

Chapter 4

The Evolution of Retinol Metabolism
and Implications for the Origin of
Vision

Abstract

Vision in animals fundamentally relies on a light-sensitive molecule, an opsin protein bound to a derivative of vitamin A, typically 11-cis-retinal. Upon light absorption, 11-cis-retinal undergoes a conformational change to its trans-state. To regain light sensitivity, the 11-cis configuration must be restored, a process dependent on retinol metabolism, which produces 11-cis-retinal from dietary vitamin A. This metabolic pathway is vital for visual functionality, therefore, understanding its evolution offers insights into the origin of vision itself. Despite its significance, the evolution and diversity of enzymes integral to this pathway remain elusive.

Thus, the aim of this chapter was to investigate the evolution of the retinol metabolism pathway by identifying the orthogroups (phylogenetically defined gene families) its enzymes belong to and characterizing their evolutionary history across diverse eukaryotic lineages. The results identified 12 overarching orthogroups encompassing the enzymes involved in the retinol metabolism. These orthogroups are generally very ancient and span the eukaryotic domain. Phylogenetic analyses uncovered intricate substructures within each orthogroup, revealing multiple sub-families. Intriguingly, the orthogroups containing some of most quintessential enzymes of the retinol metabolism (e.g., BCMO1 and RPE65) exhibit a pattern where the specific subfamily engaged in the pathway is found exclusively in animals, despite the wider eukaryotic distribution of the overarching orthogroup. Such findings allude to animal-specific expansions of these gene families, concurrent with the emergence of vision.

Introduction

The retinol metabolism comprises a series of enzymatic reactions that convert dietary vitamin A into various bioactive compounds, primarily retinal for vision and retinoic acid for gene regulation, ensuring the proper functioning of visual processes and other physiological roles in the body (Blomhoff and Blomhoff 2006; Dewett, Lam-Kamath, et al. 2021).

Retinol (Vitamin A₁) is an essential micronutrient derived primarily from diet. It can be obtained directly from animal sources as retinyl esters or indirectly from plant sources as pro-vitamin A carotenoids, which are then converted into retinol in the body (Trifiletti 2014). In turn retinol can be esterified to retinyl ester by the enzyme lecithin retinol acyltransferase (LRAT) allowing for its storage (Batten et al. 2004). When needed, retinyl ester is hydrolysed back to retinol (Moiseyev et al. 2005). Retinol is oxidized to retinal by retinol dehydrogenases (RDHs). Several other enzymes are involved in various steps of the retinol metabolism pathway as schematically shown in Figure 4.1 that summarizes what is known about the pathway according to the KEGG Pathway Database (Kanehisa et al. 2021). In addition, Table 4.1 provides a comprehensive list of these enzymes ranked by the number of pathways they participate in according to KEGG. Involvement in one or few pathways serves as an indicator of enzyme specificity to the retinol metabolism, as opposed to broad spectrum enzymes.

Retinal, particularly 11-cis-retinal, plays a crucial role in vision (Palczewski and Kiser 2020). 11-cis-retinal binds to the protein opsin in photoreceptor cells forming rhodopsin. Upon absorbing a photon, 11-cis-retinal is isomerized to all-trans-retinal, leading to a conformational change in opsin, and initiating a cascade of events called phototransduction (Hardie and Juusola 2015; Lamb 2020) (see Chapter 3). After light exposure, all-trans-retinal is reduced to all-trans-retinol and then converted back to 11-cis-retinal through a series of enzymatic reactions. This part of the visual cycle is essential as it ensures the retina's responsiveness to light (Palczewski and Kiser 2020). The regulation of the metabolic steps ensures sufficient 11-cis-retinal availability and prevents toxic build-up of intermediates. Additionally, retinal can be further oxidized to retinoic acid by retinaldehyde dehydrogenase (RALDH1). Retinoic acid serves as a signalling

molecule that regulates gene expression and is critical for numerous developmental processes (Blomhoff and Blomhoff 2006).

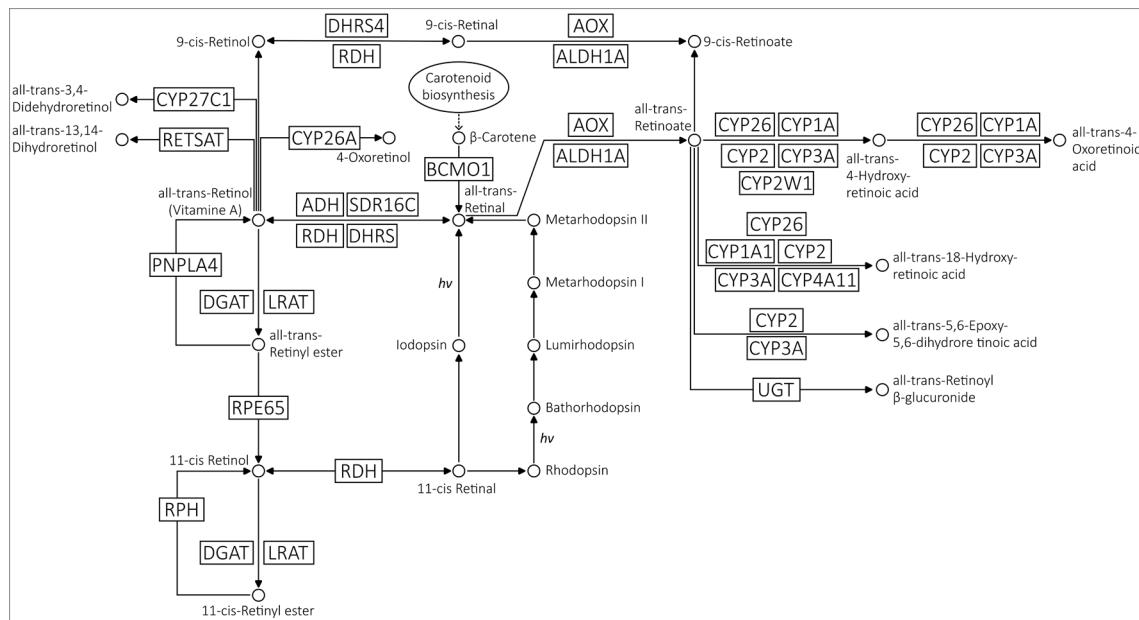


Figure 4.1. Retinol metabolism pathway. The pathway used as reference in this study is based on the KEGG map00830.

The retinol metabolism, particularly as it relates to vision, has been primarily studied in vertebrates, especially mammals, with mouse (*Mus musculus*) and human being the most extensively characterized due to their relevance in medical research (Trifiletti 2014; Widjaja-Adhi and Golczak 2020). Some aspects of retinol metabolism have been studied in the invertebrate model *Drosophila melanogaster* (Dewett, Labaf, et al. 2021; Dewett, Lam-Kamath, et al. 2021). Outside of animals, carotenoid biosynthesis pathways, producing retinol precursors such as beta-carotene, have received more attention than the retinol metabolism itself, especially in plants (Hirschberg 2001).

Given the importance of retinol metabolism, it is compelling to explore its evolutionary history and potential diversity outside of traditional model organisms, especially in the wider context of the evolution of vision. Hence, the work presented in this chapter aimed to unravel this intricate history. The initial step was to identify the genetic components involved and determine their evolutionary relationships to answer questions such as: Do the gene families belong to overarching orthogroups? How closely related are they? The subsequent objective was to uncover the distribution of these components across the animal kingdom and, more broadly, within eukaryotes, to pinpoint the specific point in time when all the components came into place. The final endeavour was to delineate the

main evolutionary events characterizing each orthogroup, to discern, for instance, if certain gene families have undergone a greater number of evolutionary events and contextualizing them within the evolutionary tree of life.

Table 4.1. Enzymes involved in retinol metabolism listed in order of number of pathways they are involved in according to KEGG as a measure of their specificity to the retinol metabolism pathway.

Gene Family	Kegg map00830			No Kegg Pathways
	Entry	Symbol	Name	
RETSAT	K09516	RETSAT	all-trans-retinol 13,14-reductase	1
RPH	EC3.1.1.63	?	11-cis-retinyl-palmitate hydrolase	1
PNPLA4	K11157	PNPLA4	patatin-like phospholipase domain-containing protein 4	2
RDH	K00061	RDH5	11-cis-retinol dehydrogenase	2
	K11150	RDH8	retinol dehydrogenase 8	2
	K11151	RDH10	retinol dehydrogenase 10	2
	K11152	RDH11	retinol dehydrogenase 11	3
	K11153	RDH12	retinol dehydrogenase 12	3
	K11161	RDH13	retinol dehydrogenase 13	3
ALDH1	K07249	ALDH1A	retinal dehydrogenase	2
RPE65	K11158	RPE65	retinoid isomerohydrolase / lutein isomerase	2
BCMO1	K00515	BCMO1	beta-carotene 15,15'-dioxygenase	3
LRAT	K00678	LRAT	phosphatidylcholine-retinol O-acyltransferase	3
DHRS	K11148	DHRS4L2	dehydrogenase/reductase SDR family member 4-like protein 2	2
	K11149	DHRS9	dehydrogenase/reductase SDR family member 9	2
	K11146	DHRS3	short-chain dehydrogenase/reductase 3	3
	K11147	DHRS4	dehydrogenase/reductase SDR family member 4	3
	K15734	SDR16C5	all-trans-retinol dehydrogenase (NAD+)	3
	K13369	HSD17B6	17beta-estradiol 17-dehydrogenase / all-trans-retinol dehydrogenase (NAD+)	4
DGAT	K11156	GAT2L4	diacylglycerol O-acyltransferase 2-like protein 4	3
	K11155	DGAT1	diacylglycerol O-acyltransferase 1	4
CYP	K17951	CYP27C1	all-trans-retinol 3,4-desaturase	1
	K07411	CYP2A	cytochrome P450 family 2 subfamily A	2
	K07423	CYP2W1	cytochrome P450 family 2 subfamily W1	2
	K07437	CYP26A	cytochrome P450 family 26 subfamily A	2
	K12664	CYP26B	cytochrome P450 family 26 subfamily B	2
	K12665	CYP26C	cytochrome P450 family 26 subfamily C	2
	K07420	CYP2S1	cytochrome P450 family 2 subfamily S1	3
	K17691	CYP3A7	cytochrome P450 family 3 subfamily A7	4
	K17720	CYP2C18	cytochrome P450 family 2 subfamily C18	4
	K07412	CYP2B	cytochrome P450 family 2 subfamily B	5
	K07424	CYP3A	cytochrome P450 family 3 subfamily A	6
	K17690	CYP3A5	cytochrome P450 family 3 subfamily A5	6
	K07425	CYP4A	long-chain fatty acid omega-monooxygenase	7
	K17709	CYP2B6	cytochrome P450 family 2 subfamily B6	7
	K07413	CYP2C	cytochrome P450 family 2 subfamily C	8
	K17683	CYP2A6	cytochrome P450 family 2 subfamily A6	8
	K17718	CYP2C8	cytochrome P450 family 2 subfamily C8	8
	K17719	CYP2C9	cytochrome P450 family 2 subfamily C9	9
AOX	K07408	CYP1A1	cytochrome P450 family 1 subfamily A1	10
	K17689	CYP3A4	cytochrome P450 family 3 subfamily A4	10
	K07409	CYP1A2	cytochrome P450 family 1 subfamily A2	12
	K00157	AOX	aldehyde oxidase	10
ADH	K13980	ADH4	alcohol dehydrogenase 4	9
	K13951	ADH1_7	alcohol dehydrogenase 1/7	10
	K13952	ADH6	alcohol dehydrogenase 6	10
	K00001	adh	alcohol dehydrogenase	13
	K13953	adhP	alcohol dehydrogenase, propanol-preferring	13
	K00121	ADH5	alcohol dehydrogenase 5	16
UGT	K00699	UGT	glucuronosyltransferase	14

Results and Discussion

Enzymes involved in retinol metabolism belong to 12 major orthogroups.

To gain insights into the evolution of retinol metabolism, it is essential to trace the evolutionary history of each of the enzymes in the pathway. For this I used as reference the pathway described by KEGG (Kanehisa et al. 2021) (Figure 4.1 and Table 4.1) and explored the genes encoding these enzymes across 101 species spanning all of Eukarya (Table 4.2 and Extended Table 4.2).

Although many enzymes partake in the pathway, some might be part of a larger gene family. Therefore, to study their evolution, the initial task was to identify their respective orthogroup – a collection of orthologs and paralogs that originated from the same initial gene duplication. An orthogroup can be considered as a phylogenetically defined gene family.

KEGG ortholog lists (Kanehisa 2019) for each enzyme were used as starting point for each enzyme (see more details in Methods). It is worth noting that the only enzyme from the KEGG pathway excluded from this analysis was RPH (11-cis-retinyl-palmitate hydrolase) (Figure 4.1 and Table 4.1). Despite its hypothesized role in hydrolysing stored 11-cis-retinyl esters to 11-cis retinol is pertinent to vision (Blaner et al. 1984; Blaner et al. 1987), there is a significant knowledge gap surrounding this putative enzyme. The human gene encoding it remains unidentified, and KEGG does not list any orthologs for it. Given the nebulous nature of this enzyme, this study chose to prioritize better-understood enzymes, including RPE65 that catalyses the extremely similar reaction of hydrolysing all-trans-retinyl esters to 11-cis retinol (Moiseyev et al. 2005).

Orthogroup inference methods often rely on computing sequence similarity scores amongst sequences as a measure of protein distances and then using these scores for clustering the sequences (e.g., OrthoMCL (Li et al. 2003)). Here, two alternative software for orthogroup inference were used to independently infer orthogroups (see details in Methods). The first was OrthoFinder that implements a method that eliminates gene length bias during similarity score assessment (Emms and Kelly 2015) and uses a

Table 4.2. List of species used in this study with respective proteome BUSCO scores.

Clades						Species	% Complete BUSCOs (tot) (eukaryota odd10)	
Amorphea	Obazoa	Opisthokonta	Holozoa	Metazoa	Bilateria	Mollusca Brachiopoda Annelida Bryozoa Acoelomorpha Platyhelminthes Nematoda Rotifera Tardigrada Arthropoda Vertebrate Urochordata Cephalochordata Echinodermata Hemichordata	<i>Octopus bimaculoides</i> <i>Lottia gigantea</i> <i>Lingula anguis</i> <i>Capitella teleta</i> <i>Helobdella robusta</i> <i>Bugula neritina</i> <i>Hofstenia miamia</i> <i>Echinococcus multilocularis</i> <i>Caenorhabditis elegans</i> <i>Loa loa</i> <i>Brachionus plicatilis</i> <i>Ramazzottius varieornatus</i> <i>Daphnia pulex</i> <i>Drosophila melanogaster</i> <i>Calanus glacialis</i> <i>Homo sapiens</i> <i>Mus musculus</i> <i>Danio rerio</i> <i>Eptatretus burgeri</i> <i>Ciona intestinalis</i> <i>Oikopleura dioica</i> <i>Branchiostoma belcheri</i> <i>Acanthaster planci</i> <i>Strongylocentrotus purpuratus</i> <i>Saccoglossus kowalevskii</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> >90% >80% >70% >60% >50% <50% </div> </div>
						Cnidaria	<i>Fungia scutaria</i> <i>Montastrea cavernosa</i> <i>Madracis auretenra</i> <i>Sylophora pistillata</i> <i>Astreopora sp</i> <i>Porites australiensis</i> <i>Acropora digitifera</i> <i>Anthopleura elegantissima</i> <i>Nematostella vectensis</i> <i>Gorgia ventinalis</i> <i>Clytia hemisphaerica</i> <i>Hydra magnapapillata</i> <i>Aurelia sp</i> <i>Thelohanellus kitaei</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 84.40% 80.40% 68.30% 85.10% 69.80% 81.60% 42.80% 62.00% 93.40% 56.50% 84.70% 70.90% 67.40% 29.40% </div> </div>
						Placozoa	<i>Trichoplax adhaerens (sPH2)</i> <i>Haitiangia hongkongensis</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 96.10% 96.80% </div> </div>
						Ctenophora	<i>Mnemosynopsis leidyi</i> <i>Pleurobrachia bachei</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 83.60% 47.50% </div> </div>
						Porifera	<i>Leucosolenia complicata</i> <i>Sycon ciliatum</i> <i>Styliosa carteri</i> <i>Oscarella pearsei</i> <i>Amphimedon queenslandica</i> <i>Haliclona tubifera</i> <i>Ephydatia muelleri</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 94.90% 97.30% 43.60% 94.90% 92.50% 75.30% 93.40% </div> </div>
						Choanoflagellata	<i>Acanthoeeca spectabilis</i> <i>Helgoeca nana</i> <i>Diaphanoeca grandis</i> <i>Didymoea costata</i> <i>Choanoeeca perplexa</i> <i>Monosiga brevicollis</i> <i>Mylnosiga fluctuans</i> <i>Salpingoeca kvevrii</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 93.30% 94.10% 92.20% 94.90% 92.10% 78.80% 93.30% 94.90% </div> </div>
						Ichthyosporea	<i>Amoebidium parasiticum</i> <i>Sphaeroforma arctica</i> <i>Sphaerothecum destruens</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 89.90% 63.20% 68.60% </div> </div>
						Pluriformea	<i>Corallochytrium limacisporum</i>	90.20%
						Filastrea	<i>Capsaspora owczarzaki</i>	93.70%
						Nucleomyceta (Holomycota)		
Diaphoreticces	Excavata	Archaeplastida	Malawimonadidae	Chloroplastida	Stramenopiles	Fungi	<i>Fusarium oxysporum</i> <i>Saccharomyces cerevisiae</i> <i>Ustilago maydis</i> <i>Mortierella elongata</i> <i>Rhizopus microsporus</i> <i>Spizellomyces punctatus</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 97.30% 93.30% 98.80% 98.40% 97.30% 94.90% </div> </div>
						Rotospaerida	<i>Parvularia atlantis</i> <i>Fonticula alba</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 75.70% 71.00% </div> </div>
						Apusomonadida	<i>Thecamonas trahens</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 81.60% </div> </div>
						Breviata	<i>Pygusul biforma</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 62.40% </div> </div>
						Amoebozoa	<i>Dictyostelium discoideum</i> <i>Vermamoeba vermiformis</i> <i>Cunea spBSH02190019</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 94.10% 89.00% 76.50% </div> </div>
						CRuMs	<i>Diphyllidea rotama</i>	90.60%
						Rigfilida	<i>Rigifila ramosa</i>	86.30%
						Incertae sedis	<i>Gefionella okellyi</i>	78.50%
						Archaeplastida	<i>Arabidopsis thaliana</i> <i>Marchantia polymorpha</i> <i>Chlamydomonas reinhardtii</i> <i>Chloropicon primus</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 99.60% 99.60% 91.80% 89.80% </div> </div>
						Glaucoma	<i>Gloeochaete wittrockiana</i>	85.50%
						Rhodophyta	<i>Porphyridium purpureum</i> <i>Galdieria sulphuraria</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 72.60% 77.30% </div> </div>
						Rhodelphis	<i>Rhodelphis marinus</i>	89.80%
						Haptista	<i>Prymnesium parvum</i> <i>Phaeocystis globosa</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 71.40% 75.30% </div> </div>
						Cryptista	<i>Centroplastelida</i> <i>Cryptophyceae</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 88.20% 78.90% </div> </div>
						Sar	<i>Baffinella spCCMP2293</i> <i>Guillardia theta</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 97.90% 95.60% </div> </div>
						Stramenopiles	<i>Aphanomyces astaci</i> <i>Phytophthora infestans</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 90.20% </div> </div>
						Rhizaria	<i>Plasmidophora brassicaceae</i>	89.00%
						Alveolata	<i>Colponemidia spColp10</i> <i>Tetrahymena thermophila</i>	<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> 93.30% 76.40% </div> </div>
						Telomelia	<i>Telomelina subtile</i>	82%
						Euglenozoa	<i>Euglena longa</i>	83.50%
						Heterolobosea	<i>Pharyngomonas kirbyi</i>	85.10%
						Jakobida	<i>Andalucia godoyi</i>	79.20%

phylogenetic framework to detect orthologs (Emms and Kelly 2019); the second was Broccoli that uses phylogenetic relationships instead of protein distances for clustering sequences and then applies machine learning algorithms to extract orthologous relationships from this network (Derelle et al. 2020). By comparing results from these distinct strategies, the chances of comprehensively identifying orthogroups for retinol metabolism enzymes was enhanced.

OrthoFinder identified a total of 50 orthogroups, while Broccoli provided 58. After annotating the orthogroups and filtering out unrelated ones (see Methods), we were left with 14 OrthoFinder and 21 Broccoli orthogroups (Figure 4.2). Results were compared by assessing the percentage of shared sequences between OrthoFinder and Broccoli orthogroups. Generally, there is substantial agreement between OrthoFinder and Broccoli results, with many orthogroups displaying one-to-one correspondence. However, while OrthoFinder yielded fewer, larger orthogroups, Broccoli in some cases produced more and smaller ones. As a result, some gene families were fragmented into multiple smaller orthogroups exclusively in Broccoli's output. Collectively, the OrthoFinder and Broccoli results delineated 12 orthogroups encompassing retinol metabolism enzymes (Table 4.3). While the primary purpose of the orthogroup inference step was to identify gene families to investigate further with phylogenetic analyses, it also provided some preliminary insights into the evolution of some of the enzymes involved in retinol metabolism. For example, both OrthoFinder and Broccoli place DGAT1 and DGAT2L4 into distinct orthogroups. Additionally, RDH and DHRS enzymes, subfamilies of a larger group, display a complex substructure, suggesting intricate phylogenetic relationships.

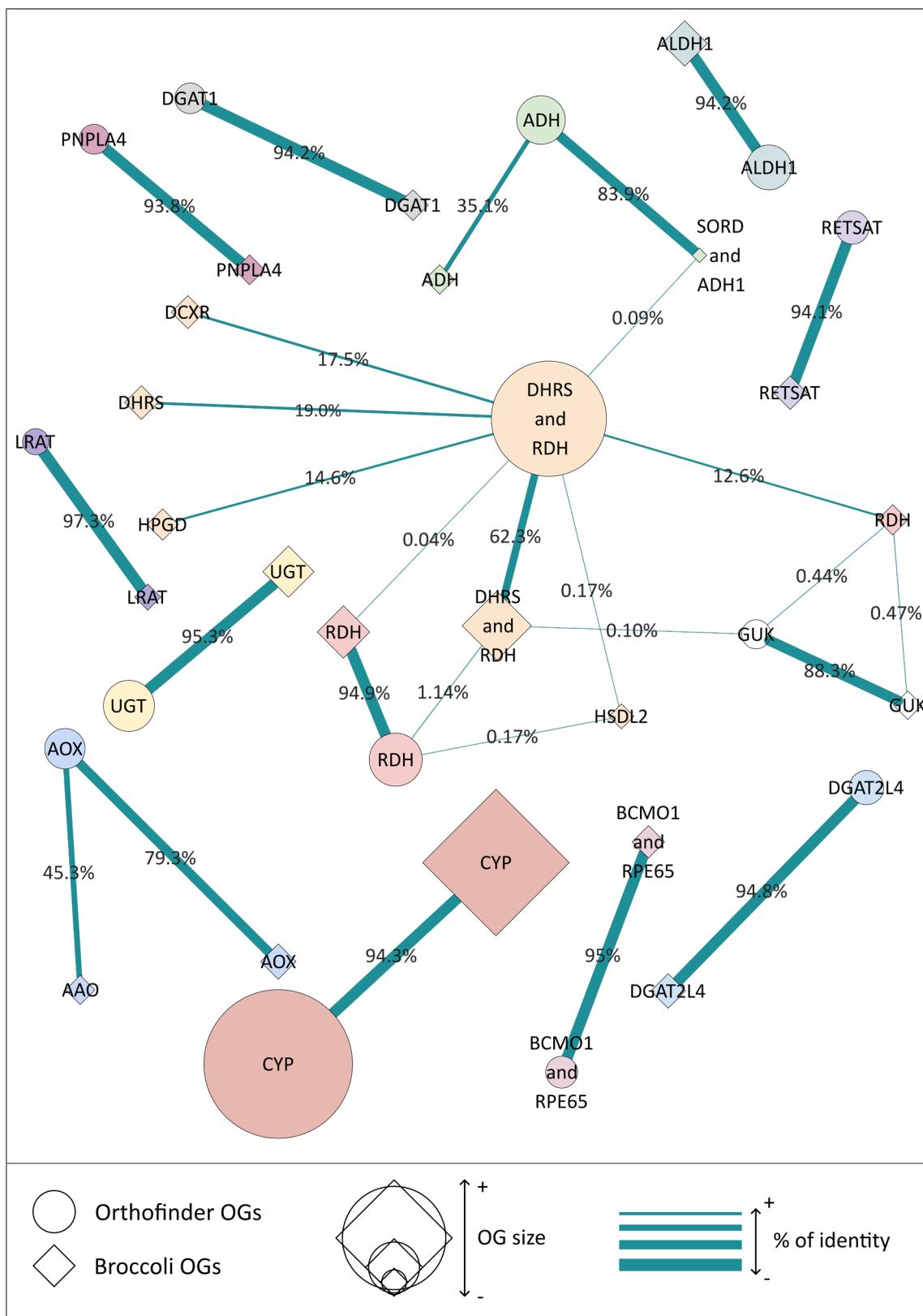


Figure 4.2. Orthogroup inference analysis. Orthogroups inferred from two different software (OrthoFinder and Broccoli) are compared.

Table 4.3. Summary of the comparison between OrthoFinder and Broccoli orthogroup inference results. In the last column, the final orthogroups used for phylogenetic analyses are shown.

Kegg Groups	Orthofinder Orthogroups (OOG)	Broccoli Orthogroups (BOG)	Consensus Groups
RETSAT	OOG16-RETSAT	BOG37-RETSAT	RETSAT
PNPLA4	OOG20-PNPLA4	BOG51-PNPLA4	PNPLA4
ALDH1	OOG6-ALDH1	BOG33-ALDH1	ALDH1
BCMO1 RPE65	OOG17-BCMO1-RPE65	BOG49-BCMO1-RPE65	BCMO1-RPE65
LRAT	OOG27-LRAT	BOG52-LRAT	LRAT
DHRS	OOG1-DHRS-RDH	BOG18-DHRS BOG3-DHRS-RDH BOG17-HPGD BOG22-DCXR BOG19 -HSDL2	RDH-DHRS
RDH	OOG3-RDH	BOG5-RDH BOG15-RDH	
DGAT	OOG19-DGAT1 OOG14-DGAT2L4	BOG27-DGAT1 BOG46-AWAT2	DGAT1 DGAT2L4
CYP	OOG0-CYP	BOG-CYP	CYP
AOX	OOG7-AOX	BOG38-AOX BOG39-AAO3	AOX
ADH	OOG5-ADH	BOG14-ADH BOG11-SORD-ADH1	ADH
UGT	OOG4-UGT	BOG30-UGT	UGT

Reconstructing phylogenetic histories of retinol metabolism orthogroups.

All retinol metabolism related orthogroups were further examined with phylogenetic analyses to understand the details of their evolutionary histories. After constructing phylogenetic trees (see Methods), two distinct but complementary approaches were applied for analysing each orthogroup tree. The first, using Possvm software (Grau-Bové and Sebé-Pedrós 2021), identifies orthologs within the gene tree, defines sub-orthogroups within the primary orthogroup, and annotates the tree based on these sub-orthogroups. The second employs GeneRax (Morel et al. 2020), which reconciles the gene tree to a species tree using a maximum-likelihood framework. Possvm offers the advantage of swiftly annotating large trees, facilitating their interpretation. As it infers orthologs using implicit taxonomic information from the gene tree, it eliminates the need for a species tree and avoids potential biases from contentious species relationships. GeneRax, in contrast, delivers a precise reconciled tree detailing speciation, duplication, and loss events at each node. However, it demands more computation time and necessitates a species tree.

The detailed results for each orthogroup, presented in order of specificity to the retinol metabolism, are described below.

RETSAT

The Retinol Saturase (RETSAT) enzyme catalyses the reaction that saturates the 13-14 double bond of all-trans-retinol to produce all-trans-13,14-dihdriretinol (Moise et al. 2004) (Figure 4.1). This enzyme appears to be involved only in retinol metabolism according to the KEGG Database (Table 4.1), meaning it is very specific to this pathway.

The orthogroups identified for RETSAT by OrthoFinder and Broccoli present a clear one-to-one relationship with high degree of identity (Figure 4.2), indicating no mixture with any other orthogroup examined. The merged RETSAT orthogroup contained 338 sequences distributed throughout all major eukaryotic clades (Figure 4.3A).

Phylogenetic analysis identified a monophyletic clade containing RETSAT genes from various species of eukaryotes, as well as other clades of related enzymes (Figure 4.3 B and C). Ortholog sorting with Possvm identified 7 orthogroups within the RETSAT family, with one orthogroup containing RETSAT, PYRD2 (Pyridine Nucleotide-

Disulphide Oxidoreductase Domain 2) and CRT enzymes that are involved in carotenoid metabolism (Figure 4.3B). Gene tree to species tree reconciliation with GeneRax confirmed the overall topology and revealed a high number of evolutionary events (especially losses) in proportion to the size of the orthogroup (Figure 4.3C).

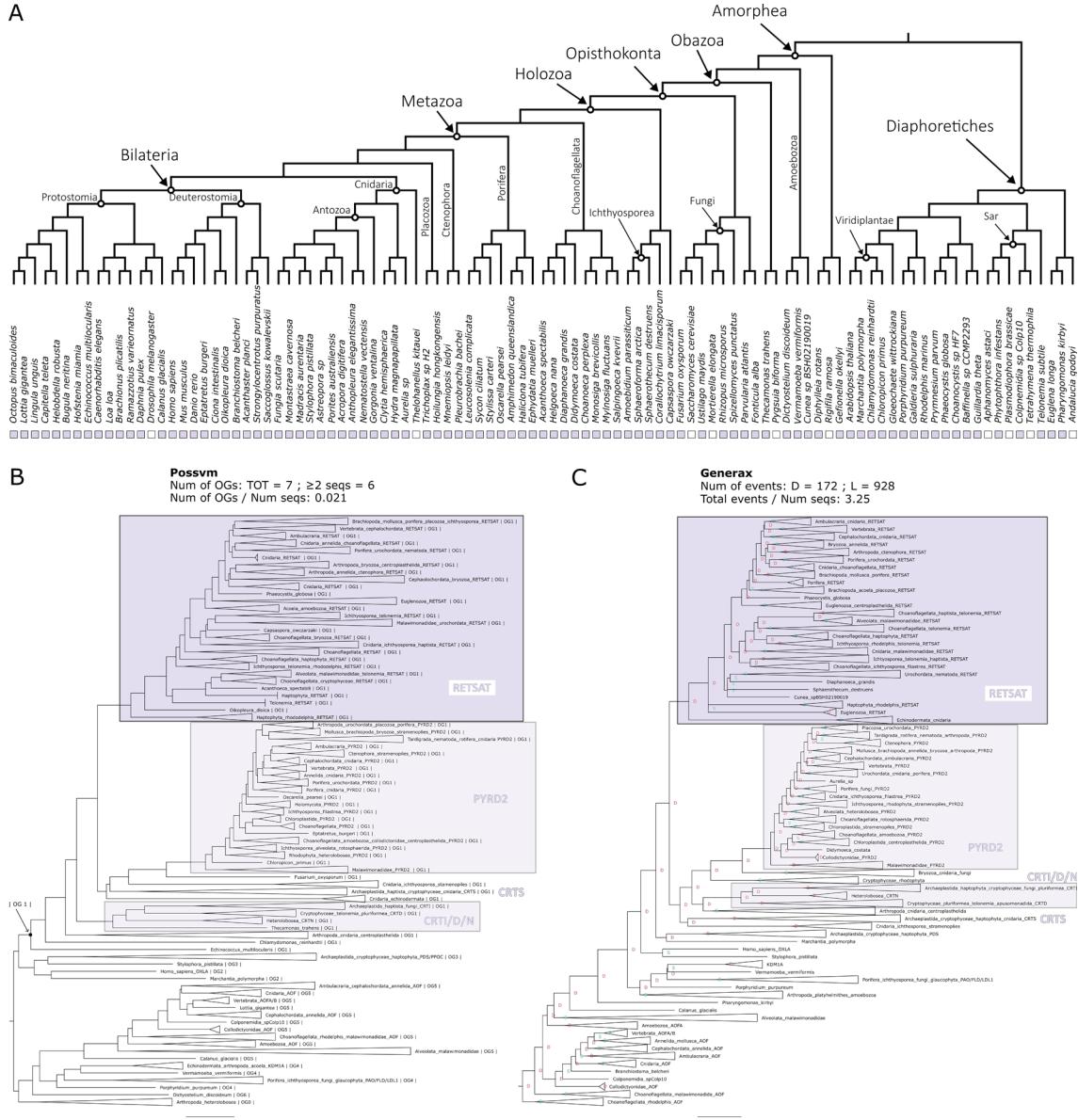


Figure 4.3. Phylogenetic analysis for the RETSAT orthogroup. **A.** Distribution of the orthogroup throughout Eukarya. **B.** Phylogenetic tree analysed with Possvm for identification of sub-orthogroups. Number of Possvm orthogroups identified (Total, and with ≥ 2 sequences) are provided, as well as the ratio between number of Possvm sub-orthogroups per number of sequences in full orthogroup are indicated. **C.** Gene tree to species tree reconciliation using GeneRax highlighting speciation and duplication events at each node. Numbers of duplications (D) and losses (L) are provided, as well as the ratio of total events (D+L) per number of sequences in the orthogroup.

PNPLA4

The Patatin Like Phospholipase Domain Containing 4 (PNPLA4) enzyme plays a role in the hydrolysis of retinyl esters to retinol (Holmes 2012; Schreiber et al. 2012) (Figure 4.1). It is involved in one other pathway according to KEGG (Table 4.1).

Both OrthoFinder and Broccoli identify one distinct orthogroup for PNPLA4 independent from all other orthogroups (Figure 4.2). The final PNPLA4 orthogroup contains 215 sequences. While being present in both major eukaryotic clades, this orthogroup appears to be missing in basal groups of Amorphea, such as the Holomycota branch that includes Fungi (Figure 4.4A).

The phylogenetic analysis clarified the relationship between PNPLA4 and other PNPLA enzymes present in the orthogroup (Figure 4.4 B and C). Possvm identified 9 orthogroups within this family. PNPLA1-5 belonged to the same orthogroup, with PNPLA4 being sister group to the other genes (Figure 4.4B). The GeneRax reconciled tree recovered the same topology and identified a moderate number of events (Figure 4.4C). The phylogenetic analysis also revealed that while the broad PNPLA4 orthogroup included sequences from a wide range of eukaryotic organisms, the PNPLA1-5 sub-clades contained primarily animal sequences. The tight relationship between PNPLA4 and other PNPLA genes is in accordance with evidence suggesting that some of them are also involved in retinol metabolism (Kienesberger et al. 2009; Pingitore and Romeo 2019). Similarly, one cannot rule out the possibility that even more distantly related sequences from non-animal species within the overarching orthogroup might also perform similar functions.

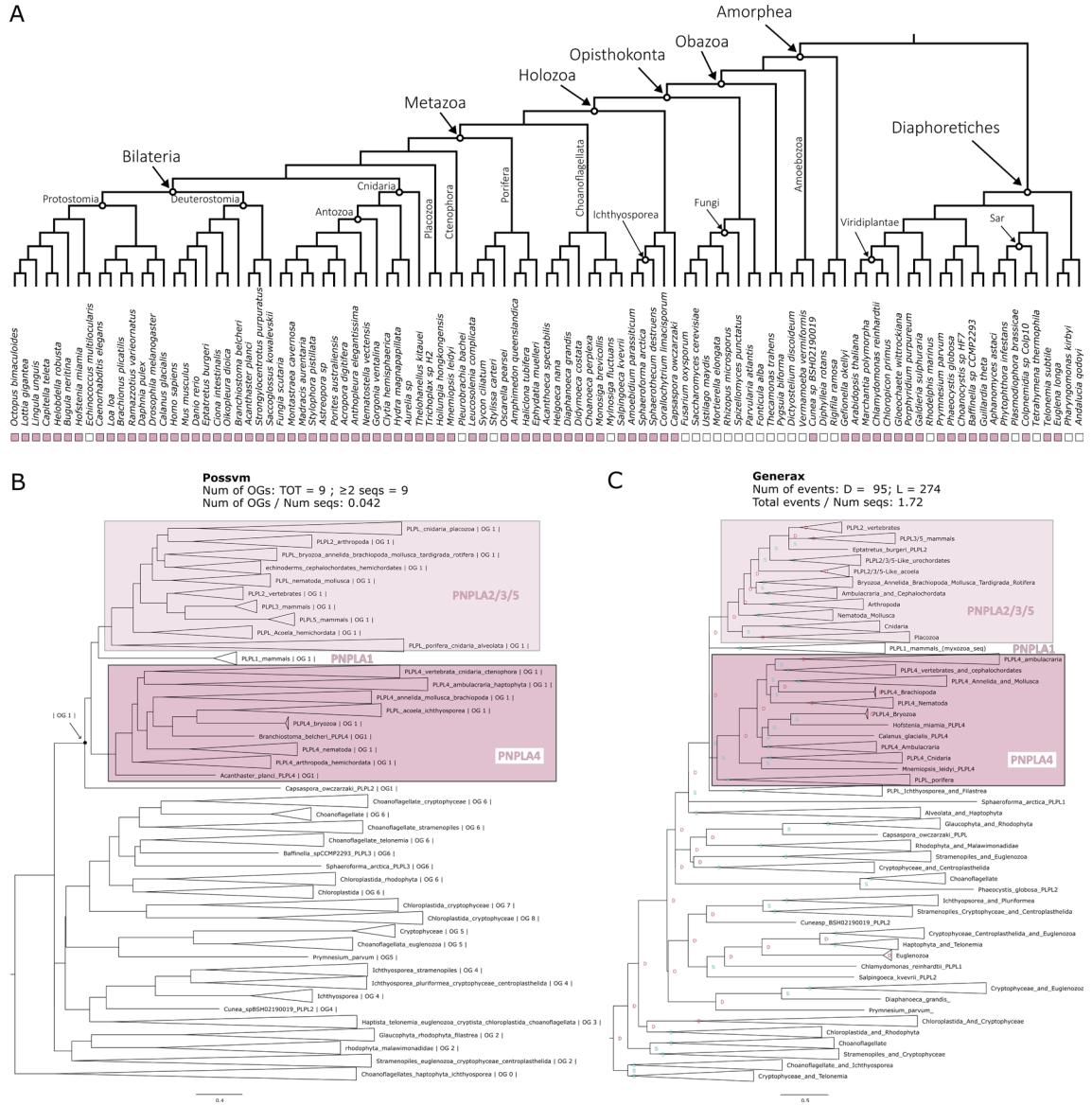


Figure 4.4. Phylogenetic analysis for the PNPLA4 orthogroup. **A.** Distribution of the orthogroup throughout Eukarya. **B.** Phylogenetic tree analysed with Possvm for identification of sub-orthogroups. Number of Possvm orthogroups identified (Total, and with ≥ 2 sequences) are provided, as well as the ratio between number of Possvm sub-orthogroups per number of sequences in full orthogroup are indicated. **C.** Gene tree to species tree reconciliation using GeneRax highlighting speciation and duplication events at each node. Numbers of duplications (D) and losses (L) are provided, as well as the ratio of total events (D+L) per number of sequences in the orthogroup.

ALDH1

Aldehyde Dehydrogenase 1 Family Member A1 (ALDH1A1 or ALDH1), also known as Retinaldehyde Dehydrogenase 1 (RALDH1), is an enzyme that can catalyse the oxidation of retinal to retinoic acid (or retinoate) (Duester 2000) (Figure 4.1). ALDH1 is involved in two KEGG pathways (Table 4.1).

Both OrthoFinder and Broccoli identify ALDH1 as its own distinct orthogroup (Figure 4.2) and the final merged orthogroup consists of 765 sequences. This orthogroup is ubiquitous, with only a handful of eukaryotic species lacking it (Figure 4.5A).

The phylogenetic analyses revealed a complex substructure within the ALDH1 orthogroup (Figure 4.5 B and C), with Possvm subdividing it into 44 orthogroups, a high number relative to total sequences. ALDH1A, ALDH1B and ALDH2 all coalesce to a same Possvm orthogroup. While the full orthogroup includes other aldehyde dehydrogenases, including ALDH1L, ALDH8A1, ALDH16A1, ALDH9A1 and ALDH5A1. The GeneRax reconciled tree found a very similar topology and identified a relatively high number of evolutionary events (Figure 4.5C). Interestingly, the ALDH1/2 sub-orthogroup predominantly features animal sequences, whereas other ALDH clades encompass a diverse range of eukaryotic species. This suggests a link between the ALDH1/2 expansion within animals and the emergence of vision in these organisms.

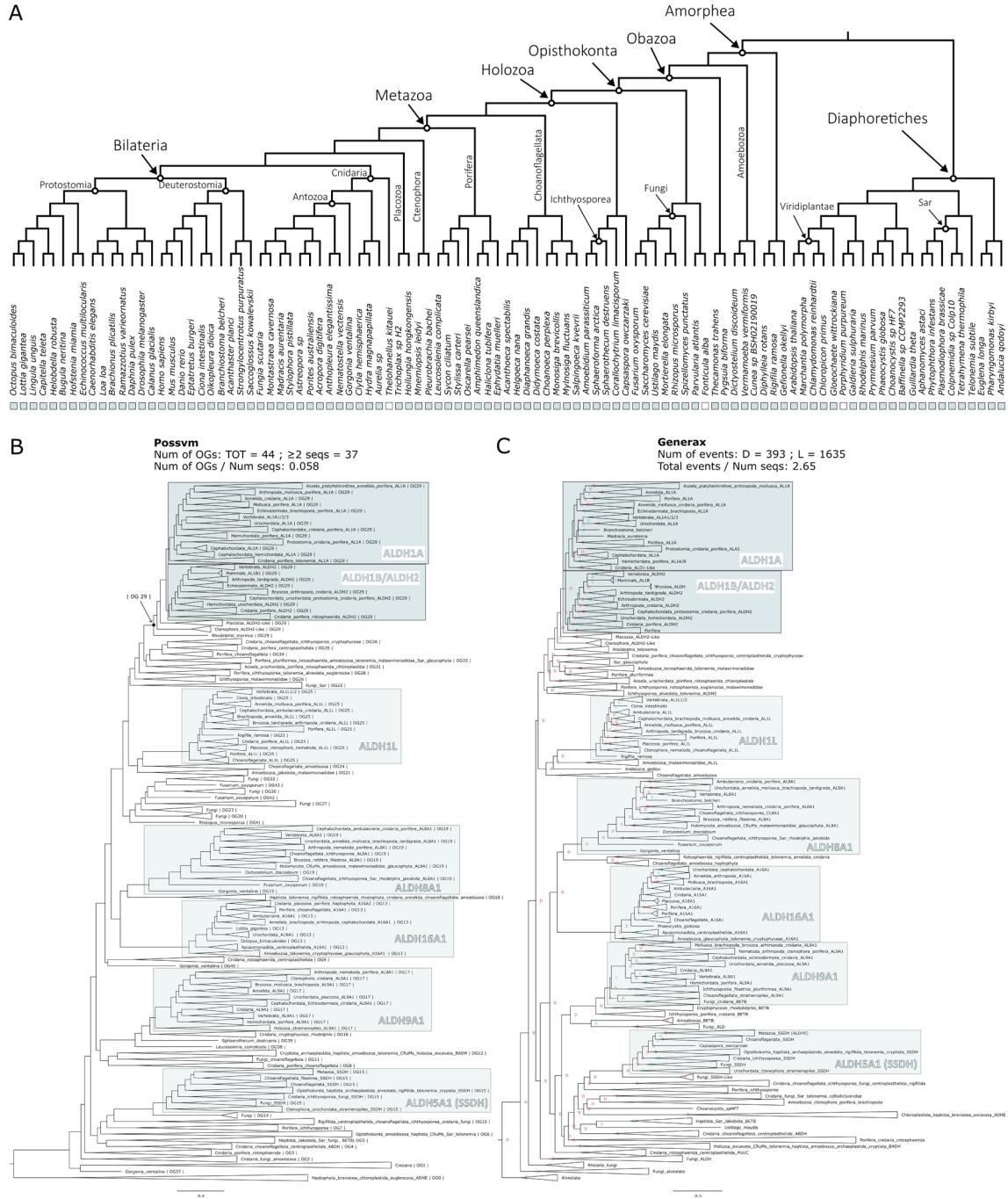


Figure 4.5. Phylogenetic analysis for the ALDH1 orthogroup. **A.** Distribution of the orthogroup throughout Eukarya. **B.** Phylogenetic tree analysed with Possvm for identification of sub-orthogroups. Number of Possvm orthogroups identified (Total, and with ≥ 2 sequences) are provided, as well as the ratio between number of Possvm sub-orthogroups per number of sequences in full orthogroup are indicated. **C.** Gene tree to species tree reconciliation using GeneRax highlighting speciation and duplication events at each node. Numbers of duplications (D) and losses (L) are provided, as well as the ratio of total events (D+L) per number of sequences in the orthogroup.

BCMO1/RPE65

Beta-carotene 15–15'-monooxygenase (BCMO1), more recently known as Beta-Carotene Oxygenase 1 (BCO1) (Seña et al. 2014), plays a crucial role in converting dietary beta-carotene into retinal by catalysing the symmetric cleavage of beta-carotene to produce two all-trans-retinal molecules (Harrison 2012) (Figure 4.1). Another carotenoid cleavage oxygenase (CCO) enzyme is Retinoid Isomerohydrolase RPE65. RPE65 is expressed in retinal pigment epithelium (RPE) cells where it catalyses the conversion of all-trans-retinyl ester to 11-cis-retinol (Jin et al. 2005; Moiseyev et al. 2005; Redmond et al. 2005). These two essential enzymes are both quite specific to the pathway, with RPE65 being present in only two KEGG pathways and BCMO1 in three (Table 4.1).

BCMO1 and RPE65 are placed in the same orthogroup both by OrthoFinder and by Broccoli (Figure 4.2) confirming that they belong to the same family of enzymes. The complete orthogroup consists in 322 sequences. This orthogroup has a patchy presence throughout most eukaryotic clades (Figure 4.6A).

The phylogenetic analysis for this orthogroup revealed several subfamilies. Possvm identified 16 orthogroups within this family, with BCO1, RPE65, as well as BCO2, belonging to the same orthogroup (Figure 4.6B). GeneRax recovers a fairly similar topology and a moderately high number of events (Figure 4.6C). Also in this case, the BCMO1/RPE65 specific subclade appears to be animal-specific; while other subgroups are either widely distributed (like ACOX) or specific to eukaryotic clades distantly related to animals (such as CCD8 and NCED/CCD1).

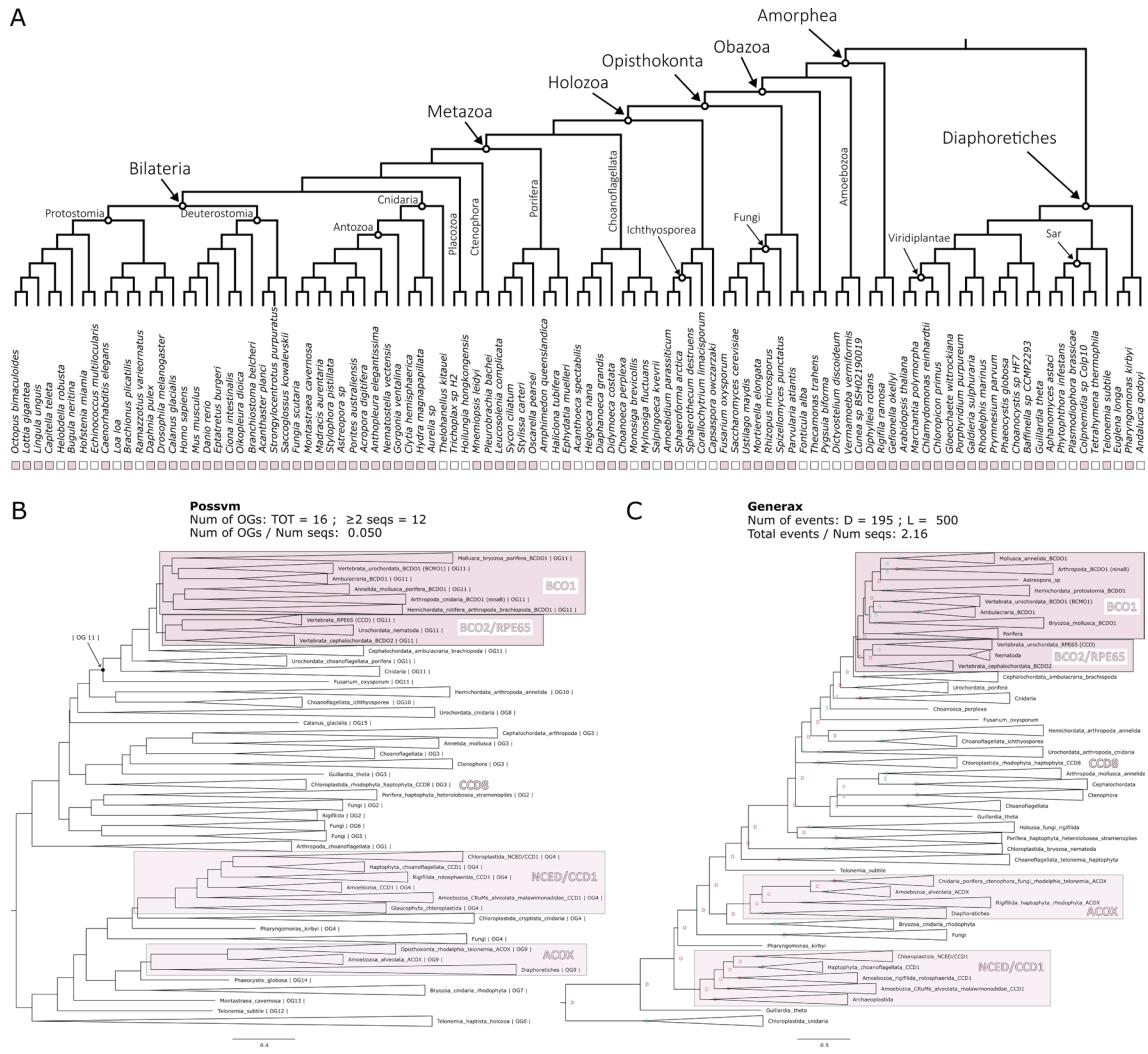


Figure 4.6. Phylogenetic analysis for the BCMO1/RPE65 orthogroup. **A.** Distribution of the orthogroup throughout Eukarya. **B.** Phylogenetic tree analysed with Possvm for identification of sub-orthogroups. Number of Possvm orthogroups identified (Total, and with ≥ 2 sequences) are provided, as well as the ratio between number of Possvm sub-orthogroups per number of sequences in full orthogroup are indicated. **C.** Gene tree to species tree reconciliation using GeneRax highlighting speciation and duplication events at each node. Numbers of duplications (D) and losses (L) are provided, as well as the ratio of total events (D+L) per number of sequences in the orthogroup.

LRAT

Lecithin Retinol Acyltransferase (LRAT), also known as Phosphatidylcholine--Retinol O-Acyltransferase, catalyses the esterification of all-trans-retinol into all-trans-retinyl ester (Ruiz et al. 1999; Batten et al. 2004) (Figure 4.1). It belongs to three KEGG pathways (Table 4.1).

OrthoFinder and GeneRax orthogroups for this enzyme correspond to each other with high identity (Figure 4.2). The LRAT orthogroup is the smallest, including only 93 sequences. This is reflected in its limited distribution throughout eukaryotes. It is present in most animal clades, with exception of placozoans and ctenophores. However, outside of animals there seems to be very sparse and uneven distribution (Figure 4.7A).

Possvm identifies only 6 orthogroups within LRAT (Figure 4.7B). Interestingly, apart from the orthogroup containing LRAT, there is also an orthogroup containing the related Phospholipase A And Acyltransferase (PLAAC) family of enzymes (Hussain et al. 2017). GeneRax confirms a similar tree topology and identifies a rather high number of events relative to the low number of sequences (Figure 4.7C). The few non-metazoan sequences within the LRAT orthogroup belong neither to the LRAT nor the PLAAC clades in the tree, making it another case in which one of the enzymes most specific to retinol metabolism appears animal specific.

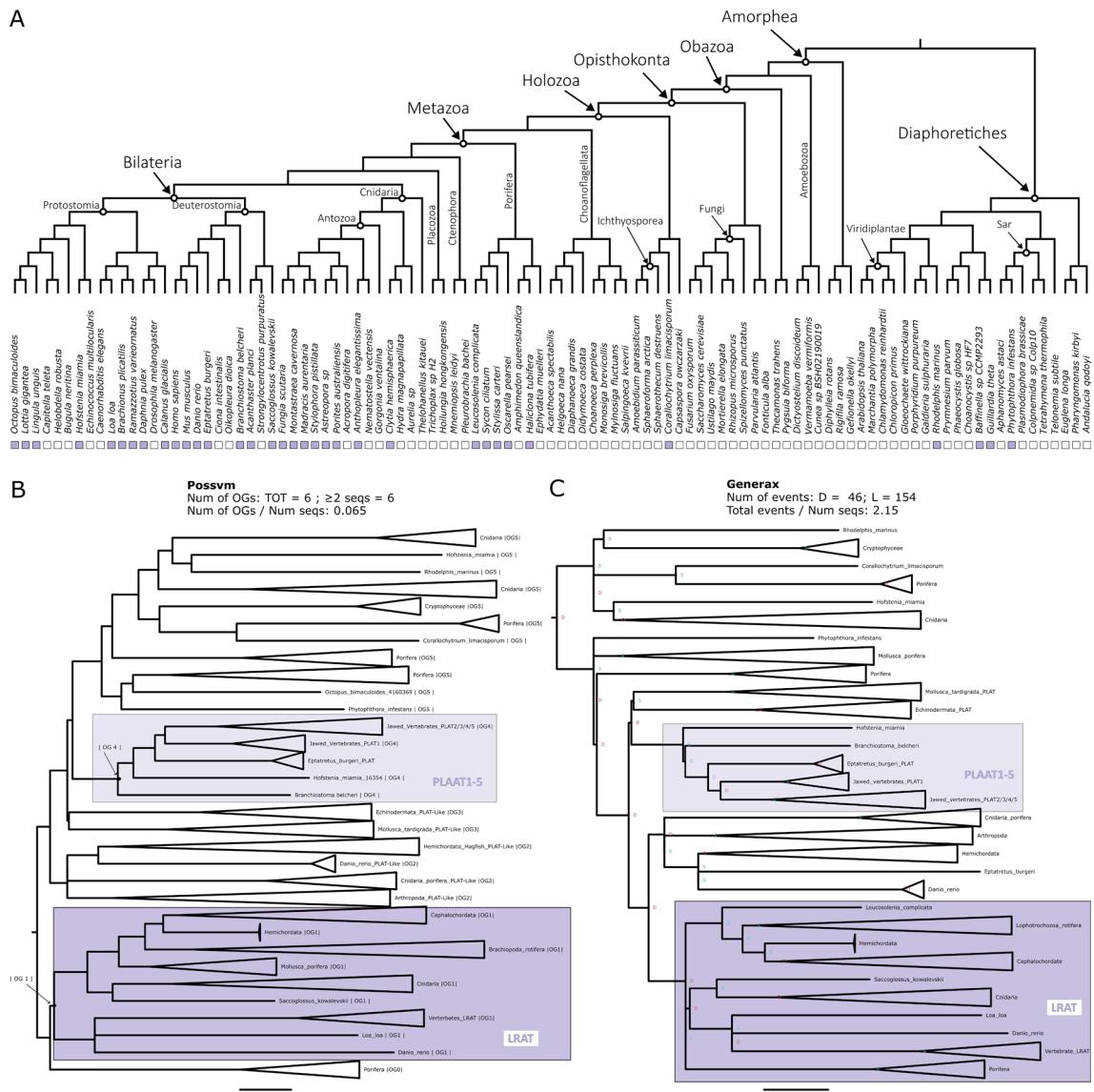


Figure 4.7. Phylogenetic analysis for the LRAT orthogroup. **A.** Distribution of the orthogroup throughout Eukarya. **B.** Phylogenetic tree analysed with Possvm for identification of sub-orthogroups. Number of Possvm orthogroups identified (Total, and with ≥ 2 sequences) are provided, as well as the ratio between number of Possvm sub-orthogroups per number of sequences in full orthogroup are indicated. **C.** Gene tree to species tree reconciliation using GeneRax highlighting speciation and duplication events at each node. Numbers of duplications (D) and losses (L) are provided, as well as the ratio of total events (D+L) per number of sequences in the orthogroup.

RDH/DHRS

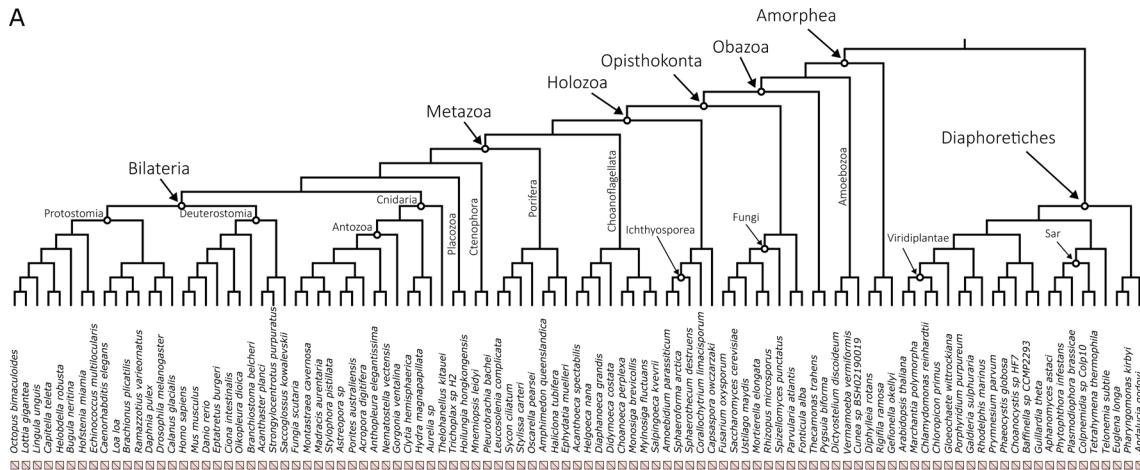
Retinol Dehydrogenase (RDH) enzymes are responsible for the oxidation of retinol to retinal (Sahu and Maeda 2016). RDH5 in particular is responsible for the conversion of 11-cis-retinol to 11-cis-retinal, the visual chromophore (Duester 2000). Other RDHs involved in the retinol metabolism are listed in Table 4.1. These enzymes are quite specific to retinol metabolism, being involved in either two or three KEGG pathways (Table 4.1). RDHs are in turn classified within the broader short-chain dehydrogenases/reductases (SDR) family (Duester 2000; Lhor and Salesse 2014). Other enzymes within this family include members of the Dehydrogenase/Reductases SDR family (DHRS), several of which are also implicated in retinol metabolism (Figure 4.1 and Table 4.1). The DHRS enzymes involved in retinol metabolism belong to a minimum of two up to a maximum of four KEGG pathways (Table 4.1).

The orthogroup analyses reveals a very complex situation for RDH and DHRS enzymes (Figure 4.2). First, there is a substantial difference in results between OrthoFinder that identifies two orthogroups and Broccoli that identifies seven orthogroups containing RDH and DHRS enzymes. Both methods pinpointed two primary orthogroups: one consisting solely of RDH enzymes and another comprising a mix of RDH and DHRS enzymes. Beyond these, Broccoli discerned several smaller orthogroups, some leaning towards an RDH profile, while others were more DHRS-specific. Two of the Broccoli orthogroups even share a very small number of sequences with the GUK orthogroup which, being unrelated to the retinol metabolism, was discarded from further analysis. Furthermore, the OrthoFinder DHRS+RDH orthogroup had a small connection with the ADH orthogroup. However, this was negligible (0.09% of identity) and ADH can confidently be regarded as a distinct orthogroup. All these considerations led to the decision to include all RDH and DHRS orthogroups into one big orthogroup for phylogenetic analysis, even when this meant dealing with a large number of sequences. This is in fact the second largest orthogroup examined in this study with a total of 4476 sequences and the only one that is present in every single species examined (Figure 4.8A).

The complexity outlined by the OrthoFinder and Broccoli orthogroup detection is reflected in the complexity of the phylogenetic tree (Figure 4.8 B and C). 207 Possvm orthogroups were defined (Figure 4.8B). The RDH and DHRS enzymes described by KEGG to be involved in retinol metabolism (Table 4.1) are distributed across 6 different

possvm orthogroups, which further clade with other members of this expansive family. GeneRax recovered a largely compatible substructure and revealed a very large number of evolutionary events even for the size of the orthogroup (Figure 4.8C). Overall, not all RDH enzymes belong to a monophyletic clade, and neither do all DHRS enzymes. Instead, monophyletic clades within this broad gene family include enzymes that have been described (based primarily on structure and function) to belong to different subfamilies. This underscores the need for a phylogenetic approach to clarify the evolutionary relationships among these enzymes. As mentioned, RDH and DHRS families are part of the extensive SDR superfamily. Delving deeper into the relationships within other SDR members might shed more light on subfamily connections. However, that would present an extremely challenging task as already the current orthogroup touched the maximum number of sequences with which this type of detailed phylogenetic analysis is feasible. Finally, regarding the distribution of specific subgroups, while most subgroups spanned eukaryotes, a handful were animal specific, such as RDH11/12, RDH13, and RDH16/H17B6/DRC7/RDH5/DHRS9. Yet, examining the larger clades these smaller orthogroups are part of reveals the presence of other eukaryotes.

Figure 4.8. Phylogenetic analysis for the RDH/DHRS orthogroup. **A.** Distribution of the orthogroup throughout Eukarya. **B.** Phylogenetic tree analysed with Possvm for identification of sub-orthogroups. Number of Possvm orthogroups identified (Total, and with ≥ 2 sequences) are provided, as well as the ratio between number of Possvm sub-orthogroups per number of sequences in full orthogroup are indicated. **C.** Gene tree to species tree reconciliation using GeneRax highlighting speciation and duplication events at each node. Numbers of duplications (D) and losses (L) are provided, as well as the ratio of total events (D+L) per number of sequences in the orthogroup.



DGAT1

Diacylglycerol O-Acyltransferase 1 (DGAT1) is known primarily for its role in triacylglycerol synthesis (Bhatt-Wessel et al. 2018). However, it has also been implicated in the retinol metabolism as an alternative to LRAT in the esterification of retinol to retinyl esters (Orland et al. 2005) (Figure 4.1). DGAT1 is involved in four metabolic pathways according to KEGG (Table 4.1).

KEGG proposes that both DGAT1 and DGAT2L4 (see below) occupy the same position in the pathway (Figure 4.1 and Table 4.1); however, the orthogroup detection analysis clearly indicates that DGAT1 and DGAT2L4 are independent orthogroups, with both OrthoFinder and Broccoli keeping them separate (Figure 4.2). Therefore, the phylogenetic analysis was performed separately for these two orthogroups. The DGAT1 orthogroup contains 246 sequences and appears to be present throughout all Eukarya with only a handful of species missing it (Figure 4.9A).

The Possvm analyses revealed a relatively simple substructure with only 7 orthogroups (Figure 4.9B). DGAT1 itself is monophyletic and belonging to one orthogroup. The Sterol O-Acyltransferase (SOAT) family appears to be closely related to DGAT1. The same substructure was described by GeneRax that also revealed a relatively low number of evolutionary events within this orthogroup (Figure 4.9C). The DGAT1 sub-orthogroup defined by Possvm includes sequences from across eukaryotes.

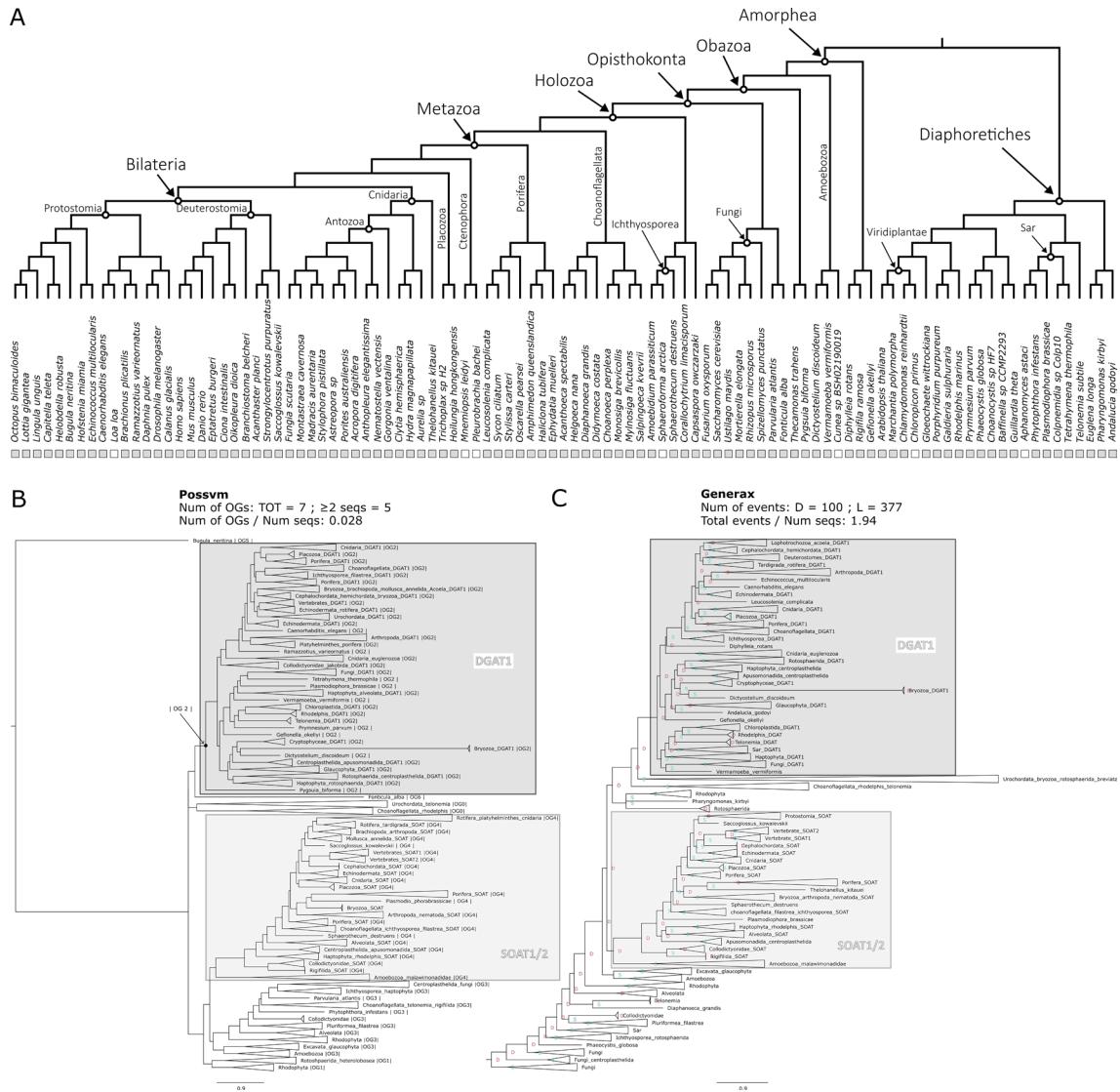


Figure 4.9. Phylogenetic analysis for the DGAT1 orthogroup. **A.** Distribution of the orthogroup throughout Eukarya. **B.** Phylogenetic tree analysed with Possvm for identification of sub-orthogroups. Number of Possvm orthogroups identified (Total, and with ≥ 2 sequences) are provided, as well as the ratio between number of Possvm sub-orthogroups per number of sequences in full orthogroup are indicated. **C.** Gene tree to species tree reconciliation using GeneRax highlighting speciation and duplication events at each node. Numbers of duplications (D) and losses (L) are provided, as well as the ratio of total events (D+L) per number of sequences in the orthogroup.

DGAT2LA4

Diacylglycerol O-Acyltransferase 2-Like Protein 4 (DGAT2L4), also known as Acyl-CoA Wax Alcohol Acyltransferase 2 (AWAT2), is primarily known for its role in the production of wax esters (Cheng and Russell 2004). It has also been recently implicated in the conversion of retinol to retinyl ester (Kaylor et al. 2014; Arne et al. 2017; Blaner 2017) (Figure 4.1). According to KEGG this enzyme is involved in three metabolic pathways (Table 4.1).

Although DGAT2LA4 indeed seems to be involved in the same step as DGAT1 (and LRAT), it appears to form its own distinct orthogroup (see above) (Figure 4.2). This orthogroup includes 372 sequences and is present in all eukaryotes with few species missing it (Figure 4.10A).

Possvm identified 23 orthogroups, which quite high for the number of sequences (Figure 4.10B). DGAT2L4 forms a monophyletic clade with DGAT2L2, DGAT2L3, DGAT2L6 and DGAT2. While DGAT2L1 and DGAT2L5 form another monophyletic clade, sister group to the previous one (Figure 4.10B). Both clades, together with other less well characterized sequences, belong to one Possvm orthogroup. The same relationships are maintained in the reconciled tree by GeneRax that calculated quite a high number of events (Figure 4.10C). While the clades encompassing the DGAT2L and DGAT2 genes are specific to animals, the same Possvm orthogroup contains various non-metazoan sequences. This implies that this gene family existed anciently, even if animal-specific expansions gave rise to the recognized enzymes with a marginal role in retinol metabolism.

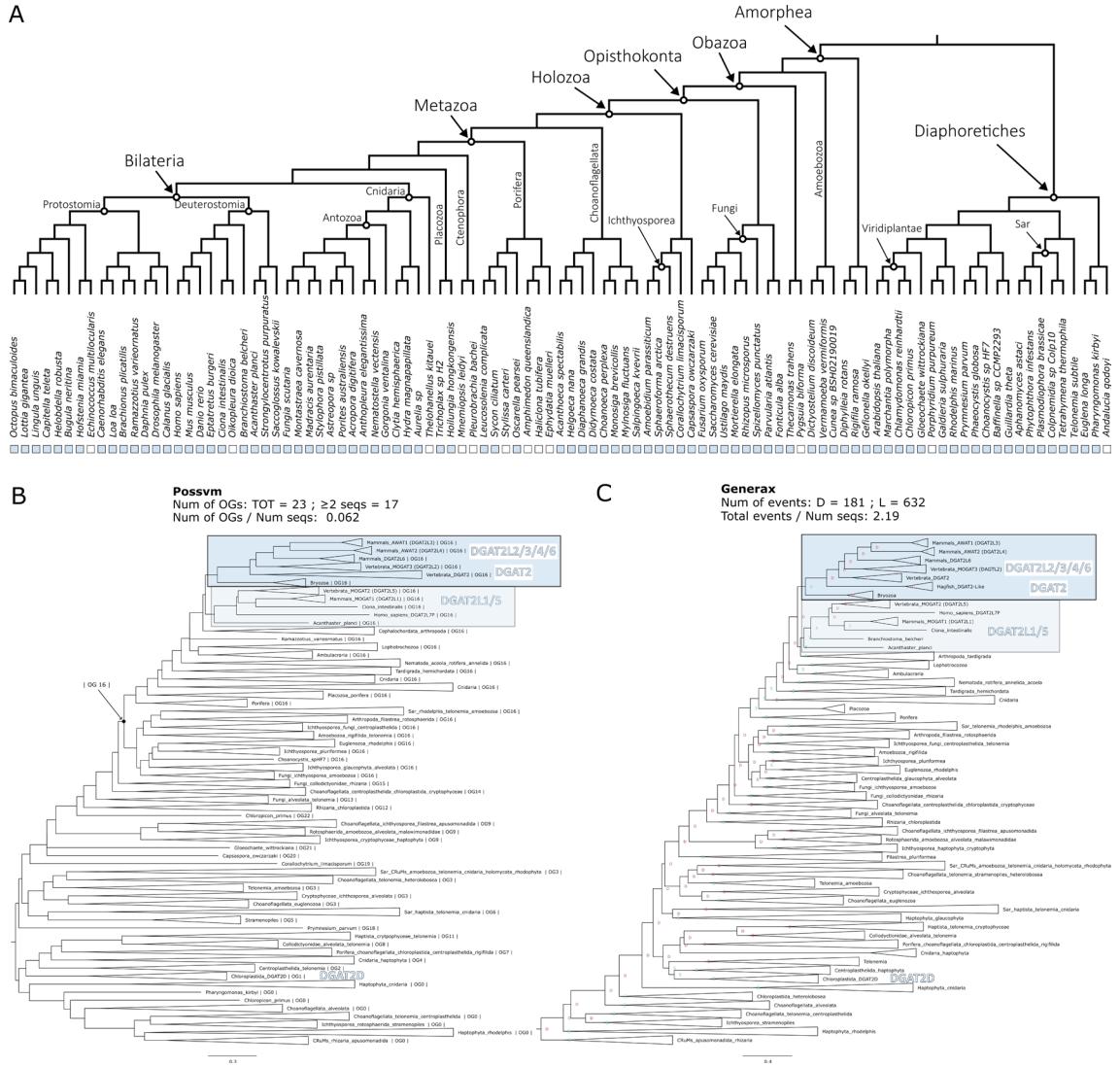


Figure 4.10. Phylogenetic analysis for the DGAT2L4 orthogroup. **A.** Distribution of the orthogroup throughout Eukarya. **B.** Phylogenetic tree analysed with Possvm for identification of sub-orthogroups. Number of Possvm orthogroups identified (Total, and with ≥ 2 sequences) are provided, as well as the ratio between number of Possvm sub-orthogroups per number of sequences in full orthogroup are indicated. **C.** Gene tree to species tree reconciliation using GeneRax highlighting speciation and duplication events at each node. Numbers of duplications (D) and losses (L) are provided, as well as the ratio of total events (D+L) per number of sequences in the orthogroup.

CYP

Cytochrome P450 (CYP) enzymes represent a large and diverse family of heme-containing enzymes involved in the synthesis and metabolism of a wide range of compounds (Zhao et al. 2021). The number of CYP enzymes is so vast that it is generally considered to be a super family in turn subdivided into families and subfamilies (Nelson 2018). For example, the CYP27C1 enzyme, the most specific to the retinol metabolism (Table 4.1), belongs to the family 27, subfamily C, and is the member 1. It catalyses the 3,4 desaturation of all-trans-retinol to all-trans-3,4-didehydroretinol (Enright et al. 2015; Kramlinger et al. 2016; Corbo 2021) (Figure 4.1). The other CYP enzymes involved in the retinol metabolism have varied degree of specificity and are listed in Table 4.1.

While being a vast family, the orthogroup identification was straightforward, with OrthoFinder and Broccoli results coinciding (Figure 4.2). The total orthogroup contained 4499 sequences, making it the largest group examined in this study. The distribution also spans all of Eukarya with only three species of the 101 examined lacking it (Figure 4.11A).

Possvm identified 74 orthogroups (Figure 4.11B), meaning that while being slightly larger than the RDH/DHRS orthogroup, it is overall much less fragmented. Nevertheless, the CYP enzymes described to be involved in the retinol metabolism (Table 4.1) are not all belonging to the same Possvm orthogroup, nor to one monophyletic clade, but rather span 5 separate monophyletic clades. These groups are confirmed with the GeneRax reconciliation (Figure 4.11C) that also identifies a relatively low amount of duplication and loss events considering the number of sequences in the orthogroup. Overall, monophyletic clades encompassing CYP enzymes implicated in retinol metabolism contain sequences spanning most eukaryotic groups.

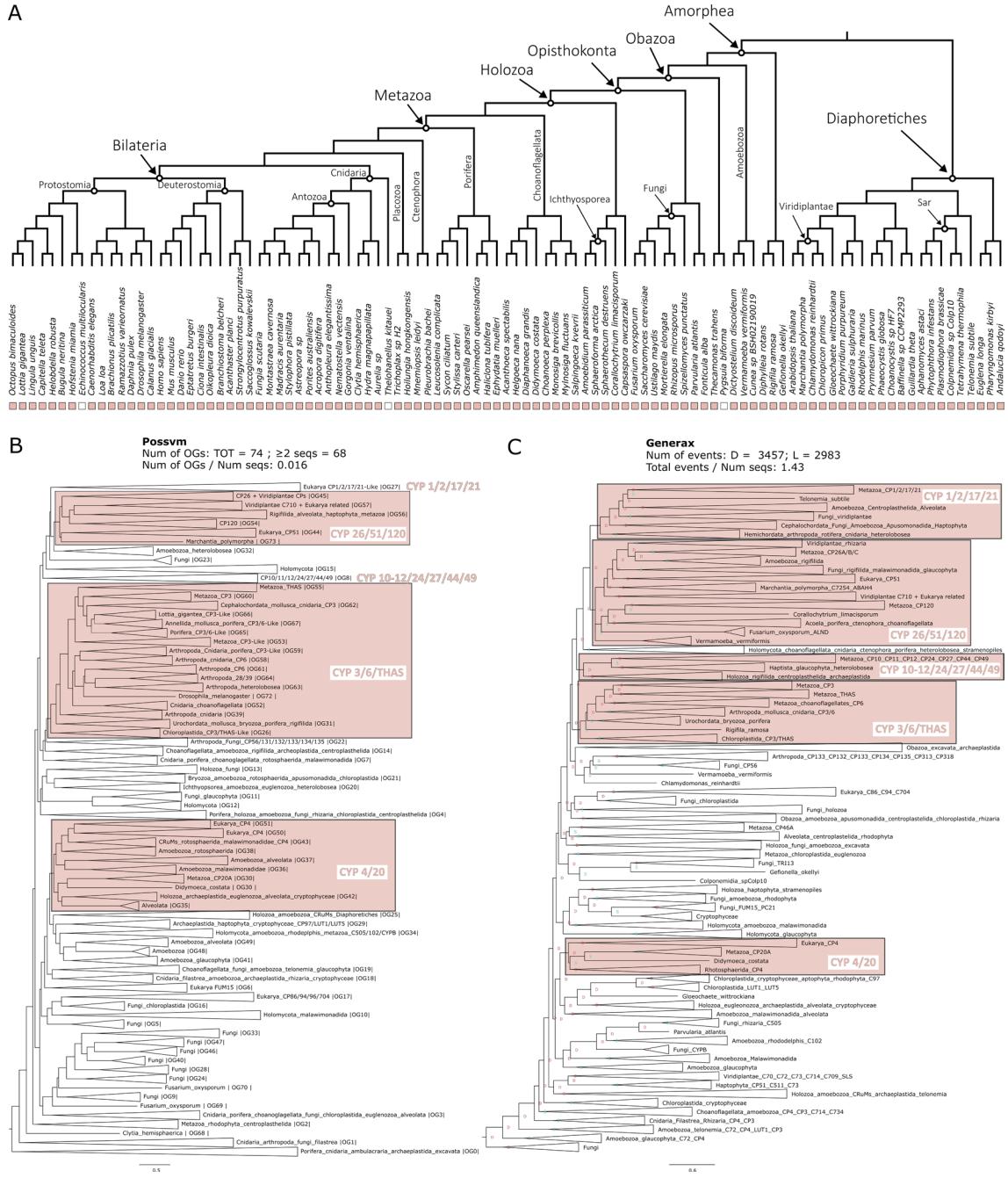


Figure 4.11. Phylogenetic analysis for the CYP orthogroup. **A.** Distribution of the orthogroup throughout Eukarya. **B.** Phylogenetic tree analysed with Possvm for identification of sub-orthogroups. Number of Possvm orthogroups identified (Total, and with ≥ 2 sequences) are provided, as well as the ratio between number of Possvm sub-orthogroups per number of sequences in full orthogroup are indicated. **C.** Gene tree to species tree reconciliation using GeneRax highlighting speciation and duplication events at each node. Numbers of duplications (D) and losses (L) are provided, as well as the ratio of total events (D+L) per number of sequences in the orthogroup.

AOX

Aldehyde Oxidase 1 (AOX1) is responsible for the oxidation of a wide variety of aldehydes to their corresponding carboxylic acids (Terao et al. 2016). Within the retinol metabolism it is able to oxidise retinal to retinoate (Terao et al. 2016) (Figure 4.1), although the primary enzyme for this is ALDH1 (see above). Overall AOX1 is not to be considered specific to the retinol metabolism (Table 4.1).

The identification of the AOX orthogroup presented slight differences between OrthoFinder, which found one orthogroup, and Broccoli, that split the family into two orthogroups, the AOX and the AAO (Abscisic-aldehyde oxidase), a group of aldehyde oxidases primarily known in plants (Seo et al. 2000). The total orthogroup of AOX includes 599 sequences. It is overall present in all eukaryotes with some exceptions, e.g., ctenophores (Figure 4.12A).

Possvm identified 25 orthogroups (Figure 4.12B). The phylogenetic analysis uncovered how the Xanthine Dehydrogenase (XDH) family is closely related to the AOX. While the AAO (present primarily in Diaphoretiches) is more distantly related. This is confirmed in the reconciled GeneRax tree that also revealed a moderate number of events (Figure 4.12C). Interestingly, while the AOX clade is limited to a specific subset of animal species, the closely related XDH clade encompasses sequences from a diverse array of eukaryotes.

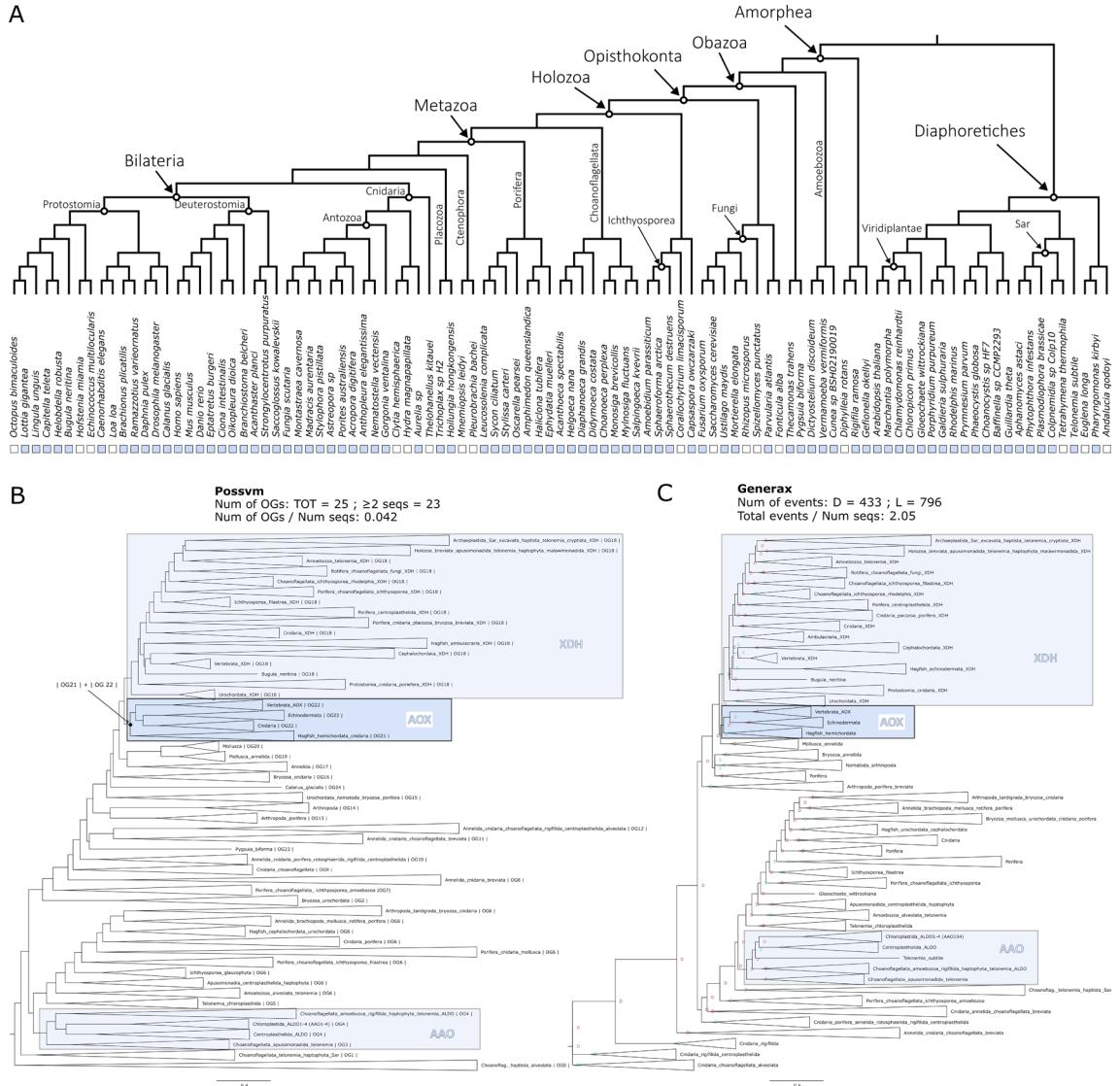


Figure 4.12. Phylogenetic analysis for the AOX orthogroup. **A.** Distribution of the orthogroup throughout Eukarya. **B.** Phylogenetic tree analysed with Possvm for identification of sub-orthogroups. Number of Possvm orthogroups identified (Total, and with ≥ 2 sequences) are provided, as well as the ratio between number of Possvm sub-orthogroups per number of sequences in full orthogroup are indicated. **C.** Gene tree to species tree reconciliation using GeneRax highlighting speciation and duplication events at each node. Numbers of duplications (D) and losses (L) are provided, as well as the ratio of total events (D+L) per number of sequences in the orthogroup.

ADH

Alcohol dehydrogenase (ADH) enzymes play crucial roles in the metabolism of alcohols (Edenberg 2007). In the retinol metabolism ADHs play a role in the oxidation of retinol to retinal (Duester 2000) (Figure 4.1). Although RDH is the primary enzyme for this reaction, particularly within the retina, ADHs can also contribute and within humans are especially used in non-visual related tissues such as the liver (Duester 2000). Seeing as ADHs are involved in metabolising a wide variety of alcohols it is not surprising that they are involved in numerous other pathways other than the retinol metabolism (Table 4.1).

During the identification of an orthogroup for ADH, OrthoFinder placed all sequences in one orthogroup, while Broccoli split the family into two orthogroups. One primarily comprised ADH sequences, while the other was a mixed group that incorporated the related Sorbitol Dehydrogenase (SORD) (Figure 4.2). The merged orthogroup consisted of 955 sequences and was present in all but one species (Figure 4.13A).

Ortholog analysis with Possvm revealed a complex substructure, with 59 orthogroups identified (one of the highest numbers relative to orthogroup size) (Figure 4.13B). Possvm split the various ADH enzymes into different orthogroups, with ADH5 being the most distantly related. Nevertheless, all ADHs belonged to a larger monophyletic group. Other families picked up in this broad orthogroup are Cinnamyl alcohol dehydrogenase (CADH), Succinate-semialdehyde dehydrogenase (SUCD), and Sorbitol Dehydrogenase (SORD). The GeneRax reconciled tree maintains the same overall topology and a large number of events were calculated, one of the highest relative to number of sequences (Figure 4.13C). The ADH1/4/6/7 group seems to represent a mammalian-specific expansion within the family. In contrast, ADH5 appears ancient, comprising sequences from many different eukaryotic groups.

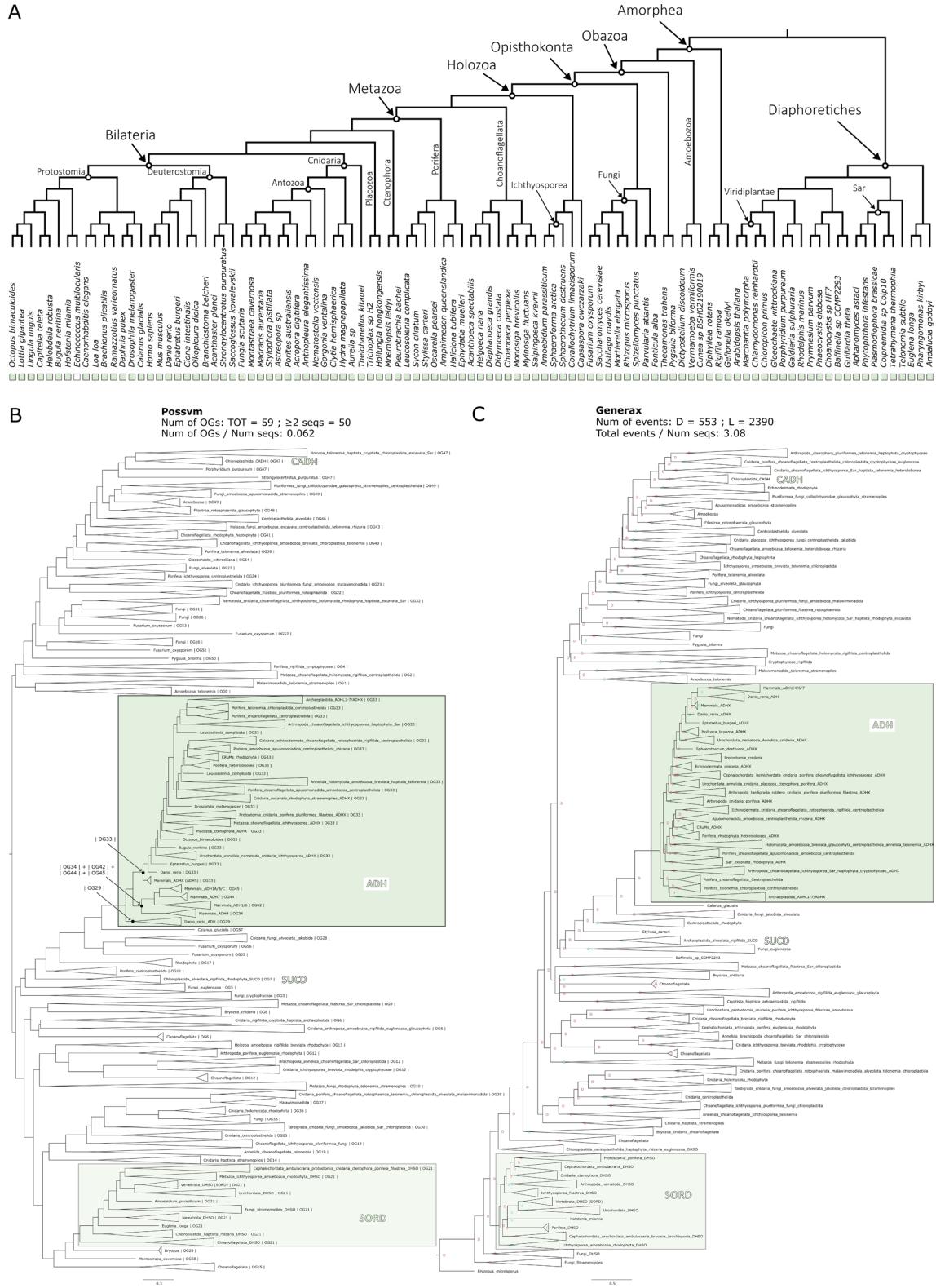


Figure 4.13. Phylogenetic analysis for the ADH orthogroup. **A.** Distribution of the orthogroup throughout Eukarya. **B.** Phylogenetic tree analysed with Possvm for identification of sub-orthogroups. Number of Possvm orthogroups identified (Total, and with ≥ 2 sequences) are provided, as well as the ratio between number of Possvm sub-orthogroups per number of sequences in full orthogroup are indicated. **C.** Gene tree to species tree reconciliation using GeneRax highlighting speciation and duplication events at each node. Numbers of duplications (D) and losses (L) are provided, as well as the ratio of total events (D+L) per number of sequences in the orthogroup.

UGT

UDP-glucuronosyltransferase (UGT) enzymes are involved in the process of glucuronidation of small lipophilic molecules, whereby a glucuronic acid is transferred from a UDP-glucuronic acid to the small molecule, making it more water soluble and therefore easier to excrete from the body (Rowland et al. 2013). In mammals there are four UGT families: UGT1; UGT2; UGT3; and UGT8 (Meech et al. 2019). UGTs are involved in the regulation of retinoid levels in the body; by glucuronidating all-trans-retinoate to all-trans-retinoyl beta-glucuronide it facilitates the excretion of this molecule (Meech et al. 2019) (Figure 4.1). Overall, this enzyme family is very broad spectrum (Table 4.1) and involved only marginally in the retinol metabolism, nevertheless we included it in our evolutionary study.

UGTs are clearly identified as being an independent orthogroup by both OrthoFinder and Broccoli (Figure 4.2). This orthogroup consists of many sequences (1005 sequences). Interestingly, while present in both major branches of eukaryotes, it appears to be missing in several clades, including several unicellular holozoans (such as ichthyosporeans) that are closely related to animals, although it is present in the sister group to animals, the choanoflagellates (Figure 4.14A).

The phylogenetic analysis uncovers that UGT1 and UGT2 are closely related to each other, as are UGT3 and UGT8. However, all of them belong to a single monophyletic clade, which Possvm identifies as one orthogroup (Figure 4.13B). The GeneRax reconciled tree maintains this topology (Figure 4.14C). Overall, Possvm identifies a total of 21 orthogroups and GeneRax identifies the lowest ratio of events to sequences from all orthogroups examined. Collectively, this indicates that the UGT orthogroup is rather conserved. The UGT1/2/3/8 monophyletic clade predominantly consists of deuterostome (vertebrates and their close relatives) sequences within a Possvm orthogroup that includes only animal sequences. Nevertheless, the rest of the broad orthogroup contains a diverse array of eukaryotic sequences, including an apparently plant specific clade of UGTs.

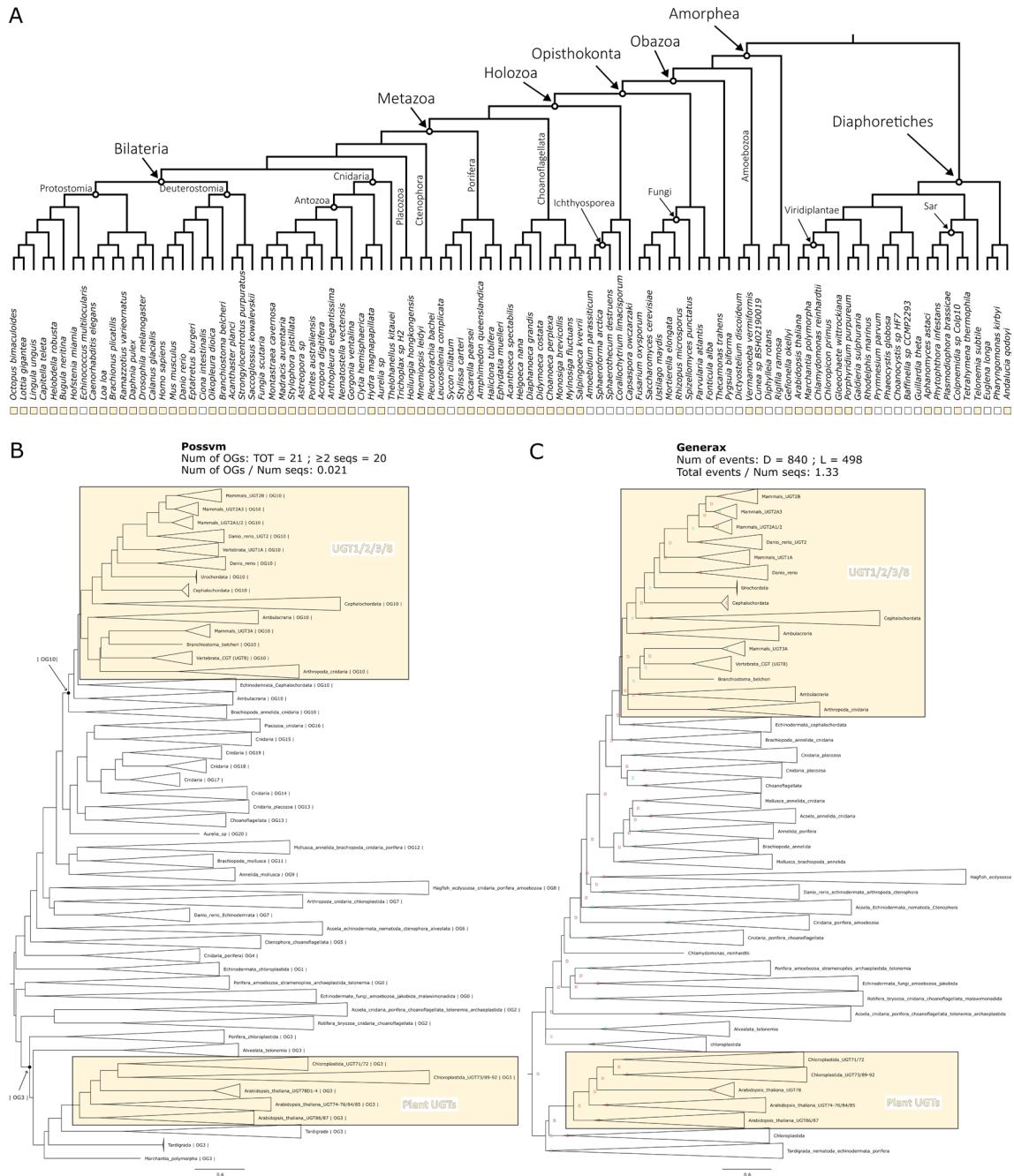


Figure 4.13. Phylogenetic analysis for the UGT orthogroup. A. Distribution of the orthogroup throughout Eukarya. B. Phylogenetic tree analysed with Possvm for identification of sub-orthogroups. Number of Possvm orthogroups identified (Total, and with ≥ 2 sequences) are provided, as well as the ratio between number of Possvm sub-orthogroups per number of sequences in full orthogroup are indicated. C. Gene tree to species tree reconciliation using GeneRax highlighting speciation and duplication events at each node. Numbers of duplications (D) and losses (L) are provided, as well as the ratio of total events (D+L) per number of sequences in the orthogroup.

Conclusions

Vision, a distinguishing feature of the animal kingdom, hinges on a specific light-sensitive molecule initiating the phototransduction pathway. This molecule is the visual chromophore 11-cis-retinal bound to the membrane protein opsin in photoreceptor cells. When 11-cis-retinal absorbs light, it isomerises into all-trans-retinal, setting off the phototransduction process (Lamb 2020). Continuous light detection demands that this chromophore be perpetually restored to its original 11-cis state. This is obtained through the retinol metabolism, a pathway essential to both vision and other biological functions (Blomhoff and Blomhoff 2006). Thus, understanding the origins and evolution of vision necessitates exploring the evolution of retinol metabolism, which sustains our light sensitivity.

The aim of this chapter was to comprehensively explore the evolution of retinol metabolism genes to provide insights on the origin of the pathway. Initially, I explored the relationships among the enzymes integral to the pathway, grouping them into overarching gene families or orthogroups. Subsequently, I analysed the distribution of these orthogroups across eukaryotes. Finally, I outlined the evolutionary events that have shaped the history of each respective orthogroup.

Enzymes can be classified into families based on several criteria, with enzymatic activity being one of the most predominant (Webb 1992). Through my orthogroup analyses, I sought to determine if the enzymes involved in retinol metabolism could be grouped according to their evolutionary relationships and how these groups align with established enzymatic families. My results suggest that enzymes integral to retinol metabolism can be categorized into 12 distinct orthogroups (Figure 4.2, Table 4.3). Some of these enzymes play pivotal roles in the critical steps for recycling 11-cis-retinal, while others have more peripheral functions (Figure 4.15A). This analysis generally aligned with the established enzymatic families, yet it also shed light on some unexpected findings. For instance, while Diacylglycerol O-Acyltransferase enzymes have conventionally been classified under a single overarching family, my findings provide compelling evidence that DGAT1 exhibits significant evolutionary divergence from DGAT2 and its related molecules, such as DGAT2L4, warranting their categorization into distinct orthogroups.

(Figure 4.2). Another example involves the enzymes responsible for converting retinol to retinal, which belong to subfamilies within the expansive SDR family. While prevailing nomenclatures suggest a distinction between RDH and DHRS, the orthogroup analyses suggested a more complex web of relationships (Figure 4.2), subsequently corroborated by phylogenetic analyses (Figure 4.8). Ultimately all this suggests that within the SDR family, phylogenetic relationships may define different subfamilies compared to the currently established ones.

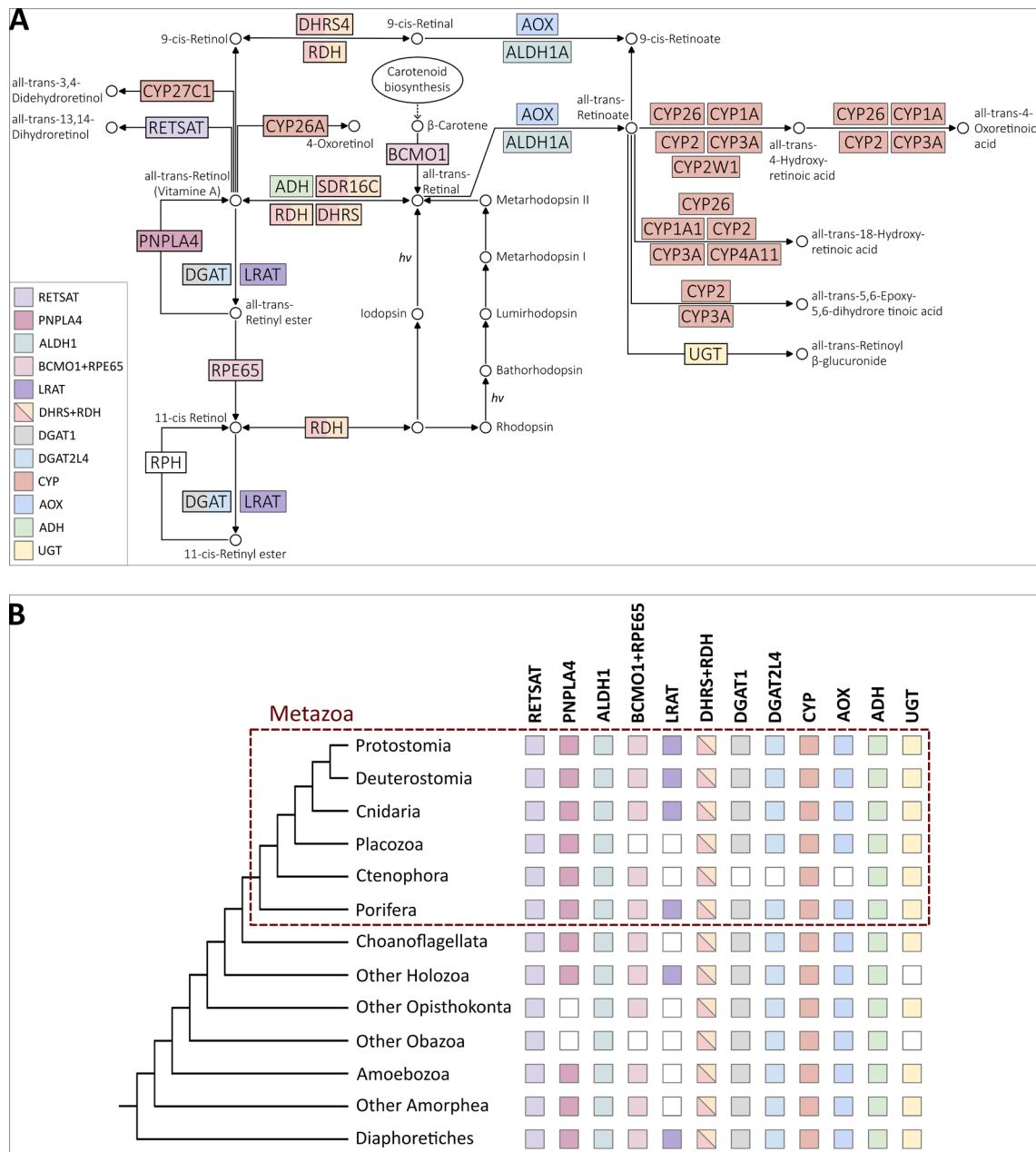


Figure 4.15. Summary of the orthogroups to which enzymes involved in retinol metabolism belong to and their distribution. A. Overlap of the retinol metabolism related orthogroups identified in this study with the KEGG map00830 pathway components. Some enzymes belong to the same orthogroup (e.g., RDH

and DHRS) while some are split into different orthogroups (e.g., DGAT1 and DGAT2L4). **B.** Orthogroups are ancient and widely distributed throughout Eukarya, except LRAT.

Examining the distribution of orthogroups throughout Eukarya, the first consideration is that most orthogroups have ancient origins, spanning the major eukaryotic clades (Figure 4.15B). The only exception is LRAT that appears to be present primarily in animals (except placozoans and ctenophores) and in a handful of other species. While orthogroups tend to be present in all major eukaryotic clades, only one orthogroup (RDH/DHRS) was present in every single species examined (Figure 4.8). Within the realm of animals, only the two non-bilaterian phyla of placozoans and ctenophores seem to lack some key orthogroups, notably BCMO1/RPE65 and LRAT, which are integral to the retinol metabolism pathway. This observation is intriguing for two main reasons. First, the absence of these enzymes might suggest that these two phyla have evolved a variant of the retinol metabolism where these missing enzymes are replaced by others. For instance, the function of LRAT can also be carried out by DGAT1 and DGAT2L4. Secondly, even though sponges lack opsins, they possess all the retinol metabolism orthogroups (although not every sponge species has all orthogroups). This has interesting implications in understanding the origin of vision. It reinforces the growing interest surrounding potential opsin-independent light-sensing processes in sponges (Wong et al. 2022), as discussed in Chapter 3.

Further insights came with the detailed phylogenetic analyses. The approach of using both Possvm and GeneRax provided a dual advantage: Possvm facilitated the swift identification and annotation of sub-orthogroups; while GeneRax ensured a thorough and accurate description of evolutionary events of duplication, speciation, and losses. These events serve as a gauge of the evolutionary intricacies (as opposed to a more linear progression) and can also account for the degree of subfamily diversity observed within orthogroups. These analyses facilitated the identification of orthogroups with particularly complex evolutionary histories. For instance, the ADH orthogroup showcased one of the highest ratios of sub-orthogroups to sequences and also had one of the highest ratios of evolutionary events to sequences. Furthermore, through these analyses, we can juxtapose orthogroups of comparable sizes. This allows, for example, to discriminate between large orthogroups with simple subgroup structure, such as CYP, and large orthogroups with more intricate substructures, like RDH/DHRS.

Another compelling insight from the phylogenetic analyses was the refined understanding of distribution patterns within Eukarya, allowing for a detailed mapping of subgroup distributions beyond just the overarching orthogroups. An intriguing observation that emerged was that precisely the enzymes that are most specific to retinol metabolism (PNPALA4, ALDH1, BCMO1/RPE65) presented a distinctive pattern where the overarching orthogroups spanned across Eukarya, but the subfamilies actively utilized in the pathway were predominantly animal specific. This suggests that animal specific expansions of these enzyme families may have evolved concurrently with the development of vision. However, it should be noted that other subfamilies within those orthogroups might still hold the potential to execute similar functions.

To conclude, the ancient origin of the enzyme families involved in the replenishment of 11-cis-retinal suggest that the foundational molecular framework was already in place well before the emergence of vision. Expansions in these enzymes families during early animal evolution could have steered this pathway towards its specialized role in visual functions. This study's detailed reconstruction of retinol metabolism enzyme families paves the way for exploring a whole new set of questions regarding the evolution of vision. One vital question that arises is whether the enzymes identified bioinformatically within early-branching animals, truly operate in a physiological setting to facilitate the recycling of 11-cis-retinal in these animals. This is an especially fascinating enquiry to pursue for organisms like sponges which, lacking opsins, in theory should not rely on this pathway. Another intriguing avenue for future research is to understand the evolution of the cell types involved in this pathway. In the human retina, for instance, several cell types, including the Retinal Pigment Epithelium (RPE) cells and Mueller cells, play roles in the visual chromophore replenishment (Arshavsky 2002; Mata et al. 2002; Thompson and Gal 2003; Moiseyev et al. 2005; Strauss 2005). It would be compelling to determine whether homologous functions are carried out by homologous cell types throughout animals, including in early-branching animals. Alternatively, different animals might manage parts of the pathway using unrelated cell types or execute the functions within a single cell type. Thus, future research on the evolution of vision should expand their focus beyond just the evolution of photoreceptor cells, which perform phototransduction, to also include cell types implicated in retinol metabolism.

Methods

Identification of orthogroups for retinol metabolism enzymes.

Species list and species tree

To understand the evolution of the retinol metabolism, I selected 101 eukaryotic species (Table 4.2 and Extended Table 4.2) in which to search for the genes involved in the pathway. The choice of species was based on a combination of balanced taxonomic sampling throughout Eukarya and quality of the proteomes. The latter was assessed using BUSCO (v4.0.6) (Simão et al. 2015; Waterhouse et al. 2018) with the eukaryota_odb10 database. The final selection included 50 animals, of which 25 non-bilaterians, 13 unicellular holozoans closely related to animals, and various other species from all major eukaryotic clades.

The single-copy BUSCO genes obtained from the BUSCO analysis were also used to construct a species tree. This is because knowledge of species relationships can be used both for orthogroup inference with the OrthoFinder software (Emms and Kelly 2015; Emms and Kelly 2019) and to construct species-tree-aware gene trees (Boussau and Scornavacca 2020) using software such as GeneRax (Morel et al. 2020) (see more details below). The species tree was constructed by: aligning single-copy BUSCO genes with MAFFT v7.470 (--auto) (Katoh et al. 2002; Katoh and Standley 2013); trimming alignments with Trimal v1.4.rev22 (-automated1) (Capella-Gutiérrez et al. 2009); concatenating alignments into a super-matrix using FASconCAT v1.11 (Kück and Meusemann 2010); maximum-likelihood tree construction using IQTREE v2.0.6 (Hoang et al. 2018; Minh et al. 2020) after identifying the best-fitting phylogenetic model with the IQTREE2 Model Finder (Kalyaanamoorthy et al. 2017). The resulting tree was inspected to confirm that species and phyla relationships were compatible with the known literature and where necessary Mesquite v3.6.1 (Maddison and Maddison 2008) was used to correct branch positions. The species tree used in this chapter (available on GitHub) places sponges as sister-group to all other animals as this is one of the currently accepted scenarios (Feuda et al. 2017; Schultz et al. 2023). Furthermore, my previous work presented in Chapter 3 showed that no substantial difference was detected between sponge-first and ctenophore-first scenarios when performing gene-tree to species-tree reconciliations using a eukaryotic-wide set of organisms (see Supplementary Table S3.2).

Data mining

Enzymes for the retinol metabolism were chosen based on the pathway described on KEGG Database (KEGG map00830) (Kanehisa et al. 2021). Queries for BLASTP were collected from the KEGG Orthology lists (Kanehisa 2019) for each component of the pathway. BLASTP (Camacho et al. 2009) was conducted (with e-value threshold of 1e-5) for each query against the species database. To provide a preliminary annotation also for sequences from non-annotated non-model organisms, these were BLASTed versus the SwissProt Database (Poux et al. 2017) and the top hit was used as an approximate annotation.

Orthogroup inference

The results from BLASTP, organised by species, were used as “mini-proteomes” for orthogroup inference. By having reduced species proteomes by narrowing down to sequences with sequence similarity with the target enzymes of interest, it is in fact possible to reduce the computational load which is quite extensive for this type of analysis on large numbers of species. Two alternative methodologies for orthogroup inferences were used and compared in this work. In this way it was possible to verify the consistency of results when using different software. It also allowed to make sure not to miss out any potential sequences belonging to the orthogroups for the enzymes under investigation.

OrthoFinder

To insure best possible accuracy, OrthoFinder v.2.5.4 (Emms and Kelly 2015; Emms and Kelly 2019) was run with BLAST search (instead of default DIAMOND) and with the MSA workflow (using the default MAFFT for alignment and FastTree for tree inference). Furthermore, the species tree was provided (see above) rather than inferred by OrthoFinder. The inflation parameter used for MCL clustering was 1.3.

Broccoli

Broccoli v1.2.1 (Derelle et al. 2020) was run with kmer length for sequence clustering set to 80 to account for the distantly related species analysed; for the phylogeny step, maximum likelihood was chosen to maximise accuracy. Finally, regarding the species

overlap parameter, several values were tested and finally the value of 0.9 was found to be the best compromise between orthogroup accuracy (usually obtained with lower values) and avoidance of orthogroup fragmentation.

Filtering and annotation of orthogroups

To reach the goal of identifying orthogroups for the enzymes involved in the retinol metabolism, the orthogroups inferred by OrthoFinder and Broccoli must be annotated and potential unrelated orthogroups discarded. As a first step, all orthogroups that contained less than 4 sequences, or less than 4 species were discarded. Then, all sequences from each orthogroup were annotated using EggNog mapper (Cantalapiedra et al. 2021). One of the annotation fields outputted by EggNog is KEGG_pathways. Therefore, this was exploited to filter out any orthogroup that did not contain at least one sequence that obtained the KEGG map00830 (retinol metabolism) annotation. In this way it was possible to narrow down the number of orthogroups to analyse to identify orthogroups for our target enzymes. The remaining orthogroups were annotated by identifying the human sequences contained in them.

Comparison of OrthoFinder and Broccoli results and definition of final orthogroups

All enzymes known to be involved in the retinol metabolism were recovered as one or more orthogroup by both OrthoFinder and Broccoli. To assess the consistency between the results of the two methods, the next step was to compare the orthogroups by checking percentage of shared identical sequences amongst all OrthoFinder and Broccoli orthogroups (Figure 4.2). This comparison was visualised using Cytoscape v3.9.1 (Shannon et al. 2003), where orthogroups are represented as nodes and edges connecting the nodes represent the percentage of identical sequences shared between orthogroups. One-to-one correspondence with high percentage of identity was recovered in most cases and overall, it was possible to clearly establish the correspondence between OrthoFinder and Broccoli orthogroups. Final orthogroups used for subsequent phylogenetic analyses were the combined sequences collected with OrthoFinder and Broccoli. Cd-hit (Li et al. 2001; Fu et al. 2012) was used to remove duplicates with 100% identity after merging OrthoFinder and Broccoli orthogroups.

Reconstructing the evolutionary history for each orthogroup.

Phylogenetic Trees

A phylogenetic analysis was conducted for each orthogroup separately. Sequences from each orthogroup were aligned using MAFFT (--auto) (Katoh et al. 2002; Katoh and Standley 2013) and then trimmed using Trimal (with -gt 0.3 to remove columns with more than 70% gaps) (Capella-Gutiérrez et al. 2009). Resulting multiple sequence alignments were used for phylogenetic tree construction under maximum-likelihood using IQTREE2 (Hoang et al. 2018; Minh et al. 2020) after best-fit model testing (Kalyaanamoorthy et al. 2017).

Identifying clusters of orthologs with Possvm

The resulting gene trees were then further examined with Possvm (Grau-Bové and Sebé-Pedrós 2021), a tool that aids in identifying clusters of orthologs within gene trees facilitating the annotation process which, especially for large trees, can be very time consuming. A further advantage of this method is that it does not require a species tree as input for the ortholog sorting, eliminating potential biases related to disputed species relationships. Possvm was run using default parameters. As a result, each orthogroup corresponding to a broad enzyme family was further subdivided into smaller orthogroups corresponding to specific subfamilies.

Reconstructing evolutionary events with GeneRax

Each gene tree was also reconciled to a species tree using GeneRax (Morel et al. 2020) enabling tree rooting and the discerning of speciation, duplication and loss events characterising each gene tree. The species tree used for reconciliation places sponges as sister-group to all other animals (see above) as this is one of the current accepted scenarios. Moreover, by comparing the reconciled trees with the Possvm-annotated tree, it is possible to control for potential inconsistencies and further investigate if the placement of sponges influenced them. Before running GeneRax, any polytomy in the gene trees were randomly resolved using ETE3 (Huerta-Cepas et al. 2016). GeneRax was

run with the UndatedDL model that accounts for duplication and losses but not horizontal gene transfer events.

Data Availability

Additional supplementary material and raw output files are available at the GitHub repository: [put link](#).

Acknowledgements

For this chapter I would like to thank Riccardo Kyriacou who during his time as a summer intern assisted me in optimising Broccoli parameters for orthogroup detection and then comparing Broccoli orthogroups with those from OrthoFinder using Cytoscape. My thanks also go to Julien Devilliers for his invaluable coding assistance, which facilitated the automation of various steps within this chapter.

References

- Arne JM, Widjaja-Adhi MAK, Hughes T, Huynh KW, Silvaroli JA, Chelstowska S, Moiseenkova-Bell VY, Golczak M. 2017. Allosteric modulation of the substrate specificity of acyl-CoA wax alcohol acyltransferase 2. *Journal of Lipid Research* 58:719–730.
- Arshavsky VY. 2002. Like Night and Day: Rods and Cones Have Different Pigment Regeneration Pathways. *Neuron* 36:1–3.
- Batten ML, Imanishi Y, Maeda T, Tu DC, Moise AR, Bronson D, Possin D, Gelder RNV, Baehr W, Palczewski K. 2004. Lecithin-retinol Acyltransferase Is Essential for Accumulation of All-trans-Retinyl Esters in the Eye and in the Liver *. *Journal of Biological Chemistry* 279:10422–10432.

- Bhatt-Wessel B, Jordan TW, Miller JH, Peng L. 2018. Role of DGAT enzymes in triacylglycerol metabolism. *Archives of Biochemistry and Biophysics* 655:1–11.
- Blaner WS. 2017. Acyl-CoA wax alcohol acyltransferase 2: its regulation and actions in support of color vision1. *Journal of Lipid Research* 58:633–635.
- Blaner WS, Das SR, Gouras P, Flood MT. 1987. Hydrolysis of 11-cis- and all-trans-retinyl palmitate by homogenates of human retinal epithelial cells. *Journal of Biological Chemistry* 262:53–58.
- Blaner WS, Prystowsky JH, Smith JE, Goodman DS. 1984. Rat liver retinyl palmitate hydrolase activity. Relationship to cholestryloleate and triolein hydrolase activities. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* 794:419–427.
- Blomhoff R, Blomhoff HK. 2006. Overview of retinoid metabolism and function. *Journal of Neurobiology* 66:606–630.
- Boussau B, Scornavacca C. 2020. Reconciling Gene trees with Species Trees. In: Scornavacca C, Delsuc F, Galtier N, editors. *Phylogenetics in the Genomic Era*. No commercial publisher | Authors open access book. p. 3.2:1-3.2:23. Available from: <https://hal.science/hal-02535529>
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
- Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Available from: <https://www.biorxiv.org/content/10.1101/2021.06.03.446934v2>
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
- Cheng JB, Russell DW. 2004. Mammalian Wax Biosynthesis: II. EXPRESSION CLONING OF WAX SYNTHASE cDNAs ENCODING A MEMBER OF THE

ACYLTRANSFERASE ENZYME FAMILY *. *Journal of Biological Chemistry* 279:37798–37807.

Corbo JC. 2021. Vitamin A1/A2 chromophore exchange: Its role in spectral tuning and visual plasticity. *Developmental Biology* 475:145–155.

Derelle R, Philippe H, Colbourne JK. 2020. Broccoli: Combining Phylogenetic and Network Analyses for Orthology Assignment. *Molecular Biology and Evolution* 37:3389–3396.

Dewett D, Labaf M, Lam-Kamath K, Zarringhalam K, Rister J. 2021. Vitamin A deficiency affects gene expression in the *Drosophila melanogaster* head. *G3 Genes|Genomes|Genetics* 11:jkab297.

Dewett D, Lam-Kamath K, Poupault C, Khurana H, Rister J. 2021. Mechanisms of vitamin A metabolism and deficiency in the mammalian and fly visual system. *Developmental Biology* 476:68–78.

Duester G. 2000. Families of retinoid dehydrogenases regulating vitamin A function. *European Journal of Biochemistry* 267:4315–4324.

Edenberg HJ. 2007. The Genetics of Alcohol Metabolism: Role of Alcohol Dehydrogenase and Aldehyde Dehydrogenase Variants. *Alcohol Res Health* 30:5–13.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* 16:157.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* 20:238.

Enright JM, Toomey MB, Sato S, Temple SE, Allen JR, Fujiwara R, Kramlinger VM, Nagy LD, Johnson KM, Xiao Y, et al. 2015. Cyp27c1 Red-Shifts the Spectral Sensitivity of Photoreceptors by Converting Vitamin A1 into A2. *Current Biology* 25:3048–3057.

- Feuda R, Dohrmann M, Pett W, Philippe H, Rota-Stabelli O, Lartillot N, Wörheide G, Pisani D. 2017. Improved Modeling of Compositional Heterogeneity Supports Sponges as Sister to All Other Animals. *Current Biology* 27:3864–3870.e4.
- Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150–3152.
- Grau-Bové X, Sebé-Pedrós A. 2021. Orthology Clusters from Gene Trees with Possvm. *Molecular Biology and Evolution* 38:5204–5208.
- Hardie RC, Juusola M. 2015. Phototransduction in *Drosophila*. *Current Opinion in Neurobiology* 34:37–45.
- Harrison EH. 2012. Mechanisms involved in the intestinal absorption of dietary vitamin A and provitamin A carotenoids. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1821:70–77.
- Hirschberg J. 2001. Carotenoid biosynthesis in flowering plants. *Current Opinion in Plant Biology* 4:210–218.
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* 35:518–522.
- Holmes RS. 2012. Vertebrate patatin-like phospholipase domain-containing protein 4 (PNPLA4) genes and proteins: a gene with a role in retinol metabolism. *3 Biotech* 2:277–286.
- Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution* 33:1635–1638.
- Hussain Z, Uyama T, Tsuboi K, Ueda N. 2017. Mammalian enzymes responsible for the biosynthesis of N-acylethanolamines. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1862:1546–1561.
- Jin M, Li S, Moghrabi WN, Sun H, Travis GH. 2005. Rpe65 Is the Retinoid Isomerase in Bovine Retinal Pigment Epithelium. *Cell* 122:449–459.

- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589.
- Kanehisa M. 2019. Toward understanding the origin and evolution of cellular organisms. *Protein Science* 28:1947–1951.
- Kanehisa M, Sato Y, Kawashima M. 2021. KEGG mapping tools for uncovering hidden features in biological data. *Protein Science* [Internet] n/a. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.4172>
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30:3059–3066.
- Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30:772–780.
- Kaylor JJ, Cook JD, Makshanoff J, Bischoff N, Yong J, Travis GH. 2014. Identification of the 11-cis-specific retinyl-ester synthase in retinal Müller cells as multifunctional O-acyltransferase (MFAT). *Proceedings of the National Academy of Sciences* 111:7302–7307.
- Kienesberger PC, Oberer M, Lass A, Zechner R. 2009. Mammalian patatin domain containing proteins: a family with diverse lipolytic activities involved in multiple biological functions. *Journal of Lipid Research* 50:S63–S68.
- Kramlinger VM, Nagy LD, Fujiwara R, Johnson KM, Phan TTN, Xiao Y, Enright JM, Toomey MB, Corbo JC, Guengerich FP. 2016. Human cytochrome P450 27C1 catalyzes 3,4-desaturation of retinoids. *FEBS Letters* 590:1304–1312.
- Kück P, Meusemann K. 2010. FASconCAT, Version 1.0, Zool. Forschungsmuseum A. Koenig, Germany, 2010.
- Lamb TD. 2020. Evolution of the genes mediating phototransduction in rod and cone photoreceptors. *Progress in Retinal and Eye Research* 76:100823.

- Lhor M, Salesse C. 2014. Retinol dehydrogenases: Membrane-bound enzymes for the visual function. *Biochem. Cell Biol.* 92:510–523.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* 13:2178–2189.
- Li W, Jaroszewski L, Godzik A. 2001. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* 17:282–283.
- Maddison W, Maddison D. 2008. Mesquite: A modular system for evolutionary analysis. *Evolution* 62:1103–1118.
- Mata NL, Radu RA, Clemons RS, Travis GH. 2002. Isomerization and Oxidation of Vitamin A in Cone-Dominant Retinas: A Novel Pathway for Visual-Pigment Regeneration in Daylight. *Neuron* 36:69–80.
- Meech R, Hu DG, McKinnon RA, Mubarokah SN, Haines AZ, Nair PC, Rowland A, Mackenzie PI. 2019. The UDP-Glycosyltransferase (UGT) Superfamily: New Members, New Functions, and Novel Paradigms. *Physiological Reviews* 99:1153–1222.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* 37:1530–1534.
- Moise AR, Kuksa V, Imanishi Y, Palczewski K. 2004. Identification of All-trans-Retinol:All-trans-13,14-dihydroretinol Saturase *. *Journal of Biological Chemistry* 279:50230–50242.
- Moiseyev G, Chen Y, Takahashi Y, Wu BX, Ma J. 2005. RPE65 is the isomerohydrolase in the retinoid visual cycle. *Proceedings of the National Academy of Sciences* 102:12413–12418.
- Morel B, Kozlov AM, Stamatakis A, Szöllősi GJ. 2020. GeneRax: A Tool for Species-Tree-Aware Maximum Likelihood-Based Gene Family Tree Inference under

Gene Duplication, Transfer, and Loss. *Molecular Biology and Evolution* 37:2763–2774.

Nelson DR. 2018. Cytochrome P450 diversity in the tree of life. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1866:141–154.

Orland MD, Anwar K, Cromley D, Chu C-H, Chen L, Billheimer JT, Hussain MM, Cheng D. 2005. Acyl coenzyme A dependent retinol esterification by acyl coenzyme A:diacylglycerol acyltransferase 1. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1737:76–82.

Palczewski K, Kiser PD. 2020. Shedding new light on the generation of the visual chromophore. *Proc Natl Acad Sci U S A* 117:19629–19638.

Pingitore P, Romeo S. 2019. The role of PNPLA3 in health and disease. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1864:900–906.

Poux S, Arighi CN, Magrane M, Bateman A, Wei C-H, Lu Z, Boutet E, Bye-A-Jee H, Famiglietti ML, Roechert B, et al. 2017. On expert curation and scalability: UniProtKB/Swiss-Prot as a case study. *Bioinformatics* 33:3454–3460.

Redmond TM, Poliakov E, Yu S, Tsai J-Y, Lu Z, Gentleman S. 2005. Mutation of key residues of RPE65 abolishes its enzymatic role as isomerohydrolase in the visual cycle. *Proceedings of the National Academy of Sciences* 102:13658–13663.

Rowland A, Miners JO, Mackenzie PI. 2013. The UDP-glucuronosyltransferases: Their role in drug metabolism and detoxification. *The International Journal of Biochemistry & Cell Biology* 45:1121–1132.

Ruiz A, Winston A, Lim Y-H, Gilbert BA, Rando RR, Bok D. 1999. Molecular and Biochemical Characterization of Lecithin Retinol Acyltransferase *. *Journal of Biological Chemistry* 274:3834–3841.

Sahu B, Maeda A. 2016. Retinol Dehydrogenases Regulate Vitamin A Metabolism for Visual Function. *Nutrients* 8:746.

- Schreiber R, Taschler U, Preiss-Landl K, Wongsiriroj N, Zimmermann R, Lass A. 2012. Retinyl ester hydrolases and their roles in vitamin A homeostasis. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1821:113–123.
- Schultz DT, Haddock SHD, Bredeson JV, Green RE, Simakov O, Rokhsar DS. 2023. Ancient gene linkages support ctenophores as sister to other animals. *Nature*:1–8.
- Seña C dela, Riedl KM, Narayanasamy S, Curley RW, Schwartz SJ, Harrison EH. 2014. The Human Enzyme That Converts Dietary Provitamin A Carotenoids to Vitamin A Is a Dioxygenase *. *Journal of Biological Chemistry* 289:13661–13666.
- Seo M, Koiwai H, Akaba S, Komano T, Oritani T, Kamiya Y, Koshiba T. 2000. Abscisic aldehyde oxidase in leaves of *Arabidopsis thaliana*. *The Plant Journal* 23:481–488.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13:2498–2504.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
- Strauss O. 2005. The Retinal Pigment Epithelium in Visual Function. *Physiological Reviews* 85:845–881.
- Terao M, Romão MJ, Leimkühler S, Bolis M, Fratelli M, Coelho C, Santos-Silva T, Garattini E. 2016. Structure and function of mammalian aldehyde oxidases. *Arch Toxicol* 90:753–780.
- Thompson DA, Gal A. 2003. Vitamin A metabolism in the retinal pigment epithelium: genes, mutations, and diseases. *Progress in Retinal and Eye Research* 22:683–703.

Trifiletti RR. 2014. Vitamin A. In: Aminoff MJ, Daroff RB, editors. Encyclopedia of the Neurological Sciences (Second Edition). Oxford: Academic Press. p. 717–718.

Available from:

<https://www.sciencedirect.com/science/article/pii/B9780123851574001172>

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol Biol Evol* 35:543–548.

Webb EC. 1992. Enzyme nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes. *Enzyme nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes*. [Internet]. Available from: <https://www.cabdirect.org/cabdirect/abstract/19930457289>

Widjaja-Adhi MAK, Golczak M. 2020. The molecular aspects of absorption and metabolism of carotenoids and retinoids in vertebrates. *Biochim Biophys Acta Mol Cell Biol Lipids* 1865:158571.

Wong E, Anggono V, Williams SR, Degnan SM, Degnan BM. 2022. Phototransduction in a marine sponge provides insights into the origin of animal vision. *iScience* 25:104436.

Zhao M, Ma J, Li M, Zhang Y, Jiang B, Zhao X, Huai C, Shen L, Zhang N, He L, et al. 2021. Cytochrome P450 Enzymes and Drug Metabolism in Humans. *International Journal of Molecular Sciences* 22:12808.