12-2016

Punctuated Evolution Within a Eurythermic Genus (Mesenchytraeus) of Segmented Worms: Genetic Modification of the Glacier Ice Worm F1F0 ATP Synthase

Shirley A. Lang
Rowan University

PUNCTUATED EVOLUTION WITHIN A EURYTHERMIC GENUS (*MESENCHYTRAEUS*) OF SEGMENTED WORMS: GENETIC MODIFICATION of the GLACIER ICE WORM F_1F_O ATP SYNTHASE

Shirley A. Lang, A.S., B.A.

A Dissertation submitted to the Graduate School of Biomedical Sciences, Rowan University in partial fulfillment of the requirements for the Ph.D. Degree.

Stratford, New Jersey 08084

December 2016

Table of Contents

Acknowledgements

This thesis is dedicated to my grandmother, Jeannette Harriet Hardie – a woman possessed of great natural intelligence and common sense whose lifelong regret was that the politics of gender and class in 20th century America prevented her from achieving her dream of a college education.

This journey has not been without difficulties, sacrifices and personal loss, but at no point did I wish to quit because I always had people in my corner, listening, encouraging, consoling and believing that I could do this. I would first like to thank my mentor, Dr. Dan Shain, who encouraged me to enroll in the Ph.D program, gave me the opportunity to work with a fascinating organism, and allowed me the freedom to follow my interests. He has been the "big picture" guy who knew when to point out the forest as I was counting each tree.

I would also like to thank friends and colleagues at Rutgers University – Camden – my 'home' for the past decade. My friend and collaborator, Joseph Kawash, who set up our databases, performed Blast searches, and patiently listened when I ranted, has been an invaluable asset. To Dr. Naim Saglam, who performed the morphological identification of specimens, sağ olun! Sincere thanks to my friend, Katarina Slobodenko, whom I introduced to worm hunting on a memorable car trip to Florida, for her wit, conversation and commiseration. Kudos to fellow Shain lab

member, Ralph Saunders, the ‘leech whisperer’, for his eternal optimism and good humor, even though we rarely knew what the other was talking about.

My deepest thanks I extend to my family – to my husband, Jim, you have been my ‘rock’; I could not have done this without your unwavering love, support and encouragement. I know you often felt helpless, but without your faith in me, none of this would have been possible. To my children, Shaun, Colin and Jolie, you are, and always will be, my greatest accomplishment. Despite many fewer home-cooked meals since I began this pursuit, you never complained, and willingly picked up my slack. Your belief in me is profound and humbling.

Final thanks to the Rowan community – My introduction to hands-on research was as an undergraduate in Dr. Ron Ellis’ lab as a participant in the SURE program (Summer Undergraduate Research Experience). This was such a positive experience that I committed to a future in science; thank you Dr. Ellis. To my chairman, Dr. Eric Moss, I learned so much from my time in your lab and courses you taught; thank you for your encouragement and guidance. Lastly, thank you to Rowan University -GSBS for their years of support that enabled me to pursue this degree.

“Nature uses only the longest threads to weave her patterns, so that each small piece of her fabric reveals the organization of the entire tapestry”

Richard P. Feynman (1965)

Abstract

Segmented worms (Annelida) are among the most successful animal inhabitants of extreme environments worldwide. An unusual group of *Mesenchytraeus* worms endemic to the Pacific Northwest of North America occupy geographically proximal ecozones ranging from low elevation temperate rainforests to high altitude glaciers. Along this altitudinal transect, *Mesenchytraeus* representatives from disparate habitat types were collected and subjected to deep mitochondrial and nuclear phylogenetic analyses. Evidence presented here employing modern bioinformatic analyses (i.e., maximum likelihood, Bayesian inference, multi-species coalescent) supports a *Mesenchytraeus* "explosion" in the upper Miocene (5-10 million years ago) that gave rise to ice, snow and terrestrial worms, derived from a common aquatic ancestor. Among these ecologically-disparate but genetically-close worms, those maintaining the highest intracellular ATP levels reside permanently on glacier ice (i.e., *M. solifugus*). A comparative molecular analysis of 11 core structural subunits of the F_1F_o-ATP synthase revealed extraordinary conservation across species, with a few notable exceptions. Most strikingly, the ice worm mitochondrial-encoded ATP6 (a) subunit – the "ATP throttle" known to regulate proton flux, hence ATP synthesis – encoded a highly basic, 15 amino acid carboxy-terminal extension likely to have been acquired by lateral gene transfer from an ancestral prokaryote. This insertion is supported by transcriptome raw read reconstruction and independent PCR amplifications from three geographically-distinct ice worm populations, and

represents a rare example of a mitochondrial-based gene transfer event. The position and biochemical properties of the extension domain suggest a role in ATP synthase dimerization and/or proton shuttling, both of which would predictably enhance ATP production.

Introduction

Bioenergetics and temperature

Earth's myriad life forms are heterogeneously, but not randomly, distributed. The observation that species richness declines as one moves latitudinally from the equator was noted more than 200 years ago (Forster, 1778), and a number of theories seeking to explain this asymmetry have been promulgated, chief among them, energy availability (Hutchinson, 1959; Connell and Orias, 1964). While this theory assesses energy in terms of net primary productivity (NPP) and population density, both factors are inextricably linked, via the Arrhenius equation, to temperature (Fig. 1), and the concept that energy supply may be a limiting factor in cold tolerance was proposed more than half a century ago (Luyet and Gehenio, 1940).

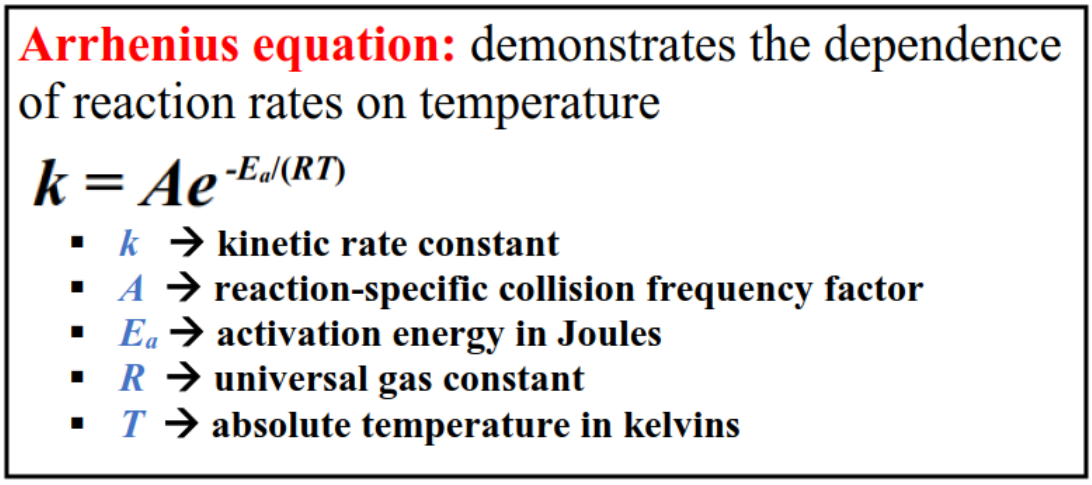


Figure 1. The Arrhenius equation

Indeed, temperature is considered the most pervasive abiotic stressor affecting living organisms in the short term, and a potent force for adaptive selection on an evolutionary time scale (Clarke, 2003; Hoffmann et al., 2003; David, 2005). Its evolutionary effects are evidenced by the observation that thermal limits define the biogeography of all extant organisms and local adaptations have allowed colonization of widespread and diverse habitats. As the majority of Earth's biosphere is permanently cold (<5°C), genetic adaptations driven by low temperature selection are found in numerous organisms comprising all three domains of life (Feller and Gerday, 2003; Somero, 2004; Siddiqui and Cavicchioli, 2006). Even humans, despite their ability to modify most climatic stressors (e.g. via clothing, shelter, migration), accumulated a variety of cold-induced genetic adaptations after migrating northward from their tropical African origins, and genes involved in energy metabolism are most reflective of these adaptations (Mishmar et al., 2003; Ruiz-Pesini, 2004; Hancock et al., 2008; Gibbons, 2010).

The Electron Transport Chain

Organisms are 'open' systems, perpetually exchanging energy with their environment. As such, they require a constant influx of free energy to maintain homeostasis against the unrelenting force of thermodynamic entropy. In aerobic eukaryotes, the majority of that energy is supplied by mitochondria, intracellular organelles of endosymbiont origin, via a process known as chemiosmotic coupling (i.e., oxidative phosphorylation, OXPHOS). Mitochondria are highly dynamic – moving, reshaping, fusing and dividing in response to local energetic requirements. In the decades since Peter Mitchell first

proposed the chemiosmotic theory, whereby cellular energy in the form of ATP (adenosine triphosphate) is generated by an electrochemical gradient across a semi-permeable membrane (Mitchell, 1961), efforts to elucidate the mechanism by which the proton gradient is coupled to ATP synthesis have been ongoing. With the axiom 'structure is function' in mind, the multi-meric molecular machines comprising the electron transport chain (ETC), which generates the gradient, and the F-type ATP synthase, which exploits the gradient to generate ATP, the individual subunits have been extracted, sequenced, visualized and reassembled. For the sake of clarity, all further descriptions of OXPHOS will be limited to aerobic eukaryotes except as noted.

The protein complexes of the ETC are embedded in the inner mitochondrial membrane where they self-assemble and function somewhat like a "bucket brigade" to pass electrons down their energy gradient in a series of redox reactions. As these reactions are exergonic, energy is released and used to pump protons (H^+) from the mitochondrial matrix to the intermembrane space (IMS), creating an electrochemical gradient termed the 'protonmotive force' (pmf; Mitchell, 1961). Complex I (*NADH dehydrogenase*), by far the largest of the ETC components (~1,000 kD), is the primary entry point into the ETC. It binds NADH and transfers 2 electrons (e^-) to ubiquinone, a lipid-soluble carrier. The energy generated by reduction of ubiquinone is used to pump 4 H^+ to the IMS. Complex II (*succinate dehydrogenase*), a secondary entry point to the ETC, accepts succinate as its substrate and, like Complex I, passes electrons to ubiquinone, but it does not pump H^+, so it does not contribute to pmf. Complex III (coQ-*cytochrome c reductase*) functions as a homodimer and accepts ubiquinone from both Complexes I and II, which it reduces, then passes the stripped e^- to the water-

soluble carrier, cytochrome *c*, and pumps 4 H^+ to the IMS. The final ETC macromolecule, Complex IV (*cytochrome c oxidase*), accepts e^- from cytochrome *c* and transfers them to molecular oxygen to form water; 2 H^+ are pumped to the IMS in the process.

The F_1F_o-ATP synthase

Subunit Structure and Function

F_1F_o-ATP synthase, also called Complex V, is not part of the ETC, but it exploits the pmf generated by the ETC to synthesize ATP, and is likewise embedded within the mitochondrial inner membrane. Responsible for producing greater than 90% of cellular energy (Boyer, 1997), ATP synthase is an astonishingly efficient multi-subunit nano-machine capable of synthesizing >100 ATP molecules/second (Matsuno-Yagi and Hatefi, 1988; Etzold et al., 1997). The structure of ATP synthase is modular in design, with separable domains, a soluble catalytic domain exposed to the matrix called F_1 (factor 1), a membrane-intrinsic domain called F_o (oligomycin-sensitive factor), and a peripheral stalk that joins the two. Assembly of the enzyme structure is highly choreographed and not completely understood, requiring a number of accessory proteins, many of which remain uncharacterized (Rühle and Leister, 2015, and references therein). In eukaryotes, assembly is further complicated by the fact that most subunits are nuclear-encoded, but two, *atp6* and *atp8* (also *atp9*(*c*) in yeast) are mitochondrial-encoded.

Bacterial ATP synthase is composed of only 8 different subunits, with the F_1 domain comprising $\alpha_3\beta_3\gamma_1\delta_1\varepsilon_1$, and F_o comprising 10-15 c-subunits, a, and b_2 (Foster

and Fillingame, 1982) (Fig. 2A). The mitochondrial ATP synthase has a few additional subunits in its core structure (plus many more accessory proteins that vary by taxon), making it a bit more complex, but the 8 bacterial subunits are all conserved (Fig. 2B).

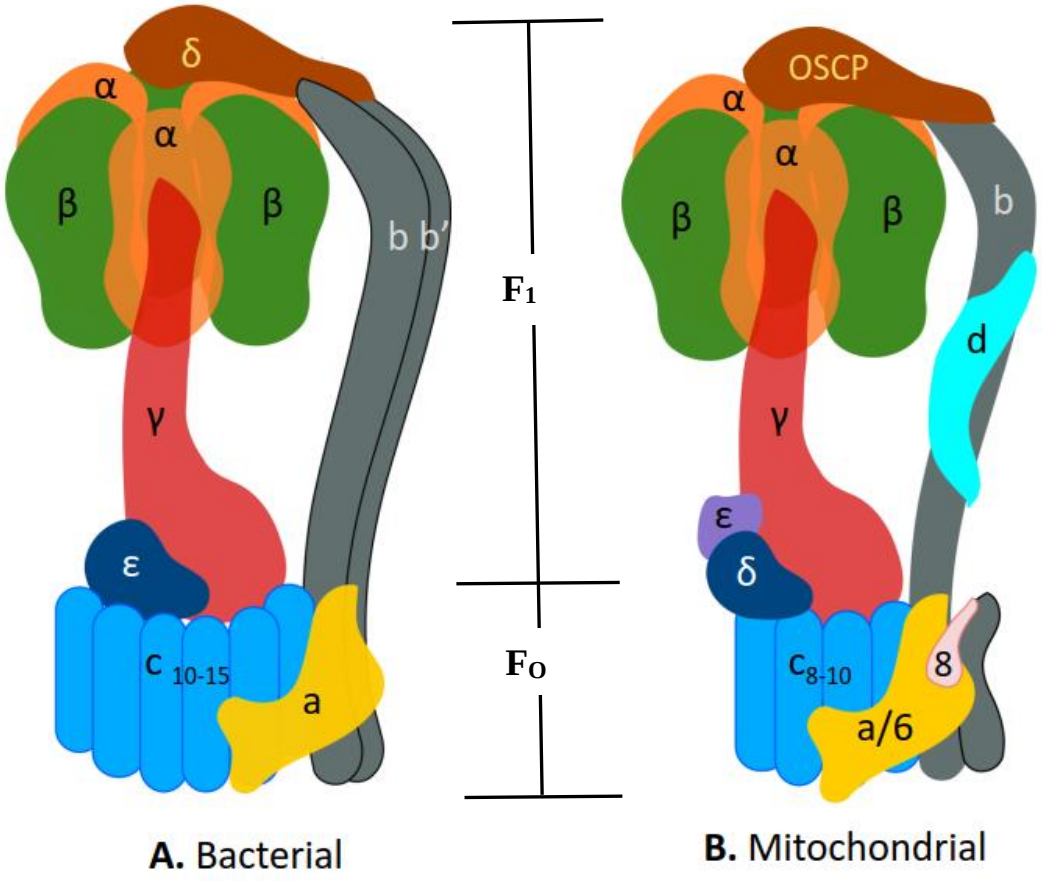


Figure 2. ATP synthase structure cartoons for (A) bacteria and (B) eukaryotic mitochondria. Core homologous subunits are shown in the same colors for both cartoons. Note that subunit delta in bacteria is homologous to mitochondrial OSCP, and bacterial epsilon is homologous to mitochondrial delta. The upper F_1 subcomplex, comprising $\alpha_3\beta_3\gamma_1$, is the core catalytic component where conformational changes in the $\alpha_3\beta_3$ ring, due to rotation of the asymmetrical γ within the ring, drive phosphorylation of ADP to produce ATP. Subunits δ(OSCP), ε(δ), and the membrane-extrinsic domain of b form the remaining F_1 elements common to both bacteria and mitochondria, while the latter has, in addition, ε, d, and F6 (not shown). In bacteria, the peripheral stator comprises a homo- or hetero-dimer of b, plus δ.

The mitochondrial stator has only a single b subunit, but additional proteins d and F6 may serve the structural role of the second b. The stator, anchored in the inner mitochondrial membrane at its base, and extending into the matrix, is attached to the $\alpha_3\beta_3$ ring to prevent it from turning in response to γ rotation. The membrane-intrinsic F_o domain is where proton translocation across the membrane generates torque, driving γ rotation and coupling proton movement to ATP synthesis. The homo-multimeric c-ring functions as a Brownian ratchet, with a strictly-conserved aspartate or glutamate alternately protonated/deprotonated as it is exposed to a strictly-conserved arginine in ATP6 (a), and the hydrophilic proton half-channel, respectively. ATP6 spans the membrane and provides the two half-channels necessary for protons to move down their gradient. ATP8 may be derived from bacterial b and plays a role in monomer assembly and stability (Hahn et al., 2016).

Confusingly, however, subunit names are not consistent between the two – subunit δ in bacteria is homologous to mitochondrial OSCP (oligomycin sensitivity-conferring protein, a likewise misleading name), and bacterial ε is homologous to mitochondrial δ. In addition, assignment of subunits into F_1 and F_o domains differs for mitochondrial ATP synthase, with F_1 elements $\alpha_3\beta_3\gamma_1\delta_1\varepsilon_1$, and F_o elements c_8, ATP6, ATP8, and two membrane-intrinsic helices of b. Several more subunits (e, f, g, DAPIT, MLQ) are considered mitochondrial F_o constituents, but these are not part of the core group considered here. The peripheral stalk components: OSCP, d, soluble domain of b, and F6 (not shown), are not assigned to F_1 or F_o (Zhou et al., 2015). Despite these differences, ATP synthase is an evolutionarily ancient machine with its core structure and function conserved from bacteria to humans (Abrahams et al., 1994; Walker, 1998).

The F_1 subcomplex, comprising $\alpha_3\beta_3\gamma_1$, is the core catalytic component where conformational changes in the $\alpha_3\beta_3$ ring, due to rotation of the asymmetrical γ within

the ring, drive phosphorylation of ADP to produce ATP (Boyer, 1997; Walker, 1998). However, only the β subunits catalyze the ADP + P_i → ATP reaction. The peripheral stator is anchored in the inner mitochondrial membrane at its base and extends into the matrix where it is attached to the $\alpha_3\beta_3$ ring. This coupling serves to prevent the $\alpha_3\beta_3$ ring from turning in response to γ rotation and wasting proton efflux. Additional F_1 constituents, δ and ε, play a critical role in clamping γ to the c-ring via extensive protein-protein interactions with both (Stock et al., 1999; Duvezin-Caubet et al., 2003). The F_1 subunits, α and β in particular, are the most highly conserved components of the ATP synthase.

The membrane-intrinsic F_o domain is where proton translocation across the membrane generates torque, driving γ rotation and coupling proton movement to ATP synthesis. This region is the only part of the ATP synthase holoenzyme that, until very recently, did not have a well-resolved structure as hydrophobic membrane proteins are recalcitrant toward crystallization (Carpenter et al., 2008). As the motor of the enzyme, without which ATP cannot be synthesized, resolution of F_o domain architecture has been a priority. Fortunately, advancements in cryo-electron microscopy (cryo-EM), enabling visualization of frozen specimens in their natural state without requirement of crystallization, have finally yielded details of the structure and arrangement of this crucial domain (Allegretti et al., 2015; Zhou et al., 2015; Hahn et al., 2016).

The F_o domain is regularly described as a 'rotary motor', and in keeping with that metaphor, ATP6 may be termed the 'ATP throttle' due to its function in regulating the H^+ ion flow (and hence ATP production). It comprises six transmembrane α-helices, arranged in a spatially and structurally conserved pattern, that span the inner

membrane; the amino-terminal (N-terminal) is exposed to the IMS and the carboxy-terminal (C-terminal) protrudes into the matrix (Fig. 3). Helix 1 is situated vertically within the membrane, farthest from the c-ring interface. Helices 2-6 assume tilted horizontal positions, with helix 2 at the matrix surface and the remainder within the membrane. Helices 3/4 and 5/6 form hairpins that follow the curve of the c-ring, with 5/6 most proximal. The arrangement of the helices provides for formation of the two slightly offset aqueous half-channels necessary for protons to move down their gradient. Helix 5 contains the strictly conserved arginine residue that interacts with a conserved aspartate/glutamate in the c subunit at their aqueous interface to ensure forward rotation of the c-ring (Allegretti et al., 2015; Zhou et al., 2015; Hahn et al., 2016).

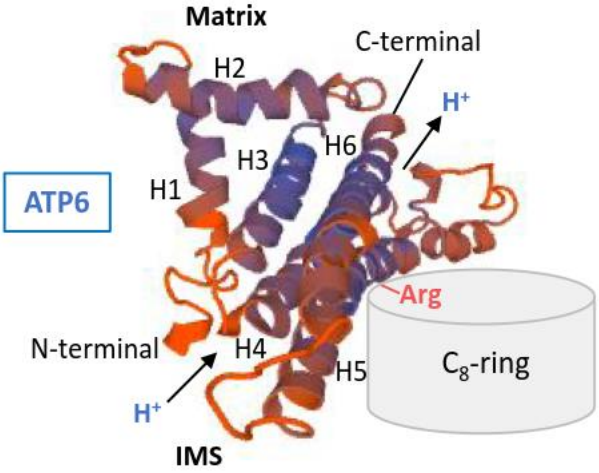


Figure 3. ATP6 structure. Bovine, yeast and alga show a conserved number and arrangement of α-helices for ATP6. The N-terminal is in the IMS and the C-terminal extends into the matrix. The skewed arrangement of the helices form the aqueous half-channels that allow proton translocation. Helix 5 curves around the c-ring and contains the strictly conserved arginine. PDB template: 5lqx.1.2.

ATP8, a very small protein (~53 aa) has its N-terminal in the IMS, one membrane-embedded α-helix, and a matrix-exposed C-terminal (Stephens et al., 2000; Hahn et al., 2016). In bovine, the C-terminal interacts with subunits b and d (Lee et al., 2015), and in yeast, the N-terminal may serve to stabilize ATP6 (Hahn et al., 2016).

The F_o homo-multimeric c-ring, the rotary motor of ATP synthase, functions as a Brownian ratchet (Zhou et al., 2015). It has a conserved aspartate (Asp) or glutamate (Glu) residue that, as it is exposed to proton flow at one half-channel, its negatively charged carboxyl group becomes protonated, allowing it to rotate through the hydrophobic core of the ring. When it reaches the second half-channel, the conserved ATP6 Arg and the aqueous environment destabilize the membrane, promoting deprotonation. Consequently, regeneration of the charged form of Asp prevents it from moving backward into the hydrophobic ring, thus assuring movement in the direction of ATP synthesis rather than hydrolysis (Allegretti et al., 2015). Since the c-ring is physically coupled to the γ rotor, each 360° rotation generates 3 molecules of ATP and the bioenergetic cost, given that all metazoans seem to have C_8-rings (Walpole et al., 2015), is 8/3 = 2.6 protons/ATP.

Dimerization

One key difference between bacterial and mitochondrial ATP synthases is that, in eukaryotes, the functional form of the enzyme is a dimer (Arnold et al., 1998; Schagger and Pfeiffer, 2000; Eubel et al., 2003). Recognition of this fact is relatively recent, yet diverse organisms spanning the eukaryotic realm, from paramecia (Muhleip et al. 2016), to alga (Dudkina et al., 2005), to bovine (Schagger and Pfeiffer, 2000), display these duplex structures. In all species, the individual monomers have their stators at the interface and are tethered through protein-protein interactions in their F_o sectors (Fig. 4).

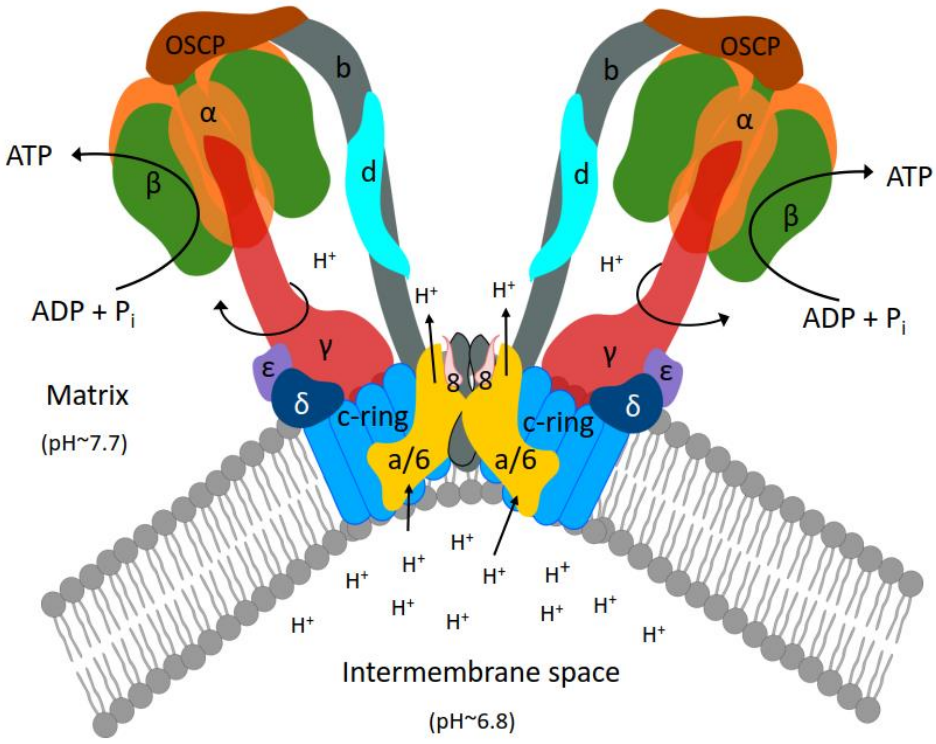


Figure 4. Mitochondrial ATP synthase dimer and subunit interactions. Dimerized ATP synthase monomers seem to be the primary functional form in eukaryotes with ATP6 (a), ATP8, and the membrane-embedded domain of the peripheral stator (b), playing essential roles. Dimerization forces membrane curvature and is thought to drive formation of characteristic mitochondrial cristae (Davies et al. 2012). A variety of additional proteins have been identified associating within the dimer interface region, but many are species-specific, suggesting ATP synthase dimerization as an evolutionary driver.

The particular subunits involved are largely unknown, though accessory proteins, e, f, g and DAPIT seem necessary, and ATP6, ATP8 and the b membrane-helices are also considered to have a role. In metazoans, the dimers all form characteristic stator-to-stator V-shapes, but in alveolates such as *Paramecia* and *Tetrahymena* the dimers are interdigitated U-shapes (Nina et al., 2010; Muhleip et al., 2016). Interestingly, metazoan mitochondria have cristae with deeply invaginated

folds and acute angle edges, while alveolates have tubular cristae. Furthermore, in both groups, dimers self-assemble into rows along the sharpest curves of the cristae (Fig. 5).

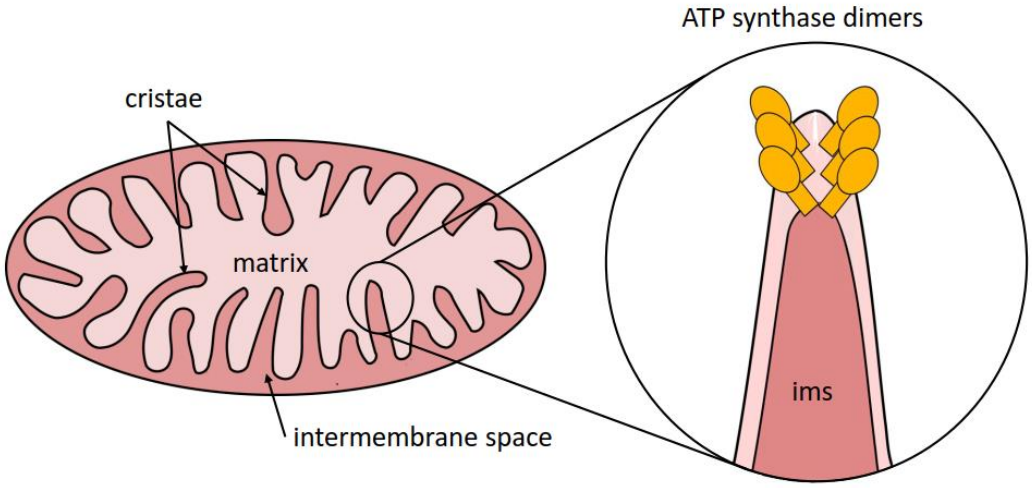


Figure 5. ATP synthase dimer arrangement and role in cristae formation. The mitochondrial inner membrane is a dynamic and highly folded structure in which the components of the electron transport chain (ETC) are embedded. These folds, known as 'cristae', bear rows of ATP synthase dimers at their most highly-curved edges and evidence suggests that dimerization itself induces cristae formation and self-assembly (Davies et al., 2012).

These observations, plus experimental studies of yeast knockouts lacking genes necessary for dimerization, suggest that dimer formation, in and of itself, drives formation of cristae (Davies et al., 2012), though there is some disagreement regarding their primacy (Cogliati, 2016).

Reasons for the ubiquity of ATP synthase dimers in eukaryotes are still unclear, but certainly dimerization must confer some significant biological advantage for it to have arisen and/or been retained in such a diverse and distantly-related group of organisms. Undisputedly, highly-folded cristae increase the surface area of the inner

membrane providing for a greater number of OXPHOS complexes, and the idea that deep invaginations may trap and concentrate protons in the vicinity of ATP synthase is also well regarded. Thus, if dimerization is responsible for cristae formation, these factors would likely be sufficiently advantageous for this trait to become fixed. Another theory is that dimers stabilize the enzyme. The opposing arrangement of the individual monomers in the dimer means that their rotors are turning in opposite directions, a situation that would minimize membrane stress resulting from the torque. Another argument for stabilization is that accessory proteins implicated in dimer formation/stabilization vary greatly in number and diversity from one taxon to another, indicative of species-specific adaptations to their particular niche (Chaban et al., 2014).

Adaptive and Pathological Mutations

In metazoans, *atp6* and *atp8* are the only two mitochondrial-encoded ATP synthase subunits and, unsurprisingly given the higher mutation rate of mtDNA (versus nuclear DNA), they display the greatest sequence divergence among the core structural components (Tzen et al., 2001; Mishmar et al., 2003). The majority of human diseases involving ATP synthase deficiencies are attributable to *atp6* mutations, primarily single nucleotide polymorphisms (SNPs) resulting in substitution of a single amino acid (Holt et al., 1990). Among the best described conditions are NARP (neuropathy, ataxia, retinitis pigmentosa), MILS (maternally-inherited Leigh syndrome) (Tatuch and Robinson, 1993; Trounce et al., 1994, Morava et al., 2006) and FBSN (familial bilateral striatal necrosis) (Dionisis-Vici et al., 1998). In addition, *atp6* mutations are currently implicated in autism, diabetes, multiple sclerosis and blindness. At present, dozens of

atp6 SNPs have been clinically described, yet only three for *ATP8*, generally involving cardiomyopathies (Jonckheere et al., 2008; Ware et al., 2009), one for ε, causing peripheral neuropathy (Mayr et al., 2010), and two for α, resulting in loss of mtDNA (Lieber et al., 2013) or neonatal encephalopathy (Jonckheere et al., 2013).

Despite these afflictions, ATP6 sequence variation plays a significant role in ecological adaptation, from bacteria to humans. For example, extreme alkaliphilic *Bacillus* bacteria have a G>K substitution, just downstream of the strictly conserved Arg, that is required for ATP synthesis at pH >10.5 (Wang et al., 2004). In humans, mitochondrial DNA haplotypes have strong geographic associations that cannot be accounted for by genetic drift. Most noticeably, haplotypes from arctic regions have considerably more ATP6 substitutions than equatorial or temperate haplotypes (Mishmar et al., 2003). In general, these substitutions cause partial decoupling of proton passage and ATP synthesis resulting in greater heat output.

Cold-adapted organisms

Interest in cold-adapted organisms (psychrophiles) has grown, and the genetic strategies enabling them to inhabit icy, inhospitable environments are both fundamentally and practically intriguing. As biological systems are understood to function in accordance with the laws of physics and chemistry, at temperatures approaching 0°C, phase changes and drastic reductions in molecular motion and enzyme kinetics exert profoundly negative effects on the structural integrity and metabolic capabilities of cells. While early research on psychrophiles focused on novel proteins, such as antifreeze (glyco) proteins [AF(G)Ps], first isolated and described

more than 40 years ago (Devries et al., 1970), and unusual adaptations, such as hemoglobin loss in Antarctic icefishes (family Channichthyidae; Ruud, 1954), at present the spotlight is on common metabolic enzymes. Their unique attributes – high specific activity at low temperatures and mild thermic lability – make them attractive targets for developing applications translatable to agriculture, industry, and medicine (Gerday et al., 2000). This line of research has benefitted greatly from recent advancements in sequencing technology, which have enabled the sequencing of full genomes from a wide variety of ecological models. These investigations have provided insight into genetic pathways most responsive to environmental stress and, regardless of the specific environment, energy metabolism is always a significant factor (e.g. Jung et al., 2003; Baena-González, 2010; Lisarre et al., 2010; Yampolsky et al., 2014; Malecki et al., 2016). At present, prokaryotes dominate the cold-adaptation research field, but they are not ideal models for understanding metazoan mitochondrial dysfunction due to significant differences in composition and regulation of OXPHOS components. In this regard, development of new ecological models for cold-temperature adaptation would be informative.

Experimental system

Among Animalia, representatives of the phylum Annelida occupy some of the most diverse habitats on Earth, ranging from geothermal black smokers on the ocean floor to glacier ice (Shain, 2009). At the latter end of this spectrum, oligochaetes of genus *Mesenchytraeus* are exceptionally well represented along coastal regions of the Pacific Northwest, from northern California to central Alaska. This geographically complex

region is characterized by vast temperate rainforests accommodating many of the world's largest trees, snow-fed waterways, and rugged mountains (many volcanically active) that harbor lower elevation transient snowfields and higher altitude glacier ice (Littell et al., 2009). These maritime peaks are altitudinally stratified into zones, each characterized by distinct thermal maxima and minima, vegetation, and faunal communities (Taylor, 1922). Within this geographic ecozone, *Mesenchytraeus* species have evolved to occupy remarkably diverse environmental niches.

Comprising >700 species and 33 genera, Enchytraeidae are predominantly small, inconspicuous soil-dwelling worms with a worldwide distribution (Schmelz and Collado, 2015; Erséus et al., 2010). Yet *Mesenchytraeu*s, one of its most speciose genera, is distinguished by extraordinary intra-generic variation, not only in habitat (aquatic, terrestrial, snow, glaciers), but also in size (4 to >60 mm long), and coloration (transparent, white, yellow, red, tawny, brown, gray, black, uniform, variegated) (Eisen, 1904; Welch, 1916; Welch, 1919; Rota and Brinkhurst, 2000; Schmelz and Collado, 2010). *Mesenchytraeus solifugus* (Emery, 1898), the glacier-obligate ice worm, was the first enchytraeid described from the Pacific Northwest, and over the next two decades numerous mesenchytraeid species/subspecies, occupying the full spectrum of ecozones, were formally described (Eisen, 1904; Welch, 1916; Welch, 1919). Mesenchytraeid dominance, in terms of species richness and ecological specialization, within the Beringian coastal region suggests this as its point of origin (Piper et al., 1982; Rota and Brinkhurst, 2000). Although other Enchytraeidae genera were likewise discovered in the Pacific Northwest, none are as diverse or extreme in habitat choice.

Despite their abundance and unusual phenotypic plasticity, mesenchytraeids of this region received little attention in the ensuing 50 years until Tynen (1970a, b) and Goodman (1971) undertook investigations into the distribution, habitat, behavior and physiology of *M. solifugus*. More recently, we have demonstrated that glacier ice worms, *M. solifugus* and subspecies *M. solifugus rainierensis*, display atypically high intracellular ATP levels that paradoxically increase as temperatures decline (Napolitano and Shain, 2004; Napolitano et al., 2004). In combination with experimental manipulations in *E. coli* (Morrison and Shain, 2008; Parry and Shain, 2011), this physiological capacity appears to be a critical component, and bioenergetic signature, of cold adaptation, but identification of specific adaptations has proven elusive.

That *M. solifugus* was the focus of virtually all recent investigations can be ascribed to their rare and compelling status as glacier-obligate macrofauna. For most animals, extended exposure to temperatures ~0°C results in cellular dysfunction and metabolic failure (Thauer, 1962; Kruuv et al., 1983; Zachariassen, 1991). Thus, the ability of ice worms to overcome this physiological barrier garnered intense interest. Unfortunately, its congeners have been virtually ignored and little is known about lifestyles and genealogical relationships among these worms. Modern biology recognizes that comparative trait analyses, informed by a credible phylogenetic framework, provides the best methodology for identification of ecologically adaptive traits (Felsenstein, 1985; Martins and Hansen, 1997; Maddison et al., 2007; Rezende and Diniz-Filho, 2012).

Rationale

In nearly all eukaryotic cells, intracellular organelles called mitochondria produce the majority of cellular adenosine triphosphate (ATP), the most pervasive high energy carrier molecule, and a critical metabolic cofactor. Previous work in our lab has firmly established that diverse psychrophilic (cold-adapted) organisms maintain relatively high ATP levels (compared to non-psychrophilic counterparts), and we have proposed that increased energy availability acts as a compensatory mechanism against inherent reductions in molecular motion and enzyme kinetics at low temperatures.

To gain insight into evolutionary aspects of cold temperature adaptation, I chose to examine a group of closely-related segmented worms (i.e., annelids) in the genus *Mesenchytraeus*, which collectively occupy a continuum of habitats ranging from temperate soil to glacier ice within a small geographic range (i.e., Pacific Northwest). One species of particular interest, *Mesenchytraeus solifugus*, the glacier ice worm, is among only a few described metazoan psychrophiles, and thus represent a valuable data point for comparative studies ranging from the mechanisms of cold tolerance in eukaryotic cells (e.g., hibernation, diapause) to the genetic adaptations of extremophiles in general.

Given the previously established unusual energy profile found in ice worms (i.e., high ATP), I chose to perform comparative transcriptome analyses, focusing on bioenergetic genes, in an effort to identify genetic changes in fundamental cellular machinery associated with energy metabolism. The goals of this investigation were to (1) establish the genetic relationships among a phenotypically diverse selection of

annelid worms (genus *Mesenchytraeus*) endemic to the Pacific Northwest, (2) use comparative bioinformatic analyses to investigate possible cold-temperature adaptations in F-type ATP synthase subunits, and (3) encourage use of these species, the glacier-obligate ice worm, *M. solifugus*, in particular, as a new ecological model for investigating cold-adaptive strategies.

Materials and methods

Specimens

Mesenchytraeus specimens were collected from four distinct habitat types along the Pacific northwestern coast of North America between 2011-2014. These comprised 1) cold, outflow rivers from coastal mountains (*M. pedatus;* aquatic), 2) low elevation temperate rainforests (*M. antaeus;* terrestrial), 3) transient snowfields shaded by coniferous stands (*M. gelidus*, *M. hydrius;* snow), and 4) coastal, maritime glaciers (*M. solifugus;* ice) (Table 1, Fig. 6). Additionally, two confamilials, *Enchytraeus albidus* and *E. crypticus*, both well-studied European species, were included as outgroups.

Table 1. Description and collection sites of Enchytraeidae species.

Species	Overt phenotype (color, length x diameter)	Collection location & date	Habitat Description (at time of collection)
Mesenchytraeus solifugus Emery 1898	Dark brown to black 10-20 mm x 0.5-0.7 mm	60.762003° N, 148.846545° W July 2012	0°C (constant), snow overlying glacier ice; alt: 122 m
Mesenchytraeus pedatus Eisen 1904	White 10-20 mm x 0.3-0.5 mm	40.5937222° N, 122.39919° W June 2011	8-14°C, cold mountain streams/rivers; alt: 148 m
Mesenchytraeus gelidus Welch 1916	Dark red-brown to black 20-38 mm x 1-2 mm	47.288126° N, 121.263348° W June 2014	0-2°C, transient snow shaded by coniferous trees, alt: 1353 m
Mesenchytraeus hydrius Welch 1919	White 15-35 mm x 0.5-0.9 mm	46.746679° N, 121.790266° W June 2014	0-2°C, transient snow shaded by coniferous trees, alt: 1332
Mesenchytraeus antaeus Rota & Brinkhurst 2000	White to yellowish 30-60 mm x 1.8-2.5 mm	48.874057° N, 124.701034° W May 2014	10-15°C, temperate rainforest, alt: 149 m
Enchytraeus crypticus Westheide & Graefe 1992	White 3-12 mm x 0.2-0.4 mm	Commercially purchased lab stock	22°C, moist, acidic, peat-enriched soil; alt: 5 m

Specimens were express-shipped live to Rutgers University, and maintained in plastic containers according to their endogenous habitat/temperature (e.g., soil, refrigerated, etc.) as described (Napolitano et al., 2004). *Enchytraeus crypticus* were from laboratory maintained stock, and *E. albidus* and *Mesenchytraeus solifugus* (Davidson glacier) were ethanol-preserved samples used for DNA analysis.

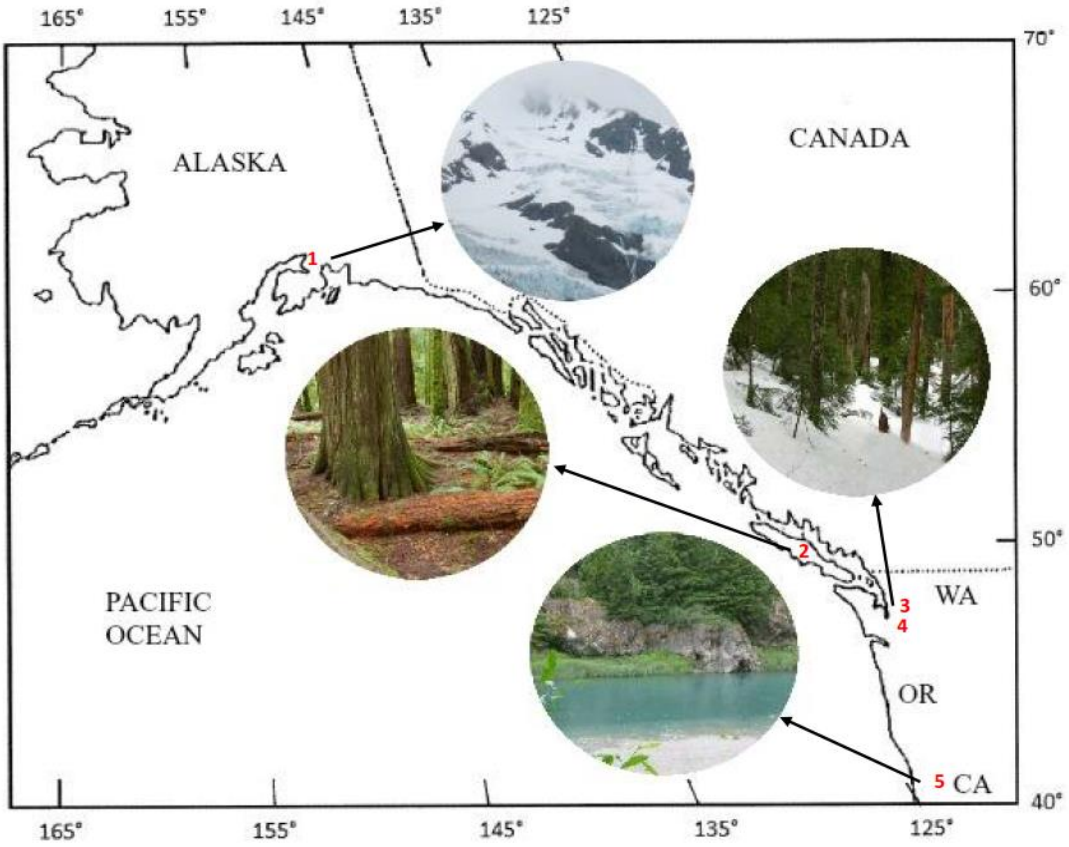


Figure 6. Mesenchytraeid collection sites and habitats. Site 1: Chugach Mountains, Alaska; *Mesenchytraeus solifugus*, glacier-obligate. Site 2: Vancouver Island, BC; *M. antaeus*, temperate forests. Sites 3, 4: Cascade Mountains, WA; *M. gelidus* (3) and *M. hydrius* (4), transient snow beneath conifers. Site 5: Sacramento River, CA; *M. pedatus*, aquatic.

Species were identified by comparison to molecular data, when available, and/or morphological examination. Enchytraeidae, most of which are <15 mm long and transparent or white, are primarily identified by internal structures using live specimens examined with a light microscope (Schmelz and Collado, 2010). However, this technique is often not practical for mesenchytraeids due to their size and/or pigmentation. Thus, morphological identifications were performed on freshly

sacrificed specimens prepared as follows. Worms were placed in 70% ethanol for 24 hours, then dehydrated through an ethanol series: 80% (10 min), 90% (10min), 100% (2 x 10 min). Dehydrated specimens were transferred to a 1:2 ratio of benzyl alcohol:benzyl benzoate (Sigma-Aldrich) and incubated until transparent (~20 min). Internal structures were viewed with a stereoscope and sketched and photographed for comparison to published descriptions.

DNA processing and PCR

Single-worm genomic DNA extractions were performed using an E.Z.N.A. Tissue DNA Isolation Kit (Omega Bio-tek, Inc.) following the tissue protocol. With the exception of *E. albidus*, extractions were from live worms. The use of single worms was primarily to reduce sequencing uncertainty, a concern only for *M. pedatus* with 0-15% variability at the *cytochrome oxidase 1* (*cox1*) locus (N=20; data not shown). Admittedly, cryptic speciation is possible within this population, however, all haplotypes proved equidistant to the considered congeners. For *E. albidus* and *M. hydrius*, only a single worm was available; the remaining species were effectively clonal.

Nuclear *28S* ribosomal RNA was PCR-amplified using annelid-specific primers designed for this study: 28sC1-F (**5′-ACCCGYTGAAYTTAAGCATAT-3′**) / 28sC4-R (**5′-TTCGATTRGTCTTTCGCCCCT-3′**) amplifies a 5'-end fragment of ~1180 bp. For some species, additional internal primers were necessary: 28sC2-F (**5′-CAAGTACCGTGAGGGAAAGTTG-3′**) / 28sC3-R (**5′-CCGTGTTTCAAGACGGGTCG-3′**). The final *28S* dataset comprised 600 bp encompassing the D2 variable region. For

mitochondrial genes, primer pairs were designed to amplify multiple overlapping mitochondrial genome (mt-genome) fragments, 2-4 Kb in length, based on gene size and order as found in *Lumbricus terrestris* (Fig. 7A, B) (Boore and Browne, 1995). Additional species-specific primers were designed as needed based on sequenced PCR products (primer walking) to fill genome gaps (see Appendix Table S1 for mitochondrial primers). The primers for amplification of full-length *atp6* are shown in the alignment in figure 15.

PCR reactions contained 34.5 µl water, 200 µM dNTPs, 600 nM primers, 10 µl of 5X Phusion HF buffer, 1 U Phusion Hot Start II DNA Polymerase (Thermo Scientific), and 50-200 ng total genomic DNA. Reaction conditions were: denaturation at 98°C for 30 seconds, followed by 35 cycles of 98°C for 10 s, 47-60°C for 20 s, and 72°C for 0.5-2 min; final extensions were 5 min at 72°C. Single-band PCR products were purified using DNA Clean and Concentrator (Zymo Research); multi-band products were electrophoresed on 0.8% agarose gels and desired bands excised (See Fig. 7C). Gel-excised DNA was purified using a Zymoclean Gel DNA Recovery Kit (Zymo Research) according to the manufacturer's protocol. Fragments were cloned into pMiniT vectors (New England Biolabs), transformed into NEB 10-beta chemically-competent *E. coli*, and 3-5 clones/fragment/species were selected for purification using Zyppy Plasmid Miniprep (Zymo Research). Sanger DNA sequencing was performed by GENEWIZ, Inc. (South Plainfield, NJ). Initial sequencing was performed using vector-specific forward and reverse primers. Resequencing was performed as necessary using original and internal primers, to ensure recovery of overlapping sequences. For regions spanning multiple consecutive tRNAS, KAPA2G

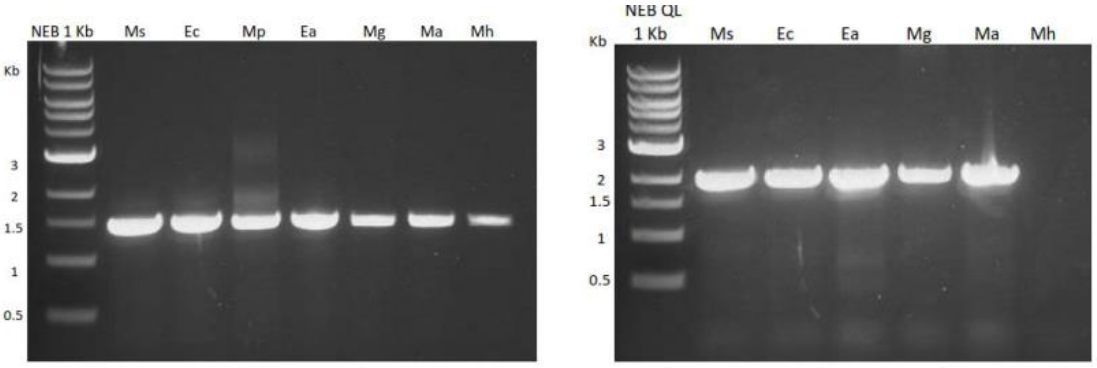
Robust polymerase (KAPA Biosystems) was used according to the manufacturer's protocol.

A.

B.

Set	Span	Forward Primer	Degen	Tm (°C)	Reverse Primer	Degen	Tm (°C)	Anneal (°C)	Extend (sec)	Expect (bp)
A	CO1 → CO2	Cox5.1	nd	60.1	Cox2-R1	8	63	58/55	30	1230
B.1	CO2 → CO3	Cox2-F2	8	60.1	Cox3-R1	24	70	55/53	30	1360
C	CO3 → CytB	Cox3-F2	24	71.6	CytB-R2	72	63.6	58/55	30	1235
CB1.1	CytB → CytB	CytB-F1	18	53.3	CytB-R1	18	63.2	47	15	667
D	Nad5 → Nad5	Nad5-F1	48	50.5	Nad5-R1	6	58.3	47	15	600
E	Nad5 → Nad4	Nad5-F2	48	67.1	Nad4-R1	12	60.9	58/55	60	2842
F	Nad4 → 12S	Nad4-F1	24	55.2	12S-B1	nd	71.3	50	40	1669
G	12S → 16S	12S-A1	nd	53.6	16S-R2	nd	61.7	53/50	30	1156
H	16S → Nad1	16S-F1	nd	61	Nad1-R1	12	53.6	53/50	30	1469
I	Nad1 → Nad3	Nad1-F1	8	56.7	Nad3-R1	96	59.6	53/50	30	1370
J	Nad3 → Nad2	Nad3-F1	8	62.4	Nad2-R1	6	52.6	50	20	851
I/J	Nad1 → Nad2	Nad1-F1	8	56.7	Nad2-R2	6	52.6	52/48	20/10	1892
K	Nad2 → CO1	Nad2-F1	32	59.7	IWHCO	nd	60.9	53/50	30	1543
L	Nad3 → COI	Nad3-F1	8	62.4	IWHCO	nd	60.9	53/50	45	2024
M	16S → Nad3	16S-F1	nd	61	Nad3-R1	96	59.6	53/50	45	2201
P	CO1→ CO3	Cox5.1_in	nd	57.9	Cox3-R1	16	57.7	58/55	60	2665

C.



Set **B.1**: 0.8% gel; 70V/75 min; 5 µl/sample

Set **I/J**: 0.8% gel; 70V/75 min; 5 µl/sample

Figure 7. PCR amplification of mitochondrial genes. (A) Linear schematic showing annelid mitochondrial gene order. Gray bars indicate location of tRNAs. (B) Major primer sets used for amplifications with primer characteristics, PCR parameters and expected product sizes. (C) Sample gel images for two primer set amplifications, left (B.1), right(I/J).

Mitochondrial sequence alignment and annotation

DNA sequence chromatograms were evaluated and trimmed using Chromas 2.4.3 (Technelysium). Overlapping sequence fragments were aligned in MEGA version 6 (Tamura et al., 2013) and assembled into consensus sequence contigs using Mesquite v.3.04 (Maddison and Maddison, 2015). Protein coding genes (PCGs) and ribosomal RNAs (rRNA) were identified using NCBI's BLASTx and BLASTn algorithms. Individual genome contigs were uploaded to DOGMA (http://dogma.ccbb.utexas.edu) for identification of tRNAs and whole genome annotation using default parameters for invertebrate mitochondrial genomes (tRNA data is not discussed in this manuscript).

Transcriptome generation

Transcriptomes were generated at different times and at different facilities, thus preparation details varied and are described separately. For *M. solifugus, M. pedatus* and *E. crypticus*, pooled worms (10-12) were immersed in RNA*later* (Ambion) before transport to the Halanych Lab at Auburn University where total RNA was extracted using TRIzol (Invitrogen) according to the manufacturer's protocol, and purified using the RNeasy kit (Qiagen) with on-column DNase digestion. The SMART cDNA library construction kit (Clontech) was used for first-strand cDNA synthesis, followed by synthesis and amplification of double-stranded cDNA using the Advantage 2 PCR Kit (Clontech). Illumina sequencing was conducted at the Genomics Services Laboratory at Hudson Alpha Institute for Biotechnology (Huntsville, AL). For *M. pedatus* and *M. solifugus*, sequencing runs were performed on an Illumina HiSeq 2000 platform with 2 x 100 bp paired-end reads using the TruSeq v3 (Illumina) protocol. *Enchytraeus*

crypticus sequencing employed the Illumina HiSeq 2500 platform with 2 x 150 bp paired-end reads, also using the TruSeq v3 (Illumina) protocol. Paired-end reads were demultiplexed and *de novo* assembly was performed using the Trinity software package (Grabherr et al., 2011), version October 2012 (*M. pedatus, M. solifugus*) or February 2013 (*E. crypticus*) with default settings. See Waits et al. (2016) for additional details.

For *M. antaeus, M. gelidus,* and *M. hydrius,* total RNA was extracted from single worms using TRIzol Plus RNA Purification Kit (Invitrogen) according to the manufacturer's protocol. RNA samples were express-shipped on dry ice to the Giribet lab at Harvard University's Department of Organismic and Evolutionary Biology and purified using the Dynabeads mRNA Purification Kit (Invitrogen) following manufacturer's instructions. cDNA libraries were constructed in the Apollo 324 automated system using the PrepX mRNA kit (IntegenX). Further details about the protocols can be found in Fernández et al. (2014). Samples were run using the Illumina HiSeq 2500 platform with 2 x 150 paired-end reads at the FAS Center for Systems Biology at Harvard University. Demultiplexed sequencing results, in FASTQ format, were retrieved, each sample being quality-filtered according to a threshold average quality score of 30 based on a Phred scale and adaptor trimmed using Trimgalore! 0.3.3 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore). Ribosomal RNA and mitochondrial DNA were filtered out using Bowtie v. 1.0.0 (Langmead et al., 2009). Strand specific de novo assemblies were done individually in Trinity (Haas et al., 2013) using paired read files, a strand specificity flag and path reinforcement distance enforced to 75. Access to the *E. albidus* transcriptome was provided by Christer Erséus (U. of Gothenburg, Sweden).

Gene extraction and verification from transcriptomes

Multiple loci with varying levels of conservancy were chosen for the phylogenetic analyses in an effort to effectively infer the most plausible species tree. Well-conserved nuclear housekeeping genes, *actin*, *α-tubulin*, *EF-1α*, *histone H3*, and *GAPDH* were extracted, as well as core subunits of ATP synthase: *alpha (α), beta (β), gamma (γ), delta (δ), epsilon (ε), b, c, d and OSCP*. The latter genes, though nuclear-encoded, function within the mitochondria in conjunction with mitochondrial-encoded genes to produce the bulk of cellular energy. The utmost importance of this metabolic process imposes mito-nuclear coevolutionary constraints that are linked to speciation (Schmidt et al., 2001; Gershoni et al., 2009; Bar-Yaacov et al., 2012; Hill, 2016), thus making them good candidates for species delimitation analysis. In addition, all 13 mitochondrial PCGs and *12S* and *16S* rRNAs were extracted for comparison to PCR-amplified sequences (see Table 2 for gene details).

Local BLAST searches were performed using NCBI's BLAST 2.2.26 utilizing the aforementioned Trinity assemblies as databases. Queries for nuclear-encoded genes used annelid gene sequence when available, but otherwise queries were made using mollusc, lancelet (*Branchiostoma floridae*), and acorn worm (*Saccoglossus kowalevskii*) sequence. Extracted transcripts were subjected to GenBank searches to verify gene identity, translated to ensure that they were not pseudogenes, and aligned to each other to assess relatedness. Extracts with questionable or incomplete sequence were re-BLAST using genus-specific queries and/or subjected to raw-read mapping to rule out assembly error. At present, no non-hirudinid oligochaete genome is publicly available to facilitate gene identification.

Phylogenetic analyses

Individual gene sequences were aligned using Muscle (Edgar, 2004) under default parameters, as implemented in MEGA version 6 (Tamura et al., 2013), with protein coding genes (PCGs) aligned by codons. Statistical analyses to assess phylogenetic informativeness and divergence were also performed in MEGA (Table 2, Appendix table S2). Gene trees were inferred using Maximum-likelihood (ML) and Bayesian inference (BI) for each gene. Prior to ML analysis, jModelTest 2.1.7 (Darriba et al., 2012, Guindon and Gascuel, 2003) was used to select the best-fitting evolutionary model for each locus according to the Akaike Information criterion (AIC; Akaike, 1974). ML analyses were conducted via the PHYLIP-like interface of PhyML v3.1 (Guindon et al., 2010) using the recommended best-fit substitution model. The subtree pruning and regrafting (SPR) option was used to more thoroughly explore the space of tree topologies and five random starting trees were generated to avoid local maxima. Branch support was estimated using the non-parametric Shimodaira–Hasegawa-like approximate likelihood ratio test (SH-aLRT, Guindon et al., 2010), as this method provides superior performance, relative to bootstrapping, for datasets with short branches (Guindon et al., 2010) and topological conflict (Simmons and Norton, 2014). Analyses were run with, and without, outgroup designation to test their suitability.

Bayesian analyses were performed in MrBayes v.3.2 (Ronquist et al., 2012) using Bayes-block nexus files created in Mesquite v.3.04 (Maddison and Maddison, 2015), with default parameters modified as described below. Each PCG dataset was analyzed partitioned (by codon position) and unpartitioned; non-coding genes were unpartitioned. Model settings for each gene were: nst = mixed, rates = gamma. The

'mixed' model allows the analysis to integrate across the full general time-reversible (GTR, Tavaré 1986) model space and avoids specifying a particular, possibly inappropriate, evolutionary model of nucleotide substitution (Huelsenbeck et al., 2004). The 'gamma' parameter models per site mutation rate heterogeneity with the assumption that it is drawn from a random statistical distribution (Yang, 1994). For partitioned analyses, partition parameters were unlinked and the 'ratepr' parameter of the 'prset' command was set to variable to allow rates across partitions to differ. All analyses were run for ≥ 1,000,000 generations (x2) with a relative burn-in fraction of 25%, sample frequency = 1000 and diagnostic frequency = 1000. All other parameters were default values. The 'sum parameters' (sump) command was used to assess run convergence by examining the generation versus log likelihood plot, average deviation of split frequencies (<0.01), the Potential Scale Reduction Factor (PSRF, ~1) and the Effective Sample Size (ESS, >200) (See Appendix Figure 1).

Phylogenetic reconstruction was further explored using various combinations of genes concatenated into supermatrices using Mesquite v.3.04 (Maddison and Maddison 2015). Specifically, five combined-gene datasets were examined: (1) HK6 = nuclear housekeeping genes (*actin, α-tubulin, EF-1α, histone H3, GAPDH, 28S* rRNA), (2) nuASU = nuclear-encoded ATP synthase subunits (*alpha, beta, gamma, delta, b, c, d, OSCP*), (3) MtPCGs = mitochondrial protein-coding genes (*cox 1-3, cytb, nad 1-6, nad 4L, atp6, atp8*), (4) Mt13+RNAs = 13 MtPCGs + *12S-(tRNA-Val)-16S*, and (5) Mt4 = mitochondrial subset (*cox1, cytb, 12S, 16S*) (Table 3). These were subjected to ML and BI analyses as described above with the following changes. For ML analyses, the designated model was GTR+I+G as this was the most commonly recommended for

the individual genes, and PhyML does not currently support partitioning. Prior to Bayesian analysis, PartitionFinder v1.1.0 (Lanfear et al., 2012) was used to find the best-fit partitioning scheme for each dataset under the following criteria: branch lengths = linked, models = mrbayes, model selection = BIC, search = greedy. This scheme was then incorporated into the Bayes blocks and analyses proceeded as described above (See appendix Table S3 for partitioning schemes).

Table 2. Combined-gene datasets and analyses

Dataset	Genes	Size (bp)	Analysis	Framework
HK6	*actin, α-tubulin, EF-1α, GAPDH, histone3, 28S rRNA*	5,916	Phylogenetic inference	ML/BI
Mt4	*cox1, cytb, 12S rRNA, 16S rRNA*	4,600	Phylogenetic inference	ML/BI
MtPCGs	*cox1-3, cytb, nad1-6 + 4L, atp6, atp8*	11,107	Phylogenetic inference	ML/BI
Mt13+RNAs	*cox1-3, cytb, nad1-6 + 4L, atp6, atp8, 12S, tRNA-V, 16S*	13,381	Phylogenetic inference	ML/BI
nuASU	*alpha, beta, gamma, delta, b, c, d, OSCP*	6,894	Phylogenetic inference	ML/BI
Coal_set1	*cox1-3, cytb, nad1-6, alpha, beta, b, actin, α-tubulin, EF-1α, GAPDH, 28S*	18,429	Multi-species coalescent	BI
Coal_set2	*cytb, nad2/4/5, 12S, alpha, gamma, b, c, delta, actin, EF-1α, α-tubulin, histone3, 28S*	14,232	Multi-species coalescent	BI
Diverge	*cox1, cytb, nad4, actin, EF-1α, histone3, 28S*	7,478	Divergence dating	BI

Next, species trees were inferred using the BEST multi-species coalescent (MSC) algorithms (Liu and Pearl, 2007) as implemented in MrBayes v.3.2. (Ronquist et al., 2012). The MSC models the inherent stochasticity in genetic drift to infer a probabilistic species tree by reconciliation of topological discordance among gene trees

(Rannala and Yang, 2003). Gene tree discordance is particularly problematic in phylogenetic analyses of closely-related species due to incomplete lineage sorting (ILS) (Maddison 1997), thus likely a more realistic fit given these taxa. I used two datasets, one comprising 18 genes (Coal set 1) and the second with 15 genes (Coal set 2), with overlap between them (Table 3). For set 1, 18 gene trees were estimated (1/gene), whereas for set 2, the five mitochondrial genes were grouped as a single locus, thus 11 gene trees were estimated. Prior settings were: brlenspr = clock:speciestree, topologypr = speciestree, popvarpr = variable, and popsizepr = gamma (1,100). Model and rate were as above (mixed/gamma) with topologies unlinked; mitochondrial gene ploidy was set to haploid. Four independent analyses of 30 million generations/set, with burn-in fraction = 0.33, were run via the CIPRES-Portal v.3.3 (Miller et al., 2010) and convergence was assessed as previously described. Coalescence analyses excluded *E. albidus.*

Finally, estimation of divergence time was performed in BEAST v1.8.2 (Drummond et al., 2012) and in MrBayes v3.2.6 (Ronquist et al., 2012) using a dataset comprising *cox1, cytb, nad4, actin, EF-1a, histone H3* and *28s* rRNA (Table 3). For these analyses, a minimum age bound for family Enchytraeidae was 35 million years before present based on a fossilized specimen found in Baltic amber (Ulrich and Schmelz, 2001). *Lumbricus terrestris* was used as the outgroup with sequences retrieved from GenBank (*cox1, cytb, nad4*: U24570.1; *28S*:HQ691218.1; *actin*:X96514.1; *EF-1α*:DQ813372.1; histone3:FJ214241.1). For the BEAST analysis, the XML file specified application of the GTR substitution model, 4 gamma categories, proportion invariant (0-1), and an uncorrelated relaxed clock (Drummond et al., 2006)

with lognormal distribution. For ‘time to most recent common ancestor’ (tmrca), a uniform age constraint prior on the root of 35/300 or 44/500 (mya) was applied. The lower bounds represent the youngest (35) and mean (44) estimated ages of Baltic amber (Larsson, 1978; Standke, 1998), while the upper bounds represent the estimated ages of oligochaetes (300) and annelids (500) (Bomfleur et al., 2012; Parry et al., 2014). Each age constraint file had 5 independent runs of 100 million generations, sampling every 10,000, executed via the CIPRES-Portal v.3.3 (Miller et al., 2010). Convergence, ESS, 95% HPD, and ucld.stdev were assessed in Tracer v1.6 (Rambaut et al., 2014). MrBayes analysis employed the independent gamma rates (IGR) relaxed clock model with a GTR + gamma substitution model and variable rates across partitions (determined with PartitionFinder v1.1.0; Lanfear et al., 2012). A uniform root constraint (35/300 mya) was set for calibration, and a partial backbone constraint was specified to separate the mesenchytraeids from the enchytraeids. Four independent runs of 5 million generations were performed and convergence was assessed as described previously.

Consensus trees were visualized and annotated using FigTree v.1.4.2 (http://tree.bio.ed.ac.uk/software/figtree) with species names color-coded according to habitat: aqua = aquatic, blue = glacier, green = snow, brown = terrestrial. DensiTree2 v.2.2.4 (Bouckart and Heled, 2014) was used to analyze and portray the degree of uncertainty present in consensus trees. Whereas consensus trees provide a quantitative measure of phylogenetic analyses, DensiTree offers a qualitative assessment by overlaying the full set of Bayesian trees, using transparent lines, to reveal areas of

uncertainty. Areas of agreement appear darker and thicker, while those with topological conflict appear more diffuse.

ATP assays

Intracellular ATP levels were assayed for *Mesenchytraeus solifugus, M. pedatus, M. antaeus* and *Enchytraeus crypticus.* Lack of sufficient numbers of specimens for *M. gelidus* and *M. hydrius* prevented their inclusion. The requirement for assays to be performed on freshly collected specimens meant that they were done at different times and, unavoidably, different luminometers. *Mesenchytraeus solifugus* and *M. pedatus* are aquatic and *M. antaeus* and *E. crypticus* are terrestrial so acclimation media differed, but all other aspects of the protocol were the same.

For each species, multiple worms were removed from their normal media, rinsed in sterile "leech water" (0.02% Instant Ocean/0.01% non-ionic aquarium salt) and divided into 3 multi-worm groups (4 for *E. crypticus*). Each group was placed into separate sterile 6 cm glass dishes. For aquatic worms, each dish contained 20 ml of temperature-equilibrated leech water, and for the terrestrial worms, leech water-soaked filter paper lined the bottoms. Dishes were covered with parafilm, placed in a light-free container, and then placed at desired assay temperatures (0, 4, 8, 16, 22°C) for 24-27 hours. For each assay temperature, 3 biological replicates of ~2-5 mg (wet weight of Kim wipe-blotted worms) were tested (3-8 worms/sample). All weights were measured using a Mettler Toledo AG285 scale accurate to 0.01 mg.

After incubation, blotted, weighed worms were placed in microfuge tubes containing 90 µl lysis buffer (50 mM HEPES, 1mM Ca_2Cl, 2 mg/ml proteinase K) pre-

heated to 50°, and incubated at 50° for 30-45 minutes, vortexing every 10 minutes. When worms were fully degraded, tubes were placed at 80°C for 5 minutes to stop proteinase K activity. Samples were chilled briefly on ice, centrifuged 20 seconds, then returned to ice. ATP assays were performed using StayBrite Highly Stable ATP Bioluminescence Assay Kit (BioVision) according to the manufacturer's protocol with 10 µl of sample. For each species, at each temperature, 3 technical replicates were performed. Two ATP standard curves (10^{-8} – 10^{-12} moles ATP) were performed per assay and averaged; this value was used to calculate ATP/mg. For *M. solifugus* and *M. pedatus*, luminescence readings were performed with a BetaScout luminometer (Perkin-Elmer) set to a 4 second delay. For *M. antaeus* and *E. crypticus*, a 20/20n luminometer (Turner Systems) was used, also with a 4 second delay.

Prediction of N-terminal mitochondrial targeting sequence

For the phylogenetic analyses, full-length coding sequences of nuclear-encoded ATP synthase subunits were evaluated, as the mitochondrial targeting pre-sequences were phylogenetically informative. However, assessment of potential impacts due to non-conserved amino acid substitutions requires comparisons between mature protein sequences trimmed of signal sequence. First, full coding sequences for transcriptome-extracted ATP synthase subunit genes were translated into amino acids (aa) and aligned in MEGA version 6 (Tamura et al., 2013). Outgroup species, *Enchytraeus albidus*, was excluded, as it was missing some subunit data, and its congener, *E. crypticus*, already provided the necessary outgroup representative. Next, three programs employing different algorithms for prediction of mitochondrial targeting sequences were applied:

two ExPASy (Artimo et al., 2012) suite programs, MitoProt (Claros and Vincens, 1996) and MitoFates (Fukasawa et al., 2015), plus TPpred 2.0 (Savojardo et al., 2014). These programs consider a variety of parameters (e.g. physico-chemical, aa position, sequence motif) to predict sites where enzymatic cleavage releases the mature subunit protein from the mitochondrial signal sequence. Lastly, the output predicted cleavage sites were evaluated by (1) conservation in protein alignments of the considered species and (2) comparison to experimentally validated subunit sizes for humans and yeast (*Saccharomyces cerevisiae*). WebLogo 2.8.2 (Crooks et al., 2004) was used to display the cleavable mitochondrial targeting sequences for the nuclear-encoded subunits. All consequent protein analyses employed trimmed, mature sequences.

Sequence characterization of ATP synthase subunit proteins

The following ExPASy bioinformatics resource proteomic tools were used to interrogate sequence data: ProtParam (Gasteiger et al., 2005) was used to extract physico-chemical properties of proteins such as molecular weight, charge, iso-electric point, and hydrophobicity (GRAVY). Prediction of membrane protein topology (ATP6, ATP8, c, b) was performed in TopPred 1.10 (Claros and von Heine, 1994), which uses hydrophobicity and the 'positive inside' rule to assess the likelihood of a trans-membrane α-helix. NCBI's protein BLAST was used to search for proteins with sequence similarity to the ATP6 carboxy-extension.

Determination of singleton sites in aa alignments was performed in MEGA version 6 (Tamura et al., 2013). Species-specific aa substitutions were counted as such if all species, except one, contained an identical amino acid in a given position;

multivariable positions were not included. Conservation was assessed by Blosum62 scores (Henikoff and Henikoff, 1992) as implemented in NCBI's Amino Acid Explorer: scores 1-10 = conserved, <1 = non-conserved; highly non-conserved scores (≤ -2) were noted. Boxshade v.3.21 was used for displaying alignment conservation.

Molecular modeling

Mesenchytraeus solifugus subunits showing the greatest divergence (Gamma, b and ATP6, based on non-conserved substitutions) were chosen for molecular modeling to determine the spatial position and possible interactions of the substitutions. SWISS-MODEL Workspace provided template selection and generation of molecular models (Arnold et al., 2006; Benkert et al., 2011; Biasini et al., 2014). DeepView (Swiss-Pdb-Viewer) was used for analysis, annotation and imaging of models. Models are presented in stick format with van der Waals interactions depicted as space-filling spheres. For subunits ATP6 and b, fully resolved templates were not available, thus some positions could not be visualized.

Testing for signatures of adaptive selection (d_N/d_S)

To test for adaptive selection, the value of ω, the ratio of non-synonymous to synonymous substitutions, was determined for each subunit using Codeml in the PAML v4.9a software package (Yang, 2007). Values of ω >1 are evidence of positive (adaptive) selection; ω = 1 is neutral; ω < 1 is negative (purifying) selection. Input trees were those used for the phylogenetic analyses in MrBayes v3.2.6 (Ronquist et al., 2011). All 11 subunits were tested using the "Sites" model, which allows the ω ratio to

vary among codons; (model = 0, NSites = 1 2 7 8, fix_omega = 0); ATP6 and ATP8 were also evaluated using the "Branch-Site" model, which allows the ω ratio to vary among codons and lineages [model = 2, NSites = 2, fix_omega = 1 (null) or 0 (alternative)]. Each analysis was run 3 times and results were evaluated by posterior probabilities calculated by Bayes Empirical Bayes (BEB).

Transcriptome raw read mapping

Sequence analyses revealed an anomalous 50 bp extension on the 3'-end of *M. solifugus atp6* transcripts. In order to verify that that the extension was not merely an artifact arising from transcriptome assembly error, pre-assembly raw reads were mapped to the recovered transcript. Paired end reads were aligned to the *atp6* reference sequence using BWA-mem (Li and Durbin, 2010), followed by implementation of SAMtools (Li et al., 2009) to process and visualize the alignment file; processing included removal of duplicate reads from the bam file. Default parameters were used for both BWA and SAMtools. Visualization and subsequent images were created using IGV (Robinson et al., 2011).

Results

Transcriptome cDNA sequence characteristics

Illumina sequencing produced >49 million paired-end reads per transcriptome, resulting in >102,000 contigs with an average size of 934 bp. The average G+C content (~41%) was uniform across species, including the outgroup. All gene queries were recovered, and raw-read mapping indicated that depth of coverage was more than sufficient. This description does not apply to the *E. albidus* transcriptome which was not generated for this investigation. Of 30 genes extracted for analysis, 29 were full length, with *28S* rRNA the only exception (Table 2). Sequence alignment was unambiguous for PCGs, but *12S* and *16S* rRNA required trimming. PCR-amplified mitochondrial genes matched those extracted from the respective species transcriptome. Neither ATP synthase subunit (ASU) *gamma*, nor *epsilon*, was recovered from the *E. albidus* transcriptome and were treated as missing data in phylogenetic analyses. GenBank accession numbers for novel sequences can be found in Appendix Table S4.

Among mesenchytraeids, the average nucleotide percent ID within concatenated datasets was (1) housekeeping genes: 87.17 (± 1.93 sd), (2) ATP synthase subunits: 85.14 (± 2.05 sd), (3) mitochondrial PCGs: 76.87 (± 1.52 sd), and (4) mitochondrial subset *cox1-cytb*-rRNAs: 83.8 (± 1.1 sd). Pairwise divergence values were highly similar within each dataset, displaying no obvious inter-specific

relationship trends. Percent identity at the protein level was ~10% greater (Appendix Table S2).

Gene Category	Gene	Alignment positions	Conserved		Variable		Parsimony inform.		Parsimony inform.	
			All	Mes	All	Mes	All	%	Mes	%
Nuclear-encoded housekeeping genes	actin	1131	878	911	253	220	157	13.9	94	8.3
	α-tubulin	1353	1026	1080	327	273	207	15.3	95	7.0
	EF-1α	1398	1041	1155	357	243	216	15.5	77	5.5
	histone H3	408	267	305	141	103	80	19.6	47	11.5
	GAPDH	1011	582	615	429	396	315	31.2	209	20.7
Nuclear-encoded ATP synthase subunits	alpha	1656	1133	1284	523	372	389	23.5	116	7.0
	beta	1311	958	1019	353	292	229	17.5	105	8.0
	gamma	840	417	599	423	241	198	23.6	83	9.9
	delta	525	319	357	206	153	159	30.3	61	11.6
	b	927	463	614	461	307	355	38.3	110	11.9
	c	486	296	350	187	124	152	31.3	41	8.4
	d	516	267	360	249	156	199	38.6	52	10.1
	OSCP	633	366	438	267	195	206	32.5	72	11.4
Mitochondrial protein-coding OXPHOS genes	COI	1536	1008	1079	528	457	344	22.4	199	13.0
	Cox2	678	414	460	264	218	183	27.0	95	14.0
	Cox3	777	472	513	305	264	222	28.6	125	16.1
	Cytb	1137	677	754	460	383	323	28.4	170	15.0
	Nad1	936	503	596	433	340	316	33.8	153	16.3
	Nad2	1005	419	532	586	470	440	43.8	200	19.9
	Nad3	354	171	203	183	148	125	35.3	46	13.0
	Nad4	1359	612	777	747	579	534	39.3	226	16.6
	Nad4L	297	141	183	156	111	118	39.7	52	17.5
	Nad5	1746	800	977	946	769	698	40.0	334	19.1
	Nad6	474	197	248	277	220	223	47.0	100	21.1
	Atp6	699	341	450	358	249	284	40.6	113	16.2
	Atp8	165	78	92	87	67	69	41.8	32	19.4
Mitochondrial RNA genes	12s	802	508	587	294	195	198	24.7	61	7.6
	16s	1164	744	898	420	266	287	24.7	79	6.8

Table 3. **Descriptive statistics of genetic loci.** 'All' refers to analyses including data from all 7 considered species; 'Mes' refers to analyses limited to the mesenchytraeids.

Expectedly, the housekeeping (HK) genes were the least parsimony informative (PI), with a range of 2.2 (*28S*) to 11.5% (*histone H3*) (Table 3). Interestingly, *histone H3* is the only gene in the dataset that is 100% conserved at the protein level. Furthermore, the low phylogenetic signal for *28S* within the genus jumps to >21% when comparing inter-generic data, indicating its utility for delimiting taxa at a higher level. Likewise, the ATP synthase subunits (ASU) had a PI range of 6.5 – 11.9% among

mesenchytraeids, but a few ~3-4-fold higher when the enchytraeids are included, and may prove useful in resolving annelid phylogeny questions. The mitochondrial PCGs had an average PI of ~16%, with *nad6* highest at 21.1%, and *cox1* and *nad3* lowest at 13%. Ribosomal RNA genes, *12S* and *16S*, were the most highly conserved and least phylogenetically informative in the mt-genome with PI s of 7.6 and 6.8%, respectively.

Phylogenetic relationships

Individual gene tree results were comparable with regard to partitioning scheme (partitioned vs. un-partitioned) and framework (Bayesian vs. ML), but topological discord was significant (Appendix Table S5). Of the 30 gene trees generated, 23 distinct topologies were present: HK genes (6/6), ASU genes [8/9; (*alpha*, *delta*)], mitochondrial genes [11/15; (*cytb, atp8*) (*nad2, 16S*) (*nad3, nad6, 12S*)]. The two enchytraeids consistently appeared outside the mesenchytraeid clade, even when outgroup status was not specified, supporting their utility in these analyses. Mesenchytraeid monophyly was violated only for ASU *gamma* and the ML analysis of the 15 mitochondrial genes (Mt13+RNAs), placing *M. pedatus* as an outgroup. The Bayesian analysis of the latter dataset placed *M. pedatus* as sister taxon to its four congeners with 100% support (Fig. 8). Indeed, this sister taxon positioning was the second most frequently observed grouping, occurring in five gene trees and the HK dataset. By far, the most common sister taxa pair was the ice worm, *M. solifugus*, and the white snow worm, *M. hydrius*, present in 16 gene trees, the ASU tree and the coalescent tree (Fig. 9). The two largest species, *M. antaeus* and *M. gelidus*, appeared

in four gene trees and the ASU tree as sister taxa, making them the third most common pair.

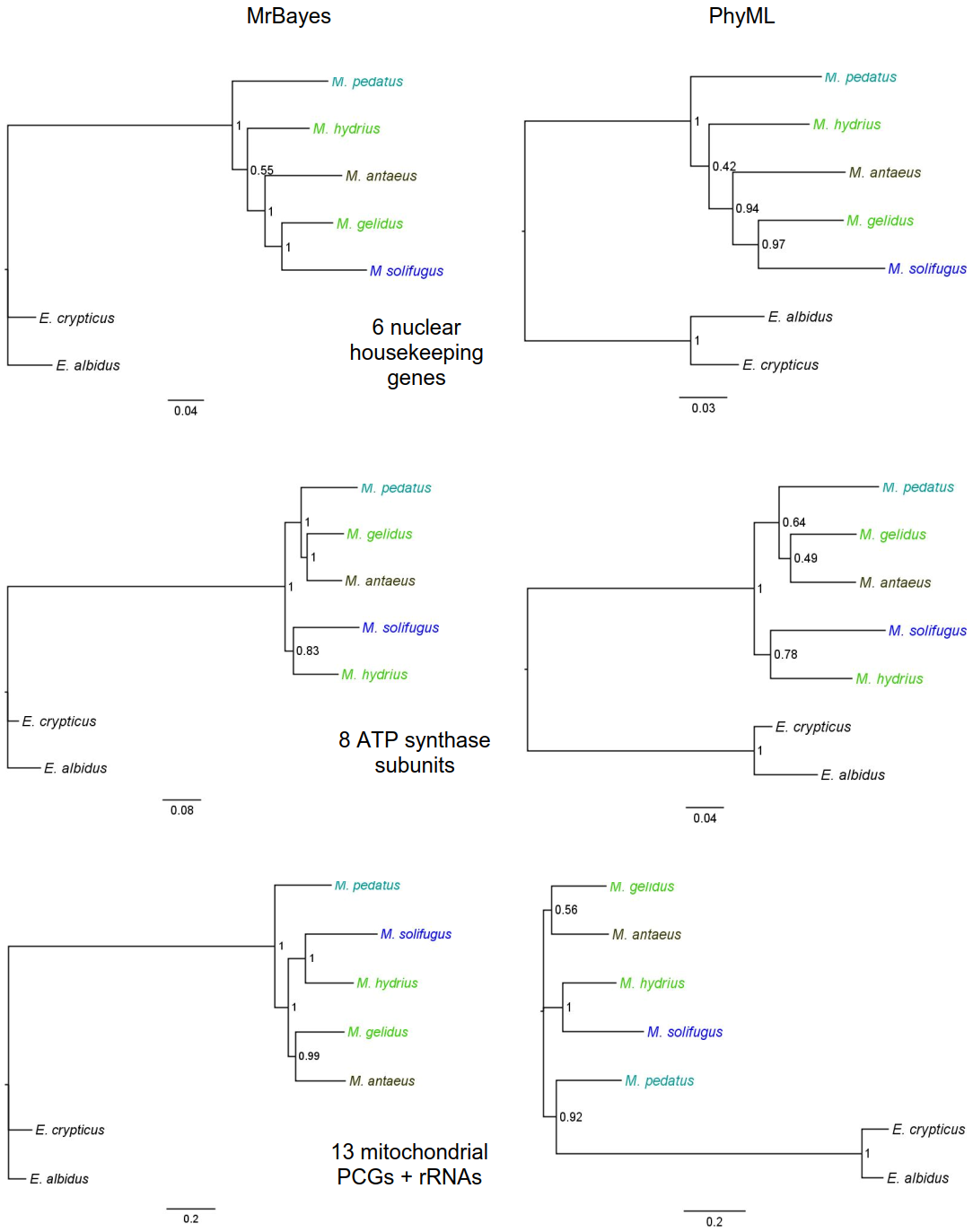
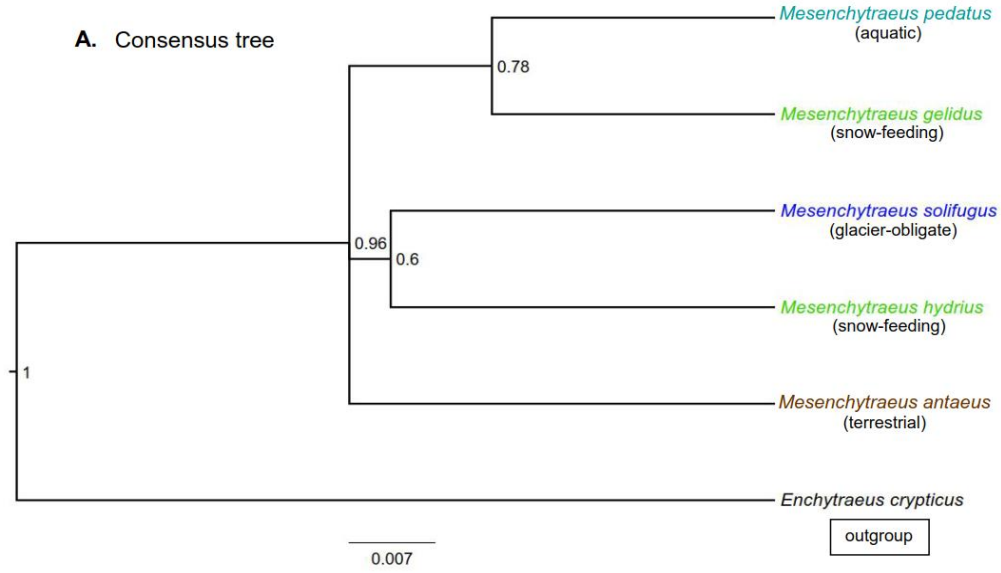


Figure 8. Bayesian and maximum likelihood consensus trees for concatenated datasets. Species names are color-coded to habitat: dark blue (glacier), aqua (aquatic), brown (terrestrial) and green (snow/soil). The outgroup species are in black. Numbers represent confidence scores: posterior probability (pp) for MrBayes and SH-aLRT for PhyML. Each dataset returned a well-supported bifurcating tree in the Bayesian analyses, yet they are topologically discordant. The ML trees show lower support, the mitochondrial dataset is unresolved, and again, the tree topologies differ.

The concatenated datasets, HK6, nuASU and Mt13+RNA, did little to resolve interspecific relationships, each producing topologically different consensus trees (the mitochondrial subsets, MtPCGs and Mt4, reiterated the Mt13+RNA tree). Unlike the individual gene trees, partitioning significantly affected topological outcomes. While no difference was seen for the HK dataset, un-partitioned analyses of the ASU and Mt13+RNA resulted in lower probability values and polytomies (data not shown).

The multi-species coalescent (MSC) analyses resulted in similar consensus trees for dataset 1 and 2, both with polytomies and low posterior probabilities (Set 2, Fig. 9A). Figure 9B is a DensiTree (Bouckart and Heled, 2014) visualization of the uncertainty present in the consensus tree. By overlaying the set of consensus trees (i.e. the average of all trees with the same topology) with the sum total of all trees generated in the analysis, areas of topological conflict become apparent. Blue lines represent the highest probability topology, thus explaining the consensus tree pairing of *M. pedatus/M. gelidus* and *M. solifugus/M. hydrius*. However, red lines, used for the second highest probability trees, and green, third highest, show a significant number of trees conflict with those groupings, hence the low posterior probabilities. The nearly equal-intensity red and blue lines for *M. antaeus* indicate considerable relationship

ambiguity. Attempts to establish divergence times using a fossil-calibrated root node were unsuccessful for both BEAST and MrBayes analyses. Output file assessments indicated that convergence was not achieved and ESS values were frequently below 200, even for combined runs, thus the resulting consensus trees were deemed unreliable.



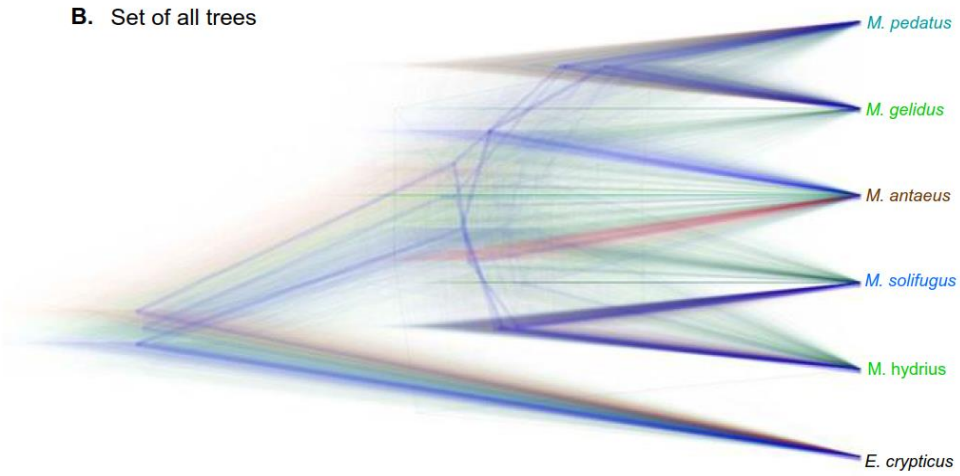


Figure 9. Bayesian multi-species coalescent (MSC) trees (set 2). Species are color-coded as in Fig. 6 and confidence values are pp. The MSC models the inherent stochasticity in genetic drift to infer a probabilistic species tree by reconciliation of topological discordance among gene trees due to incomplete lineage sorting (ILS). This type of discordance is commonly seen in analyses between closely-related species when speciation events and gene coalescence are out of sync. In (A) the consensus tree shows that even this paradigm was unable to resolve relationships among the species. The DensiTree (B) visualizes the uncertainty in the consensus tree by overlaying the full set of trees generated in the analysis; dark blue lines are trees that appeared most frequently, with decreasing frequency from blue > red > green.

Intracellular ATP levels

Consistent with previous studies that revealed elevated ATP in glacier ice worms (Napolitano et al. 2004; Napolitano and Shain 2004), results here indicated that ice worms maintain ATP significantly higher than all mesenchytraeids examined. ATP assays for the evaluated *Mesenchytraeus* species revealed significant differences within and between assays. The aquatic species, *M. solifugus* (ice worm) and *M. pedatus* (Fig. 10A) have ATP levels about an order of magnitude higher than the terrestrial species. Comparisons between them show ice worms with >2- and 4-fold higher ATP at 0 and 8°C, respectively, while ATP levels in *M. pedatus* show no significant differences over a 22°C range. For the terrestrial species (Fig. 10B), *M. antaeus* displayed significantly higher levels of ATP at each temperature tested than its confamilial, *Enchytraeus crypticus*. ATP levels in both species were unchanged between 4 and 8°C, below-normal temperatures for each, but at its optimum of 16°C, *M. antaeus* displayed significantly higher ATP than *E. crypticus* at its optimum of 22°C.

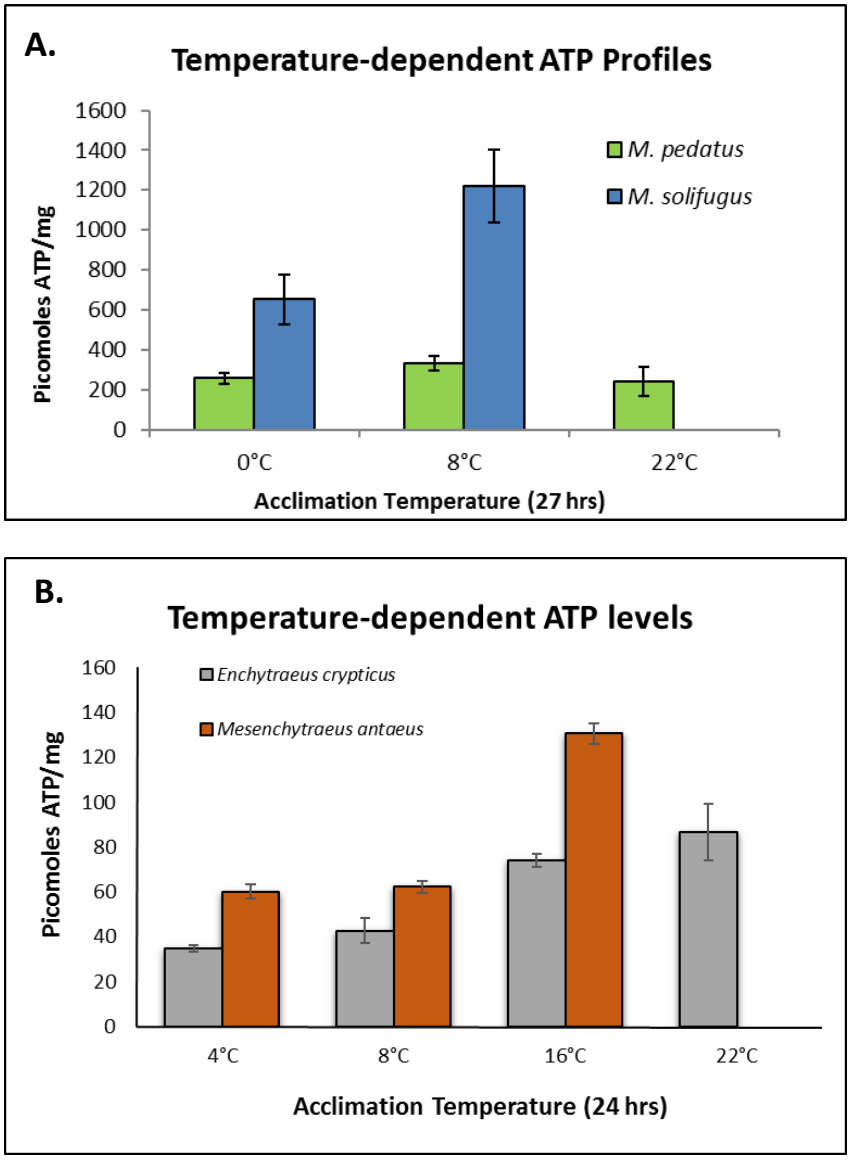


Figure 10. ATP profiles for four Enchytraidae species. Multiple worms were acclimated to different temperatures within their viable range for 24-27 hours before ATP extraction. Graphs represent the average of ≥3 biological and 3 technical replicates/species/ temperature. (A) Aquatic worms, *Mesenchytraeus pedatus* and *M. solifugus* have unusually high levels of ATP, with *M. solifugus* 2- and 4-fold higher at 0 and 8°C, respectively. (B) Terrestrial species, *M. antaeus* and *Enchytraeus crypticus*, have ~10-fold lower ATP levels compared to the aquatic worms, yet the mesenchytraeid's ATP is still significantly higher than its confamilial, *E. crypticus*, particularly at its optimum temperature of 16°C. Error bars are SEM.

Mitochondrial target sequence prediction

Despite using three different programs for mitochondrial target sequence prediction (TPpred, MitoFates, MitoProt), results were inconsistent and unreliable (Table 4). For α and OSCP, 2 of 3 predictions were in agreement and seemed correct, and for γ and δ at least one prediction was likely correct, but the remainder were dubious. Thus, resulting predictions were further evaluated by (1) conservation in protein alignments of the considered species and (2) comparison to experimentally validated subunit sizes for humans and yeast (*Saccharomyces cerevisiae*). For β, no predictions met these criteria and a cleavage site at ~57X/58G is postulated. Subunit-b (SynB) has only a single likely prediction; all others would result in a mature protein significantly larger than human or yeast proteins. Epsilon does not have cleavable transit peptide in human or yeast, yet two programs predicted presequence for this subunit. However, the N-terminal motif [W(R/K)X_5YX_2(Y/F)X_3(C/A)X_4RX_3K], highly conserved among metazoans (Tu et al., 2000), is present in the worm sequences, so cleavage predictions can be dismissed. This makes the worm Epsilon relatively long at 70 aa residues. Worm subunit-d (SynD), like human and yeast, does not possess a predicted presequence. In Table 4, highlighting indicates cleavage sites deemed correct, the forward slash indicates the cleavage site between the given aa's, and confidence scores are in parentheses. Figure 11 is a WebLogo depiction of the target sequences with an arrow showing where the presequence was removed to generate the mature protein.

Species	Prediction program	Nuclear-encoded ATP synthase subunits								
		Alpha	Beta	Gamma	Delta	Epsilon-L	OSCP	SynB	SynC/9	SynD
M. solifugus	TPpred	44L/45A (0.99)	35C/36A (0.64)	109N/110I (0.5)	34Y/35A (0.99)	26L/27K (0.69)	14F/15S (0.99)	44A/45A (0.84)	28L/29S (0.98)	NP
	MitoFates	44L/45A (0.96)	35C/36A (1.00)	24N/25M (0.67)	19T/20G (0.97)	NP	21Q/22G (0.98)	31S/32V (0.99)	28L/29S (0.87)	NP
	MitoProt	52E/53V (0.99)	71V/72D (0.97)	35L/36K (0.97)	19T/20G (0.99)	17V/18C (0.99)	21Q/22G (0.95)	31S/32V (0.99)	35R/36F (0.75)	NP
M. hydrius	TPpred	44L/45A (0.99)	47P/48T (0.81)	109N/110I (0.5)	34Y/35A (0.99)	26F/27K (0.58)	14F/15S (0.96)	35S/36V (0.57)	82R/83D (0.93)	NP
	MitoFates	44L/45A (0.89)	34A/35V (1.00)	24N/25M (0.57)	20T/21G (0.97)	NP	21R/22S (1.00)	35S/36V (0.96)	27L/28T (0.90)	NP
	MitoProt	52E/53V (0.99)	73V/74D (0.99)	35L/36K (0.96)	20T/21G (1.00)	17I/18C (1.00)	38Y/39A (1.00)	35S/36V (0.96)	74L/75F (0.38)	NP
M. gelidus	TPpred	44L/45A (0.99)	47P/48T (0.84)	109N/110I (0.5)	34Y/35A (0.99)	26L/27K (0.61)	14F/15S (0.98)	35C/36V (0.64)	82R/83D (0.92)	NP
	MitoFates	44L/45A (0.95)	34A/35A (1.00)	24N/25M (0.65)	19T/20G (0.97)	NP	21R/22A (1.00)	35C/36V (0.97)	27L/28S (0.89)	NP
	MitoProt	52E/53V (0.99)	71V/72D (0.96)	35L/36K (0.96)	19T/20G (0.99)	17I/18C (0.99)	21R/22A (0.98)	35C/36V (0.99)	74L/75F (0.77)	NP
M. antaeus	TPpred	44L/45A (0.99)	48A/49A (0.88)	109N/110I (0.5)	34Y/35A (0.99)	26L/27K (0.62)	21R/22A (0.97)	72L/73S (0.69)	27L/28S (0.96)	NP
	MitoFates	44L/45A (0.95)	43S/44D (0.98)	24N/25M (0.69)	34Y/35A (0.97)	NP	21R/22A (1.00)	35S/36V (0.97)	27L/28S (0.91)	NP
	MitoProt	52E/53V (0.99)	72V/73D (1.00)	35L/36K (0.98)	19S/20G (0.99)	17V/18C (0.98)	21R/22A (0.95)	35S/36V (0.98)	74L/75F (0.41)	NP
M. pedatus	TPpred	44L/45A (0.99)	40R/41S (0.80)	109N/110I (0.5)	34Y/35A (0.99)	26L/27K (0.58)	14F/15S (0.99)	35S/36V (0.99)	82R/83D (0.88)	NP
	MitoFates	44L/45A (0.96)	41S/42D (0.98)	24N/25M (0.67)	19T/20G (0.96)	NP	21R/22S (0.96)	35S/36V (0.50)	27L/28S (0.90)	NP
	MitoProt	52E/53V (0.99)	68V/69D (0.99)	35L/36K (0.96)	19T/20G (0.99)	17V/18C (0.99)	21R/22S (0.99)	35S/36V (0.93)	74L/75F (0.73)	NP
E. crypticus	TPpred	37L/38S (1.00)	44A/45A (0.82)	116N/117I (0.5)	39Y/40A (0.99)	NP	14F/15S (0.99)	34N/35A (0.76)	NP	NP
	MitoFates	44N/45A (1.00)	42C/43A (0.99)	31N/32M (0.54)	39Y/40A (1.00)	NP	21Q/22G (0.92)	34N/35A (1.00)	28L/29S (0.48)	NP
	MitoProt	44N/45A (0.99)	Error	42L/43K (0.91)	Error	17I/18C (0.99)	21Q/22G (0.98)	34N/35A (0.80)	Error	NP
Mature protein (AA):	Worms:	508	472-474?	273	136	50-70?	190	233?	75	172
	Human:	510	482	273	146	50	190	214	75	160
	Yeast:	510	478	278	138	61	195	209	76	173

Table 4. Predicted cleavage site for nuclear-encoded mitochondrial ATP synthase subunits. Three programs employing different prediction algorithms were used to determine sites where enzymatic cleavage releases the mature subunit protein from the mitochondrial signal sequence. Predicted cleavage sites were evaluated by (1) conservation in protein alignments of the considered species and (2) comparison to experimentally validated subunit sizes for humans and yeast (*Saccharomyces cerevisiae*). Highlighting indicates sites most likely correct and the predicted mature protein size for the worms is shown at bottom. The forward slash indicates the cleavage site between the given amino acids, followed by confidence scores (in parentheses) generated by the prediction programs.

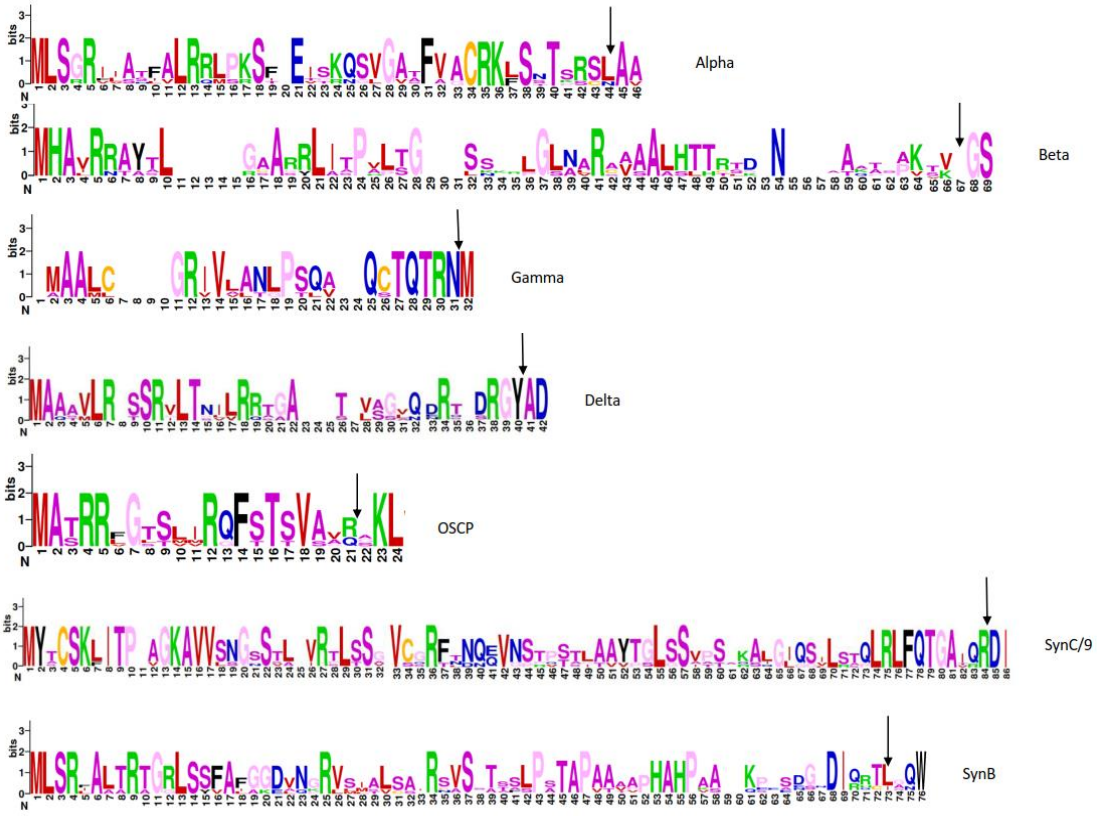


Figure 11. WebLogo representations of predicted targeting signal sequence for nuclear-encoded ATP synthase subunits. Arrows indicate sites of enzymatic cleavage to produce mature protein. All comparative sequence analyses and molecular modeling considered only the mature protein with cleavable signal sequences removed at the shown positions.

Physico-chemical characteristics

For each species, physico-chemical characteristics – molecular weight (kD), isoelectric point (pI), charge, and hydrophobicity values – were determined for each of the 11 subunits, and then summed to assess them in the context of the holoenzyme (Table 5). For subunits Alpha and Beta, contributing 3 copies each to the ATP synthase monomer, and subunit c with 8 copies, values for molecular weight and charge were adjusted to reflect their stoichiometry. The average MW was similar for all, ~529 kD, which is a bit low compared to mammals (~600 kD), but a few small subunits (F6, e, g) are not included here. Likewise, isoelectric points are similar across species, ~86, but charge and hydrophobicity show a great deal more variation. For mesenchytraeids, the overall charge of the ATP synthase monomer is >2-fold more positive than for *Enchytraeus crypticus.* This may be compensatory, in light of the colder temperatures inhabited by mesenchytraeids, since ionic interactions are strengthened at low temperature. The relatively high hydrophobicity value for *Mesenchytraeus hydrius* (0.599) however, is less easily explained. Although it is ~2-fold higher than *E. crypticus* (0.3), it is greater still for its congener, *M. gelidus* (0.282), and ~1.5 times that of *M. pedatus* (0.38) and *M. antaeus* (0.37). Hydrophobicity in the ice worm, *M. solifugus*, is 71% of *M. hydrius*.

Species		Alpha*	Beta*	Gamma	Delta	Epsilon	OSCP	SynB	SynC/9*	SynD	ATP6/a	ATP8	**Totals**
M. solifugus	MW	164.34	151.56	30.41	14.35	7.97	20.92	26.85	60.16	19.54	27.41	6.12	529.63
	pI	8.58	4.99	9.6	4.61	10.28	9.32	7.97	6.11	9	6.78	9.31	86.55
	+/-	+9	-51	+14	-8	+8	+6	+1	0	+3	-1	+2	-17
	HΦ	-0.024	0.041	-0.164	0.075	-0.344	-0.087	-0.358	1.003	-0.644	0.831	0.096	0.425
M. hydrius	MW	164.46	152.04	30.59	14.13	7.79	21.06	26.97	60.16	19.6	25.8	6.09	528.69
	pI	8.57	5.08	9.69	4.61	10.28	9.5	6.73	6.11	8.94	6.56	8	84.07
	+/-	+9	-48	+14	-8	+8	+8	-1	0	+3	-1	+1	-15
	HΦ	-0.047	0.026	-0.214	0.19	-0.51	-0.114	-0.399	1.003	-0.658	1.016	0.306	0.599
M. gelidus	MW	164.55	151.83	30.52	14.28	7.71	21.11	27.2	60.16	19.58	25.84	6.05	528.83
	pI	8.75	5.04	9.53	4.58	10.17	9.56	8.48	6.11	8.96	6.26	9.79	87.23
	+/-	+12	-48	+12	-8	+8	+9	+2	0	+3	-2	+3	-9
	HΦ	-0.048	0.026	-0.18	0.105	-0.503	-0.191	-0.451	1.003	-0.691	1.029	0.183	0.282
M. antaeus	MW	164.61	151.92	30.53	14.32	7.74	21.12	27.26	60.16	19.54	25.8	6.08	529.08
	pI	8.76	5.04	9.6	4.58	10.28	9.48	7.25	6.11	8.98	6.26	9.79	86.13
	+/-	+12	-48	+12	-8	+8	+8	0	0	+3	-2	+3	-12
	HΦ	-0.059	0.031	-0.185	0.074	-0.451	-0.231	-0.488	1.003	-0.599	1.042	0.233	0.37
M. pedatus	MW	164.43	151.92	30.62	14.19	7.78	21.15	27.04	60.16	19.51	25.96	6.11	528.87
	pI	8.76	5.04	9.58	4.56	10.28	9.5	7.95	6.11	8.98	6.88	9.63	87.27
	+/-	+12	-48	+13	-8	+8	+9	+1	0	+3	0	+3	-7
	HΦ	-0.036	0.02	-0.251	0.105	-0.394	-0.202	-0.409	1.003	-0.607	1.053	0.098	0.38
E. crypticus	MW	164.25	151.65	30.53	14.16	7.73	20.93	26.97	60.16	19.39	26.22	6.45	528.44
	pI	6.58	4.93	9.36	4.51	10.35	9.62	9.07	6.11	7.84	5.78	10.6	84.75
	+/-	-3	-57	+10	-9	+7	+11	+5	0	+1	-4	+5	-38
	HΦ	-0.014	0.036	-0.21	0.006	-0.736	-0.117	-0.484	1.003	-0.509	0.968	0.357	0.3

Table 5. **Physico-chemical properties of ATP synthase subunits.** Molecular weight (MW) in kiloDaltons, isoelectric point (pI), overall charge (+/-), and hydrophobicity (HΦ) as determined by the 'Grand Average of Hydropathy' (GRAVY) score. Parameter values are totaled across subunits for each species, and significantly different sums are highlighted. *Note that for Alpha, Beta and subunit SynC(9), MW and charge values are adjusted for their stoichiometry – Alpha and Beta x 3, c/9 x 8. The mesenchytraeid ATP synthase is >2-fold more positive than that of *Enchytraeus crypticus*, and total hydrophobicity in *Mesenchytraeus hydrius* is 2-fold greater than congener, *M. gelidus,* and confamilial, *E. crypticus.*

Subunit sequence polymorphisms

As a whole, the ATP synthases among the mesenchytraeids are highly conserved – approximately 92 % aa identity (Appendix Table S2) – but examination of individual subunits revealed striking differences, both between proteins and among species. Counts of species-specific unique singleton sites (i.e. only counted if a single species differed from consensus in the other four) demonstrated that aa substitutions in the ice worm (*M. solifugus*) far exceeded those of its congeners (Table 6). For the 11 subunits, ice worms had 21 conserved, versus 46 non-conserved, unique substitutions (conservation according to Blosum62 scores), including 3 insertion events. Of the 46 non-conserved changes, 6 were highly non-conserved (Blosum62 $\leq$ -2). By contrast, conserved vs. non-conserved for the other mesenchytraeids were: *M. hydrius* (21/14), *M. gelidus* (11/6), *M. antaeus* (6/5), and *M. pedatus* (31/17).

Gene	Species	Conserved	Total Conserved	Non-conserved	Total Non-conserved
Alpha	*M. solifugus*		0	**N113I**	1
	M. hydrius		0		0
	M. gelidus	L121V, S123N	2		0
	M. antaeus		0		0
	M. pedatus	D190E, V444I, V455I	3	T491A	1
Beta	*M. solifugus*	V101L, , R405K, M442L, I448V	4	V367T, V460A, **K461L**	3
	M. hydrius		0	**A105H**	1
	M. gelidus		0		0
	M. antaeus		0		0
	M. pedatus		0		0
Gamma	*M. solifugus*	N49S, I58V, L119I, S139A	4	I69T, V94A, **P95L**, M103T, **R121P**, N137Q	6
	M. hydrius	I148V, R166K, K172R, I209L, D244E, V263M	6	S8Q, **E96L**, A99P, T145A, A170T, L213F	6
	M. gelidus	F214Y	1		0
	M. antaeus	K150R, L184V, Q186E, L190I	4		0
	M. pedatus	A117S, E154D, E199D	3	**G158C**	1
Delta	*M. solifugus*	D134E	1		0
	M. hydrius	R47K, D57E, S93A, V133I	4	S120G	1
	M. gelidus	L108I	1		0
	M. antaeus	A11S	1	T19A	1
	M. pedatus	E98D	1		0
Epsilon	*M. solifugus*		0	P28A, 58Sins, T60Q, V62A, K67T, 69-70KAins	7
	M. hydrius	N13H, Y14F	2	S3G, **L6Q**, L26F	3
	M. gelidus		0		0
	M. antaeus	A63S	1	A64T	1
	M. pedatus		0		0
SynB	*M. solifugus*	V182I	1	A5V, G35A, F87L, R92T **K100G**, E130A, A204K	7
	M. hydrius	H7N, I37V, I85L, D107E	4	R187T	1
	M. gelidus	L34I, E124D	2	T28R, P31A	2
	M. antaeus		0	Q196N	1
	M. pedatus	D103N, Q110E, D111E, D167E, D200N, A229S	6	T40P, **G95F**, S117P, A121T, T186V, H190Q, H192E	7
SynC	*M. solifugus*	100% AA identity after signal sequence removed	0		0
	M. hydrius		0		0
	M. gelidus		0		0
	M. antaeus		0		0
	M. pedatus		0		0
SynD	*M. solifugus*	L37I, S73A, K89R, I92V, K118Q	5	T2A, H169Q	2
	M. hydrius	R152K, N48S	2		0
	M. gelidus	R25K, Q104E	2	F17L, I113T, A129P	3
	M. antaeus		0	A102V, K118T	2
	M. pedatus	A78S, I107V	2	**R147L**, A170T	2
OSCP	*M. solifugus*	V4I, Q42K, V163I	3	Q70A, K75T, K97M, R135T	4
	M. hydrius	I138L	1	**H19T**, N103G	2
	M. gelidus		0		0
	M. antaeus		0		0
	M. pedatus	V81I, I117L	2	L142F	1
ATP6	*M. solifugus*	T59S, N65S, L118M	3	**G192T**, 233-48ins(15 aa)	16
	M. hydrius		0		0
	M. gelidus	V18I, F58Y, V112I	3		0
	M. antaeus		0		0
	M. pedatus	I41M, S47T, E52K, Q55E, V67L, L75I, V81I, T107S, T122S, S176N, M195L, I201M, L204I	13	F49L, F89M, V178A, A230V	4
ATP8	*M. solifugus*		0		0
	M. hydrius		0		0
	M. gelidus		0	L53Y	1
	M. antaeus		0		0
	M. pedatus	S34T	1	S26E	1

Table 6. Unique aa substitutions in mature sequence of ATP synthase subunits. Data consider only singleton sites where four species have the same amino acid and one species differs; multi-variable sites are not included. Substitutions within identified cleavable signal sequence are also excluded. Conservation is assessed by Blosum62 scores: 1-10 = conserved, <1 = non-conserved (scores ≤ -2 in bold). Total non-conserved substitutions in 11 subunits per species: *M. solifugus* (46), *M. hydrius* (14), *M. gelidus* (6), *M. antaeus* (5), *M. pedatus* (17). Insertions and deletions are in red.

There are two unusual aspects to these data – the first being that normally, the ratio of conserved/non-conserved aa changes is >1, which is seen here for all but the ice worm. Secondly, in light of our phylogenetic analyses indicating that these worms speciated within a very narrow timeframe, you would expect to see a very similar number of changes among them under a neutral evolution model. However, here we see a low of 11 total changes for *M. antaeus* (6 + 5), and a high of 67 for *M. solifugus.* Although the total for *M. pedatus* is fairly high (48), almost half are non-conserved and it is likely that these worms are a mixed population with possible cryptic speciation. This discordance in mutation rates is consistent with the typical pattern of accelerated evolution following a rapid radiation event as each species adapts to a new environment.

In general, the conserved changes conform to known trends in cold-adapted proteins – reductions in size and potential hydrogen bonds (e.g. N>S, I>V) and/or increased hydrophobicity (K>Q) – or they are virtually interchangeable (D>E, L>I, T>S). Many of the non-conserved changes reflect these proclivities as well: F>L, H>Q, R>T, K>M, T>V, K>T, but some, such as the K>G seen in ice worm SynB, are so

drastically different (size, H-bonds, pI, hydrophobicity) that, unless located in a peripheral loop, are likely to have a large impact on structure/function.

Another interesting observation in these data is the widely varying levels of conservation among different subunits. For example, the aa sequence for SynC is 100% identical for all of the mesenchytraeids, and for the enchytraeids as well; even human c differs at only 6 of 75 residues (Fig. 12). The c (A9) subunit protein assembles into a homopolymer forming the critical rotary motor, with the number of subunits affecting the 'cost' of ATP synthesis (fewer = lower cost). Walpole and colleagues (2015) found that diverse metazoans possess a c_8-ring that can be predicted from 3 conserved alanine residues in the inner helix that, if changed, would sterically inhibit an 8-mer. Since the considered species have alanine at the specified positions, they too must have a c_8-ring.

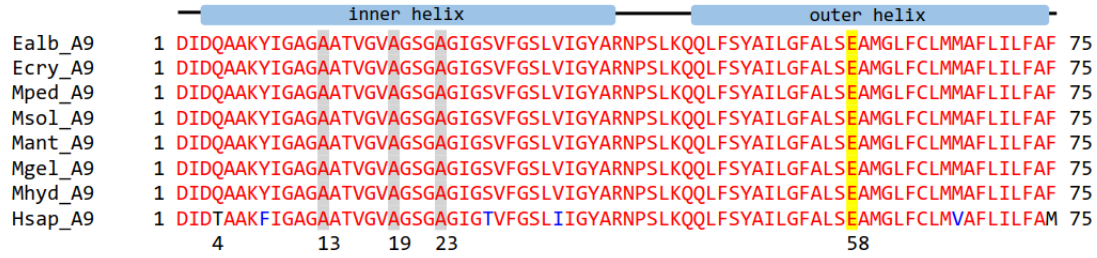


Figure 12. Sequence alignment of enchytraeid and human subunit c/9 mature protein. Among metazoans, the very high aa conservation in the c subunit is thought indicative of a common C_8-ring stoichiometry. In particular, steric constraints resulting from the hourglass shape found in the bovine c-ring limit positions 13, 19, and 23 of the inner helix to alanine (Walpole et al., 2015). The glutamine (Q) at position 4 is characteristic of Lophotrochozoans, while Phe75 (F) is found in diverse invertebrates. Glutamate-58 in the outer helix is the strictly-conserved proton acceptor responsible for forward rotation of the c-ring. (Hsap = human).

Alpha, Beta, Delta and ATP8 have only a few non-conserved changes, as is expected for proteins in critical metabolic pathways due to strong negative selection to preserve function. Both Alpha and Beta have canonical P-loop (phosphate-binding loop) motifs [GDRQTGKT] and [GGAGVGKT], respectively, identical to bovine (Walker et. al., 1982). The Beta 'DELSEED' loop that, through contact with Gamma, transmits the rotational torque, has the motif 'DELSEDD' in these worms (Hara et al., 2001). Neither Delta, a component of the rotor, nor ATP8, part of the membrane-intrinsic stator, display any non-conserved changes in the ice worm.

Three subunits, Gamma, SynB, and ATP6, account for more than half of all changes and, as all play integral roles in ATP synthase structure and function, these changes would seem to be adaptive. In an effort to assess the impact of the non-conserved changes, based on their location within the tertiary structure, molecular models for ice worm ATP6, Gamma and SynB were generated (Fig. 13: A. Gamma, B. SynB; the ATP6 model is discussed separately) In Gamma (A) the shaft, formed by coiled-coils comprised of the N- and C-terminals, has no non-conserved changes. The most highly non-conserved substitutions, R>P145 and P>L119 (orange), as well as I>T93, V>A118, and M>T127 (green), are grouped on the upper aspect (foot), distal to the shaft, a region proposed to be in contact with Epsilon and Delta residues. The N>Q161 residue, located in a loop region on the underside of the foot, is proximal to the c-ring interface.

Subunit b (SynB) does not have a fully resolved structure, thus four non-conserved changes (A>V5, G>A35, F>L87, R>T92) located in the membrane helices could not be modeled as residues 1-98 are missing, but this domain is known to interact

with ATP6 and ATP8, and to participate in the dimer-dimer interface (Davies et al., 2012). The K>G mutation, shown at the N-terminal of the model, is predicted to be near the matrix-membrane interface, proximal to the Gamma foot area where the cluster of changes is shown in A. The E>A residue is likely in contact with SynD, while A>K204 (blue), near the C-terminal, is within the OSCP stator region. Protein Databank templates used for modeling were 2w6j.1.G (Gamma) and 2cly.1.A (SynB).

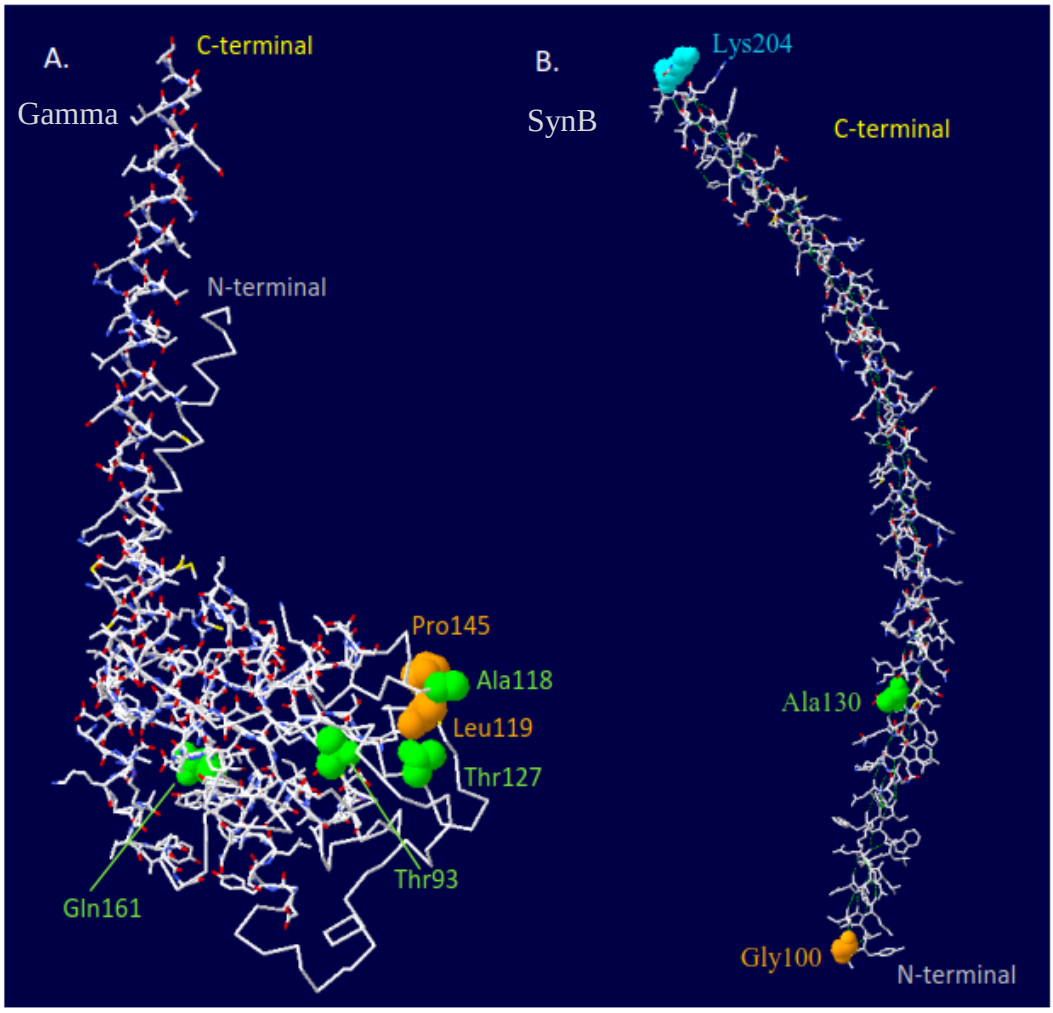


Figure 13. Molecular model of *M. solifugus* ATP synthase subunits Gamma (A) and SynB (B) showing non-conserved amino acid changes; van der Waals surfaces are rendered in 3D, highly non-conserved substitutions are in orange, remainder in green. (A) For Gamma, all

non-conserved changes are found in the 'foot' domain which interacts with the c-ring to form the rotor. (B) Subunit b, of the peripheral stator, does not have a fully resolved structure, thus the model lacks residues 1-98 and 205-234. The N-terminal has two predicted membrane-intrinsic helices that interact with ATP6 and ATP8, and the ice worm has four non-conserved changes within this region: A>V5, G>A35, F>L87, R>T92. The highly non-conserved K>G100 is localized in the matrix near the membrane surface at the juncture between ATP synthase dimers, while E>A130 is likely in contact with stator subunit d. The final substitution, A>K204 (blue), near the C-terminal, is within the OSCP stator region. Protein Databank templates used for modeling were 2w6j.1.G (gamma) and 2cly.1.A (b).

One final, but compelling, observation is the presence of three insertion (indel) events in ice worm subunits, two in Epsilon and one in ATP6. Indels, much more so than aa substitutions, affect the tertiary structure of a protein and alter its fitness landscape (Gillespie, 1984; Leushkin et al., 2012). Effects due to altered structure are compounded in proteins that function within multimeric complexes due to quaternary structure interactions (Yang et al., 2012; Odokonyero et al., 2014). As a consequence, indels create mutational 'hotspots' in adjacent protein residues to provide for compensatory adjustments necessary to maintain function (Tian et al., 2008; Zhang et al., 2011).

Subunit Epsilon, part of the F_1 rotor, is not well characterized, and has no bacterial counterpart, but is proposed to play a critical role in attachment of the F_1 domain to the c-ring rotor (Havlickova et al., 2010). It forms a hair-pin shaped heterodimer with Delta, both making extensive contact with the Gamma 'foot'. Of the considered species, all but *M. solifugus*, have two Epsilon isoforms, one short (58 aa) and one long (66 or 67 aa). However, the ice worm has only a long form with 3

additional amino acids – a serine at position 58, and two additional C-terminal residues, KA (69-70). A Blast search of the *Helobdella robusta* (leech) genome returned a single *epsilon* isoform containing a 283 bp intron that, when spliced out, produces an mRNA similar in size to the Enchytraeidae short transcript (171:177 bp).

In figure 14, the *H. robusta* genome sequence is shown aligned with the short and long isoforms of the enchytraeids. The codon spanning the splice site translates to an isoleucine for all species, and long and short isoforms are identical up to base pair 163, and also for the final 14 base pairs. This suggests that a second or alternative splice site is responsible for the transcript variants, though multi-copy genes may also be culpable. The lack of variants in the ice worm may reflect functional differences in the isoforms related to environmental conditions such as thermal constraints.

The most significant and unexpected finding of these comparative analyses was the unusually long *M. solifugus atp6* transcript recovered from Blast searches of our transcriptome databases. Both independently-generated ice worm transcriptomes contained the same 750 bp mitochondrial-encoded transcript – 48-54 bp longer than any of the other considered species. An insertion of this magnitude is quite extraordinary given, (1) the structural and functional importance of the ATP6 protein in the F_1F_O-ATP synthase, and (2) the evolutionary trend toward compaction in the mitochondrial genome (Selosse et al., 2001; Adams and Palmer, 2003). Thus, additional efforts to ensure that the transcript was real, and not due to assembly error, are described below.

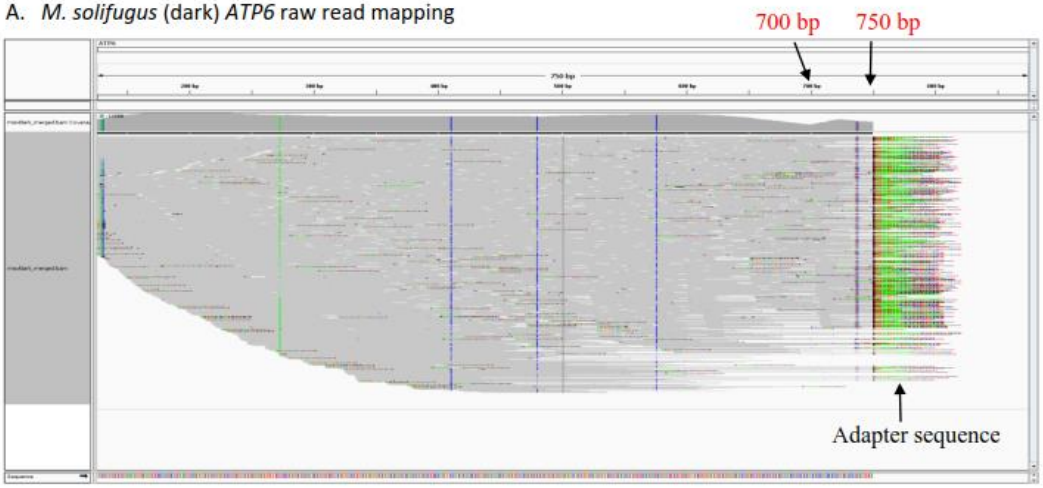
>**Helobdella robusta** gi|675893983:523768-524221 [intron in yellow]
ATGTCCGGTTGGAGGCAAGCAGGACTCAA**GT**AACTCTTACTTACTTAATTACTTCTTATTTACGATTAAA
TTTTACTTTATTTGGTCCTCATCTGGCGAACTAAAACTAAAAGTGAAACTAAATCAAAACTGTATTAACT
GAAAGTTAAGACTTAAAAACTAAAACCAAAACAATACAATTCATCAGGCTTTTTATCATTAGAATTATTT
ATAATTATATTATTAAAAAAAATTATTTACATATGTTTTTTCCACTTTTCAAAACTTTTTCATTTTAATT
TTATTTTACATTTTCTTTATTTTATTCAAC**AG**TTACATCCAGTTTTCAAGAATATGTGCTCGAATCGTGA
GGAGGTGTTTGAAGCCAGAACAACAGGCTGAAGCCATGAAAAGAGATGAAGGTCTCATCAGGGCTATCCA
TTGGAGAGATGGCAAACCGATCAGTGAGAAATAA
>**Ecry_epsilonL** 201 bp ▼
ATGTCAGCGTGGAGACAGGCCGGACTCAATTATATCAACTATTCGCGCATATGTGCGCGCGTGGTGCGTC
GATCTCTCAAGCCAGAGCTGCGCACGGAGGCTCTAAAGCGCGATGAGGGGACGATTCGCCCAGTGTTCTG
GAAGGATGGGCAGCCTATTCAGA TCAAGAAAACACAAGTTGCTGCTG AGGAAAGGACCTGA
>**Ecry_epsilonS** 177 bp ▲
ATGTCAGCGTGGAGACAGGCCGGACTCAATTATATCAACTATTCGCGCATATGTGCGCGCGTGGTGCGTC
GATCTCTCAAGCCAGAGCTGCGCACGGAGGCTCTAAAGCGCGATGAGGGGACGATTCGCCCAGTGTTCTG
GAAGGATGGGCAGCCTATTCAGA AGGAAAGGACCTGA
>**Mant_epsilonL** 204 bp
ATGTCTTCTTGGAGGCTGGCAGGACTTAACTACATCAACTATTCTCGTGTATGCGCACGTATTGTACGGC
GGGCCCTGAAGCCAGAGCTACGTGTGGACGCAATGAAGCGAGATGAAGGTTTTATTCGACCAATCATGTG
GAAAGATGGAAAGGCCGCCGTCA TCAAGAAAAGTACACAAGTATCCACCG AGGAGAAAGCTTAA
>**Mant_epsilonS** 177 bp
ATGTCTTCTTGGAGGCTGGCAGGACTTAACTACATCAACTATTCTCGTGTATGCGCACGTATTGTACGGC
GGGCCCTGAAGCCAGAGCTACGTGTGGACGCAATGAAGCGAGATGAAGGTTTTATTCGACCAATCATGTG
GAAAGATGGAAAGGCCGCCGTCA AGGAGAAAGCTTAA
>**Mgel_epsilonL** 204 bp
ATGTCTTCGTGGAGGCTCGCAGGACTCAACTACATCAATTATTCTCGCATTTGTGCACGCGCCGTCCGGC
GGTCCCTTAAGCCAGAGGTTCGCGTGGATGCAATGAAGCGAGATGAAGGATTCATTCGACCAATCTACTG
GAAGGATGGAAAGGCCGCCGTCA TCAAGAAACCTACACAAGTAGCTGCTG AGGAGAAAGCCTGA
>**Mgel_epsilonS** 177 bp
ATGTCTTCGTGGAGGCTCGCAGGACTCAACTACATCAATTATTCTCGCATTTGTGCACGCGCCGTCCGGC
GGTCCCTTAAGCCAGAGGTTCGCGTGGATGCAATGAAGCGAGATGAAGGATTCATTCGACCAATCTACTG
GAAGGATGGAAAGGCCGCCGTCA AGGAGAAAGCCTGA
>**Myhd_epsilonL** 204 bp
ATGTCCGGGTGGAGGCAAGCAGGACTCAACTACATCCACTTTTCTCGCATATGCGCTCGTGCTGTTCGGC
GGTCCTTTAAGCCAGACATCCGCGTGGAAGCAATGAAGCGAGATGAAGGATTCATTCGGCCGATCTACTG
GAAGGATGGAAAGGCCATTGTCA TCAAGAAACCTACACAAGTAGCTGCTG AGGAGAAAGCTTGA
>**Mhyd_epsilonS** 177 bp
ATGTCCGGGTGGAGGCAAGCAGGACTCAACTACATCCACTTTTCTCGCATATGCGCTCGTGCTGTTCGGC
GGTCCTTTAAGCCAGACATCCGCGTGGAAGCAATGAAGCGAGATGAAGGATTCATTCGGCCGATCTACTG
GAAGGATGGAAAGGCCATTGTCA AGGAGAAAGCTTGA
>**Mped_epsilonL** 204 bp
ATGTCTTCGTGGAGACTCGCAGGACTCAACTACATCAATTATTCTCGCGTTTGCGCGCGTATTGTTCGGC
GGTCACTTAAACCAGATATTCGTGTGGAAGCAATGAAGCGAGACGAAGGATTCATTCGACCAATCTTCTG
GAAGGATGGAAAGGCCGTCGTCA TCAAGAAAAATACACAAGTTGCAGCGG AGGAGAAAGCTTGA
>**Mped_epsilonS** 177 bp
ATGTCTTCGTGGAGACTCGCAGGACTCAACTACATCAATTATTCTCGCGTTTGCGCGCGTATTGTTCGGC
GGTCACTTAAACCAGATATTCGTGTGGAAGCAATGAAGCGAGACGAAGGATTCATTCGACCAATCTTCTG
GAAGGATGGAAAGGCCGTCGTCA AGGAGAAAGCTTGA
>**MsolD_epsilon** 213 bp
ATGTCTTCTTGGAGGCTAGCAGGACTAAACTACATCAACTATTCTCGTGTGTGCGCGCGCGTTGTTCGAC
GGGCCCTCAAGGCAGATCTGCGTGTCGATGCGATGAAGCGAGACGAAGGATTCATTCGACCAATCTTCTG
GAAGGATGGAAAAGCAATCGTCATCAAGAAAAGTCCACAACAAGCTGCCGCTGAGGAGACTGCGAAGGCT
TGA

Figure 14. **DNA alignment of *epsilon* subunit isoforms.** The Enchytraeidae, with the exception of the ice worm, *Mesenchytraeus solifugus*, have two isoforms of the ATP synthase *epsilon* subunit, one long and one short. The *Helobdella robusta* (leech) genome has a single isoform containing a 283 bp intron (highlighted in yellow) that, when spliced out, produces an mRNA similar in size to the Enchytraeidae short transcript (171:177 bp). Trans-splice-site codon (isoleucine) is in red, and down-facing arrowhead indicates where that splice site is in the Enchytraeidae. Intra-specific variants are identical up to base pair 163 (green), as well as the final 14 base pairs (gray), while the long isoform has an additional 24 or 27 bp situated in between (underlined). Upward-facing arrowhead indicates a possible additional splice site. The ice worm has only a single isoform, 9 bp longer than the longest of its congeners; the insertions are highlighted in blue.

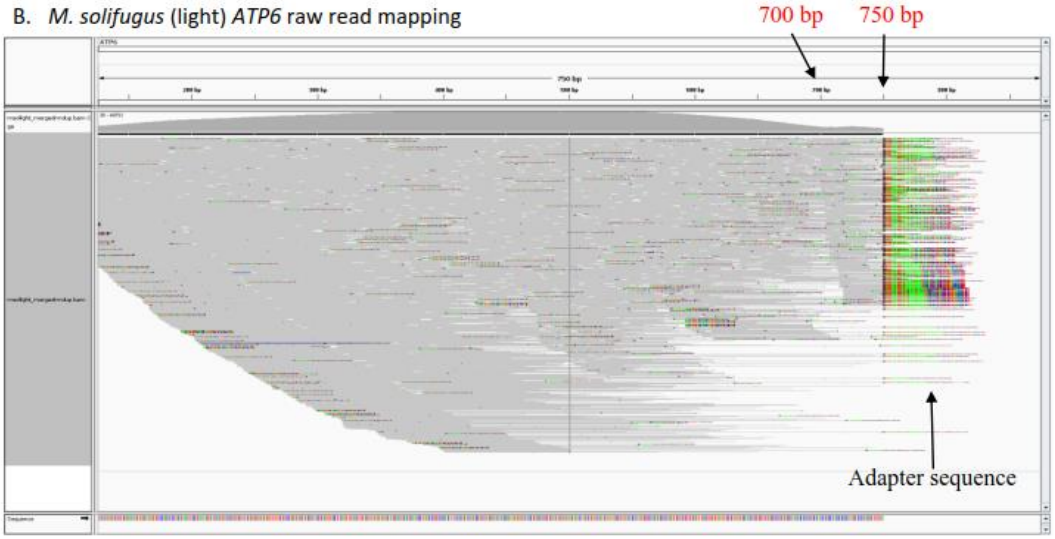
Verification of *Mesenchytraeus solifugus atp6* extension

Transcriptome raw read mapping to the *M. solifugus atp6* transcript extracted from the transcriptomes indicates that the ~50 bp extension is not likely to be an assembly error (Fig. 15). The two ice worm transcriptomes, *M. solifugus*-dark (A) and *M. solifugus*-light (B) were independently generated and assembled, yet the same *atp6* transcript was found in each. The 700 bp mark identifies where all other worm *atp6* transcripts end, but the ice worm sequence extended to 750 bp where it ended at the sequencing adapter. Furthermore, the depth of transcript coverage at that point was still strong. As an additional test, the ice worm extension was appended to the *M. hydrius atp6* transcript and used as a raw read mapping template. (C) Raw read mapping of the *M. hydrius*-*M. solifugus* transcript showed no transcripts extending beyond 700 bp.

A. *M. solifugus* (dark) *ATP6* raw read mapping



B. *M. solifugus* (light) *ATP6* raw read mapping



C. *Mesenchytraeus hydrius* transcriptome raw read mapping using *M. solifugus ATP6* extension

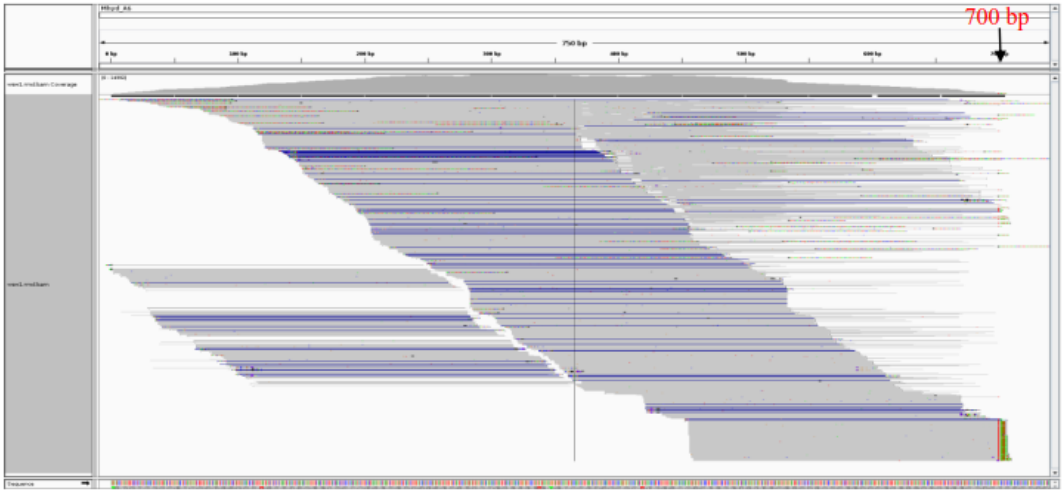


Figure 15. Transcriptome raw reads were mapped to Trinity-assembled *atp6* transcripts derived from two independently-generated ice worm (*M. solifugus*) transcriptomes, (A) dark and (B) light. Both show deep coverage of the 3'-end extension from 700-750 bp and ending with the RNA-seq adapters. In (C), the ice worm 3'-extension was added to the *M. hydrius atp6* transcript before mapping; all transcripts end at 700 bp, thus no extension is present.

To further ensure the authenticity of the extension, PCR amplifications of full-length *atp6* were performed from genomic DNA template using a forward primer designed to amplify all considered species, and an ice worm extension-specific reverse primer (Fig. 16A). In addition to the six considered species, two additional ice worms from widely-separated populations (Davidson Glacier, Mariner Mt. Glacier) were evaluated to determine whether the *atp6* extension was common to ice worms, or an isolated population anomaly (Fig. 16B). Figure 16C shows the translated protein alignment of the PCR-amplified *atp6* sequences.

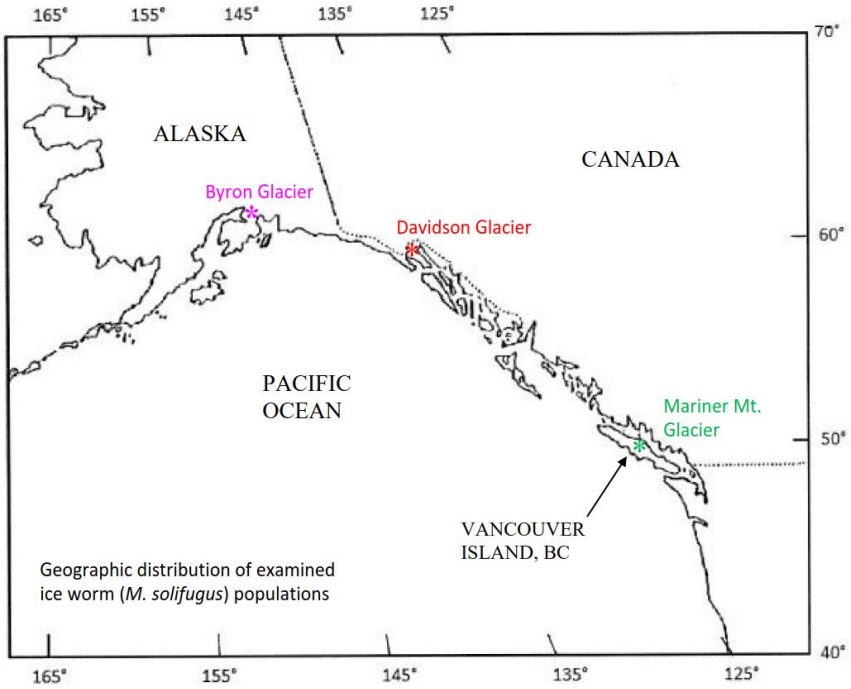
For all species, other than ice worm, the extension primer failed to amplify, but robust amplification occurred when the same forward primer was paired with a species-specific 3'-end primer. Mariner Mt. ice worms displayed a 3'-end extension of the same length as the Byron Glacier population, differing only in a single aa (V/P). Davidson Glacier ice worms, members of the putative ancestral clade (Dial et al., 2012), displayed a truncated version of the extension – the final 13 aa, with a single conserved change (S/A), but lacking 6 aa upstream. Note that we were unable to make a primer downstream of the extension as this region could not be amplified from gDNA (See Fig. 7A), and is not present in any of our transcriptomes. Difficulties amplifying this

region are well known in the annelid community (Boore and Brown, 2000; Boore, 2001; Jennings and Halanych, 2005; Bleidorn et al., 2006a; Bleidorn et al., 2006b).

A.

```
                    Atp6_start primer
                    ATGATACCWGATATTTTYTCMTCATTYGAYCC
Ecry  1 ATGATACCTGATATTTTCTCATCATTTGACCCATC…//…GACCACGCACAT---------- 693
Mped  1 ATGATGCCTGATATCTTTTCCTCATTCGACCCATT…//…GACCACGTTTTTTCT---------- 696
Mhyd  1 ATGATACCTGATATTTTCTCCTCATTCGACCCATT…//…GACCACGCGCACAGTTTT---------- 699
Mant  1 ATGATACCTGATATTTTTTCATCATTCGATCCATT…//…GATCACGCCCATGTTTTT---------- 699
Mgel  1 ATGATACCTGATATTTTTTCCTCCTTCGATCCATT…//…GACCACGCCCATGTTTTT---------- 699
MsDav 1 ATGATACCTGATATTTTCTCATCATTTGACCCTTT…//…GACCACGCTCAAGCT----------CACGCCCCAGGTCATGCCCACAGTCACGGTTT 729
MsByr 1 ATGATACCAGATATTTTTTCATCATTTGATCCTTT…//…GACCACGCTCACGTCCAAGACCGCTCCCAAGCCCACTCACCAGGTCATGCCCACAGTCACGGTTT 747
MsMMt 1 ATGATACCAGATATTTTTTCATCATTTGACCCTTT…//…GACCATGCTCACCCCCAAGACCGCTCCCAAGCTCACTCCCCAGGTCATGCCCACAGTCACGGTTT 747
                                                                    TCATGCCCACAGTCACGGTTTTAA
                                                                    Ms_Atp6_end primer
```

B.



C.

```
Ecry  1 MMPDIFSSFDPSSFNSLIPSMTIMFMINIIYMFLVDRNIW…//…GYLLFEVAICMIQGYIFFLLLTLYANDHAH---------- 231
Mped  1 MMPDIFSSFDPFMFNTLVPYNSMFLTINLFLILLMQSSFW…//…GYILFEVAICMIQAYIFCLLLSLYSDDHVFS---------- 232
Mhyd  1 MMPDIFSSFDPFMFNSVVPSNSVFLCINILLILFMQSSFW…//…GYLLFEVAICMIQAYIFCLLLSLYSDDHAHSF---------- 233
Mant  1 MMPDIFSSFDPFMFNTLVPSNSVFLCINVMLILFMQSSFW…//…GYLLFEVAICMIQAYIFCLLLSLYSDDHAHVF---------- 233
Mgel  1 MMPDIFSSFDPFMFNTLIPSNSIFLCINVLLILFMQSSFW…//…GYLLFEVAICMIQAYIFCLLLSLYSDDHAHVF---------- 233
MsDav 1 MMPDIFSSFDPFMFNSVVPSNSIFLCINILLILSMQSSFW…//…GYLLFEVAICMIQAYIFCLLLSLYSDDHA------QAHAPGHAHSHGF 243
MsByr 1 MMPDIFSSFDPFMFNSVVPSNSIFLCINVLLILSMQSSFW…//…GYLLFEVAICMIQAYIFCLLLSLYSDDHAHVQDRSQAHSPGHAHSHGF 249
MsMMt 1 MMPDIFSSFDPFMFNSVVPSNSIFLCINVLLILSLQSSFW…//…GYLLFEVAICMIQAYIFCLLLSLYSDDHAHPQDRSQAHSPGHAHSHGF 249
```

Figure 16. Verification of *Mesenchytraeus solifugus Atp6* 3'-extension in different ice worm populations. Full-length *atp6* sequence was PCR amplified from genomic DNA using an Enchytraeid-specific forward primer and the ice worm-specific 3'-extension primer and/or a

species-specific end primer (A). In addition to the Northern clade (Byron Glacier, MsByr) ice worms used throughout this study, two additional ice worm populations, Davidson Glacier (MsDav), considered ancestral (Dial et al., 2012), and Mariner Mountain Glacier (MsMMt) were also examined (B). The 3'-extension primer amplified only ice worm samples, but the forward primer amplified for all when paired with species-specific end primers. The translated protein alignment (C) shows the conserved block of amino acids (highlighted) found in all three ice worm populations, but not found in any other species. The 3'-extension is 6 aa shorter in the Davidson population. The alignment in (A) is shown with stop codons removed.

Evolutionary origin of the ATP6 extension

The histidine-rich motif present in the ice worm ATP6 suggested some functional importance, as histidine plays a significant role in protein interactions and catalysis due to its ability to be both a proton donor and acceptor within a physiological pH range. Protein and nucleotide Blast searches against our transcriptomes using the 50-base pair extension as query failed to produce any matches other than *M. solifugus atp6*, so it does not appear to be endogenous in origin. However, protein Blasts against the NCBI nr-database proved more fruitful (Fig. 17). Functionally diverse histidine-rich proteins, primarily from prokaryotes, share ≥ 68% identity to the ice worm ATP6 extension motif. Interestingly, one species, *Aureobasidium subglacialis*, is a glacier-dwelling yeast, and among members of the genus *Thalassospira*, are species living in cold deep-sea niches. It is well-established that glacier-obligate microorganisms are the primary food source for ice worms (Goodman, 1971), and recently published work identified a new genus and species of bacteria living as endosymbionts in ice worm gut (Murakami et al., 2014). These data suggest that the novel ATP6 extension found in the ice worm may have been acquired by horizontal gene transfer.

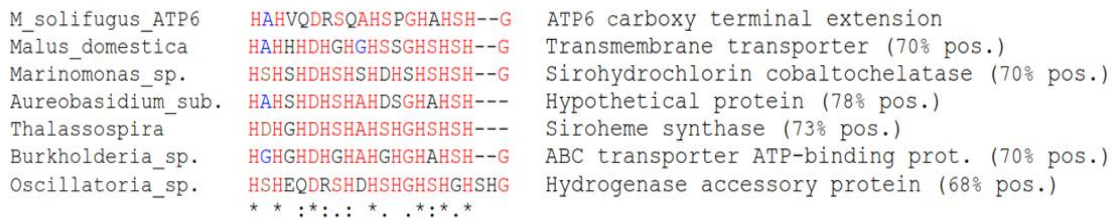
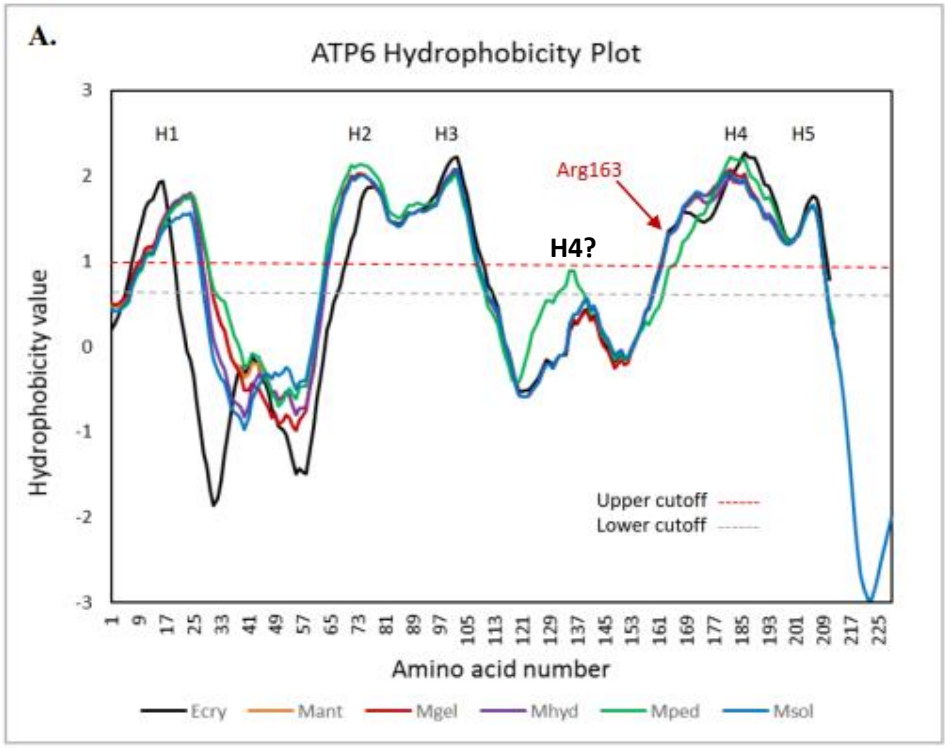


Figure 17. Protein alignment of NCBI nr-pBLAST results using *Mesenchytraeus solifugus* ATP6 C-extension as query. Histidine-rich domains are involved in numerous cellular processes, but membrane transport is a well-represented function. *Malus domestica* (apple), *Marinomonas* (marine bacteria), *Aurobasidium subglaciale* (glacier fungus), *Thalassospira* (marine bacterioplankton), *Burkholderia* (beta proteobacteria), *Oscillatoria* (filamentous cyanobacteria).

Secondary structure analysis of ATP6

Hydrophobicity plots for prediction of secondary structure indicated that only five membrane α-helices are present in the worm ATP6 (Fig. 18A). Recent studies in alga, yeast and bovine (Allegretti et al., 2015; Zhou et al., 2015; Hahn et al., 2016) using cryo-EM show six transmembrane helices for ATP6, so the weak peak between H3 and H4 is likely a real membrane α-helix. The hydrophilic C-terminal is predicted to be in the mitochondrial matrix, while helix 4(5) containing the highly-conserved arginine necessary for proton transfer, and helix 5(6), are proximal to the c-ring subunit within the inner membrane. The N-terminal extends slightly into the IMS. (B) Molecular model of *M. solifugus* ATP6; helices are red, sheets are yellow and disordered loops are gray. Arginine163 is shown in blue and the ice worm-specific non-conserved G>T substitution is in pink. Ala82 (green) is the N-terminal and Asp227 is the C-terminal. Residues 1-81 and 228-249 were not present in the PDB template file (1C17).

Collectively, these data provide unequivocal evidence that the 18 aa C-terminal extension of the ice worm ATP6 subunit is not artefactual, but rather functions in association with the ice worm ATP synthase holo-complex. Moreover, this fragment of genetic information appears to have been acquired by lateral gene transfer from a psychrophilic microorganism, either through dietary consumption or via an endosymbiotic relationship.


A.
ATP6 Hydrophobicity Plot
H1
H2
H3
H4
H5
Arg163
H4?
Upper cutoff
Lower cutoff
Hydrophobicity value
Amino acid number
Ecry
Mant
Mgel
Mhyd
Mped
Msol

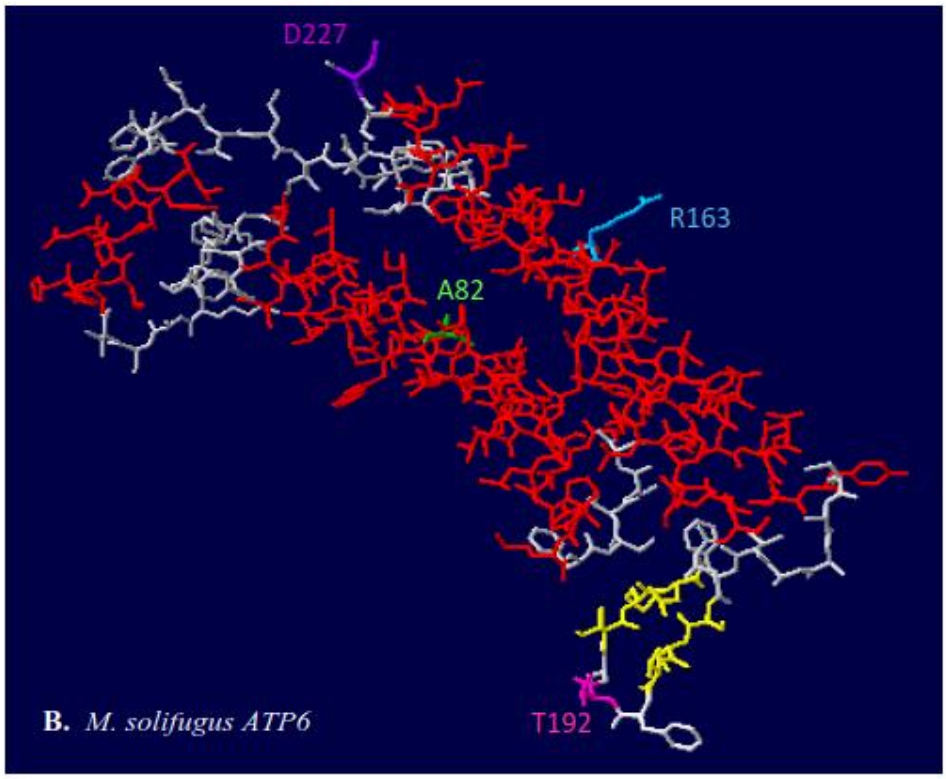

D227
R163
A82
T192
B. M. solifugus ATP6

Figure 18. ATP6 secondary structure prediction and model. (A) ATP6 hydrophobicity plot for the considered species. Peaks above the upper cutoff are predicted transmembrane helices. Recent Cryo-EM studies in alga, yeast and bovine show 6 transmembrane helices for ATP6, so the weak peak between H3 and H4 is likely a real membrane α-helix. The hydrophilic C-terminal is predicted to be in the matrix, while helix 4(5) containing the highly-conserved Arg necessary for proton transfer, and helix 5(6), are proximal to the c-ring subunit within the inner membrane. B. Molecular model of *M.solifugus* ATP6*;* helices are red, sheets are yellow and disordered loops are gray. Arg163 is shown in blue and the ice worm-specific non-conserved G>T substitution is in pink. Ala82 (green) is the N-terminal and Asp227 is the C-terminal; amino acids 1-81 and 228-249 were not present in the PDB template file (1C17).

Adaptive selection analyses (d_N/d_S)

Despite the number and diversity of changes found in the ATP synthase subunits, tests of selection (sites and branch/sites) found only a single change that achieved a p-value indicative of positive selection. Position 185 of the mature Gamma aa sequence had a Bayes Empirical Bayes probability of $\omega > 1 = 0.957$ with a post-mean of 2.794 ± 1.879 SE; this value conforms to a 95% confidence interval. This site was not included in Table 6 because it was a multi-variable site – the mesenchytraeids had either S or A, a highly-conserved interchange, while *E. crypticus* had an N – conserved relative to S. The failure to find evidence of positive selection is not surprising. Firstly, d_N/d_S is a very conservative measure of adaptation, thus changes must be very pronounced to be detected. Secondly, our datasets were very small (7 species) and closely-related, thus the differences between them were few and lacked the statistical power gained from large datasets. Lastly, this type of analyses strictly discards sites with missing data, thus indels have no effect on outcome.

Discussion

The current study is the first molecular phylogenetic reconstruction of an unusually diverse clade of Pacific Northwest-endemic mesenchytraeid worms that inhabit disparate ecological regimes (from temperate soil to glacier ice). This approach employed tree estimates of 30 individual genes, five concatenated supermatrices, and two multi-species coalescent datasets. *A priori* divergence rates were not known among the considered species, thus the genes chosen for analyses span a wide conservation range. Clearly, these results indicate that housekeeping genes are too highly conserved to provide resolution for these closely-related species. Individually, these six genes (*actin, α-tubulin, EF-1α, GAPDH, histone H3, 28S*) produced six topologically-distinct trees, often with very high branch support, and the concatenated dataset yielded yet a seventh topology (Fig. 8). Gene tree discord is a commonly observed phenomenon attributed to incomplete lineage sorting (Pamilo and Nei, 1988; Maddison, 1997; Degnan and Rosenberg, 2009), mitochondrial DNA introgression (Chan and Levin, 2005), horizontal gene transfer (Kidwell, 1993; Maddison, 1997), gene duplication/loss (Fitch, 1970; Goodman et al., 1979), and adaptive selection (Bazin et al., 2006; Toews and Brelsford, 2012). Incomplete lineage sorting (ILS) occurs when DNA sequence coalescence is temporally out of step with speciation, a situation more often found in species separated by relatively few genetic changes (Maddison, 1997). Thus, given the short branch lengths observed and the extraordinary level of discordance, the failure of

the housekeeping gene dataset to resolve these interspecific relationships is likely attributable to ILS.

Likewise, the ATP synthase subunit (ASU) genes produced discordant gene trees – eight topologies for nine genes – and ILS may underlie this lack of agreement as well, but other factors merit consideration. Generally, gene products functioning in critical metabolic pathways are under strong negative (purifying) selection because almost all mutations that arise are deleterious. This is certainly the case for ASU Alpha and Beta with > 78% amino acid identity across diverse metazoan phyla. However, nuclear-encoded oxidative phosphorylation gene products that functionally interact with faster-evolving mitochondrial-encoded proteins experience additional selective pressure. This 'mito-nuclear' association creates an "arms race" whereby the nuclear genes must adapt to, and compensate for, mutations in mitochondrial genes to maintain optimum energy production (Bayona-Bafaluy et al., 2005; Mishmar et al., 2006; Bar-Yaacov et al., 2012). Therefore, ASU gene comparisons from species evolving in ecologically diverse niches may support relationships incongruent with true phylogenetic relationships (Hill, 2016).

Mitochondrial genomes have been considered the best dataset for discernment of species level relationships because of their ~10-fold greater mutation rate, lack of recombination, conserved gene content, and the perception that they evolve under a neutral equilibrium model. While the first three attributes are true in most cases, it has become clear that selection plays a non-trivial role in mt-genome evolution and these effects can confound phylogenetic interpretations (Bazin et al., 2006; Dowling et al., 2008; Galtier et al., 2009; Toews and Brelsford, 2012). Although BI analysis of the

concatenated mitochondrial genes (Mt13 + RNAs, Fig. 6) produced a highly supported (posterior probabilities ≥0.99) bifurcating tree, 10 of the 15 individual gene trees had polytomies and much lower support values.

This dichotomy between a well-supported concatenated multi-gene tree and discordant individual gene trees is not surprising. Numerous studies have demonstrated that phylogenetic inference based on concatenation of genes evolving under different coalescent properties can produce incorrect, but well supported trees (Degnan and Rosenberg, 2006; Edwards et al., 2007; Kubatko and Degnan, 2007). This effect is particularly prevalent in species-level analyses where few genetic changes separate the taxa (Kubatko and Degnan, 2007). Recognition of this problem has led to coalescent-based methodologies that model emergent relationship properties within a group of individual gene trees to produce a more realistic species tree (Edwards et al., 2007; Heled and Drummond, 2010; Yang and Rannala, 2010). Unfortunately, even this framework failed to resolve relationships among the mesenchytraeids, as demonstrated by the multifurcating consensus tree (Fig. 9A), and the considerable ambiguity present in the full set of trees (Fig. 9B). This failure may be due to limitations of the MSC in that gene tree incongruities due to factors other than ILS (e.g. hybridization, adaptive selection) are not accounted for.

Among the proposed causative factors of gene tree/species tree discord, incomplete lineage sorting is a likely agent, given the short branch lengths within the mesenchytraeid clade, and their extraordinary habitat diversity suggests adaptive selection may also play a role. Horizontal gene transfer, thought to be primarily confined to prokaryotes, is increasingly found to be the source of novel traits in a

variety of eukaryotes (Moran and Jarvik, 2010; Boschetti et al., 2012; Boto, 2014; Crisp et al., 2015) and cannot be dismissed. Finally, mitochondrial introgression resulting from gene flow (i.e. hybridization) between species may explain the considerable incongruities seen in the mitochondrial datasets and the inefficacy of the MSC to resolve species relationships. Interestingly, mitochondrial introgression has a putative role in rapid adaptive radiations (Grant et al., 2005) and energy metabolism (Boratyński et al., 2011). Indeed, the overwhelmingly discordant trees, short branch lengths and overall equidistant divergence rates among the mesenchytraeids examined here argue persuasively for a recent and rapid adaptive radiation event. Furthermore, the proximity of their niches would have enhanced opportunities for hybridization during nascent speciation.

The *Mesenchytraeus* species considered in this study are highly endemic within the cool, wet ecozones characterizing the Pacific Northwest. Approximately 85 species of *Mesenchytraeus* have been described worldwide (Schmelz and Collado 2012), and although many occur in colder, subarctic climates (Piper et al. 1992, Birkemoe et al. 2000), glacier ice worms are known only from the western coast of North America (Tynan 1970b, Hartzell et al. 2005), and a single location on the Tibetan Plateau (Liang et al. 1979). Snow worms, too, are predominantly found in the PNW, though a few undescribed mesenchytraeid snow worms have been found in eastern Asia (Murakami et al., 2015). The putative predisposition of these worms to ice/snow habitats is not known, but unusually high ATP levels observed in glacier ice worms suggest that enhanced energy metabolism is a critical component of this adaptation (Napolitano and Shain 2004; Napolitano et al. 2004). This notion is further supported by experiments

that correlate elevated intracellular ATP levels with cold tolerance (Southard et al., 1985; English and Storey, 2000; Morrison and Shain, 2008; Amato and Christner, 2009; Marotta et al., 2009, Parry and Shain, 2011).

Morphological structure of nephridia – organs that regulate osmolarity in worms – strongly suggest that mesenchytraeids arose relatively recently from an aquatic ancestor (Christensen and Glenner, 2010), thus species that invaded terrestrial habitats (e.g., soil, snow) were likely derived from an aquatic worm. In the context of this data, an aquatic stock with high standing variation would permit a source from which terrestrial, snow and ice worms could evolve. Note that the latter (*M. solifugus*) are fully aquatic, residing permanently between ice crystals in liquid-filled channels that characterize temperate, maritime glaciers (Patterson, 1994; Nye, 1999), while snow worms (*M. hydrius, M. gelidus*) inhabit a similar microenvironment for part of the year before residing in moist, temperate soil for the remainder (comparable with the terrestrial, *M. anteaus*).

Mountain building in the Pacific northwestern region occurred 5-10 mya (late Miocene to early Pliocene (Hammond, 1979; Parrish, 1983; Montgomery, 2000), thus the worms considered in this study are unlikely to predate this time period. Periodic glacial cycles and volcanic activity characterize the region (Bjornstad, 2006), and the continuity of glacier ice/snow is supported by the persistence of ice worms (Dial et al., 2012), which are a stenothermic, glacially-obligate species (Tynan, 1970a, b; Goodman, 1971; Shain et al., 2001). Further, documented aquatic refugia have been described throughout this region and time period, providing a genetic pool from which the diverse mesenchytraeids could have arisen (Bernatchez and Wilson, 1998; Wilke

and Duncan, 2004; Steele and Storfer, 2006; Shafer et al., 2010). Previous analyses estimate ice worm origins between ~4.9 - 9 mya (Dial et al., 2012), which may approximate the time frame of the mesenchytraeid radiation proposed here. Geological uplifting in the region is hypothesized to have created diverse ecozones into which mesenchytraeids invaded, and thus these two events (i.e., mountain building/ mesenchytraeid radiation) are likely coincident. Regardless of timing, however, these data support a short time window of rapid speciation in which environmental conditions were favorable for a cold adapted, aquatic mesenchytraeid to invade the open habitats of glacier ice/snow and soil (Fig.19).

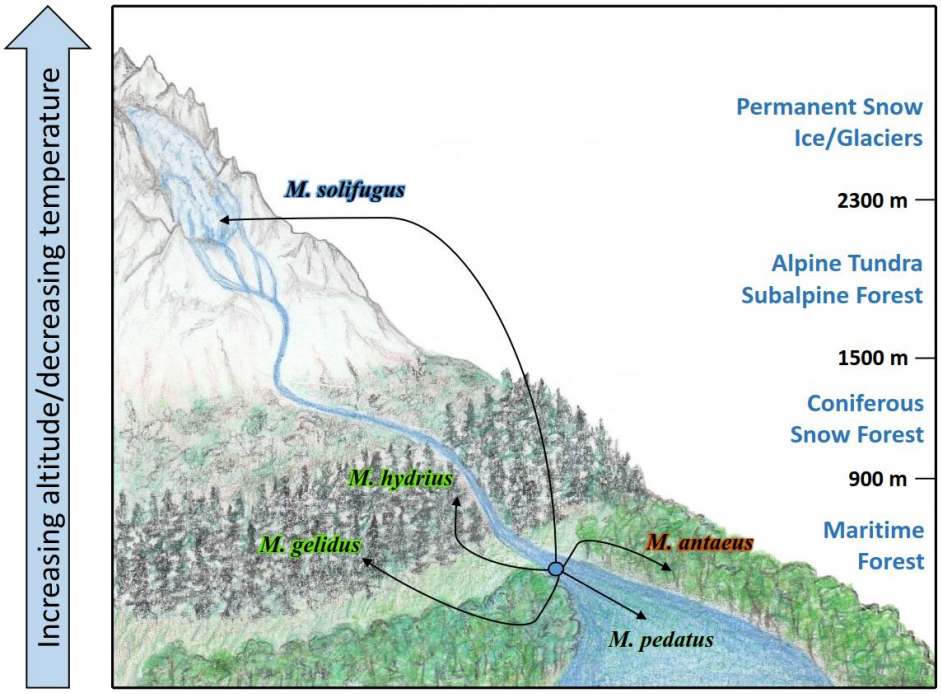


Figure 19. **Proposed radiation of *Mesenchytraeus* species along an altitudinal gradient from an aquatic ancestor.**

Their origins notwithstanding, extant *Mesenchytraeus* species in the Pacific Northwest offer a unique animal group to dissect the relationship between genes and

the environment, particularly in the context of habitat diversity and thermal regimes. We anticipate that life histories of other regional mesenchytraeids, and possibly other psychrophilic fauna in the Pacific Northwest, may share a similar evolutionary paradigm.

The invasion of *M. solifugus* into glacier ice above the equilibrium line altitude (ELA) is a novelty among Animalia with few known exceptions, e.g., a Himalayan glacier ice worm (Liang et al., 1979) and Icelandic glacier rotifers (Shain et al., 2016). An important component of this adaptation within *M. solifugus* appears to be unusually high levels of intracellular ATP levels (see Fig. 10), which paradoxically increase as temperatures fall (Napolitano et al., 2004). This phenomenon has also been reported in other cold-adapted taxa (Napolitano and Shain, 2005; Morrison and Shain, 2009; Amato and Christner, 2009) and appears to be a metabolic signature of psychrophily. Importantly, genetic manipulations of *Escherichia coli* to reflect a similar energetic profile (i.e., elevated ATP) led to enhanced cold tolerance (Morrison and Shain, 2009; Parry and Shain, 2011). Collectively, these observations target intracellular ATP production as a primary mechanism associated with cold adaption, and possibly as a general strategy for coping with thermal stress.

Specifically, the F_1F_O-ATP synthase complex is the major cellular machinery responsible for ATP production under aerobic conditions across domains of life. Importantly, the electrochemical gradient across the inner mitochondrial membrane that drives ATP synthesis is not a diffusion-dependent process (e.g., as in glycolysis), and therefore is less susceptible to changes in temperature in comparison with other

ATP synthetic processes. Taken together, ATP synthase is a logical target for an organism to modify in the context of energy metabolism and cold adaptation.

Interestingly, most subunits of the ATP synthase complex display very high conservation across the disparate mesenchytraeids examined here with regard to overall sequence similarity, size, isoelectric point, and hydrophobicity (see Tables 5, S2). Regarding the amino acid substitutions, in general, they conform to a standard paradigm of cold-adapted proteins in terms of increased flexibility at the expense of thermal stability (Smalas et al., 2000; Siddiqui and Cavicchioli, 2006; Feller, 2010). A few exceptions are notable, however, and may underlie the ice worm's paradoxical ability to sustain atypically high ATP levels at cold physiological temperatures.

Most strikingly, the ice worm ATP6 subunit diverges dramatically from its congener's subunits by having acquired an unusual histidine-rich, 18 amino acid C-terminal extension: VQDRSQAHSPVHAHSHGF. This modification is particularly noteworthy because ATP6 forms the critical half-channels permitting the proton flux that drives c-ring rotation (Allegretti et al., 2015; Hahn et al., 2016; Kühlbrandt and Davies, 2016), and thus ATP synthesis. Moreover, a trend toward mitochondrial genome compaction is well supported among metazoans (i.e., insertions are generally not tolerated; Selosse et al., 2001; Adams and Palmer, 2002), and the ice worm C-terminal extension, which significantly increases its length, hydrophilicity, and molecular weight, is unprecedented among metazoan ATP6 sequences currently deposited in genome databases.

Collectively, its presence and maintenance across geographically distinct ice worm populations over geological time (see Fig. 16B), coupled with enhanced ATP

production in the species, appears to be more than coincidental. In addition, sequence similarity between the ATP6 extension, and a domain found in cold-adapted bacteria, suggests that it may have been acquired via horizontal gene transfer (HGT). The explosion of whole genome sequences available today has firmly established that HGT, once thought to only occur in bacteria, is an ancient and continuing process across all domains of life (Crisp et al., 2015; Soucy et al., 2015). Furthermore, research in this area shows that the most common sources of HGT in eukaryotes are endosymbionts and diet (Doolittle,1998; Yue et al., 2013; Grant and Katz, 2014), and the majority of genes transferred function in metabolic pathways (Crisp et al., 2015). It has also been noted that organisms occupying extreme environments are enriched in foreign genes, and these genes play a role in ecological adaptation (Keeling and Palmer, 2008; Schönknecht et al., 2014).

The significance of these facts regarding HGT is that all of these conditions are met in the glacier ice worm, *M. solifugus* – it lives on glaciers, its diet is primarily composed of glacier microbes and fungi (Goodman, 1972), it harbors a bacterial endosymbiont (Murakami et al., 2014), and it has an unusual metabolic profile (i.e., high ATP). Thus, the conjecture that the novel ATP6 extension derives from HGT is a reasonable explanation for the data. A point of contention is that ATP6 is a mitochondrial-encoded gene, and HGT seems limited to nuclear integration. However, HGT is widespread in plant mitochondria and chloroplasts (Marienfeld et al., 1999), and the metazoan coral, *Sarcophyton glaucum*, carries a bacterial DNA repair gene in its mitochondrial genome (Pont-Kingdon et al., 1998; McFadden et al., 2006).

The putative role of the ice worm ATP6 extension remains undetermined, but its position and physiochemical properties suggest a few possible functions. First, its projection into the mitochondrial matrix coincides with a region known to facilitate ATP synthase dimerization (Poumard et al., 2003; Wittig and Schägger, 2008). The repetitive His residues contained within the ice worm C-terminal extension iterate the sequence of other worm ATP6 C-termini (i.e. HAH, see fig. 16C), and therefore may amplify the dimerization signal through interactions with other accessory proteins. Secondly, the linear sequence of the extension and regularly-spaced His residues share properties with several ion-transport domains found in other species (Hara, et al., 2005; Nishida et al., 2008). In this scenario, local positive charges may be spatially oriented to deflect incoming H^+ ions away from the pore, so as to increase the local pH thereby increasing H^+ ion flow (Fig. 20). Importantly, the average pH of the mitochondrial matrix is ~7.7 and the isoelectric point of histidine is 7.6, thus within that environment, the His residues will rapidly fluctuate between protonated and deprotonated.

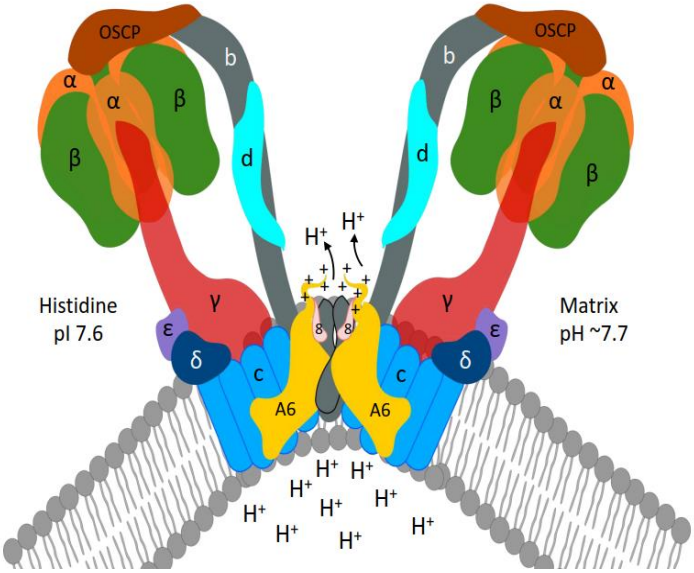


Figure 20. The histidine-rich ATP6 extension may function as a proton shuttle.

Regardless of the putative function or mechanism of action of the ice worm ATP6 extension, it is noteworthy that most human mitochondrial diseases attributed to ATP synthase dysfunction are caused by mutations within the ATP6 subunit (Brandon, et al., 2006, and references therein), and a single point mutation can enable a bacterium to thrive in a highly alkaline environment (Wang et al., 2004). Clearly, ATP6 is a sensitive component of the F_1F_O ATP synthase machinery that strongly correlates with bioenergetic efficiency. In effect, ATP6 can be thought of as the genetic 'throttle' for the ATP synthase machine, whereby small sequence tweaks can modulate energy performance and output in a changing environment.

In addition to anomalies of the ATP6 subunit, the ice worm Epsilon subunit displays a short C-terminal extension and an upstream serine insertion. Likewise, ice worm-specific aa substitutions were clustered in SynB and Gamma subunits (Fig. 13) such that structural changes appear likely within a region proximal to ATP6 and the dimer interface. Taken together, it appears that ice worm evolution has targeted the inner matrix region of the F_O pore for structural modification, which likely influences the efficacy of the ATP synthase complex at cold temperature. Note that ice worms have accumulated 46 species-specific non-conserved aa substitutions within the ATP synthase complex, more than 2.5 times any of its congeners (Table 6), and thus have undergone a period of accelerated evolution since diverging from the other mesenchytraeids 5-10 mya.

The evolutionary relationships established here provide a platform for exploring the molecular adaptations that have facilitated genetically-related worms to occupy disparate environmental niches, ranging from temperate soil to glacier ice. The

current study has identified significant modifications in key regulatory subunits of the ice worm F_1F_O ATP synthase complex (e.g., C-terminal extension of ATP6, epsilon insertions, conserved and variable domains within individual subunits). Manipulating these and other subunits of the ATP synthase complex in a tractable eukaryotic system will lead to a mechanism for enhancing mitochondrial ATP synthesis, and potentially correcting energy-based deficiencies in a number of human conditions.

Summary and Conclusions

Deep nuclear/mitochondrial sequencing followed by extensive phylogenetic analyses revealed the punctuated and simultaneous appearance of at least five *Mesenchytraeus* species between 5-10 million years ago that currently occupy water, soil, snow and ice, respectively, along altitudinal transects in the PNW. Comparisons of intracellular ATP levels among these worms demonstrated that the glacier ice worm consistently maintained ATP well above its congeners, even though it resides permanently in ice at 0˚C. Strikingly, the ice worm mitochondrial ATP6 subunit of the F_1F_o- ATP synthase encoded a highly basic, 18 amino acid carboxy-terminal extension likely to have been acquired by lateral gene transfer from an ancestral prokaryote. The position and biochemical properties of the extension domain suggest a role in ATP synthase dimerization and/or proton shuttling, both of which would predictably enhance ATP production.

References

Abrahams, J.P., Leslie, A.G.W., Lutter, R., Walker, J.E. (1994) Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* 370: 621-628.

Adams, K.L. and Palmer, J.D. (2003) Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol* 29: 380-395.

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6): 716–723, doi:10.1109/TAC.1974.1100705, MR 0423716.

Allegretti, M., Klusch, N., Mills, D.J., Vonck, J., Kühlbrandt, W., Davies, K.M. (2015) Horizontal membrane-intrinsic α-helices in the stator a subunit of an F-type ATP synthase. *Nature*. 521: 237–240.

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

Arnold, I., Pfeiffer, K., Neupert, W., Stuart, R.A., Schägger, H. (1998) Yeast mitochondrial F_1F_0-ATP synthase exists as a dimer: identification of three dimer-specific subunits. *EMBO J* 17: 7170–7178.

Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195-201.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., Stockinger, H. (2012) ExPASy: SIB bioinformatics resource portal, *Nucleic Acids Res* 40(W1):W597-W603.

Baena-González, E. (2010) Energy signaling in the regulation of gene expression during stress. *Mol Plant* 3(2): 300-13. doi: 10.1093/mp/ssp113.

Bar-Yaacov, D., Blumberg, A., Mishmar, D. (2012) Mitochondrial-nuclear co-evolution and its effects on OXPHOS activity and regulation. *Biochim Biophys Acta* 1819: 1107-1111.

Bayona-Bafaluy, M.P., Muller, S., Moraes, C.T. (2005) Fast adaptive evolution of nuclear and mitochondrial subunits of ATP synthase in Orangutan. *Mol Biol Evol* 22(3): 716-724.

Bazin, E., Glemin, S., Galtier, N. (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science* 312: 570–572.

Benkert, P., Biasini, M. and Schwede, T. (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27: 343-350.

Bernatchez, L. and Wilson, C.C. (1998) Comparative phylogeography of nearctic and palearctic fishes. *Mol Ecol* 7: 431–452.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 42 (W1): W252-W258; doi: 10.1093/nar/gku340.

Birkemoe, T., Coulson, S.J., Somme, L. (2000) Life cycles and population dynamics of Enchytraeids (Oligochaeta) from the high arctic. *Can J Zool* 78: 2079-2086.

Bjornstad, B.N. (c. 2006). On the trail of the Ice Age floods: a geological field guide to the mid-Columbia basin. Sandpoint, Idaho: Keokee Books. p. 4.

Bleidorn, C., Podsiadlowski, L., Bartolomaeus, T. (2006a) The complete mitochondrial genome of the orbiniid polychaete *Orbinia latreillii* (Annelida, Orbiniidae) – A novel gene order for Annelida and implications for annelid phylogeny. *Gene* 370: 96–103.

Bleidorn, C., Kruse, I., Albrecht, S., Bartolomaeus, T. (2006b) Mitochondrial sequence data expose the putative cosmopolitan polychaete *Scoloplos armiger* (Annelida, Orbiniidae) as a species complex. *BMC Evol Biol* 6: 47.

Bomfleur, B., Kerp, H., Taylor, T., Moestrup, O., Taylor, E. (2012) Triassic leech cocoon from Antarctica contains fossil bell animal. *PNAS* 109: 20971–20974.

Boore, J.L., Brown, W.M. (1995) Complete sequence of the mitochondrial DNA of the Annelid worm *Lumbricus terrestris*. *Genetics* 141(1): 305-319.

Boore, J.L. and Brown, W.M. (2000) Mitochondrial genomes *of Galathealinum, Helobdella, and Platynereis*: sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. *Mol Biol Evol* 17(1): 87-106.

Boore, J.L. (2001) Complete Mito chondrial Genome Sequence of the Polychaete Annelid *Platynereis dumerilii*. *Mol Biol Evol* 18 (7): 1413-1416.

Boratyński, Z., Alves, P.C., Berto, S., Koskela, E., Mappes, T., Melo-Ferreira, J. (2011) Introgression of mitochondrial DNA among *Myodes* voles: consequences for energetics? *BMC Evol Biol* 11:355. DOI: 10.1186/1471-2148-11-355.

Borda, E. and Siddall, M.E. (2004) Arhynchobdellida (Annelida: Oligochaeta: Hirudinida): phylogenetic relationships and evolution. *Mol Phylogenet Evol* 30: 213-25.

Boto, L. (2014) Horizontal gene transfer in the acquisition of novel traits by metazoans. *Proc R Soc B* 281: 20132450.

Boschetti, C., Carr, A., Crisp, A., Eyres, I., Wang-Koh, Y., Lubzens, E. et al. (2012) Biochemical diversification through foreign gene expression in bdelloid rotifers. *PLoS Genet* 8:e1003035.

Bouckaert, R.R. and Heled, J. (2014) DensiTree 2: Seeing Trees Through the Forest. bioRxiv, http://dx.doi.org/10.1101/012401.

Boyer, P.D. (1997) The ATP synthase – a splendid molecular machine. *Annu Rev Biochem* 66: 717-749.

Brandon, M., Baldi, P., Wallace, D.C. (2006) Mitochondrial mutations in cancer. *Oncogene* 25: 4647-4662.

Carpenter, E. P., Beis, K., Cameron, A. D., & Iwata, S. (2008). Overcoming the challenges of membrane protein crystallography. *Curr Opin Struc Biol 18*(5), 581–586. http://doi.org/10.1016/j.sbi.2008.07.001.

Chaban, Y., Boekema, E.J., Dudkina, N.V. (2014) Structures of mitochondrial oxidative phosphorylation supercomplexes and mechanisms for their stabilization. *Biochim Biophys Acta* 1837: 418-426.

Chan, K.M. and Levin, S.A. (2005) Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution* 59(4): 720-9.

Christensen, B. and Glenner, H. (2010) Molecular phylogeny of Enchytraeidae (Oligochaeta) indicates separate invasions of the terrestrial environment. *J Zool Syst Evol Res* doi: 10.1111/j.1439-0469.2009.00558.x.

Clarke, A. (2003) Costs and consequences of evolutionary temperature adaptation. *Trends Ecol Evol* 18:11, 573-581.

Claros, M.G. and von Heijne, G. (1994) TopPred II: An improved software for membrane protein structure predictions. *CABIOS* 10: 685-686.

Claros, M.G., Vincens, P. (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* 241, 779-786.
Cogliati, S., Enriquez, J.A., Scorrano, L. (2016) Mitochondrial cristae: where beauty meets functionality. *Trends Biochem Sci* 41(3) 261 – 273.

Connell, J., and Orias, E. (1964) The ecological regulation of species diversity. *Am Nat* 98:399–414.

Crisp, A., Boschetti, C., Perry, M., Tunnacliffe, A., Micklem, G. (2015) Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. Genome Biol 16: 50. doi 10.1186/s13059-015-0607-3.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E. (2004) WebLogo: A sequence logo generator, *Genome Res* 14:1188-1190.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9(8): 772.

David, J.R., Araripe, L.O., Chakir, M., *et al.* (2005) Male sterility at extreme temperatures: a significant but neglected phenomenon for understanding *Drosophila* climatic adaptations. *J Evol Biol* 18: 838–846.

Davies, K., Anselmi, C., Wittig, I., Faraldo-Gomez, J., and Kühlbrandt, W. (2012) Structure of the yeast F_1F_o-ATP synthase dimer and its role in shaping mitochondrial cristae. *Proc Natl Acad Sci* USA, 109(34): 13602-7.

Degnan, J.H., Rosenberg, N.A., (2006) Discordance of species trees with their most likely gene trees. *PLOS Genet* 2: 762–768.

DeVries, A.L., Komatsu, S.K., Feeney, R.E. (1970) Chemical and physical properties of freezing point-depressing glycoproteins from Antarctic fishes. *J. Biol. Chem* 245(11): 2901–8.

Dial, C.R., Dial, R.J., Saunders, R., Lang, S.A., Lee, B, Wimberger, P., Dinapoli, M.S., Egiazarov, A.S., Gipple, S.L., Maghirang, M.R., Swartley-M.cArdle, D.J., Yudkovitz, S.R., Shain, D.H. (2012) Historical biogeography of the North American glacier ice worm, *Mesenchytraeus solifugus* (Annelida: Oligochaeta: Enchytraeidae). *Mol Phylogenet Evol* 63(3): 577-84. doi: 10.1016/j.ympev.2012.01.008.

Dionisi-Vici, C., Seneca, S., Zeviani, M., et al. (1998) Fulminant Leigh syndrome and sudden unexpected death in a family with the T9176C mutation of the mitochondrial ATPase 6 gene. *J Inherit Metab Dis* 21: 2–8.
Doolittle, W.F. (1998) You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet* 14: 307-311.

Dowling, D.K., Friberg, U., Lindell, J., (2008) Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends Ecol Evol* 23: 546-554.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A. (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4: e88.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7 *Mol Biol Evol* 29: 1969-1973.

Dudkina, N.V., Heinemeyer, J., Keegstra, W., Boekema, E. J., Braun, H. P. (2005) Structure of dimeric ATP synthase from mitochondria: an angular association of monomers induces the strong curvature of the inner membrane. *FEBS Lett* 579: 5769–5772.

Duvezin-Caubet, S., Caron, M., Giraud, M-F., Velours, J., di Rago, J-P. (2003) The two rotor components of yeast mitochondrial ATP synthase are mechanically coupled

by subunit δ. *Proc Natl Acad Sci* 100(23): 13235–13240. doi: 10.1073/pnas.2135169100.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5): 1792-1797.

Edwards, J.S., (1986) How small ectotherms thrive in the cold without really trying. *Cryo Lett* 6: 388–390.

Edwards, S.V., Liu, L., Pearl, D.K. (2007) High-resolution species trees without concatenation. *Proc Natl Acad Sci USA* 104:5936-5941.

Eisen, G. (1904) Enchytraeidae of the West Coast of North America. In Harriman Alaska Expedition, v. 12 The Annelids, Smithsonian Institution, Washington.

Emery, C. (1898) Diagnosi di un nuovi genore e nuovi specie di Annellidi della famiglia deli Enchytraeidae. R. Accademia nazionale dei Lincei. Classe di scienze fisishe matematiche e naturali.

English, T. E., and Storey, K. B. (2000) Enzymes of adenylate metabolism and their role in hibernation of the white-tailed prairie dog, *Cynomys leucurus*. *Arch Biochem Biophys* 376: 91-100.

Erséus, C., Rota, E., Matamoros, L., De Wit, P. (2010) Molecular phylogeny of Enchytraeidae (Annelida, Clitellata). *Mol Phylogen Evol* 57: 849-858.

Etzold, C., Deckers-Hebestreit, G., Altendorf, K. (1997) Turnover number of *Escherichia coli* F_OF_1- ATP synthase for ATP synthesis in membrane vesicles. *Eur J Biochem*, 243(1-2):336-343.

Eubel, H., Jansch, L., Braun, H.P. (2003) New insights into the respiratory chain of plant mitochondria: supercomplexes and a unique composition of complex II. *Plant Physiol* 133: 274-286.

Feller, G. and Gerday, C. (2003) Psychrophilic enzymes: hot topics in cold adaptation. *Nat Rev Microbiol* 1: 200–8.

Felsenstein, J. (1985) Phylogenies and the comparative method. *Am Nat* 125(1):1-15.

Fernández, R., Laumer, C.E., Vahtera, V., Libro, S., Kaluziak, S., Sharma, P.P., Pérez-Porro, A.R., Edgecombe, G.D., Giribet, G. (2014) Evaluating topological conflict in centipede phylogeny using transcriptomic data sets. *Mol Biol Evol* 31: 1500–1513. doi:10.1093/molbev/msu108.

Feynman, R.P. (1965) *The Character of Physical*, 28. Quoted in William H. Cropper, *Great Physicists* (2004), 397.

Fitch, W.M. (1970) Distinguishing homologous from analogous proteins. *Syst Zool* 19: 99-113.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 3: 294–299.

Forster, J.R. (1778) Observations made during a voyage round the world, on physical geography, natural history and ethic philosophy. G. Robinson. London.

Foster, D.L. and Fillingame, R. H. (1982) Stoichiometry of subunits in the H^+-ATP synthase complex of *Escherichia coli*. *J Biol Chem* 257: 2009-2015.

Fujisawa, M., Fackelmayer, O., Liu, J., Krulwich, T. and Hicks, D. (2010) The ATP synthase a-subunit of extreme alkaliphiles is a distinct variant: mutations in the critical alkaliphile-specific residue Lys-180 and other residues that support alkaliphile oxidative phosphorylation. *J Biol Chem* 285(42), 32105-15.

Fukasawa, Y., Tsuji, J., Fu, S-C., Tomii, K., Horton, P., Imai, K. (2015) MitoFates: Improved Prediction of Mitochondrial Targeting Sequences and Their Cleavage Sites. *Mol Cell Proteomics* 14(4): 1113-1126, 2015

Galtier, N., Nabholz, B., Glemin, S., Hurst G.D.D. (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol Ecol* 18: 4541-4550.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server. (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005). pp. 571-607.

Gerday, C., Aittaleb, M., Bentahir, M., *et al*. (2000) Cold-adapted enzymes: from fundamentals to biotechnology. *Trends Biotechnol* 18: 103–107.

Gershoni, M., Templeton, A.R., Mishmar, D. (2009) Mitochondrial bioenergetics as a major motive force of speciation. *Bioessays* 31: 642-650.
Goodman, D. (1971) Ecological investigations of ice worms on Casement Glacier, Southeast Alaska. Ohio State Univ. Res. Foundation, Institute of Polar Studies, Report 39:1-59.

Gibbons, A. (2010) Human Evolution. Tracing evolution's recent fingerprints. *Science* 329(5993): 740-2.

Gillespie, J.H. (1984) Molecular evolution over the mutational landscape. *Evolution* 38: 1116–1129.

Goodman, D. (1971) Ecological investigations of ice worms on Casement glacier, Southeastern Alaska. Ohio State Univ. Res. Found., Institute of Polar Studies. Report 39.

Goodman, M., Czelusniak, J., Moore, G.W., Romero-Herrera, A.E., Matsuda G. (1979) Fitting the gene lineage into its species lineage, a parsimony strategy illustrated by cladograms constructed from globin sequences. *Syst Zool* 28:132-163.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A. (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol*. 29(7): 644-52. doi: 10.1038/nbt.1883.

Grant, J.R. and Katz, L.A. (2014) Phylogenomic study indicates widespread lateral gene transfer in Entamoeba and suggests a past intimate relationship with parabasilids. *Genome Biol Evol* 6: 2350-2360.

Grant, P.R., Grant, B.R., Petren, K. (2005). Hybridization in the recent past. *Amer Nat* 166: 56–67.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59(3): 307-21.

Guindon, S. and Gascuel, O. (2003). A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Syst Biol* 52: 696-704.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. Aug;8(8): 1494-512. doi: 10.1038/nprot.2013.084.

Hahn, A., Parey, K., Bublitz, M., Mills, D.J., Zickermann, V., Vonck, J., Kühlbrandt, W., Meier, T., (2016) Structure of a complete ATP synthase dimer reveals the molecular basis of inner mitochondrial membrane morphology. *Molec Cell*, 63: 445-456. doi.org/10.1016/j.molcel.2016.05.037.

Hammond, P.E. (1979), "A tectonic model for evolution of the Cascade Range", in Armentrout, J.M., et al. The Cenozoic paleogeography of the Western United States, Society of Economic Paleontologists and Mineralogists, pp. 219–237.

Hancock, A.M., Witonsky, D.B., Gordon, A.S., et al. (2008) Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genet* 4(2): e32. doi:10.1371/journal.pgen.0040032.

Hara, K.Y., Kato-Yamada, Y., Kikuchi, Y., Hisabori, T., Yoshida, M. (2001) The role of the βDELSEED motif of F_1-ATPase: propagation of the inhibitory effect of the ε subunit. *J Biol Chem* 276: 23969–23973.

Hara, M., Fujinaga, M., Kuboi, T. (2005) Metal binding by citrus dehydrin with histidine-rich domains. *J Exp Bot* 56(420): 2695–2703.

Hartzell, P.L., Nghiem, J.V., Richio, K.J., Shain, D.H. (2005) Distribution and phylogeny of glacier ice worms (*Mesenchytraeus solifugus* and *Mesenchytraeus solifugus rainierensis*). *Can J Zool* 83(9):1206-1213, 10.1139/z05-116.

Havlíčková, V., Kaplanová, V., Nůsková, H., Drahota, Z., Houštěk, J. (2010) Knockdown of F(1) epsilon subunit decreases mitochondrial content of ATP synthase and leads to accumulation of subunit c. *Biochim. Biophys. Acta* 1797: 1124-1129.

Heled, J. and Drummond, A.J. (2010) Bayesian inference of species trees from multilocus data. *Mol Biol Evol* 27(3): 570–580.

Henikoff, S. and Henikoff, J.G. (1992). Amino Acid Substitution Matrices from Protein Blocks. *Proc Natl Acad Sci* 89 (22): 10915–10919.

Hill, G.E. (2016) Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecol Evol* 6:5831–5842. doi: 10.1002/ece3.2338.

Huelsenbeck, J.P., Larget, B., Alfaro, M.E. (2004) Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Mol Biol Evol* 21: 1123-1133.

Hoffmann, A.A., Sørensen, J.G., Loeschcke, V. (2003) Adaptation of *Drosophila* to temperature extremes: bringing together quantitative and molecular approaches. *J Therm Biol* 28: 175–216.

Holt, I.J., Harding, A.E., Petty, R.K., Morgan-Hughes, J.A. (1990) A new mitochondrial disease associated with mitochondrial DNA heteroplasmy. Am J Hum Genet 46: 428-433.

Hutchinson, G. (1959) Homage to Santa Rosalia, or why are there so many kinds of animals? *Am Nat* 93:145–159.

Jennings, R.M. and Halanych, K.M. (2005) Mitochondrial Genomes of *Clymenella torquata* (Maldanidae) and *Riftia pachyptila* (Siboglinidae): Evidence for Conserved Gene Order in Annelida. *Mol Biol Evol* 22(2): 210-222.

Jonckheere, A.I., Hogeveen, M., Nijtmans, L.G., et al. (2008) A novel mitochondrial ATP8 gene mutation in a patient with apical hypertrophic cardiomyopathy and neuropathy. *J Med Genet* 45: 129–133.

Jonckheere, I. and Smeitink, J. (2012) Mitochondrial ATP synthase: architecture, function and pathology. *J Inherit Metab Dis* 35: 211-225.

Jonckheere, A.I., Renkema, G.H., Bras, M., van den Heuvel, L.P., Hoischen, A., Gilissen, C., Nabuurs, S.B., Huynen, M.A., de Vries, M.C., Smeitink, J.A.,

Rodenburg, R.J. (2013) A complex V ATP5A1 defect causes fatal neonatal mitochondrial encephalopathy. *Brain* 136(Pt 5): 1544-54.

Jung, S.H., Lee, J.Y., Lee, D.H. (2003) Use of SAGE technology to reveal changes in gene expression in *Arabidopsis* leaves undergoing cold stress. *Plant Mol Biol* 52(3): 553-67.

Keeling, P.J. and Palmer, J.D. (2008) Horizontal gene transfer in eukaryotic evolution. *Nature Rev Genet* 9: 605-618.

Kidwell, M.G. (1993) Lateral transfer in natural populations of eukaryotes. *Annu Rev Genet* 27:235-256.

Kruuv, J., Glofcheski, D., Cheng, K.H., Campbell, S.D., Al-Qysi, H.M., Nolan, W.T., Lepock, J.R. (1983). Factors influencing survival and growth of mammalian cells exposed to hypothermia. I. Effects of temperature and membrane lipid perturbers. *J Cell Physiol* 115:179–185.

Kubatko, L.S. and Degnan, J.H. (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst Biol* 56(1): 17-24.

Kühlbrandt, W. and Davies, K.M. (2016) Rotary ATPases: a new twist to an ancient machine. *Trends Biochem Sci* 41: 106-116.

Lanfear, R., Calcott, B., Ho, S., Guindon, S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol* 29: 1695–1701.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10. doi:10.1186/Gb-2009-10-3-R25.

Larsson, S.G. (1978) Baltic Amber - A palaeontological study. Entomograph, Vol. I. Scandinavian Scientific Press, Klampenborg.

Lee, J., Ding, S., Walpole, T.B., Holding, A.N., Montgomery, M.G., Fearnley, I.M., Walker, J.E. (2015) Organization of subunits in the membrane domain of the bovine F-ATPase revealed by covalent cross-linking. *J Biol Chem* 290: 13308–13320.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009)

The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25: 2078-9. [PMID: 19505943].

Li, H. and Durbin, R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler Transform. *Bioinformatics*, Epub. [PMID: 20080505].

Liang, Y., Hsu, C., Chang, T. (1979) A new genus and species of Enchytraeidae from Tibet. *Acta Zootax Sinica* 4: 312-317.

Lieber, D.S., Calvo, S.E., Shanahan, K., Slate, N.G., Liu, S., Hershman, S.G., Gold, N.B., Chapman, B.A., Thorburn, D.R., Berry, G.T., Schmahmann, J.D., Borowsky, M.L., Mueller, D.M., Sims, K.B., Mootha, V.K. (2013) Targeted exome sequencing of suspected mitochondrial disorders. *Neurology* 80(19): 1762-70.

Lissarre, M., Ohta, M., Sato, A., & Miura, K. (2010) Cold-responsive gene regulation during cold acclimation in plants. *Plant Signal Behav* 5(8): 948–952. doi.org/10.4161/psb.5.8.12135.

Liu, L. and Pearl, D.K. (2007) Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Syst Biol* 56: 504-514.

Littell, J.S., McGuire, M., Elsner, L.C. Whitely Binder, and A.K. Snover (eds). (2009) The Washington climate change impacts assessment: Evaluating Washington's future in a Changing climate, Climate Impacts Group, University of Washington, Seattle, Washington.

Luyet, B.J. and Gehenio, P.M. (1940) *Life and Death at Low Temperatures*. Biodynamica, Normandy, MO.

Maddison, W.P. (1997) Gene trees in species trees. *Syst Biol* 46: 523–536.

Maddison, W.P., Midford, P.E., Otto, S.P. (2007) Estimating a binary character's effect on speciation and extinction. *Syst Biol* 56(5): 701-710.

Maddison, W.P. and Maddison, D.R. (2015) Mesquite: a modular system for evolutionary analysis, Version 3.04. http://mesquiteproject.org.

Malecki, M., Bitton, D.A., Rodríguez-López, M., Rallis, C., Calavia1, N.G., Smith, G.C., Bähler, J. (2016) Functional and regulatory profiling of energy metabolism in fission yeast. *Genome Biol* 17: 240.

Marienfeld, J., Unseld, M., Brennicke, A. (1999) The mitochondrial genome of *Arabidopsis* is composed of both native and immigrant information. *Trends Plant Sci* 4(12): 495-502.

Marotta, R., Parry, B.R., Shain D.H. (2009) Divergence of AMP deaminase in the ice worm *Mesenchytraeus solifugus* (Annelida, Clitellata, Enchytraeidae). *Int J Evol Biol* Volume 2009, Article ID 715086.

Martins, E.P. and Hansen, T.F. (1997) Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am Nat* 149(4):646-667.

Matsuno-Yagi, A. and Hatefi, Y. (1988) Estimation of the turnover number of bovine heart F_1F_o complexes for ATP synthesis. *Biochemistry* 27:14 748-14752.

Mayr, J.A., Havlickova, V., Zimmermann, F., et al. (2010) Mitochondrial ATP synthase deficiency due to a mutation in the *ATP5E* gene for the F1 epsilon subunit. *Hum Mol Genet* 19: 3430–3439.

McFadden, C.S., France, S.C., Sánchez, J.A., Alderslade, P. (2006) A molecular phylogenetic analysis of the Octocorallia (Cnidaria: Anthozoa) based on mitochondrial protein-coding sequences. *Mol Phylogenet Evol* 41(3): 513-27.

Miller, M.A., Pfeiffer, W., Schwartz, T. (2010) "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" in Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1-8.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., … Wallace, D.C. (2003). Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci* USA 100(1): 171–176. doi.org/10.1073/pnas.0136972100.

Mishmar, D., Ruiz-Pesini, E., Mondragon-Palomino, M., Procaccio, V., Gaut, B., Wallace, D.C. (2006) Adaptive selection of mitochondrial complex I subunits during primate radiation. *Gene* 378:11–18.

Mitchell, P. (1961) Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* 191 (4784): 144–148.

Montgomery, D.R. (2000) Coevolution of the Pacific salmon and Pacific Rim topography. *Geology* 28(12): 1107-1110.

Moran, N.A. and Jarvik, T. (2010) Lateral transfer of genes from fungi underlies carotenoid production in ahids. *Science* 328: 624-627.

Morava, E., Rodenburg, R.J., Hol, F., et al. (2006) Clinical and biochemical characteristics in patients with a high mutant load of the mitochondrial T8993G/C mutations. *Am J Med Genet A* 140: 863–868.

Morrison, B.A. (2007) Elevated ATP levels increase cold tolerance in *Escherichia coli*. Rutgers The State University of New Jersey, MS Thesis.

Morrison, B.A. and Shain, D.H. (2008) An *AMP nucleosidase* gene knockout in *Escherichia coli* elevates intracellular ATP levels and increases cold tolerance. *Proc R Soc Lond B* (Suppl) 4: 53-56.

Mühleip, A.W., Joos, F., Wigge, C., Frangakis, A.S., Kühlbrandt, W., Davies, K.M. (2016) Helical arrays of U-shaped ATP synthase dimers form tubular cristae in ciliate mitochondria *Proc Natl Acad Sci* USA 113: 8442-8447.

Murakami, T., Segawa, T., Bodington, D., Dial, R., Takeuchi, N., Kohshima, S., Hongoh, Y. (2015) Census of bacterial microbiota associated with the glacier ice worm *Mesenchytraeus solifugus*. *FEMS Microbiol Ecol* 91. doi: 10.1093/femsec/fiv003.

Napolitano, M.J., Nagele, R.O. and Shain, D.H. (2004) The ice worm, *Mesenchytraeus solifugus*, elevates adenylate nucleotides at low physiological temperature. *Comp Biochem Physiol* Part A 137: 227-235.

Napolitano, M.J. and Shain, D.H. (2004) Four kingdoms on ice: convergent energetic processes boost energy levels at low physiological temperatures. *Proc R Soc Biol B* (Suppl.) 271: S273-S276.

Nina, P.B., Dudkina, N.V., Kane, L.A., van Eyk, J.E., Boekema, E.J., Mather, M.W., et al. (2010) Highly Divergent Mitochondrial ATP Synthase Complexes in *Tetrahymena thermophila*. *PLoS Biol* 8(7): e1000418.

Nishida, S., Mizuno, T., Obata, H. (2008) Involvement of histidine-rich domain of ZIP family transporter TjZNT1 in metal ion specificity. *Plant Physiol Biochem* 46(5-6): 601-6. doi: 10.1016/j.plaphy.2008.02.011.

Nye, J.F. (1999) Lecture notes on water in ice: microscopic and geophysical scales. In: Wettlaufer, J.S., Dash, J.G., Untersteiner, N. (Eds.), Ice Physics and the Natural Environment. Springer International Publishing AG, pp. 273–279.

Odokonyero, D., Sakai, A., Patskovsky, Y., Malashkevich, V.N., Fedorov, A.A., Bonanno, J.B., Fedorov, E.V., Toro, R., Agarwal, R., Wang, C., Ozerova, N.D.S., Yew, W.S., Sauder, J.M., Swaminathan, S., Burley, S.K., Almo, S.C., Glasner, M.E. (2014) Loss of quaternary structure is associated with rapid sequence divergence in the OSBS family. *Proc Natl Acad Sci* USA 111(23): 8535-8540.

Pamilo, P., Nei, M. (1988) Relationships between gene trees and species trees. *Mol Biol Evol* 5: 568–583.

Parrish, R.R. (1983) Cenozoic thermal evolution and tectonics of the Coast Mountains of British Columbia: 1. Fission track dating, apparent uplift rates, and patterns of uplift. *Tectonics* 2: 601-631.

Parry, B.R. and Shain, D.H. (2011) Manipulations of AMP metabolic genes increase growth rate and cold tolerance in *Escherichia coli*: implications for psychrophilic evolution. *Mol Biol Evol* doi:10.1093/molbev/msr038.

Parry, L., Tanner, A., Vinther, J. (2014) The origin of annelids. *Palaeontology* 57: 1091-1103.

Patterson, W.S.B. (1994) *The Physics of Glaciers*, third ed. Elsevier, Oxford, New York and Tokyo, pp. 480.

Piper, S.R., MacLean, S.F., Christensen, B. (1982) Enchytraeidae (Oligochaeta) from the taiga and tundra habitats of northeastern U.S.S.R. *Can J Zool* 60: 2594-2609.

Pont-Kingdon, G., Okada, N.A., Macfarlane, J.L., Beagley, C.T., Watkins-Sims, C.D., Cavalier-Smith, T., Clark-Walker, G.D., Wolstenholme, D.R. (1998)

Mitochondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue of bacterial *MutS*: a possible case of gene transfer from the nucleus to the mitochondrion. *J Mol Evol* 46(4): 419-31.

Paumard, P., Vaillier, J., Coulary, B., Schaeffer, J., Soubannier, V., Mueller, D.M., Brèthes, D., di Rago, J.P., Velours, J. The ATP synthase is involved in generating mitochondrial cristae morphology. *EMBO J* 21: 221–230.

Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J. (2014) Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer.

Rannala, B.H. and Yang, Z. (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 164: 1645–1656.

Rezende, E.L. and Diniz-Filho, J.F.A. (2012) Phylogenetic analyses: comparing species to infer adaptations and physiological mechanisms. *Compr Physiol* 2:639-674.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P. (2011) Integrative Genomics Viewer. *Nature Biotechnol* 29: 24–26.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61(3):539-42. doi: 10.1093/sysbio/sys029.

Rota, E. and Brinkhurst, R.O. (2000) *Mesenchytraeus antaeus*, a new giant enchytraeid (Annelida, Clitellata) from the temperate rainforest of British Columbia, Canada, with a revised diagnosis of the genus *Mesenchytraeus*. *J Zool*, London 252: 27-40.

Rühle, T. and Leister, D. (2015) Assembly of F_1F_o-ATP synthases. *Biochim Biophys Acta Bioenergetics*, 1847(9): 849-860. http://dx.doi.org/10.1016/j.bbabio.2015.02.005.

Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V., Wallace, D.C. (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* 303(5655): 223-226.

Ruud, J.T. (1954) Vertebrates without erythrocytes and blood pigment. *Nature* 173(4410): 848–850.

Savojardo, C., Martelli, P.L., Fariselli, P., Casadio, R. (2014) TPpred2: improving the prediction of mitochondrial targeting peptide cleavage sites by exploiting sequence motifs. *Bioinformatics* 30(20): 2973-2974.

Schagger, H. and Pfeiffer, K. (2000) Supercomplexes in the respiratory chain of yeast and mammalian mitochondria. *EMBO J* 19: 1777-1783.

Schmelz, R.M., and Collado, R. (2010) A guide to European terrestrial and freshwater species of Enchytraeidae (Oligochaeta). *Soil Organisms*, 82: 1-176.

Schmelz, R.M., and Collado, R. (2012) An updated checklist of currently accepted species of Enchytraeidae (Oligochaeta, Annelida). *vTI Agriculture and Forestry Research*, Special Issue 357; Newsletter on Enchytraeidae No. 12. 67-87.

Schmelz, R.M., and Collado, R. (2015) Checklist of taxa of Enchytraeidae (Oligochaeta): an update. *Soil Organisms*, 87(2), 149-152.

Schmidt, T.R., Wu, W., Goodman, M., Grossman, L.I. (2001) Evolution of nuclear- and mitochondrial-encoded subunit interaction in cytochrome c oxidase. *Mol Biol Evol* 18(4):563-569.

Schönknecht, G., Weber, A.P., Lercher, M.J. (2014) Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. *Bioessays* 36: 9 –20.

Selosse, M-A., Albert, B., Godelle, B. (2001) Reducing the genome size of organelles favours gene transfer to the nucleus. *Trends Ecol Evolut* 16(3): 135-141.

Shafer, A.B.A., Cullingham, C.I., Cote, S.D., Coltman, D.W. (2010) Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Mol Ecol* 19: 4589-4621.

Shain, D.H., Mason, T.A., Farrell, A.H., Michalewicz, L.A. (2001) Distribution and behavior of ice worms (*Mesenchytraeus solifugus*) in south-central Alaska. Canadian Journal of Zoology 79(10): 1813-1821, 10.1139/z01-143.

Shain, D. H. (2009). Annelids in modern biology. Hoboken, N.J.: Wiley-Blackwell.

Shain, D.H., Halldórsdóttir, K., Pálsson, F., Aðalgeirsdóttir, G., Gunnarsson, A., Jónsson, Þ., Lang, S.A., Pálsson, H.S., Steinþórssson, S., Arnason, E. (2016) Colonization of maritime glacier ice by bdelloid Rotifera. *Mol Phylogenet Evol* 98: 280-287. doi: 10.1016/j.ympev.2016.02.020.

Shaw, K.L. (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proc Natl Acad Sci USA* 99:16122-16127.

Siddiqui, K.S. and Cavicchioli, R. (2006) Cold-adapted enzymes. *Annu Rev Biochem* 75:403-433.

Simmons, M.P., Norton, A.P. (2014) Divergent maximum-likelihood-branch-support values for polytomies.
Mol Phylogenet Evol 73: 87-96. doi: 10.1016/j.ympev.2014.01.018.

Somero, G.N. (2004) Adaptation of enzymes to temperature: searching for basic "strategies" *Comp Biochem Physiol B* 139:321–33.

Soucy, S.M., Huang, J., Gogarten, J.P. (2015) Horizontal gene transfer: building the web of life. *Nature Rev Genet* 16: 472-482.

Southard, J.H., Rice, M.J., Ametani, M.S., Belzer, F.O. (1985) Effects of short-term hypothermic perfusion and cold storage on function of the isolated-perfused do kidney. *Cryobiology* 22: 147-155.

Smalas, A.O., Leiros, H.K., Os, V. et al., (2000) Cold adapted enzymes. *Biotechnol Annu Rev* 6: 1–57.

Standke, G. (1978) Die Tertiarprofile der Samlandischen Bernsteinkuste bei Rauschen. *Schriftenreihe Geowiss* 7: 93-103.

Steele, C.A. and Storfer, A. (2006) Coalescent-based hypothesis testing supports multiple Pleistocene refugia in the Pacific Northwest for the Pacific giant salamander (*Dicamptodon tenebrosus*). *Mol Ecol* 15: 2477-2487.

Stephens, A.N., Roucou, X., Artika, I.M., Devenish, R.J., and Nagley, P. (2000) Topology and proximity relationships of yeast mitochondrial ATP synthase subunit 8 determined by unique introduced cysteine residues. *Eur J Biochem* 267: 6443–6451.

Stock, D., Leslie, A.G., Walker, J.E. (1999) Molecular architecture of the rotary motor in ATP synthase. *Science* 286: 1700-1705.

Sullivan, J., Swofford, D.L., Naylor, G.J.P. (1999) The effect of taxon sampling on estimating rate heterogeneity parameters of maximum-likelihood models. *Mol Biol Evol* 16: 1347–1356.

Tamura, K; Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30: 2725-2729.

Tatuch, Y., Robinson, B.H. (1993) The mitochondrial DNA mutation at 8993 associated with NARP slows the rate of ATP synthesis in isolated lymphoblast mitochondria. *Biochem Biophys Res Commun* 192: 124–128.

Tavare, S. (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures on Mathematics in the Life Sciences (American Mathematical Society) 17: 57–86.

Taylor, W. P. (1922) A distributional and ecological study of Mount Rainier, Washington. *Ecology* 3(3): 214–236. http://doi.org/10.2307/1929036.

Thauer, R. and Brendel, W. (1962) Hypothermie *Prog Surg* 2:73–271.

Tian, D., Wang, Q., Zhang, P., Araki, H., Yang, S., Kreitman, M., Nagylaki, T., Hudson, R., Bergelson, J., Chen, J.Q. (2008) Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. *Nature* 455(7209):105–108.

Toews, D.P.L., Brelsford, A. (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Mol Ecol* 21: 3907–3930.

Trounce, I., Neill, S., Wallace, D.C. (1994) Cytoplasmic transfer of the mtDNA nt 8993T–>G (ATP6) point mutation associated with Leigh syndrome into mtDNA-less cells demonstrates cosegregation with a decrease in state III respiration and ADP/O ratio. *Proc Natl Acad Sci USA* 91: 8334–8338.

Tu, Q., Yu, L., Zhang, P., et al. (2000) Cloning, characterization and mapping of the human *ATP5E* gene, identification of pseudogene ATP5EP1, and definition of the *ATP5E* motif. *Biochem J* 347: 17–21.

Tynan, M.J. (1970a) Worms on ice. *Nature* 225: 587.

Tynan, M.J. (1970b) The geographical distribution of ice worms (Oligochaeta: Enchytraeidae). *Can J Zool* 48: 1363-1367.

Tzen, C.Y., Wu, T.Y., Liu, H.F. (2001) Sequence polymorphism in the coding region of mitochondrial genome encompassing position 8389-8865. *Forensic Sci Int* 120: 204-209.

Ulrich, H. and Schmelz, R.M. (2001) Enchytraeidae as prey of Dolichopodidae, recent and in Baltic amber. *Bonn Zool Beitr* 50: 89-101.

Waits, D.S., Santos, S.R., Thornhill, D.J., Li, Y., Halanych, K.M. (2016) Evolution of sulfur binding by hemoglobin in Siboglinidae (Annelida) with special reference to bone-eating worms, *Osedax. J Mol Evol* 82: 219-229. doi 10.1007/s00239-016-9739-7.

Walker, J.E., Saraste, M., Runswick, M.J., Gay, N.J. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J*. 1(8): 945–51.

Walker, J.E. (1998) ATP synthesis by rotary catalysis. *Angew Chem Int Ed* 37: 5000-5011.

Walpole, T.B., Palmer, D.N., Jiang, H., Ding, S., Fearnley, I.M., Walker, J.E. (2015) Conservation of complete trimethylation of lysine-43 in the rotor ring of c-subunits of metazoan adenosine triphosphate (ATP) synthases. *Mol Cell Proteomics* 14: 828 - 840.

Wang, Z., Hicks, D.B., Guffanti, A.A., Baldwin, K., Krulwich, T.A. (2004) Replacement of amino acid sequence features of a- and c-subunits of ATP synthases of alkaliphilic Bacillus with the Bacillus consensus sequence results in defective oxidative phosphorylation and non-fermentative growth at pH 10.5. *J Biol Chem* 279: 26546–26554.

Ware, S.M., El-Hassan, N., Kahler, S.G., et al. (2009) Infantile cardiomyopathy caused by a mutation in the overlapping region of mitochondrial ATPase 6 and 8 genes. *J Med Genet* 46: 308–314.

Welch, P.S. (1916) Snowfield and glacier Oligochaeta from Mt. Rainier. *Science* 43: 143.

Welch, P.S. (1919) Further studies on North American Mesenchytraeids (Oligochaeta). *Trans. Am. Microscop Soc*. 38: 175-188.

Wilke, T., Duncan, N. (2004) Phylogeographical patterns in the American Pacific Northwest: lessons from the arionid slug *Prophysaon coeruleum. Mol Ecol* 13: 2303–2315.

Wittig, I. and Schägger, H. (2008) Structural organization of mitochondrial ATP synthase. *Biochim Biophys Acta* 1777: 592–598.

Wu, M.T., Chatterji, S., Eisen, J.A. (2012) Accounting for alignment uncertainty in phylogenomics. *PLoS ONE* 7, 30288. doi:10.1371/journal.pone.0030288.
Yampolsky, L.Y., Zeng, E., Lopez, J., Williams, P.J., Dick, K.B., Colbourne, J.K., Pfrender, M.E. (2014) Functional genomics of acclimation and adaptation in response to thermal stress in *Daphnia*. *BMC Genomics* 15: 859.

Yang, J.R., Liao, B.Y., Zhuang, S.M., Zhang, J. (2012) Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *Proc Natl Acad Sci USA* 109(14): E831–E840.

Yang, Z. (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* 39: 306–314.

Yang, Z. (2007) PAML 4: A program package for phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24: 1586-1591.

Yang, Z. and Rannala, B. (2010) Bayesian species delimitation using multilocus sequence data. *Proc Natl Acad Sci USA*, 107: 9264–9269

Zachariassen, K.E. (1991). Hypothermia and cellular physiology. *Arctic Med Res* 50(6): 13–17.

Zhang, C., Rannala, B., Yang, Z. (2012) Robustness of compound Dirichlet priors for Bayesian inference of branch lengths. *Syst Biol* 61: 779-784.

Zhang, Z., Huang, J., Wang, Z., Wang, L., Gao, P. (2011) Impact of indels on the flanking regions in structural domains. *Mol Biol Evol* 28(1):291–301.

Zhou, A., Rohou, A., Schep, D.G., Bason, J.V., Montgomery, M.G., Walker, J.E., Grigorieff, N., Rubinstein, J.L. (2015) Structure and conformational states of the bovine mitochondrial ATP synthase by cryo-EM. Elife.2015;4:e10180.

Appendix

Table S1. Primers used for mitochondrial genome amplification and sequencing. The approximate location of primers within each gene is based on *M. solifugus* sequence.

Primer	Approx. location	Sequence (5' → 3')	Source
Nad1-F1	5' end	GCYATAGCATTTTACACCYTWATAG	1
Nad1-R1	237 bp from 3' end	AATRTTDGCATATTCWGCTAT	1
Nad4-F1	588 bp from 5' end	GGATTTHTAGTAAAACTWCCWATATTYTC	1
Nad4-R1	1110 bp from 5' end	TTAATDGATGGRGGMGCAGCTAT	1
Nad5-F1	443 bp from 5' end	TACTATCARAAYCCWAAATCTYTDGC	1
Nad5-R1	1044 bp fro 5' end	TTKCCCATTCATCGTARRTCTTG	1
Nad5-F2	530 bp from 5' end	GCHTGAACYTTAAAYCARGGKCA	1
Nad5-R2	1072 bp from 5' end	TGWDGCTACTGGYATTTG	1
Cox2-F1	5' end of COII	TGAGGWCAAYTAAYATTYCAAGACG	1
Cox2-R1	Middle of COII	TAYTCRTATCTTCARTATCATTG	1
Cox2-F2	Middle of COII	CAATGATAYTGAAGATAYGARTA	1
16S-F1	Middle of 16S	GTATCCTAACCGTGCAAAGG	1
16S-R1	3' end of 16S	CTTACGCCGGTCTGAACTCAG	1
16S-R2	Middle of 16S	TACCTTTGCACGGTTAGGATAC	1
Cox3-F1	Middle of COIII	AGGNGTDACHGTWACMTGAGC	1
Cox3-R1	3' end of COIII	ACRTCWACAAARTGTCARTATC	1
Cox3-F2	170 bp 3' of Cox3-F1	TTTGTAGCHACWGGWTTCCAYGG	1
Cox3-R2	170 bp 3' of Cox3-F1	CCRTGGAAWCCWGTDGCTACAAA	1
Nad6-F1	178 bp from 5' end	GGYATATTAGTAATATTTKCWTAYTTYGT	1
Nad6-R2	178 bp from 5' end	GCYACRAARTAWGMAAATATTACTAATATRCC	1
Nad2-F1	163 bp from 5' end	GAAGCAKCARTTAARTAYTTMTTA	1
Nad2-R1	Middle of Nad2	CATCCTATRTGRGTAATWGATG	1
Nad2-F2	150 bp 3' of Nad2-F1	GGWATRGCHCCMTGYCAYYTATG	1
Nad2-R2	150 bp 3' of Nad2-F1	CATARRTGRCAKGGDGCYATWCC	1
Nad3-F1	112 bp from 5' end	CCATTTGARTGTGGATTYGAYCC	1
Nad3-R1	3' end	CADTYTAATGANCCYTWATTTCATTC	1
CytB-F1	370 bp from 5' end	ACHATRGCHACAGCATTTATAGG	1
CytB-R1	3' end	TCTTCTACTGGBCGSCCHCCAATTC	1
CytB-F2	120 bp 3' of CytB-F1	AATWTGAGGHGGNTTTGCHGTAG	1
CytB-R2	120 bp 3' of CytB-F1	CTACDGCAAANCCDCCTCASATTC	1
CytB-F3	307 bp from 3' end	GAGTDTATGCWATTYTACGATCWATTCC	1
CytB-R3	307 bp from 3' end	GGAATSGATCGTARAATSGCATAHACTC	1
Atp6-F496	496 bp from 5' end	GCWAAYATAAGAGCNGGBCAYATYGT	1
Atp-R521	496 bp from 5' end	ACRATRTGVCCNGCTCTTATRTTSGC	1
12S-A1	5' end	AAACTAGGATTAGATACCCTATTAT	2
12S-R1	3' end	GAGAGYGACGGGCGATGTGT	1
LCO1490	5' end of COI	GGTCAACAAATCATAAAGATATTGG	3

HCO2198	Middle of COI	TAAACTTCAGGGTGACCAAAAAATCA	3
COI_r	3' end of COI	CCDCTTAGWCCTARRAARTGTTGNGG	1
Cox5.1	Middle of COI	GATTCTTTGGACATCCAGAAG	4
IW-LCO	5' end of COI	ACTCAACTAATCACAAAGACATTG	1
IW-HCO	Middle of COI	ACTTCTGGATGTCCAAAGAATC	1

Primer sources: 1 = this study; 2 = Borda and Siddall, 2004; 3 = Folmer, *et al.*, 1994; 4 = Hartzell, *et al.,* 2005.

Table S2. Pairwise percent identity matrices for combined datasets. Lower diagonal is nucleotide data; upper (red) diagonal is amino acid. The housekeeping gene dataset excludes 28S rRNA.

Dataset		*E. albidus*	*E. crypticus*	*M. pedatus*	*M. solifugus*	*M. antaeus*	*M. gelidus*	*M. hydrius*
5 nuclear house-keeping genes 5304 bp	*E. albidus*		98.64	92.75	92.52	94.67	95.13	94.84
	E. crypticus	93.1		92.8	92.8	94.7	95.0	95.07
	M. pedatus	82.3	82.6		96.3	93.6	93.5	92.81
	M. solifugus	82.1	82.6	86.7		94.7	94.9	93.49
	M. antaeus	84.0	84.7	85.1	86.0		98.1	96.66
	M. gelidus	83.9	84.5	85.3	86.9	90.0		97.17
	M. hydrius	84.5	84.8	85.9	86.4	89.5	90.0	
8 nuclear ATP synthase subunits 6894 bp	*E. albidus*		93.7	78.03	77.2	77.72	77.76	77.58
	E. crypticus	92.7		80.49	79.31	79.78	80	79.68
	M. pedatus	70.6	72.1		91.58	94.31	93.83	92.64
	M. solifugus	69.9	70.9	82.5		92.15	91.49	92.71
	M. antaeus	70.0	71.3	86.9	83.4		95.89	92.99
	M. gelidus	70.7	71.7	86.6	83.4	89.1		92.64
	M. hydrius	70.4	71.6	83.9	84.0	85.8	85.8	
13 mito - chondrial PCGs 11,160 bp	*E. albidus*		96.75	69.22	69.43	69.97	70.15	70.29
	E. crypticus	88.4		69.43	69.61	70.21	70.14	70.28
	M. pedatus	69.4	69.2		84.27	86.54	87.39	85.16
	M. solifugus	68.3	67.9	75.9		85.5	86.37	88.29
	M. antaeus	70.0	69.5	78.4	76.9		89.58	86.95
	M. gelidus	66.8	66.6	75.7	74.1	77.0		87.77
	M. hydrius	69.5	69.2	77.6	78.1	79.2	75.8	
Mito-chondrial COI-CytB 12s-16s rRNAs 4600 bp	*E. albidus*							
	E. crypticus	91.7						
	M. pedatus	75.2	74.9					
	M. solifugus	75.3	75.0	81.9				
	M. antaeus	75.1	74.6	83.1	83.5			
	M. gelidus	74.3	74.5	83.4	83.7	85.7		
	M. hydrius	73.9	73.9	82.9	84.3	84.8	85.0	

Table S3. PartitionFinder recommended partitioning schemes used for concatenated datasets.

Dataset	Genes	No. of partitions	Partition scheme
HK6	Actin, a-tubulin, EF1a, GAPDH, Histone3, 28S rRNA	5	(EF1a_pos1, H3_pos1, aTub_pos1, actin_pos1, GAPDH_pos1) (EF1a_pos2, H3_pos2, aTub_pos2, actin_pos2, GAPDH_pos2) (EF1a_pos3, aTub_pos3, actin_pos3) (H3_pos3, GAPDH_pos3) (28S)
Mt4	COI, Cytb, 12S rRNA, 16S rRNA	5	(Cytb_pos1, COI_pos1) (COI_pos2) (Cytb_pos3, COI_pos3) (Cytb_pos2) (rRNAs)
MtPCGs	Cox1-3, Cytb, Nad1-6 + 4L, Atp6, Atp8	6	(COI_p1, Cox2_p1, Cox3_p1, CytB_p1) (COI_p2) (Atp6_p3, Atp8_p3, COI_p3, Cox2_p3, Cox3_p3, CytB_p3, Nad1_p3, Nad2_p3, Nad3_p3, Nad4L_p3, Nad4_p3, Nad5_p3, Nad6_p3) (Cox2_p2, Cox3_p2, CytB_p2, Nad1_p2, Nad4L_p2) (Atp6_p1, Atp8_p1, Nad1_p1, Nad2_p1, Nad3_p1, Nad4L_p1, Nad4_p1, Nad5_p1, Nad6_p1) (Atp6_p2, Atp8_p2, Nad2_p2, Nad3_p2, Nad4_p2, Nad5_p2, Nad6_p2);
Mt13+RNAs	Cox1-3, Cytb, Nad1-6 + 4L, Atp6, Atp8, 12S, tRNA-V, 16S	7	As above, plus one additional partition for the RNAs
nuASU	Alpha, Beta, Gamma, Delta, b, c, d, OSCP	7	(alpha_pos1, beta_pos1) (alpha_pos2, beta_pos2) (alpha_pos3, delta_pos3, oscp_pos3, syn-b_pos3, syn-c_pos3, syn-d_pos3) (delta_pos1, gamma_pos1, oscp_pos1, syn-b_pos1, syn-c_pos1, syn-d_pos1) (gamma_pos2, oscp_pos2, syn-b_pos2, syn-d_pos2) (beta_pos3, gamma_pos3) (delta_pos2, syn-c_pos2)

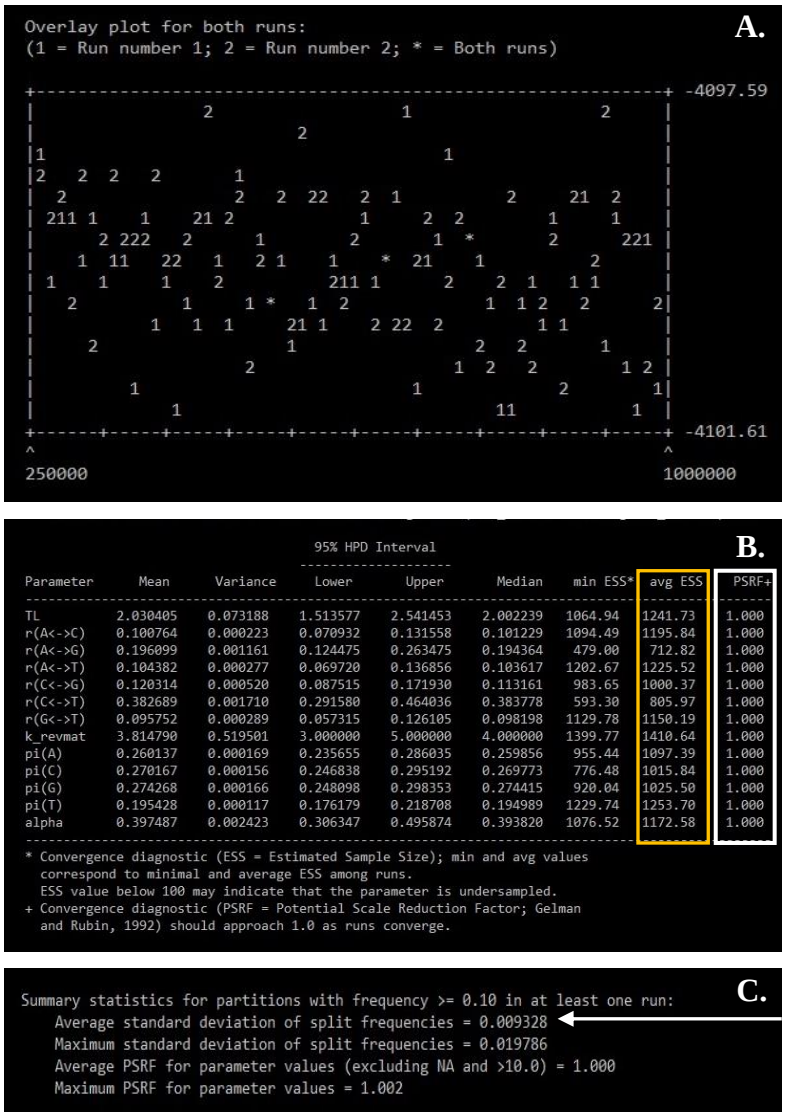
Table S4. GenBank accession numbers for novel transcripts. Epsilon and 28S are pending.

Gene Category	Gene	*E. albidus*	*E. crypticus*	*M. solifugus*	*M. pedatus*	*M. antaeus*	*M. gelidus*	*M. hydrius*
Nuclear-encoded housekeeping genes	*actin*	KU728801	KU728802	KU728803	KU728804	KU728805	KU728806	KU728807
	α-tubulin	KU728808	KU728809	KU728810	KU728811	KU728812	KU728813	KU728814
	EF-1α	KU728815	KU728816	KU728817	KU728818	KU728819	KU728820	KU728821
	histone3	KU728829	KU728830	KU728831	KU728832	KU728833	KU728834	KU728835
	GAPDH	KU728822	KU728823	KU728824	KU728825	KU728826	KU728827	KU728828
	28s rRNA							
Nuclear-encoded ATP synthase subunits	*alpha*	KU728752	KU728753	KU728754	KU728755	KU728756	KU728757	KU728758
	beta	KU728766	KU728767	KU728768	KU728769	KU728770	KU728771	KU728772
	gamma	No Data	KU728788	KU728789	KU728790	KU728791	KU728792	KU728793
	delta	KU728780	KU728781	KU728782	KU728783	KU728784	KU728785	KU728786
	epsilon							
	b	KU728759	KU728760	KU728761	KU728762	KU728763	KU728764	KU728765
	c	KU728745	KU728746	KU728747	KU728748	KU728749	KU728750	KU728751
	d	KU728773	KU728774	KU728775	KU728776	KU728777	KU728778	KU728779
	OSCP	KU728794	KU728795	KU728796	KU728797	KU728798	KU728799	KU728800
Mitochondrial protein-coding genes	*COI*	KU728850	KU728851	KU728852	KU728853	KU728854	KU728855	KU728856
	Cox2	KU728857	KU728858	KU728859	KU728860	KU728861	KU728862	KU728863
	Cox3	KU728864	KU728865	KU728866	KU728867	KU728868	KU728869	KU728870
	Cytb	KU728843	KU728844	KU728845	KU728846	KU728847	KU728848	KU728849
	Nad1	KU728871	KU728872	KU728873	KU728874	KU728875	KU728876	KU728877
	Nad2	KU728878	KU728879	KU728880	KU728881	KU728882	KU728883	KU728884
	Nad3	KU728885	KU728886	KU728887	KU728888	KU728889	KU728890	KU728891
	Nad4	KU728892	KU728893	KU728894	KU728895	KU728896	KU728897	KU728898
	Nad4L	KU728899	KU728900	KU728901	KU728902	KU728903	KU728904	KU728905
	Nad5	KU728906	KU728907	KU728908	KU728909	KU728910	KU728911	KU728912
	Nad6	KU728913	KU728914	KU728915	KU728916	KU728917	KU728918	KU728919
	Atp6	KU728836	KU728837	KU728838	KU728839	KU728840	KU728841	KU728842
	Atp8	KU639961	KU639962	KU639963	KU639964	KU639965	KU639966	KU639967
Mitochondrial rRNA genes	*12s*	KU728920	KU728921	KU728922	KU728923	KU728924	KU728925	KU728926
	16s	KU728927	KU728928	KU728929	KU728930	KU728931	KU728932	KU728933

Table S5. Bayesian inference Newick trees for all genes and datasets.

Loci	Bayesian Inference Newick Trees
12s	(Ealb:0.006770898,Ecry:0.03538963,(Mped:0.1356714,(Msol:0.09907861,Mhyd:0.08418686):0.03167925,(Mant:0.07932975,Mgel:0.09419123):0.02030384):0.761206);
16s	(Ealb:0.006770898,Ecry:0.03538963,(Mped:0.1356714,(Msol:0.09907861,Mhyd:0.08418686):0.03167925,(Mant:0.07932975,Mgel:0.09419123):0.02030384):0.761206);
28s	(Ealb:0.009272668,Ecry:0.003186291,((Mped:0.01092156,((Msol:0.02990588,Mhyd:0.03602543):0.00784675 6,Mant:0.01396765):0.008108208):0.007772201,Mgel:0.01116978):0.3247203);
Actin	(Ealb:0.006634816,Ecry:0.01077796,(Mped:0.02552142,(((Msol:0.06971496,Mgel:0.04039868):0.03221018,Mant:0.1219178):0.05155338,Mhyd:0.02191136):0.01205492):0.08816325);
Alpha	(Ealb:0.008025507,Ecry:0.03388649,(((Mped:0.08920178,Mant:0.04654193):0.007337236,Mgel:0.05208632):0.04008744,Msol:0.08265423):0.02041718,Mhyd:0.07423455):0.4128043);
Atp6	(Ealb:0.1006109,Ecry:0.06697589,((Mped:0.53591 91,(Msol:0.2669599,Mhyd:0.1365852):0.07045715,Mgel:0.13741161):0.1030408,Mant:0.03734736):1.667381);
Atp8	(Ealb:0.04558744,Ecry:0.05204976,(Mped:0.08935008,(((Msol:0.2350216,Mhyd:0.06556376):0.2291701,Mant:0.04424696):0.05427059,Mgel:0.1921267):0.1337614):0.7579237);
aTub	(Ealb:0.0280425,Ecry:0.05940868,((Mped:0.1844469,Mant:0.047322):0.0229864,((Msol:0.05237005,Mgel:0.06000996):0.01533611,Mhyd:0.05423477):0.02439432):0.1095459);
Beta	(Ealb:0.009431517,Ecry:0.01898405,(((Mped:0.09013813,Msol:0.1703738):0.04445966,Mgel:0.04070362):0.0203751,(Mant:0.06328075,Mhyd:0.08784072):0.0186773):0.2052972);
COI	(Ealb:0.04967886,Ecry:0.08279515,(Mped:0.1265964,(Msol:0.1823573,(Mant:0.1022228,(Mgel:0.1406486,Mhyd:0.1248605):0.02142548):0.0308885):0.01767774):0.166163);
Cox2	(Ealb:0.05148498,Ecry:0.07268306,(Mped:0.2101649,((Msol:0.2368364,Mhyd:0.07374255):0.07473834,Mant:0.1347546):0.04838382,Mgel:0.1468044):0.4432648);
Cox3	(Ealb:0.1100225,Ecry:0.08157745,(Mped:0.1309824,(Msol:0.3234226,(Mant:0.2194696,Mhyd:0.2078351):0.04899946,Mgel:0.1683284):0.06326157):0.3822504);
CytB	(Ealb:0.1899507,Ecry:0.1136645,(Mped:0.2216164,(((Msol:0.2761503,Mhyd:0.2795215):0.08509904,Mant:0.2359159):0.0655646 2,Mgel:0.2144372):0.1312525):0.7417922);
Delta	(Ealb:0.02311254,Ecry:0.007912509,((((Mped:0.1132083,Mant:0.1007896):0.0255964,Mgel:0.08999871):0.05545259,Msol:0.07418853):0.04720779,Mhyd:0.08393262):0.5105228);
EF1a	(Ealb:0.06155822,Ecry:0.05248541,(Mped:0.05515498,(((Msol:0.05667427,Mhyd:0.06419409):0.02041683,Mant:0.03704231):0.01661503,Mgel:0.03527733):0.03079627):0.1483257);
GAPDH	(Ealb:0.02965163,Ecry:0.01018119,(Mped:0.1023636,(Msol:0.1110167,Mhyd:0.07963731):0.02086502,Mant:0.08317348,Mgel:0.08051307):0.3720227);
His3	(Ealb:0.4618253,Ecry:0.1657169,((Mped:0.09145983,Msol:0.2043889):0.06660217,Mant:0.05458691,Mgel:0.1081008,Mhyd:0.08565743):0.07861968);
Nad1	(Ealb:0.06557506,Ecry:0.1439481,(Mped:0.3475321,((Msol:0.2614244,Mhyd:0.2768651):0.06552739,Mgel:0.2453313):0.05995 69,Mant:0.14646591):1.106616);
Nad2	(Ealb:0.1166942,Ecry:0.09196712,(Mped:0.3152666,(Msol:0.4380866,Mhyd:0.269151):0.1388592,(Mant:0.2681369,Mgel:0.2627457):0.07513771):2.38637);
Nad3	(Ealb:0.08377358,Ecry:0.09204989,(Mped:0.2560777,Msol:0.2309416,Mant:0.1874927,Mgel:0.1822523,Mhyd:0.2224673):0.8099003);
Nad4	(Ealb:0.1254377,Ecry:0.06112254,(Mped:0.186308,(Msol:0.2832701,Mhyd:0.2412424):0.1010536,Mant:0.2694843,Mgel:0.2732755):1.464979);
Nad4L	(Ealb:0.09111356,Ecry:0.06359623,(Mped:0.2287165,Msol:0.3784204,(Mant:0.169601,Mhyd:0.2387357):0.1138872,Mgel:0.09100801):1.544231);
Nad5	(Ealb:0.06632153,Ecry:0.1155907,(((Mped:0.2878917,Mgel:0.1724932):0.07616429,Mant:0.2707214):0.06106031,(Msol:0.3768889,Mhyd:0.1754368):0.08591947):1.147521);
Nad6	(Ealb:0.09655636,Ecry:0.1306967,(Mped:0.5228473,Msol:0.3751322,Mant:0.2812365,Mgel:0.3089747,Mhyd:0.3991824):3.104096);
OSCP	(Ealb:0.007914523,Ecry:0.01891683,((((Mped:0.11301,(Mant:0.07559698,Mgel:0.06912302):0.02215433):0.05354757,Mhyd:0.06046808):0.05404024,Msol:0.1235045):0.4891871);
b	(Ealb:0.03214107,Ecry:0.0252717,((Mped:0.1431415,(Mant:0.05179623,Mgel:0.09555153):0.01966341):0.03779533,(Msol:0.1404522,Mhyd:0.09338837):0.05312848):0.9416359);
c	(Ealb:0.01380941,Ecry:0.02547177,(Mped:0.03104475,((Msol:0.07769612,Mhyd:0.08155793):0.04698265,Mant:0.0831281,Mgel:0.09443937):0.06045314):0.6834484);
d	(Ealb:0.01097478,Ecry:0.027533,((((Mped:0.09748302,Mgel:0.0635017):0.02492598,Mant:0.07978162):0.05921604,Mhyd:0.08366598):0.07140724,Msol:0.05010475):0.751968);
Gamma	(Ecry:0.5470455,Mped:0.06583683,(((Msol:0.1688175,Mhyd:0.1053316):0.02367685,Mgel:0.06574272):0.02567122,Mant:0.07674278):0.03466705);
Epsilon	(Ecry:0.4470465,(Mped:0.05858022,(Msol:0.1491025,Mant:0.1020348):0.052111):0.04423441,(Mgel:0.04290376,Myhd:0.09981654):0.04159724);
HK6part	(Ealb:0.04803389,Ecry:0.03066387,(((Mped:0.1238827,Msol:0.09472495):0.07403191,(Mant:0.08173273,Mgel:0.06410131):0.01586885):0.01923323,Mhyd:0.06359047):0.2257389);
ASUpart	(Ealb:0.06869,Ecry:0.022726,((Mped:0.118345,(Mant:0.072395,Mgel:0.076172):0.012759):0.034114,(Msol:0.13808,Mhyd:0.095012):0.017211):0.580004);
Mt13_RNA	(Ealb:0.07511753,Ecry:0.09892344,(Mped:0.23282,((Msol:0.2945894,Mhyd:0.1970226):0.07184251,(Mant:0.2071693,Mgel:0.2009015):0.03017737):0.05486687):1.096201);
Mt4part	(Ealb:0.4925207,Ecry:0.5486095,(Mped:1.016442,((Msol:1.561981,Mhyd:1.264297):0.3147588,(Mant:1.219553,Mgel:1.193196):0.245599):0.7346435):6.150887);

Figure S1. Screen shots of MrBayes output for assessment of a successful analysis. The output resulting from addition of the 'sump' command to a phylogenetic analysis in MrBayes (Ronquist et al., 2012) includes several parameters by which the success of the run can be assessed. (A) The plot of generation (x-axis) vs. log probability (y-axis) indicates convergence if there are no obvious trends (e.g. rising or falling). (B) In the 95% highest probability density (HPD) table, the potential scale reduction factor (PSRF) should be ~1 when you have a good sample from the posterior probability (white box), and the effective population size should be >200 (yellow box). (C) A final parameter for diagnosis of success is the standard deviation of split frequencies (arrow), which should approach 0 as the 2 runs converge on the same stationary distribution of probabilities.



Abbreviations

+/-	charge	kD	kilo Daltons
aa	amino acid	Lter	*Lumbricus terrestris*
AIC	Akaike information criterion	Mant	*Mesenchytraeus antaeus*
Arg	arginine	Mgel	*Mesenchytraeus gelidus*
ASU	ATP synthase subunit	Mhyd	*Mesenchytraeus hydrius*
Asp	aspartic acid	ML	*maximum likelihood*
ATP	adenosine triphosphate	Mped	*Mesenchytraeus pedatus*
BEB	Bayes Empirical Bayes	mRNA	*messenger RNA*
BI	Bayesian inference	MSC	*multi-species coalescent*
bp	base pairs	Msol	*Mesenchytraeus solifugus*
cDNA	complementary DNA	Mt	mitochondrial
Cryo-EM	cryogenic electron microscopy	MW	molecular weight
C-terminal	carboxy-terminal	mya	million years ago
e-	electron	NADH	nicotinamide adenine dinucleotide
Ealb	*Enchytraeus albidus*	N-terminal	amino-terminal
Ecry	*Enchytraeus crypticus*	OSCP	oligomycin sensitivity-conferring protein
ELA	equilibrium line altitude	OXPHOS	oxidative phosphorylation
ESS	effective sample size	PCG	protein coding gene
ETC	electron transport chain	PCR	polymerase chain reaction
F_1	factor 1 domain of ATP synthase	PI	phylogenetically informative
F_O	factor O domain of ATP synthase	pI	iso-electric point
Glu	glutamic acid	pmf	proton motive force
GRAVY	grand average of hydropathy	PSRF	potential scale reduction factor
GTR	general time reversible	rRNA	ribosomal RNA
HΦ	hydrophobicity	sd	standard deviation
HGT	horizontal gene transfer	SEM	standard error of the mean
His	histidine	SH-aLRT	Shimodaira-Hasagawa approximate likelihood ratio test
HK	house keeping genes	SPR	subtree pruning and regrafting
Hsap	*Homo sapians*	SynB	ATP synthase subunit b
IGR	independent gamma rate	SynC	ATP synthase subunit c
ILS	incomplete lineage sorting	SynD	ATP synthase subunit d
IMS	inter-membrane space	tmrca	time to most recent common ancestor
Kb	kilo base pairs	tRNA	transfer RNA

Attributes

Figure 1 – Arrhenius equation box created by Shirley A. Lang

Figure 2 – ATP Synthase cartoon created by Shirley A. Lang (Inkscape 0.91)

Figure 3 – ATP6 ribbon-model created by Shirley A. Lang from *M. solifugus* aa sequence and PDB model 5lqx.1.2 in Swiss-Model.

Figure 4 – Structure of ATP synthase dimers created by Shirley A. Lang

Figure 5 – Mitochondrial cristae with dimers created by Shirley A. Lang

Figure 6 – Specimen collection sites and habitats by Shirley A. Lang

Figure 7 – Mitochondrial genome PCR scheme and gel images by Shirley A. Lang

Figure 8 – ML and BI dataset trees by Shirley A. Lang

Figure 9 – MSC tree and DensiTree by Shirley A. Lang

Figure 10 – ATP assay graphs by Shirley A. Lang

Figure 11 – WebLogo image of transit sequences by Shirley A. Lang

Figure 12 – Subunit c/9 multi-species alignment by Shirley A. Lang

Figure 13 – Molecular models of Gamma and SynB by Shirley A. Lang

Figure 14 – *Epsilon* isoform alignment by Shirley A. Lang

Figure 15 – Raw read maps for ATP6 by Joseph Kawash at request of S. Lang

Figure 16 – Alignment from PCR amplification of ATP6 by Shirley A. Lang

Figure 17 – Result of ATP6 extension protein BLAST by Shirley A. Lang

Figure 18 – ATP6 hydrophobicity plot and molecular model by Shirley A. Lang

Figure 19 – Proposed radiation of examined species drawn by Shirley A. Lang, as conceived by Daniel Shain.

Figure 20 – ATP synthase dimer showing ATP6 extension and proposed role in proton shuttling by Shirley A. Lang

Table 1 – Specimen descriptions and collection sites by Shirley A. Lang

Table 2 – Concatenated dataset descriptions by Shirley A. Lang

Table 3 – Descriptive statistics of genetic loci by Shirley A. Lang

Table 4 – Target sequence prediction by Shirley A. Lang

Table 5 – Physico-chemical properties by Shirley A. Lang

Table 6 – Unique amino acid changes by Shirley A. Lang

Table S1 – Mitochondrial primers by Shirley A. Lang

Table S2 – Pairwise identity of datasets by Shirley A. Lang

Table S3 – Partitioning schemes by Shirley A. Lang

Table S4 – GenBank accession numbers by Shirley A. Lang

Table S5 – Newick trees of all genes and datasets by Shirley A. Lang

Figure S1 – Screen-shots of MrBayes diagnostics for assessment of the success of a phylogenetic analysis created by Shirley A. Lang