Genome analysis reveals unexpected absence of oxygen metabolic capacity in helminths

by

Emma Collington

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Biology

Waterloo, Ontario, Canada, 2022

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

I personally performed all analyses described in this document, except for the set-up of the proteome library and development of the phylogenetic tree. Nooran Mazen assisted in the collection and organization of the proteome library, and with running Pfam Scan on the files. Dr. Andrew Doxey wrote the post-processing script allowing us to convert Pfam files to lists of architectures. Briallen Lobb wrote the code to determine the top 20 most conserved architectures across our species of interest, and assisted me with the alignments used to develop the phylogenetic tree used to display data.

Abstract

Oxygen metabolism in parasitic helminths differs significantly from that of most aerobic eukaryotes, as these organisms must be able to switch between aerobic and anaerobic metabolisms depending on their life-cycle stage. These parasites also require robust antioxidant defense systems that allow them to survive bursts of reactive oxygen species (ROS) released by their hosts. Early metabolism studies suggested a lack of cytochrome *c* oxidase (COX) activity in certain parasitic helminths, and the role of COX in helminth mitochondria remains unclear. To determine whether a functional COX is widely present in helminths, we analyzed the phylogenetic distribution of oxygen metabolism systems across 129 helminth genomes, investigating three distinct sets of protein-coding genes involved in different aspects of oxygen metabolism: COX and its assembly factors, peroxisomal genes, and genes coding for the most abundant ROS-metabolizing proteins. While glycolytic and citric acid cycle enzymes are highly conserved in helminthic species, in tracking the presence of genes associated with the electron transport chain we observed that certain lineages of parasitic worm have lost genes associated with COX and its assembly factors. Although most common genes encoding proteins involved in the defense against ROS are maintained across virtually all lineages, many species exhibit complete absence of peroxisomal metabolic pathways, including an absence of catalase. Our results suggest that a subset of parasitic nematodes and platyhelminths utilize oxygen differently from related, non-parasitic worms such as *C. elegans*, with significant differences in their mitochondrial electron transport chains and peroxisomes. The identification of substantive metabolic differences between the parasitic helminths and their mammalian hosts offers a new avenue for the development of anthelminthic therapeutics.

Acknowledgements

I would like to thank my supervisors, Dr. Moira Glerum and Dr. Andrew Doxey for their support, and for allowing and encouraging me to pursue a really ‘out there’ topic! Thank you for pushing me to take extra computer science courses, I’ve learned many new transferrable skills that will assist me in my future research endeavors.

To my committee members, thank you for your enthusiasm in my project and for asking (not just difficult) but also interesting questions, I’ve really enjoyed our discussions during committee meetings.

To the Doxey and Glerum lab groups, thank you for being a sounding board for my ideas, and for being a wonderful audience for practice presentations. Special thanks to Briallen Lobb for her assistance in assembling the phylogenetic tree, and for generously checking over my code whenever I ran into trouble. Thank you to Nooran Mazen for all her help assembling the proteome library and for running Pfam scan on all those endless files with me!

Thank you to Lucy Collington for testing out different graphics programs and teaching me how to use Google Drawings to produce the figures in the introduction, you saved me a lot of time and frustration!

To the ‘worm people’: Dr. Tim Geary and Dr. Andrew Peregrine, thank you for your interest and enthusiasm, and for taking time to meet with us to discuss the quirky aspects of worm metabolism. Thank you for your introductions to other researchers in the field, and for helping me realize that this is a topic worth pursuing further!

A heartfelt thankyou to my family for their love and support, and for allowing me to dedicate my meager allowance of spoons to completing my research! I would not have been able to do any of this without you. Thanks are also in order for my service dog Marlowe, who sat with his head in my lap through every committee meeting and presentation, and who has been by my side (literally) every step of the way.

Finally, thank you to Dr. Danescu for keeping me alive and functional enough to complete this project!

Dedication

In loving memory of Danielle Tench, dance among the stars my friend.

Also, you owe me a hot chocolate.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Acetyl CoA	Acetyl coenzyme A
ATP	Adenosine triphosphate
CO_2	Carbon dioxide
Coa	Cytochrome *c* oxidase assembly factor protein
COX	Cytochrome *c* oxidase, Complex IV
Cox	Cytochrome *c* oxidase protein
Cu	Copper
ETC	Electron transport chain
$FADH_2$	Flavin adenine dinucleotide
GTP	Guanosine-5'-triphosphate
H_2	Hydrogen gas
H_2O	Water
iTOL	Interactive Tree of Life
KEGG	Kyoto Encyclopedia of Genes and Genomes
MUSCLE	MUltiple Sequence Comparison by Log-Expectation
NAD^+	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide
NCBI	National Center for Biotechnology Information
Nduf	NADH:ubiquinoneoxidoreductase subunits
O_2	Oxygen
PEP	Phosphoenolpyruvate
Pex	Peroxin protein
ROS	Reactive oxygen species
Sdh	Succinate dehydrogenase
tBLASTn	Translated nucleotide basic local alignment search tool
TCA	Tricarboxylic acid cycle (citric acid cycle)

Chapter 1: Introduction

Estimated to infect 25% of the human population (see Table 1 for global prevalence), parasitic helminths cost global economies billions of dollars through human disability and through their impact on livestock and crop industries[1]. Helminths have been present over the course of human evolution: helminth eggs have been detected in mummified human feces dating back thousands of years; and human type 2 immunity is suspected to have evolved specifically to deal with helminthic pathogens[2]. Because the distribution of helminths is highly dependent on climate and topography, developing countries, especially Sub-Saharan Africa, carry the highest disease burden[2]. Pre-school and school-aged children are most susceptible to these infections, and because infections can result in chronic illnesses that persist long after the elimination of the infective agent, helminth infections have poverty promoting effects[2]. Although many anthelminthic agents exist and there have been global efforts to implement mass drug administration strategies[2], many of the current nematicides used are toxic to the host and environmentally damaging, and overuse of antiparasitic agents to treat livestock infections has resulted in the development of widespread resistance to these medications[1]. Identifying key differences between parasite and host metabolism is therefore of utmost importance for the development of new therapeutic targets.

Table 1: Major helminth-related diseases in humans

The most common human diseases caused by helminths, the causative organism for each condition, global prevalence, and distribution of infections. Despite the high infection rates by this type of organism, these are considered neglected tropical diseases. Adapted from Hotez *et. al*., 2008[2].

Diseases	Causative agent	Number of cases worldwide
Nematodes		
Ascariasis	*Ascaris lumbricoides,*	807 million
Trichuriasis	*Trichuris trichiura*	604 million
Hookworm	*Necator americanus, Ancylostoma duodenale*	576 million
Strongyloidiasis	*Strongyloides stercoralis*	30 – 100 million
Liver filariasis	*Wuchereria bancrofti, Brugia malayi*	120 million
Onchocerciasis	*Onchocerca volvulus*	37 million
Loiasis	*Loa loa*	13 million
Dracunculiasis	*Dracunculus medinensis*	0.01 million
Platyhelminths		
Schistosomiasis	*Schistosoma haematobium, Schistosoma mansoni, Schistosoma japonicum*	207 million
Trematodiasis	*Clonorchis sinensis, Opisthorchis viverinni, Paragonimus spp., Fasciolopsis buski, Fasciola hepatica*	>40 million
Cysticercosis	*Taenia solium*	0.4 million

1.1 Aerobic eukaryotic metabolism

Cells require energy in the form of ATP to perform their biological functions. In non-photosynthetic eukaryotic organisms, this energy is produced through the breakdown of complex nutrients into amino acids, sugars, fatty acids, and glycerol[3]. The free energy released from the gradual oxidation of these molecules, particularly the breakdown of glucose by aerobic respiration, is used to synthesize energy (ATP) and potential energy (NADH) for use in downstream anabolic pathways[4] (see Figure 1).

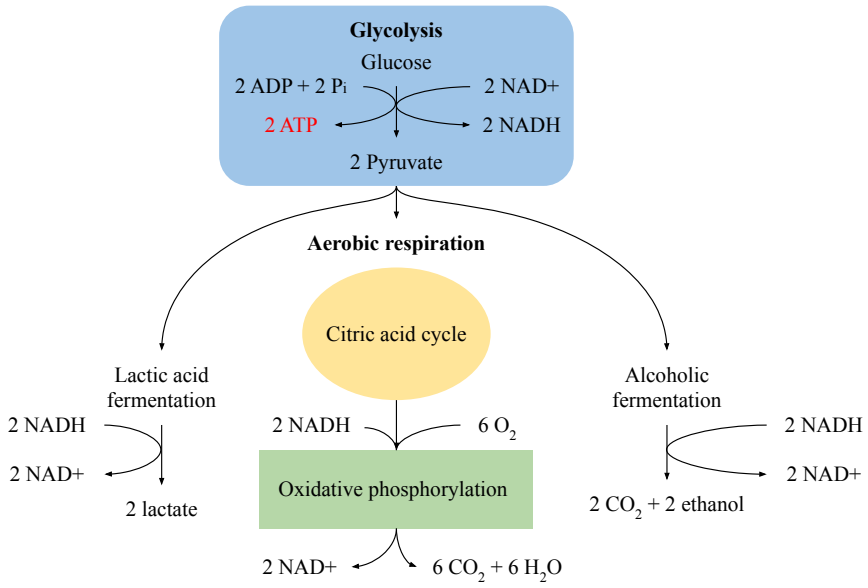


Figure 1: Glucose metabolism

Glycolysis converts one molecule of glucose into two molecules of pyruvate. Under aerobic conditions, pyruvate is then oxidized to water and carbon dioxide through the citric acid cycle and oxidative phosphorylation. Under anaerobic conditions, pyruvate can undergo fermentation to produce either lactate or ethanol. Adapted from Voet & Voet, 2010[4].

The process of converting the glucose into energy begins with glycolysis, a cytosolic process which produces ATP under both aerobic and anaerobic conditions[3]. One molecule of glucose is broken down through a stepwise series of ten enzymatic reactions, producing different sugar intermediates that are eventually converted into two molecules of pyruvate[3]. This process produces two molecules each of ATP and NADH.

After the production of pyruvate molecules by glycolysis, glucose metabolism splits into two pathways depending on the availability of environmental oxygen. Glycolysis is the main source of ATP for many anaerobic organisms, with the pyruvate being fermented into different end products depending on the organism (usually lactate or ethanol) in a process that uses the NADH produced in glycolysis to regenerate the NAD^+ needed for the next glycolytic reaction cycle[3].

Under aerobic conditions, pyruvate is oxidized by the pyruvate dehydrogenase complex to produce acetyl CoA and carbon dioxide[4]. Within the mitochondria, acetyl CoA enters the tricarboxylic acid cycle (TCA cycle, also known as the citric acid cycle or the Krebs cycle), where it combines with oxaloacetate to form tricarboxylic acid[3]. The TCA cycle generates high energy electrons in the form of NADH, and regenerates oxaloacetate which allows the cycle to continue[3]. The NADH produced is kept within the mitochondria where it enters the electron transport chain (ETC), which produces the majority of cellular ATP in aerobic organisms[3].

The final stage of aerobic respiration occurs in the mitochondrial ETC, which is made up of five complexes that ultimately transfer electrons to molecular oxygen to form water and

produce cellular energy in the form of ATP[4]. In order for the transfer of electrons to occur between complexes of the chain (see Figure 2), each complex must have a higher affinity for electrons than the previous one, allowing electrons to be transferred between complexes successively until they reach oxygen, which has the highest affinity for electrons[3]. This process results in the pumping of protons across the inner mitochondrial membrane, which produces an electrochemical gradient[3]. The transport of electrons between ETC complexes relies on the use of cofactors (cytochromes and quinones) that contain transition metals such as iron and copper, which exist in different oxidation states, allowing for electrons to pass between them[4].

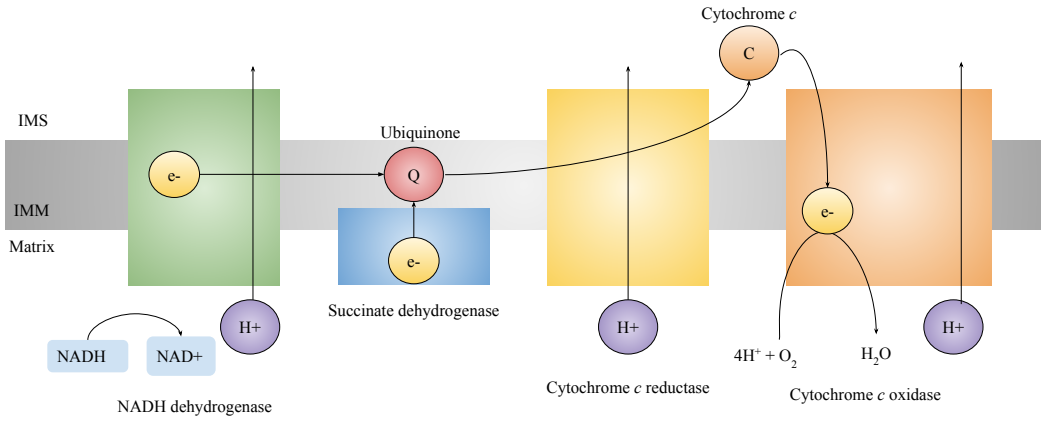


Figure 2: The mitochondrial electron transport chain

Complex I (NADH dehydrogenase) in the inner mitochondrial membrane (IMM) accepts electrons from NADH and passes them to ubiquinone (Q), reducing it to ubiquinol. This then transfers electrons to Complex III (cytochrome *c* reductase), which passes electrons one at a time to another cofactor, cytochrome *c* in the mitochondrial intermembrane space (IMS). From here electrons are moved to the heme and copper centers of Complex IV (cytochrome *c* oxidase), where they are finally transferred to molecular oxygen. Each of these complexes uses the energy released by the transfer of electrons to a higher affinity molecule to pump protons across the inner membrane, producing an electrochemical gradient. Adapted from Alberts *et al.*, 2015[3].

The overall process of cellular respiration produces ATP, water, and carbon dioxide in the reaction shown below, with approximately thirty molecules of ATP being produced from a single molecule of glucose[3]. Additional byproducts from the TCA cycle and ETC include NADH, $FADH_2$, and GTP[3].

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O + ATP$$

1.2 Anaerobic eukaryotic metabolism

Although many features of eukaryotic intermediary metabolism are strongly conserved from the unicellular yeast, *Saccharomyces cerevisiae*, through mammals, several deviations have been identified in recent years, many of which center around the differences between aerobic and anaerobic metabolism. Five subclassifications of mitochondrial organelles have been recently described (see Figure 3): four of these are ATP-producing, including standard aerobic mitochondria, which utilize an ETC and oxygen as a final electron acceptor; mitochondria which do not utilize oxygen as a final electron acceptor and function anaerobically; H_2 producing mitochondria, which still contain an ETC but do not utilize oxygen; and hydrogenosomes, which lack an ETC entirely[5]. An additional class has also been identified, the non-ATP producing mitosomes[5].

Peroxisomes have also been determined to have anaerobic and aerobic organellar variants: anaerobic peroxisomes, which have lost all oxygen utilizing pathways, have been described in eukaryotes with reduced mitochondria, for example, in the protist *Mastigamoeba balamuthi*[6]. Despite the existence of these alternate mitochondrial-derived and peroxisomal organelles, certain standard means of oxygen consumption are present in

most eukaryotic organisms, and an estimated >90% of cellular oxygen is consumed by cytochrome *c* oxidase (COX, Complex IV), of the mitochondrial respiratory chain[7].

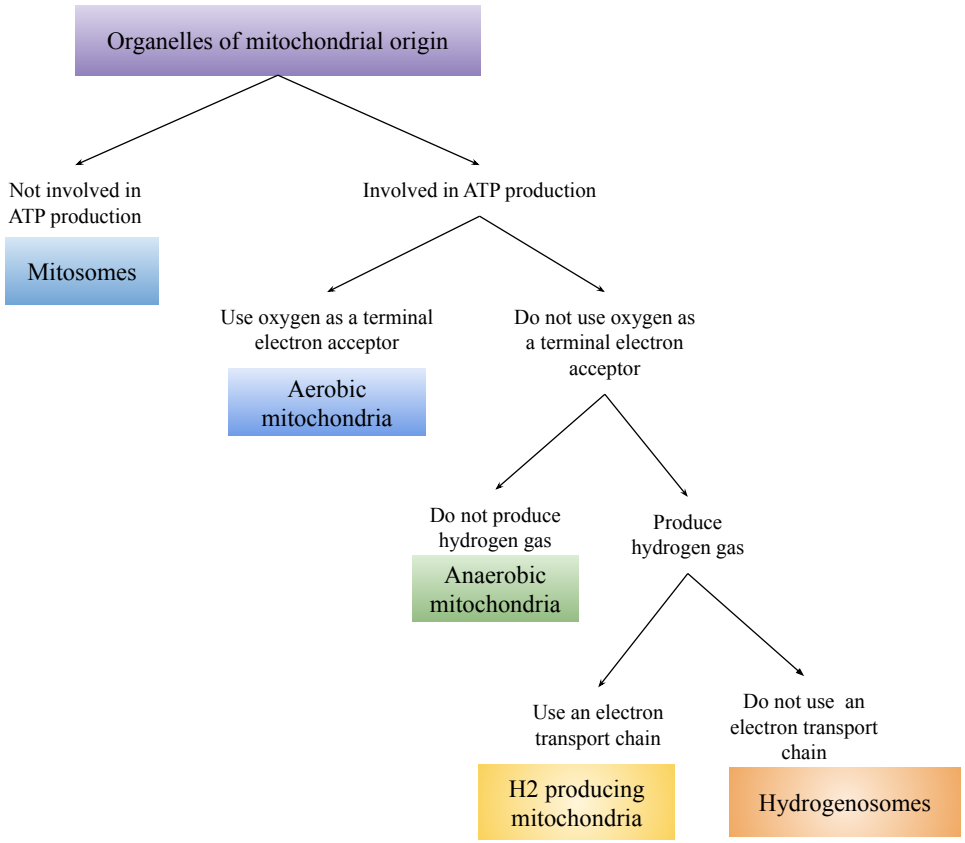


Figure 3: Organelles of mitochondrial origin

This chart outlines the criteria for each of the different classifications of organelles of mitochondrial origin. The three overarching categories of these organelles are mitosomes, which do not produce ATP; aerobic mitochondria, which generate ATP and use oxygen as a final electron acceptor; and mitochondria which generate ATP but do not utilize oxygen as the terminal electron acceptor. The latter group includes anaerobic mitochondria which use different terminal electron acceptors, H_2 producing mitochondria which use an electron transport chain, and hydrogenosomes which have lost the electron transport chain complexes altogether. Adapted from Muller *et. al.*, 2012[5].

1.3 Parasitic Helminths

Parasitic helminths are found in a wide variety of environments and infect all types of multicellular organisms[8]. Because there are so many different types of helminths, it is difficult to make generalizations about them, however they tend to have at least two distinct life-cycle stages: one within a host organism, where reproduction occurs; and a brief free-living stage after the eggs have exited the host[8]. Reinfection occurs when a host accidentally ingests either eggs or larva, although some species use an additional arthropod vector with infection occurring through the ingestion of an infected host[8].

Helminths fall into two different phyla: nematodes (roundworms), which are divided into five distinct clades of which four are parasitic to humans; and platyhelminths (flatworms), including the monogeneans, trematodes, and cestodes[1]. The exclusively parasitic cestodes, the simplest of the platyhelminths, have lost all internal digestive systems: their outer tegument layer has evolved to perform the functions of intestinal tissue, including nutrient absorption and digestion[9]. The parasitic trematodes and ectoparasitic monogeneans, both have an outer syncytial tegument layer in addition to an internal body cavity with a single opening, the gastrodermis, which is involved in digestion[9]. In contrast, nematodes can be either parasitic or free-living, and have a fully developed gut with openings at each end of the organism[9].

Recent bioinformatic investigations into helminths have uncovered a large amount of novel gene content, including gene families with no known functional annotations, and gene families with significant expansions[1]. Parasitic worms also have significant variation in

genome size and gene number, and their extended phylogenic branches are the locations of rapid molecular evolution[8]. All parasitic worms are descended from free living-ancestors, and many of the different clades within the parasitic worm phyla contain both free-living and parasitic members[8]. Parasitism evolved independently within each of these phyla on up to eighteen separate occasions, although there are many basic adaptations to parasitic lifestyle that are shared between species that infect different hosts in different manners[8].

1.4 Chromatin diminution in helminths

Although recent bioinformatics-based studies have revealed that although helminths are undergoing evolution and acquiring novel gene families at an accelerated rate[1], some species of helminth additionally perform the very rare and not well understood process of chromatin diminution[10]. In most eukaryotic cells, the genomic content of germline and somatic cells is identical, and remains so throughout the life cycle of the organism; however, there are at least 100 known eukaryotic species across nine major taxonomic groups that alter the DNA content of certain lineages of cell[10]. DNA can be lost through chromatin diminution, or via the complete elimination of chromosomes, although the latter is very rare[10]. Chromatin diminution was first discovered in the parasitic helminth *Parascaris univalens* by Theodor Boveri in 1887[11], and is known to occur in at least ten other species of parasitic helminth[12] (see Table 2 for a complete list).

Table 2: Chromatin diminution in nematodes

A list of nematode species known to undergo chromatin diminution, and the embryonic cleavage division at which the diminution event occurs. Most of these species belong to either the family *Ascarididae* or *Toxocaridae*. Although *Strongyloides papillosus* does undergo chromatin diminution, its process differs from the other species listed here and results in the development of free-living males. It is currently unknown whether additional species also undergo this process; however, *Caenorhabditis elegans* and *Panagrellus redivivus* have been confirmed to not perform diminution. Adapted from Tobler *et. al.*, 1992[12].

Species	Number of cleavage divisions at which chromatin diminution occurs
Parascaris univalens	(2 or 3) to 5
Ascaris lumbricoides	3 to 5
Ascaris suum	3 to 5
Opbidascaris filaria	(2 or 3) to 5
Contracaecum incurvum	3 to 5
Toxocara canis	2 to 6
Toxocara cati	(2 or 3) to 6
Toxocara vulpis	3 to 6
Cosmocerca sp.	3 to 8
Physaloptera Indiana	1 to 3
Strongyloides papillosus	During mitotic parthenogenesis

In *P. univalens* and other species of helminth, chromatin diminution occurs during early embryogenesis in presomatic cells (see Figure 4) and changes the genomic content and organization of somatic cells relative to germ-line cells[13]. Diminution happens in regions of chromatin termed CBRs (chromatin breakage regions), which are target sites for yet-to-be-determined elimination factors that cause the chromatin to be broken up into smaller fragments[14]. During subsequent rounds of cell division, certain regions of fragmented

chromatin fail to migrate to the spindle poles and are instead degraded in the cytoplasm[13]. *P. univalens* is estimated to eliminate 80-90% of its nuclear germline DNA, however this amount varies depending on helminth species[13].

All heterochromatin is eliminated during the chromatin diminution process, as well as germline-specific chromatin containing highly repetitive satellite DNA sequences (although the exact function of satellite DNA in eukaryotic cells is currently unknown, its loss during chromatin diminution suggests a germline-restricted function)[13]. Some single copy genes have also been reported to be eliminated as well, including *aleg-3*, *fert-1*, and *alep-1*[13]. Diminution could occur in helminths as a silencing mechanism of germline specific sequences in somatic cells, or it could produce changes in chromatin structure resulting in positional effects of other genes[13]. It has additionally been suggested that diminution may be a way for these organisms to deal with a partial genome duplication event that occurred in a common helminth ancestor[13].

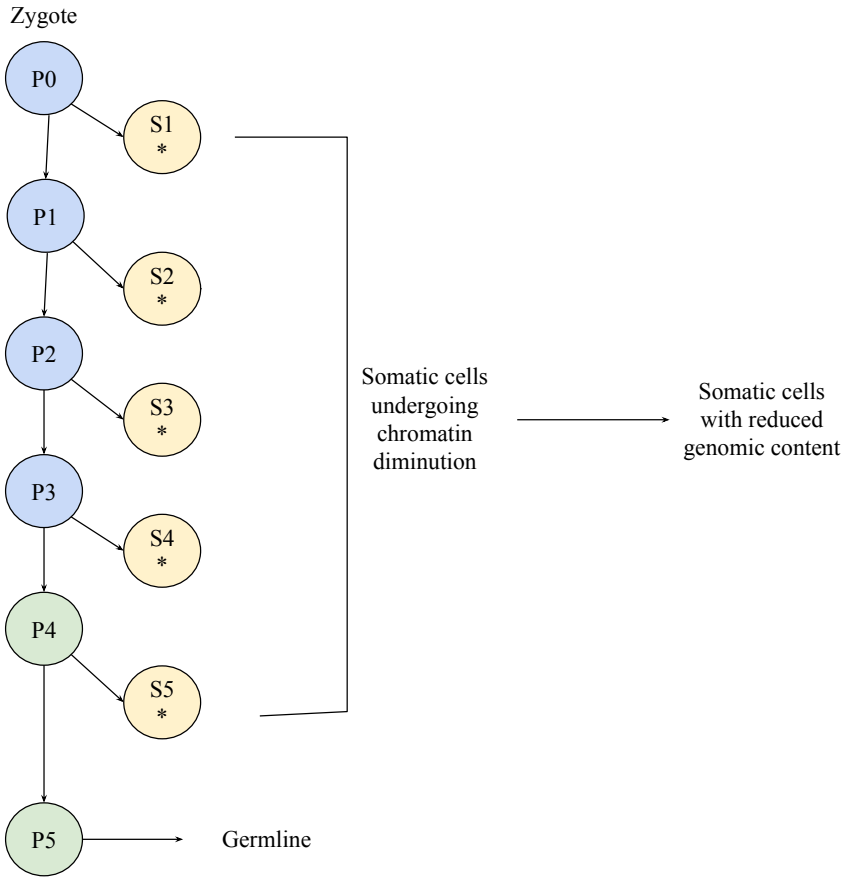


Figure 4: Chromatin diminution in *Parascaris univalens*

During *P. univalens* early embryonic development, the presumptive primordial germ cells (designated P0 – P3 and shown as blue circles) give rise to presomatic cells (yellow circles) and germline cells (green circles). Chromatin diminution occurs in presomatic cells S_1-S_5, indicated by the asterisk (*), and all cells descended from these lineages will inherit the reduced genomic content. A similar process occurs during the development of *A. suum,* but in this species diminution occurs in presomatic cells S_2-S_4, and S_{1a} and S_{1b}. Adapted from Muller & Tobler, 2000[13].

1.5 Anthelminthic agents

The treatment of infections by helminth parasites differs from methods used to treat infections by other types of microorganisms. For infections of bacterial or viral origin, therapeutic agents target mechanisms of multiplication and cause the infective organism to die off. The target of anthelminthic agents, however, is the adult worm, which does not rely on multiplication for its survival[15]. Therapeutic avenues instead target parasite locomotion and nerve transmission, which cause the worm to detach from its infection site to be expelled by the host[16]. There are six classes of anthelminthic agents[17], with the three most commonly used mechanisms including beta-tubulin binding, which disrupts helminth cytoskeleton formation and nutrient uptake; and spastic and flaccid paralytic agents, which prevent neurotransmission within the helminth[16]. Other less commonly used anthelminthic agents target unique aspects of helminth metabolism[15].

1.6 Helminth mitochondrial metabolism & oxygen requirements

Parasitic helminths have high ATP requirements, as energy is necessary to establish and maintain an infection within a host, as well as to produce enough eggs to ensure that transmission occurs[18]. Early biochemical studies on helminth metabolism revealed that, although oxygen use is a key feature of mitochondrial metabolism, anthelminthic agents such as cyanine dyes (which inhibit oxygen uptake) have no impact on either the survival or rates of glycolysis in certain species of filarial parasities[19]. Further investigations on this subject were incongruous with these findings: it was determined that this type of anthelminthic agent

does still have a broad spectrum of activity against intestinal nematodes that utilize anaerobic metabolism[15]. From this it was determined that different species of helminth have different oxygen requirements, for example, *Nippostrongylus brasiliensi* has aerobic requirements, as do *Ascaris* larva. Conversely, *Ascaris* adults function anaerobically, and for another species, *Hymenolepis diminuta*, oxygen is inhibitory[15]. All species of helminth appear to be capable of utilizing oxygen to some degree if the conditions are appropriate, however these species are mostly unable to fully oxidize substrates to carbon dioxide and water, with terminal respiration either absent or rate limiting in these organisms[15]. Helminth species range from having strict aerobic requirements to being able to function completely anaerobically, with oxygen even being inhibitory to certain species[15]. Other species are suspected of having an oxygen requirement for biosynthesis and development, but not for the generation of energy in the adult organism[15].

Inside the helminth mitochondria, oxygen has two very different roles: in aerobic metabolism, oxygen is a necessary metabolite for the ETC and is involved in the production of cellular energy in the form of ATP; but incomplete reduction leads to the formation of reactive oxygen species (ROS) that can result in oxidative damage within the mitochondria and diffuse into the rest of the cell[20]. In addition to mechanisms for dealing with ROS generated from their own metabolism and released by their hosts, parasitic helminths must also be able to adapt to environments with varying oxygen availability throughout their different life-cycle stages, resulting in several unique adaptations in their oxygen metabolizing machinery. Many parasitic helminths can therefore use aerobic metabolism, but

can also undergo a metabolic shift to anaerobic metabolism for parasitic life-cycle stages where there is limited or no oxygen available[21]. In general, adult parasites do not use aerobic oxidation of carbohydrates and do not use oxygen as a final electron acceptor in their ETCs, regardless of oxygen availability[21]. Immediately after a change from an oxygen-rich to oxygen-depleted environment, carbohydrates are initially degraded to lactate and ethanol through anaerobic glycolysis[21]. Given that helminths are incapable of fully oxidizing substrates to carbon dioxide and water, fermentation products are also produced in addition to lactate and ethanol, including acetate, succinate, propionate, other volatile fatty acids, acetoin, pyruvate, and glycerol[15].

Long term exposure to anaerobic environments results in changes that lead to the degradation of carbohydrates to phosphoenolpyruvate (PEP), which is the point at which helminth metabolism deviates from that of their hosts and other eukaryotes[15]. PEP is gradually reduced to malate which gets transported to the mitochondria, where the pathway splits into two branches: some malate is oxidized to acetate, and the rest is reduced to succinate or propionate and then excreted[21] (see Figure 5). Some types of helminths (including species of *Ascaris*, see Figure 6) have additional metabolic steps wherein acetyl CoA and propionyl CoA get further reduced to branched chain fatty acids in a process resembling a reversal of beta oxidation[5]. It is not yet fully clear why *Ascaris* performs these additional reactions when the formation of propionate and acetate should be sufficient to meet energy requirements, but it has been suggested that the fatty acids produced are transported to reproductive tissues and incorporated into forming eggs; the fatty acids could

then be used to generate energy aerobically during development[18]. If this turns out to be the case, helminths are capable of an interesting cycle where anaerobic products are synthesized and then later used aerobically[18]. This could also account for the formation of the variety of fermentative end products that accumulate in adult worms regardless of oxygen availability[18].

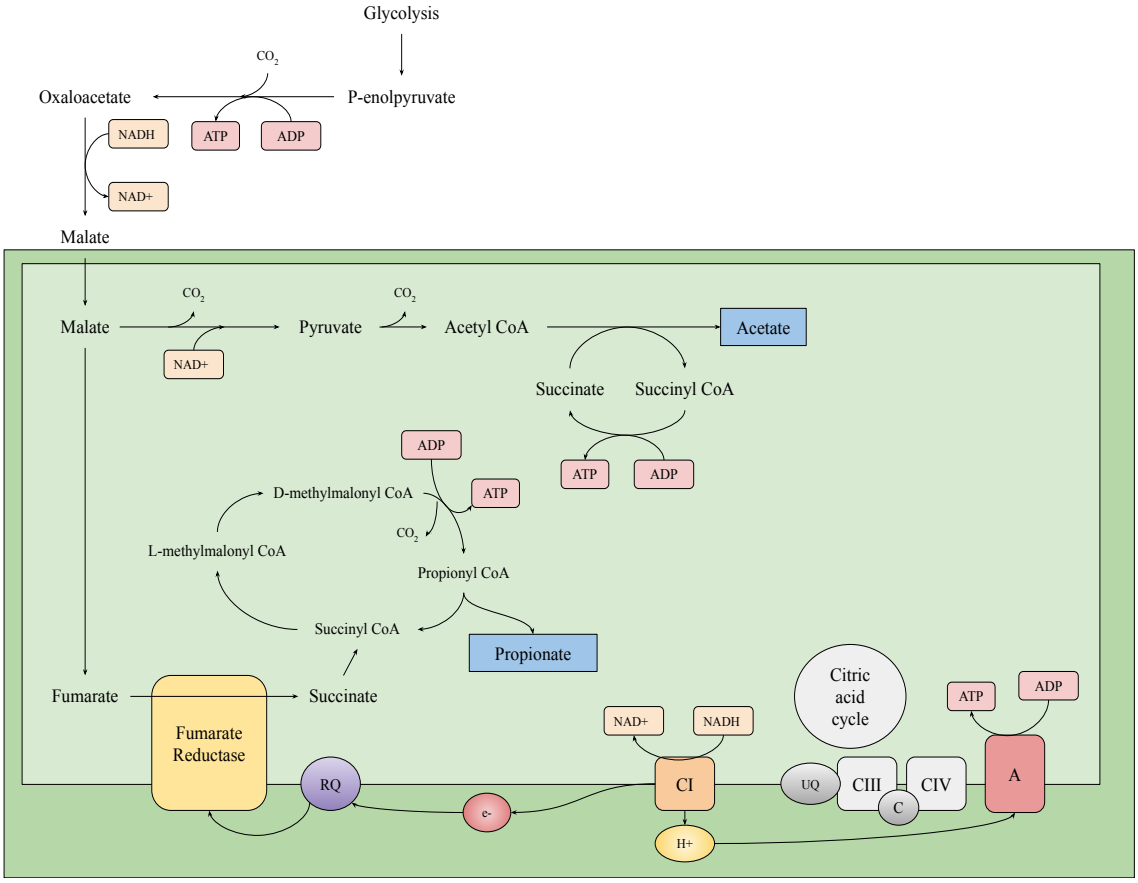


Figure 5: Anaerobic metabolism of the parasitic helminth *Fasciola hepatica*

The anaerobic mitochondrial pathways of *F. hepatica*, showing malate dismutation within the mitochondrial matrix (light green) to two end products (propionate and acetate, shown in blue boxes). The citric acid cycle and ETC complexes after NADH dehydrogenase have been bypassed and are shown in a grey and no longer play a role in ATP production. Complex II has been replaced by a fumarate reductase complex, and ubiquinone has been replaced with rhodoquinone. Adapted from Müller *et. al.*, 2012[5].

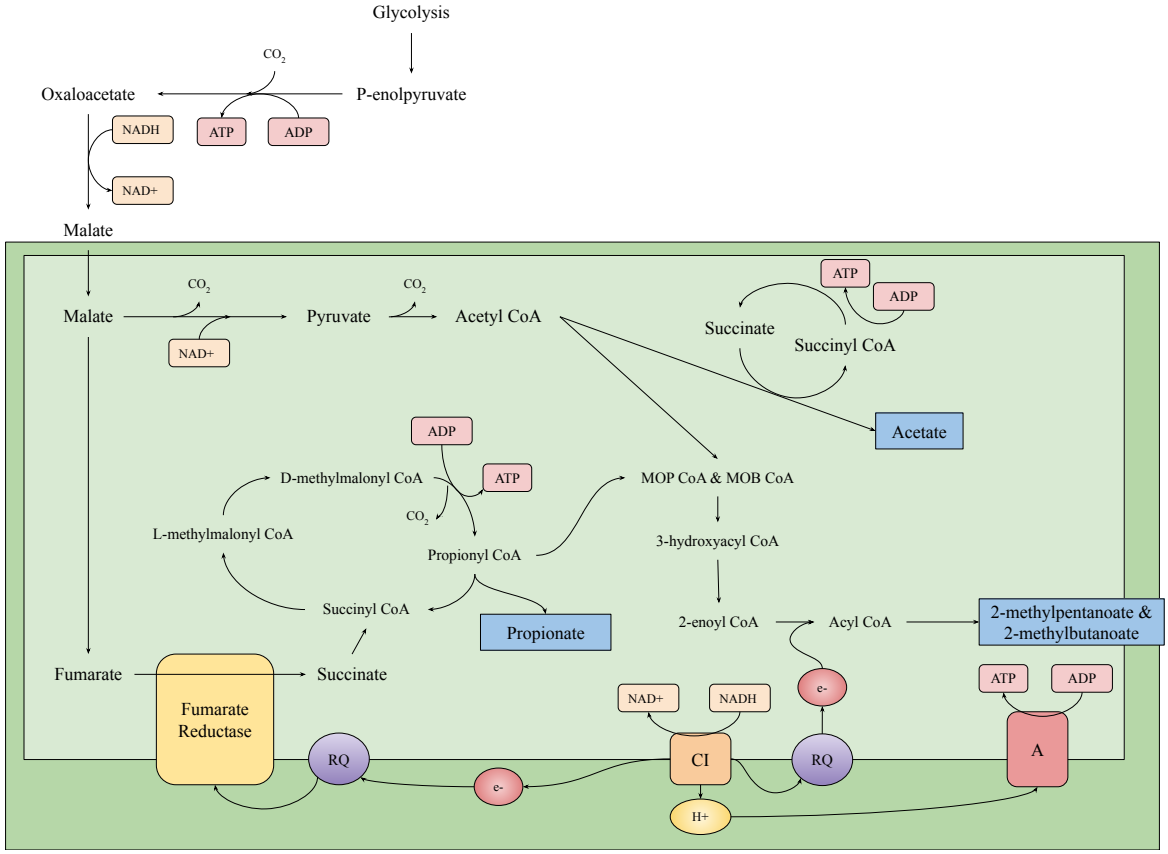


Figure 6: Anaerobic metabolism of the parasitic helminth *Ascaris lumbricoides*

The anaerobic metabolic pathways of *A. lumbricoides*, which differ from those of *F. hepatica*. *A. lumbricoides* also performs malate dismutation, and uses rhodoquinone for fumarate reduction, however this species additionally produces two branched-chain fatty acids from propionyl CoA and acetyl CoA using electrons provided from Complex I via rhodoquinone. All end products are shown in boxes. Electron transport chain Complexes III and IV are no longer in use and are not shown on this diagram, nor is the citric acid cycle. Adapted from Muller *et. al.*, 2012[5].

Along with these metabolic adaptations, parasitic helminths utilize their ETCs differently from other aerobic eukaryotes as they can reverse succinate dehydrogenase (Complex II) and reduce fumarate to succinate[22]. The genes controlling the directional switch

for this reaction are differentially expressed throughout development, where adult helminths express rhodoquinone instead of ubiquinone for transporting electrons to Complex II[21].

Parasitic helminths have been proposed to have differing cytochrome systems from their hosts: even under aerobic conditions, their oxygen uptake is lower than would be expected and early studies of *Ascaris lumbricoides* and *Schistosoma mansoni* showed little or no detectable COX activity[22]. The biogenesis of COX is dependent on the assistance of a set of highly conserved assembly factors that ensure that the nuclear and mitochondrially encoded subunits, together with the requisite prosthetic groups (heme A and copper), are assembled into a functional holoenzyme[23]. The absence of any of these assembly factors, whether providing copper or heme A, or chaperoning the mitochondrially-encoded subunits, results in the degradation of the enzyme[24].

1.7 Essential cytochrome *c* oxidase assembly factors

Although little research has been conducted into COX assembly factors in helminths, three assembly factors - Cox11p, Cox17p, and Sco1p[25–27]- that are involved in providing copper to the active centers of Cox1p and Cox2p are present in *Caenorhabditis elegans*[28]. These assembly factors are believed to function in a bucket brigade manner, passing copper ions between them (see Figure 7).

Cox17 is a cysteine-rich protein that is highly conserved in all eukaryotes and was discovered by Glerum et al. in 1996[27]. It was initially proposed to be responsible for

delivering copper ions to the mitochondria, however subsequent research demonstrated that the protein remained functional when tethered to the inner mitochondrial membrane, indicating that its function is localized to the intermembrane space[29]. Once copper enters the mitochondrial intermembrane space, it gets picked up by Cox17, which then passes it to Sco1 and Cox11, both of which are inner mitochondrial membrane-anchored proteins[30].

Sco1 goes on to deliver copper ions to the Cu_A site of Cox2[31], while Cox11 delivers its copper to the Cu_B site on Cox1[30]. The incorporation of copper into COX subunits is a vital stage to its assembly, and if the copper ions do not insert correctly, the Cox1 and Cox2 subunits are rapidly degraded [24].

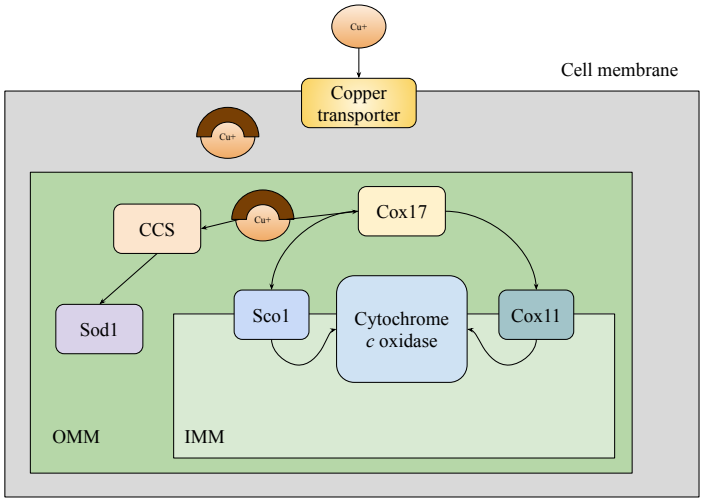


Figure 7: Copper trafficking in the mitochondria

Reduced copper enters the cell via specific or non-specific copper transporters embedded in the plasma membrane, where it is picked up by an unknown transporter (brown). Copper passes through the outer mitochondrial membrane (OMM) to the intermembrane space, where it gets picked up by CCS or Cox17. CCS delivers copper to Sod1, while Cox17 delivers it to either Sco1 or Cox11, which load copper ions into subunits of the assembling COX holoenzyme in the mitochondrial inner membrane (IMM). Adapted from Baker *et. al.*, 2017[32].

1.8 Peroxisomes and oxygen

Aside from mitochondria, peroxisomes are the other major oxygen-consuming organelles, using both oxygen and hydrogen peroxide for oxidation reactions[3]. Peroxisomes and mitochondria are recognized to have many similarities, both in terms of metabolic function and in their ability to adapt in number and morphology to cell conditions[33]. From an evolutionary standpoint, one hypothesis on the origin of this organelle is that peroxisomes may have performed all oxygen metabolism in ancient eukaryotic ancestors, and were later rendered largely obsolete by the development of the mitochondrion, which shares many of the same biological pathways[3]. Peroxisomes are primarily responsible for scavenging hydrogen peroxide through the activity of catalase. This enzyme performs two important roles: it is involved in the breakdown of hydrogen peroxide into water and oxygen (first reaction); catalase can additionally act in the oxidation of electron donors in the breakdown of peroxides (second reaction)[9].

$$2H_2O_2 \rightarrow 2H_2O + O_2$$

$$ROOH + AH_2 \rightarrow H_2O + ROH + A$$

In addition to this critical antioxidant role, peroxisomes are involved in several other biological pathways. Peroxisomes contain enzymes that perform beta oxidation reactions, gradually breaking down blocks of fatty acid molecules to form acetyl CoA, which is then exported to the cytosol for biosynthetic reactions[3]. Beta oxidation is one of the processes shared between the mitochondrion and the peroxisome[3], however certain types of fatty acids are handled by each of these organelles (for example, some very-long chain fatty acids can

only be oxidized by the peroxisome)[34]. In the case of fatty acids that are unable to undergo beta oxidation, peroxisomes also have enzymes for alpha oxidation[34].

The peroxisome performs other, non-oxygen dependent functions, including etherphospholipid biosynthesis, which requires an enzyme exclusively located in the peroxisome[34]. It additionally houses other pathways, including amino acid, retinol, glutathione, and purine metabolisms[35].

1.9 Reactive oxygen species and ROS management in helminths

In addition to the essential organelles involved in oxygen consumption, cells also contain a conserved set of proteins that are involved in the management of ROS. The final transfer of electrons to molecular oxygen by COX must be carefully controlled, as oxygen can only accept electrons one at a time[36]. After receiving its first electron, oxygen forms a superoxide radical anion ($O_2^{\cdot -}$) which is highly reactive and will take up additional electrons from anywhere it can get them[36]. To avoid this, cells carefully control the rate of transfer of electrons to oxygen, with COX clamping oxygen molecules between its metal centers until the oxygen has picked up a total of four electrons, however, some of the superoxide ions escape this clamp[36]. Other scenarios and conditions can also occur within the mitochondria where production of ROS is unavoidable[36]. For example, if no ADP is available to be converted into ATP, then a build-up of the hydrogen gradient across the inner mitochondrial membrane results in a more reducing environment that slows the travel of electron flow through the chain and favours the formation of superoxide ions[36]. In other situations where

an inhibitor of one of the ETC complexes is introduced, all upstream complexes become reduced as electrons cannot be transferred to the next complex, which additionally favours ROS production[36]. Electrons can also leak prematurely out of ETC Complexes I and III, which are able to reduce molecular oxygen to the superoxide anion[20].

The rate of oxygen consumption can also impact ROS production, as an increased rate of oxygen metabolism results in the increased production of ROS, which the mitochondria can only cope with to a certain extent[36]. Because mitochondria contain nitric oxide synthases, the presence of superoxide anions can easily result in the production of the even more oxidizing reactive nitrogen species, peroxynitrite, so it is in the best interest of mitochondria to rid themselves of superoxide ions[36]. This is especially relevant to helminths during larval stages, when high levels of superoxide dismutase expression have been observed, likely to combat increased ROS production[37].

The management of ROS in parasitic helminths is of particular importance to the survival of these organisms: like other eukaryotes, they must be able to contend with ROS produced as a biproduct of their own oxygen metabolisms, but they are also exposed to additional ROS in the form of hydrogen peroxide released by their host as part of the immune response to parasitic infection[21] (See Figure 8).

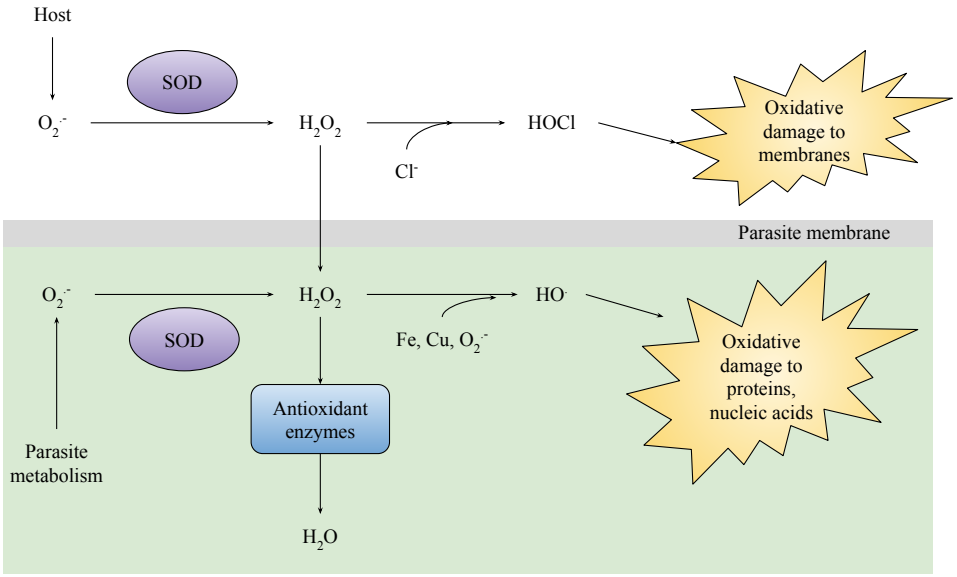


Figure 8: Reactive oxygen species generation in parasitic helminths

Parasitic helminths must be able to survive not only their own ROS production, but exposure to additional ROS released by their host as a primary immune response to infection. Infected hosts release superoxide anions ($O_2^{\cdot-}$), which are converted to hydrogen peroxide (H_2O_2) by host superoxide dismutase proteins, and then into hypochlorous acid (HOCl) by host myeloperoxidase enzymes (MPO). The hydrogen peroxide can also diffuse across parasite membranes (grey) where it joins the pool of ROS produced by the host itself. Some of the hydrogen peroxide in the parasite gets detoxified and converted into water by antioxidant enzymes, while the rest can potentially react with other superoxide anions, iron, or copper to form a hydroxyl radical ($HO^{\cdot}$). The extracellular hypochlorous acid and intracellular hydroxyl radicals can cause oxidative damage to parasite membranes, proteins, and nucleic acids. Adapted from Guevara-Flores *et. al.*, 2017[38].

Some of the important proteins involved in the management of ROS include superoxide dismutases, glutathione peroxidases, thiolases, peroxidases, thioredoxins, and glutaredoxins. Superoxide dismutases exist in different forms with various metal cofactors (copper-zinc, manganese, iron)[39], and scavenge superoxide radicals, catalyzing their dismutation into molecular oxygen and hydrogen peroxide in the reaction shown below[9]:

$$2H^+ + 2O_2^- \rightarrow H_2O_2 + O_2$$

The hydrogen peroxide produced by superoxide dismutation is then dealt with by other antioxidant proteins such as catalase. In addition to their intracellular superoxide dismutase proteins, helminths have been noted to have a secreted version of the protein, likely to combat ROS produced by their hosts[9].

Peroxidases are a broad class of enzymes that reduce peroxides, using a variety of electron donors[40]. Glutathione peroxidase, for example, catalyzes the reaction of reduced glutathione (GSH) and a non-specific hydroperoxide into oxidized glutathione disulfide (GSSG) and water (see first reaction below). The reduced glutathione can then be regenerated in a reaction of the disulfide form and NADPH in the second reaction shown below[9].

$$ROOH + 2GSH \rightarrow ROH + GSSG + H_2O$$

$$GSSG + NADPH + H^+ \rightarrow 2GSH + NAD^+$$

Other types of peroxidases rely on redoxin proteins as their electron donors[41]. Helminths have both thioredoxins and glutaredoxins and have been documented to have a unique system that links these two redoxins in a multidomain architecture, wherein the thioredoxin and glutaredoxin domains are able to function separately or in concert[42].

In this work, we identify previously undescribed alterations to oxygen metabolic capacity in the genomes of 129 helminth species, including the widespread absence of genes encoding key subunits of cytochrome *c* oxidase and related assembly factors and the apparent

absence of peroxisomes in many parasitic helminths. This large-scale metabolic restructuring of helminths could provide a method by which future anthelminthic agents could distinguish between the host and parasite, allowing for the development of new therapeutic approaches.

Chapter 2: Methods

2.1 Construction of architecture database and phylogenetic tree

We utilized genomic data and gene predictions for 129 species of nematode and platyhelminth that are available on WormBase ParaSite[43], including sequence data from the 50 Helminth Genomes Project[1] and other previously published helminth genomes[44–103]. These included 102 parasitic species and 27 free-living species (for a full list of species, see Appendix for Supplementary Table 1). In cases where multiple genomes were available for a single species of helminth, the one with more predicted protein-coding genes was selected for further analysis based on the assumption that the larger genome would be more complete. Complete BUSCO values were obtained directly from WormBase ParaSite to assess genome completeness, with an average complete BUSCO score across our 129 species of 83% (see Appendix for Supplementary Figure 2).

Pfam Scan[104] (with default settings in version 1.6, against the Pfam database version 35.0) was applied to each of the predicted proteomes to functionally annotate each protein sequence by domain architecture. A post-processing script was applied to the Pfam files to convert them into an easily searchable list of architectures.

A phylogenetic tree was constructed using abundant single copy proteins, defined based on unique Pfam domain architectures identified within our library of predicted proteins. We identified the twenty most abundant single-copy protein architectures each for our free-living and parasitic species (see Table 3 for architectures used). We then generated a concatenated multiple sequence alignment of these 39 architectures in all our species of

interest using MUSCLE[104] (version 5, with default settings), although not every species had all 39 architectures present, and for species for which the architecture was present in multiple copies, the multi-copy architecture was excluded from the alignment. The aligned sequences were trimmed with GBlocks[105] (version 0.91b), and a phylogenetic tree with 100 bootstrap replicates was constructed using FastME[106] (version 2.0). We then used this tree to display protein presence/absence data using iTOL[107].

Table 3: Conserved protein architectures

Architectures used to construct alignment for phylogenetic tree. The 20 most frequently expressed single copy conserved architectures from parasitic and free-living species of helminth were used to build our phylogenetic tree. Architectures shown in italics were common to the top 20 lists from both the parasitic and free-living species.

Most frequent single copy architectures in parasitic species	Most frequent single copy architectures in free-living species
Ccdc124	*DUF2615*
Secretogranin_V	Ribosomal_L30_N,Ribosomal_L30
NDUF_B7	SRPRB
MRP-L51	PRCC
RNA_pol_Rpc34	BCL_N
Mago_nashi	TBCA
RNA_pol_Rpb8	F-actin_cap_A
COX5A	Endosulfine
Peptidase_C65	ATP-synt_D
Mt_ATP-synt_D	Ribosomal_S7e
UFC1	Fer4_7
APG5	GCN5L1
CybS	Sec62

G10	MRP-L28
DUF2615	Spt4
Ribosomal_L6,Ribosomal_L6	SF3a60_bindingd,SF3A3,Telomere_Sde2_2,DUF3449
Complex1_30kDa	Ssu72
PTPS	Rab3-GTPase_cat
RNA_pol_Rpb6	Fer4_20,Pyr_redox_2,DHO_dh,Fer4_21
AAA_18	CRAL_TRIO,Motile_Sperm

2.1 Tracking the presence of proteins of interest

Model sequences for proteins of interest were identified based on known proteins from *C. elegans* from the Universal Protein Knowledgebase[28]. We utilized three separate methods to assess whether a target protein was present within each of our species. First, we performed tBLASTn searches (with default settings in May and June of 2021) in the genomes of our species of interest directly in the WormBase ParaSite[43] DNA database using an *E*-value threshold of 1e-3. Species for which there were 'hits' at this threshold were considered to have a gene encoding the sequence of interest, while we performed further analysis for species with no hits. For these, we searched the WormBase Parasite entries directly for orthologs to genes encoding the *C. elegans* proteins of interest. We also searched for our proteins of interest based on Pfam domain architectures. All the proteins of interest for this project could be represented by a single conserved Pfam domain, however, given that previous research in helminth proteomics has uncovered the presence of many unusual protein architectures, especially those containing multiple, unrelated domains 'fused' into single architectures[1], we chose to additionally examine the domain architecture library we

constructed to see if our proteins of interest were occurring in combination with other conserved protein families. We used the NCBI Conserved Domains search tool[108] (version 3.18, searching CDD v.3. 19 58235 PSSMs) to identify conserved protein families within the *C. elegans* target protein. We then searched for architectures in our protein library that matched that of the *C. elegans* protein. In many species we identified the sequence of interest only in multidomain architectures with additional conserved domains that are not present in the target *C. elegans* protein.

Our proteins of interest to this project included enzymes involved in glycolysis, the citric acid cycle (see Table 4 for abbreviations used), and the five complexes of the electron transport chain; peroxisomal proteins; and proteins involved in defense against ROS.

We used the KEGG peroxisomal pathway[35] and the Universal Protein Knowledgebase[28] to identify peroxisomal proteins in *C. elegans* and produced a presence/absence profile in iTOL for visualization. For a full list of peroxisomal proteins identified, along with their associated metabolic pathways within the peroxisome and the abbreviations used to refer to them, see Table 5. Please note that the abbreviation *prx* is used in many eukaryotes for genes associated with peroxiredoxin, however in *C. elegans* it is used to denote peroxin protein-coding genes.

As a definitive marker for the presence of peroxisomal organelles, we looked specifically at four peroxin protein-coding genes Pex3, Pex10, Pex12, and Pex19, which encode a core set of proteins that are conserved in all eukaryotic lineages[109]. We selected these as they cover two peroxin functions essential to the biogenesis of the organelle itself:

Pex3 and Pex19 are necessary for the sorting of peroxisomal membrane proteins; and Pex10 and Pex12 function in receptor recycling and ubiquitination[109].

Table 4: Glycolytic and citric acid cycle proteins

Sequences from major glycolytic and citric acid cycle enzymes from *C. elegans* were used to search for genes encoding these proteins in our helminth species of interest. Abbreviations are shown where applicable.

Pathway	Enzyme	Abbreviation
Glycolysis	Hexokinase	HK
	Phosphoglucose isomerase	PGI
	Phosphofructokinase	PFK
	Fructose bisphosphate aldolase	Aldolase
	Triosephosphate isomerase	TPI
	Glyceraldehyde phosphate dehydrogenase	GAPDH
	Phosphoglycerate kinase	PGK
	Phosphoglycerate mutase	PGM
	Enolase	
	Pyruvate kinase	PK
Citric acid cycle	Citrate synthase	CS
	Aconitase	
	Isocitrate dehydrogenase	IDH
	Succinyl-CoA synthetase	SCS
	Alpha ketoglutarate	AKG
	Fumarase	
	Malate dehydrogenase	MDH

Table 5: Peroxisomal proteins in *C. elegans*

Proteins involved in different peroxisomal pathways from *C. elegans* were identified and their sequences used to determine the presence of genes encoding these proteins in other species of helminth.

Pathway	Protein	Abbreviation
Peroxins	Pex3	*prx-3*
	Pex10	*prx-10*
	Pex12	*prx-12*
	Pex19	*prx-19*
Unsaturated fatty acid beta oxidation	Delta3 5-delta2 4-dienoyl-coa	ECH
	ATP-binding cassette, subfamily	ABCD
	2,4-dienoyl-CoA reductase	PDCR
	Solute carrier family 27 member	VLACS
Beta oxidation	Acyl-CoA oxidase	ACOX
	Alpha-methylacyl-CoA racemase	AMACR
	Hydroxymethylglutaryl-CoA lyase	HMGCL
	Sterol carrier protein	SPCX
Alpha oxidation	Phytanoyl-CoA hydroxylase	PHYH
	2-hydroxyacyl-CoA lyase	HPCL2
Etherphospholipid biosynthesis	Alkyldihydroxyacetonephosphate synthase	AGPS
	Alcohol-forming fatty acyl-CoA reductase	FAR
	Glyceronephosphate O-acyltransferase	GNPAT
Amino acid metabolism	D-aspartate oxidase	DDO
	Alanine-glyoxylate transaminase	AGXT
	(S)-2-hydroxy-acid oxidase	HAO
	Isocitrate dehydrogenase	IDH
	N1-acetylpolyamine oxidase	PAOX

	Sarcosine oxidase	PIPOX
Reactive oxygen species metabolism	Catalase	CAT
	Protein Mpv17	MPV17
	Peroxisomal membrane protein 2	PXMP4
Other oxidation	Carnitine O-acetyltransferase	CRAT
	Carnitine O-octanoyltransferase	CROT
	Malonyl-CoA decarboxylase	MLYCD
	NAD+ diphosphatase	NUDT12
	Nucleoside diphosphate-linked moiety X motif	NUDT19
Retinol metabolism	Dehydrogenase/reductase SDR family member 4	DHRS4
Organelle fission	Peroxisome fission protein	FIS1
Glutathione metabolism	Glutathione S-transferase kappa 1	GSTK1
Sterol precursor biosynthesis	Mevalonate kinase	MVK
	Phosphomevalonate kinase	PMVK
Purine metabolism	Xanthine dehydrogenase	XDH

Chapter 3: Results

3.1 Construction of phylogenetic tree and architecture database

The inability to detect COX activity in parasitic helminths was initially reported by Bueding and Charms in 1952[22]. The recent availability of helminth genomes might allow us to readdress the issue of COX usage by helminths, and to investigate at a genomic level those initial observations made decades ago[22]. To assess the patterns of presence and absence of protein-coding genes of interest, we began by constructing a phylogenetic tree (see Figure 9) of our 129 helminth species, including members of the two major classifications of helminth worms: platyhelminths (monogeneans, cestodes, and trematodes); and nematodes (clades I, III, IV, and V). No genomic sequence data is currently available for the free-living clade II nematodes, and thus they are entirely absent from this tree.

Generating this tree then allowed us to examine the presence and absence of protein-coding genes of interest according to their lineage and evolutionary relationships when visualized using iTOL. Species in our tree cluster correctly according to their clade of nematode or class of platyhelminth in agreement with previous publications, particularly the tree generated by the International Helminth Genomes Consortium in their 2019 paper[1].

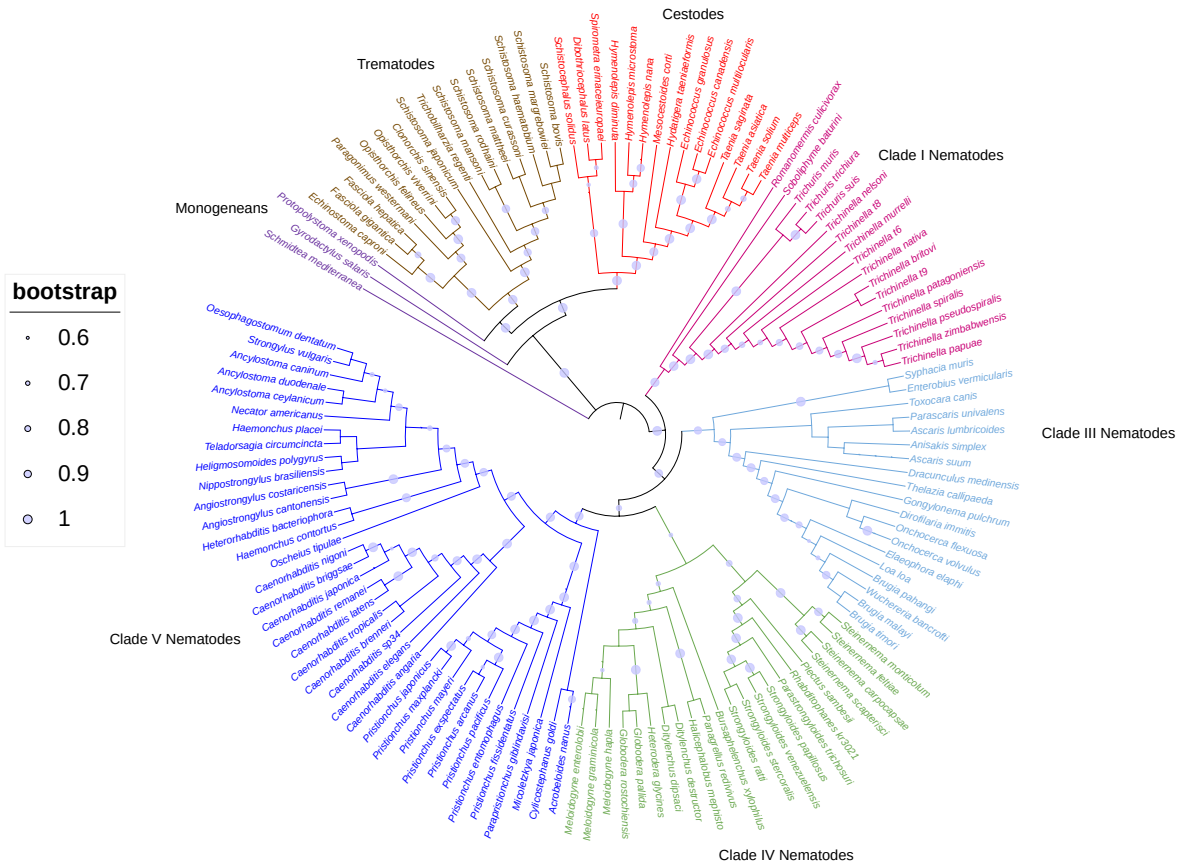


Figure 9: Phylogeny of 129 species of helminth

This tree is based on an alignment of the 20 most frequently expressed single copy protein architectures identified for the free-living and parasitic species. One of these architectures was common to the top 20 lists for both sets of species, and the final, trimmed alignment was based on a total of 39 protein sequences, amounting to approximately 3200 amino acid positions. Bootstrap support values between 60 (0.6) and 100 (1) are indicated on branches for a total of 100 replicates. Major classifications of helminth and their associated branches are colour-coded based on whether they are platyhelminths (monogeneans, trematodes and cestodes), or nematodes (clades I, III through V). See Appendix Supplementary Figure 1 for branch lengths.

3.2 Tracking the presence/absence of mitochondrial aerobic metabolism

As part of the Glerum Lab's ongoing work on understanding COX assembly in human health and disease, we first asked whether COX and its assembly factors were present in our species of interest. Because COX is the aerobic endpoint of oxidative metabolism, we started our investigation by looking at the overall process of cellular respiration. We checked each of the 129 species of nematode and platyhelminth for genes encoding major proteins involved in glycolysis, the citric acid cycle, and core subunits essential for each of the five complexes in the mitochondrial electron transport chain, using the three methods described above and the results illustrated in Figure 10.

With few exceptions (mainly phosphoglycerate mutase in platyhelminths), all the major glycolytic and citric acid cycle protein-coding genes were present across all our species of interest, as were the markers chosen to indicate the presence of electron transport chain members Complex I (*nduf-2.2*), Complex II (*sdh-1*), and Complex III (cytochrome C1). However, we noticed an apparent decrease in the number of species that were positive for *cox*-1, the subunit we chose to use as a marker for the presence of cytochrome *c* oxidase, and ATP6, the marker used for ATP synthase. Given the multi-subunit nature of COX, and the fact that *cox-1* is a mitochondrially encoded subunit which could account for its absence from the genomes investigated, we chose to perform an in-depth analysis of protein-coding genes associated with this complex.

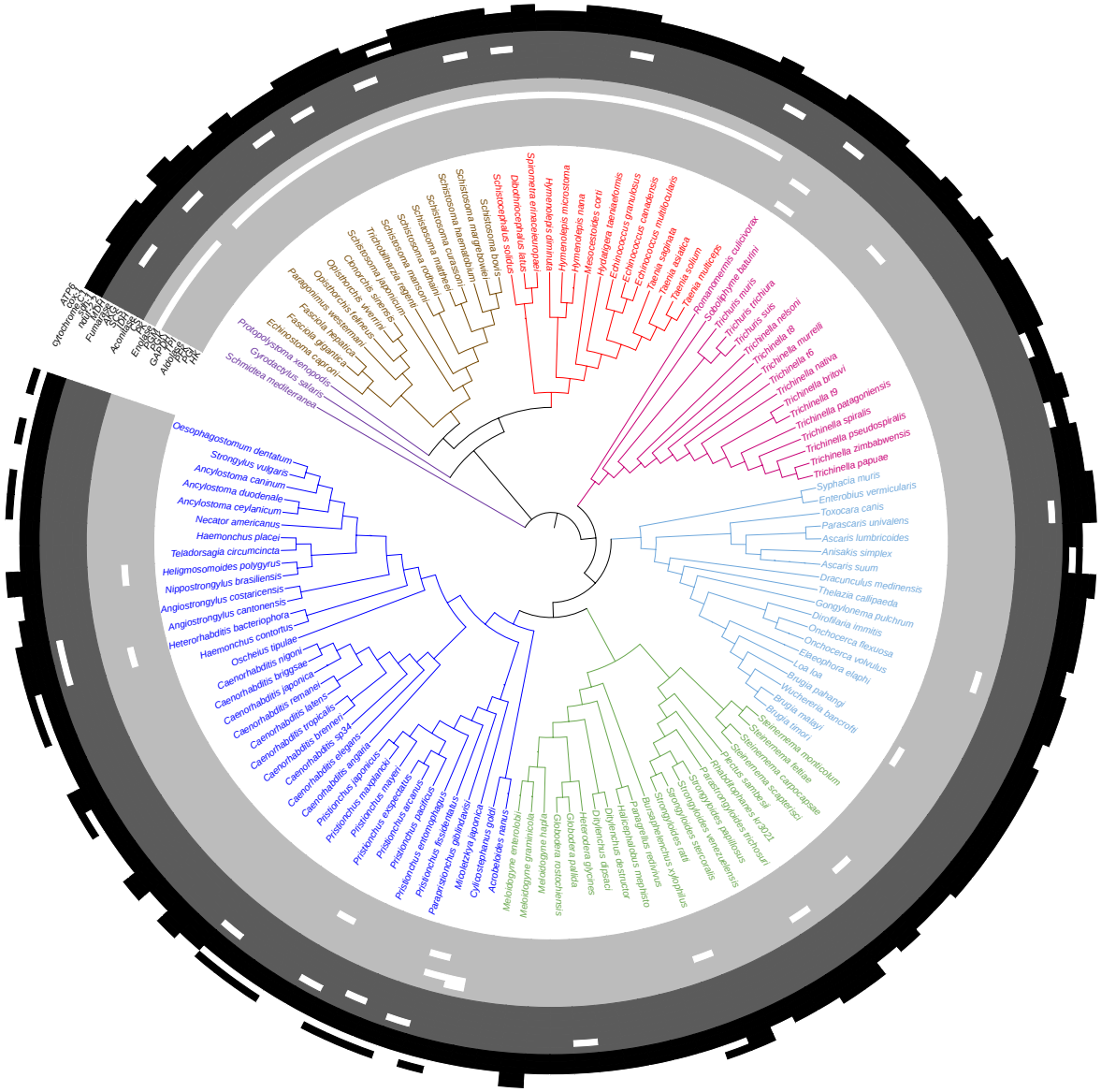


Figure 10: Tracking the presence of genes encoding proteins involved in glycolysis, the TCA cycle, and the electron transport in 129 species of helminth

Protein-coding genes include, from inside to outside: glycolysis in light grey, HK, PFK, aldolase, GAPDH, PGK, PGM, enolase, and PK; TCA cycle in medium gray, CS, aconitase, IDH, SCS, AKG, fumarase, and MDH; and the ETC in black, *nduf-2.2* (Complex I marker), *sdh-2* (Complex II marker), cytochrome C1 (Complex III marker), *cox-1* (Complex IV marker), and ATP6 (ATP synthase marker).

To further assess the presence of COX in our species, we identified the core set of COX proteins that are present in *C. elegans* using the Universal Protein Knowledgebase[28]. *C. elegans* COX has three catalytic subunits[110] encoded by mitochondrial genes *cox-1*, *cox-2*, and *cox-3*; and subunits encoded by the nuclear genes *cox-4*, *cox-5a*, *cox5b*, *cox-6a*, *cox-6b*, *cox-6c*, and *cox-7c.* We additionally identified the presence of a set of assembly factors: *coa1, coa3-7*, *cox11*, *cox14-19*, and *sco1* (see Figure 11). Although *cox1-3* encode for the catalytic subunits and are therefore essential to COX function, we were unable to verify whether the genomic data available from WormBase Parasite includes the full mitochondrial genomes because the sequences could be confounded by the presence of pseudogenes of mitochondrial origin in the nuclear genome, and because of the technical difficulties associated with sequencing helminth mitochondrial DNA (insufficient material, lack of high quality reference sequences, small mitochondrial genome size, and the AT-rich nature of helminth mitochondrial DNA)[111]. Instead, we considered the presence of the three nuclear-encoded COX assembly factors Cox17, Cox11, and Sco1, which are essential for the provision of copper for the assembly of the catalytic core and thereby indicative of a functional COX holoenzyme (see Figure 12).

The core subunits are absent in many species from different lineages of the tree; however, there are more frequent absences in the platyhelminth lineages and clade I nematodes relative to the other three nematode clades. The only species found to have all the COX subunits and associated assembly factors was *C. elegans*, with other members of the *Caenorhabditis* genus and other clade V nematodes having more hits for these protein-coding

genes than the other lineages of the tree. To avoid the possibility that other species of helminth possess assembly factors that are not present in *C. elegans*, we ran additional searches in our proteome library and in WormBase Parasite using the keywords ‘cox’ and ‘cytochrome oxidase’, which would allow us to identify any additional subunits, however we did not find any besides the original set we identified, which should rule out the possibility of known assembly factors that are not present in *C. elegans*.

As with the core mitochondrial subunits, the platyhelminths and clade I nematodes have fewer hits for the assembly factors, and most species of these lineages are lacking *cox-11*and *cox-17*, without which the COX complex cannot assemble[24]. The absence of these protein-coding genes thus predicts a complete absence of COX, which corroborates the initial biochemical results from the mid-20th century.

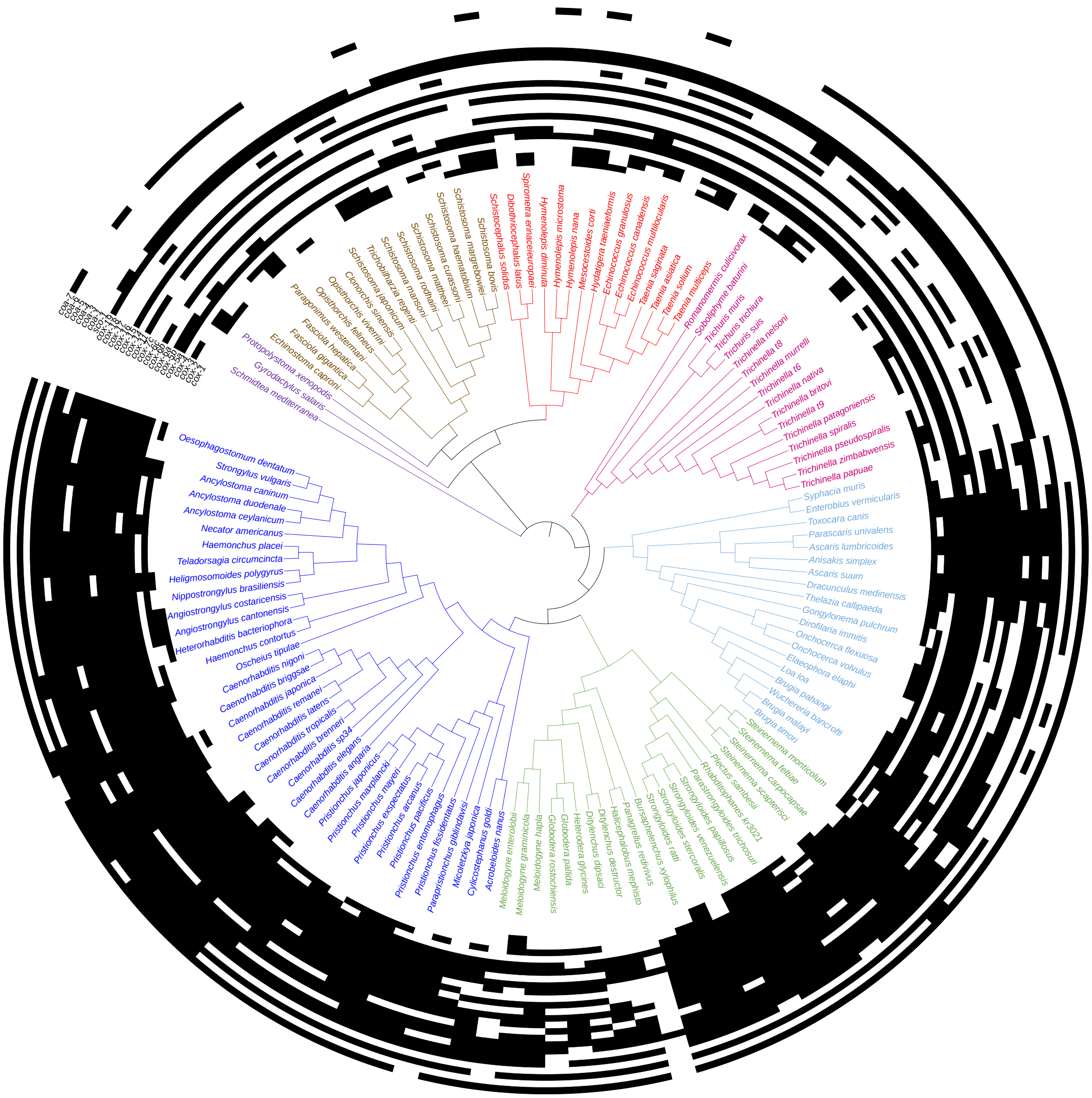


Figure 11: The presence of major cytochrome *c* oxidase protein-coding genes in 129 species of helminth.

After identifying COX subunits and assembly factors present in *C. elegans*, we used these sequences to search for associated genes in our species of interest. Protein-coding genes used include (from inside to outside): mitochondrial subunits *cox-1*, *cox-2*, and *cox-3*; nuclear encoded subunits *cox-4*, *cox-5a*, *cox5b*, *cox-6a*, *cox-6b*, *cox-6c*, and *cox-7c*; and assembly factors *cox-11*, *cox-14*, *cox-15*, *cox-16*, *cox-17*, *cox-18*, *cox-19*, *sco1, coa-1*, *coa-3*, *coa-4*, *coa-5*, *coa-6*, and *coa-7*.

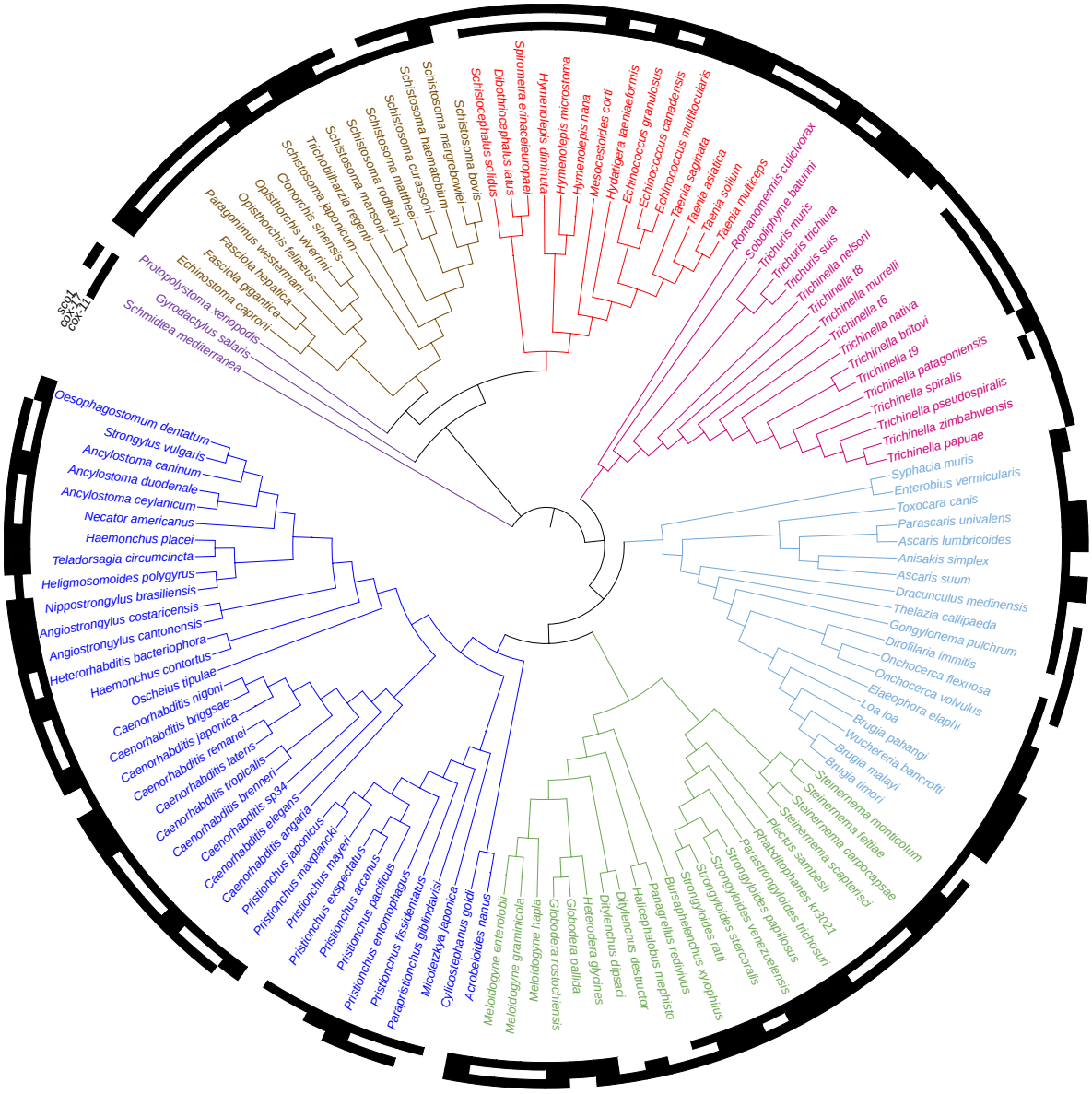


Figure 12: Presence of essential COX assembly factors in 129 species of helminth.

Essential assembly factors necessary for formation of the COX catalytic core include, (from inside to outside), *cox-11*, *cox-17*, and *sco1*.

3.3 Investigating the presence of genes encoding peroxisomal proteins

Because we found such a striking absence of the terminal oxidase for the mitochondrial respiratory chain, we asked whether the other major oxygen-consuming organelle, the peroxisome, was similarly impacted, and if so, was it impacted in the same lineages on the tree? As we found for the COX-related protein-coding genes, there also appears to be an absence of peroxisomal protein-coding genes in the species in our collection (see Figure 13). *C. elegans* and other members of the *Caenorhabditis* genus, as well as free-living clade V nematodes, have the most complete set of these peroxisomal genes. The platyhelminth and clade I nematodes, however, have lost nearly all their peroxisomal genes - not just those with antioxidant function, but also those related to other peroxisomal metabolic pathways. Very few species on the tree have the four genes encoding peroxin proteins that are essential for organelle formation (peroxin proteins Pex3, Pex10, Pex12, and Pex19)[109], strongly suggesting that the entire organelle is missing from these species (see Figure 14).

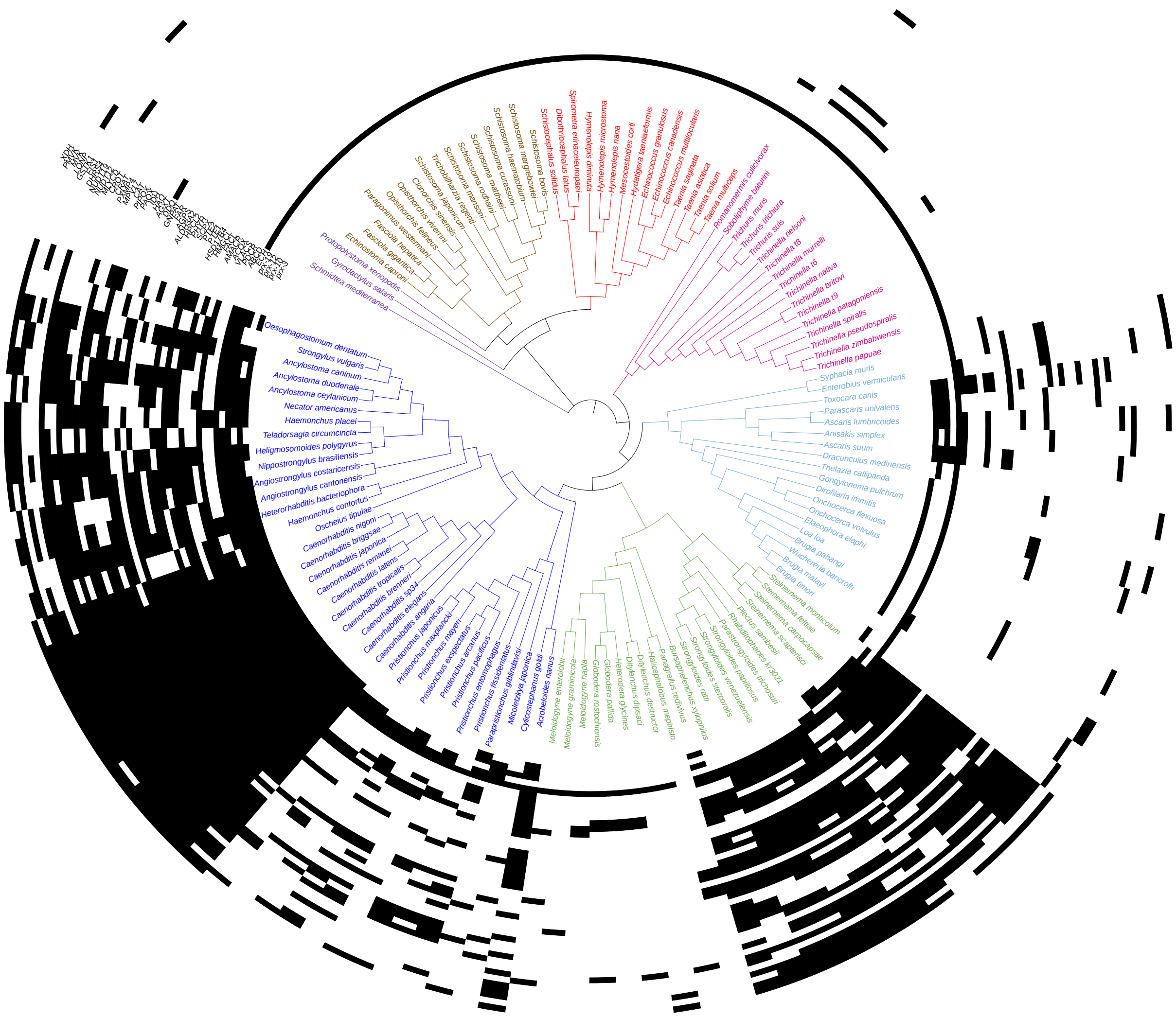


Figure 13: Absence of peroxisomal protein-coding genes in 129 species of helminth

Genes used include (from inside to outside): *prx-3*, *prx-10*, *prx-12*, *prx-19*, ECH, ABCD, PDCR, VLACS, ACOX, AMACR, HMGCL, SPCX, PHYH, HPCL2, AGPS, FAR, GNPAT, DDO, AGXT, HAO, IDH, PAOX, CAT, MPV17, PXMP4, CRAT, CROT, MLYCD, NUDT12, NUDT19, DHRS4, FIS1, GSTK1, MVK, PMVK, and XDH. For a full list of non-abbreviated protein-coding genes and the associated peroxisomal pathway, see Table 5.

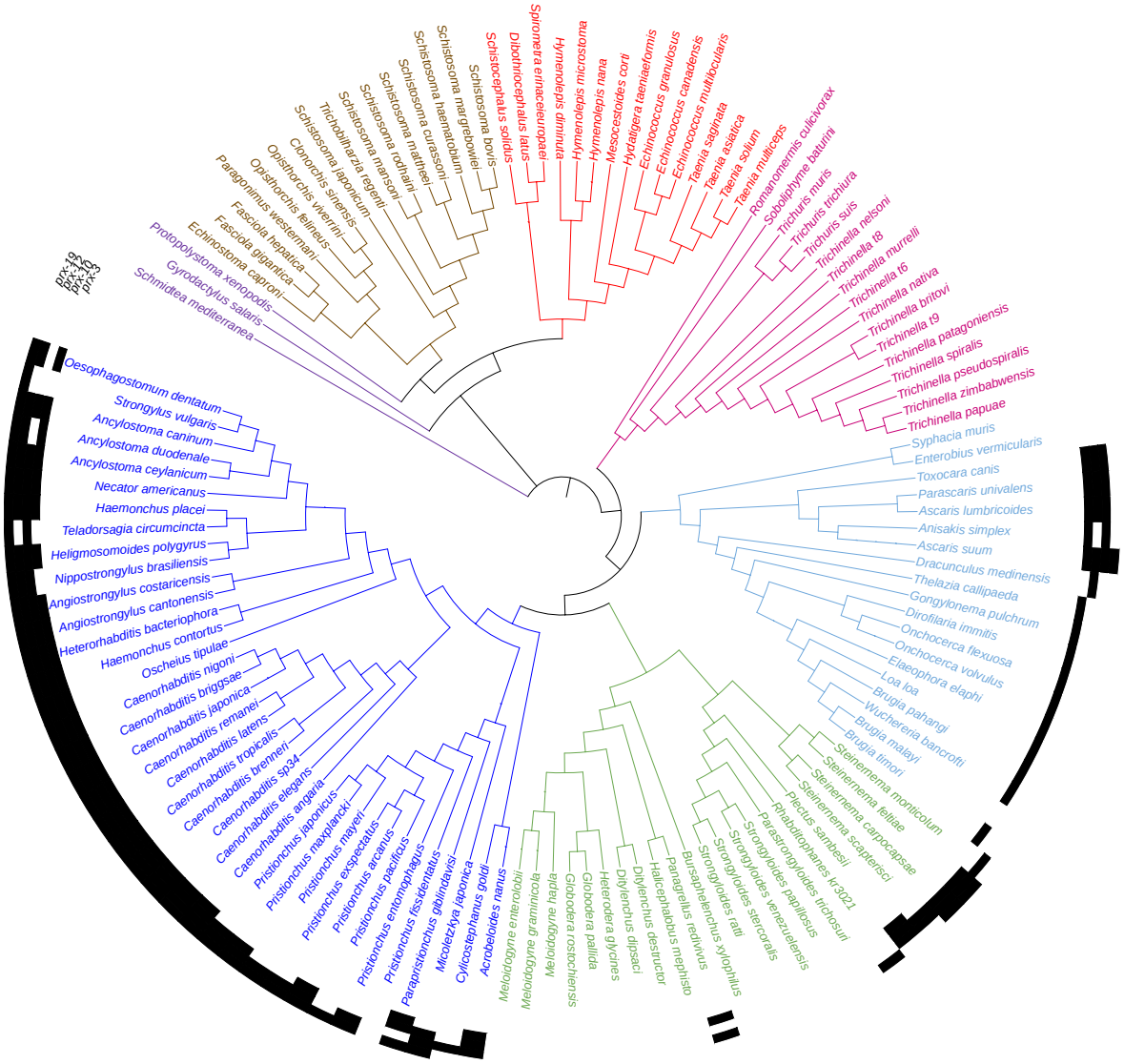


Figure 14: Presence of essential peroxin protein-coding genes in 129 species of helminth. Genes include (from inside to outside) *prx-3*, *prx-10*, *prx-12*, *prx-19*.

Given that the wide-spread absence of peroxisomal biogenesis genes appeared to occur in a similar subset of species in which we observed an absence of COX-related genes, we then determined whether those absences occurred in the same worm lineages by overlaying the presence/absence profiles for each set of genes (see Figure 15). Many of the parasitic species that are predicted to have no functional COX based on the presence of essential assembly factors are also predicted to lack the peroxins required for peroxisome biogenesis.

Figure 16 shows the percentage of species found to have all essential peroxin and COX assembly factors, as well as the species that have lost either one or both sets of genes. It is clear that more than half the species examined here have lost both the terminus of the canonical eukaryotic mitochondrial respiratory chain and the typical suite of peroxisomal functions that are typically considered essential in higher eukaryotes.

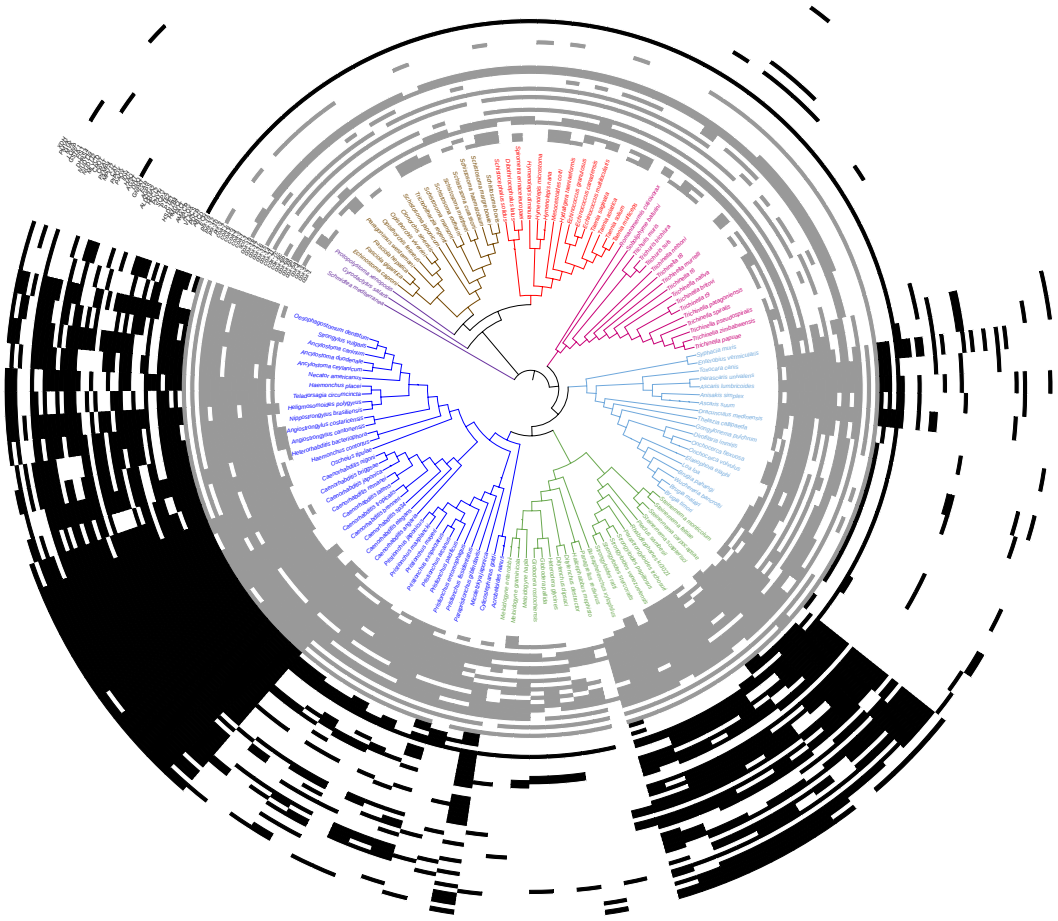


Figure 15 Presence/absence profile showing the distribution of the absences of genes encoding peroxisomal and COX-related proteins

Genes include (from inside to outside): COX related genes in grey, *cox-1*, *cox-2*, and *cox-3*, *cox-4*, *cox-5a*, *cox5b*, *cox-6a*, *cox-6b*, *cox-6c*, and *cox-7c*, *cox-11*, *cox-14*, *cox-15*, *cox-16*, *cox-17*, *cox-18*, *cox-19*, *sco1, coa-1*, *coa-3*, *coa-4*, *coa-5*, *coa-6*, *coa-7*; and peroxisomal genes in black, *prx-3*, *prx-10*, *prx-12*, *prx-19*, ECH, ABCD, PDCR, VLACS, ACOX, AMACR, HMGCL, SPCX, PHYH, HPCL2, AGPS, FAR, GNPAT, DDO, AGXT, HAO, IDH, PAOX, PIPOX, CAT, MPV17, PXMP4, CRAT, CROT, MLYCD, NUDT12, NUDT19, DHRS4, FIS1, GSTK1, MVK, PMVK, and XDH. A full list of non-abbreviated protein-coding genes can be found in Table 5.

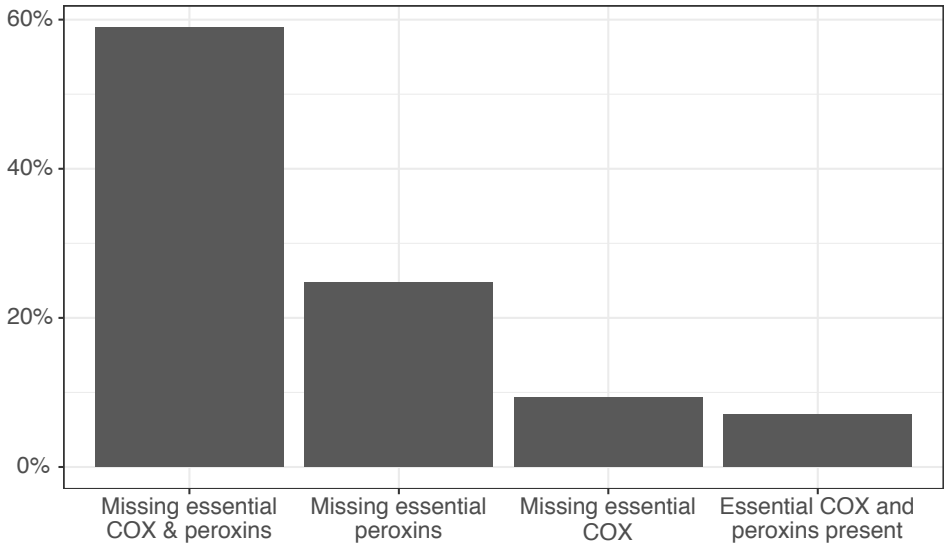


Figure 16: Percent of species of interest that have lost protein-coding genes essential to COX and peroxisomes

Genes essential to the formation of COX include *cox-11*, *cox-17*, and *sco1*; genes essential for peroxisomal biogenesis include *prx3, prx10, prx12,* and *prx19.*

3.4 Investigating the presence of antioxidant protein-coding genes

Given the absence of two significant sources of oxygen consumption, we sought to determine whether there would be any consequences for the proteins associated with cellular antioxidant defenses. We identified superoxide dismutase protein-coding genes in helminths with copper (Sod_Cu) and iron (Sod_Fe) cofactors, as well as other redox proteins-coding genes (glutathione peroxidase, thiolase, peroxidase, thioredoxin, and glutaredoxin) present in our species of interest. Although catalase is involved in defense against reactive oxygen species, we chose to include it as a peroxisomal gene and so it does not appear in this dataset.

Figure 17 demonstrates that, unlike the peroxisomal genes and COX assembly factor genes, the antioxidant protein-coding genes we searched for, which represent a broad swath of cellular antioxidant defenses, are widely present in the helminth species investigated here.

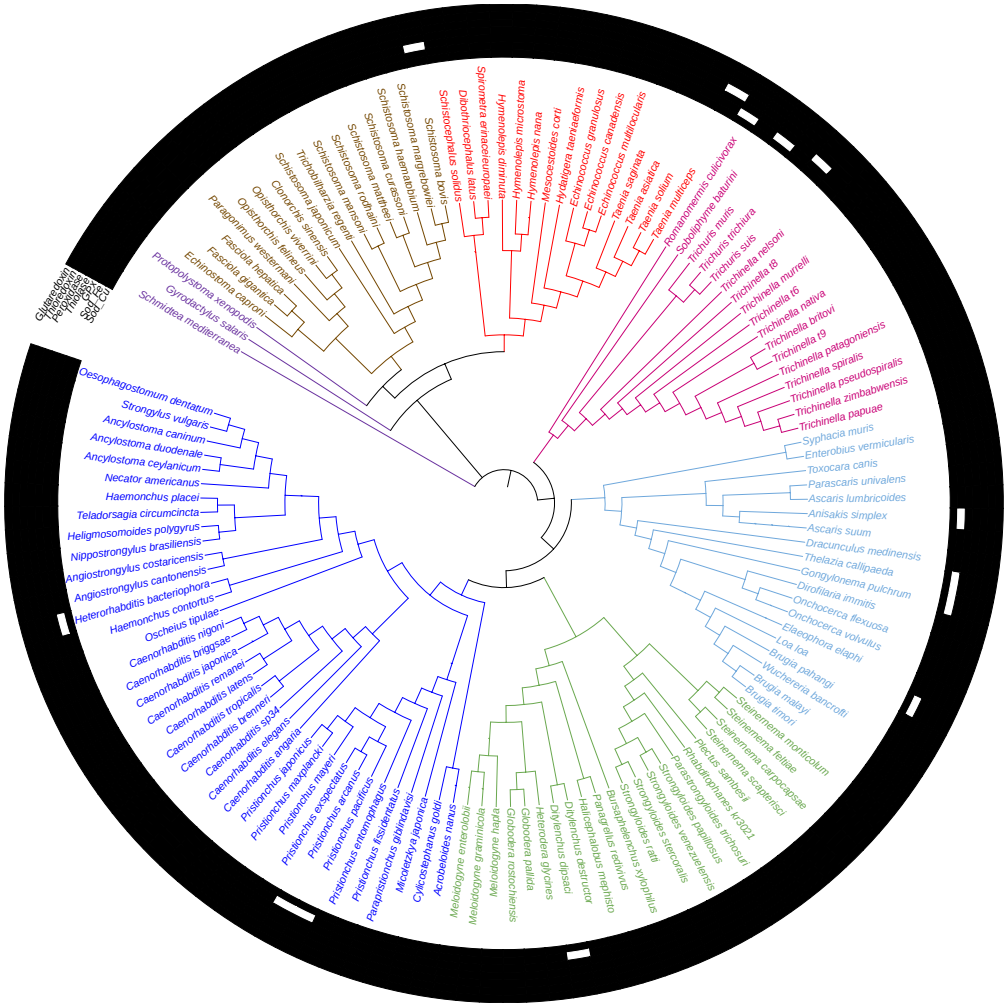


Figure 17: Antioxidant protein-coding genes in 129 species of helminth. Genes include (from inside to outside): superoxide dismutases Sod_Cu and Sod_Fe, glutathione peroxidase, thiolase, peroxidase, thioredoxin, and glutaredoxin.

Chapter 4: Discussion

Helminths are known to be rapidly evolving, with each parasitic species having undergone specific genomic adaptations that allow it to infect, exist, and reproduce within its host[112]. Parasitic flatworms have undergone extensive genome reduction relative to their nematode counterparts, resulting in the loss of many metabolic pathways, especially those related to digestion and auxiliary metabolism, with the most extreme example of this being the cestodes, which have lost their internal body cavity and digestive systems[112]. Given the recent availability of newly sequenced genomes from the 50 Helminth Genomes Project[1], and the fact that helminths are unusual among eukaryotes for their ability to switch between aerobic and anaerobic metabolism, we chose to investigate the presence of genes associated with the metabolism of molecular oxygen in these organisms.

Helminths have already been reported to have adaptations to their mitochondria which allow for the switch between aerobic and anaerobic life-cycle stages. Stage-specific expression of ubiquinone and rhodoquinone allows for a reversal in the reaction direction of succinate dehydrogenase, resulting in the reduction of fumarate to succinate in anaerobic conditions[22]. Early metabolic studies on helminths had also suggested that these organisms have differing cytochrome systems, with adult *Ascaris lumbricoides* and *Schistosoma mansoni* having no detectable COX activity even in aerobic conditions[22]. Alternate mechanisms for energy production are known to exist, via fermentation reactions and the incomplete oxidation of carbon compounds, however the presence of a complete cytochrome oxidase complex had not been investigated in helminth species. We used this research as a

starting point for our bioinformatics-based investigations into helminth oxygen usage in the mitochondria by COX.

To fully assess the presence of COX in our species of interest, we first looked for major genes encoding proteins involved in the overall process of cellular respiration, allowing us to determine whether electron transport chain-related genes were being lost from the genomes at a higher rate than other genes involved in the respiratory process. We searched for genes related to all major glycolytic and TCA enzymes, and found that, although there is an occasional missing gene, these enzymes were present in almost all our species of interest. We did observe an absence of phosphoglycerate mutase (PGM) in many of the species of platyhelminth, but this could be attributed to the use of a *C. elegans* protein as our target sequence for tBLASTn searches. *C. elegans* and other free-living nematodes possess only an independent PGM (iPGM) protein which operates without cofactors[113]. The species where we observed an absence of iPGM may instead have a cofactor dependent PGM enzyme (dPGM), which would not have been detected in our search.

The cases where a gene is missing from these datasets could potentially be attributed to incomplete genomic data, however, when we ran our preliminary investigation of ETC complexes, we noticed that genes related to COX and ATP synthase were absent at a far higher rate than any other gene we searched for, suggesting that, in certain lineages of helminth, these complexes are potentially being lost. In the case of incomplete genomes, we would expect to see an equal distribution of genes absent across all the pathways investigated, and beginning our analysis by looking at genes related to glycolysis and the

TCA cycle, processes which should be highly conserved in helminths, allowed us to establish a baseline level of gene absence/presence for a metabolic process. The absence of genes related to COX is significantly higher than the baseline set by the other cellular respiration processes.

To further assess the absence of COX specifically, we identified and performed tBLASTn searches for the entire complement of genes encoding subunits and assembly factors required for the formation of a functional COX in *C. elegans*. Our results show a widespread absence of many of these subunits and assembly factors from species in all the major lineages of helminth. Although subunits encoded by *cox-1, cox-2*, and *cox-3* are essential to the formation of the complex and should be sufficient to determine whether the enzyme is present, we were unable to rule out the possibility that data relating to these subunits was absent because they are encoded by the mitochondrial genomes, which may not have been fully covered, depending on sequencing methods used. Also, because of the possibility of mitochondrial pseudogenes, it is difficult to determine whether BLAST hits for mitochondrial encoded genes are representative of the mitochondrial genome[114].

The potential lack of mitochondrial sequences could additionally account for the apparent absence of ATP6, which we used as a marker for the presence of ATP synthase. ATP6 is mitochondrial encoded, and we cannot yet determine from our data whether its absence in certain species is indicative of an absence of ATP synthase itself, or if the sequences were not available. Interestingly, ATP8, another component of ATP synthase, is absent from the mitochondrial genomes of both *C. elegans* and *A. suum*[110], so even if the

mitochondrial genomes are included in our dataset, the absence of ATP6 may not correspond to an absence of ATP synthase. Future analysis will need to be conducted to assess the presence of nuclear-encoded subunits of this ETC complex to determine if it is present in these species.

We chose to use the presence of three nuclear encoded assembly factors, *cox-11*, *cox-17*, and *sco1*, as a definitive marker for the presence of COX, as the proteins they encode are involved in the assembly of subunits 1-3 and are essential to the proper formation of the enzyme's catalytic core[24]. Only 31.2% of our species of interest were found to have all three of the genes encoding these essential assembly factors, suggesting that COX is absent from many species of helminth (see S. Fig. 2). Our findings agree with the metabolic studies done in adult *Ascaris lumbricoides* and *Schistosoma mansoni*[22] where little or no COX activity was observed: in our dataset, *S. mansoni* was missing not just the three major mitochondrial subunits, but most of its assembly factors as well; and although all the subunits were present in *A. lumbricoides*, it too was missing assembly factors which could account for the absence of enzymatic activity.

This suggests that helminths are utilizing a different final electron acceptor in their mitochondrial electron transport chains, although we were only able to identify the presence of an alternative oxidase in two species, *Ditylenchus destructor* and *Ditylenchus dipsaci*. It could be that, because helminths are able to utilize anaerobic metabolism, mitochondrial genes involved in aerobic respiration are becoming redundant and are gradually being lost. It is also possible that different species of helminths require a different set of core COX genes

and assembly factors than *C. elegans*, although *cox-17, cox-11,* and *sco1* are highly conserved and required for the formation of COX in other eukaryotes[23], making this unlikely. We did note many species that had one or two of the three assembly factors, perhaps suggesting that the loss of these genes is a gradual process. The presence of assembly factors when essential subunits are absent, especially when these assembly factors are found within multi-domain protein architectures, could also suggest additional roles for these factors outside of COX assembly.

The absence of COX-related genes may, in fact, be more extensive than shown here: many of the Cox17p, Cox11p, and Sco1p proteins in our predicted protein library were part of multi-domain architectures, in combinations with other protein families which function in other metabolic processes and localize outside of the mitochondria (data not shown). The ability of these domains to fulfill their normal roles in these multi-architectural proteins will require further investigation, but for the purposes of this study, we consider them at least indicative of some level of function, given that the sequences from COX-related domains are maintained. If it is later revealed that these multi-domain proteins do not function in COX assembly, or that the COX-related domains within them are functioning in a different pathway, then the presence/absence profiles displayed here would need to be reconstructed to exclude these species.

In addition to the mitochondria, we were interested in investigating peroxisome metabolism, as peroxisomes both utilize oxygen in oxidation pathways and act as a scavenger of reactive oxygen species, making them important for antioxidant defense. Previous studies

have already documented the genomic loss of peroxisomal catalase in platyhelminths, and a complete loss of the organelle has been suspected in nine species of nematode and all platyhelminths[115]. Other non-helminth anaerobic eukaryotes have been observed to lose their peroxisomes, or to contain so called 'anaerobic peroxisomes', which lose their oxygen-utilizing metabolic pathways while retaining other peroxisome functions[6]. The present study expands on this data, investigating the major peroxisomal pathways in a broader selection of helminth species. We searched for genes involved in antioxidant defense and oxidation pathways (alpha, beta, unsaturated fatty acid, and other types of oxidation), as well as non-oxygen utilizing pathways (etherphospholipid biosynthesis, amino acid metabolism, retinol metabolism, glutathione metabolism, sterol precursor biosynthesis, and purine metabolism).

We had initially expected to observe the presence of anaerobic peroxisomes based on the ability of helminth mitochondria to function anaerobically. The absence of peroxisomal genes we observed goes beyond just oxygen utilizing pathways, and occurs in more species of helminth than previously demonstrated. Platyhelminth lineages, in agreement with previous studies[115], were missing all peroxisomal protein-coding genes (except for enoyl CoA hydratase, which encodes a beta oxidation protein that can also be present in the mitochondria[35], which would account for its presence here). We also observed an absence of most peroxisomal genes in all clade I and clade III nematodes in our dataset, and most clade IV and V nematodes are missing at least some of the genes.

As a definitive marker for the presence of peroxisomal organelles, we looked specifically at four peroxin genes, encoding Pex3, Pex10, Pex12, and Pex19[109]. While there

are many different peroxin proteins in eukaryotes including some that are species specific, these are considered part of a core set of proteins that are conserved in all eukaryotic lineages[109]. We selected these four as they cover two major peroxin functions essential to the biogenesis of the organelle itself. Of our 129 species of helminth worm, we determined that 83.7% of species were lacking essential peroxins. As with the absence of COX assembly factors, we identified species that had different combinations of these four essential peroxins, but the only species we identified with all four were clade V nematodes, mainly of the genus *Caenorhabditis*, and *Ascaris suum*. The different combinations of peroxins present suggests that peroxisome loss may be an ongoing process.

When considering the absence of peroxisomal and COX protein-coding genes together, we observed that the lineages of helminth with missing protein-coding genes overlapped between the two datasets, and 58.9% of species investigated were missing both their essential COX assembly factors and essential peroxins. Only 7% of species on our phylogenetic tree had all our essential marker genes, and another 24.8% were missing their essential peroxins but retained their essential COX assembly. Although our findings are preliminary, we propose that, after peroxisomal loss in the reductive genome evolution of anaerobic eukaryotes such as helminths, mitochondrial genes that encode proteins involved in aerobic metabolism are the next set of genes to be lost.

The final set of protein-coding genes of interest to this project were antioxidant genes, given that parasitic helminths must have robust antioxidant defense systems to cope with bursts of reactive oxygen species released by their hosts[21]. If peroxisomes are being lost,

along with their antioxidant pathways and catalase, other antioxidant proteins will be of increasing importance to the survival of these organisms. This is supported by our data that shows that copper and iron superoxide dismutases, glutathione peroxidase, thiolase, peroxidase, glutaredoxin, and thioredoxin genes are all highly maintained across our species of interest. When looking in our proteome library for our dataset, we additionally identified hundreds of protein architectures containing amino acid sequences from these antioxidant proteins, many of which are entirely unique to parasitic helminths (data not shown).

The absence of protein coding genes described here could be accounted for by either evolutionary reductive genome evolution, or it could be a technical artifact. We have chosen to discuss the former possibility in detail, given that the complete BUSCO scores are within a reasonable range for genome completeness (with an average score of 83%). Although there is some variation in the scores, with the platyhelminth lineages having slightly lower scores overall than the nematodes, this variation is not enough to account for the significant amount of absent genes. For example, the members of *Trichinella* in the clade I nematodes have complete BUSCO values comparable to those of *Caenorhabditis* in clade V, however we were unable to detect any peroxisomal genes in *Trichinella*, and all associated species were missing at least one essential COX assembly factor, while these genes are retained in the clade V nematodes.

The absence of genes could also be a result of the technical difficulties that come with sequencing small genomes, particularly that of the mitochondria[111]. The life cycle stage of helminth, tissue type, and DNA extraction method utilized by the research groups who

conducted the sequencing experiments could additionally result in incomplete genomes. If gene absences are determined to be a technical artifact and not the result of genome evolution, then the genomes available on WormBase ParaSite, especially the reference genomes, are less complete and accurate than thought.

The possibility that some species are undergoing chromatin diminution could also result in gene absences observed, although this phenomenon has only been documented so far in certain species of clade III nematode (*Ascaris suum*, *Ascaris lumbricoides*, *Toxocara canis*, and *Parascaris univalens*) and could only account for losses in these species[12].

We also note that it is easier to demonstrate the absence of a gene than it is to prove the presence of one, and while we are confident in our findings of absent protein-coding genes, it is difficult to verify whether the 'hits' we identified for our genes of interest are able to encode functional proteins, especially given the possibility of pseudogenes (both of mitochondrial and nuclear origin) and the presence of multi-domain fusion architectures whose functions are currently unknown.

Assuming that the absent genes identified here are the result of genome evolution and gene loss, taking these three sets of oxygen-handling genes together, it can be concluded that helminths utilize oxygen differently from other eukaryotic organisms. With their ability and apparent preference for utilizing anaerobic pathways and ongoing rapid, reductive genome evolution, these organisms have either lost, or are undergoing the process of losing their peroxisomes, and are also losing genes related to cytochrome *c* oxidase. Over the course of helminth evolution oxygen is becoming less of a necessary metabolite, but it still poses a

potential danger in the form of host-derived hydrogen peroxide, resulting in the retention of non-peroxisomal protein-coding genes involved in antioxidant defense. The absence of peroxisomes and catalase may also account for the presence of unique architectures containing other antioxidant protein families, which we propose could be serving a protective role. Furthering the understanding of helminth metabolism and identifying key differences from host metabolism could provide future therapeutic targets to be exploited in the development of new anthelminthic agents. For example, if the absences of genes encoding COX subunits and assembly factors observed here correspond to a complete loss of the enzyme, which would restrict helminths to using only anaerobic metabolism, then enzymes that allow for malate dismutation could be targeted. Therapies that inhibit rhodoquinone or fumarate reductase could prevent parasitic helminths lacking COX from utilizing malate dismutation and would deprive them of their cellular energy without impacting the host. Similarly, with the loss of peroxisomes and catalase, other helminth antioxidant systems could be under increased pressure to manage ROS, and targeting other helminth-specific antioxidant systems could also combat infections by this type of organism.

References

1. Coghlan, A. *et al.* Comparative genomics of the major parasitic worms. *Nature Genetics* 51, 163–174 (2019).

2. Hotez, P. J. *et al.* Helminth infections: the great neglected tropical diseases. *Journal of Clinical Investigation* 118, 1311–1321 (2008).

3. Alberts, B. *et al. Molecular Biology of the Cell, 6th ed.* (Garland Science, 2015).

4. Voet, D. & Voet, J. *Biochemistry, 4th ed.* (Wiley Global Education, 2010).

5. Müller, M. *et al.* Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiology and Molecular Biology Reviews* 76, 444–495 (2012).

6. Le, T. *et al.* Anaerobic peroxisomes in *Mastigamoeba balamuthi*. *Proceedings of the National Academy of Sciences* 117, 2065–2075 (2020).

7. Wikström, M., Krab, K. & Sharma, V. Oxygen activation and energy conservation by cytochrome *c* oxidase. *Chemical Reviews* 118, 2469–2490 (2018).

8. Viney, M. The genomic basis of nematode parasitism. *Briefings in Functional Genomics* 17, 8–14 (2018).

9. Thompson, D. & Geary, T. G. Biochemistry and Molecular Biology of Parasites. in 203–226 (Elsevier, 1995). doi:10.1016/B978-0-12-473345-9.X5000-4.

10. Wang, J. & Davis, R. E. Programmed DNA elimination in multicellular organisms. *Curr Opin Genet Dev* 27, 26–34 (2014).

11. Müller, F., Wicky, C., Spicher, A. & Tobler, H. New telomere formation after developmentally regulated chromosomal breakage during the process of chromatin diminution in *Ascaris lumbricoides*. *Cell* 67, 815–822 (1991).

12. Tobler, H., Etter, A. & Müller, F. Chromatin diminution in nematode development. *Trends in Genetics* 8, 427–432 (1992).

13. Müller, F. & Tobler, H. Chromatin diminution in the parasitic nematodes *Ascaris suum* and *Parascaris univalens*. *International Journal for Parasitology* 30, 391–399 (2000).

14. Bachmann-Waldmann, C., Jentsch, S., Tobler, H. & Müller, F. Chromatin diminution leads to rapid evolutionary changes in the organization of the germ line genomes of the parasitic nematodes A. suum and P. univalens. *Molecular and Biochemical Parasitology* 134, 53–64 (2004).

15. Saz, H. J. Comparative energy metabolisms of some parasitic helminths. *The Journal of Parasitology* 56, 634 (1970).

16. Bharti, B., Bharti, S. & Khurana, S. Worm infestation: diagnosis, treatment and prevention. *The Indian Journal of Pediatrics* 85, 1017–1024 (2018).

17. Kaminsky, R. *et al.* Worms—A "license to kill." *Veterinary Parasitology* 195, 286–291 (2013).

18. Saz, H. J. Helminths: Primary models for comparative biochemistry. *Parasitology Today* 6, 92–93 (1990).

19. Wang, E. J. & Saz, H. J. Comparative biochemical studies of *Litomosoides carinii*, *Dipetalonema viteae*, and *Brugia pahangi* adults. *The Journal of Parasitology* 60, 316 (1974).

20. Paulsen, C. E. & Carroll, K. S. Cysteine-mediated redox signaling: Chemistry, biology, and tools for discovery. *Chemical Reviews* 113, 4633–4679 (2013).

21. Tielens, A. G. M. Energy generation in parasitic helminths. *Parasitology Today* 10, 346–352 (1994).

22. Bueding, E. & Charms, B. Cytochrome *c*, cytochrome oxidase, and succinoxidase activities of helminths. *J Biol Chem* 196, 615–27 (1952).

23. Watson, S. A. & McStay, G. P. Functions of cytochrome *c* oxidase assembly factors. *International Journal of Molecular Sciences* 21, 7254 (2020).

24. Glerum, D. M. & Tzagoloff, A. Submitochondrial distributions and stabilities of subunits 4, 5, and 6 of yeast cytochrome oxidase in assembly defective mutants. *FEBS Letters* 412, 410–414 (1997).

25. Carr, H. S., George, G. N. & Winge, D. R. Yeast Cox11, a protein essential for cytochrome *c* oxidase Assembly, Is a Cu(I)-binding Protein. *Journal of Biological Chemistry* 277, 31237–31242 (2002).

26. Brière, J.-J. & Tzagoloff, A. The Scoop on Sco. *Molecular Cell* 25, 176–178 (2007).

27. Glerum, D. M., Shtanko, A. & Tzagoloff, A. Characterization of COX17, a yeast gene involved in copper metabolism and assembly of cytochrome oxidase. *Journal of Biological Chemistry* 271, 14504–14509 (1996).

28. Bateman, A. *et al.* UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* 49, D480–D489 (2021).

29. Maxfield, A. B., Heaton, D. N. & Winge, D. R. Cox17 is functional when tethered to the mitochondrial inner membrane. *Journal of Biological Chemistry* 279, 5072–5080 (2004).

30. Horn, D. & Barrientos, A. Mitochondrial copper metabolism and delivery to cytochrome *c* oxidase. *IUBMB Life* 60, 421–429 (2008).

31. Balatri, E., Banci, L., Bertini, I., Cantini, F. & Ciofi-Baffoni, S. Solution structure of Sco1. *Structure* 11, 1431–1443 (2003).

32. Baker, Z. N., Cobine, P. A. & Leary, S. C. The mitochondrion: a central architect of copper homeostasis. *Metallomics* 9, 1501–1512 (2017).

33. Schrader, M. & Yoon, Y. Mitochondria and peroxisomes: Are the 'Big Brother' and the 'Little Sister' closer than assumed? *BioEssays* 29, 1105–1114 (2007).

34. Wanders, R. J. A. Metabolic functions of peroxisomes in health and disease. *Biochimie* 98, 36–44 (2014).

35. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Research* 49, D545–D551 (2021).

36. Turrens, J. F. Mitochondrial formation of reactive oxygen species. *The Journal of Physiology* 552, 335–344 (2003).

37. Wu, W. K., Mak, C. H. & Ko, R. C. Cloning and differential expression of manganese superoxide dismutase (Mn-SOD) of *Trichinella pseudospiralis*. *Parasitology Research* 102, 251–258 (2007).

38. Guevara-Flores, A., Martínez-González, J., Rendón, J. & del Arenal, I. The architecture of thiol antioxidant systems among invertebrate parasites. *Molecules* 22, 259 (2017).

39. Fridovich, I. Superoxide Dismutases. in 35–97 (2006). doi:10.1002/9780470122860.ch2.

40. Margis, R., Dunand, C., Teixeira, F. K. & Margis-Pinheiro, M. Glutathione peroxidase family - an evolutionary overview. *FEBS Journal* 275, 3959–3970 (2008).

41. Arnér, E. S. J. & Holmgren, A. Physiological functions of thioredoxin and thioredoxin reductase. *European Journal of Biochemistry* 267, 6102–6109 (2000).

42. Salinas, G., Selkirk, M. E., Chalar, C., Maizels, R. M. & Fernández, C. Linked thioredoxin-glutathione systems in platyhelminths. *Trends in Parasitology* 20, 340–346 (2004).

43. Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P. & Berriman, M. WormBase ParaSite − a comprehensive resource for helminth genomics. *Molecular and Biochemical Parasitology* 215, 2–10 (2017).

44. Rödelsperger, C. *et al.* Single-molecule sequencing reveals the chromosome-scale genomic architecture of the nematode model organism *Pristionchus pacificus*. *Cell Reports* 21, 834–844 (2017).

45. Grohme, M. A. *et al.* The genome of *Schmidtea mediterranea* and the evolution of core cellular mechanisms. *Nature* 554, 56–61 (2018).

46. Rošić, S. *et al.* Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. *Nature Genetics* 50, 452–459 (2018).

47. Srinivasan, J. *et al.* The draft genome and transcriptome of *Panagrellus redivivus* are shaped by the harsh demands of a free-living lifestyle. *Genetics* 193, 1279–1295 (2013).

48. Besnard, F., Koutsovoulos, G., Dieudonné, S., Blaxter, M. & Félix, M.-A. Toward universal forward genetics: using a draft genome sequence of the nematode *Oscheius tipulae* to identify mutations affecting vulva development. *Genetics* 206, 1747–1761 (2017).

49. Prabh, N. *et al.* Deep taxon sampling reveals the evolutionary dynamics of novel gene families in *Pristionchus* nematodes. *Genome Research* 28, 1664–1674 (2018).

50. Wasik, K. *et al.* Genome and transcriptome of the regeneration-competent flatworm, *Macrostomum lignano*. *Proceedings of the National Academy of Sciences* 112, 12462–12467 (2015).

51. Weinstein, D. J. *et al.* The genome of a subterrestrial nematode reveals adaptations to heat. *Nature Communications* 10, 5268 (2019).

52. Fradin, H. *et al.* Genome architecture and evolution of a unichromosomal asexual nematode. *Current Biology* 27, 2928-2939.e6 (2017).

53. Kanzaki, N. *et al.* Biology and genome of a newly discovered sibling species of *Caenorhabditis elegans*. *Nature Communications* 9, 3216 (2018).

54. Yin, D. *et al.* Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins. *Science,* 359, 55–61 (2018).

55. Fierst, J. L. *et al.* Reproductive mode and the evolution of genome size and structure in *Caenorhabditis* nematodes. *PLOS Genetics* 11, e1005323 (2015).

56. Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. *Science* 282, 2012–2018 (1998).

57. Schiffer, P. H. *et al.* The gene regulatory program of *Acrobeloides nanus* reveals conservation of phylum-specific expression. *Proceedings of the National Academy of Sciences* 115, 4459–4464 (2018).

58. Small, S. T. *et al.* Population genomics of the filarial nematode parasite *Wuchereria bancrofti* from mosquitoes. *Molecular Ecology* 25, 1465–1477 (2016).

59. Jex, A. R. *et al.* Genome and transcriptome of the porcine whipworm *Trichuris suis*. *Nature Genetics* 46, 701–706 (2014).

60. Foth, B. J. *et al.* Whipworm genome and dual-species transcriptome analyses provide molecular insights into an intimate host-parasite interaction. *Nature Genetics* 46, 693–700 (2014).

61. Mitreva, M. *et al.* The draft genome of the parasitic nematode *Trichinella spiralis*. *Nature Genetics* 43, 228–235 (2011).

62. Korhonen, P. K. *et al.* Phylogenomic and biogeographic reconstruction of the *Trichinella* complex. *Nature Communications* 7, 10513 (2016).

63. Borvinskaya, E. *et al.* Comparative analysis of proteins of functionally different body darts of the fish parasites *Triaenophorus nodulosus* and *Triaenophorus crassus*. *Acta Parasitologica* (2021) doi:10.1007/s11686-021-00384-6.

64. Wang, S. *et al.* Comparative genomics reveals adaptive evolution of Asian tapeworm in switching to a new intermediate host. *Nature Communications* 7, 12845 (2016).

65. Li, W. *et al.* The genome of tapeworm *Taenia multiceps* sheds light on understanding parasitic mechanism and control of coenurosis disease. *DNA Research* 25, 499–510 (2018).

66. Hunt, V. L. *et al.* The genomic basis of parasitism in the *Strongyloides* clade of nematodes. *Nature Genetics* 48, 299–307 (2016).

67. Nemetschke, L., Eberhardt, A. G., Hertzberg, H. & Streit, A. Genetics, chromatin diminution, and sex chromosome evolution in the parasitic nematode genus *Strongyloides*. *Current Biology* 20, 1687–1696 (2010).

68. Dillman, A. R. *et al.* Comparative genomics of *Steinernema* reveals deeply conserved gene regulatory networks. *Genome Biology* 16, 200 (2015).

69. Serra, L. *et al.* Hybrid assembly of the genome of the entomopathogenic nematode *Steinernema carpocapsae* identifies the X-chromosome. *G3 Genes|Genomes|Genetics* 9, 2687–2697 (2019).

70. Bennett, H. M. *et al.* The genome of the sparganosis tapeworm *Spirometra erinaceieuropaei* isolated from the biopsy of a migrating brain lesion. *Genome Biology* 15, 510 (2014).

71. Protasio, A. v. *et al.* A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases* 6, e1455 (2012).

72. Luo, F. *et al.* An improved genome assembly of the fluke *Schistosoma japonicum*. *PLOS Neglected Tropical Diseases* 13, e0007612 (2019).

73. Stroehlein, A. J. *et al.* High-quality *Schistosoma haematobium* genome achieved by single-molecule and long-range sequencing. *Gigascience* 8, (2019).

74. Oey, H. *et al.* Whole-genome sequence of the bovine blood fluke *Schistosoma bovis* supports interspecific hybridization with *S. haematobium*. *PLOS Pathogens* 15, e1007513 (2019).

75. Schiffer, P. H. *et al.* The genome of *Romanomermis culicivorax*: revealing fundamental changes in the core developmental genetic toolkit in Nematoda. *BMC Genomics* 14, 923 (2013).

76. Hunt, V. L. *et al.* The genomic basis of parasitism in the *Strongyloides* clade of nematodes. *Nature Genetics* 48, 299–307 (2016).

77. Oey, H. *et al.* Whole-genome sequence of the oriental lung fluke *Paragonimus westermani*. *Gigascience* 8, (2019).

78. Young, N. D. *et al.* The *Opisthorchis viverrini* genome provides insights into life in the bile duct. *Nature Communications* 5, 4378 (2014).

79. Ershov, N. I. *et al.* New insights from *Opisthorchis felineus* genome: update on genomics of the epidemiologically important liver flukes. *BMC Genomics* 20, 399 (2019).

80. Tang, Y. T. *et al.* Genome of the human hookworm *Necator americanus*. *Nature Genetics* 46, 261–269 (2014).

81. Opperman, C. H. *et al.* Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences* 105, 14802–14807 (2008).

82. Somvanshi, V. S., Tathode, M., Shukla, R. N. & Rao, U. Nematode genome announcement: A draft genome for rice root-knot nematode, *Meloidogyne graminicola*. *Journal of Nematology* 50, 111–116 (2018).

83. Szitenberg, A. *et al.* Comparative genomics of apomictic root-knot nematodes: Hybridization, ploidy, and dynamic genome change. *Genome Biology and Evolution* 9, 2844–2861 (2017).

84. Blanc-Mathieu, R. *et al.* Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLOS Genetics* 13, e1006777 (2017).

85. Desjardins, C. A. *et al.* Genomics of *Loa loa*, a *Wolbachia*-free filarial parasite of humans. *Nature Genetics* 45, 495–500 (2013).

86. Nowak, R. M. *et al.* Hybrid de novo whole-genome assembly and annotation of the model tapeworm *Hymenolepis diminuta*. *Scientific Data* 6, 302 (2019).

87. Bai, X. *et al.* A lover and a fighter: The genome sequence of an entomopathogenic nematode *Heterorhabditis bacteriophora*. *PLoS ONE* 8, e69618 (2013).

88. Masonbrink, R. *et al.* The genome of the soybean cyst nematode (*Heterodera glycines*) reveals complex patterns of duplications involved in the evolution of parasitism genes. *BMC Genomics* 20, 119 (2019).

89. Schwarz, E. M. *et al.* The genome and developmental transcriptome of the strongylid nematode *Haemonchus contortus*. *Genome Biology* 14, R89 (2013).

90. Hahn, C., Fromm, B. & Bachmann, L. Comparative genomics of flatworms (platyhelminthes) reveals shared genomic features of ecto- and endoparastic *Neodermata*. *Genome Biology and Evolution* 6, 1105–1117 (2014).

91. Eves-van den Akker, S. *et al.* The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence. *Genome Biology* 17, 124 (2016).

92. Cotton, J. A. *et al.* The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biology* 15, R43 (2014).

93. Tsai, I. J. *et al.* The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* 496, 57–63 (2013).

94. Zheng, H. *et al.* The genome of the hydatid tapeworm *Echinococcus granulosus*. *Nature Genetics* 45, 1168–1175 (2013).

95. Maldonado, L. L. *et al.* The *Echinococcus canadensis* (G7) genome: a key knowledge of parasitic platyhelminth human diseases. *BMC Genomics* 18, 204 (2017).

96. Mimee, B. *et al.* The draft genome of *Ditylenchus dipsaci*. *Journal of Nematology* 51, 1–3 (2019).

97. Zheng, J. *et al.* The *Ditylenchus destructor* genome provides new insights into the evolution of plant parasitic nematodes. *Proceedings of the Royal Society B: Biological Sciences* 283, 20160942 (2016).

98. Godel, C. *et al.* The genome of the heartworm, *Dirofilaria immitis* , reveals drug and vaccine targets. *The FASEB Journal* 26, 4650–4661 (2012).

99. Young, N. D. *et al.* High-quality reference genome for *Clonorchis sinensis*. *Genomics* 113, 1605–1615 (2021).

100. Kikuchi, T. *et al.* Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus. PLoS Pathogens* 7, e1002219 (2011).

101. Ghedin, E. *et al.* Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* 317, 1756–1760 (2007).

102. Wang, J. *et al.* Comparative genome analysis of programmed DNA elimination in nematodes. *Genome Research* 27, 2001–2014 (2017).

103. Schwarz, E. M. *et al.* Erratum: The genome and transcriptome of the zoonotic hookworm *Ancylostoma ceylanicum* identify infection-specific gene families. *Nature Genetics* 47, 689–689 (2015).

104. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* 47, W636–W641 (2019).

105. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17, 540–552 (2000).

106. Lemoine, F. *et al.* NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Research* 47, W260–W265 (2019).

107. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* 49, W293–W296 (2021).

108. Lu, S. *et al.* CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Research* 48, D265–D268 (2020).

109. Jansen, R. L. M., Santana-Molina, C., van den Noort, M., Devos, D. P. & van der Klei, I. J. Comparative genomics of peroxisome biogenesis proteins: Making sense of the PEX proteins. *Front Cell Dev Biol* 9, 654163 (2021).

110. Okimoto, R., Macfarlane, J. L., Clary, D. O. & Wolstenholme, D. R. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics* 130, 471–498 (1992).

111. Jia, W. *et al.* Advances in the study of helminth mitochondrial genomes and their associated applications. *Chinese Science Bulletin* 57, 54–67 (2012).

112. Zarowiecki, M. & Berriman, M. What helminth genomes have taught us about parasite evolution. *Parasitology* 142, S85–S97 (2015).

113. Dhamodharan, R., Hoti, S. L. & Sankari, T. Characterization of cofactor-independent phosphoglycerate mutase isoform-1 (Wb-iPGM) gene: A drug and diagnostic target from human lymphatic filarial parasite, *Wuchereria bancrofti*. *Infection, Genetics and Evolution* 12, 957–965 (2012).

114. Benesh, D. P., Hasu, T., Suomalainen, L.-R., Valtonen, E. T. & Tiirola, M. Reliability of mitochondrial DNA in an acanthocephalan: The problem of pseudogenes. *International Journal for Parasitology* 36, 247–254 (2006).

115. Žárský, V. & Tachezy, J. Evolutionary loss of peroxisomes – not limited to parasites. *Biology Direct* 10, 74 (2015).

Appendix

Supplementary Figures

S. Table 1: Genome assemblies of 129 species of helminth

Species	Type	Genome assembly ID	Host
Ancylostoma caninum	Parasitic	A_caninum_9.3.2.ec.cg.pg	Dogs
Ancylostoma ceylanicum	Parasitic	GCA_000688135.1	Humans, hamsters
Ancylostoma duodenale	Parasitic	GCA_000816745.1	Humans
Angiostrongylus cantonensis	Parasitic	GCA_000950995.1	Rats, humans, snails, slugs, crab, shrimp
Angiostrongylus_costaricensis	Parasitic	GCA_900624975.1	Rats, humans, snails, slugs, crab, shrimp
Anisakis simplex	Parasitic	GCA_900617985.1	Humans, marine mammals, fish
Ascaris lumbricoides	Parasitic	GCA_000951055.1	Humans
Ascaris suum	Parasitic	GCA_000187025.3	Pigs, humans
Brugia malayi	Parasitic	GCA_000002995.5	Mosquitoes, humans
Brugia pahangi	Parasitic	GCA_900618355.1	Mosquitoes, cats, dogs
Brugia timori	Parasitic	GCA_900618025.1	Mosquitoes, humans
Bursaphelenchus xylophilus	Parasitic	GCA_000231135.1	Pine trees
Clonorchis sinensis	Parasitic	GCA_003604175.1	Human
Cylicostephanus goldi	Parasitic	GCA_900617965.1	Horses

Dibothriocephalus latus	Parasitic	GCA_900617775.1	Fish, humans
Dirofilaria immitis	Parasitic	nDi.2.2, University of Edinburgh	Dogs, cats, mosquitoes
Ditylenchus destructor	Parasitic	GCA_001579705.1	Potatoes
Ditylenchus dipsaci	Parasitic	GCA_004194705.1	Onion, garlic
Dracunculus medinensis	Parasitic	GCA_000946415.1	Humans
Echinococcus canadensis	Parasitic	GCA_900004735.1	Dogs
Echinococcus granulosus	Parasitic	GCA_000524195.1	Dogs
Echinococcus multilocularis	Parasitic	GCA_000469725.3	Dogs, humans
Echinostoma caproni	Parasitic	GCA_900618425.1	Humans
Elaeophora elaphi	Parasitic	GCA_000499685.1	Deer, sheep
Enterobius vermicularis	Parasitic	GCA_900576705.1	Humans
Fasciola gigantica	Parasitic	GCA_006461475.1	Ruminants, snails, humans
Fasciola hepatica	Parasitic	PRJNA179522	Humans, ruminants
Globodera pallida	Parasitic	GCA_000724045.1	Potatoes
Globodera rostochiensis	Parasitic	GCA_900079975.1	Potatoes, tomatoes
Gongylonema pulchrum	Parasitic	GCA_900617915.1	Humans, insects
Gyrodactylus salaris	Parasitic	GCA_000715275.1	Freshwater fish
Haemonchus contortus	Parasitic	Hco_v4_coding_submitted, University of Melbourne	Ruminants
Haemonchus placei	Parasitic	GCA_900617895.1	Ruminants
Heligmosomoides polygyrus	Parasitic	GCA_900618505.1	Rodents
Heterodera glycines	Parasitic	GCA_004148225.1	Soybeans

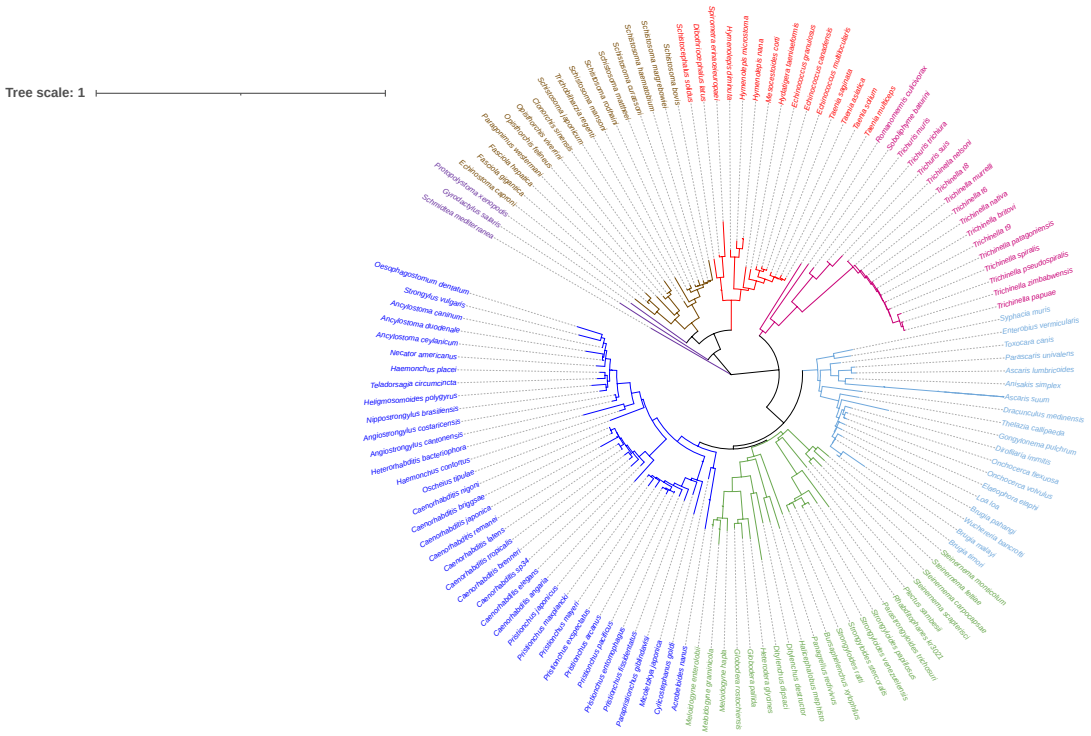
Heterorhabditis bacteriophora	Parasitic	GCA_000223415.1	Insects
Hydatigera taeniaeformis	Parasitic	GCA_900622495.1	Cats, rodents
Hymenolepis diminuta	Parasitic	GCA_902177915.1	Rodents, insects
Hymenolepis microstoma	Parasitic	GCA_000469805.3	Rodents
Hymenolepis nana	Parasitic	GCA_900617975.1	Rodents, humans
Loa loa	Parasitic	GCA_000183805.2	Humans
Meloidogyne enterolobii	Parasitic	GCA_003693675.1	Plants
Meloidogyne graminicola	Parasitic	GCA_002778205.1	Rice
Meloidogyne hapla	Parasitic	GCA_000172435.1	Plants
Mesocestoides corti	Parasitic	GCA_900604375.1	Rodents, cats, dogs, birds
Necator americanus	Parasitic	GCA_000507365.1	Humans
Nippostrongylus brasiliensis	Parasitic	GCA_900618405.1	Rats
Oesophagostomum dentatum	Parasitic	GCA_000797555.1	Pigs
Onchocerca flexuosa	Parasitic	GCA_900618345.1	Deer
Onchocerca volvulus	Parasitic	GCA_000499405.2	Humans, blackflies
Opisthorchis felineus	Parasitic	GCA_004794785.1	Humans, freshwater fish
Opisthorchis viverrini	Parasitic	GCF_000715545.1	Humans, freshwater fish
Paragonimus westermani	Parasitic	GCA_008508345.1	Humans, snails, crustaceans
Parascaris univalens	Parasitic	GCA_002259205.1	Horses
Parastrongyloides trichosuri	Parasitic	GCA_000941615.1	Possums

Protopolystoma xenopodis	Parasitic	CAAALY010000000	Xenopus frogs
Romanomermis culicivorax	Parasitic	GCA_001039655.1	Mosquitoes
Schistocephalus solidus	Parasitic	GCA_900618435.1	Fish, birds, rodents
Schistosoma bovis	Parasitic	GCA_003958945.1	Sheep, cattle, goats
Schistosoma curassoni	Parasitic	GCA_900618015.1	Sheep, cattle, goats
Schistosoma haematobium	Parasitic	GCF_000699445.1	Snails, humans
Schistosoma japonicum	Parasitic	GCA_006368765.1	Snails, humans
Schistosoma mansoni	Parasitic	GCA_000237925.3	Snails, humans
Schistosoma margrebowiei	Parasitic	GCA_900618395.1	Snails, mammals
Schistosoma mattheei	Parasitic	GCA_900617995.1	Bovid ruminants, snails, humans
Schistosoma rodhaini	Parasitic	GCA_000951475.1	Snails, rodents
Soboliphyme baturini	Parasitic	GCA_900618415.1	Holarctic mustelids
Spirometra erinaceieuropaei	Parasitic	GCA_000951995.1	Humans, domestic animals, copepods, frogs, snakes
Steinernema carpocapsae	Parasitic	GCA_000757645.3	Insects
Steinernema feltiae	Parasitic	S_felt_v1_submitted, California Institute of Technology	Insects

Steinernema monticolum	Parasitic	S_monti_v1_submitted, California Institute of Technology	Insects
Steinernema scapterisci	Parasitic	S_scapt_v1_submitted, California Institute of Technology	Insects
Strongyloides papillosus	Parasitic	GCA_000936265.1	Ruminants, pigs, rodents
Strongyloides ratti	Parasitic	GCA_001040885.1	Rats
Strongyloides stercoralis	Parasitic	GCA_000947215.1	Humans
Strongyloides venezuelensis	Parasitic	GCA_001028725.1	Mice
Strongylus vulgaris	Parasitic	GCA_900624965.1	Horses, donkeys
Syphacia muris	Parasitic	GCA_000939275.1	Rats
Taenia asiatica	Parasitic	GCA_900618005.1	Humans, pigs
Taenia multiceps	Parasitic	GCA_001923025.3	Dogs, wolves, foxes, ruminants, humans
Taenia saginata	Parasitic	GCA_001693075.2	Cattle, humans
Taenia solium	Parasitic	Tsolium_Mexico_v1, National University of Mexico	Pigs, humans
Teladorsagia circumcincta	Parasitic	GCA_002352805.1	Sheep
Thelazia callipaeda	Parasitic	GCA_900618365.1	Mammals
Toxocara canis	Parasitic	GCA_900622545.1	Dogs, humans
Trichinella britovi	Parasitic	GCA_001447585.1	Carnivores, pigs, horses

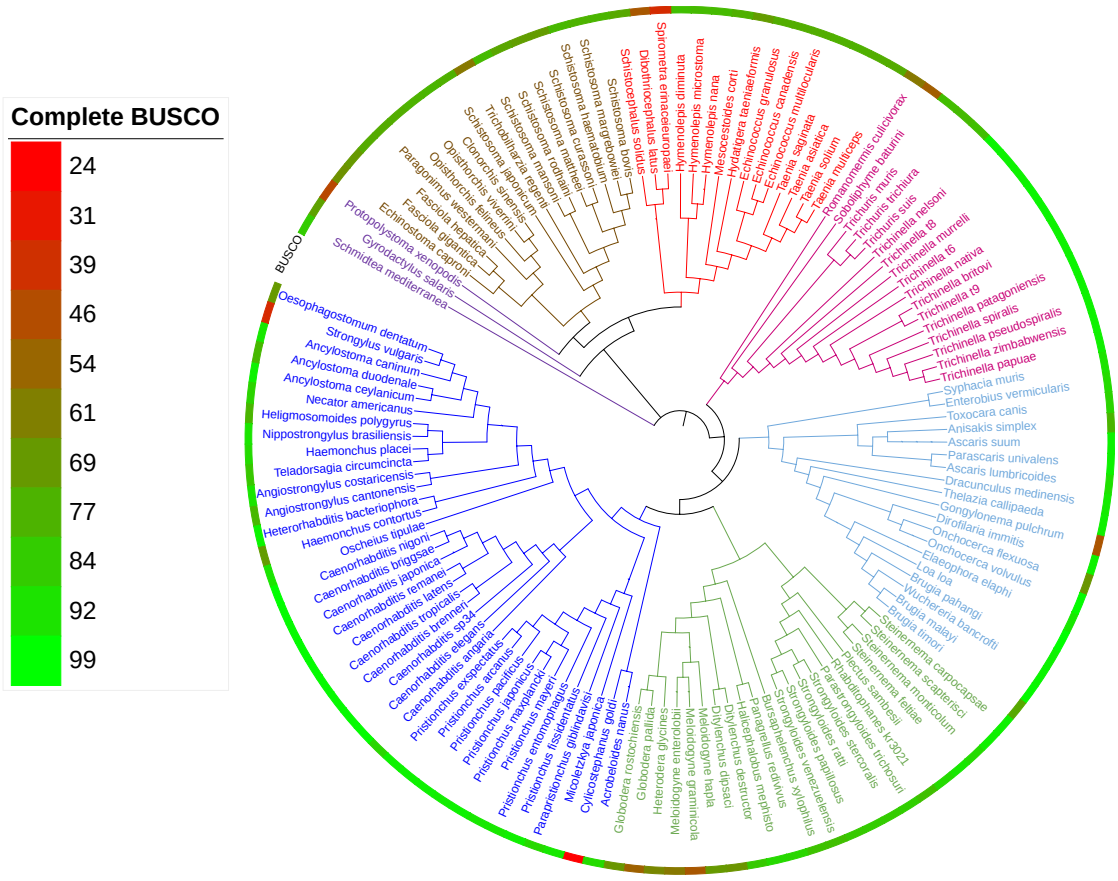
Trichinella murrelli	Parasitic	GCA_001447425.1	Bears, raccoons, foxes, bob cats, coyotes
Trichinella nativa	Parasitic	GCA_001447565.1	Mammals, birds
Trichinella nelsoni	Parasitic	GCA_001447455.1	Carnivores, scavengers
Trichinella papuae	Parasitic	GCA_001447755.1	Pigs, crocodiles
Trichinella patagoniensis	Parasitic	GCA_001447655.1	Carnivorous mammals
Trichinella pseudospiralis	Parasitic	GCA_001447675.1	Mammals, birds
Trichinella spiralis	Parasitic	GCA_001447595.1	Pigs, humans
Trichinella sp. t6	Parasitic	GCA_001447435.1	Mammals, birds
Trichinella sp. t8	Parasitic	GCA_001447745.1	Lions, hyenas
Trichinella sp. t9	Parasitic	GCA_001447505.1	Carnivores
Trichinella zimbabwensis	Parasitic	GCA_001447665.1	Crocodiles, monitor lizards
Trichobilharzia regenti	Parasitic	GCA_900618515.1	Birds, snails, humans
Trichuris muris	Parasitic	GCA_000612645.2	Mice
Trichuris suis	Parasitic	GCA_000701005.1	Pigs
Trichuris trichiura	Parasitic	GCA_000613005.1	Humans
Wuchereria bancrofti	Parasitic	GCA_000180755.1	Humans
Acrobeloides nanus	Free-living	GCA_900406225.1	

Caenorhabditis angaria	Free-living	PRJNA51225, California Institute of Technology
Caenorhabditis brenneri	Free-living	GCA_000143925.2
Caenorhabditis briggsae	Free-living	GCA_000004555.3
Caenorhabditis elegans	Free-living	GCA_000002985.3
Caenorhabditis japonica	Free-living	GCA_000147155.1
Caenorhabditis latens	Free-living	GCA_002259235.1
Caenorhabditis nigoni	Free-living	GCA_002742825.1
Caenorhabditis remanei	Free-living	GCA_000149515.1
Caenorhabditis inopinata	Free-living	GCA_003052745.1
Caenorhabditis tropicalis	Free-living	GCA_000186765.1
Halicephalobus mephisto	Free-living	SWDT00000000
Micoletzkya japonica	Free-living	GCA_900490955.1
Oscheius tipulae	Free-living	GCA_900184235.1
Panagrellus redivivus	Free-living	GCA_000341325.1
Parapristionchus giblindavisi	Free-living	GCA_900491355.1
Plectus sambesii	Free-living	GCA_002796945.1
Pristionchus arcanus	Free-living	GCA_900490705.1
Pristionchus entomophagus	Free-living	GCA_900490825.1
Pristionchus exspectatus	Free-living	GCA_900380275.1
Pristionchus fissidentatus	Free-living	GCA_900490895.1
Pristionchus japonicus	Free-living	GCA_900490845.1
Pristionchus maxplancki	Free-living	GCA_900490775.1
Pristionchus mayeri	Free-living	GCA_900490875.1
Pristionchus pacificus	Free-living	GCA_000180635.3
Rhabditophanes kr3021	Free-living	GCA_000944355.1
Schmidtea mediterranea	Free-living	GCA_002600895.1



S. Figure 1: Phylogeny of 129 species of helminth worm with branch lengths

Worm species are colour coded, with monogeneans in purple, trematodes in brown, cestodes in red, clade I nematodes in pink, clade III nematodes in light blue, clade IV nematodes in green, and clade V in dark blue. No genomic data is currently available for clade II nematodes, and they are absent from this tree.



S. Figure 2: Complete BUSCO scores for 129 species of helminth

Complete BUSCO scores for species were obtained directly from WormBase ParaSite and displayed as a heatmap. Scores ranged from 24-99%.