51774.1 SVCTP
Pocilloporidae
Merulinidae
Edwardsiidae
Pocilloporidae
Merulinidae
Edwardsiidae
Hydridae
Placozoa
Hominidae
Branchiostomidae
Harrimaniidae
Stichopodidae
Acanthasteridae
Capitellidae
Macrostomidae
Strongylocentrotidae
Acanthasteridae
Harrimaniidae
Oikopleuridae
Pectinidae
Ostreidae
Pectinidae
Ostreidae
Pectinidae
Ostreidae
Pectinidae
Ostreidae
Pectinidae
Continues next page



Continues next page



Continues next page



Supplementary Figure 7 - Non-bilaterian, protostomian and basal deuterostomian *Sodium-dependent Vitamin C transporters* (*SVCTNB*, *SVCTP* and *SVCT*) Bayesian phylogeny. The species relevant taxonomic classifications are represented next to the corresponding sequences. 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 spread across the polytomy.