Mito-nuclear compatibility boundaries across the Danionin lineage

By

Trevor J. Chamberlain

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Genetics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2025

Date of final oral examination: 07/29/2025

The dissertation is approved by the following members of the Final Oral Committee:
Francisco J. Pelegri, Professor, Laboratory of Genetics
Nicole T. Perna, Professor, Laboratory of Genetics
Christopher Hittinger, Professor, Laboratory of Genetics
Melissa Harrison, Professor, Biochemistry
Akihiro Ikeda, Professor, Laboratory of Genetics

Dissertation Abstract

Mito-nuclear compatibility boundaries across the Danionin lineage

By

Trevor J. Chamberlain

Under the supervision of Professor Francisco Pelegri

at the University of Wisconsin – Madison

Mitochondrial activity is the product of two genomes, the nuclear genome (nDNA) that contains the large majority of a cell's genetic instructions, and a small genome (mtDNA) with copies that reside within each mitochondrial matrix and code for genes involved in oxidative phosphorylation (OXPHOS) to produce ATP and translation of mitochondrial proteins. Interactions between the nDNA and mtDNA are crucial to mitochondrial function, and incompatibilities that may arise in hybrids that have a mtDNA from one parental species and nDNA from another parent could lead to decreased mitochondrial fitness or even hybrid lethality. In this dissertation I systematically study the variations that arise in mitochondrial protein sequences and structural predictions between species within the Danionin lineage which includes the model organism *Danio rerio* (zebrafish) and its relatives. Chapter 1 contains a review characterizing features and methods involved in mito-nuclear research and then applies the concepts to interspecies somatic cell nuclear transfer efforts currently used in conservation breeding efforts. Chapter 2 examines the sequence variation, phylogenetic relationships, and evolutionary histories of Danionin species and characterizes proteins including the Cox2 protein in *Devario* species and Nd5 in *Danio albolineatus* that underwent rapid evolution relative to

other Danionin species and would make suitable candidates for future functional studies characterizing activity in these subunits. In Chapter 3, I describe respiratory complex IV protein structural predictions and docking simulations between the mitochondrial-encoded Cox2 protein and the nuclear-encoded Cycsb protein. I then identify probable docking incompatibilities between *Devario aequipinnatus* Cox2 and *Danio rerio* Cycsb and the unique docking orientation that occurs when Cycsb binds to the surface of Cox2 in *Devario aequipinnatus*. Together, these works present the most likely sites of mito-nuclear incompatibility, work to establish the Danionin lineage as a convenient vertebrate model for mito-nuclear research, and provide a better understanding of the appearance of mitochondrial variants within a phylogenetic context.

Acknowledgments

First, I want to take a moment in memoriam for loved ones gone but forever cherished. To my grandmothers, Renee Austad and Karen Chamberlain, I am grateful for the moments in childhood where you inspired me and loved me. I thank Renee for her example as she sought her own path and chose to pursue an education later in life, and for the encouragement she provided me as I braved the turbulent waters of non-traditional academic life. To my in-laws, Debbie Boudreau, Marilyn Boudreau, and Leonard Boudreau, I express deep gratitude for your faith in me and your endless support as I set out to pursue a PhD. In times of trial, that lingering support has been the lasting optimism for me that has kept me pushing forward. I am sorry it took me too long to reach this point because you were always supposed to be here at the end of this chapter, especially after asking so many times when you could come to my graduation.

I thank my doctoral advisor Dr. Francisco Pelegri for taking me into his lab, providing me with an exciting and new project, and for inspiring me with his enthusiasm for the work our lab does. I was asked by a journalist for Grow magazine in a piece she wrote on our work if I could describe Francisco in one word and the answer was easy: “Passionate.” I have been grateful to share in the passion Francisco has for conservation and genetics during my graduate work. In addition, I would like to thank the colleagues I worked with in the Pelegri Lab: Dr. Christina Hansen, Dr. Ryan Trevena, Caroline Barry, Gabby Voit, Erin Miners, and Sara Fateriоun. You’ve all inspired me and assisted me in countless ways as we collectively braved our research. I’m especially thankful to Christina, who often commiserated with me, uplifted me, and provided an amazing example of incredible scientific work. And to Ryan who joined me on a new research

path for the lab and worked with me to establish care for new species and to design new protocols.

I would like to thank Dr. Nicole Perna, Dr. Melissa Harrison, Dr. Chris Hittinger, Dr. Akihiro Ikeda, and Dr. Dave Pagliarini for generously sharing their time, knowledge, and support throughout my graduate training. I also thank the countless other individuals at the University of Wisconsin – Madison who helped me get to this point, including Martha Reck, Annie Anderson, and Errol Wizda. My doctoral cohort in the Laboratory of Genetics is full of amazing individuals, and I am grateful to also have been influenced by their presence.

I am grateful for the individuals in the UW – Madison Biocore program where I was graciously welcomed as a teacher's assistant for three semesters. Among others I thank Dr. Heidi Horn, Dr. Anna Kowalkowski, Seth McGee, Dr. Shelby O'Connor, Dr. Bill Bement, Dr. Trina McMahon, Dr. Anne Griep, and Carol Borcherding. For offering a strong environment for growth as an instructor, I am truly grateful. Being part of the Biocore was the highlight of the final stages of my PhD experience and I thank you for the opportunity I had to belong with all of you.

Thank you to my mentors and professors at Utah Valley University, including Dr. Olga Kopp, Dr. Eric Domyan, Dr. Eddy Cadet, Dr. Devin Taylor, and Dr. Dani Taylor who taught me the essentials to succeed as a scientist and supported my progress at the onset of my career. I also thank Mark Cooper at Mt. San Antonio college for instructing the biology course that sparked my desire to pursue a career studying biology and evolution and for the engagement you offered me as your student. Thank you to all other individuals who imbued in me a love for discovery,

especially those I had in my formative years in grade school and high school. There are too many of you to mention, but I owe so much of who I am to the experiences you provided.

To Bishop Mark Nolte, Bro. James McKay, Bro. Chad Shorter, Bro. Bruce Matthews, and Bro. Karl Miller, it was a privilege serving alongside you in the Madison 1st Ward Bishopric of The Church of Jesus Christ of Latter-day Saints. Your camaraderie was and continues to be an enormous support to me and my family and I cherish your examples and advice I've received while serving with you. In addition, I recognize the friendship and support I have received from AJ Sorensen and Dave Armstrong who I have looked forward to sharing time with each week for more than two years as we served the financial needs of our congregants. I have also felt the love and support from the rest of my ward family over the past eight years, and I am grateful to each one of you for your encouragement and fellowship.

Thank you to my friends, neighbors, and family for being there during both joyful and challenging times. I have been influenced and supported by innumerable people – thank you especially to my parents, grandparents, aunts, uncles, cousins, in-laws, extended family, and my beloved friends Keith Jackson and Dakota Nuttal who have been as close to family as anyone could be.

To Chris and Kent Chamberlain, thank you so much for everything. You've done well with me and in instilling the traits that matter in my character. Mom, I love you and I love your belief in me. Dad, thank you for teaching me an incredible work ethic and for sharing a curiosity in the natural world and the mysteries that we collectively marvel at just waiting to be explored.

Candice, you are everything to me. I love you beyond my ability to express in words. No one could believe in anyone more than the way you have believed in me. I am the luckiest person to have a companion who wholly supports me and works with me to pursue a future together. Thank you for being my best friend, for accepting all of me, and for taking this journey with me.

To El and Iris, I could not be prouder to be your dad. You are both wonderful individuals and I love you both so much. I am thankful for your love and support and for the daily reminders of how I can do better because both of you impress me at every turn.

Table of Contents

CHAPTER 1

Concerning mito-nuclear compatibilities during cloning with interspecies nuclear transfer for conservation-focused species remediation

Chamberlain, T. J.

Laboratory of Genetics

University of Wisconsin – Madison

425-G Henry Mall

Madison, WI 53706

USA

1.1 Abstract

Mitochondrial-encoded proteins do not function without interacting with nuclear-encoded subunits. These interactions are crucial for the functions that mitochondria perform but are especially necessary for oxidative phosphorylation (OXPHOS) for the generation of cellular energy as ATP. However, mito-nuclear compatibility boundaries sometimes arise in diverged populations in which the mitochondrial-derived subunits of one species can no longer interact with nuclear-derived subunits of another species which can lead to a loss of mitochondrial fitness or even embryonic lethality in hybrids between the two species. In this sense, mito-nuclear incompatibilities can be a source of postzygotic isolation, a mechanism described in the Bateson-Dobzhansky-Muller hybrid incompatibility (BDMI) with the potential to affect speciation between populations. In this review, I cover how BDMIs arise in more detail, how mito-nuclear incompatibilities are studied, and how they relate to conservation research and cloning efforts involving interspecies somatic cell nuclear transfer (isSCNT). Finally, I describe some current needs for mito-nuclear research and what the future may hold with the promise that present endeavors hint at.

1.2 The basics of mito-nuclear interactions

While mitochondria are responsible for a variety of critical cellular functions, including but not limited to regulation of apoptosis (Green & Reed, 1998; Tait & Green, 2010), metabolic processes including TCA cycle and fatty acid β-oxidation (McBride et al., 2006), cell cycle regulation (Owusu-Ansah & Banerjee, 2009), innate immunity (West et al., 2011) and thermogenesis (Cannon & Nedergaard 2004), most notably mitochondria carry out aerobic

respiration to generate ATP, the energy currency of cells (Zhao et al., 2019; Nolfi-Donegan et al., 2020).

Respiration in mitochondria is conducted through oxidative phosphorylation (OXPHOS) in which respiratory complexes I–IV, which are involved in the electron transport chain (ETC), perform redox reactions driven by electron transfer (ET) from complex to complex to provide the energy for pumping protons into the intermembrane space (Mitchell, 1961; Saraste, 1999; Zhao et al., 2019; Nolfi-Donegan et al., 2020). Then, in complex V, the ATP synthase, the proton motive force generated by the protons crossing the gradient between the intermembrane space and the mitochondrial matrix, drives the phosphorylation of ADP to ATP.

Interestingly, mitochondria have their own genome that encodes subunits that are exclusively involved in the ETC and ATP synthase processes that drive OXPHOS and some of the genes needed for mitochondrial gene translation. In vertebrates, this circular genome is usually 16-17 kilobases and normally contains the same 13 protein-coding genes, 22 tRNAs, and 2 rRNAs (Figure 1A) (Attardi & Schatz, 1988; van Gisbergen et al., 2015). Broken down by respiratory complex, there are seven complex I genes (*ND1*, *ND2*, *ND3*, *ND4L*, *ND4*, *ND5*, *ND6*), one complex III gene (*CYTB*), three complex IV genes (*COX1*, *COX2*, *COX3*), and two complex V genes (*ATP6*, *ATP8*). Among these, nearly every gene encodes a core structural subunit in their respective complexes upon which other proteins rely upon foundationally (Signes & Fernandez-Vizarra, 2018). In addition to structural purposes, several proteins also serve as functional components of each complex, performing essential tasks involved in the ETC or in ATP synthesis. In complex I, *ND2*, *ND4*, and *ND5* each encode the active portions of proton pumps

involved in shuttling protons from NADH to the mitochondrial intermembrane space which helps create the necessary OXPHOS proton gradient and maintains mitochondrial membrane potential (Hunte et al., 2010; Efremov and Sazanov, 2011). Additionally, both *COX1* and *COX2* mitochondrial-encoded proteins make up the binding domain in complex IV where cytochrome c docks and transfers electrons during the final transfer of the ETC (Roberts & Pique, 1999; Sato et al., 2016).

Furthermore, all mitochondrial roles rely on the structural and functional tasks that mitochondrial-encoded proteins perform. One scenario occurs because of increased proton leakage in which protons are prematurely siphoned from the intermembrane space into the mitochondrial matrix leading to membrane depolarization and loss of membrane potential (Rego et al., 2001; Divakaruni & Brand, 2011). Also, in experiments examining OXPHOS complex formation defects, transmission electron microscopy (TEM) has been used to observe disruption to the integrity of mitochondrial cristae which shape the organelle and facilitate all functions performed by mitochondria in addition to OXPHOS-related activities (Vincent et al., 2016). These are just a couple of examples demonstrating that ultimately, the precise mechanisms that involve mitochondrial-encoded genes have implications on the function and survival of each mitochondrion in a cell.

While the thirteen mitochondrial-encoded subunits are each necessary to mitochondrial function, they are not sufficient to perform OXPHOS or to regulate mtDNA replication, transcription or translation on their own. Most of the components of mitochondrial structure and function are carried out by nuclear-encoded genes that are transported into the mitochondria. For example, the

majority of proteins in OXPHOS complexes are from nuclear-encoded protein-coding genes. In Figure 1B I have depicted the five mitochondrial complexes involved in ETC and ATP synthesis with mitochondrial-encoded subunits depicted in colors corresponding to the mtDNA genes found in Figure 1A while all nuclear-encoded components are shown in gray. It is the interactions between mitochondrial-encoded components, including mtDNA itself, and nuclear-encoded genes and subunits that constitute mito-nuclear interactions. These interactions perform functions related to OXPHOS activity, mtDNA replication, mtDNA transcription, and mtDNA translation. Among a few key noteworthy characters to mention here are every gene of respiratory complex II, which have all completely migrated to nDNA (Ma et al., 2011). Additionally, cytochrome c is essential to electron transfer between complex III and complex IV among other activities, and its protein subunit, *CYCS*, is encoded by the nucleus (Scharlau et al., 2019). To replicate mtDNA, mitochondria have no replication machinery within mtDNA and rely on nuclear-encoded players including *POLG* and *TFAM* to make mtDNA copies (Spikings et al., 2007; Rahn et al., 2015). Similarly, transcription can only occur within mitochondria through uses of *mtRPOL* and *TFAM* as well (Burton & Barreto, 2012). Regarding translation, 12s and 16s rRNAs are found within mtDNA along with all 20 amino acid tRNAs (both tRNA-Leu and tRNA-Ser have two copies in mtDNA) required for mitochondrial gene translation, but the rest of the protein and RNA factors making up mitochondrial ribosomes are derived from nDNA (Burton & Barreto, 2012). While I have not included an exhaustive list, these examples punctuate the importance of mito-nuclear interactions and provide context for the details contained throughout this review.

1.3 Where mito-nuclear incompatibilities come from

The simple answer to how mito-nuclear incompatibilities arise is mutation. However, there are several aspects of mutation which I will discuss here. The first type of mutations in mito-nuclear interactions I will consider are genetic polymorphisms leading to heritable traits. This form of mito-nuclear incompatibility is often the cause of disease phenotypes and has been the subject of a justifiably large number of studies for its applications in human health (Sharma et al., 2009; Calvo & Mootha, 2010; Johnson et al., 2017).

The other origins of mito-nuclear incompatibility are of intrigue for studies considering species relationships which will be the focus for the remainder of this review. The mechanism from which mito-nuclear incompatibilities arise is described in the Bateson-Dhobzhansky-Muller hybrid incompatibility (BDMI) model (Welch, 2004; Burton & Barreto, 2012; Sunnucks et al., 2017). This model posits that a population with co-evolved mtDNA and nDNA that becomes geographically split will begin to have mitochondrial mutations accumulate first in mtDNA with compensatory mutations in the nDNA arising to match. Over time, the number of these mutations grows and the mito-nuclear interactions in each population grow distinct to each population. Finally, if the two populations are re-introduced to each other, the co-evolved mtDNA and nDNA of each species will no longer be adequate for correct mito-nuclear pairing and either mitochondrial dysfunction leading to fitness loss or even embryonic lethality may occur in inter-population hybrids, which indicates a mechanism of postzygotic speciation.

Of course, further evolutionary considerations also play a role in developing mito-nuclear interactions. mtDNA is typically under tight negative selection constraints (Aledo et al., 2014; Shtolz & Mishmar, 2019), but sites still undergo positive selection when it provides a functional

advantage, usually related to environmental adaptation. Matching signatures of positive selection are then adopted by nuclear genes at compensatory sites to further increase fitness (Osada & Akashi, 2012). Genetic drift also plays a role here when relaxation of purifying selection occurs, either in response to abundance in resources for a population (Zheng et al., 2024), or sometimes due to low effective population size (Paplova et al., 2017). Once selection has been relaxed, genetic drift is then able to increase rates of variant accumulation and evolution in mtDNA. This relaxation may even temporarily increase variation followed by later rounds of positive selection acting on those variants, especially where mito-nuclear interactions are more favorable (Zhao et al., 2024; Liu et al., 2025).

Going forward, I will be focused on details relevant to mitochondrial research and apply it to conservation studies. I will first describe the focus and goals of mitochondrial research in conservation, then experimental and modeling techniques used for studies, relevant experimental systems or animal models used to study mitochondrial applications in conservation, and the current courses of direction involving mitochondrial studies in this field.

1.4 The value of mito-nuclear studies insc conservation

The most immediate need to study mito-nuclear interactions in the field of conservation comes from recent improvements to cloning techniques for use in breeding programs for endangered species. In this rapidly advancing technology, it is increasingly more possible to perform interspecies somatic cell nuclear transfer (isSCNT) by taking somatic cells from an endangered species, preferably non-invasively, then transfering the nucleus of somatic cells to the egg of a domestic species that can be used as a low-risk surrogate to rear the endangered offspring

(Lagutina et al, 2013; Moulavi et al., 2017; Wani et al., 2017). Recent isSCNT successes have included cloning three black-footed ferrets that offer hope to introduce genetic diversity if they can successfully breed since the entire population of black-footed ferrets dwindled to only seven individuals in the 1980s (Novak et al., 2024a). One can imagine that as more isSCNT successes occur, possibilities exist to breed numerous other animals that are currently backsliding in increasing situations of habitat loss such as orangutans in rain forest ecosystems undergoing agricultural-derived deforestation (Ancrenaz et al., 2006), and spreading diseases including chytrid fungi that is harming frog populations worldwide (Turner et al., 2021).

However, while isSCNT seems beneficial in conservation approaches, this techniqe depends on the “stuff” in the egg, such as RNA, proteins, and maternally inherited mitochondria, having compatibility with the foreign DNA that is introduced to those maternal factors. While technology is rapidly improving and successful isSCNT procedures become more common, more basic research should be performed to understand the underlying risks of using isSCNT to boost populations of already threatened creatures. For example, all three of the cloned black-footed ferrets are females which means their mitochondria will be passed on to their offspring. This could be problematic if the domestic ferret harbors a currently unseen mitochondrial substitution that leads to a mito-nuclear incompatibility and introduces a mitochondrial disorder to the critically threatened ferret population. Fortunately, this scenario is fairly easy to overcome if breeders are cognizant that male ferrets will not pass on their mitochondria and by selecting male progeny from the cloned ferrets, there is no chance to introduce a mitochondrial defect to future ferret generations. Nevertheless, this example highlights the importance of research into mito-nuclear interactions that are presented during isSCNT.

As it stands now, work on understanding interspecies mito-nuclear incompatibilities is at minimum. One thing we currently know is that species must be closely related for isSCNT to work. For example, experiments testing nuclear transfer in cat species have been successful (Gómez et al., 2003; Gómez et al., 2004). Other compatible crosses have included wolf pups reared in dog surrogates (Oh et al., 2008), mouflon into sheep (Loi et al., 2001), and Przewalski's horses into domestic horses (Novak et al., 2024b). The cases that tend to fail seem like they would be obviously incompatible, but experiments have attempted isSCNT between rabbits and mice (Huang et al., 2017), giant pandas and rabbits (Jiang et al., 2004), and even whales and pigs (Lee et al., 2009), primarily to study epigenetic nuclear reprogramming, but these types of interspecies experiments have demonstrated that great distances between species inhibit successful nuclear transfer attempts.

It stands to reason that questions can be addressed through mito-nuclear compatibility research that could be done in a systematic fashion with attention to finer details all while practicing responsible research stewardship. Specifically, topics to address include the following: Are there rules to guide relatedness assessments between species considered for isSCNT? What techniques would be most appropriate for reversing a mito-nuclear incompatibility if a domestic species is not available with close enough relatedness? How do we address issues specific to mitochondria such as heteroplasmy which may disrupt mitochondrial fitness in currently unforeseen fashion? What kinds of incompatibilities are most common in mito-nuclear incompatibilities? What are the most efficient ways to characterize whether isSCNT could lead to unfit mito-nuclear interactions in species of interest for the procedure? Ultimately, intraspecies incompatibilities

that lead to disease phenotypes have been studied more intensely and for a considerably longer time than dysfunctions derived from interspecies mito-nuclear pairings. It is with these questions in mind that I continue my dive into methodology for studying mito-nuclear relationships.

1.5 Methods used in mito-nuclear conservation studies

Early studies that addressed relatedness for mito-nuclear pairings did so in a fascinating manner. Researchers performed experiments in transmitochondrial cybrid cells which are a fusion between cells from one species that have their mitochondria depleted, called rho-zero cells, and cells from another species that are enucleated (King & Attardi, 1996). Through these experiments, researchers found that a complete loss of competent function occurred in crosses with orangutan mitochondria and a human nucleus in primate assays (Kenyon & Moraes, 1997) and in scenarios with Norwegian rat mitochondria and a mouse nucleus in rodents (McKenzie & Trounce, 2000). In both experiments cells with mitochondrial donors that were closer related to either human or mouse, respectively, were able to survive with varying degrees of mitochondrial function. Taken together, these results suggest that a divergence time of about 10–12 million years is sufficient to lose all degrees of mito-nuclear compatibility. Up to the present, these classic experiments continue to be one of the only sources of information with systematically tested lineages that experimentally defined compatibility boundaries between species and there is little to go on for evidence involving mito-nuclear compatibility boundaries *in vivo*.

More presently, mitochondrial functional assays have benefitted from advances developed in the biotechnology field. First off, the common biotechnological technique qPCR can be used to perform mitochondrial copy number counts by comparing the threshold cycle (Ct) of a sample to

a known standard or a nuclear gene Ct can be used to get a relative mtDNA copy number after dividing the nuclear gene measurement in half to account for diploid nDNA (Rooney et al., 2015; Gonzalez-Hunt et al., 2016). Additionally, methods involving polyacrylamide gels (PAGE), such as blue-native PAGE (BN-PAGE) can be used to investigate respiratory compound stability or protein-protein interactions because it uses detergents that do not denature protein structures and keep their compounds or protein-protein interactions intact (Na Ayutthaya et al., 2020).

Additionally, microscopy is commonly used to assess mitochondrial function, with a variety of approaches available. To investigate mitochondrial structure or function, transmission electron microscopy can be used to examine mitochondrial shape, number of cristae, and cristae integrity which are all affected by changes in mitochondrial biochemistry, loss of structural integrity, or fluctuations in membrane potential (Vincent et al., 2016). Alternatively, membrane-permeable dyes in are useful for compound or confocal microscopy where dyes localize to mitochondrial membranes following the chemiosmotic gradient established by the ETC that maintains mitochondrial membrane potential. Many dyes exist that make mitochondria visible for localization studies, but fluorescent rhodamine-based dyes, including tetramethylrhodamine methyl ester (TMRM), are positively charged and cross membranes to accumulate in healthy mitochondria which maintain negatively charged membrane potentials (Sukumar et al., 2016). In cases of mitochondrial stress, damage, or dysfunction, the membrane depolarizes and the TMRM dye is not retained leading to a decreased fluorescent signal compared to healthy organelles.

Next, the prevailing biotechnological technique for assessing mitochondria is the Seahorse Extracellular Flux analyzer by Agilent. This machine uses special cell culture plates that have

fluorescent reporters built into them with one that fluoresces in the presence of oxygen and the other fluoresces according to pH (van der Windt et al., 2016). During a run, spectrograph rods drop down into the plate and take fluorescence readings which measure oxygen consumption rates (OCR) and extracellular acidification rates (ECAR). This technique has the ability to distinguish levels of respiration due to the consumption of oxygen from the plate media and compare it to the rate of acidification in the media due to glycolysis. These readouts, which correlate with specific mitochondrial inhibitors provide detailed information about max and min respiratory rates as well as premature proton leakage in the ETC. This method has even been adapted by researchers to work with zebrafish embryos directly (Gilbert et al., 2013; Rollwitz & Jastroch, 2021), which provides a specialized approach to studying some *in vivo* mitochondrial characteristics.

Finally, thinking further from the laboratory bench, mito-nuclear research is perfectly suited to the abundance of sequencing techniques and databases that currently exist. For a conservation researcher, the accessibility and inexpensive cost of sequencing methods such as Illumina NGS approaches along with publicly available sequence and reads databases such as the NCBI Sequence Read Archive (SRA) provide ways to examine a wide variety of organisms from diverse locations around the world. Indeed, whole lineages can be examined computationally using phylogenetics, selection assays, coevolution predictions, and even powerful tools such as Alphafold (Jumper et al., 2021) which can turn a protein sequence into a three-dimensional model to analyze protein structure and function within a matter of hours. These approaches have opened a variety of research avenues in which genetic and proteomic signatures can be used for predictive purposes to locate ideal candidates for experimental lab work.

1.6 Current and future directions of mito-nuclear conservation studies

As previously described, some of the most interesting mito-nuclear research currently resides with groups like the San Diego Frozen Zoo, the San Diego Wild Animal Park, and Revive and Restore which are responsible for recent successes in isSCNT within their organizations. While I stand by the argument that systematic approaches to studying interspecies mito-nuclear compatibility are essential to safeguarding animals and improving the efficacy of isSCNT, the real-world results produced by cloning three critically endangered black-footed ferrets (Novak et al., 2024a) and two Przewalski's horses (Novak et al., 2024b) is awe-inspiring and will lead to an abundance of data on *in vivo* interspecies mito-nuclear interactions that no other current approach comes close to.

Still, previous experimental systems for this field have included the intertidal copepod, *Tigriopus californicus*, which has been carrying the mito-nuclear research field for decades and has contributed data regarding intraspecies mito-nuclear incompatibilities from various populations located off the coast of California (Rawson & Burton, 2002; Barreto et al., 2018), but lacks the breadth of a lineage with a variety of species for investigation. Other models have included *Drosophila* species and have discovered exciting research related to mito-nuclear incompatibilities, but again, most of the work has been performed in lab strains and *Drosophila* also exhibit sexual dimorphism in mitochondrial fitness which deviates from most vertebrates in which sex does not play a direct role in mitochondrial function (Rand et al., 2001). Yeast is also a highly useful system for mito-nuclear studies, especially given that their mtDNA undergoes

recombination and allows for mtDNA editing site-specific investigations of mitochondrial activity, a tool that is currently absent in vertebrate mitochondrial research (Wolters et al., 2018).

Therefore, as I've alluded to, the field of mito-nuclear compatibility research could use a new experimental system ideal for asking questions about mitochondrial function, interspecies relationships and timescales, and is genetically tractable while easy to raise in a laboratory environment. For these purposes, our research group has worked to establish fish within the Danionin lineage as a system for studying interspecies developmental compatibilities that arise from mismatching DNA and maternal factors including mitochondria. Danioninae is a subfamily of Cyprinid fish which includes zebrafish (*Danio rerio*), a well-studied, genetically tractable model organism with a variety of established tools for research in developmental biology. We have even found that various zebrafish-specific antibodies used in research applications such as fluorescence microscopy work across several species within the Danionin lineage due to high sequence conservation (Hansen et al., 2021). Additionally, we have found that hybrids between *Danio rerio* and *Devario aequipinnatus* exhibit an embryonic lethal phenotype with 100% penetrance that appears to correlate with high levels of protein sequence variation in the mitochondrial-encoded Cox2 protein that provides an ideal situation for studying mito-nuclear interactions within the lineage (Trevena et al., 2025; Chamberlain & Pelegri, in review). Moreso, the mitochondrial-encoded Nd5 protein in *Danio albolineatus* contains high numbers of amino acid substitutions (Chamberlain & Pelegri, in review) and another research group has reported a hybrid lethality that exists in crosses involving *D. albolineatus* females with *D. rerio* males which may be related to mito-nuclear deficiencies in respiratory complex I (Delomas &

Dabrowski, 2022). Overall, the introduction of Danionin fish as model system offers a promising avenue for mito-nuclear research and conservation.

In summary, while mitochondrial research is sometimes compared to studying the surface of Mars, which we can see but it is typically out of reach for us to probe, current advancements are promising to break that old perception in respect to mito-nuclear research applications and their utility for conservation. The abundance of tools for researchers to perform *in silico* studies is a promising avenue as we strive to understand mitochondria. Indeed, we live in an exciting time when isSCNT cloning approaches have the potential to rescue species that we were previously hopeless to provide a lifeline for, and the future looks bright.

1.7 Figures

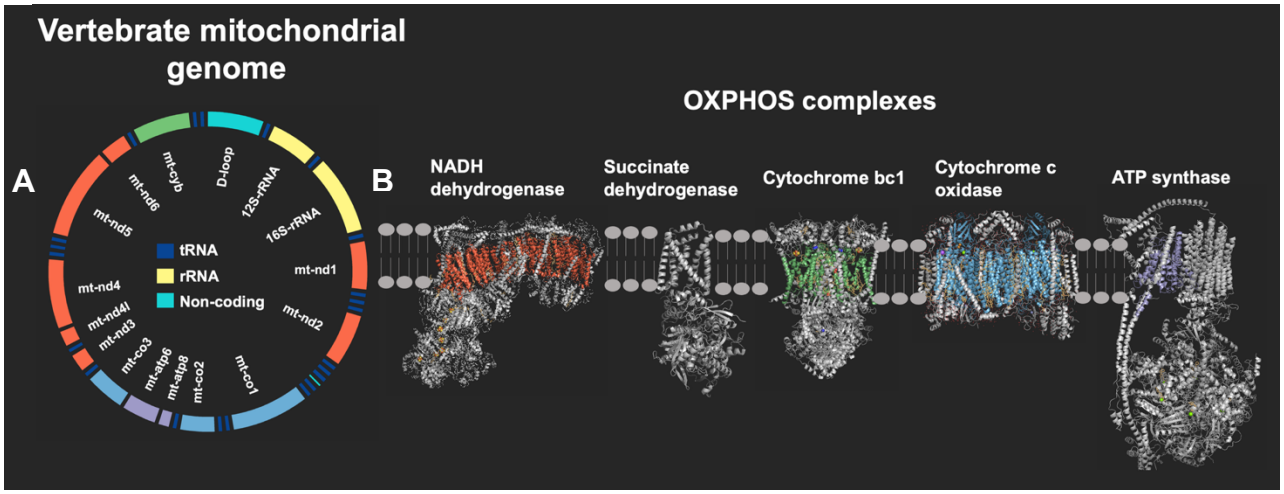


Figure 1. A vertebrate mitochondrial genome and the respiratory complex proteins.

A. The layout of a typical mitochondrial genome containing all coding features with color coding based on feature type or respiratory complex it functions in. *B*. The respiratory complex proteins depicted within the mitochondrial membrane. The subunits of each complex in color are mitochondrial-encoded genes with colors corresponding to the genome in *A*. All gray components are nuclear-encoded subunits. This visualization in meant to help understand the degrees to which mitochondrial-derived proteins and nuclear-derived proteins participate in OXPHOS processes and what it could mean for these subunits to interact at this level.

CHAPTER 2

Appearance of mitochondria genetic variants in the Danionin lineage is both gradual and punctuated

Chamberlain, T. J.[1*], Pelegri, F.[1]

Laboratory of Genetics

University of Wisconsin – Madison

425-G Henry Mall

Madison, WI 53706

USA

* Author for correspondence, email: trevor.chamberlain@bsu.edu

Manuscript in review at *Genome Biology and Evolution*

2.1 Abstract

Barriers to successful interbreeding arise as differential evolutionary forces act on diverging species. One possible barrier is postzygotic isolation, in which mating and fertilization occur, but embryos remain inviable. Members of the cyprinid subfamily Danioninae, which includes zebrafish (*Danio rerio*), comprise a model phylogeny for studying biological mechanisms in a well understood developmental genetics framework. These may include hybrid and nucleocytoplasmic hybrid (cybrid) function in inter-species pairs that are relevant to postzygotic isolation. A source of postzygotic species isolation lies in incompatibilities between the mitochondrial genome of one species and the nuclear genome of another. To begin to study mito-nuclear incompatibilities, we sequenced mitochondrial DNA from zebrafish and related species *D. aesculapii*, *D. kyathit*, *D. nigrofasciatus*, *D. tinwini*, *D. albolineatus*, *D. margaritatus*, *D. erythromicron*, *D. choprae*, *Devario sondhii*, *Devario aequipinnatus*, and *Microdevario kubotai*, assembled their mitochondrial genomes, and identified single amino acid variants (SAVs) among mitochondrial proteins. Our analyses show episodic bursts of SAVs in mitochondrial respiratory complex I proteins in *D. albolineatus* as well as in complex IV proteins in species belonging to *Devario* and *Microdevario* genera. Similarly, we found above-trend, evolutionary rates in *D. albolineatus* complex I and in *Devario* and *Microdevario* complex IV, suggesting differential evolutionary pressures specific to these mitochondrial loci. These proteins and complexes also show signals for relaxed and positive selection during lineage divergence. These studies shed light on evolving mitochondrial genes that underly mitochondrial fitness in hybrids and which likely contribute to mechanisms of post zygotic isolation leading to speciation.

Keywords: mitochondria, zebrafish, speciation, mito-nuclear, evolution, BDMI

2.2 Significance statement

This work uses the Danionin family of fish, which includes the well-established model organism *Danio rerio* (zebrafish), to explore mitochondrial variation in a phylogenetic context, altogether constituting a vertebrate model lineage to investigate mitochondrial evolution and mitochondrial-nuclear incompatibilities. Our initial work in this system compares mitochondrial genomes and highlights both gradual and episodic gains in mitochondrial variation across the phylogenetic clade. The work also allows pinpointing candidate mitochondrial components, whose interactions with nuclear counterparts may lead to mitochondrial dysfunction and developmental inviability in inter-species Danionin hybrids. Our studies contribute to understanding the appearance of mitochondrial variation in phylogenetic space and will facilitate the use of the Danionin model lineage to understand the role of mito-nuclear interactions in reproductive isolation and speciation.

2.3 Introduction

Mechanisms that impact developmental potential include genetic incompatibilities that arise between genetically isolated populations. These incompatibilities are known as Bateson-Dobzhansky-Muller incompatibilities (BDMIs) and arise when alleles in interacting genes gain co-evolving compensatory mutations (Bateson, 1909; Dobzhansky, 1936; Muller, 1942; Burton & Barreto, 2012). One type of BDMI involves nucleocytoplasmic compatibility, in particular involving mitochondrial-nuclear (mito-nuclear) interactions. Due to the rapid evolutionary rate of mtDNA, BDMIs tend to accumulate and result in interspecies incompatibilities (Burton, 2022). These BDMIs are thought to arise as mtDNA evolves new mutations in a population and

interacting nuclear DNA (nDNA) correspondingly acquires compensatory mutations. Over time, mutations accumulate in separated populations until they develop mito-nuclear incompatibilities between one another. Such incompatibilities lead to embryonic inviability and provide the basis for post-zygotic reproductive isolation in hybrids, where half of the genome corresponds to one species and the other half plus cytoplasmic components to another. Additionally, BDMIs including mito-nuclear incompatibilities may interfere with the function and viability of nucleocytoplasmic hybrids (cybrids), where the entire genome corresponds to one species and cytoplasmic components from another, generated in reproductive and regenerative approaches (Hirano, 2001; Nakada & Hayashi, 2011; Patananan et al., 2016).

Interspecies mito-nuclear incompatibilities have notably been studied for decades in various organisms including primates, rodents, copepods, *Drosophila*, wasps, chameleons, dolphins, and even mammoths. Several systems have allowed the studies of incompatibilities in (conventional) hybrids, involving the genome of two species. Copepods hybrids from geographically distinct populations, for example, have allowed identifying multiple single-base changes involved in mito-nuclear incompatibilities leading to a decline in fitness (Ellison & Burton, 2006; Ellison & Burton, 2008; Healy & Burton, 2020). Other notable studies involving detrimental effects on hybrids have been carried in the Drosophilidae family (Meiklejohn et al., 2013), such as a nuclear-encoded mitochondrial tRNA-synthetase from *Drosophila melanogaster* that is unable to attach a tyrosine to the mitochondrial-encoded $tRNA^{Tyr}$ from *D. simulans*. Studies in wasps, including in chameleons, dolphins, and mammoths have also shown sets of amino acid sequence variation arising across taxa that suggest adaptation to varied habitats and likely underly

speciation boundaries between isolated populations (Breeuwer & Werren, 1995; Bar-Yaacov et al., 2015; Caballero et al., 2015; Ngatia et al., 2019).

In work involving primate and rodent cells, interspecies cybrid cell lines were created by fusing mitochondria-less (rho-zero) human or mouse cells to enucleated mitochondria-containing cytoplasmic fragments from related species in their respective clades (Kenyon & Moraes, 1997; Barrientos et al., 1998; McKenzie & Trounce, 2000). This work revealed interspecies mtDNA and nDNA incompatibilities in cybrids, resulting in defects in energy production and leading to cell inviability, between humans and orangutans and between mice and Norwegian rats. These deficiencies revealed changes in mito-nuclear interacting partners that result in functional incompatibilities between lineages, or incompatibility boundaries, corresponding to identifiable degrees of phylogenetic divergence. Cell fusions from closer-related species pairs, those with divergence less than that corresponding to the appearance of an incompatibility boundary ("phylogenetically proximal" species), showed unaffected mito-nuclear interactions and normal mitochondrial function. Conversely, cell fusions from species more distantly related than the incompatibility boundary ("phylogenetically distal" species) exhibited defective mitochondrial function. At the same time, although initial studies indicated catastrophic loss of mito-nuclear interactions at an incompatibility boundary and no significant defects in cybrids of more phylogenetically proximal species (Kenyon & Moraes, 1997; McKenzie & Trounce, 2000), more sensitive assays have shown that effects in the latter can also be more nuanced, with varying, intermediate degrees of functional loss (Barrientos et al., 1998; Dey et al., 2000; Bayona-Bafaluy et al., 2005).

To further study mito-nuclear interaction divergence across phylogenetic space in a vertebrate species, we focus on fish species from the cyprinid subfamily Danioninae, which includes the model species zebrafish (*Danio rerio*) of the genus *Danio*, as well as species in other genera including *Devario* and *Microdevario* (Bleeker, 1863). This family includes species with a high degree of relatedness, with the whole of the *Danio* genus sharing a common ancestor at 13 million years ago (MYA), and the common ancestor of all Danionins at 31 MYA (Rüber et al., 2007; McCluskey & Postlethwait, 2014), allowing to interpret findings in the context of extensive knowledge in the zebrafish system (Streisinger et al., 1981; Corley-Smith et al., 1996; Westerfield, 2007). Moreover, genetic methods developed in zebrafish involving in vitro fertilization (IVF) techniques allow the ready generation of hybrid and cybrids embryos (Trevena et al., 2025; Trevena et al, in preparation). These combined features make the Danionin clade an excellent "model lineage" system to understand the evolution of genomes and their function (McCluskey & Postlethwait, 2014), including the divergence of functionally interacting products after reproductive isolation.

Here, we used the Danionin phylogeny to investigate patterns of mitochondrial genetic variation and possible sources of mito-nuclear incompatibilities that appear during evolutionary divergence. Specifically, we sequenced the mitochondrial genome from twelve species of Danionin fish, namely *Danio rerio*, *D. aesculapii*, *D. kyathit*, *D. nigrofasciatus*, *D. tinwini*, *D. albolineatus*, *D. margaritatus*, *D. erythromicron*, *D. choprae*, *Devario aequipinnatus*, *Devario sondhii*, and *Microdevario kubotai*. Subsequently, we compared and analyzed mitochondrial-encoded protein-coding genes across species to examine patterns of amino acid variability across phylogenetic space. We find expected patterns of gradual variation that follow phylogenetic

divergence but also identify events of larger-than-expected species- and genus-specific episodic variation. Generally, both gradual and episodic variation are more prevalent in complex I and IV, at the beginning and end of the OXPHOS electro transfer chain. Proteins exhibiting larger, episodic variation also show evidence for both relaxed and positive selection. This work sheds light on multiple, superimposed patterns of variation in mitochondrial-encoded proteins in a vertebrate model lineage and suggests candidate mitochondrial genes that may drive interspecies mito-nuclear functional incompatibilities.

2.4 Results

2.4.1 Inference of phylogenetic relationships via mtDNA sequencing in twelve Danionin species

We isolated mitochondrial DNA (mtDNA) via PCR, using five overlapping amplicons as previously described (Dames et al., 2013; Supplemental Figure 1; Supplemental Table 1), and performed de novo sequencing of mtDNA from twelve Danionin species, specifically from the Danioninae subfamily, representing three genera, *Danio*, *Devario* and *Microdevario*. These species included *Danio rerio* (*D. rerio,* zebrafish), *Danio aesculapii* (*D. aesculapii,* panther danio), *Danio kyathit* (*D. kyathit,* orange finned danio), *Danio nigrofasciatus* (*D. nigrofasciatus,* dwarf spotted danio), *Danio tinwini* (*D. tinwini,* gold ring danio), *Danio albolineatus* (*D. albolineatus,* pearl danio), *Danio margaritatus* (*D. margaritatus*, celestial pearl danio), *Danio erythromicron* (*D. erythromicron,* emerald dwarf danio), *Danio choprae* (*D. choprae,* glowlight danio), *Devario sondhii* (*Dev. sondhii,* fireline devario), *Devario aequipinnatus* (*Dev. aequipinnatus,* giant danio), and *Microdevario kubotai* (*M. kubotai*), We successfully assembled

complete mitochondrial genomes from amplified DNA for all species analyzed (GenBank PV357751-PV357794). An exception was *D. erythromicron* (GenBank PV367247-PV367250) where the assembled sequence corresponded to 93.1% the expected full mitochondrial sequence and excluded the D-loop, which could not be amplified (see Methods). Subsequently, *de novo* sequences for the twelve above species were used to create assembled and annotated mitochondrial genomes followed by multiple sequence alignments (MSAs). Mitochondrial sequences for three additional cyprinid species outside the Danioninae subfamily, obtained from NCBI and GenBank databases, *Sundadanio axelrodi* (*S. axelrodi*, Axelrod's rasbora, GenBank: AP011431.1), *Trigonostigma espei* (*T. espei*, lambchop rasbora, NCBI Reference Sequence: NC_015535.1), and *Boraras maculatus* (*B. maculatus*, dwarf rasbora, GenBank: AP011420.1) were included in the analysis as outgroups.

Using these MSAs, we constructed a phylogenetic tree by estimating maximum likelihood to infer relationships and to use in downstream selection analyses. The resulting mtDNA-based phylogenetic tree (Figure 1) reconstructed many of the same species' relationships from previous work exclusively based on whole genome analysis of nuclear markers (McCluskey & Postlethwait, 2014), such as the arrangement of genera within the tree and the assignment of *D. rerio* and *D. aesculapii* as sister species. However, the mitochondrial-based phylogeny differs in that *D. kyathit* shares a clade with *D. rerio* and *D. aesculapii,* instead of with *D. nigrofasciatus* and *D. tinwini* as in McCluskey & Postlethwait (2014) (Figure 1; Supplemental Figure 2A, Supplemental Table 2). We also constructed trees for each protein-coding gene and found varying topologies among the 13 inferred trees (Supplemental Figure 3), suggesting potential

introgression between the species (McCluskey & Postlethwait, 2014; McCluskey et al., 2024; Ren et al., 2025).

Further, we constructed a time-calibrated tree with a Markov chain Monte Carlo (MCMC) Bayesian analysis to assess substitution rates as a function of time (Supplemental Figure 2B). Calibrating using a previously reported estimate of 13-million-year (MY) for the most recent common ancestor (MRCA) at the base of the *Danio* genus (Rüber et al., 2007), we determined the MRCA of the *Danio* and *Devario* clades at a mean height of 25.5 MY and the root of all species in our tree, including all three cyprinid outgroups, at 34.3 MY.

2.4.2 Amino acid conservation and variance within mitochondrial DNA-encoded OXPHOS complexes

To compare variation among mitochondrial-encoded proteins from the twelve Danionin species, we translated the annotated nucleotide sequences, carried out MSAs and identified single amino acid variants (SAVs) in each species, relative to *D. rerio* (for a complete list of SAVs, see Supplemental File 1). Initially, we pooled SAVs according to OXPHOS complexes with mitochondrial-encoded proteins (I, III, IV or V) and calculated the total percent of SAVs, relative to total number of residues per complex (Figure 2A). We find that as a group and when compared to *D. rerio*, OXPHOS complex I proteins and V had the highest and second highest, respectively, percent of SAVs across analyzed Danionin species. An exception of this pattern was complex IV in *Devarios* and *Microdevarios*, which exhibits increased variation compared to that across *Danios*, in particular exhibiting the highest percent of SAVs among all complexes in *M.*

kubotai, and the second highest (after complex I) in *Dev. aequipinnatus*. Overall, the observed patterns are consistent with previous studies generally showing greater inter-species protein variation in OXPHOS complex I and/or V proteins (da Fonseca et al., 2008; Castellana et al., 2011) and also reveals genus- and species-specific exceptions in protein variation per complex.

2.4.3 Multiple patterns of variation in mitochondrial DNA-encoded OXPHOS genes

We used positional data of SAVs from the MSAs as input for Circos, a graphical representation of all SAVs for each species (Krzywinski et al., 2009), in relation to *Danio rerio.* This allowed us to examine mitochondrial variation signals at the amino acid level, with *D. rerio* positioned on the outermost ring and the remaining species moving inward in order of phylogenetic distance (Figure 2B).

Using the Circos plot, we observed a pattern of gradual accumulation of SAVs, relative to *D. rerio*, among Danionin mtDNA-encoded proteins, with an increasing number of SAVs following increasing inter-species phylogenetic divergence (Figure 2B,C). This gradual increase appears to occur in all proteins, although is particularly apparent in complex I proteins Nd1, Nd2, Nd4 and Nd5. For Nd5, additional variation overlying the gradual increase is apparent for the *D. albolineatus* protein (see below). For the highly conserved complex IV proteins, SAVs are sparsely distributed in Cox2 or Cox3 or largely absent in Cox1 within the *Danio* genus, with variation, in comparison to *D. rerio* and other *Danios*, appearing in the sister *Devario* and *Microdevario* genera (see below).

The Circos visualization shows a significantly high number of SAVs in *D. albolineatus* complex I Nd5 protein, in relation to *D. rerio* Nd5, that stands out above the gradual trend of variant increase (Figure 2B,D). Similarly, a dramatic increase in SAVs in complex IV COX proteins between *Devario*/*Microdevario* and *Danio* genera contrasts sharply with the low number of SAVs in these proteins across the *Danio* genus (Figure 2B,E). Such apparent sharp increases indicated potential events of rapid variation in specific mitochondrial genes and lineages.

We used the program SIFT (Sorting Intolerant From Tolerant), again with each *D. rerio* protein sequence as input, as a proxy for changes likely leading to significant protein conformational variation (Ng & Henikoff, 2003), even if these changes are presumed to be functional in all Danionin lineages due to ongoing selection for mitochondrial function. Plotting of high SIFT score residues onto the SAV Circos plot (Figure 2B, gray highlights; Supplemental File 1, asterisks), revealed that candidate conformational variants generally followed the observed patterns of gradual and rapid variation shown by all SAVs.

We repeated the analysis relative to *D. aesculapii*, the sister species of *D. rerio,* instead of relative to *D. rerio,* with the patterns of SAV accumulation remaining consistent (Supplemental Figure 4). Thus, the analysis shows a gradual trend of accumulation of variation that is gene-specific and which directly correlates to phylogenetic divergence. Additionally, the data reveals multiple instances where, on a gene- and lineage-specific manner, mtDNA genes acquire a more rapid, or episodic, increase in the number of changes above a baseline corresponding to a gradual accumulation trend.

2.4.4 Detection of episodic variation events using regression analyses

To quantitatively analyze differences in variant accumulation in Danionins, we performed a regression analysis for each mtDNA-encoded protein across phylogenetic space. Specifically, we represented the number of SAVs in a scatterplot of all pairwise combinations of genes and species as a function of phylogenetic distance within the lineage (Figure 3A, see Methods for details). Use of all pairwise combinations allowed us to visualize SAV variation relative to inter-species phylogenetic distance across Danionins without a bias arising from using the *D. rerio* sequence as a reference. Additionally, we plotted all regression lines using a sliding window of specified amino acid stretches (see Figure 3A, Methods) across the length of the protein to increase permutations in the analysis and reduce pseudoreplication bias that can arise through sequence analysis in closely related lineages (Garland, 2001; Millar & Anderson, 2004; Maddison & Fitzjohn, 2015). The regression trend confirms the pattern of gradual SAV accumulation observed in the Circos plot exhibiting a range of rates for SAV accumulation, with the highest rate in complex I Nd5, the lowest rates in complex III Cytb and complex IV Cox1 and Cox3 genes, and intermediate rates for other proteins.

Regression trends also allowed identifying points corresponding to pairwise combinations for specific genes that fell above or below the range of expected regression lines within Danionins, excluding outgroups used for regression trajectories (Figure 3A, see Supplemental Table 3 for a list of all above- and below-range pairwise comparisons). Notably and consistent with episodic variation, most pairwise outliers were due to a ratio of SAVs to protein length that is significantly

higher than expected based on inter-species phylogenetic distance. Outlier pairwise combinations with higher-than-expected SAVs, given phylogenetic distance, include complex I Nd5 and complex IV Cox2. For Nd5, all (8 of 8) possible pairwise comparisons involving *D. albolineatus* and other *Danio* species in the analysis showed pairwise outlier variation values, confirming inferences from the Circos plot of rapid appearance of variation in *D. albolineatus* Nd5 (Figure 3B; Supplemental Figure 5A). For Cox2, all (27 of 27) possible pairwise comparisons involving inter-genus pairings between *Danio* genus species and *Devario* or *Microdevario* species, as well as all (2 of 2) *Devario* to *Microdevario* combinations, showed higher than expected pairwise outlier variation values (Figure 3C; Supplemental Figure 5B). Variation also appeared above average regression lines for Cox1 and Cox3, showing an apparent general trend for other complex IV proteins, although for these two genes variation still fell withing the regression line range. Other instances of outlier variation, either above-regression range (Cytb, Cox2, Atp 6 & Atp8) or below-regression range (Cox2) also appeared, albeit solely in pairwise comparisons with outlier comparators.

Thus, multi-species regression analysis, not biased on the sequence of any given species, corroborates and expands on observations in the *D. rerio*-centered Circos plot alignment, providing further evidence for gradual and episodic variation within the Danionin phylogeny. Most strikingly, overlayed onto gradual trends of divergence expected from increasing phylogenetic distance, we identify foci of above-background divergence in multiple mtDNA-encoded genes in various lineages, particularly Nd5 in *D. albolineatus*, and Cox proteins in *Devario* and *Microdevario*.

2.4.5 Increased rates of mitochondrial gene variation in Danionin gene tree branches

To gather further support for incidences of above-background variation in Danionins, we assessed mutation rates along branches of their phylogeny (Figure 4). Briefly, we used Bayesian estimates through the software BEAST 2 to estimate mutation rates for each branch of a tree constructed for each gene (Bouckaert et al., 2019), then compared the distribution of mutation rates per million years among branches of the phylogenetic tree corresponding to the twelve Danionins assessed in our study (see Methods, Figure 4). As expected, complex I Nd5 shows above-background variation in *D. albolineatus*. Additionally, the analysis also shows faster rates of evolution in complex IV cox genes in branches leading to *Devario* and *Microdevario* lineages. Unexpectedly, many genes (Nd1, Nd2, Nd3, Nd4l, Nd4, Nd6, Cox1 and Atp6) exhibit above-background variation in *D. rerio*, an effect that may not have been apparent in the Circos plot and regression analysis due to being based in relation to *D. rerio* itself. To test if the increased mutation rate in these *D. rerio* genes was due to domestication of the analyzed laboratory strain, we repeated the analysis using the mitochondrial sequence for the Cooch Behar (CB) *D. rerio* strain, recently derived from the wild (Suurväli et al., 2020) (BankIt submission identifier 2974502). We found protein variation rates in the DB strain similar to those observed in the AB strain (Supplemental Figure 6), indicating that the identified trends also occur in the wild *D. rerio* population.

To further assess rates of genetic variation, we incorporated values from maximum-likelihood trees for each mitochondrial-encoded protein-coding gene into constrained backbone tree topologies, each of which excluded the corresponding analyzed gene to reduce an intrinsic bias introduced by the same gene. A similar approach was carried out on a concatenation of all genes

per respiratory complex, with the backbone tree excluding the genes for the corresponding analyzed complex. For each gene or concatenated complex tree, log ratios were binned, and histograms were made to search for length ratios over a 3 mean absolute deviation (MAD) upper limit, with a mean representing a log ratio of substitution rates expected among all mitochondrial gene trees (Supplemental Figure 7). Subsequently, branches with log ratios greater than a 3 MAD upper limit were highlighted on each respective gene tree denoting above-average increase in variation along those branches (Figure 5). This analysis showed that *D. albolineatus* accumulated above-average variation in branches leading to Nd5 in the gene tree as well as to OXPHOS complex I in the concatenated complex tree. Additionally, branches with significantly high substitution rates appear in the Cox2 gene tree and complex IV tree in branches leading to *Devario* and *Microdevario* genera. Thus, the observed signals of increased rates of variation in Nd5 and complex IV are consistent with SAV assays and mutation rate analyses.

2.4.6 Identification of mitochondrial gene episodic positive selection

To explore whether branches with inferred episodic variation within the Danionin lineage may have experienced bursts of positive selection, we used BUSTED (Branch-Site Unrestricted Statistical Test for Episodic Diversification) (Pond, et al., 2005; Murrell, 2015) setting paths from the *Danio and Devario* MRCA to each species as foreground branches. We performed this analysis on each mitochondrial-encoded protein-coding gene, a concatenation of all genes per respiratory complex, and in a concatenation of all mitochondrial-encoded H-strand protein-coding genes. BUSTED scores that were positive for episodic diversifying selection were then visualized on each corresponding tree (Figure 6). The analysis indicates signals of episodic diversifying selection in *D. albolineatus* complex I, which contains Nd5, as well as in *Devario*

and *Microdevario* Cox genes and more broadly within the Cox protein-containing complex IV. Concatenation of all H-strand mitochondrial protein-coding genes highlighted diversifying selection in *D. albolineatus* and *Devario* and *Microdevario* genera, consistent with the observed increased variation in Nd5 and Cox genes, respectively, in those lineages.

2.4.7 Detecting relaxed selection among Danionin species

We also used the software RELAX to calculate relaxed or intensifying selection across the analyzed Danionins (Pond, et al., 2005; Wertheim et al., 2015). Relaxed selection results from the weakening of selective pressure on a protein leading to genetic drift, hypothesized to drive evolutionary innovation. On the other hand, intensifying selection reflects the strengthening of selective pressure in rapid adaptation towards specific trait variation. We performed these tests on trees for each gene, each concatenated complex, and a concatenation of all H-strand protein-coding genes, setting the branches in each path as the foreground against all other species and visualizing k scores associated with relaxation and intensification (Figure 7). The analysis reveals signals for strong relaxation in Nd5 and complex I in *D. albolineatus*, as well as in the branch leading to Nd5 in the tree for all concatenated mitochondrial genes. Similarly, strong relaxation is detected in Cox1, complex IV and concatenated genes for the *Devario* and *Microdevario* clade. Thus, signals associated with increased mitochondrial gene variation and episodic positive selection appear to also be associated with relaxed selection. These findings are consistent with a role for increased random variation as a substrate for subsequently selected novel functional adaptations in the OXPHOS chain.

2.5 Discussion

Our studies identify patterns of protein variation among mitochondrial genes within the Danionin family, revealing not only the expected gradual accumulation of variation in association with phylogenetic divergence but also rapid, episodic changes of increased appearance of genetic variation in specific genes and lineages.

2.5.1 Gradual and punctuated patterns of protein diversification in Danionins

Protein variation is predicted to arise gradually during evolution, through the combined action of mutation and fixation. Fixation of new variants, in turn, is dependent on constraints on variation through purifying selective pressure due to the need to maintain functionality. Given their involvement in energy production this selective constraint is particularly important for components of the OXPHOS electron chain reaction, leading to overall relatively high rates of conservation for mitochondrial gene products (Stewart et al., 2008). Despite such high conservation, we find a gradual appearance of variation in mitochondrial protein genes across all complexes that broadly correlates with phylogenetic divergence within Danionins, consistent with the expected accumulation of gene variants through evolutionary time.

Superimposed on the expected pattern of gradual divergence, we also identify dramatic increases in variation in some mitochondrial protein genes and lineages. This rapid, episodic variation can appear relatively confined, as observed in complex I Nd5 protein alone and solely in the *D. albolineatus* lineage. Rapid variation can also be more extensive in terms of affected genes and/or lineages, such as complex IV Cox2 protein and to a lesser extent Cox1 and Cox3 in the

genera *Devario* and *Microdevario* in relation to Danios. The latter case also highlights rapid fixation of variation in phylogenetically basal lineages, which may have had a role in the appearance of entire genera. Of interest, while variation in complex IV cox proteins within the *Danio* lineage is minimal, consistent with high constraint in their function, episodic variation in basal lineages for *Devario* and *Microdevario* appears to overcome such strong constraints during lineage diversification. The observation that inferred episodic variation events range from involving a single species to multi-genera lineages likely reflects the sporadic yet regular appearance of such events within an evolutionary time frame, coupled to their identification in our analysis at different stages of their occurrence in the phylogeny. It is tempting to infer that such events promote lineage diversification, as in the case of changes leading to *Devarios* and *Microdevarios*, although more studies will be needed to validate this hypothesis.

In previous work, we found loss of fitness in hybrids between species within the Danionin clade, namely hybrids of *D. rerio* eggs and sperm of species from increasing phylogenetic distances (Trevena et al., 2025). Specifically, hybridization within the *Danio* genus, while still yielding a fraction of viable hybrid adults, shows embryonic and larval lethality whose expressivity and penetrance increases with phylogenetic distance. Furthermore, there was more consistent, fully penetrant embryonic lethality in the inter-genus cross between *D. rerio* and *Dev. aequipinnatus*. The gradual increase in mitochondrial protein variation we observe mirrors the previously observed loss of fitness in hybrids and is consistent with the gradual accumulation of BDMIs following reproductive isolation and species divergence. Within *Danios*, *D. rerio/D. albolineatus* hybrids from *D. rerio* eggs and *D. albolineatus* sperm do not show evidence of increased lethality above expected trends (Trevena et al, 2025). However, the reciprocal cross, from *D.*

albolineatus eggs and *D. rerio* sperm, results in fully penetrant embryonic lethality (Delomas & Dabrowski, 2022). These observations are consistent with known asymmetric effects on hybrid fitness and more specifically with a BDMI leading to inviability caused by Nd5 episodic variation in the *D. albolineatus* lineage. Moreover, the fully penetrant embryonic lethality in *Danio-Devario* hybrids correlates with associated rapid, episodic variation in complex IV cox genes, known to be essential for embryonic viability (Baden et al., 2007). Thus, our findings in mitochondrial protein episodic variation in Danionins generally correlate with known phenotypic outcomes in their hybrids, although molecular and cellular studies will be needed to confirm relevant BDMI-causing gene interactions.

2.5.2 Increased evolutionary rates in specific OXPHOS pathway complexes

All mitochondrial protein-coding genes, which have not migrated to the nucleus after mitochondrial evolution from an endosymbiont (Gray et al., 2001), participate in the OXPHOS cascade. This has led to the hypothesis that they are selected as an evolutionary correlated set, perhaps to better adapt to different environmental conditions that benefit from close regulation of energy production pathways (Ballard et al., 2007; Castellana et al., 2011). In this context it is of interest that this work and that of others identify higher rates of mitochondrial protein variation in both complex I and IV (Adkins & Honeycut, 1994; Barrientos et al., 1998; Grossman et al., 2001; Gershoni et al., 2010), which constitute the initial and final steps of the electron cascade. Similarly, we see apparently correlated changes in lineages undergoing episodic variation, with greater variation for the sets of complex I genes in *D. albolineatus* and complex IV genes in the lineage leading to *Devarios* and *Microdevarios*. Complex I is a primary entry point for electrons from NADH using the complex I/III/IV pathway that contributes to the energy-generating

mitochondrial protein gradient (Nolfi-Donegan et al., 2020; Okoye et al., 2023). On the other hand, complex IV is the final complex involving electron transport to its final acceptor, molecular oxygen, to generate water, the last step contributing to the H+ gradient used to generate ATP (Sharma et al., 2009; Zhao et al., 2019), and has been implicated in feedback inhibition of OXPHOS activity through allosteric binding by ATP (Kadenbach, 2021). An alternate electron donor pathway involves electrons from succinate via the coenzyme Q-junction into complex II/III (Tretter et al., 2016; Nesci et al., 2023), yet this electron donor step is not associated with proton transport and thus is thought to contribute less energy to the overall electron chain reaction process.

Higher rates of variation in the initial and final steps of a pathway concurs with characteristic patterns of regulation within enzymatic pathways, where regulation often occurs at the initial, first step of the pathway to control commitment into the pathway, and the final step may be involved in feedback inhibition after sufficient product generation. Thus, increased rates of mitochondrial protein variation in Danionins occur in complexes expected to have key roles in the regulation of the electron transport chain, inherited as a linked gene set for those protein components encoded in the mitochondrial genome. Complex I is also thought to constitute a primary site for electron leakage, a process where electron donor activity becomes uncoupled to energy production and which is thought to allow oxygen sensing and redox signaling to influence cellular outcomes (Nolfi-Donegan et al., 2020; Okoye et al., 2023), providing potential additional mechanisms that may drive observed increased variation in this complex.

2.5.3 Similarities to episodic events in other lineages

Our findings related to episodic variation are reminiscent of previous works in primates, which identified a rapid increase of amino acid variants in specific complexes (Kenyon & Moraes, 1997; Barrientos et al., 1998). In the case of primates, the inherently high rate of mitochondrial genome evolution obscured a signal for rapid evolution in complex I genes encoded in mtDNA, but the relative rates of the nuclear-encoded interacting subunits were sufficiently increased to be detectable, and led to findings that mitochondrial ND4, ND5, and ND1 were coevolving with these rapidly changing nuclear subunits (Mishmar et al., 2006; Gershoni et al., 2010). The high rate of non-synonymous substitutions observed in *D. albolineatus* Nd5 is notable because the signal is readily detectable and not obscured by the underlying mtDNA variation.

While complex IV genes were not initially noticeable in any great ape sequence comparisons, a later study noted that monkey and ape COX2 sequences had an amino acid substitution rate that was nearly doubled compared to other genes (Adkins & Honeycutt, 1994). Further analyses revealed that COX1 and COX2 exhibited rapid evolution with several interacting nuclear-encoded complex IV subunits exhibiting signs of compensatory mutations (Wu et al., 2000; Grossman et al., 2001; Doan et al., 2004; Osada & Akashi; 2012). Thus, rapid complex IV gene variation in primates also parallels observed elevated rates of evolution in *Devario* and *Microdevario*.

Future studies on Danionins will focus on identifying nuclear-encoded mitochondrial products that coevolve with identified rapidly evolving mtDNA-encoded products such as Nd5 in *D.*

albolineatus and Cox genes in Devarios and Microdevarios, representing interacting sets of products that underly mito-nuclear incompatibilities in hybrids.

2.5.4 Variation, selection and genetic history

Our phylogenetic analysis reveals increased rates of variation corresponding to episodic events in complex I Nd5 for *D. albolineatus*, and complex IV Cox genes for *Devarios* and *Microdevarios*. Intriguingly, branches for these specific genes and lineages also show signals of increased relaxed purifying selection and positive selection. These two seemingly opposite processes have been shown to coexist (Zhao et al., 2024; Liu et al, 2025), though to proceed by an initial relaxation of selection leading to an increase in new variants including potentially advantageous mutations, with specific residues or domains being subsequently selected for through positive selection.

The initial step in this sequence, relaxed selection, is thought to appear due to changes in the environment that result in decreased need for function. In the context of mitochondrial function, relaxation of purifying selection may occur under low-oxygen conditions that favor glycolysis, which can occur anaerobically, over oxidative phosphorylation (Henderson, 1969; Kierans & Taylor, 2021). Other conditions such as increased reliance on fast-twitch muscle fibers required for burst swimming behavior may also favor anaerobic energy-generation pathways (Picard et al., 2012). In this context, it is interesting that *D. albolineatus*, which exhibits relaxation of purifying selection in Nd5 at the initiating step of the electron chain, is the only *Danio* species of the analyzed set that populates not only South and Southeast Asia continental areas as other

Danios but also Southeast Asia islands including Sumatra and Phu Quoc Island. These islands became separated from the mainland well before (65-250 MYA) the appearance of the *D. albolineatus* lineage, suggesting that this species has traversed water straits separating the continental land from this island, likely facilitated by the intermittent formation of land bridges (Sholihah et al., 2021) and/or glacial cycles (de Beaufort, 1951; McConnell, 2004), and possibly taking advantage of low water salinity in particular during the monsoon season (Steinke et al., 2006). It is possible that navigation in lower oxygen-carrying capacity of brackish water and/or fast currents may have selected for increased reliance on anaerobic energy production, leading to relaxation of purifying selection in complexes involved in oxidative phosphorylation.

A similar pattern of episodic increase in variation, including signals of relaxation of selection and bursts of positive selection can be detected in *Devario* and *Microdevario* complex IV Cox genes. Factors such as oxygen availability or swimming patterns may have similarly resulted in relaxation of selection. Of interest, many *Devarios* (though not *Microdevarios*) have an increased size compared to *Danios*, and larger fish can be associated with lower basal metabolic rate and/or increase ability to carry out glycolysis (Urbina & Glover, 2013; Müller et al., 2023). It is tempting to speculate that changes in oxygen utilization patterns in lineages leading to *Devario* and *Microdevario* may have primed OXPHOS chain activity and basal metabolism to facilitate the appearance of larger body size through a greater ability to use glycolytic pathways, additionally associated with relaxation of selection and the accumulation of new variants.

Small effective population sizes alone are also linked with mechanisms of relaxed selection (Kliman et al., 2000; Woolfit & Bromham, 2003), in particular for cellular organelle genomes because of their haploid component and uniparental inheritance (Lynch & Blanchard, 1998; Neiman & Taylor, 2009). It is thus also possible that historical small effective population size, in particular in species with significant representation in islands such as *D. albolineatus,* contributed to the observed signals of relaxed selection, similar to other cases of increased relaxed selection in fragmented populations (Iverson et al., 2024).

Relaxation of purifying selection may have been concurrently or subsequently coupled to the detected positive selection at specific amino acids or domains (Rawson & Burton, 2006; Hunt et al., 2011; Zheng et al., 2024), which we observe in *D. albolineatus* Nd5 and *Devario/Microdevario* cox genes. Altogether, our results associate relaxed and positive selection as important mechanisms for diversification, as previously suggested (Lahti et al., 2009; Zhao et al., 2024).

In summary, analysis of the Danionin lineage mitochondrial genomes reveal patterns of accumulation of new variation and inter-species sequence divergence, in particular a combination of gradual and episodic increase in variation that is gene- and lineage-specific. Our studies also provide support for relaxed selection together with positive selection as mechanisms for the acquisition of new traits that can confer adaptive properties to new ecosystem or organismal conditions.

2.6 Methods

2.6.1 Animal husbandry

Danionin fish were raised and maintained in standard conditions including a temperature of 28.5° Celsius (C) (Brand et al., 2002, Hansen et al., 2021, Trevena et al., 2025). All procedures were conducted according to the University of Wisconsin-Madison and Institutional Animal Care and Use Committee (IACUC) guidelines (Protocol ID: M006714).

2.6.2 Amplicon-based sequencing and de novo assembly of Danionin mtDNA

We sequenced mtDNA from four individuals including two males and two females of each species including *D. rerio*, *D. aesculapii*, *D. kyathit*, *D. nigrofasciatus*, *D. tinwini*, *D. albolineatus*, *D. margaritatus*, *D. erythromicron*, *D. choprae*, *Dev. aequipinnatus*, *Dev. sondhii*, and *Microdevario kubotai*. Genomic DNA (gDNA) extraction was performed by anesthetizing each fish with 0.014% tricaine (ethyl 3-aminobenzoate methanesulfonate salt; MilliporeSigma, Burlington, MA, USA, catalog no. A5040), performing a tail fin transection using a razorblade, followed by DNA extraction using a DNeasy® Blood and Tissue kit (QIAGEN Inc., Venlo, The Netherlands, catalog no. 69504) into molecular-grade nuclease-free water (Invitrogen™, Waltham, MA, USA, catalog no. AM9938). The tail fin transection was distal to the base of the tail to allow for full tail regeneration.

We isolated mtDNA from gDNA using PCR amplification of overlapping amplicons. Each mtDNA was amplified twice to yield one pair of overlapping amplicons, named zmt1 and zmt2,

and a separate trio of overlapping amplicons, named mito1, mito2, and mito3, with amplicons ranging from 4,436 bp to 9,257 bp (Supplemental Figure 1; Supplemental Table 1; Dames et al., 2013). We used a 3-step PCR method with Primestar® GXL DNA Polymerase (Takara Bio USA, Jose, CA, USA, catalog no. R050B) to perform PCR on each amplicon with an initial denaturation step of 2 minutes at 98ºC, then 30 cycles of denaturation (98ºC for 10 seconds), annealing (15 seconds with temperatures range from 55ºC to 60ºC), and extension (68ºC for 1 minute/kB), a final extension step for 5 minutes at 68ºC, then held at 4ºC. In many species, we identified conserved primers that could be shared across species, although annealing temperatures differed slightly between some primer pairs (Supplemental Table 1). One primer pair, named zmt2 forward and zmt2 reverse, was conserved across all species and allowed the amplification of roughly half of the mtDNA and could be used with Sanger sequencing to elucidate new primer pairs for other amplicons. In cases where sequence conservation was not sufficient to amplify certain amplicons, we performed Sanger walks along working amplicons to design primer pairs for additional locations. For example, by performing Sanger sequencing at either end of the zmt2 amplicon, we designed species-specific primer pairs for zmt1using reverse complement sequences of zmt2, allowing to amplify the additional mitochondrial genome regions. Using this approach, we amplified the entire mtDNA for all species. One exception was the case of *D. erythromicron,* where we were unable to amplify the D-loop region, reaching instead an expected 93.1% mitochondrial genome coverage.

Once PCR was complete, we purified each amplicon using a DNA Clean & Concentrator™ -5 kit (Zymo Research, Irvine, CA, USA, catalog no. D4014) then pooled all five amplicons equimolarly. DNA Libraries were prepared by The University of Wisconsin – Madison

Biotechnology Center's DNA Sequencing Facility (Research Resource Identifier – RRID:SCR_017759) with Celero EZ DNA-Seq (Tecan Genomics, Redwood, CA, USA, catalog no. 0568, 0569) using 100 ng of the pooled amplicon DNA. Enzyme fragmentation was performed according to kit protocols with the parameters set for 25°C for 10 minutes, 55°C for 10 minutes, then held at 4°C. Ligation parameters were 25°C for 30 minutes, 70°C for 10 minutes, and 10°C hold. PCR parameters for amplification of adapter ligated libraries were 72°C for 2 minutes, 95°C for 3 minutes, 6 cycles (98°C for 20 seconds, 65°C for 30 seconds, 72°C for 1 minute), 72°C for 1 minute, then held at 10°C. Sequencing was then performed on a NovaSeq 6000 S4 flow cell (Illumina, Inc., San Diego, CA, USA) using paired-end sequencing (2x150 bp) for 1 million reads (Modi et al., 2021).

Following sequencing, reads in the fastq files were assessed for quality using FastQC v0.11.7 (Andrews, 2010), then trimmed for adapters and quality using fastp v0.20.0 (Chen et al., 2018). SPAdes v3.14.1 was then used to perform de novo assembly from these reads by running a paired-end assembly using the --cov-cutoff option to account for high levels of initial PCR-amplified DNA (Prjibelski et al., 2020). The threshold levels specified for a coverage cutoff tended to range from 2,000-fold coverage to 15,000-fold coverage and were considered adequate when the assembly reached a single contig encompassing a complete sequence matching expected mtDNA genomic features including bp length and gene order. Additionally, visual analysis of the SPAdes assembly .gfa files was done using the software Bandage v0.8.1 to check for uniform genome assembly at the set threshold levels and confirm identity to *D. rerio* mtDNA (NCBI Reference Sequence: NC_002333.2) when using the BLAST tool built into Bandage (Wick et al., 2015).

For *Dev. aequipinnatus* and *Dev. sondhii*, the D-loop would not resolve using typical de novo assembly methods. Therefore, we used the software aTRAM which iteratively blasts sequencing reads to a reference sequence to generate a new reads file, then performs an assembly with the given reads using SPAdes for assembly (Allen et al., 2015; Prjibelski et al., 2020). First, we followed the aTRAM protocol to generate a BLAST database from respective reads files. Next, we used aTRAM to perform the D-loop assembly using the *Dev. interruptus* (NCBI Reference Sequence: NC_066649.1) corresponding flanking coding regions as a reference. aTRAM was allowed to run for up to five iterations and a coverage threshold was set in SPAdes for reads reaching between 1000- and 10000-fold coverage. The top-scoring contigs were then chosen as the representative D-loop sequence, the flanking coding regions were removed, and the resulting D-loop sequence was incorporated into the final mitochondrial genome.

For *D. erythromicron*, the same protocol was followed, but without amplicons covering the D-loop, the tRNA-Phe gene, and most of the 12s rRNA gene. The obtained amplicons covered all 35 other mitochondrial genes including all 13 protein-coding genes, 21 tRNA genes, and the 16s rRNA gene.

Following sequencing, genome annotation was performed using a combination of manual identification of coding features such as start codons and identity to tRNAs as well as comparison to annotations output from the online tool MitoAnnotator (Iwasaki et al., 2013; Zhu et al., 2023). For further downstream analyses, each sequence was also translated to a protein

sequence using the transeq tool in the Emboss v6.6.0.0 software suite using codon table 2 for vertebrate mitochondrial codons (Rice et al., 2000). Consensus sequences for each protein and nucleotide were generated for subsequent analyses using the conseq tool in Emboss using default settings where each bp or nucleotide consensus is considered at a 70% level of identity.

To assemble the mtDNA of the *D. rerio* CB strain, we accessed whole genome Illumina HiSeq X Ten paired-end NGS sequencing reads (accession ERR3332331) through the NCBI Sequence Read Archive (SRA) (Leinonen et al., 2011; Sayers et al., 2024). Next, we used fastq-dump in the SRA Toolkit to unpack the SRA data, split reads into their respective paired-end files, and assign sequence IDs to each read (SRA Toolkit Development Team, n.d.). Quality was then checked with FastQC v0.11.7 (Andrews, 2010) and the reads were trimmed for potential adapters and quality using fastp v0.20.0 (Chen et al., 2018). Then, mitochondrial sequence reads were isolated from whole genome data and assembled into a mitochondrial genome using a two-step procedure in MITObim v.1.9.1 (Hahn et al., 2013). First, a reference file was generated by mapping reads to *D. rerio* mtDNA (NCBI Reference Sequence: NC_002333.2) using MIRA v.4.0.2 (Chevreux et al., 2004; Hahn et al., 2013). Then, MITObim was run for a maximum of ten iterations to assemble the reads from the reference file. The resulting sequence file was then annotated using MitoAnnotator (Iwasaki et al., 2013; Zhu et al., 2023).

2.6.3 Maximum likelihood tree construction

We constructed an MSA using mitochondrial genomes from representative samples including four newly sequenced genomes for each of the selected Danioninae fish species plus previously

available sequences from *Sundadanio axelrodi* (GenBank: AP011431.1), *Boraras maculatus* (GenBank: AP011420.1), and *Trigonostigma espei* (NCBI Reference Sequence: NC_015535.1), which we used as outgroup species. For the MSA, only genes coded on the mtDNA H-strand were used, which included the 16S rRNA, 13 tRNAs ($tRNA^{val}$, $tRNA^{leu1}$, $tRNA^{ile}$, $tRNA^{met}$, $tRNA^{trp}$, $tRNA^{asp}$, $tRNA^{lys}$, $tRNA^{gly}$, $tRNA^{arg}$, $tRNA^{his}$, $tRNA^{ser2}$, $tRNA^{leu2}$, and $tRNA^{thr}$), and 12 protein-coding genes (*mt-nd1*, *mt-nd2*, *mt-nd3*, *mt-nd4l*, *mt-nd4*, *mt-nd5*, *mt-cyb*, *mt-co1*, *mt-co2*, *mt-co3*, *mt-atp6*, and *mt-atp8*). The H-strand was used for the analysis given its coding for all but one (*mt-nd6*) mitochondrial-encoded OXPHOS protein genes, which made annotation more efficient and provided less partitions and evolutionary modeling for tree estimation. For gene pairs with open reading frame overlap (*mt-atp6* and *mt-atp8*, *mt-nd4l* and *mt-nd4*, and *mt-nd5* and *mt-nd6*), both nucleotide and amino acid protein-coding sequences were trimmed to remove the overlaps from the analyses. Each translated amino acid sequence as described above was concatenated into a fasta file and aligned using ClustalW2 v2.1 (Larkin et al., 2007). Positions containing insertions and deletions or ambiguous amino acids were removed from the amino acid alignment and corresponding sites in each nucleotide sequence. Then, pal2nal v14 was used to generate codon alignments from inputs of both the amino acid alignment and protein-coding nucleotide sequence files (Suyama et al., 2006). Once all codon alignments were generated for all 26 H-strand genes, they were concatenated into one sequence per sample with the genes in the same order as they appear in a canonical mitochondrial genome sequence with $tRNA^{val}$ first and $tRNA^{thr}$ last, and clustalW2 was used to perform a final MSA for tree construction.

Once the MSA file was generated, separate partition files were generated, one for all tRNA genes, one for the 16s rRNA, and three partitions for each protein-coding gene, each partition

representing the three amino acid codon positions. Next, we used ModelTest-NG v0.1.7 to estimate evolutionary models for each partition (Darriba et al., 2020). We used the AICC partition model from ModelTest-NG and used RAxML-NG v1.0.2 to generate a tree using maximum likelihood estimates (Kozlov et al., 2019). RAxML-NG was run with a TBE bootstrap cutoff of 0.01 and options were given for 100 both of random and parsimony starting trees. Following tree construction, FigTree v1.4.4 was then used to visualize tree results (FigTree, 2006). To ensure tree accuracy, tree construction was followed up with topology testing using IQ-TREE v2.1.2 (Minh et al., 2020). Alternative tree topologies assembled using various partition schemes and the published Danionin tree from McCluskey & Posthlethwait (2014) were tested using -zb 10000, -zw, and -au tree statistical comparison tools to ensure the assembled tree had the highest likelihood for accurate topology (Supplemental Table 3).

2.6.4 SAV comparisons to *Danio rerio*

The translated consensus sequences for each mitochondrial-encoded protein-coding sequence (Nd1, Nd2, Nd3, Nd4l, Nd4, Nd5, Nd6, Cytb, Cox1, Cox2, Cox3, Atp6, and Atp8) from each of four individuals per species were used to generate MSAs for each protein using ClustalW2 v2.1 (Larkin et al., 2007). After alignment, each species was compared pairwise to *D. rerio* and SAVs were recorded, with indels and ambiguous positions removed to avoid gaps in sequence comparisons. Next, the positions of all SAVs were visualized using Circos v0.69 (Krzywinski et al., 2008). Following analysis relative to *D. rerio*, the same procedure was performed relative to *D. aesculapii*. Protein-coding sequences were also pooled according to respiratory complex and the percentage of each amino acid sequence containing SAVs relative to *D. rerio* was quantified.

2.6.5 SIFT analysis

Each amino acid variant from the SAV comparisons was also converted into a format readable by SIFT v6.2.1 (Sorting Intolerant From Tolerant; Ng & Henikoff, 2003). Then, we used each of our *D. rerio* translated mitochondrial-encoded protein sequences along with the variant files as input to perform a SIFT analysis using the UniRef90 protein database (The UniProt Consortium, 2025). Each residue that was recorded as "intolerant" in the SIFT results was output to a Circos input file and added to the Circos SAV plot described above.

2.6.6 SAV x distance analysis between all pairwise species combinations

Using the maximum likelihood tree, we used the ape v5.7-1 package in R v4.3.0 to create a collapsed tree with only one terminal branch per species (Paradis & Schliep, 2019; R Core Team, 2023). Next, we used ape to extract pairwise distances between all Danionin species into a distance matrix. Then, we used the MSAs of consensus sequences for each translated amino acid sequence described above to record the SAVs in all pairwise species comparisons. Finally, we graphed the ratio of SAVs to the amino acid length over the genetic distance for each pairwise comparison in a dot plot using ggplot2 v3.4.3 in R (Wickham, 2016). This method of plotting SAV ratios to genetic distance is similar to p-distance x phylogenetic distance methods used in other studies (Nabholz et al., 2013; Nikelski & Weir, 2025).

To determine an expected range of SAV accumulation, we graphed regression lines of all SAV ratios by genetic distance comparisons in 60 bp windows along all positions for each protein sequence, with the sliding window correcting for varying protein-coding domains and topology along each protein (Simakova & Simakov, 2014; Raimondi et al., 2019). For the smaller proteins Nd3 and Nd4l (116 and 98 aa, respectively, compared to other mitochondrial proteins in the range of 172 - 606 aa), a large fraction of pairwise comparison data points fell outside regression lines using a 60 aa window, with the widespread range likely due to too large of a sampling window relative to overall protein length. For these proteins, a similar analysis was carried out using a composite including Nd3 and Nd4l and a 60 bp window. After this adjustment, pairwise comparison data now fell within the regression line range as expected. For a similar reason, the small Atp8 protein (54 aa) was pooled in the analysis with the larger Atp6 protein (227 aa) and data for both proteins presented together as a composite.

The resulting areas displayed by these regression lines then denoted domains of expected SAV levels relative to genetic distance and allowed highlighting pairwise comparisons landing above or below the expected range.

2.6.7 Mutation rate analysis and distribution from BEAST-generated gene trees

We used a Bayesian approach to phylogenetic inference in BEAST to create a time-calibrated tree for Danionin fish. First, we constructed a tree representing whole mtDNA and an MSA of consensus sequences and features on the H-strand, using amino acid consensus sequences after editing to remove overlapping ORFs, indels, and ambiguous positions as described in the

maximum-likelihood estimate analysis, then aligned with Clustalw2 (Larkin et al., 2007). We input the consensus amino acid sequence MSA for the 26 H-strand coding features into pal2nal and generated a codon-aligned consensus nucleotide MSA (Suyama et al., 2006). Next, we generated a partition scheme by further condensing partitions from previous maximum-likelihood tree estimates using PartitionFinder in the IQ-TREE v2.1.2 software package (Lanfear et al., 2017; Minh et al., 2020). Then, BEAST v2.7.7 was used for phylogenetic inference with evolutionary models formulated using the software package bModelTest v1.3.2 (Bouckaert et al., 2019; Bouckaert & Drummond, 2017). The analysis was conducted using a MCMC chain length of 250 million samples ran on BEAST 2 using a Yule model time calibrated to 13 MY for the most recent common ancestor (MRCA) of the *Danio* genus as previously described (Rüber et al., 2007; McCluskey & Postlethwait, 2014). Following estimation in BEAST we checked for coalescence of priors and posteriors using Tracer v1.7.1 (Rambaut et al., 2018), then used TreeAnnotator v2.7.6 in the BEAST software package to concatenate tree data into a single consensus tree (Bouckaert et al., 2019).

Following construction of the mtDNA time-calibrated tree we used a similar Bayesian approach to construct individual trees for all 13 protein-coding mitochondrial genes (Nd1, Nd2, Nd3, Nd4l, Nd4, Nd5, Nd6, Cytb, Cox1, Cox2, Cox3, Atp6, and Atp8). MSAs were created by using ClustalW2 on the nucleotide and amino acid sequences for each gene, with pal2nal used to generate codon alignments (Suyama et al., 2006; Larkin et al., 2007). Subsequently, partitions were created specifying the three codon positions and PartitionFinder used to condense the partitioning scheme (Lanfear et al., 2017). Evolutionary models were estimated using the bModelTest package in BEAST, and the topology of each tree was constrained to that of the full

mtDNA tree by defining each tree node with its respective time-calibrated length as MRCA priors under a monophyletic uniform model (Bouckaert & Drummond, 2017; Bouckaert et al., 2019). These topology and time-calibration constraints allowed for BEAST to estimate rates of evolution along each branch in each gene tree without changing the assumptions of relationships indicated by the previous mtDNA analysis. Each tree was then estimated by running BEAST using a calibrated Yule model with a MCMC chain length of 250 million samples followed by Tracer analysis and TreeAnnotator concatenation (Rambaut et al., 2018; Bouckaert et al., 2019). Next, we extracted the mutation rates from each branch of each tree excluding branches from the outgroup clade containing *S. axelrodi*, *B. maculatus*, and *T. espei*. These mutation rates, measured in substitutions per site per million years, were then plotted by branch per gene in bar graphs using ggplot2 in R (Wickham, 2016; R Core Team, 2023). Following initial results that included the laboratory AB strain of *D. rerio*, we repeated the analysis replacing the AB strain with a recently developed CB *D. rerio* strain, considered representative of the wild population (Suurväli et al., 2020) (BankIt submission identifier 2974502).

2.6.8 Identifying outlier branch lengths on gene trees

To compare the rates of nucleotide substitutions along each branch of a maximum likelihood tree, we used the consensus sequence MSAs per gene (Nd1, Nd2, Nd3, Nd4l, Nd4, Nd5, Nd6, Cytb, Cox1, Cox2, Cox3, Atp6, Atp8) or concatenated OXPHOS complex (complex I, complex IV, complex V). For complex I, we left out the L-strand gene Nd6 for consistency between the H-strand-based method used to generate tree topology. We partitioned each MSA according to codon position and used ModelTest-NG to determine evolutionary models for each partition (Darriba et al., 2020). Then, we used the AICC partition model from ModelTest-NG and used

RAxML-NG v1.0.2 to generate each tree using maximum likelihood estimates with the topology constrained to our initial mtDNA maximum likelihood tree (Kozlov et al., 2019). To generate a gene-specific adjusted backbone tree that lacks an intrinsic bias for the gene of interest, we used the full mtDNA tree to constrain tree topology and created a codon-aligned MSA involving a concatenation of consensus sequences of all protein-coding genes excluding the gene of interest.

Upon finishing gene-specific adjusted backbone tree construction, we extracted the branch lengths from each gene tree and each adjusted backbone tree, then divided the gene tree branch lengths by the adjusted backbone tree branch lengths. Using these ratios, we generated a histogram for each gene using R setting “breaks” to 10 and drawing an upper limit cutoff line equal to a 3 MAD distance above the median value for each plot (R Core Team, 2023). We considered branch length ratios that were larger than the upper limit as outliers with significantly longer branch length than average for all other trees. To visualize these results, we used ggtree v3.8.2 in R to create each gene or concatenated complex tree, using branch lengths from the maximum likelihood estimates and highlighting branches identified by the branch length analysis as outliers (Yu et al., 2017).

2.6.9 Measuring selection relaxation and episodic diversification in HyPhy

To measure selection metrics within protein-coding mitochondrial genes, including relaxation of purifying selection and episodic diversification, we used, respectively, the RELAX and BUSTED packages included in the HyPhy v2.5.46 software suite (Pond, et al., 2005; Murrell, 2015; Wertheim et al., 2015). As input, we used the mtDNA tree estimated by maximum-likelihood and

the codon-aligned consensus sequence MSAs per gene (Nd1, Nd2, Nd3, Nd4l, Nd4, Nd5, Nd6, Cytb, Cox1, Cox2, Cox3, Atp6, Atp8) or concatenated OXPHOS complex (complex I, complex IV, complex V). Additionally, we combined the concatenated complex data to generate an MSA concatenated genome containing all mitochondrial protein-coding genes except the L-strand gene Nd6. Next, we ran each analysis, RELAX and BUSTED, individually on each branch. In RELAX, to test internal branches, each branch was set as the foreground, and all other internal branches of the tree were set as background. If a tested branch was terminal, we tested it against only other terminal branches on the tree to exclude polymorphisms in internal branches to avoid genetic variation that may not have had time for removal by selection already acting on those branches (Bush et al., 1999; Xie et al., 2021).

Following RELAX and BUSTED analyses, we visualized the results on phylogenetic trees using the ggtree package in R (Yu et al., 2017; R Core Team, 2023). For each gene, we used the branch lengths from the gene trees created earlier in our maximum likelihood gene tree analysis. For the concatenated complex and concatenated genome RELAX or BUSTED visualizations, we used RAxML-NG to generate trees to determine branch lengths (Kozlov et al., 2019). Values from RELAX for each branch were extracted and represented onto the tree according to a color-coded scale spanning selection intensification and relaxation. Data from BUSTED for each branch was classified as “not significant” (“n.s.”) or “Episodic Diversification” and represented onto the tree according to classification.

2.7 Acknowledgements

This work received funding from the National Institute of Food and Agriculture, United States Department of Agriculture, Hatch to the UW Madison College of Agricultural and Life Sciences (1013782), a UW2020 Round 6 award from the Office of the Provost, U. Wisconsin - Madison (135-AAH8462), a Fall Competition award from the U. Wisconsin - Madison Graduate School, and additional support from the College of Agricultural and Life Sciences and the School of Medicine and Public Health at U. Wisconsin - Madison. Additional funding was received through the NIH Institutional National Research Service Award (T32GM007133).

We thank Cécile Ané from the Statistical Consulting group at the University of Wisconsin – Madison for assistance in the regression analysis, Bayesian estimates, and selection analyses. The authors utilized the University of Wisconsin – Madison Biotechnology Center's DNA Sequencing Facility (Research Resource Identifier – RRID:SCR_017759) to provide library prep and NovaSeq DNA sequencing services.

2.8 Data availability

Complete mtDNA sequences for *D.* rerio, *D. aesculapii*, *D. kyathit*, *D. nigrofasciatus*, *D. tinwini*, *D. albolineatus*, *D. margaritatus*, *D. choprae*, *Dev. aequipinnatus*, *D. sondhii*, and *M. kubotai* are available in the GenBank Nucleotide Database at https://www.ncbi.nlm.nih.gov/nucleotide/ and can be accessed with accession numbers PV357751-PV357794 and partial mtDNA sequences for *D. erythromicron* are available at accessions PV367247-PV367250. Sequence for *D. rerio* CB strain is under review at GenBank with BankIt Submission Identifier number 2974502. Other data underlying this article will be shared on reasonable request to the corresponding author.

2.9 Figures

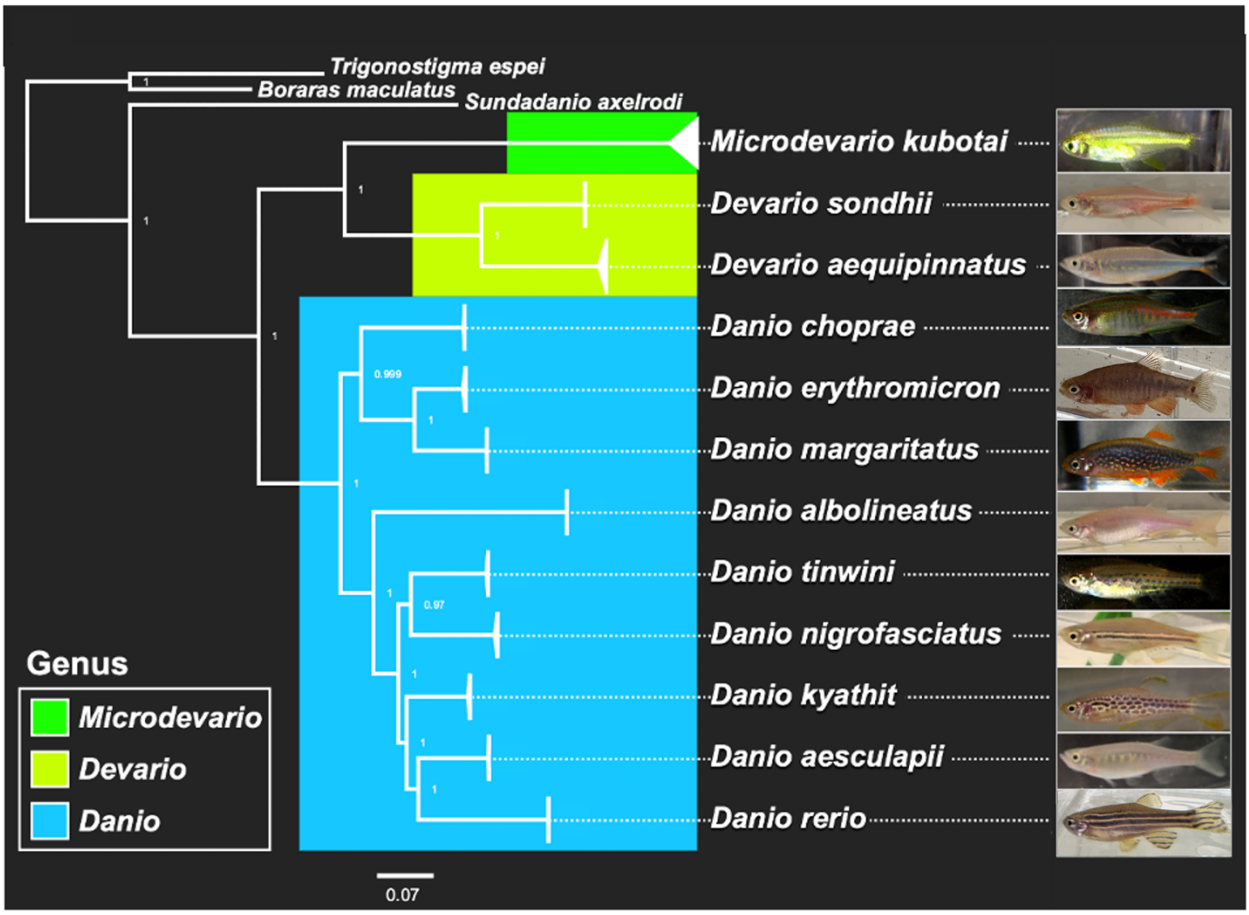


Figure 1. Phylogenetic analysis of the Danionin fish based on mitochondrial DNA.

Phylogenetic tree of fifteen Danionin fish species with representatives from six genera composed using maximum-likelihood estimates of mtDNA. Mitochondrial genomes from the twelve depicted species, representing *Danio*, *Devario*, and *Microdevario* genera, were sequenced, assembled, and annotated. The other three species were found in sequence databases and are used as outgroups. Representative fish images are not to scale.

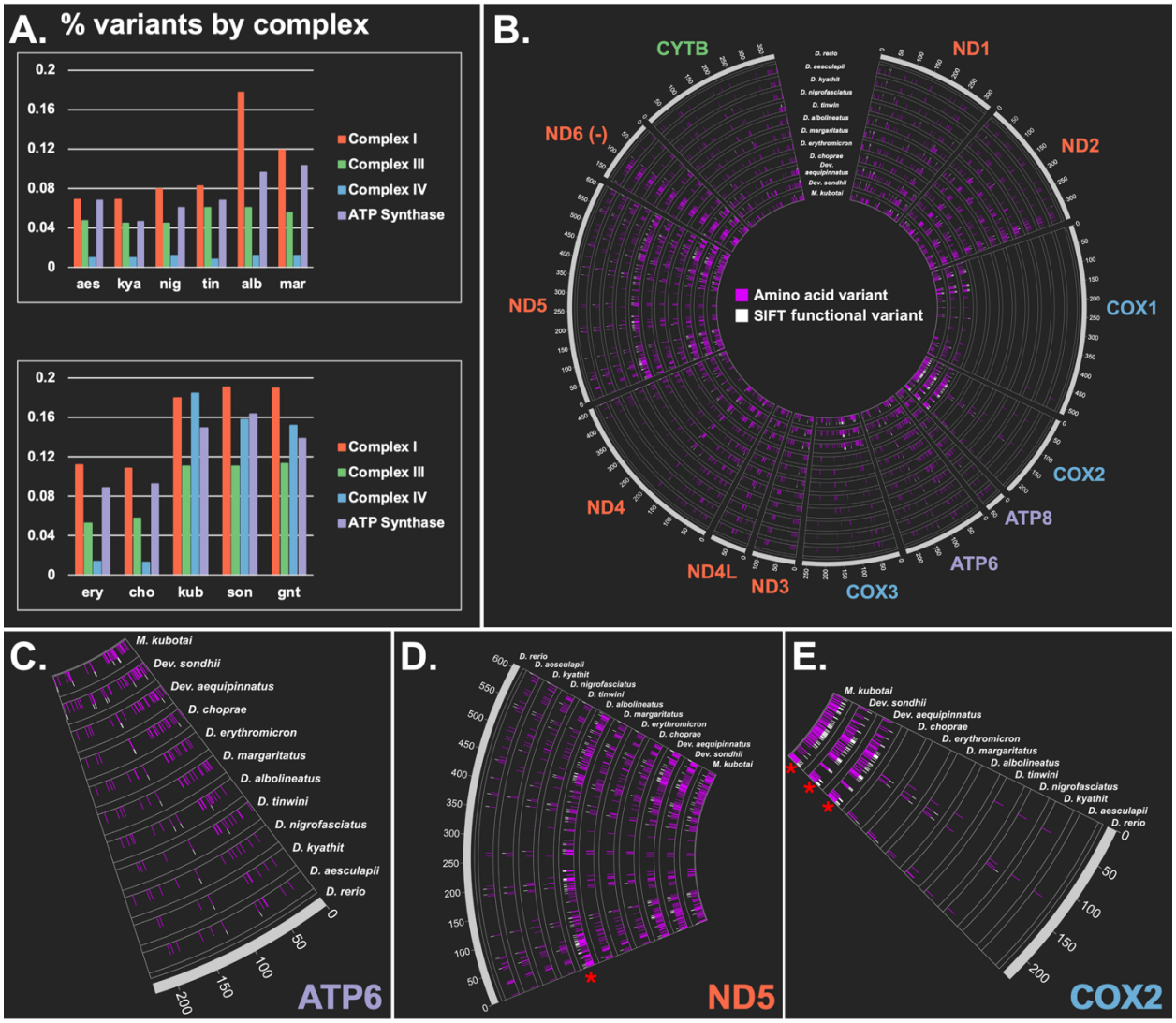


Figure 2. A complete analysis of SAVs in 12 Danionin species.

A. Percentages of SAVs to amino acid lengths of each OXPHOS complex compared to *D. rerio*. *B.* A plot representing each species in our study compared to *D. rerio* in the outermost circle, arranged from most to least related to *D. rerio* going inward. Each block of space represents one of 13 mitochondrial-H-strand-encoded proteins. Longer magenta highlight lines represent SAVs, and shorter white marks represent SIFT functional variants, both in comparison to *D. rerio*. *C.* Atp6 protein plot exemplifies an observed pattern of gradual increase in SAVs going from most related to *D. rerio* in outer circles to least related in inner circles. *D.* Nd5 exhibits a similar

gradual increase in SAVs plus the appearance of large number of SAVs in (*D*). *albolineatus* (asterisk), which stands out as above the overall gradual trend of increase. *E.* A similar above-the-trend increase in SAVs can be detected in Cox2 protein in *Devario* and *Microdevario* (asterisks), also apparent for Cox1 and Cox3 in (*B*).

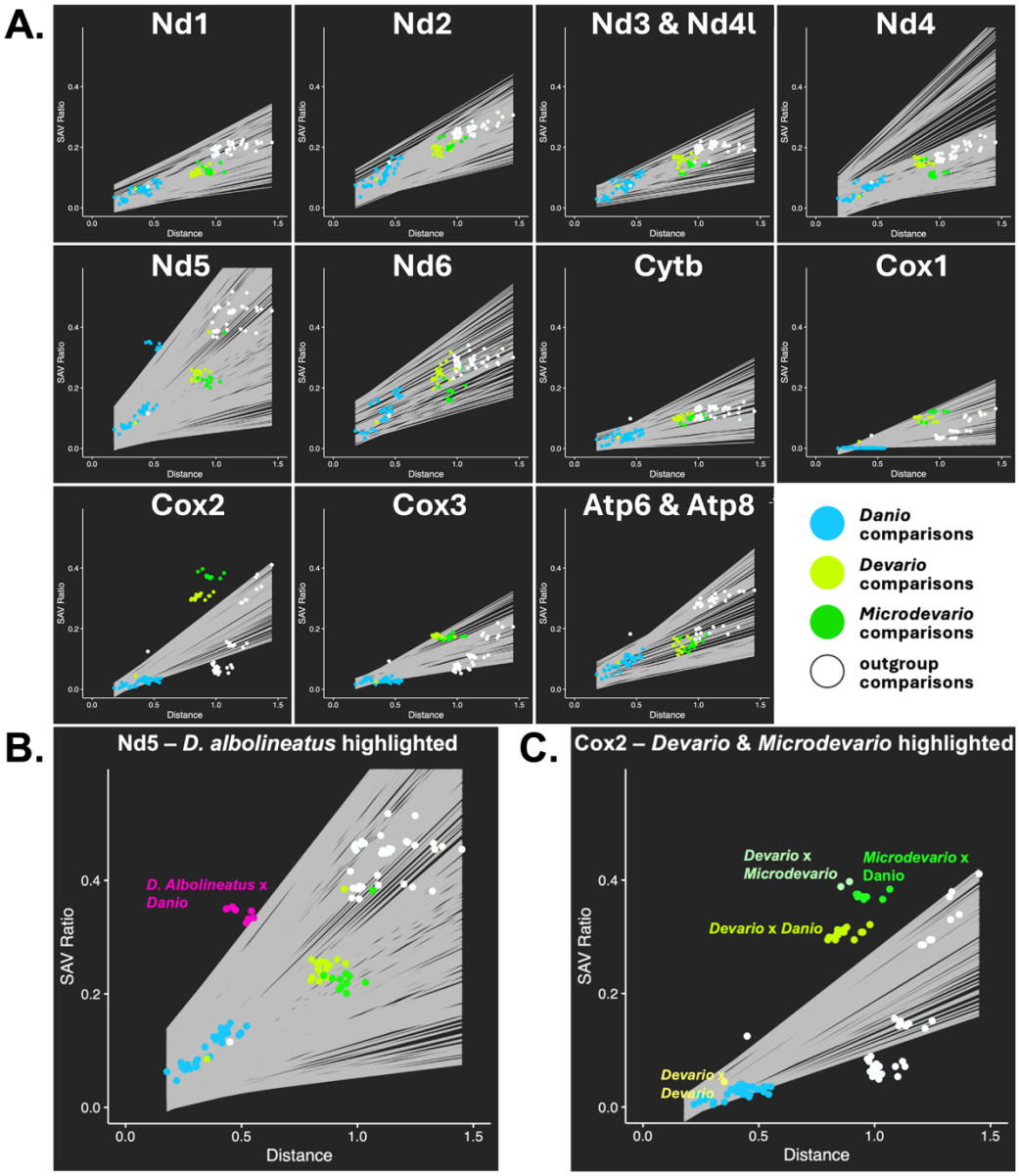


Figure 3. Pairwise comparisons of single amino acid variants (SAVs) to phylogenetic distance.

A. Scatter plots illustrating the relationship between the number of SAVs and the phylogenetic distance for mtDNA genes in pairwise comparisons of Danionin species. Each data point is color-coded by whether the pair comparison involves *Danio* species only, or whether it includes at least one species in the *Devario* or *Microdevario* genera, or an outgroup species. Gray lines

represent regression lines of SAV vs. distance plots of a subset of amino acids generated by a 60 aa sliding window along the length of the amino acid sequence. Points above or below all gray regression lines are considered outliers that vary significantly from the observed trends. *B*. Enlargement of the SAV vs. phylogenetic distance plot for Nd5 highlighting points representing comparisons of *D. albolineatus* to other *Danio* species, showing all combinations involving *D. albolineatus* and all other *Danios* fall outside of the regression lines. *C*. Enlargement of the SAV vs. phylogenetic distance plot for Cox2 highlighting points involving at least one species within *Devario* or *Microdevario* in the pair, showing all pairwise comparisons between *Danios* and *Devarios*, *Danios* and *Microdevarios,* and *Devarios* and *Microdevarios* fall outside of the regression lines.

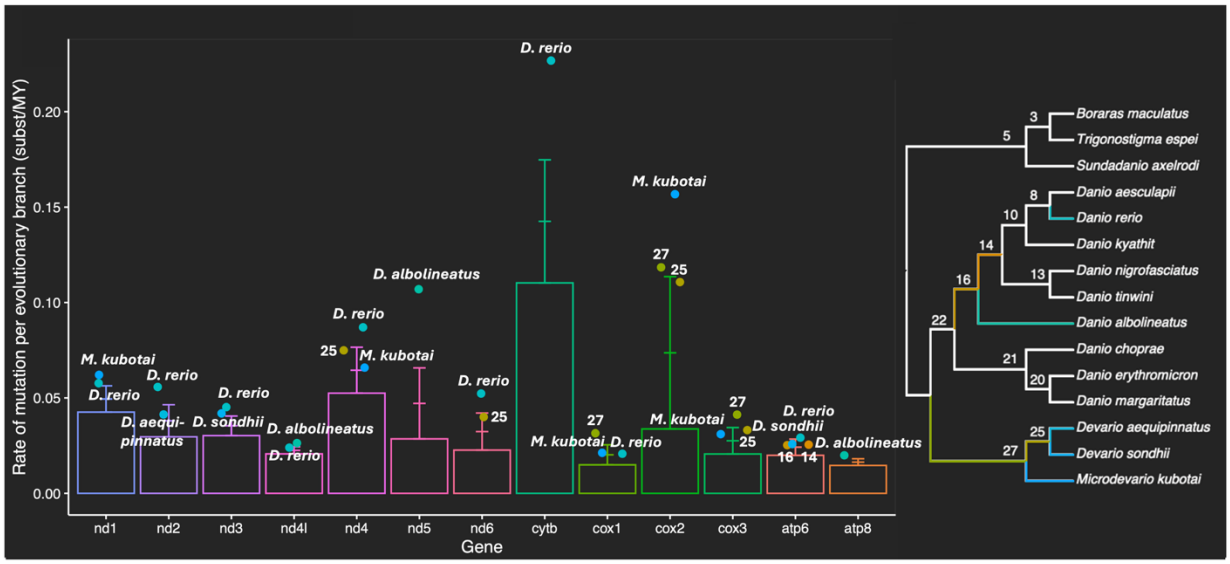


Figure 4. Comparison of mutation rates per branch across the Danionin phylogenetic tree.

A plot comparing the mutation rates per million years of each branch of each gene tree as estimated using Bayesian rate estimates. Error bars show 1 sd and 2 sds from the average for each gene tree. Points are shown that fall at least 1 sd from each mean are highlighted. A phylogenetic tree is included with each branch labeled for reference. Points on the mutation rate plot that were at least 1 sd from the mean in are represented with the same color code as corresponding lineage branches in the phylogenetic tree. Outgroups (*S. axelrodi*, *T. espei*, *B. maculatus*) were used to construct the tree but were excluded from the mutation rate distribution plot analysis.

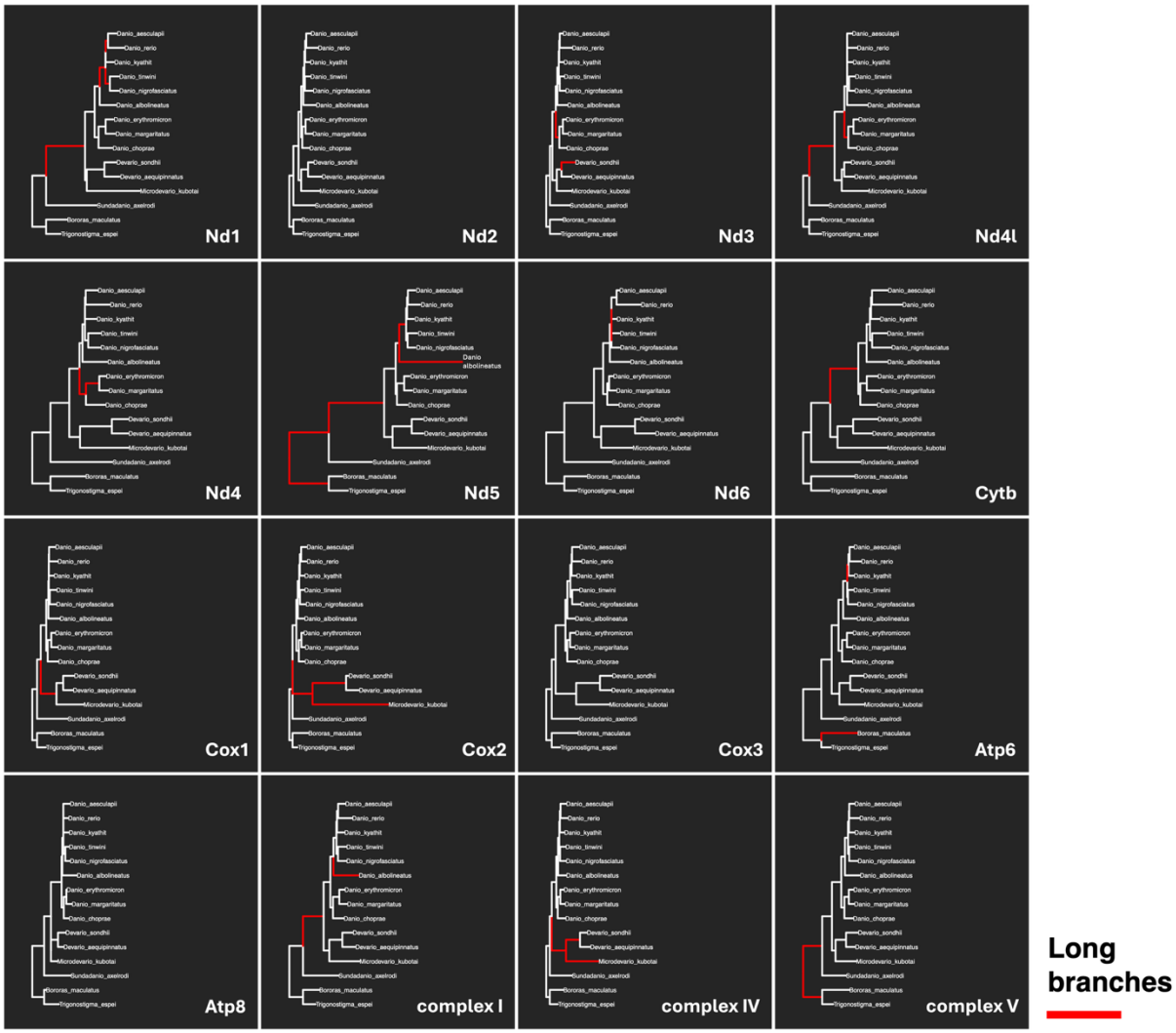


Figure 5. Long gene tree branch lengths of mtDNA genes in Danionins.

Visualization of phylogenetic branches with lengths greater than 3 MAD from the mean, per gene or complex, compared to corresponding constrained backbone trees adjusted to exclude the same genes or complex.

Figure 6. Tree branches for Danionin mtDNA genes undergoing episodic diversifying positive selection.

Gene or protein complex trees indicating episodic diversifying selection in each branch, identified through a comparison of each branch to all other branches. Highlighted branches show signals of episodic diversifying (positive) selection.

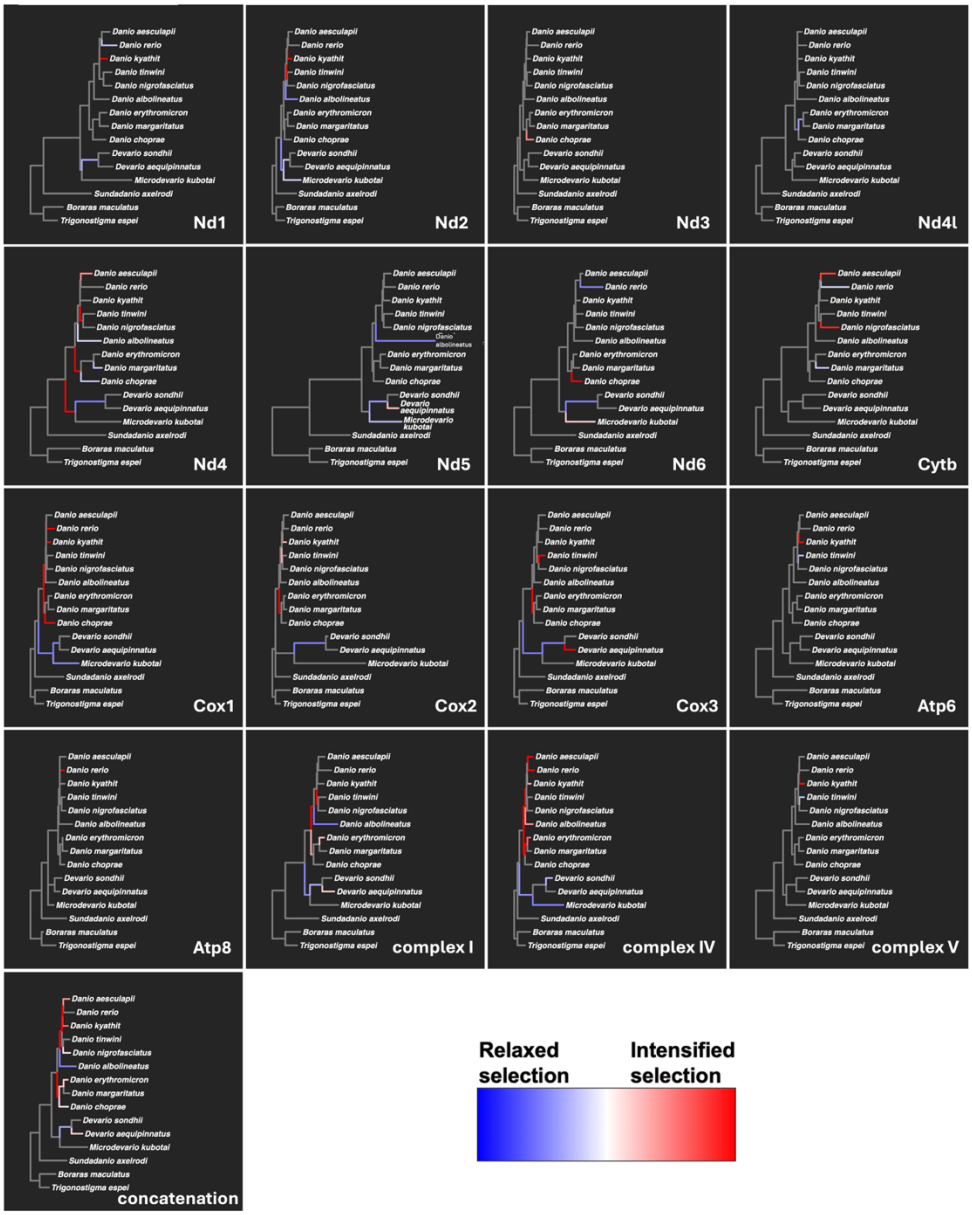
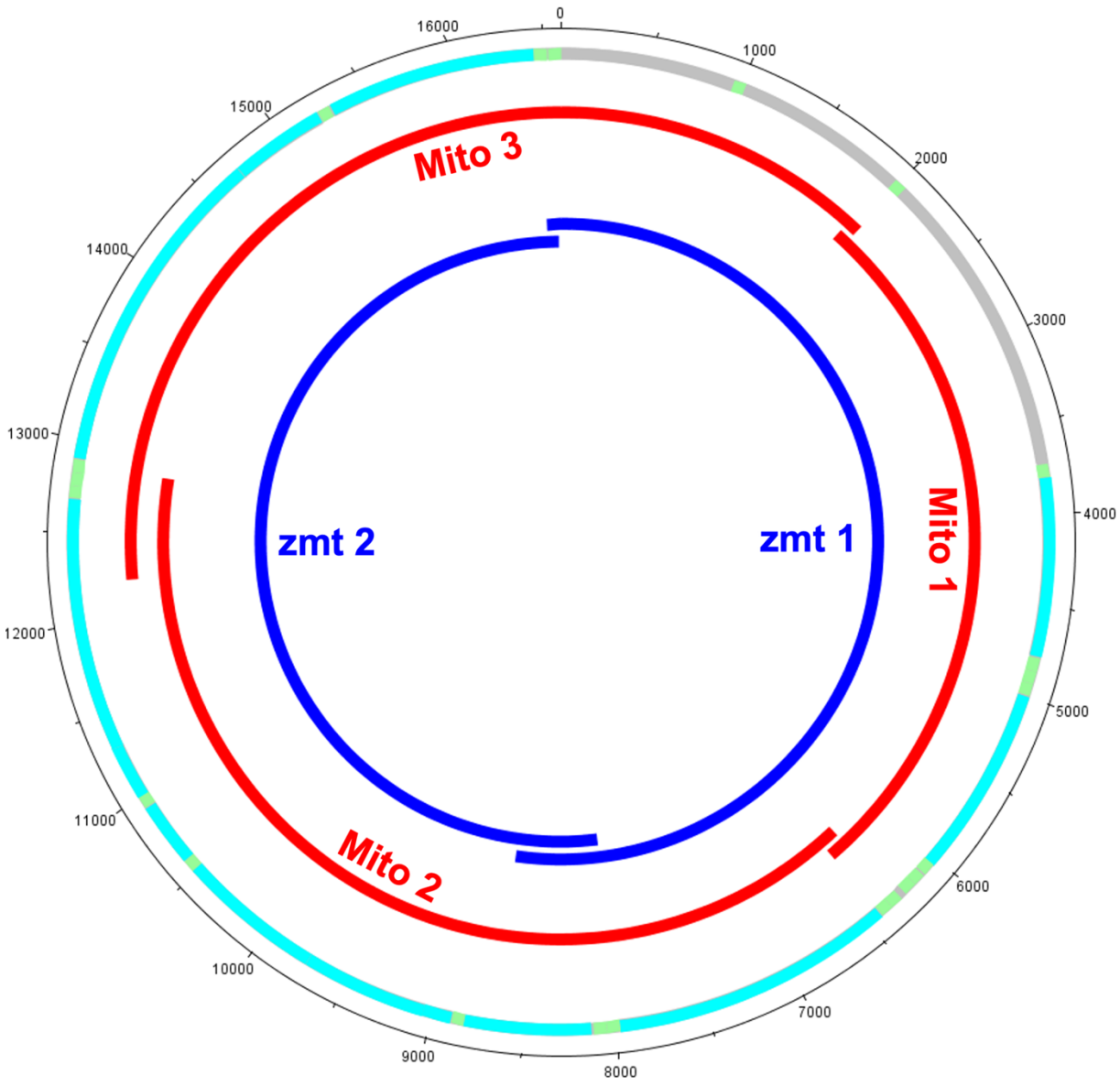


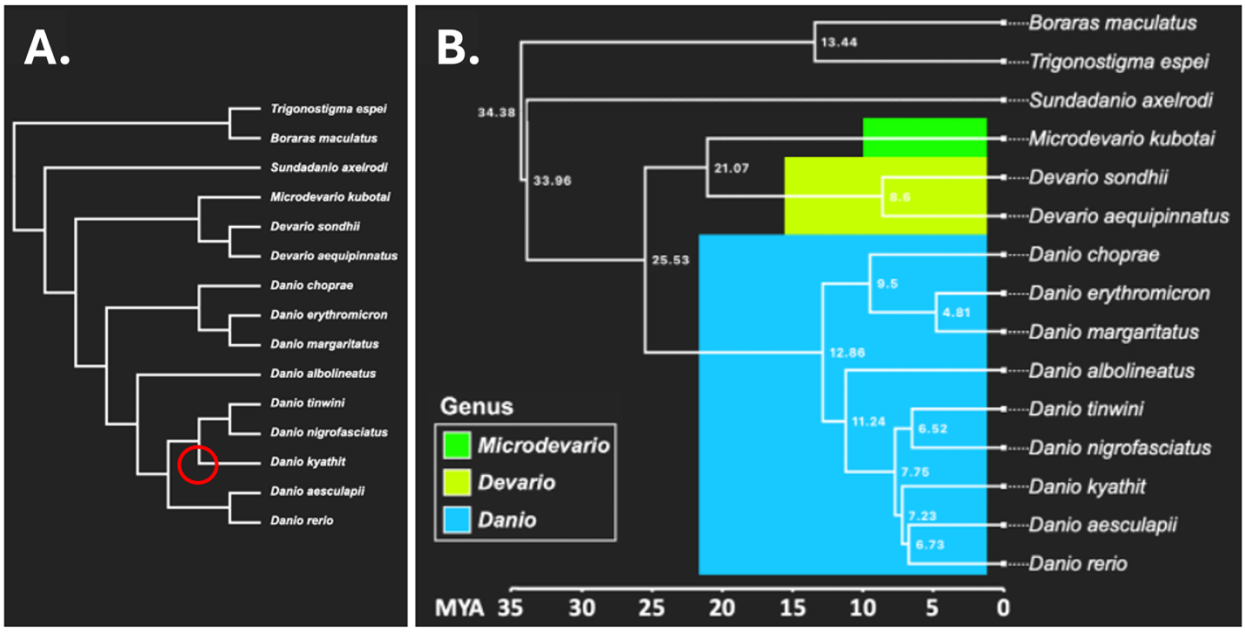
Figure 7. Tree branches for Danionin mtDNA genes with signals for relaxation or intensification of selection. Gene trees representing patterns of selection relaxation or intensification in Danionin mitochondrial proteins or protein complexes. Branches are color-coded according to relaxation (blue) or intensification (red) scores, measured branch by branch.

2.10 Supplementary materials



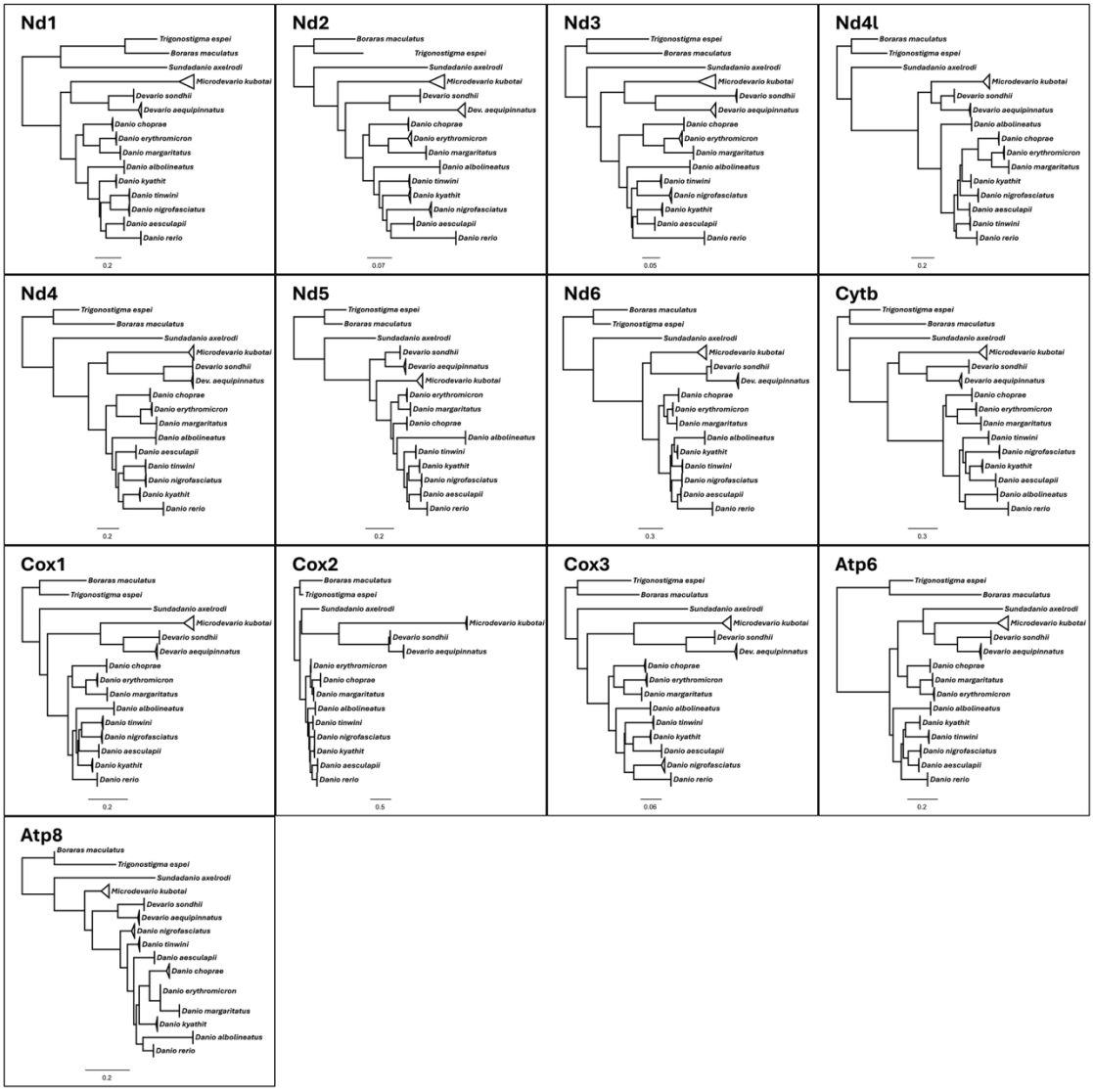
Supplemental Figure 1. Diagram of overlapping amplicons used in mtDNA sequencing.

Schematic representation of *D. rerio* mtDNA with nucleotide positions from NCBI Reference Sequence NC_002333.2. Inner lines represent two overlapping groups of amplicons used in sequencing including Mito 1, Mito 2, and Mito 3 as well as zmt 1 and zmt 2.



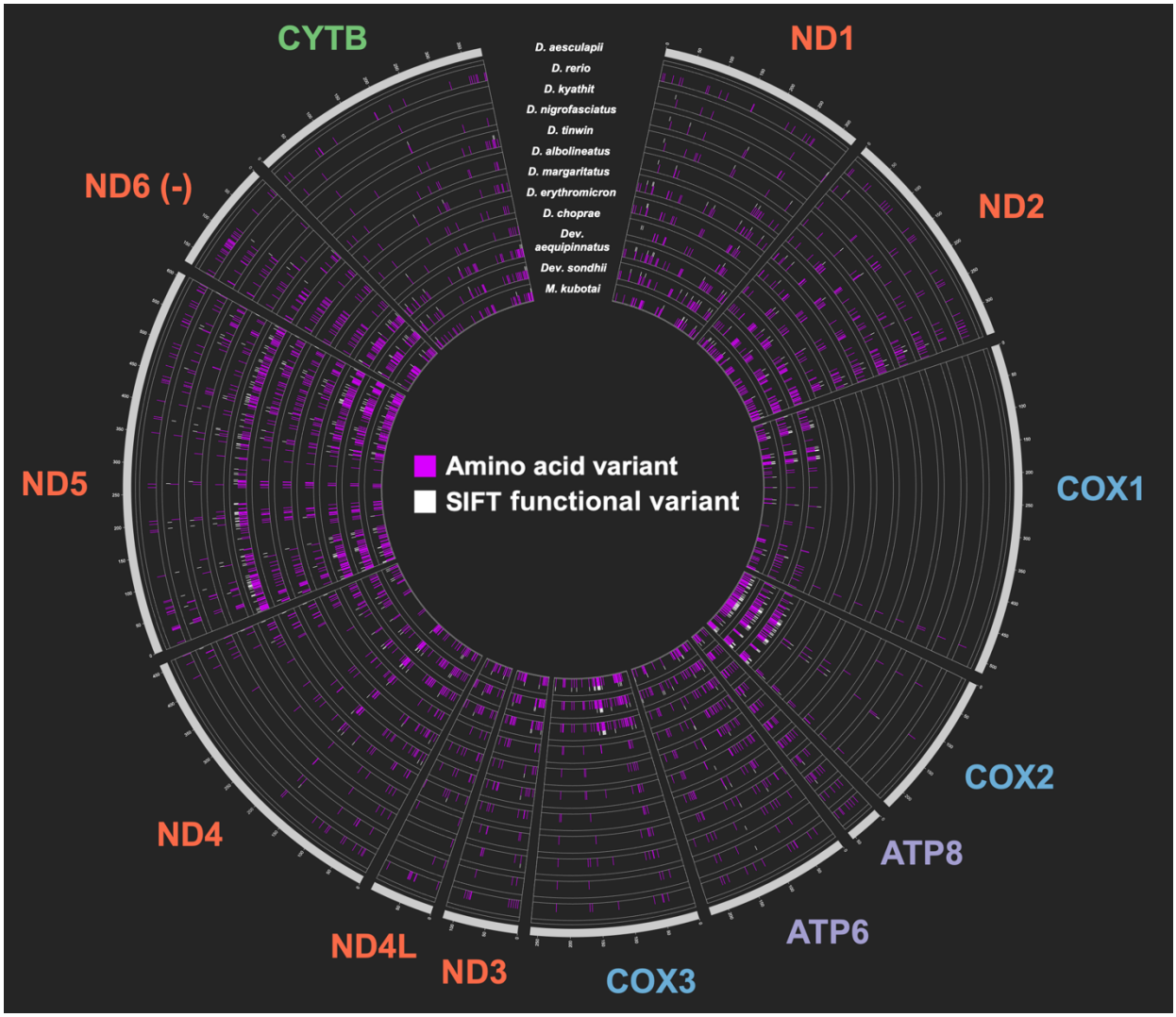
Supplemental Figure 2. Tree topology comparison to previous studies and time-calibrated tree.

A. Topology from McCluskey & Postlethwait (2014) with a circle indicating where the mtDNA-based topology in our study (Figure 1) differs from their previous tree estimates using ntDNA. In our study, *D. kyathit* shared a clade with *D. aesculapii* and *D. rerio* instead of *D. tinwini* and *D. nigrofasciatus*. *B.* Time-calibrated Bayesian-estimated phylogenetic tree for Danionins with highlights for *Danio*, *Devario*, and *Microdevario* genera. Numbers at tree nodes represent average tree height in millions of years.



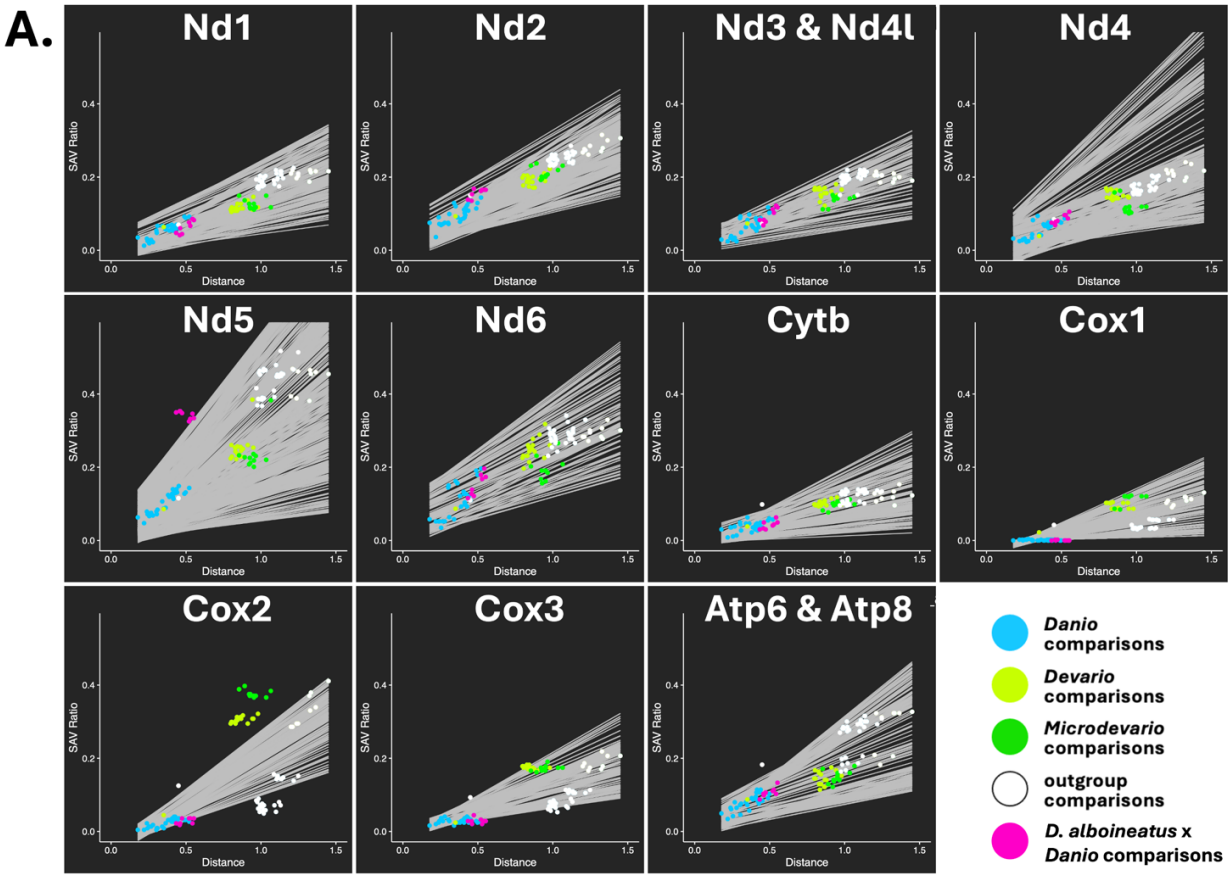
Supplemental Figure 3. Danionin mitochondrial gene tree topologies.

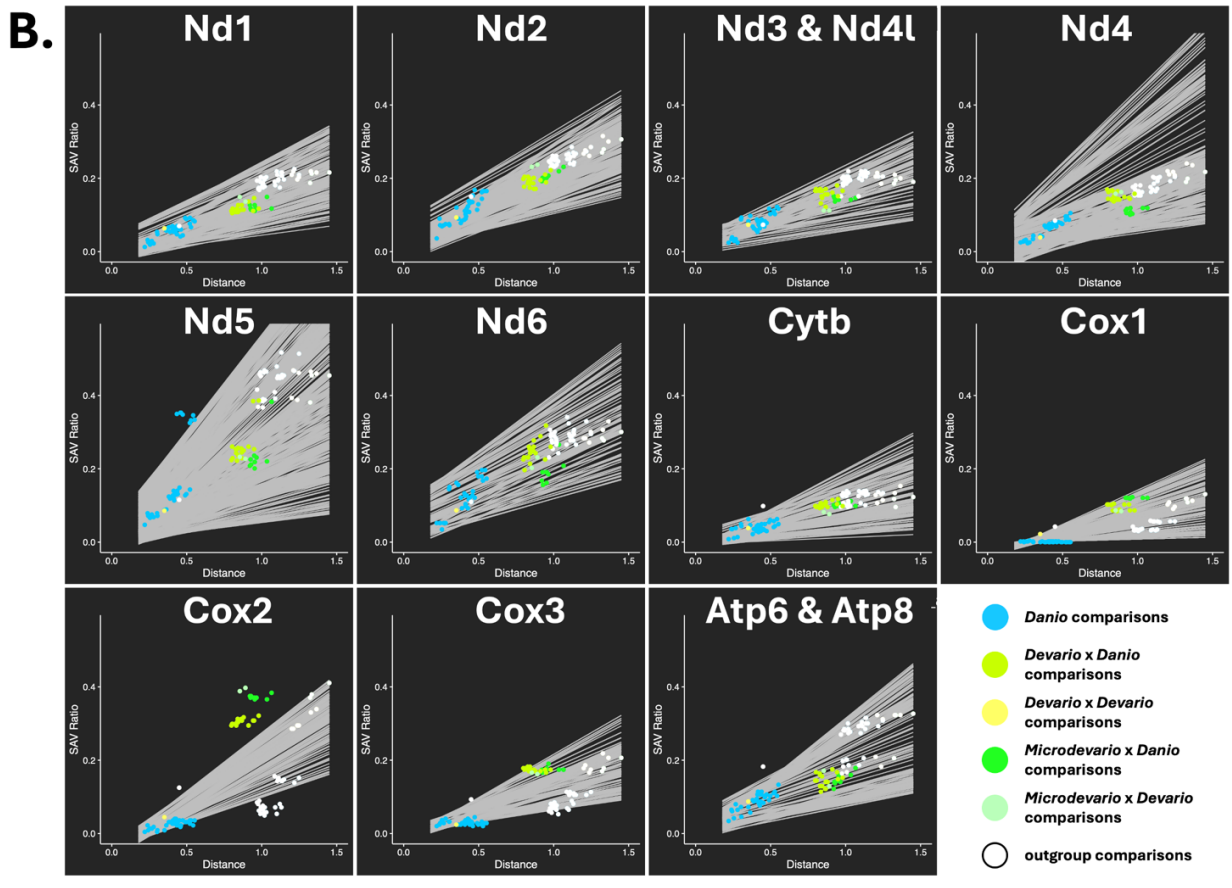
Phylogenetic trees for all 13 mitochondrial-encoded protein-coding genes, generated with maximum-likelihood methods, demonstrating distinct topologies across mitochondrial genes.



Supplemental Figure 4. SAV analysis of 12 Danionin species relative to *D. aesculapii.*

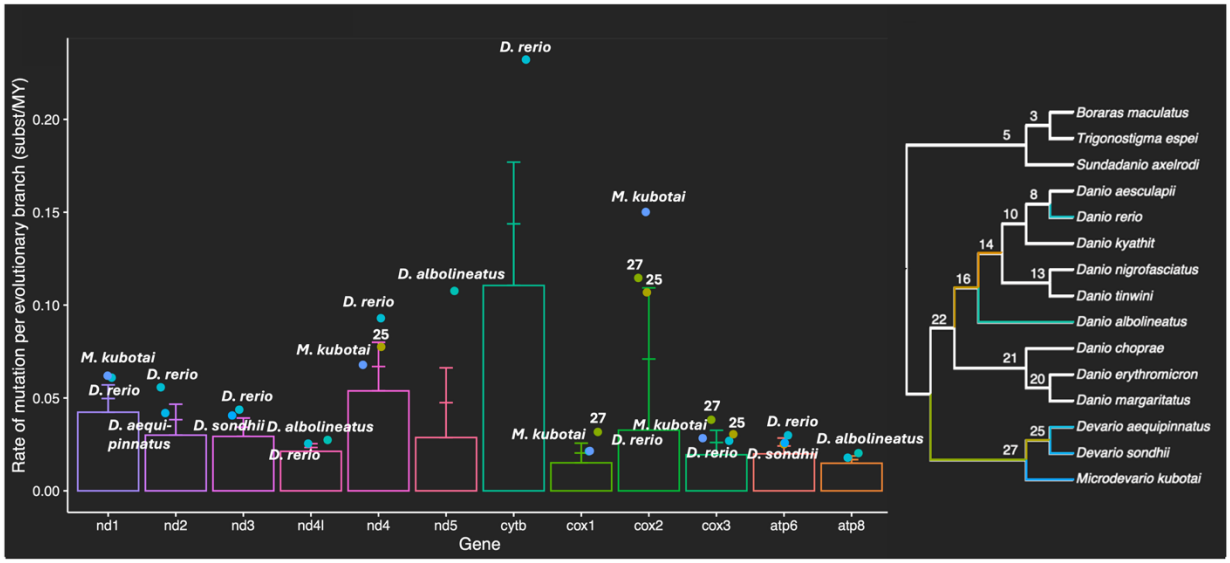
A plot representing each species in our study compared to *D. aesculapii* in the outermost circle, arranged from most to least related to *D. aesculapii* going inward. Each block of space represents one of 13 mitochondrial-H-strand-encoded proteins. Longer magenta highlight lines represent SAVs, and shorter white marks represent SIFT functional variants, both in comparison to *D. aesculapii*. The plot exhibits similar patterns as those observed when species are compared to *D. rerio.*


A.
Nd1
Nd2
Nd3 & Nd4l
Nd4
Nd5
Nd6
Cytb
Cox1
Cox2
Cox3
Atp6 & Atp8
SAV Ratio
Distance
Danio comparisons
Devario comparisons
Microdevario comparisons
outgroup comparisons
D. alboineatus x Danio comparisons


B.
Nd1
Nd2
Nd3 & Nd4l
Nd4
Nd5
Nd6
Cytb
Cox1
Cox2
Cox3
Atp6 & Atp8
SAV Ratio
Distance
Danio comparisons
Devario x Danio comparisons
Devario x Devario comparisons
Microdevario x Danio comparisons
Microdevario x Devario comparisons
outgroup comparisons

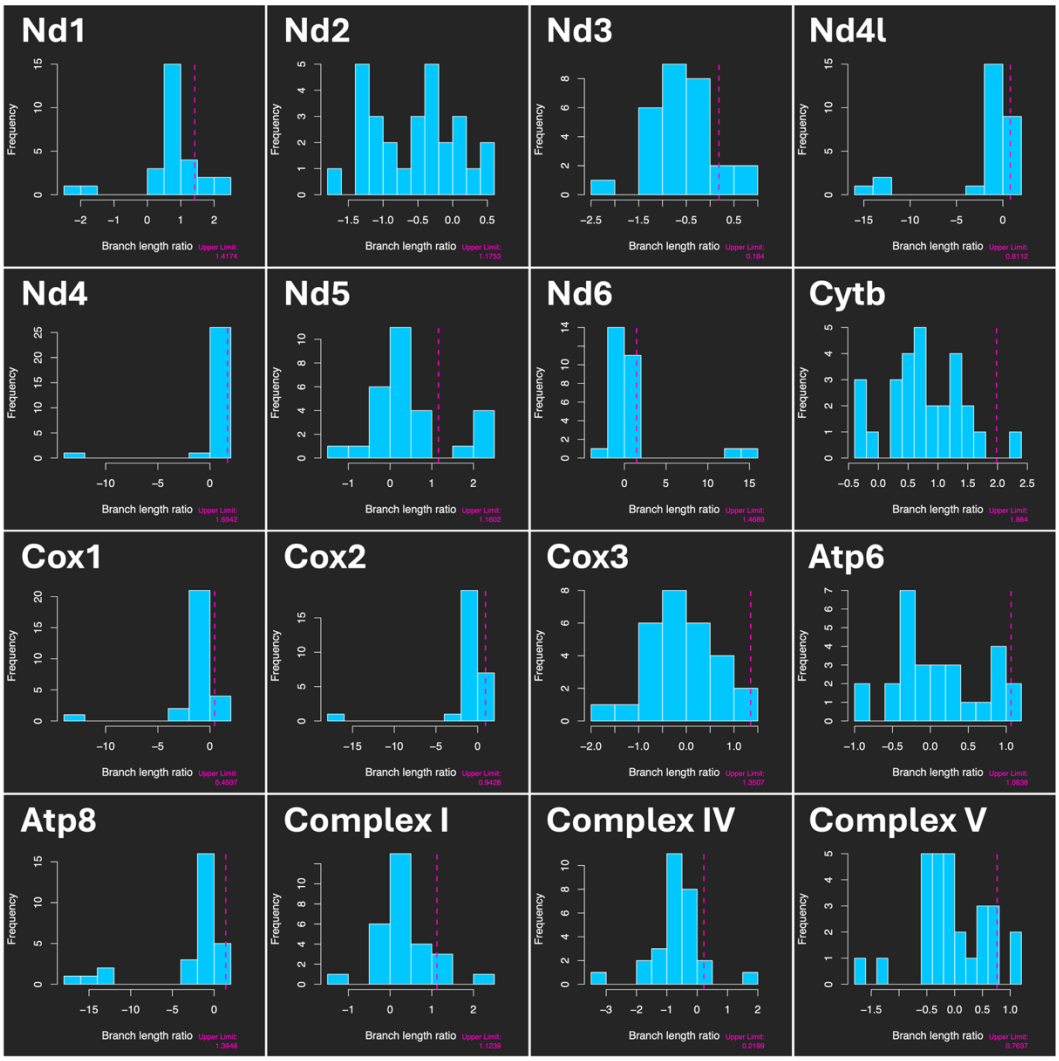
Supplemental Figure 5. Pairwise comparisons of single amino acid variations (SAVs) to phylogenetic distance emphasizing pairwise groupings.

A. Scatter plots illustrating the relationship between the number of SAVs and the phylogenetic distance for each pair of Danionin species. Each data point is color-coded by whether the species pairing is *Danio* species only, or whether the combination in the pair includes at least one species in the *Devario* or *Microdevario* genera, or an outgroup species. An additional color group is included representing all pairwise combinations involving *D. albolineatus* and another *Danio* species pairing. Gray lines represent regression lines of SAV vs. distance plots of a subset amino acids generated by a 60 aa sliding window along the length of the amino acid sequence. Points above or below all gray regression lines are considered outliers that vary significantly from the observed trends. *B.* The same scatter plots as (A) but where points that involve at least one species within *Devario* or *Microdevario* in the pair are highlighted, showing all pairwise comparisons between *Danios* and *Devarios*, *Danios* and *Microdevarios,* and *Devarios* and *Microdevarios* fall outside of the regression lines.



Supplemental Figure 6. Comparison of mutation rates per branch across the Danionin phylogenetic tree including a wild strain of *D. rerio*.

A plot comparing the mutation rates per million years of each branch of each gene tree as estimated using Bayesian rate estimates. Error bars show 1 sd and 2 sds from the average for each gene tree. Points are shown that fall at least 1 sd from each mean are highlighted. The sequences from a wild-derived CB *D. rerio* strain was used in this analysis instead of the laboratory AB strain as in Figure 4. A phylogenetic tree is included with each branch labeled for reference. Points on the mutation rate plot that were at least 1 sd from the mean in are represented with the same color code as corresponding lineage branches in the phylogenetic tree. Outgroups (*S. axelrodi*, *T. espei*, *B. maculatus*) were used to construct the tree but were excluded from the mutation rate distribution plot analysis. Trends in protein variation are similar to those obtained using the laboratory AB strain.



Supplemental Figure 7. Gene and complex tree comparisons relative to phylogenetic branches.

Histograms depicting individual branch length log ratios relative to adjusted branch lengths. For individual proteins, branch log ratios are between a tree derived for each gene and a corresponding adjusted tree using data for all other genes. For complexes, branch log ratios are between a tree for the complex (Complex I, Complex IV, or Complex V) and an adjusted tree using data for all mitochondrial genes excluding proteins in those complexes. Upper limits, represented by 3 MAD from the average log ratio, provide a measure of variation.

Supplemental Table 1. Primers used in mtDNA amplicon sequencing.

All primers are shown according to species, with distinct primer sequences, amplicon lengths, genomic positions, and thermocycler settings provided for all amplicons. The table is uniquely colored for each amplicon (mito 1: blue, mito 2: orange, mito 3: yellow, zmt 1: green, zmt 2: slate) and shading differences or variations in color indicate primers unique to species.

Amplicon	Species	Primer pair	Sequence	Position	length	thermocycler		
Mito 1	Danio rerio	niM1R1M1F1			4455	98˚C	2 min	
	GenBank: PV357751.1	Mito 1 F1	CAAGGTAAGTGTACCGGAAGG	1934-1954		98˚C	10 sec	30x
		niM1R1	GCTCCGTGGACAGAGGTTTA	6388-6370		55˚C	15 sec	
						68˚C	5 min	
						68˚C	5 min	
						4˚C	∞	
Mito 1	Danio aesculapii	niM1R1M1F1			4465	98˚C	2 min	
	GenBank: PV357755.1	Mito 1 F1	CAAGGTAAGTGTACCGGAAGG	1929-1949		98˚C	10 sec	30x
		niM1R1	GCTCCGTGGACAGAGGTTTA	6393-6374		55˚C	15 sec	
						68˚C	5 min	
						68˚C	5 min	
						4˚C	∞	
Mito 1	Danio kyathit	niM1R1M1F1			4469	98˚C	2 min	
	GenBank: PV357759.1	Mito 1 F1	CAAGGTAAGTGTACCGGAAGG	1937-1957		98˚C	10 sec	30x
		niM1R1	GCTCCGTGGACAGAGGTTTA	6405-6386		55˚C	15 sec	
						68˚C	5 min	
						68˚C	5 min	
						4˚C	∞	
Mito 1	Danio nigrofasciatus	niM1R1M1F1			4467	98˚C	2 min	
	GenBank: PV357763.1	Mito 1 F1	CAAGGTAAGTGTACCGGAAGG	1921-1941		98˚C	10 sec	30x
		niM1R1	GCTCCGTGGACAGAGGTTTA	6387-6368		55˚C	15 sec	
						68˚C	5 min	
						68˚C	5 min	
						4˚C	∞	
Mito 1	Danio tinwini	niM1R1M1F1			4464	98˚C	2 min	
	GenBank: PV357768.1	Mito 1 F1	CAAGGTAAGTGTACCGGAAGG	1893-1913		98˚C	10 sec	30x
		niM1R1	GCTCCGTGGACAGAGGTTTA	6356-6337		55˚C	15 sec	
						68˚C	5 min	
						68˚C	5 min	
						4˚C	∞	
Mito 1	Danio albolineatus	niM1R1M1F1			4436	98˚C	2 min	
	Genbank: PV357771.1	Mito 1 F1	CAAGGTAAGTGTACCGGAAGG	1908-1928		98˚C	10 sec	30x
		niM1R1	GCTCCGTGGACAGAGGTTTA	6343-6324		55˚C	15 sec	
						68˚C	5 min	
						68˚C	5 min	
						4˚C	∞	
Mito 1	Danio margaritatus	niM1R1M1F1			4475	98˚C	2 min	
	GenBank: PV357775.1	Mito 1 F1	CAAGGTAAGTGTACCGGAAGG	1936-1956		98˚C	10 sec	30x
		niM1R1	GCTCCGTGGACAGAGGTTTA	6410-6391		55˚C	15 sec	
						68˚C	5 min	
						68˚C	5 min	
						4˚C	∞	
Mito 1	Danio erythromicron	niM1R1M1F1			4474	98˚C	2 min	
	Compared to RefSeq: AP011419.1	Mito 1 F1	CAAGGTAAGTGTACCGGAAGG	988-1008		98˚C	10 sec	30x
		niM1R1	GCTCCGTGGACAGAGGTTTA	5461-5442		55˚C	15 sec	
						68˚C	5 min	
						68˚C	5 min	
						4˚C	∞	
Mito 1	Danio choprae	choM1F2R1			4546	98˚C	2 min	
	GenBank: PV357779.1	choM1F2	CCCGGTCAAACTGCATAAAA	1833-1852		98˚C	10 sec	30x
		choM1R1	GTAGGGTGGCCGAGTGTTTA	6378-6359		58˚C	15 sec	
						68˚C	5 min	
						68˚C	5 min	
						4˚C	∞	
Mito 1	Devario aequipinnatus	gntM1F1R3			5708	98˚C	2 min	
	GenBank: PV357783.1	gntM1F1	CCGAGAAGACATCTGTGCAA	2316-2335		98˚C	10 sec	30x
		gntM1R3	GTGGCATTCCTGCTAGACCT	8023-8004		55˚C	15 sec	
						68˚C	6 min	
						68˚C	5 min	
						4˚C	∞	
Mito 1	Devario sondhii	sonM1F2R1			6032	98˚C	2 min	
	GenBank: PV357787.1	sonM1F2	CCGAAAAATAGGAGGAGCAT	14262-14281		98˚C	10 sec	30x
		sonM1R1	ACGAGTGATTGCGCTACCTT	3397-3378		55˚C	15 sec	
						68˚C	6 min	
						68˚C	5 min	
						4˚C	∞	
Mito 1	Microdevario kubotai	niM1R1M1F1			4526	98˚C	2 min	
	GenBank: PV357791.1	Mito 1 F1	CAAGGTAAGTGTACCGGAAGG	2070-2090		98˚C	10 sec	30x
		niM1R1	GCTCCGTGGACAGAGGTTTA	6595-6576		55˚C	15 sec	
						68˚C	5 min	
						68˚C	5 min	
						4˚C	∞	

Amplicon	Species	Primer pair	Sequence	Position	length	thermocycler		
Mito 2	Danio rerio	reM2F1M2R1			6563	98˚C	2 min	
	GenBank: PV357751.1	reM2F1	ACCTGCGTCTTCAGATTTGC	6303-6322		98˚C	10 sec	30x
		M2R1	TTGGTTCCTAAGACCAATGGA	12865-12845		56˚C	15 sec	
						68˚C	7 min	
						68˚C	5 min	
						4˚C	∞	
Mito 2	Danio aesculapii	reM2F1M2R1			6567	98˚C	2 min	
	GenBank: PV357755.1	reM2F1	ACCTGCGTCTTCAGATTTGC	6305-6324		98˚C	10 sec	30x
		M2R1	TTGGTTCCTAAGACCAATGGA	12871-12851		56˚C	15 sec	
						68˚C	7 min	
						68˚C	5 min	
						4˚C	∞	
Mito 2	Danio kyathit	kyM2F1SM2R1			6606	98˚C	2 min	
	GenBank: PV357759.1	kyM2F1S	TTAACCAGTAAGGCGGGAAA	6280-6299		98˚C	10 sec	30x
		M2R1	TTGGTTCCTAAGACCAATGGA	12885-12865		56˚C	15 sec	
						68˚C	7 min	
						68˚C	5 min	
						4˚C	∞	
Mito 2	Danio nigrofasciatus	reM2F1M2R1			6567	98˚C	2 min	
	GenBank: PV357763.1	reM2F1	ACCTGCGTCTTCAGATTTGC	6300-6319		98˚C	10 sec	30x
		M2R1	TTGGTTCCTAAGACCAATGGA	12866-12846		56˚C	15 sec	
						68˚C	7 min	
						68˚C	5 min	
						4˚C	∞	
Mito 2	Danio tinwini	reM2F1M2R1			6566	98˚C	2 min	
	GenBank: PV357768.1	reM2F1	ACCTGCGTCTTCAGATTTGC	6270-6289		98˚C	10 sec	30x
		M2R1	TTGGTTCCTAAGACCAATGGA	12835-12815		56˚C	15 sec	
						68˚C	7 min	
						68˚C	5 min	
						4˚C	∞	
Mito 2	Danio albolineatus	reM2F1M2R1			6562	98˚C	2 min	
	Genbank: PV357771.1	reM2F1	ACCTGCGTCTTCAGATTTGC	6257-6276		98˚C	10 sec	30x
		M2R1	TTGGTTCCTAAGACCAATGGA	12818-12798		56˚C	15 sec	
						68˚C	7 min	
						68˚C	5 min	
						4˚C	∞	
Mito 2	Danio margaritatus	maM2F2M2R1			6643	98˚C	2 min	
	GenBank: PV357775.1	maM2F2	AAACCAGCGAGCATTCATCT	6247-6266		98˚C	10 sec	30x
		M2R1	TTGGTTCCTAAGACCAATGGA	12889-12869		60˚C	15 sec	
						68˚C	7 min	
						68˚C	5 min	
						4˚C	∞	
Mito 2	Danio erythromicron	maM2F2M2R1			6639	98˚C	2 min	
	Compared to RefSeq: AP011419.1	maM2F2	AAACCAGCGAGCATTCATCT	5301-5320		98˚C	10 sec	30x
		M2R1	TTGGTTCCTAAGACCAATGGA	11939-11919		60˚C	15 sec	
						68˚C	7 min	
						68˚C	5 min	
						4˚C	∞	
Mito 2	Danio choprae	maM2F2M2R1			6637	98˚C	2 min	
	GenBank: PV357779.1	maM2F2	AAACCAGCGAGCATTCATCT	6184-6203		98˚C	10 sec	30x
		M2R1	TTGGTTCCTAAGACCAATGGA	12820-12800		58˚C	15 sec	
						68˚C	7 min	
						68˚C	5 min	
						4˚C	∞	
Mito 2	Devario aequipinnatus	gntM2F1R1			5620	98˚C	2 min	
	GenBank: PV357783.1	gntM2F1	AGGAGCTGTCTTCGCCATTA	7534-7553		98˚C	10 sec	30x
		gntM2R1	TGCGAATTCTAAAATTGACCAAG	13153-13131		55˚C	15 sec	
						68˚C	6 min	
						68˚C	5 min	
						4˚C	∞	
Mito 2	Devario sondhii	sonM2F2R2			5853	98˚C	2 min	
	GenBank: PV357787.1	sonM2F2	ATCGTTATCCCCACACTGGA	3203-3222		98˚C	10 sec	30x
		sonM2R2	GATTAGCTCCGCAGATTTCG	9055-9036		55˚C	15 sec	
						68˚C	6 min	
						68˚C	5 min	
						4˚C	∞	
Mito 2	Microdevario kubotai	kubM2F2kubM2R1			6571	98˚C	2 min	
	GenBank: PV357791.1	kubM2F2	ATCTTCGGAATGCAACTCCA	6374-6393		98˚C	10 sec	30x
		kubM2R1	CGGTGAGTGAGGGGATTTTA	12944-12925		55˚C	15 sec	
						68˚C	7 min	
						68˚C	5 min	
						4˚C	∞	

Amplicon	Species	Primer pair	Sequence	Position	length	thermocycler		
Mito 3	Danio rerio	Mito 3 F1R1			6395	98°C	2 min	
	GenBank: PV357751.1	M3F1	CCCCCTGAGGATTTACAGGA	12215-12234		98°C	10 sec	
		M3R1	TCTCGGTGTAAGTGAGATGCTT	2013-1992		55°C	15 sec	30x
						68°C	7 min	
						68°C	5 min	
						4°C	∞	
Mito 3	Danio aesculapii	Mito 3 F1R1			6388	98°C	2 min	
	GenBank: PV357755.1	M3F1	CCCCCTGAGGATTTACAGGA	12223-12242		98°C	10 sec	
		M3R1	TCTCGGTGTAAGTGAGATGCTT	2008-1987		55°C	15 sec	30x
						68°C	7 min	
						68°C	5 min	
						4°C	∞	
Mito 3	Danio kyathit	Mito 3 F1R1			6397	98°C	2 min	
	GenBank: PV357759.1	M3F1	CCCCCTGAGGATTTACAGGA	12235-12254		98°C	10 sec	
		M3R1	TCTCGGTGTAAGTGAGATGCTT	2014-1995		55°C	15 sec	30x
						68°C	7 min	
						68°C	5 min	
						4°C	∞	
Mito 3	Danio nigrofasciatus	Mito 3 F1R1			6383	98°C	2 min	
	GenBank: PV357763.1	M3F1	CCCCCTGAGGATTTACAGGA	12216-12235		98°C	10 sec	
		M3R1	TCTCGGTGTAAGTGAGATGCTT	2000-1979		55°C	15 sec	30x
						68°C	7 min	
						68°C	5 min	
						4°C	∞	
Mito 3	Danio tinwini	Mito 3 F1R1			6359	98°C	2 min	
	GenBank: PV357768.1	M3F1	CCCCCTGAGGATTTACAGGA	12185-12204		98°C	10 sec	
		M3R1	TCTCGGTGTAAGTGAGATGCTT	1972-1951		55°C	15 sec	30x
						68°C	7 min	
						68°C	5 min	
						4°C	∞	
Mito 3	Danio albolineatus	kaM3F2R1			6198	98°C	2 min	
	Genbank: PV357771.1	kaM3F2	CACACCCGAGAGCATCTTTT	12544-12563		98°C	10 sec	
		kaM3R1	TGCTTTTTGGTGCTTGTACG	2210-2191		55°C	15 sec	30x
						68°C	7 min	
						68°C	5 min	
						4°C	∞	
Mito 3	Danio margaritatus	maM3F1M3R1			6362	98°C	2 min	
	GenBank: PV357775.1	maM3F1	ATTGCCCACGGACTAACATC	12277-12296		98°C	10 sec	
		M3R1	TCTCGGTGTAAGTGAGATGCTT	2015-1994		56°C	15 sec	
						68°C	7 min	
						68°C	5 min	
						4°C	∞	
Mito 3	Danio erythromicron	Not used						
		Could not amplify through the D-loop						
Mito 3	Danio choprae	maM3F1M3R1			6321	98°C	2 min	
	GenBank: PV357779.1	maM3F1	ATTGCCCACGGACTAACATC	12206-12225		98°C	10 sec	
		M3R1	TCTCGGTGTAAGTGAGATGCTT	1975-1954		56°C	15 sec	30x
						68°C	7 min	
						68°C	5 min	
						4°C	∞	
Mito 3	Devario aequipinnatus	gntM3F1R1			5760	98°C	2 min	
	GenBank: PV357783.1	gntM3F1	TCGTATCATGCGGAAATTGA	13026-13045		98°C	10 sec	
		gntM3R1	GCAAAAGGTACGAGGGTTGA	2239-2220		55°C	15 sec	30x
						68°C	6 min	
						68°C	5 min	
						4°C	∞	
Mito 3	Devario sondhii	sonM3F2R1			5744	98°C	2 min	
	GenBank: PV357787.1	sonM3F2	CGAAGACCTTACCCCAGGA	8820-8838		98°C	10 sec	
		sonM3R1	CGAGTCGTCGAATGGTGTTA	14563-14544		55°C	15 sec	30x
						68°C	6 min	
						68°C	5 min	
						4°C	∞	
Mito 3	Microdevario kubotai	Mito 3 F1R1			6519	98°C	2 min	
	GenBank: PV357791.1	M3F1	CCCCCTGAGGATTTACAGGA	12408-12428		98°C	10 sec	
		M3R1	TCTCGGTGTAAGTGAGATGCTT	2150-2129		55°C	15 sec	30x
						68°C	7 min	
						68°C	5 min	
						4°C	∞	

Amplicon	Species	Primer pair	Sequence	Position	length	thermocycler		
zmt 1	Danio rerio	A1F1R2			8791	98˚C	2 min	
	GenBank: PV357751.1	alt zmt1 F1	AATCCGAAGATCGGAGGTTA	16484-16503		98˚C	10 sec	30x
		alt zmt1 R2	AGCGGTTTGGTTTAATCGTC	8678-8659		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 1	Danio aesculapii	A1F1R2			8796	98˚C	2 min	
	GenBank: PV357755.1	alt zmt1 F1	AATCCGAAGATCGGAGGTTA	16491-16510		98˚C	10 sec	30x
		alt zmt1 R2	AGCGGTTTGGTTTAATCGTC	8684-8665		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 1	Danio kyathit	A1F1R2			8809	98˚C	2 min	
	GenBank: PV357759.1	alt zmt1 F1	AATCCGAAGATCGGAGGTTA	16505-16524		98˚C	10 sec	30x
		alt zmt1 R2	AGCGGTTTGGTTTAATCGTC	8696-8677		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 1	Danio nigrofasciatus	A1F1R2			8790	98˚C	2 min	
	GenBank: PV357763.1	alt zmt1 F1	AATCCGAAGATCGGAGGTTA	16486-16505		98˚C	10 sec	30x
		alt zmt1 R2	AGCGGTTTGGTTTAATCGTC	8677-r8658		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 1	Danio tinwini	A1F1R2			8760	98˚C	2 min	
	GenBank: PV357768.1	alt zmt1 F1	AATCCGAAGATCGGAGGTTA	16459-16478		98˚C	10 sec	30x
		alt zmt1 R2	AGCGGTTTGGTTTAATCGTC	8647-8628		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 1	Danio albolineatus	A1F1R2			8747	98˚C	2 min	
	Genbank: PV357771.1	alt zmt1 F1	AATCCGAAGATCGGAGGTTA	16419-16438		98˚C	10 sec	30x
		alt zmt1 R2	AGCGGTTTGGTTTAATCGTC	8634-8615		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 1	Danio margaritatus	maZ1F1maZ1R1			8472	98˚C	2 min	
	GenBank: PV357775.1	maZ1F1	TTGGAGGGATACCTGTGGAG	16347-16366		98˚C	10 sec	30x
		maZ1R1	TGATGCTGCGTCTTGAAATC	8195-8176		60˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 1	Danio erythromicron	Not used						
		Could not amplify through the D-loop						
zmt 1	Danio choprae	choZ1F1maZ1R1			8363	98˚C	2 min	
	GenBank: PV357779.1	choZ1F1	TTGGCCAAGTAGCATCAATTC	16313-16333		98˚C	10 sec	30x
		maZ1R1	TGATGCTGCGTCTTGAAATC	8124-8105		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 1	Devario aequipinnatus	gntZ1F1R2			8291	98˚C	2 min	
	GenBank: PV357783.1	gntZ1F1	GCAGGCCTCCTAGAAAACAA	16363-16382		98˚C	10 sec	30x
		gntZ1R2	GCGTAGGCTGTGCCATTAAG	8107-8088		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 1	Devario sondhii	sonz1F5R4			9257	98˚C	2 min	
	GenBank: PV357787.1	sonz1F5	CTGCTTGGATTCGCCATAAT	16292-16311		98˚C	10 sec	30x
		sonz1R4	TGGCAGGATGGTTCATACAA	8652-8633		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 1	Microdevario kubotai	kubz1F1R1			8598	98˚C	2 min	
	GenBank: PV357791.1	kubz1F1	GCCAGCTCTTCACACATCAA	16404-16423		98˚C	10 sec	30x
		kubz1R1	ACCGGCAACTTCATTTTGAC	8305-8286		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	

Amplicon	Species	Primer pair	Sequence	Position	length	thermocycler		
zmt 2	Danio rerio	A2F2R1			8600	98˚C	2 min	
	GenBank: PV357751.1	alt zmt2 F2	CGAGAAAGGAGGGAATTGAAC	7976-7996		98˚C	10 sec	30x
		alt zmt2 R1	ATCCTGGCTTTGGGAGTCAG	16575-16556		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 2	Danio aesculapii	A2F2R1			8601	98˚C	2 min	
	GenBank: PV357755.1	alt zmt2 F2	CGAGAAAGGAGGGAATTGAAC	7981-8001		98˚C	10 sec	30x
		alt zmt2 R1	ATCCTGGCTTTGGGAGTCAG	16581-16562		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 2	Danio kyathit	A2F2R1			8604	98˚C	2 min	
	GenBank: PV357759.1	alt zmt2 F2	CGAGAAAGGAGGGAATTGAAC	7993-8013		98˚C	10 sec	30x
		alt zmt2 R1	ATCCTGGCTTTGGGAGTCAG	16596-16577		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 2	Danio nigrofasciatus	A2F2R1			8604	98˚C	2 min	
	GenBank: PV357763.1	alt zmt2 F2	CGAGAAAGGAGGGAATTGAAC	7974-7944		98˚C	10 sec	30x
		alt zmt2 R1	ATCCTGGCTTTGGGAGTCAG	16577-16558		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 2	Danio tinwini	A2F2R1			8607	98˚C	2 min	
	GenBank: PV357768.1	alt zmt2 F2	CGAGAAAGGAGGGAATTGAAC	7944-7964		98˚C	10 sec	30x
		alt zmt2 R1	ATCCTGGCTTTGGGAGTCAG	16550-16531		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 2	Danio albolineatus	A2F2R1			8580	98˚C	2 min	
	Genbank: PV357771.1	alt zmt2 F2	CGAGAAAGGAGGGAATTGAAC	7931-7951		98˚C	10 sec	30x
		alt zmt2 R1	ATCCTGGCTTTGGGAGTCAG	16510-16491		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 2	Danio margaritatus	A2F2R1			8605	98˚C	2 min	
	GenBank: PV357775.1	alt zmt2 F2	CGAGAAAGGAGGGAATTGAAC	7998-8018		98˚C	10 sec	30x
		alt zmt2 R1	ATCCTGGCTTTGGGAGTCAG	16602-16583		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 2	Danio erythromicron				8604	98˚C	2 min	
	Compared to RefSeq: AP011419.1	alt zmt2 F2	CGAGAAAGGAGGGAATTGAAC	7049-7069		98˚C	10 sec	30x
		alt zmt2 R1	ATCCTGGCTTTGGGAGTCAG	15652-15633		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 2	Danio choprae	A2F2R1			8599	98˚C	2 min	
	GenBank: PV357779.1	alt zmt2 F2	CGAGAAAGGAGGGAATTGAAC	7932-7952		98˚C	10 sec	30x
		alt zmt2 R1	ATCCTGGCTTTGGGAGTCAG	16530-16511		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 2	Devario aequipinnatus	A2F2R1			8589	98˚C	2 min	
	GenBank: PV357783.1	alt zmt2 F2	CGAGAAAGGAGGGAATTGAAC	7937-7957		98˚C	10 sec	30x
		alt zmt2 R1	ATCCTGGCTTTGGGAGTCAG	16525-16506		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 2	Devario sondhii	A2F2R1			8585	98˚C	2 min	
	GenBank: PV357787.1	alt zmt2 F2	CGAGAAAGGAGGGAATTGAAC	8292-8312		98˚C	10 sec	30x
		alt zmt2 R1	ATCCTGGCTTTGGGAGTCAG	16875-16856		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	
zmt 2	Microdevario kubotai	A2F2R1			8574	98˚C	2 min	
	GenBank: PV357791.1	alt zmt2 F2	CGAGAAAGGAGGGAATTGAAC	8182-8202		98˚C	10 sec	30x
		alt zmt2 R1	ATCCTGGCTTTGGGAGTCAG	16755-16736		55˚C	15 sec	
						68˚C	9 min	
						68˚C	5 min	
						4˚C	∞	

Supplemental Table 2. Topology test results for maximum likelihood tree.

Table contains the statistical tests as described below demonstrating statistical differences when comparing the McCluskey & Postlethwait (2014) tree topology composed entirely of nuclear DNA fragments to our topology generated with H-strand mtDNA.

Tree	logL	deltaL	bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW	p-AU
McCluskey and Postlethwait, 2014	-92914.5737	0.062409	0.496 +	0.00099 -	0.503 +	0.00099 -	0.00118 -	0.484 +	9.68e-05 -
Current Study	-92914.5112	0	0.504 +	0.999 +	1 +	0.999 +	1 +	0.516 +	1 +

deltaL : logL difference from the maximal logl in the set.
bp-RELL : bootstrap proportion using RELL method (Kishino et al. 1990).
p-KH : p-value of one sided Kishino-Hasegawa test (1989).
p-SH : p-value of Shimodaira-Hasegawa test (2000).
p-WKH : p-value of weighted KH test.
p-WSH : p-value of weighted SH test.
c-ELW : Expected Likelihood Weight (Strimmer & Rambaut 2002).
p-AU : p-value of approximately unbiased (AU) test (Shimodaira, 2002).

Plus signs denote the 95% confidence sets.
Minus signs denote significant exclusion.
All tests performed 100000 resamplings using the RELL method.

Supplemental Table 3. Above- or below-threshold pairwise comparisons of single amino acid variations (SAVs) to phylogenetic distance.

Each line is a data point from Figure 3A that fell above or below the cutoff threshold created by the 60 bp window regression lines used to detect outliers.

Protein	Complex	Species 1	Species 2	Distance	SAVs	Ratio SAV to AA Total	Above or Below Threshold
nd5	complexI	Danio_rerio	Danio_albolineatus	0.542861	198	0.346153846	above
nd5	complexI	Danio_aesculapii	Danio_albolineatus	0.458129	202	0.353146853	above
nd5	complexI	Danio_kyathit	Danio_albolineatus	0.433969	200	0.34965035	above
nd5	complexI	Danio_nigrofasciatus	Danio_albolineatus	0.473353	199	0.347902098	above
nd5	complexI	Danio_tinwini	Danio_albolineatus	0.461375	202	0.353146853	above
nd5	complexI	Danio_albolineatus	Danio_margaritatus	0.552268	191	0.333916084	above
nd5	complexI	Danio_albolineatus	Danio_erythromicron	0.52046	186	0.325174825	above
nd5	complexI	Danio_albolineatus	Danio_choprae	0.528393	190	0.332167832	above
cox2	complexIV	Danio_rerio	Danio_albolineatus	0.542861	5	0.022321429	below
cox2	complexIV	Danio_rerio	Danio_margaritatus	0.521321	6	0.026785714	below
cox2	complexIV	Danio_aesculapii	Danio_albolineatus	0.458129	4	0.017857143	below
cox2	complexIV	Danio_tinwini	Danio_albolineatus	0.461375	5	0.022321429	below
cox2	complexIV	Danio_albolineatus	Danio_margaritatus	0.552268	8	0.035714286	below
cox2	complexIV	Danio_rerio	Devario_aequipinnatus	0.948857	69	0.308035714	above
cox2	complexIV	Danio_rerio	Devario_sondhii	0.911739	66	0.294642857	above
cox2	complexIV	Danio_aesculapii	Devario_aequipinnatus	0.864125	70	0.3125	above
cox2	complexIV	Danio_aesculapii	Devario_sondhii	0.827007	67	0.299107143	above
cox2	complexIV	Danio_kyathit	Devario_aequipinnatus	0.839965	70	0.3125	above
cox2	complexIV	Danio_kyathit	Devario_sondhii	0.802847	67	0.299107143	above
cox2	complexIV	Danio_nigrofasciatus	Devario_aequipinnatus	0.879349	71	0.316964286	above
cox2	complexIV	Danio_nigrofasciatus	Devario_sondhii	0.842231	68	0.303571429	above
cox2	complexIV	Danio_tinwini	Devario_aequipinnatus	0.867371	69	0.308035714	above
cox2	complexIV	Danio_tinwini	Devario_sondhii	0.830253	66	0.294642857	above
cox2	complexIV	Danio_albolineatus	Devario_aequipinnatus	0.979804	72	0.321428571	above
cox2	complexIV	Danio_albolineatus	Devario_sondhii	0.942686	69	0.308035714	above
cox2	complexIV	Danio_margaritatus	Devario_aequipinnatus	0.86683	69	0.308035714	above
cox2	complexIV	Danio_margaritatus	Devario_sondhii	0.829712	67	0.299107143	above
cox2	complexIV	Danio_erythromicron	Devario_aequipinnatus	0.835022	68	0.303571429	above
cox2	complexIV	Danio_erythromicron	Devario_sondhii	0.797904	66	0.294642857	above
cox2	complexIV	Danio_choprae	Devario_aequipinnatus	0.842955	69	0.308035714	above
cox2	complexIV	Danio_choprae	Devario_sondhii	0.805837	67	0.299107143	above
cox2	complexIV	Danio_rerio	Microdevario_kubotai	1.034651	82	0.366071429	above
cox2	complexIV	Danio_aesculapii	Microdevario_kubotai	0.949919	83	0.370535714	above
cox2	complexIV	Danio_kyathit	Microdevario_kubotai	0.925759	83	0.370535714	above
cox2	complexIV	Danio_nigrofasciatus	Microdevario_kubotai	0.965143	83	0.370535714	above
cox2	complexIV	Danio_tinwini	Microdevario_kubotai	0.953165	82	0.366071429	above
cox2	complexIV	Danio_albolineatus	Microdevario_kubotai	1.065598	86	0.383928571	above
cox2	complexIV	Danio_margaritatus	Microdevario_kubotai	0.952624	83	0.370535714	above
cox2	complexIV	Danio_erythromicron	Microdevario_kubotai	0.920816	84	0.375	above
cox2	complexIV	Danio_choprae	Microdevario_kubotai	0.928749	84	0.375	above
cox2	complexIV	Devario_aequipinnatus	Microdevario_kubotai	0.891148	89	0.397321429	above
cox2	complexIV	Devario_sondhii	Microdevario_kubotai	0.85403	87	0.388392857	above
cox3	complexIV	Danio_rerio	Danio_albolineatus	0.542861	7	0.028340081	below
cox3	complexIV	Danio_rerio	Danio_margaritatus	0.521321	6	0.024291498	below
cox3	complexIV	Danio_rerio	Danio_erythromicron	0.489513	7	0.028340081	below
cox3	complexIV	Danio_rerio	Danio_choprae	0.497446	6	0.024291498	below
cox3	complexIV	Danio_aesculapii	Danio_albolineatus	0.458129	7	0.028340081	below
cox3	complexIV	Danio_kyathit	Danio_erythromicron	0.380621	6	0.024291498	below
cox3	complexIV	Danio_tinwini	Danio_albolineatus	0.461375	5	0.020242915	below
cox3	complexIV	Danio_tinwini	Danio_margaritatus	0.439835	6	0.024291498	below
cox3	complexIV	Danio_tinwini	Danio_erythromicron	0.408027	6	0.024291498	below
cox3	complexIV	Danio_albolineatus	Danio_margaritatus	0.552268	7	0.028340081	below
cox3	complexIV	Danio_albolineatus	Danio_erythromicron	0.52046	8	0.032388664	below
cox3	complexIV	Danio_albolineatus	Danio_choprae	0.528393	5	0.020242915	below
cox3	complexIV	Danio_margaritatus	Danio_choprae	0.341519	4	0.016194332	below
cox3	complexIV	Danio_aesculapii	Devario_sondhii	0.827007	45	0.182186235	above
cox3	complexIV	Danio_kyathit	Devario_sondhii	0.802847	44	0.178137652	above
cox3	complexIV	Danio_nigrofasciatus	Devario_sondhii	0.842231	45	0.182186235	above
cox3	complexIV	Danio_tinwini	Devario_sondhii	0.830253	44	0.178137652	above
cox3	complexIV	Danio_erythromicron	Devario_sondhii	0.797904	44	0.178137652	above
cox3	complexIV	Danio_choprae	Devario_sondhii	0.805837	43	0.174089069	above

Supplemental File 1.

Additional data can be accessed at:

https://github.com/chamberlain-trevor/danionin_lineages

CHAPTER 3

Cytochrome c to cytochrome c oxidase docking orientation varies between *Danio* and *Devario* species

Chamberlain, T. J.

Laboratory of Genetics

University of Wisconsin – Madison

425-G Henry Mall

Madison, WI 53706

USA

3.1 Abstract

The electron transport chain (ETC) is vital to cellular respiration which produces energy for cells in the form of ATP. The roles carried out in the ETC are performed by complexes composed of subunits derived from both the mitochondrial genome and the nuclear genome. Therefore, the subunits within mitochondrial respiratory complexes tend to coevolve despite their sources of origin. However, during speciation two populations from a common ancestral origin may develop variation within their mitochondrial-encoded genes that coevolves with nuclear-encoded partners, and in hybrid crosses between the diverged populations, mitochondrial-nuclear incompatibilities may arise. In Danionin fish, a known embryonic lethal phenotype exists in hybrids between *Danio rerio* females and *Devario aequipinnatus* males. Recent examination of mitochondrial genomes has found a high level of protein variation within the respiratory complex IV gene, Cox2, which contains a binding domain where electron transfer occurs between the nuclear-encoded protein Cycsb and the mitochondrial-encoded Cox2. The high substitution rate within the Cox2 domain suggests the docking activity between it and Cycsb may be disrupted in *Danio rerio* x *Devario aequipinnatus* hybrids. This work uses predictive modeling to simulate and characterize Cycsb-Cox2 docking and examine a unique binding orientation in *Devario aequipinnatus* where Cycsb rotates to bind to Cox2. Further, this paper identifies R121 in *Devario aequipinnatus* Cox2 as necessary but insufficient to force the distinct Cycsb rotation and shows that R12 and R13 residues likely arose in Cycsb to compensate for hinderance caused by R121 and likely cause mitochondrial-nuclear incompatibility in *Danio-Devario* hybrids.

3.2 Introduction

Mitochondria are vital organelles for eukaryotic life. They perform a variety of cellular functions but are most particularly studied for their role in aerobic respiration in which oxidative phosphorylation (OXPHOS) occurs in the mitochondrial intermembrane space and produces the energy currency of the cell, ATP (Mitchell & Moyle, 1967; Saraste, 1999). The OXPHOS process has specific protein complexes devoted to either maintaining the electron transport chain (ETC) which uses redox reactions driven by electron transfer (ET) between respiratory chain complexes I-IV to pump protons into the intermembrane space, except for complex V (ATP synthase) which is devoted to using the a proton motive force generated by the proton gradient between the intermembrane space and the mitochondrial matrix to drive the phosphorylation of ADP to ATP (Zhao et al., 2019; Nolfi-Donegan et al., 2020).

A more in-depth examination of the OXPHOS process reveals a unique system that relies on protein-coding genes from two sources. In fact, complexes I, III, IV, and V in the ETC and ATP synthase are composed of subunits that originate in both the mitochondrial genome (mtDNA) and the nuclear genome (nDNA) (Levin et al., 2014; Rand, 2017; Signes & Fernandez-Vizarra, 2018). This combination of genetic origins requires that the components from either genome are compatible with each other to perform their specific functions, a state known as mitochondrial-nuclear (mito-nuclear) compatibility (Kenyon & Moraes, 1997; Hill, 2017; Burton, 2022). Often, the cooperating proteins need to bind together to form structural supports for a complex, but some proteins, such as Cox2 in complex IV which I will discuss in further detail later, have specific functional roles in the ETC (Signes & Fernandez-Vizarra, 2018).

Since mtDNA is maternally inherited, hybrids may not have the correct ratio of nuclear-encoded mitochondrial genes for mito-nuclear compatibility. These incompatibilities may arise in diverged species because of mutations that occur within the faster-evolving mtDNA that lead to compensatory mutations within nuclear-encoded subunits (Barreto & Burton, 2013; Hill, 2020). In fact, the Bateson-Dobzhansky-Muller incompatibility (BDMI) model of hybridization states that these coevolving mutations will increase over longer timespans, and when one population is reintroduced to another population of the same ancestral origin, mito-nuclear incompatibilities may lead to mitochondrial dysfunction up to embryonic lethality, becoming a source of postzygotic isolation leading to speciation (Welch, 2004; Burton & Barreto, 2012; Sunnucks et al., 2017).

One important part of the ETC is the transfer of electrons between complex III, the cytochrome bc1 complex, and complex IV, cytochrome c oxidase (*CcO*), which occurs via an intermediary electron transporter, cytochrome c (*Cc*) (Kadenbach, 2021; Signes & Fernandez-Vizarra, 2018). *Cc* is composed of two parts including a structural protein body (Cycs in mammals or Cycsb in fish) and a heme center (heme c) which stores electrons during ET (Lange & Hunte, 2002). *Cc* first docks with cytochrome bc1 at a binding site within Cyc1, a complex III protein, where it is reduced and picks up an electron and detaches (Lange & Hunte, 2002; Signes & Fernandez-Vizarra, 2018). Next, it travels to and binds to *CcO* where it is oxidized and delivers the electron from heme c to *CcO* at a docking site within Cox2, a *CcO* subunit (Scharlau et al., 2019).

Previous studies have characterized the ET pathway between *Cc* to *CcO*, a process that is generally highly conserved among most eukaryotes. The long-standing canonical ET pathway

described by Roberts and Pique (1999) was derived in a pairing of horse cytochrome c (Cycs) (PDB 1HRC) and bovine Cox2 (PDB 1OCC). In Figure 1, I have recreated that ET pathway which starts with transferring an electron from the CBC atom of heme c to W104 of Cox2. The electron then continues, transferring to M207, then to the dinuclear copper center, Cu_A, before continuing with several more electron transfers in Cox2 before the final step where oxygen acts as the final electron acceptor in the ETC in the formation of water as a byproduct of respiration (Roberts & Pique, 1999; Signes & Fernandez-Vizarra, 2018; Scharlau et al., 2019).

In one striking example of a *Cc* to *CcO* incompatibility, a single amino acid substitution was sufficient to cause a temperature-induced decrease in *CcO* activity. In a 2006 study by Harrison and Burton, cytochrome c (Cyc) amino acid substitutions were examined in two populations, SC and SD, of *Tigriopus californicus* copepods which had a known temperature-dependent hybrid incompatibility (Rawson & Burton, 2002). Harrison and Burton (2006) mutated three residues in the SC *T. californicus* Cyc protein that were suspected to alter function and found a glutamine to lysine substitution that significantly decreased the measured output of cox activity in Cyc and *CcO* combined protein isolates. They determined that the electrostatic signature of Cyc was more positively charged at the lysine substitution which changed its affinity and binding potential to Cox2. Additionally, the Cox2 protein of the SD population had two additional negatively charged acidic amino acid residues that most likely contributed to the co-evolved interaction of Cyc and Cox2 in SD *T. californicus* copepods.

In previous work, I characterized the variation between mtDNA in Danionin species (Chamberlain & Pelegri, in review). That study found that members of the genus *Devario*

underwent faster rates of evolution and increased amino acid variance in respiratory complex IV genes, particularly Cox2, relative to the *Danio* genus. This was of specific interest because our group found a lethal hybrid phenotype in crosses between *Danio rerio* (*D. rerio*) females and *Devario aequipinnatus* (*Dev. aequipinnatus*) males with 100% penetrance (Trevena et al., 2025). Furthermore, this hybrid phenotype has a striking resemblance to a cox-deficient phenotype previously characterized in *D. rerio* embryos (Baden et al, 2007).

Here, I will examine the components involved in the ET interface between cytochrome c (Cycsb) and Cox2 in comparisons with *Danio* and *Devario* protein structures. Using multiple sequence alignments (MSAs) and protein models generated in AlphaFold, I will characterize amino acid variants that lead to differences in the biochemical environment, especially regarding bonding pairs between basic and acidic amino acids. I will also use protein docking simulations to show the differential docking orientation that exists in *Devario* Cycsb to Cox2 interactions compared to the configuration of docking found in *Danio* protein pairings. This work will show that compensatory mutations in *Devario* Cox2 and Cycsb have likely led to mito-nuclear incompatibilities responsible for hybrid lethality in *Danio-Devario* hybrids.

3.3 Results

3.3.1 Cycsb and Cox2 points of variation among *Danio* and *Devario* species

I started by creating multiple sequence alignments (MSAs) at the canonical binding sites between the nuclear-encoded protein Cycsb and the mitochondrial-encoded protein Cox2 in yeast, human, bovine, *D. rerio*, and *Dev. aequipinnatus*. All protein sequences were obtained from protein sequence databases (accessions in methods) except for *Dev. aequipinnatus* Cycsb

which did not have a reference sequence, so I assembled the sequence from reads available from the NCBI Sequence Read Archives (SRA) (see methods). In previous literature, the electron transfer (ET) process between Cycsb and Cox2 was fully characterized with the point of ET occuring between the heme center (heme c) of *Cc* and W104 of Cox2 (Roberts & Pique, 1999; Schmidt et al., 2005; Scharlau et al., 2019) (Figure 2A, yellow). These studies went on to describe the rest of the pathway with canonical residues creating the binding site between Cycsb and Cox2 involving van der Waals interactions in parallel residues between a hydrophobic loop from Cox2 positions 102-105 (Figure 2A, orange) and a β-strand at Cycsb residues 80-83 (Figure 2B, green). Further, Cycsb depends on polar interactions between several positively charged lysine or arginine residues on its surface (Figure 2B, blue) and a number of negatively charged glutamic acid and aspartic acid residues in Cox2 (Figure 2A, red). With these sites in mind, I identified SAVs between *D. rerio* and *Dev. aequipinnatus* within these residues or in proximity to binding sites, namely Q12R, K13R and I81V in Cycsb and H102Y, E109D, E117A, Y121R, and E157T in Cox2 (Figure 2A, magenta). Additionally, I noted that Q12 and K13 appear to be conserved, with K13 conservation in eukaryotes examined in previous studies that found substitutions of residue K13 impact *CcO* activity (Bashir et al., 2014; Scharlau et al., 2019; Brand et al., 2022). So, I performed an MSA for the sites surrounding these amino acids with *Danio* species including *D. rerio*, *D. aesculapii*, *D. tinwini*, *D. albolineatus*, *D. margaritatus*, and *D. erythromicron* and *Devario* species *Dev. aequipinnatus*, *Dev. devario*, and *Dev. kakhienensi* (see methods for sequence acquisition and accessions). Interestingly, I found the Q12R and K13R substitutions in all three *Devario* sequences (Figure 2C). I also generated a similar MSA for Cox2 examining positions around the SAVs listed above and found a Y121R substitution in

Devario species that led to a polarity change from negative to positive within the active region of the Cox2 binding domain (Figure 2D).

3.3.2 Orientation shift of *Danio* or *Devario* Cycsb relative to *Devario* Cox2

To examine the protein structure variations between *D. rerio* and *Dev. aequipinnatus* Cycsb and Cox2, I downloaded the predicted *D. rerio* protein structures from AlphaFoldDB (Jumper et al., 2021; Varadi et al., 2024) and modeled the *Dev. aequipinnatus* proteins in AlphaFold2 in the ColabFold Google Colab notebook (Jumper et al., 2021; Mirdita et al., 2022). Then, I copied the heme center from PDB 1HRC into both *D. rerio* and *Dev. aequipinnatus* Cycsb predictions and uploaded the Cox2 and Cycsb files that contained heme centers to the HADDOCK 2.4 Webserver (Honorato et al., 2021; Honorato et al., 2024) for docking prediction using restraints to keep the heme center of Cycsb to a distance of 3.5 Å from W104 in all scenarios. In docking scenarios between all possible partners between *D. rerio* and *Dev. aequipinnatus* Cox2 and Cycsb proteins (e.g. *D. rerio* Cox2 with *Dev. aequipinnatus* Cycsb) I found that when pairing with *Dev. aequipinnatus* Cox2, Cycsb from either *Dev. aequipinnatus* or *D. rerio* were rotated in comparison to pairings between *D. rerio* Cox2 and Cycsb from either species (Figure 3, Supplemental Figure 1A). Further, I used the Measurement wizard in PyMOL v3.1.4.1 (Schrödinger, LLC., 2025) and found the angle shift of Cycsb K25 in the *Dev. aequipinnatus* configuration was 41.2° clockwise from the same amino acid in the pairing of Cox2 and Cycsb in *D. rerio*. Additionally, this orientation decreased the surface-surface binding area between Cox2 and Cycsb in the *Dev. aequipinnatus* configuration.

To test if the rotated configuration could be repeated, or if features of the specific model predicted by AlphaFold2 (Jumper et al., 2021; Mirdita et al., 2022) was responsible, I performed

an *in silico* mutagenesis trial to replicate the *Dev. aequipinnatus* rotation in *D. rerio*. I used the Mutagenesis wizard in PyMOL v3.1.4.1 (Schrödinger, LLC., 2025) to mutate 13 residues in the *D. rerio* binding site of Cox2 including V100T, H102Y, E109D, T111M, N115L, E117A, Y121R, P124D, Q126K, G132M, T138V, H140N, and E157T. Following mutagenesis, I simulated *D. rerio* Cycsb docking to the mutated Cox2 using the HADDOCK 2.4 Webserver (Honorato et al., 2021; Honorato et al., 2024) which predicted the docking orientation from *Dev. aequipinnatus* pairings with the same Cycsb rotation and decreased binding area (Supplemental Figure 1B).

3.3.3 Polar residue site variations between species

To investigate the different docking orientations of Cox2 and Cycsb I compared negatively charged acidic amino acid residues in Cox2 and basic residues with positive charges in Cycsb. After docking simulations in the HADDOCK 2.4 Webserver (Honorato et al., 2021; Honorato et al., 2024) I highlighted negatively charged aspartic and glutamic acid residues with red sticks in Cox2 and I represented positively charged lysine and arginine residues in Cycsb as blue sticks then compared the polar residue profiles between *D. rerio* and *Dev. aequipinnatus* (Figure 4). I found that the polar signature in *D. rerio* was comprised entirely of positively charged lysine residues on the interacting surface of Cycsb as described in previous literature (Roberts & Pique, 1999; Scharlau et al., 2019). The surface of Cycsb had lysine residues at positions K5, K8, K72, K86, and K87 with a single amino acid unique from *Dev. aequipinnatus* at K13 (Figure 4A). In Cox2, the acidic amino acids in the interface were located at D119, D127, D139, and D158 with several substitutions relative to *Dev. aequipinnatus* at E109, E117, E157, and D229. Three of these variants are nonpolar in *Dev. aequipinnatus* including A117, T157, and S229.

I turned my attention to *Dev. aequipinnatus* polar sites and found the architecture to be quite different. First, it was noticeable that there were two arginine residues, R12 and R13, contributing to the positively charged surface of Cycsb (Figure 4B). While these positions were already identified in previous sequence alignments, they were located in an active area of the binding site near the site of ET and were likely capable of forming strong polar bonds with electron donors during protein interactions. In addition, the side chains of arginine are longer than Lysine which suggested a change in the spatial organization of the binding site, and R12 was substituting glutamine which indicated a shift towards a strong positive charge at position 12 in *Dev. aequipinnatus* and the potential for new hydrogen bonds (H-bonds) and salt bridges to form. The rest of the Cycsb surface in *Dev. aequipinnatus* contained lysine residues in the same positions as *D. rerio*, but the surface of Cox2 had additional variations in polar residues. First, there was one acidic residue unique to *Dev. aequipinnatus* at D124 and the protein had an aspartic acid at position 109 compared with the glutamic acid at that residue in *D. rerio*. However, among the negatively charged acidic acid residues responsible for attracting the positively charged Cycsb surface, there was a positively charged arginine at position 121 in Cox2 of *Dev. aequipinnatus* that is likely to change the polar bonds and electrostatic environment of the surface of Cox2 and is centrally located near the site of electron transfer.

While investigating residues involved in polar interactions, I also examined the hydrophobic loop in *D. rerio* and *Dev. aequipinnatus* Cox2 and the parallel β-strand in Cycsb that is responsible for van der Waals interactions between the two proteins (Supplemental Figure 2). The positions involved in the Cox2 hydrophobic loop were H102, Q103, W104, and Y105, and in Cycsb the positions M80, I81, F82, and A83 made up the parallel β-strand (Supplemental

Figure 2A). In *Dev. aequipinnatus*, the positions were the same except for a H102Y substitution in Cox2 and a I81V substitution in Cycsb. Even with these changes, these regions still appeared to be similar in function and proximity to each other in either docking scenario.

3.3.4 Mapping electrostatic potential in Cox2 and Cycsb surfaces

Next, I examined the charge distributions in Cox2 and Cycsb by mapping electrostatic potentials across protein surfaces. To perform this, I used the APBS Electrostatics tool in PyMOL v3.1.4.1 (Schrödinger, LLC., 2025). This mapped predicted charge potentials on protein surfaces by coloring negative electrostatic potentials in red and positively charged regions in blue which I performed in each mito-nuclear interface of *D. rerio* and *Dev. aequipinnatus* Cox2 to *D. rerio* and *Dev. aequipinnatus* Cycsb (e.g. *D. rerio* Cox2 docking *Dev. aequipinnatus* Cycsb) (Figure 5). Using this approach, I found that mutations identified earlier while looking for differences in polar residues contributed to a varying electrostatic map in *Dev. aequipinnatus* Cox2 and Cycsb relative to the *D. rerio* proteoforms. Specifically, R121 led to a decrease in negative charge on the surface of Cox2 identified by a patch of white surface central to the Cycsb docking site and can be seen in the docking scenarios involving *Devario* Cox2 (Figure 5B, 5D). In contrast, *D. rerio* Cox2 was found to have a negative charge distributed throughout that same region (Figure 5A, 5C).

In Cycsb, the conserved Q12 and K13 residues in *D. rerio* appeared as a gray, uncharged residue in Q12 next to a blue, positively charged site at K13, with both sites located just above the ET access point where there is an opening in the surface leading to heme c (Figure 5E, 5H). However, the two basic arginine residues at Cycsb R12 and R13 in *Dev. aequipinnatus* made the

two sites appear considerably blue compared to *D. rerio*, but in addition to the color, these sites were set closer together and formed a comparatively deep trench in the protein surface. Judging by its deep blue color, this trench has a strong positive electrostatic charge and is likely involved in H-bonds and salt bridges during ET (Figure 5F, 5G).

3.3.5 Cycsb R12 and R13 and Cox2 R121 change the polar bonds near the ET site

Corresponding to the evidence found that Cycsb R12, R13, and Cox2 R121 alter the polar environment of the Cycsb-Cox2 docking site in *Dev. aequipinnatus*, I next visualized H-bonds and salt bridges involving those three residues and in all four mito-nuclear combinations between *D. rerio* and *Dev. aequipinnatus* Cox2 and Cycsb (Figure 6). In *D. rerio*, Cox2 D119 and D139 formed H-bonds and salt bridges with Cycsb K13 while Cox2 Y121 and Cycsb Q12 were uninvolved in site-specific interactions (Figure 6A). Binding in this manner was straightforward with the polar bonding directly from Cox2 straight to Cycsb itself. However, in *Dev. aequipinnatus*, D119 participated in H-bonds to Cycsb R13 and formed a salt bridge to the side chain of R12 with an additional salt bridge to Cox2 R121 (Figure 6B). Further, D139 forms a salt bridge with R121 instead of R13, and R121 spans the gap to Cycsb by forming H-bonds with the R12 main chain carbonyl oxygen. In this arrangement, R13 was stretched laterally across the gap and R12 was brought forward to connect to D119, all seemingly in a compensatory effort to circumvent the positively charged R121 side chain residing in the space.

In the cross-species interactions, *D. rerio* Cox2 also took on a more complicated arrangement when *Dev. aequipinnatus* Cycsb docked with it and the *Dev. aequipinnatus* Cox2 to *D. rerio* Cycsb architecture resembled the original *Dev. aequipinnatus* docking conformation. Indeed, the

D. rerio Cox2 formed H-bonds and salt bridges to *Dev. aequipinnatus* Cycsb R13, similar to polar bonds formed in *D. rerio* only, but the addition of Cycsb R12 caused a salt bridge to form between it and D139 (Figure 6C). This could possibly change the angle of approach or lengthen the distance between heme c and W104 in cases where a forced restraint was not applied to bring those sites within 4 Å of each other. In contrast, the docking of *D. rerio* Cycsb to *Dev. aequipinnatus* Cox2 had sites that needed to circumvent R121 in Cox2, so the same approach as the natural *Dev. aequipinnatus* interface was observed with Cox2 D119 bonding with Cycsb K13 and R121 and D139 bonding to R121 while R121 spans the gap with H-bonds to the main chain of Q12 (Figure 6D). The structure in this scenario was missing the side chain bond to R12, but this configuration also led to a rotated Cycsb in response to the R121 residue in Cox2.

3.3.6 Simulated mutations changed docking orientation between Cycsb and Cox2

Finally, I tested for residues in the *Dev. aequipinnatus* and *D. rerio* docking scenarios by performing *in silico* site mutagenesis assays to analyze docking orientations. First, I tested whether R121 in *Dev. aequipinnatus* Cox2 was responsible for the observed rotation of Cycsb by inserting a R121Y substitution using the Mutagenesis wizard tool in PyMOL v3.1.4.1 (Schrödinger, LLC., 2025) followed by a docking simulation performed in the HADDOCK 2.4 Webserver (Honorato et al., 2021; Honorato et al., 2024). In this scenario, Cox2 D119 and D139 both formed polar bonds directly to Cycsb R13 in similar fashion to *D.* rerio binding activity (Figure 7A). However, similar to *Dev. aequipinnatus* Cycsb to *D. rerio* polar bond formation, an additional salt bridge was formed between R12 and D139, plus a new H-bond was observed between Y121 and the main chain of R13. However, the single R121Y substitution led to a

docking orientation like *D. rerio* that lacked the rotation observed in *Dev. aequipinnatus* (Figure 7B).

I also performed a similar mutagenesis trial in *D. rerio* Cox2 to attempt to force a Cycsb rotation in the simulated interface. After attempting several single site mutagenesis approaches with no change to Cycsb orientation, I used the same approach as above to induce thirteen mutations in the surface of Cox2 including V100T, H102Y, E109D, T111M, N115L, E117A, Y121R, P124D, Q126K, G132M, T138V, H140N, and E157T. I then analyzed the polar bonds in this scenario which varied in several locations including Cox2 D119 forming polar bonds with Cycsb K13 and both D119 and D139 forming salt bridges to R121 which created its characteristic H-bonds to the main chain of Q12 (Figure 7C). Still, several new polar bonds were also created including a Cycsb K13-E91 salt bridge which formed another salt bridge with K87 which connected to Cox2 D109 via an H-bond with was made a salt bridge with Cox2 K98. Nevertheless, the new bonding scenario did induce the characteristic rotation observed in *Dev. aequipinnatus* docking activity (Figure 7D).

3.4 Discussion

3.4.1 Coevolution acted between Cox2 and Cycsb forming a docking pattern distinct to the *Devario* genus

In this work, I examined the binding domain variation of *Dev. aequipinnatus* Cox2 relative to *D. rerio* protein structure when binding to Cycsb during ET. My findings provide details of the interspecies biochemical and structural differences that exist in these protein-protein interactions, mainly that *Dev. aequipinnatus* contains a Y121R mutation respective to *D. rerio* that creates a

scenario where Cycsb is forced to circumvent the large, positively charged basic arginine residue in order to form polar bonds and bring Cycsb heme c in proximity to Cox2 W104 for successful ET. Bypassing R121 is accomplished through Cycsb rotating and new bonds forming with R121 to create a successful interface. This orientation is further enhanced by specific compensatory substitutions at Q12R and K13R in Cycsb which create a stronger positive electrostatic potential to direct Cycsb to the correct binding orientation and form new polar bonds that strengthen the protein-protein interface and create H-bonds between R121 so it is incorporated in the docking process rather than a hinderance.

The Cox2 R121 and Cycsb R12 and R13 residues in *Dev. aequipinnatus* are likely a result of coevolution between the mitochondrial-encoded Cox2 and the nuclear-encoded Cycsb protein. One line of evidence for this is found in the conservation of R121 in Cox2 among other *Devario* species along with conservation of R12 and R13 in their Cycsb protein sequences. Additionally, in previous work I found that mitochondrial-encoded cox proteins undergo rapid evolutionary signatures in *Devario* species and contain numerous single amino acid variants that are likely to coevolve with nuclear-encoded partners to maintain vital mitochondrial functions within the *Devario* lineage (Chamberlain & Pelegri, in review). Furthermore, in that work I found that Cox2 had undergone relaxed selection in the *Devario* genus, which could indicate how a potentially disruptive mutation such as R121 was introduced and rapidly adapted to in Cycsb with substitutions to positions 12 and 13.

Additionally, other work supports mito-nuclear coevolution of cox proteins and adds that the mitochondrial-encoded R121 likely evolved prior to the R12 and R13 substitutions. Previous

studies have found that mtDNA rapid relative to nDNA (Havird & Sloan, 2016; Hill, 2020) which generally leads to compensatory mutations arising after mtDNA substitutions begin to fix in populations to improve fitness (Barreto et al., 2018; Weaver et al., 2022). Regarding cox proteins specifically, early mito-nuclear studies in primates initially missed signatures of increased cox protein variation between primate species until they analyzed the substitution rates of nuclear-encoded cox subunits and found elevated substitution rates due to interactions with and coevolution with faster-evolving mitochondrial-encoded proteins (Wu et al., 2000; Grossman et al., 2001; Doan et al., 2004). Thus, mito-nuclear coevolution in complex IV subunits is well documented and likely to have occurred in between residues with significantly altered functional properties such as the arginine substitutions described here.

3.4.2 A mito-nuclear incompatibility likely exists between *D. rerio* Cox2 and *D. aequipinnatus* Cycsb

This work provides evidence that the mito-nuclear interaction of Cox2 and Cycsb in *D. rerio* x *Dev. aequipinnatus* hybrids likely leads to dysfunctional cox activity. Our group previously found that a hybrid incompatibility leading to embryonic lethality exists in hybrids between *D. rerio* females and *Dev. aequipinnatus* males (Trevena et al., 2025) and my work comparing protein sequences in Danionin species identified the high number of amino acid variants in cox proteins, specifically as a possible source of mito-nuclear incompatibility (Chamberlain & Pelegri, in review). In this paper, I have included a site-specific analysis of the *D. rerio* Cox2 binding domain and its interactions with Cycsb from *Dev. aequipinnatus* and it appears the two arginine residues in Cycsb positions 12 and 13 are an adaptation to the docking rotation in *Dev. aequipinnatus* but the longer extension of R13 and salt bridge formed between R12 and Cox2

D139 are likely to prevent heme c and W104 from proper encounters and could hinder cox activity in *D. rerio* x *Dev. aequipinnatus* hybrids.

Interestingly, in hybrids with *Dev. aequipinnatus* females crossed to *D. rerio* males, the mitochondrial-encoded *Dev. aequipinnatus* Cox2 would cause the rotated orientation in interactions with *D. rerio* Cycsb. However, while the salt bridge between D139 and R12 would be absent in scenario and could lead to less efficient binding, it appears access to W104 by heme c would be largely unhindered. This may lead to a higher functionality in hybrids with this parental configuration and may allow embryonic development. This is important to note because in hybrid crosses using *Dev. aequipinnatus* females and *D. rerio* males the hybrid lethality of the reciprocal cross is not present (unpublished data, Pelegri Lab, University of Wisconsin – Madison Laboratory of Genetics). Non-reciprocal crosses, in which a negative phenotype is observed with one parental configuration but is not present when switching which species are male and female, is a hallmark of BDMIs (Turelli & Moyle, 2007; Bolnick et al., 2008). This occurs after two populations from an ancestral lineage are diverged and accumulate differential mutations, but each population keeps different genetic loci from the ancestral genotype (Coyne & Orr, 2004). If these loci are recombined in hybrids, compatibility may work between the ancestral genotype conserved in one population, but that same configuration no longer exists when the partners are swapped. The conservation of Cycsb residues Q12, K13, and Cox2 Y121 seen in *D. rerio* are likely evidence of loci conserved from ancestral populations while *Dev. aequipinnatus* has a novel genotype in these residues. Therefore, the site access inhibition from R12 and R13 is a likely cause of mito-nuclear incompatibility in *Dev. aequipinnatus* paternal crosses, and the ancestral Q12 and K13 genotype is still functional in *D. rerio* paternal crosses.

3.4.3 R121 is necessary but not sufficient for the unique docking orientation seen in *Dev. aequipinnatus*

In simulated mutagenesis experiments I was able to cause the *Dev. aequipinnatus* Cycsb not to rotate during docking simulations with the simple R121Y substitution in Cox2. This indicates that R121Y is the necessary component of the distinct docking conformation observed in *Dev. aequipinnatus* Cycsb-Cox2 interactions and removing its impact prevents the rotation of Cycsb. However, I was surprised to discover that inducing the reciprocal substitution, Y121R, in *D. rerio* Cox2 did not alter the docking conformation at all. In fact, it was not until I combined the mutations from thirteen target sites that the orientation of Cycsb finally changed in *D. rerio* to match the rotation observed in *Dev. aequipinnatus* protein-protein interactions. Therefore, while 121R is necessary for the unique *Dev. aequipinnatus* docking, other residues in the surface of Cox2 also contribute to the polarity and binding potential that ultimately leads to the final orientation change. To investigate this further, I could work to reverse select mutations from the thirteen mutated sites in *D. rerio* Cox2 and assess which residues lead to a reversal in Cycsb rotation.

In summary, site-specific analyses of Cycsb-Cox2 docking variations between *D. rerio* and *Dev. aequipinnatus* reveal sites specific to the orientation and function of cox activity in a new model system for studying mito-nuclear interactions. The identification of a single residue that is responsible for Cycsb protein rotation in *Dev. aequipinnatus* docking offers an interesting target for future *in vivo* and *in silico* studies to better understand the role of mito-nuclear coevolution as a mechanism of speciation. This study also adds to the body of literature studying how mito-

nuclear incompatibilities arise with a detailed molecular-level example of a suspected hybrid lethal incompatibility.

3.5 Methods

3.5.1 Protein sequence assembly and multiple sequence alignments

To generate MSAs for amino acid comparisons I obtained Cyc/Cycs/Cycsb and Cox2 protein sequences from available databases for *S. cerevisiae* (Cyc – PDB: 5KPF, Cox2 – GenBank: AJU15444.1), human (CYCS – GenBank: AEP27189.1, COX2 – NCBI Reference Sequence: YP_003024029.1), bovine (Cycs – NCBI Reference Sequence: NP_001039526.1, Cox2 – NCBI Reference Sequence: YP_209208.1), *D. rerio* (Cycsb – NCBI Reference Sequence: NP_001002068.1, Cox2 – GenBank: XQJ35097.1) and *Dev. aequipinnatus* (Cox2 – GenBank: XQJ35513.1). However, *Dev. aequipinnatus* did not have a reference protein sequence for Cycsb, so I assembled the sequence from reads available through the NCBI Sequence Read Archive (SRA) (SRA: ERR3332370, ERR3332374). I prepared the assembly by downloading the paired-end reads and splitting them with read identification using fastq-dump available through SRA-tools (SRA Toolkit Development Team, n.d.). Next, I performed quality assessment with FastQC v0.11.7 (Andrews, 2010) and trimmed the reads for adapters and quality using fastp v0.20.0 (Chen et al., 2018). Then, I performed the assembly using HybPiper v1.3.1 (Johnson et al., 2016) with the *D. rerio* protein Cycsb sequence (NCBI Reference Sequence: NP_001002068.1) as bait, BlastX for reads database generation (Camacho et al., 2009), and SPAdes v3.14.1 for assembly (Prjibelski et al., 2020). For final curation of the protein sequence, Exonerate v2.2.0 (Slater & Birney, 2005) was used as part of the HybPiper pipeline to remove exons and assemble the completed and translated protein sequence. MSAs for Cyc/Cycs/Cycsb

and Cox2 were then generated by concatenating the sequences into a single file and aligning them using ClustalW2 v2.1 (Larkin et al., 2007). Finally, to compare conservation of sequences between *Danio* and *Devario* species I used the *D. rerio* Cycsb (NCBI Reference Sequence: NP_001002068.1) and Cox2 (NCBI Reference Sequence: NP_059334.1) reference sequences to Blast SRA runs specific to *Danio* and *Devario* species and returned reads with sequences for Cycsb in *D. tinwini* (SRA: ERR3284970), *D. albolineatus* (SRA: ERR2983444), *D. margaritatus* (SRA: ERR5838153), *D. erythromicron* (SRA: ERR5838158), *Dev. devario* (SRA: ERR3332326), and *Dev. kakhienensi* (SRA: SRR18335089). In addition, the *D. aesculapii* Cycsb sequence was available by blasting latest genome assembly of the species, fDanAes4.1. For the Cox2 sequences, I was able to use in-house mtDNA assemblies from my previous work (Chamberlain & Pelegri, in review) in *D. rerio*, *D. aesculapii*, *D. tinwini*, *D. abolineatus*, *D. margaritatus*, *D. erythromicron*, *Dev. aequipinnatus*, and *Dev. sondhii*. In addition, I blasted the SRA accessions for Cox2 sequences in *Dev. devario* (SRA: ERR3332326), *Dev. interruptus* (SRA: SRR16301567), and *Dev. kakhienensis* (SRA: SRR18335089). For these *Danio* and *Devario* MSAs, I targeted only a small region of each sequence to compare specific residues, so alignment was performed by pasting sequences into a text file and ensuring the correct positions were aligned.

3.5.2 Protein structure prediction

Protein structures for horse Cycs and bovine Cox2 were procured through the RCSB PDB database (1HRC and 1OCC, respectively) (Berman et al., 2000) and *D. rerio* Cycsb (AF-Q6IQM2-F1-v4) and Cox2 (AF-A0A0A0VG08-F1-v4) were downloaded from AlphaFoldDB (Jumper et al., 2021; Varadi et al., 2024). *Dev. aequipinnatus* did not have available structures for

either Cycsb or Cox2, so they were assembled in AlphaFold2 available through the ColabFold Google Colab notebook (Jumper et al., 2021; Mirdita et al., 2022). The protein sequences were input and AlphaFold was run in relaxed mode to accommodate protein folding conditions unique to the *Dev. aequipinnatus* biochemical signature. All proteins were then visualized in PyMOL v3.1.4.1 (Schrödinger, LLC., 2025) and differential residues in structures were identified.

3.5.3 Cycsb to Cox2 docking simulation

Before protein docking simulation, *D. rerio* and *Dev. aequipinnatus* Cycsb models were first adapted to include heme c in their structures. This was done by opening the horse Cycs model, PDB 1HRC in PyMOL v3.1.4.1 (Schrödinger, LLC., 2025) along with the Cycsb structures to adapt with the heme c residue. The structures were aligned to each other using PyMOL's Alignment plugin, then the heme c residue was selected and copied to a new object followed by copying the desired Cycsb structure into that object. Each Cycsb object with the copied heme c was then exported as a pdb file. To get horse and bovine structures for docking simulation, the PDBs 1HRC (horse Cycs) and 1OCC chain B (bovine Cox2) were imported into PyMOL and the structures were exported as pdb files.

To perform docking simulations, I used the HADDOCK 2.4 Webserver (Honorato et al., 2021; Honorato et al., 2024). I input Cox2 as the chain with identifier "A" and Cycs/Cycsb for chain "B". I then selected amino acids 102-105, 109, 119, and 158 as active residues for Cox2, and I selected the residues corresponding to 13, 72, 80-83, 86 and 87 for Cycsb active residues all based on previously published protein-protein interactions during Cycsb to Cox2 docking (Scharlau et al., 2019). I also included a distance restraint file which defined the distance

between W104 and heme c atom CBC (Cycsb residue 105) as 3.5 Å apart with an upper limit distance of 0.5 Å and a lower limit of 0.7 Å to ensure final structures prioritized the ET proximity between these sites. Protein structure predictions were then chosen based on the lowest estimated Haddock Z-score. Horse Cycs (PDB 1HRC) docking with bovine Cox2 (PDB 1OCC) was predicted followed by combinations of *D. rerio* or *Dev. aequipinnatus* Cycsb to *D. rerio* or *Dev. aequipinnatus* Cox2 docking scenarios until all combinations were predicted for *D. rerio* and *Dev. aequipinnatus* interfaces. Protein visualizations and characterizations, including H-bonds and salt bridges were then done in PyMOL v3.1.4.1 (Schrödinger, LLC., 2025).

3.5.4 Predicting electrostatic potential maps

To generate electrostatic potential maps, I used the APBS Electrostatics tool in PyMOL v3.1.4.1 (Jurrus et al., 2018; Schrödinger, LLC., 2025). The predicted structures from the HADDOCK 2.4 Webserver (Honorato et al., 2021; Honorato et al., 2024) were loaded into PyMOL, then each chain, (Chain A: Cox2, Chain B: Cycsb) was copied into a new object so the electrostatic potential could be predicted per protein and per configuration with various binding partners. The APBS Electrostatics tool was then used with all default settings to prepare the molecules, calculate an electrostatic map with APBS, then visualize the protein surface map with overlayed APBS predictions.

3.6 Acknowledgements

This work received funding from the National Institute of Food and Agriculture, United States Department of Agriculture, Hatch to the UW Madison College of Agricultural and Life Sciences (1013782), a UW2020 Round 6 award from the Office of the Provost, U. Wisconsin - Madison

(135-AAH8462), a Fall Competition award from the U. Wisconsin - Madison Graduate School, and additional support from the College of Agricultural and Life Sciences and the School of Medicine and Public Health at U. Wisconsin - Madison. Additional funding was received through the NIH Institutional National Research Service Award (T32GM007133).

The FP7 WeNMR (project# 261572), H2020 West-Life (project# 675858), the EOSC-hub (project# 777536) and the EGI-ACE (project# 101017567) European e-Infrastructure projects are acknowledged for the use of their web portals, which make use of the EGI infrastructure with the dedicated support of CESNET-MCC, INFN-LNL-2, NCG-INGRID-PT, TW-NCHC, IFCA-LCG2, UA-BITP, TR-FC1-ULAKBIM, CSTCLOUD-EGI, IN2P3-CPPM, SURFsara and NIKHEF, and the additional support of the national GRID Initiatives of Belgium, France, Italy, Germany, the Netherlands, Poland, Portugal, Spain, UK, Taiwan and the US Open Science Grid.

3.7 Figures

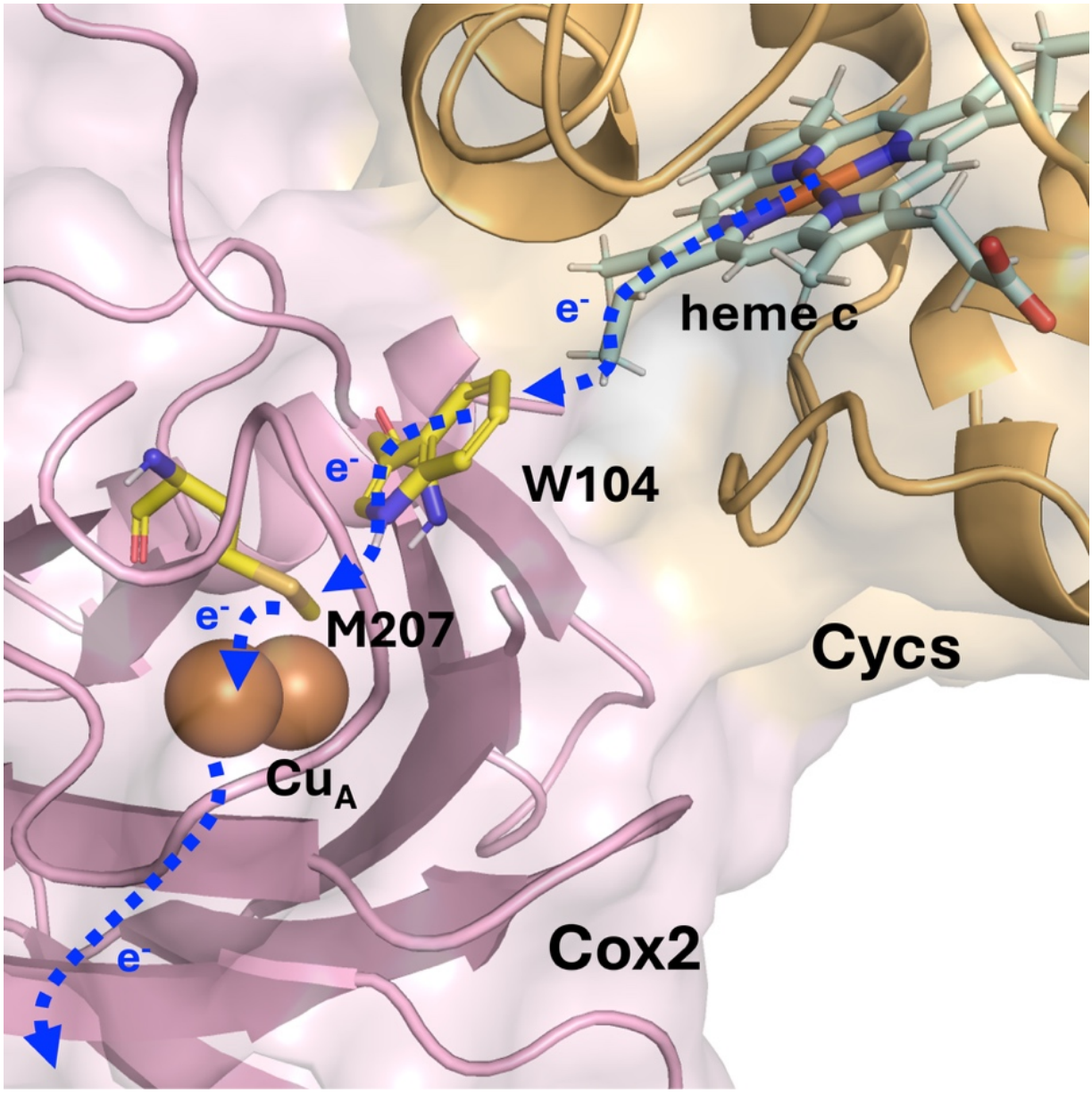


Figure 1. Electron transfer between *Cc* and *CcO*.

Electrons (blue arrows) travel between heme c in *Cc* to W104 in Cox2 of *CcO*, then pass to M207, then the dinuclear copper center, Cu_A, then continue through Cox2. PDB models 1HRC (horse Cycs) and 1OCC (bovine cytochrome c oxidase) were used to model the ET process as previously described (Roberts & Pique, 1999; Scharlau et al, 2019).

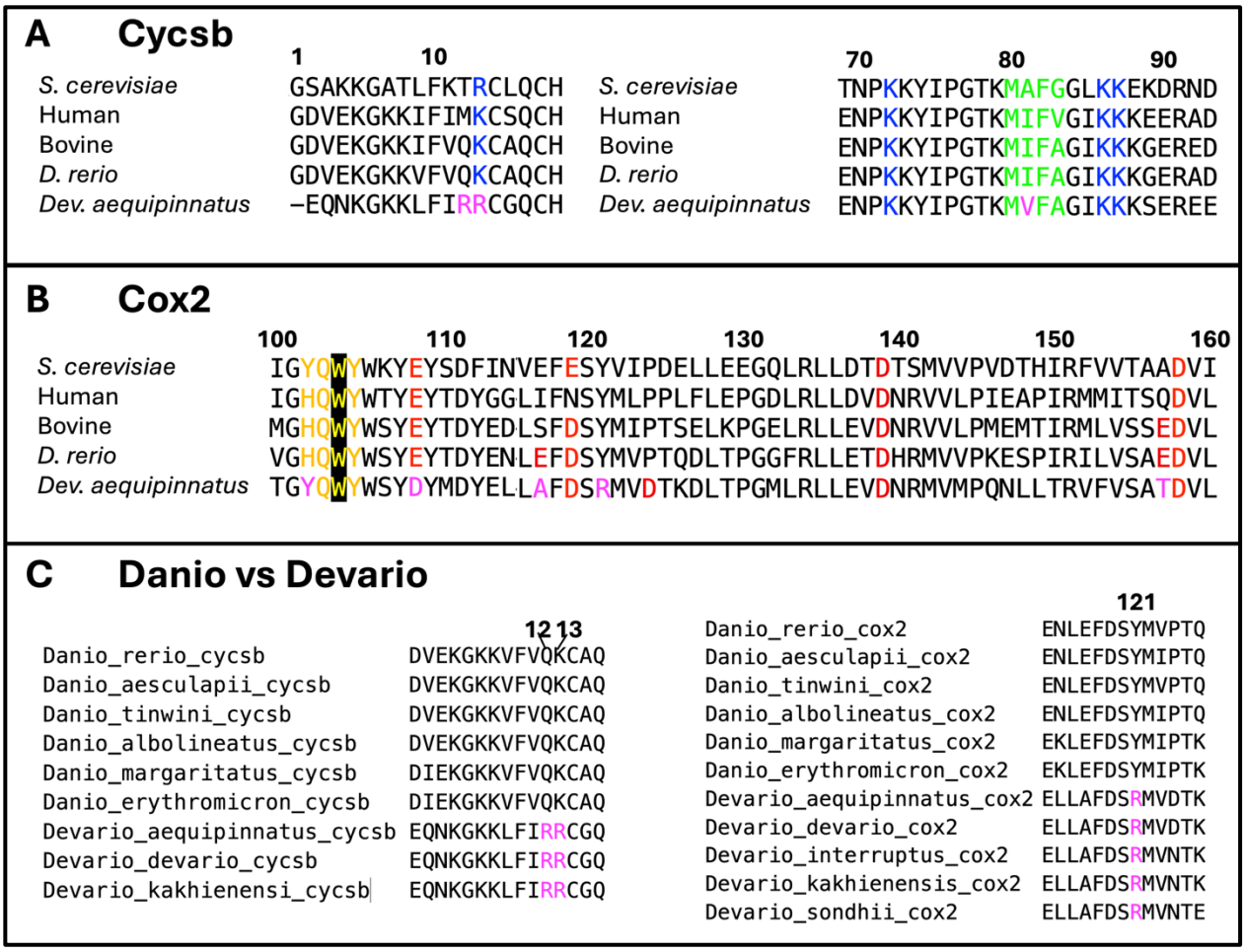


Figure 2. Multiple sequence alignments of Cycsb and Cox2 protein sequences

A. An alignment of Cycsb protein sequence that includes partial *S. cerevisiae,* Human, Bovine, *Dev. aequipinnatus*, and *D. rerio* sequences focusing on sites relative to this study. Blue sites are basic residues, green sites form a B-strand involved in van der Waals interactions with the hydrophobic patch shown orange in Cox2, and magenta sites are substitutions from *D. rerio* to Dev. *aequipinnatus*. *B.* An alignment of Cox2 where red sites are acidic residues, orange is the hydrophobic patch mentioned in *A*, magenta sites are substitutions in *Dev. aequipinnatus*, and yellow on black is W104 which is the site of electron transfer. *C.* These are alignments in only *Danio* and *Devario* species that show the conservation of residues distinct to each genus in positions 12 and 13 of Cycsb and position 121 of Cox2.

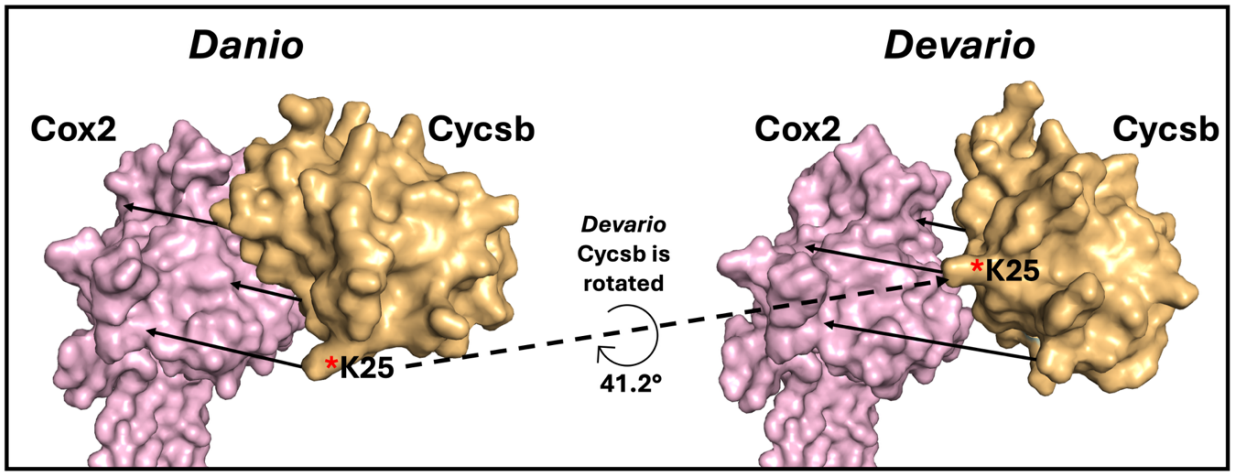


Figure 3. Cycsb to Cox2 docking orientations in *D. rerio* and *Dev. aequipinnatus*.

In the binding interface of Cycsb (bronze) to Cox2 (pink) in *Dev. aequipinnatus* (*Devario*), Cycsb approached and docked to Cox2 in a unique orientation in respect to Cycsb binding to Cox2 in *D. rerio* (*Danio*). The Cycsb protein binding to *Devario* Cox2 was rotated 41.2 degrees clockwise when comparing the angle of difference between K25 in *Danio* versus *Devario* docking scenarios. For cross-species binding conformations, see Supplemental Figure 1.

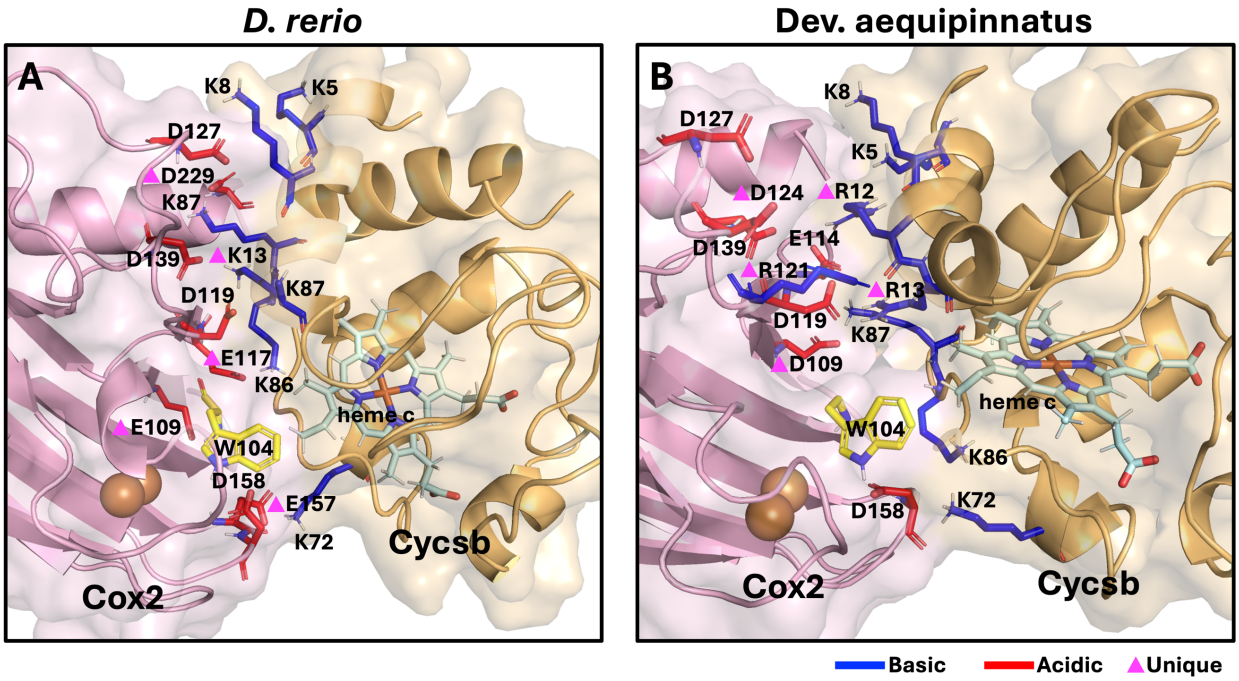


Figure 4. Polar basic and acidic residues in *D. rerio* versus *Dev. aequipinnatus* Cycsb and Cox2.

A. In *D. rerio* the canonical pattern of polar residues was largely conserved with lysine (blue) residues acting to make a positively charged basic surface in Cycsb which could pair with and bind to negatively charged acidic surface residues (red) of Cox2. Substitutions between *D. rerio* and *Dev. aequipinnatus* polar residues (magenta triangles) that may have contributed to a unique electrostatic state in the *D. rerio* site of ET and to its binding orientation were at Cycsb K13 and Cox2 E109, E117, and E157. *B.* Unique residues in *Dev. aequipinnatus* Cox2 included D109 and D124 in Cox2 with acidic residues notably absent from Cox2 positions 117 and 157 where they were found in *D. rerio*. The sites of most interest, however, were the *D. rerio* to *Dev. aequipinnatus* R12 and R13 substitutions in Cycsb which were a derivation from Q12 and K13 found in other eukaryotes, and the Y121R mutation which brought a positive charge to the surface of an otherwise negatively charged Cox2 binding domain.

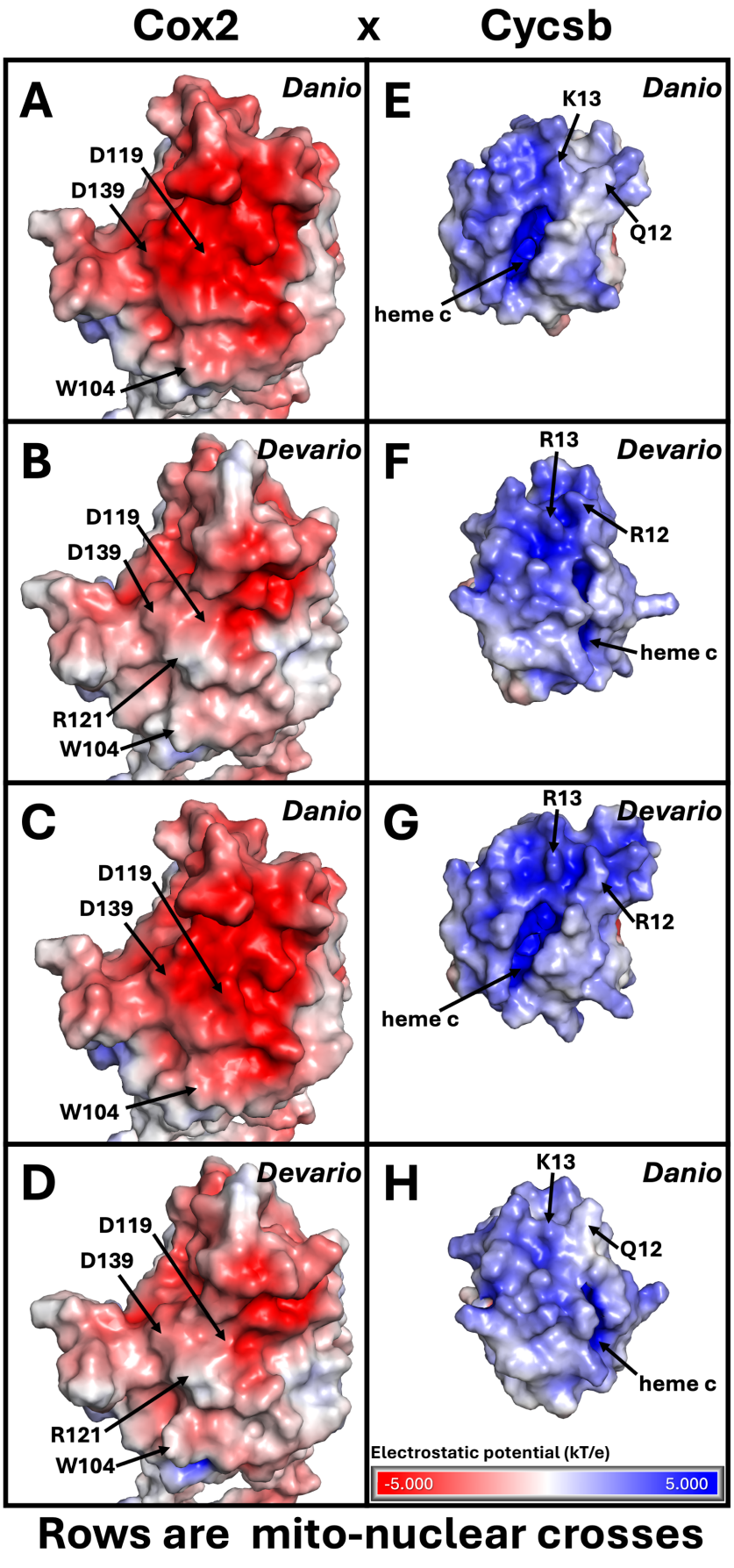

Cox2 x Cycsb
A
Danio
D119
D139
W104
E
Danio
K13
Q12
heme c
B
Devario
D119
D139
R121
W104
F
Devario
R13
R12
heme c
C
Danio
D119
D139
W104
G
Devario
R13
R12
heme c
D
Devario
D119
D139
R121
W104
H
Danio
K13
Q12
heme c
Electrostatic potential (kT/e)
-5.000
5.000
Rows are mito-nuclear crosses

Figure 5. Electrostatic potential mapping of Cox2 and Cycsb in various mito-nuclear crosses between *D. rerio* and *Dev. aequipinnatus*.

A–D. Electrostatic potential in the docking surface of Cox2 in *D. rerio* (*Danio*) or *Dev. aequipinnatus* (*Devario*) when binding to *Danio* or *Devario* Cycsb in the adjacent panels *E–H* with species as indicated. When paired with results in Figure 6, decreases in the negative charge at R121 in *Devario* Cox2, apparent by the white mapping on the protein's surface (*B*, *D*) were highlighted. D119 and D139 in panels *A–D* were labeled to demonstrate the residues responsible for variable H-bonds and salt bridges to Cycsb position 13 when docking with *Danio* Cox2 but were formed with R121 in scenarios with *Devario* Cox2. W104 was labeled to indicate the site where ET occurs. *E–H*. Positions 12 and 13 were labeled and showed differences between the proteoforms which included Q12 and K13 or R12 and R13. In the *Danio,* residues Q12 and K13 of Cycsb (*E*, *H*) bifurcated near the top of the opening for heme c access during electron transfer. Additionally, K13 appeared deeper blue indicating it carries a positive charge while Q12 remained near neutral and appeared gray. In *Devario* Cycsb (F, G) R12 and R13 both appeared as deep blue indicating positive charges and created a comparatively deep furrow between them to form a concentrated basic site for strong polar bonds with Cox2.

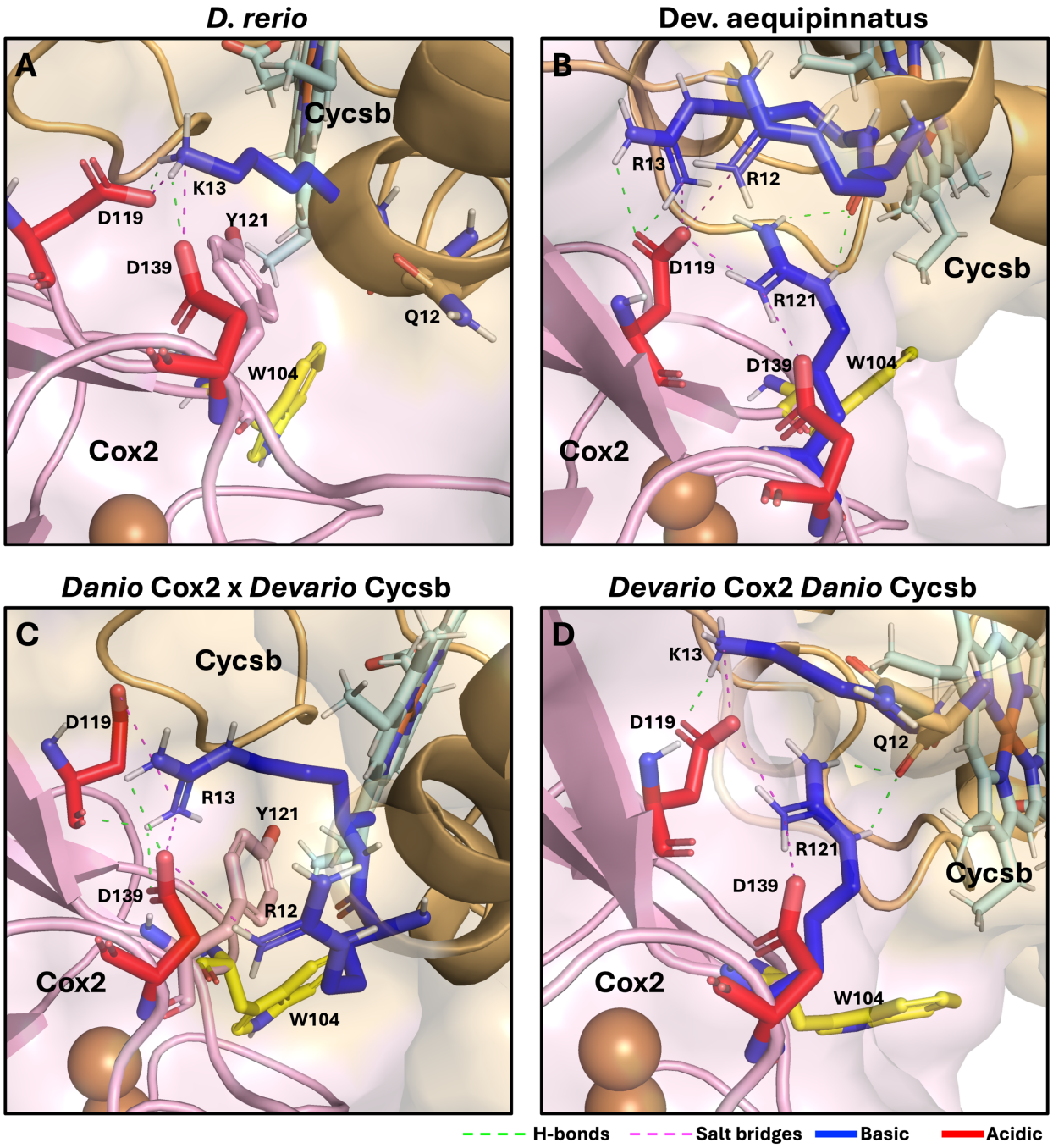


Figure 6. The effects on H-bonds and salt bridges of Cycsb Q12R, K13R, and Cox2 Y121R substitutions.

A. In *D. rerio*, D119 and D139 of Cox2 formed H-bonds (green dashes) and salt bridges (magenta dashes) with the terminal amino group of Cycsb K13 while Cox2 Y121 and Cycsb Q12

seemingly made no contribution to polar bonds in their main chains or side chains. *B*. In *Dev. aequipinnatus*, Cox2 D119 formed H-bonds and salt bridges with the Cycsb R13 guanidinium group in a location further from Cox2 D139 due to the longer R13 side chain compared to K13 in *D. rerio*. D119 also formed salt bridges with the side chains of R12 and R121 while D139 created an additional salt bridge with the side chain of R121 but did not form a polar bond with Cycsb directly. Instead, Cox2 R121 formed H-bonds with the main chain of Cycsb R12 to span the gap. *C*. When *D. rerio* (*Danio*) Cox2 bound to *Dev. aequipinnatus* (*Devario*) Cycsb, Cox2 D119 and D139 formed H-bonds and salt bridges with Cycsb R13, but there was an additional salt bridge formed between D139 and Cycsb R12 which happened near W104 and may have pulled Cycsb out of alignment between heme c and W104. In addition, Y121 formed no polar bonds in the docking situation. *D.* The *Devario* Cox2 to *Danio* Cycsb scenario had Cox2 D119 forming H-bonds and salt bridges with Cycsb K13 and both D119 and D139 formed salt bridges with R121. R121 then spanned the gap and formed H-bonds with the main chain of Cycsb Q12 similar to the binding profile in the *Devario-Devario* docking scenario.

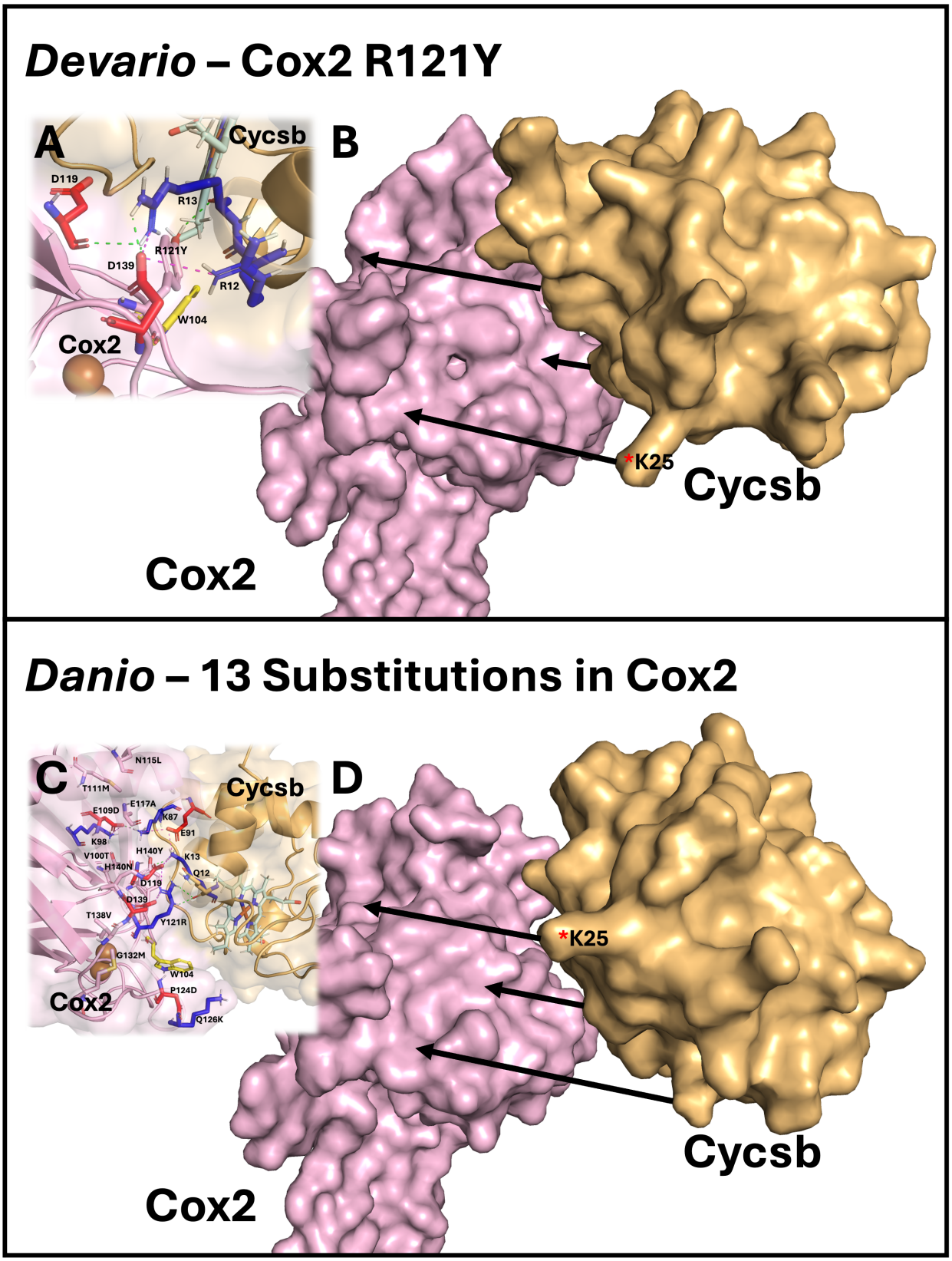
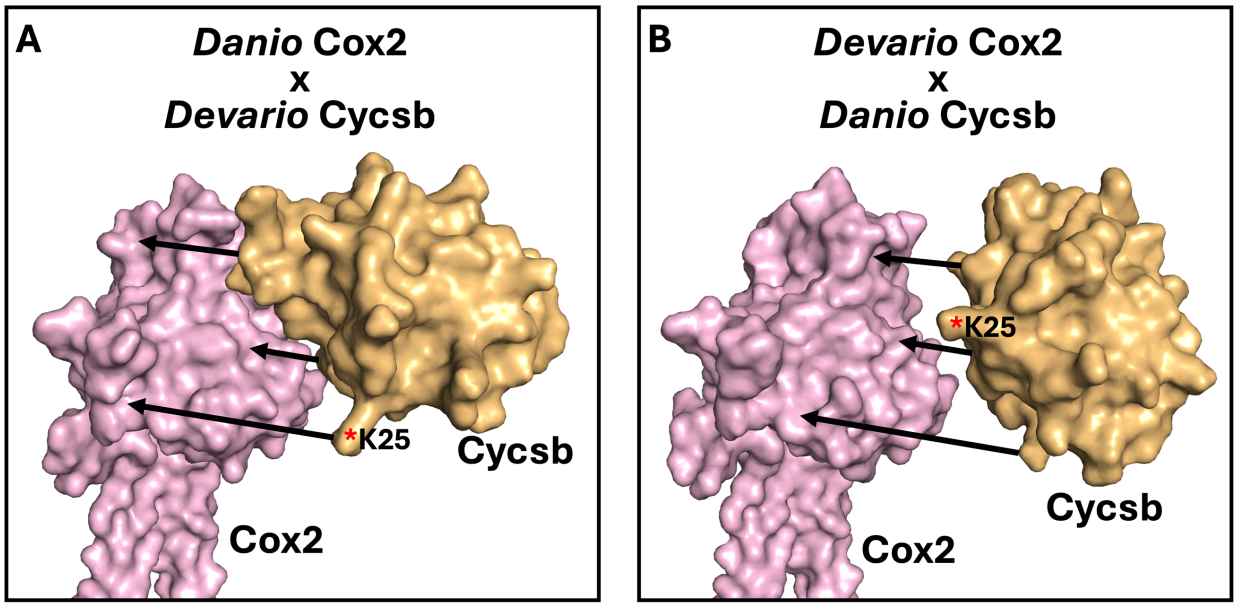


Figure 7. Simulating mutations in Cox2 can alter Cycsb-Cox2 docking orientations.

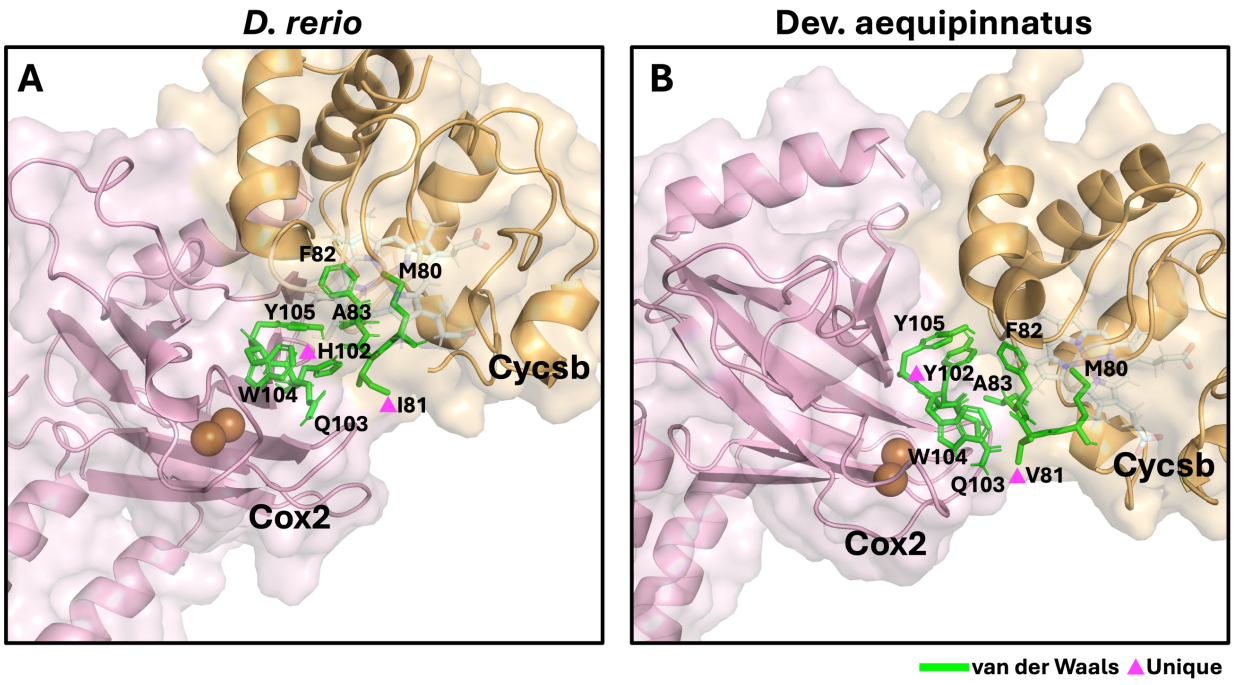
Top panel. *Dev. aequipinnatus* (*Devario*) The Cox2 R121Y mutation prevented Cycsb rotation in docking simulations. *A*. A simulated R121Y mutation stopped Cox2 D119 and D139 from forming H-bonds (green dashes) and salt bridges (magenta dashes) with Cox2 position 121. D119 formed H-bonds with Cycsb R13 and D139 formed H-bonds and a salt bridge directly with Cycsb R13 and made a salt bridge with the R12 side chain as well. Finally, a H-bond was also formed between Y121 and the main chain of R13. *B.* The single R121Y mutation changed the docking conformation in *Devario* to resemble *D. rerio* (*Danio*) and was not rotated relative to the *D. rerio* docking orientation. *C–D.* Thirteen mutations were simulated in *D. rerio* including V100T, H102Y, E109D, T111M, N115L, E117A, Y121R, P124D, Q126K, G132M, T138V, H140N, and E157T. The mutated sites and changes in polar bonds are shown in *C* where Cox2 R121 formed H-bonds with the main chain of Cycsb Q12 which spanned the interprotein gap and D119 and D139 formed salt bridges again with R121. D119 also made a salt bridge with Cycsb K13 which participated in a chain of H-bonds and salt bridges that included Cycsb E91, K87, Cox2 D109, and K98.

3.8 Supplemental materials



Supplemental Figure 1. Docking orientations of Cycsb in cross-species interactions.

A. Simulated docking between *D. rerio* (*Danio*) Cox2 (pink) and *Dev. aequipinnatus* (*Devario*) Cycsb (bronze). In this scenario, the *Devario* Cycsb matched docking orientations with *Danio-Danio* Cycsb orientation from Figure 3A. *B. Devario* Cox2 to *Danio* Cycsb docking simulation where *Danio* Cycsb matched the rotation of *Devario* Cycsb when paired to *Devario* Cox2 as in Figure 3B. K25 is labeled as a point of reference in both panels.



Supplemental Figure 2. Hydrophobic loop and β-strand comparison in parallel domains that produce van der Waals interactions.

A. The hydrophobic loop in *D. rerio* Cox2, which included H102, Q103, W104, and Y105, parallel to the β-strand of Cycsb comprising M80, I81, F82, and A83. *B.* The same domains in *Dev. aequipinnatus* which included Y102, Q103, W104, and Y105 of the Cox2 hydrophobic loop and the β-strand including M80, V81, F82, and A83. These residues are all represented as green sticks and the residues labeled with magenta triangles represent substitutions between the two species.

CHAPTER 4

Conclusion

4.1 Punctuated variation among danionins. In theories addressing the origins of mito-nuclear incompatibilities, mitochondrial genome (mtDNA) mutations accumulate over time with compensatory mutations forming in interacting proteins from the nuclear genome (nDNA). I examined the patterns of variant accumulation in mitochondrial-encoded protein-coding genes across twelve species of the Danionin lineage and found that the respiratory complex I protein, Nd5, had a rapid rate of evolution and accumulated many SAVs in sequence comparisons to other *Danio* species. Similarly, in *Devario* and *Microdevario* species, proteins from respiratory complex IV had high evolutionary rates with Cox2 exhibiting significantly more SAVs compared to other mitochondrial-encoded subunits. Taken together, these sites of heightened substitution indicated that nuclear-encoded subunits had also likely undergone rapid evolution. In my efforts to identify compatibility boundaries across Danionins species, I was successful in identifying *D. albolineatus* Nd5 and *Devario* cox proteins as candidates likely to contribute to mito-nuclear incompatibilities since the nuclear-encoded subunits that interact with these proteins have a high chance to have elevated substitution rates corresponding to coevolution. This means that in future studies, I will work to characterize the mitochondrial activity in hybrids between *Devario* species and *D. rerio* or *D. albolineatus* and *D. rerio*, starting with *Danio rerio* x *Devario aequipinnatus* hybrids. This future direction should include measuring respiration rates and membrane potentials to characterize whether mitochondrial dysfunction exists in these hybrids. Additionally, it is worth following up on my mutation rate results in work I performed using Bayesian estimates, and want to use some of the sequences that are now becoming available through sequence read databases to examine evolution rates in nuclear-encoded subunits

suspected to coevolve with *Devario* cox proteins or *D. albolineatus* Nd5 in which I want to identify more candidate proteins and sites for further study.

4.2 Devario Cox2-Cycsb docking orientation. In *Devario aequipinnatus* Cox2 I noticed the increase of variants was centralized to the binding domain between Cycsb and Cox2 where the final point of electron transfer occurs in respiratory complex IV's involvement with the electron transport chain. I was able to simulate docking between Cycsb and Cox2 in predicted models and found that Cycsb assumed a novel docking orientation when paired with Cox2 in *Devario aequipinnatus* in which it rotated clockwise relative to *Danio rerio* binding scenarios. I further examined residues R121 in *Devario aequipinnatus* Cox2 and R12 and R13 in Cycsb and determined the R121 residue probably hinders access of Cycsb to the electron transfer site in Cox2. Additionally, the R12 and R13 residues are compensatory and act to circumvent R121 and maintain a novel docking orientation. However, in the hybrid scenario between *Danio rerio* Cox2 and *Devario aequipinnatus* Cysb, R12 and R13 likely change the angle of approach in Cycsb and form additional inaccurate bonds that prevent Cycsb from reaching its destination in Cox2 to transfer electrons. In future work, I plan to attempt a genetic rescue of the *Danio rerio* x *Devario aequipinnatus* hybrids by overexpressing *Danio rerio* Cycsb in embryos to determine if that gene is the point of hybrid incompatibility alone. This will be paired with the mitochondrial activity assays described above and the results from both approaches will better characterize to what extent mito-nuclear incompatibilities contribute to hybrid lethality between these species.

References

Adkins, R. M., & Honeycutt, R. L. (1994). Evolution of the primate cytochrome c oxidase subunit II gene. *Journal of Molecular Evolution*, 38(3), 215–231. DOI: 10.1007/BF00176084

Aledo, J. C., Valverde, H., Ruíz-Camacho, M., Morilla, I., & López, F. D. (2014). Protein–protein interfaces from cytochrome c oxidase I evolve faster than nonbinding surfaces, yet negative selection is the driving force. *Genome Biology and Evolution*, 6(11), 3064–3076. DOI: 10.1093/gbe/evu240

Allen, J. M., Huang, D. I., Cronk, Q. C., & Johnson, K. P. (2015). aTRAM - automated target restricted assembly method: A fast method for assembling loci across divergent taxa from next-generation sequencing data. *BMC Bioinformatics*, 16(1), 98. DOI: 10.1186/s12859-015-0515-2

Ancrenaz, M., Lackman-ancrenaz, I., & Andau, P. (2006). Genetic signature of anthropogenic population collapse in orangutans. *PLoS Biology*, 4(2), 285–291. DOI: 10.1371/journal.pbio.0040025

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Attardi, G., & Schatz, G. (1988). Biogenesis of mitochondria. *Annual Review of Cell Biology*, 4, 289–333. DOI: 10.1146/annurev.cb.04.110188.001445

Baden, K. N., Murray, J., Capaldi, R. A., & Guillemin, K. (2007). Early developmental pathology due to cytochrome c oxidase deficiency is revealed by a new zebrafish model. *The Journal of Biological Chemistry*, 282(48), 34839–34849. DOI: 10.1074/jbc.M703528200

Ballard, J. W. O., Melvin, R. G., Katewa, S. D., & Maas, K. (2007). Mitochondrial DNA variation is associated with measurable differences in life-history traits and mitochondrial metabolism in *Drosophila simulans*. *Evolution*, 61(7), 1735–1747. DOI: 10.1111/j.1558-5646.2007.00133.x

Bar-Yaacov, D., Hadjivasiliou, Z., Levin, L., Barshad, G., Zarivach, R., Bouskila, A., & Mishmar, D. (2015). Mitochondrial involvement in vertebrate speciation? The Case of mito-nuclear genetic divergence in chameleons. *Genome Biology and Evolution*, 7(12), 3322–3336. DOI: 10.1093/gbe/evv226

Barreto, F. S., & Burton, R. S. (2013). Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA. *Molecular Biology and Evolution*, 30(2), 310–314. DOI: 10.1093/molbev/mss228

Barreto, F. S., Watson, E. T., Lima, T. G., Willett, C. S., Edmands, S., Li, W., & Burton, R. S. (2018). Genomic signatures of mitonuclear coevolution across populations of *Tigriopus*

californicus. *Nature Ecology & Evolution*, 2(8), 1250–1257. DOI: 10.1038/s41559-018-0588-1

Barrientos, A., Kenyon, L., & Moraes, C. T. (1998). Human xenomitochondrial cybrids. Cellular models of mitochondrial complex I deficiency. *The Journal of Biological Chemistry*, 273(23), 14210–14217. DOI: 10.1074/JBC.273.23.14210

Bashir, Q., Meulenbroek, E. M., Pannu, N. S., & Ubbink, M. (2014). Engineering specificity in a dynamic protein complex with a single conserved mutation. *The FEBS Journal*, 281(21), 4892–4905. DOI: 10.1111/febs.13028

Bateson, W. (2009). Heredity and variation in modern lights. In A. C. Seward (ed.), *Cambridge Library Collection - Darwin, Evolution and Genetics* (pp. 85-101). Cambridge University Press. ISBN: 9781108004350

Bayona-Bafaluy, M. P., Müller, S., & Moraes, C. T. (2005). Fast adaptive coevolution of nuclear and mitochondrial subunits of ATP synthetase in orangutan. *Molecular Biology and Evolution*, 22(3), 716–724. DOI: 10.1093/molbev/msi059

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235–242. DOI: 10.1093/nar/28.1.235

Bleeker, P. 1863. Systema cyprinoideorum revisum. *Nederlandsch Tijdschrift voor de Dierkund*, 1: 187–218.

Bolnick, D. I., Turelli, M., López-Fernández, H., Wainwright, P. C., & Near, T. J. (2008). Accelerated mitochondrial evolution and "Darwin's corollary": Asymmetric viability of reciprocal F1 hybrids in centrarchid fishes. *Genetics*, 178(2), 1037–1048. DOI: 10.1534/genetics.107.081364

Bouckaert, R. R., & Drummond, A. J. (2017). bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology*, 17(1), 42. DOI: 10.1186/s12862-017-0890-6

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., Maio, N. D., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., Plessis, L. du, Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., … Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, 15(4), e1006650. DOI: 10.1371/journal.pcbi.1006650

Brand, M., Granato, M., & Nüsslein-Volhard, C. (2002). Keeping and raising zebrafish. *Zebrafish*, 261,7–37.

Brand, S. E., Scharlau, M., Geren, L., Hendrix, M., Parson, C., Elmendorf, T., Neel, E., Pianalto, K., Silva-Nash, J., Durham, B., & Millett, F. (2022). Accelerated evolution of cytochrome

c in higher primates, and regulation of the reaction between cytochrome c and cytochrome oxidase by phosphorylation. *Cells*, 11(24). DOI: 10.3390/cells11244014

Breeuwer, J. A. J., & Werren, J. H. (1995). Hybrid breakdown between two haplodiploid species: The role of nuclear and cytoplasmic genes. *Evolution*, 49(4), 705–717. DOI: 10.2307/2410324

Burton, R. S. (2022). The role of mitonuclear incompatibilities in allopatric speciation. *Cellular and Molecular Life Sciences: CMLS*, 79(2), 103. DOI: 10.1007/s00018-021-04059-3

Burton, R. S., & Barreto, F. S. (2012). A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Molecular Ecology*, 21(20), 4942–4957. DOI: 10.1111/mec.12006

Bush, R. M., Fitch, W. M., Bender, C. A., & Cox, N. J. (1999). Positive selection on the H3 hemagglutinin gene of human influenza virus A. *Molecular Biology and Evolution*, 16(11), 1457–1465. DOI: 10.1093/oxfordjournals.molbev.a026057

Caballero, S., Duchêne, S., Garavito, M. F., Slikas, B., & Baker, C. S. (2015). Initial evidence for adaptive selection on the NADH subunit two of freshwater dolphins by analyses of mitochondrial genomes. *PLOS ONE*, 10(5), e0123543. DOI: 10.1371/journal.pone.0123543

Calvo, S. E., & Mootha, V. K. (2010). The mitochondrial proteome and human disease. *Annual Review of Genomics and Human Genetics*, 11(1), 25–44. DOI: 10.1146/annurev-genom-082509-141720

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10(1), 421. DOI: 10.1186/1471-2105-10-421

Cannon, B., & Nedergaard, J. (2004). Brown adipose tissue: Function and physiological significance. Physiological Reviews, 84(1), 277–359. DOI: 10.1152/physrev.00015.2003

Castellana, S., Vicario, S., & Saccone, C. (2011). Evolutionary patterns of the mitochondrial genome in Metazoa: Exploring the role of mutation and selection in mitochondrial protein–coding genes. *Genome Biology and Evolution*, 3, 1067–1079. DOI: 10.1093/gbe/evr040

Chamberlain, T. J., & Pelegri, F. (in review). Appearance of mitochondria genetic variants in the Danionin lineage is both gradual and punctuated. Submitted to *Genome Biology and Evolution*.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. DOI: 10.1093/bioinformatics/bty560

Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A. J., Müller, W. E. G., Wetter, T., & Suhai, S. (2004). Using the miraEST assembler for reliable and automated mRNA transcript

assembly and SNP detection in sequenced ESTs. *Genome Research*, 14(6), 1147–1159. DOI: 10.1101/gr.1917404

Corley-Smith, G. E., Lim, C. J., & Brandhorst, B. P. (1996). Production of androgenetic zebrafish (*Danio rerio*). *Genetics*, 142(4), 1265–1276. DOI: 10.1093/genetics/142.4.1265

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates.

da Fonseca, R. R., Johnson, W. E., O'Brien, S. J., Ramos, M. J., & Antunes, A. (2008). The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics*, 9(1), 119. DOI: 10.1186/1471-2164-9-119

Dames, S., Chou, L.-S., Xiao, Y., Wayman, T., Stocks, J., Singleton, M., Eilbeck, K., & Mao, R. (2013). The development of next-generation sequencing assays for the mitochondrial genome and 108 nuclear genes associated with mitochondrial disorders. *The Journal of Molecular Diagnostics*, 15(4), 526–534. DOI: 10.1016/j.jmoldx.2013.03.005

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., & Flouri, T. (2020). ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular Biology and Evolution*, 37(1), 291–294. DOI: 10.1093/molbev/msz189

de Beaufort, L. F. (1951). *Zoogeography of the land and Inland Waters*. Sidgwick and Jackson.

Delomas, T. A., & Dabrowski, K. (2022). Asymmetric viability in reciprocal crosses of zebrafish and pearl danio. *Journal of Fish Biology*, 100(1), 10–14. DOI: 10.1111/jfb.14911

Dey, R., Barrientos, A., & Moraes, C. T. (2000). Functional constraints of nuclear-mitochondrial DNA interactions in xenomitochondrial rodent cell lines. *Journal of Biological Chemistry*, 275(40), 31520–31527. DOI: 10.1074/JBC.M004053200

Divakaruni, A. S., & Brand, M. D. (2011). The regulation and physiology of mitochondrial proton leak. *Physiology*, 26(3), 192–205. DOI: 10.1152/physiol.00046.2010

Doan, J. W., R. Schmidt, T., Wildman, D. E., Uddin, M., Goldberg, A., Hüttemann, M., Goodman, M., Weiss, M. L., & Grossman, L. I. (2004). Coadaptive evolution in cytochrome c oxidase: 9 of 13 subunits show accelerated rates of nonsynonymous substitution in anthropoid primates. *Molecular Phylogenetics and Evolution*, 33(3), 944–950. DOI: 10.1016/j.ympev.2004.07.016

Dobzhansky T. (1936). Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics*, 21(2), 113–135. DOI: 10.1093/genetics/21.2.113

Efremov, R. G., & Sazanov, L. A. (2011). Structure of the membrane domain of respiratory complex I. *Nature*, 476(7361). DOI: 10.1038/nature10330

Ellison, C. K., & Burton, R. S. (2006). Disruption of mitochondrial function in interpopulation hybrids of *Tigriopus californicus*. *Evolution*, 60(7), 1382–1391. DOI: 10.1111/j.0014-3820.2006.tb01217.x

Ellison, C. K., & Burton, R. S. (2008). Interpopulation hybrid breakdown maps to the mitochondrial genome. *Evolution*, 62(3), 631–638. DOI: 10.1111/j.1558-5646.2007.00305.x

Garland T., Jr (2001). Phylogenetic comparison and artificial selection. Two approaches in evolutionary physiology. *Advances in Experimental Medicine and Biology*, 502, 107–132.

Gershoni, M., Fuchs, A., Shani, N., Fridman, Y., Corral-Debrinski, M., Aharoni, A., Frishman, D., & Mishmar, D. (2010). Coevolution predicts direct interactions between mtDNA-encoded and nDNA-encoded subunits of oxidative phosphorylation complex I. *Journal of Molecular Biology*, 404(1), 158–171. DOI: 10.1016/j.jmb.2010.09.029

Gibert, Y., Mcgee, S. L., & Ward, A. C. (2013). Metabolic profile analysis of zebrafish embryos. *Journal of Visualized Experiments*, 71(e4300), 1–5. DOI: 10.3791/4300

Gómez, M. C., Jenkins, J. A., Giraldo, A., Harris, R. F., King, A., Dresser, B. L., & Pope, C. E. (2003). Nuclear transfer of synchronized African wild cat somatic cells into enucleated domestic cat oocytes. *Biology of Reproduction*, 69(3), 1032–1041. DOI: 10.1095/biolreprod.102.014449

Gómez, M. C., Pope, C. E., Giraldo, A., Lyons, L. A., Harris, R. F., King, A. M. Y. L., Cole, A., Godke, R. A., & Dresser, B. L. (2004). Birth of African wildcat cloned kittens born from domestic cats. *Cloning and Stem Cells*, 6(3), 247–258.

Gonzalez-Hunt, C. P., Rooney, J. P., Ryde, I. T., Anbalagan, C., Joglekar, R., & Meyer, J. N. (2016). PCR-based analysis of mitochondrial DNA copy number, mitochondrial DNA damage, and nuclear DNA damage. *Current Protocols in Toxicology*, 67, 20.11.1-20.11.25. DOI: 10.1002/0471140856.tx2011s67

Gray, M. W., Burger, G., & Lang, B. F. (2001). The origin and early evolution of mitochondria. *Genome Biology*, 2(6), reviews1018.1. DOI: 10.1186/gb-2001-2-6-reviews1018

Green, D. R., & Reed, J. C. (1998). Mitochondria and Apoptosis. *Science*, 281(5381), 1309–1312. DOI: 10.1126/science.281.5381.1309

Grossman, L. I., Schmidt, T. R., Wildman, D. E., & Goodman, M. (2001). Molecular evolution of aerobic energy metabolism in primates. *Molecular Phylogenetics and Evolution*, 18(1), 26–36. DOI: 10.1006/mpev.2000.0890

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—A baiting and iterative mapping approach. *Nucleic Acids Research*, 41(13), e129. DOI: 10.1093/nar/gkt371

Hansen, C. L., Chamberlain, T. J., Trevena, R. L., Kurek, J. E., & Pelegri, F. (2021). Conserved germ plasm characteristics across the *Danio* and *Devario* lineages. *Genesis*, 59(10), e23452. DOI: 10.1002/DVG.23452

Harrison, J. S., & Burton, R. S. (2006). Tracing hybrid incompatibilities to single amino acid substitutions. *Molecular Biology and Evolution*, 23(3), 559–564. DOI: 10.1093/molbev/msj058

Havird, J. C., & Sloan, D. B. (2016). The roles of mutation, selection, and expression in determining relative rates of evolution in mitochondrial versus nuclear genomes. *Molecular Biology and Evolution*, 33(12), 3042–3053. DOI: 10.1093/molbev/msw185

Healy, T. M., & Burton, R. S. (2020). Strong selective effects of mitochondrial DNA on the nuclear genome. *Proceedings of the National Academy of Sciences of the United States of America*, 117(12), 6616–6621. DOI: 10.1073/pnas.1910141117

Henderson, A. R. (1969). Biochemistry of hypoxia: Current concepts I: An introduction to biochemical pathways and their control. *British Journal of Anaesthesia*, 41(3), 245–250. DOI: 10.1093/bja/41.3.245

Hill, G. E. (2017). The mitonuclear compatibility species concept. *The Auk*, 1341(134), 393–409. DOI: 10.1642/AUK-16-201.1

Hill, G. E. (2020). Mitonuclear compensatory coevolution. *Trends in Genetics*, 36(6), 403–414. DOI: 10.1016/j.tig.2020.03.002

Hirano, M. (2001). Transmitochondrial mice: Proof of principle and promises. *Proceedings of the National Academy of Sciences*, 98(2), 401–403. DOI: 10.1073/pnas.98.2.401

Honorato, R. V., Koukos, P. I., Jiménez-García, B., Tsaregorodtsev, A., Verlato, M., Giachetti, A., Rosato, A., & Bonvin, A. M. J. J. (2021). Structural biology in the clouds: The WeNMR-EOSC ecosystem. *Frontiers in Molecular Biosciences*, 8. DOI: 10.3389/fmolb.2021.729513

Honorato, R. V., Trellet, M. E., Jiménez-García, B., Schaarschmidt, J. J., Giulini, M., Reys, V., Koukos, P. I., Rodrigues, J. P. G. L. M., Karaca, E., van Zundert, G. C. P., Roel-Touris, J., van Noort, C. W., Jandová, Z., Melquiond, A. S. J., & Bonvin, A. M. J. J. (2024). The HADDOCK2.4 web server for integrative modeling of biomolecular complexes. *Nature Protocols*, 19(11), 3219–3241. DOI: 10.1038/s41596-024-01011-0

Huang, X., Song, L., Zhan, Z., Gu, H., Feng, H., & Li, Y. (2017). Factors affecting mouse somatic cell nuclear reprogramming by rabbit ooplasms. *Cellular Reprogramming*, 19(6), 344–353. DOI: 10.1089/cell.2017.0021

Hunt, B. G., Ometto, L., Wurm, Y., Shoemaker, D., Yi, S. V., Keller, L., & Goodisman, M. A. D. (2011). Relaxed selection is a precursor to the evolution of phenotypic plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, 108(38), 15936–15941. DOI: 10.1073/pnas.1104825108

Hunte, C., Zickermann, V., & Brandt, U. (2010). Functional modules and structural basis of conformational coupling in mitochondrial complex I. *Science*, 329(5990), 448–451. DOI: 10.1126/science.1191046

Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T. P., Sado, T., Mabuchi, K., Takeshima, H., Miya, M., & Nishida, M. (2013). MitoFish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Molecular Biology and Evolution*, 30(11), 2531–2540. DOI: 10.1093/molbev/mst141

Jiang, M. X., Yang, C. X., Zhang, L. S., Zheng, Y. L., Liu, S. Z., Sun, Q. Y., & Chen, D. Y. (2004). The effects of chemical enucleation combined with whole cell intracytoplasmic injection on panda-rabbit interspecies nuclear transfer. *Zygote*, 12(4), 315–320. DOI: 10.1017/s0967199404002941

Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., Zerega, N. J. C., & Wickett, N. J. (2016). Hybpiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences*, 4(7), 1600016. DOI: 10.3732/apps.1600016

Johnson, S. C., Gonzalez, B., Zhang, Q., Milholland, B., Zhang, Z., & Suh, Y. (2017). Network analysis of mitonuclear GWAS reveals functional networks and tissue expression profiles of disease - associated genes. *Human Genetics*, 136(1), 55–65. DOI: 10.1007/s00439-016-1736-9

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. DOI: 10.1038/s41586-021-03819-2

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., … Baker, N. A. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein Science*, 27(1), 112–128. DOI: 10.1002/pro.3280

Kadenbach, B. (2021). Complex IV – The regulatory center of mitochondrial oxidative phosphorylation. *Mitochondrion*, 58, 296–302. DOI: 10.1016/j.mito.2020.10.004

Kenyon, L., & Maraes, C. T. (1997). Expanding the functional human mitochondrial DNA database by the establishment of primate xenomitochondrial cybrids. *Proceedings of the National Academy of Sciences of the United States of America*, 94(17), 9131–9135.

Kierans, S. J., & Taylor, C. T. (2021). Regulation of glycolysis by the hypoxia-inducible factor (HIF): Implications for cellular physiology. *The Journal of Physiology*, 599(1), 23–37. DOI: 10.1113/JP280572

King, M. P., & Attardi, G. (1996). [27] Isolation of human cell lines lacking mitochondrial DNA. *Methods in Enzymology*, 264, 304–313. DOI: 10.1016/S0076-6879(96)64029-4

Kliman, R. M., Andolfatto, P., Coyne, J. A., Depaulis, F., Kreitman, M., Berry, A. J., McCarter, J., Wakeley, J., & Hey, J. (2000). The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics*, 156(4), 1913–1931. DOI: 10.1093/genetics/156.4.1913

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35(21), 4453–4455. DOI: 10.1093/bioinformatics/btz305

Krzywinski, M. I., Schein, J. E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*. DOI: 10.1101/gr.092759.109

Lagutina, I., Fulka, H., Lazzari, G., & Galli, C. (2013). Interspecies somatic cell nuclear transfer: Advancements and problems. DOI: 10.1089/cell.2013.0036

Lahti, D. C., Johnson, N. A., Ajie, B. C., Otto, S. P., Hendry, A. P., Blumstein, D. T., Coss, R. G., Donohue, K., & Foster, S. A. (2009). Relaxed selection in the wild. *Trends in Ecology & Evolution*, 24(9), 487–496. DOI: 10.1016/j.tree.2009.03.010

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34(3), 772–773. DOI: 10.1093/molbev/msw260

Lange, C., & Hunte, C. (2002). Crystal structure of the yeast cytochrome bc1 complex with its bound substrate cytochrome c. *Proceedings of the National Academy of Sciences*, 99(5), 2800–2805. DOI: 10.1073/pnas.052704699

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), 2947–2948. DOI: 10.1093/bioinformatics/btm404

Lee, E., Bhuiyan, M. M., Watanabe, H., Matsuoka, K., Fujise, Y., Ishikawa, H., & Fukui, Y. (2009). Production of cloned sei whale (*Balaenoptera borealis*) embryos by interspecies somatic cell nuclear transfer using enucleated pig oocytes. *Journal of Veterinary Science*, 10(4), 285–292. DOI: 10.4142/JVS.2009.10.4.285

Leinonen, R., Sugawara, H., & Shumway, M. (2011). The Sequence Read Archive. Nucleic Acids Research, 39(Database issue), D19–D21. DOI: 10.1093/nar/gkq1019

Levin, L., Blumberg, A., Barshad, G., & Mishmar, D. (2014). Mito-nuclear co-evolution: The positive and negative sides of functional ancient mutations. *Frontiers in Genetics*, 5. DOI: 10.3389/fgene.2014.00448

Liu, X., Blomme, J., Bogaert, K. A., D'hondt, S., Coulembier Vandelannoote, E., Wichard, T., & De Clerck, O. (2025). Positive selection and relaxed purifying selection contribute to rapid evolution of sex-biased genes in green seaweed Ulva. *BMC Ecology and Evolution*, 25(1), 44. DOI: 10.1186/s12862-025-02382-y

Loi, P., Ptak, G., Barboni, B., Fulka, J., Cappai, P., & Clinton, M. (2001). Genetic rescue of an endangered mammal by cross-species nuclear transfer using post-mortem somatic cells. *Nature Biotechnology*, 19(10), 962–964.

Lynch, M., & Blanchard, J. L. (1998). Deleterious mutation accumulation in organelle genomes. *Genetica*, 102(0), 29–39. DOI: 10.1023/A:1017022522486

Ma, Y.-Y., Zhang, X.-L., Wu, T.-F., Liu, Y.-P., Wang, Q., Zhang, Y., Song, J.-Q., Wang, Y.-J., & Yang, Y.-L. (2011). Analysis of the mitochondrial complex I–V enzyme activities of peripheral leukocytes in oxidative phosphorylation disorders. *Journal of Child Neurology*, 26(8), 974–979. DOI: 10.1177/0883073811399905

Maddison, W. P., & FitzJohn, R. G. (2015). The unsolved challenge to phylogenetic correlation tests for categorical characters. *Systematic Biology*, 64(1), 127–136. DOI: 10.1093/sysbio/syu070

McBride, H. M., Neuspiel, M., & Wasiak, S. (2006). Mitochondria: More than just a powerhouse. *Current Biology*, 16(14), R551–R560. DOI: 10.1016/j.cub.2006.06.054

McCluskey, B. M., & Postlethwait, J. H. (2014). Phylogeny of zebrafish, a "model species," within *Danio*, a "model genus." *Molecular Biology and Evolution*, 32(3), 635–652. DOI: 10.1093/molbev/msu325

McCluskey, B. M., Batzel, P., & Postlethwait, J. H. (2024). The hybrid history of zebrafish. *G3: Genes, Genomes, Genetics*, 15(2), jkae299. DOI: 10.1093/g3journal/jkae299

McConnell, S. K. J. (2004). Mapping aquatic faunal exchanges across the Sunda shelf, South-East Asia, using distributional and genetic data sets from the cyprinid fish *Barbodes gonionotus* (Bleeker, 1850). *Journal of Natural History*, 38(5), 651–670. DOI: 10.1080/002229302100003629

McKenzie, M., & Trounce, I. (2000). Expression of Rattus norvegicus mtDNA in *Mus musculus* cells results in multiple respiratory chain defects. *Journal of Biological Chemistry*, 275(40), 31514–31519. DOI: 10.1074/jbc.M004070200

Meiklejohn, C. D., Holmbeck, M. A., Siddiq, M. A., Abt, D. N., Rand, D. M., & Montooth, K. L. (2013). An Incompatibility between a mitochondrial tRNA and its nuclear-encoded tRNA synthetase compromises development and fitness in *Drosophila*. *PLOS Genetics*, 9(1), e1003238. DOI: 10.1371/journal.pgen.1003238

Millar, R. B., & Anderson, M. J. (2004). Remedies for pseudoreplication. *Fisheries Research*, 70(2), 397–407. DOI: 10.1016/j.fishres.2004.08.016

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530–1534. DOI: 10.1093/molbev/msaa015

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: Making protein folding accessible to all. *Nature Methods*, 19(6), 679–682. DOI: 10.1038/s41592-022-01488-1

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. DOI: 10.1016/j.gene.2006.03.015

Mitchell, P. (1961). Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature*, 191(4784), 144–148. DOI: 10.1038/191144a0

Mitchell, P., & Moyle, J. (1967). Chemiosmotic hypothesis of oxidative phosphorylation. *Nature*. 213, 137–139. DOI: 10.1038/213137a0

Modi, A., Vai, S., Caramelli, D., & Lari, M. (2021). *The Illumina sequencing protocol and the NovaSeq 6000 system*. Methods in Molecular Biology (Clifton, N. J.), 2242, 15–42. DOI: 10.1007/978-1-0716-1099-2_2

Moulavi, F., Hosseini, S. M., Tanhaie-Vash, N., Ostadhosseini, S., Hosseini, S. H., Hajinasrollah, M., Asghari, M. H., Gourabi, H., Shahverdi, A., Vosough, A. D., & Nasr-Esfahani, M. H. (2017). Interspecies somatic cell nuclear transfer in Asiatic cheetah using nuclei derived from post-mortem frozen tissue in absence of cryo-protectant and in vitro matured domestic cat oocytes. *Theriogenology*, 90, 197–203. DOI: 10.1016/J.THERIOGENOLOGY.2016.11.023

Muller, H. J. (1942). Isolating mechanisms, evolution and temperature. *Biological Symposia*, 6, 71–125.

Müller, J., Houben, N., & Pauly, D. (2023). On being the wrong size, or the role of body mass in fish kills and hypoxia exposure. *Environmental Biology of Fishes*, 106(7), 1651–1667. DOI: 10.1007/s10641-023-01442-w

Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A., Eren, K., Pollner, T., Martin, D. P., Smith, D. M., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Gene-wide identification of episodic selection. *Molecular Biology and Evolution*, 32(5), 1365–1371. DOI: 10.1093/molbev/msv035

Na Ayutthaya, P. P., Lundberg, D., Weigel, D., & Li, L. (2020). Blue native polyacrylamide gel electrophoresis (BN-PAGE) for the analysis of protein oligomers in plants. *Current Protocols in Plant Biology*, 5(2), e20107. DOI: 10.1002/cppb.20107

Nabholz, B., Uwimana, N., & Lartillot, N. (2013). Reconstructing the phylogenetic history of long-term effective population size and life-history traits using patterns of amino acid

replacement in mitochondrial genomes of mammals and birds. *Genome Biology and Evolution*, 5(7), 1273–1290. DOI: 10.1093/gbe/evt083

Nakada, K., & Hayashi, J.-I. (2011). Transmitochondrial mice as models for mitochondrial DNA-based diseases. *Experimental Animals*, 60(5), 421–431. DOI: 10.1538/expanim.60.421

Neiman, M., & Taylor, D. R. (2009). The causes of mutation accumulation in mitochondrial genomes. *Proceedings of the Royal Society B: Biological Sciences*, 276(1660), 1201–1209. DOI: 10.1098/rspb.2008.1758

Nesci, S., Algieri, C., Trombetti, F., Fabbri, M., & Lenaz, G. (2023). Two separate pathways underlie NADH and succinate oxidation in swine heart mitochondria: Kinetic evidence on the mobile electron carriers. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1864(3), 148977. DOI: 10.1016/j.bbabio.2023.148977

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, 31(13), 3812–3814. DOI: 10.1093/nar/gkg509

Ngatia, J. N., Lan, T. M., Dinh, T. D., Zhang, L., Ahmed, A. K., & Xu, Y. C. (2019). Signals of positive selection in mitochondrial protein-coding genes of woolly mammoth: Adaptation to extreme environments? *Ecology and Evolution*, 9(12), 6821–6832. DOI: 10.1002/ece3.5250

Nikelski, E., & Weir, J. T. (2025). Genomic analysis suggests that mitonuclear coevolution proceeds over rapid timescales in the Amazonian Pipra Manakin complex. *Molecular Ecology*, 34(12), e17802. DOI: 10.1111/mec.17802

Nolfi-Donegan, D., Braganza, A., & Shiva, S. (2020). Mitochondrial electron transport chain: Oxidative phosphorylation, oxidant production, and methods of measurement. *Redox Biology*, 37, 101674. DOI: 10.1016/j.redox.2020.101674

Novak, B. J., Gober, P., Bortner, R., Garelle, D., Wright, M., Novak, J., Houck, M. L., Ryder, O. A., Milutinovich, D., Benavidez, J., Ryan, K., Walker, S., Arenivas, S. S., Aston, L., Russell, B., Marinari, P., Crosier, A., Helmick, K., Gibson, M. R., … Phelan, R. (2024a). First endangered black-footed ferrets, *Mustela nigripes*, cloned for genetic rescue. *bioRxiv*, DOI: 10.1101/2024.04.17.589896

Novak, B. J., Ryder, O. A., Houck, M. L., Putnam, A. S., Walker, K., Russell, L., Russell, B., Walker, S., Arenivas, S. S., Aston, L., Veneklasen, G., Ivy, J. A., Koepfli, K.-P., Rusnak, A., Simek, J., Zhuk, A., & Phelan, R. (2024b). Endangered Przewalski's horse, *Equus przewalskii*, cloned from historically cryopreserved cells. *bioRxiv*, DOI: 10.1101/2023.12.20.572538

Oh, H. J., Kim, M. K., Jang, G., Kim, H. J., Hong, S. G., Park, J. E., Park, K., Park, C., Sohn, S. H., Kim, D. Y., Shin, N. S., & Lee, B. C. (2008). Cloning endangered gray wolves (*Canis lupus*) from somatic cells collected postmortem. *Theriogenology*, 70(4), 638–647. DOI: 10.1016/j.theriogenology.2008.04.032

Okoye, C. N., Koren, S. A., & Wojtovich, A. P. (2023). Mitochondrial complex I ROS production and redox signaling in hypoxia. *Redox Biology*, 67, 102926. DOI: 10.1016/j.redox.2023.102926

Osada, N., & Akashi, H. (2012). Mitochondrial–nuclear interactions and accelerated compensatory evolution: Evidence from the primate cytochrome c oxidase complex. *Molecular Biology and Evolution*, 29(1), 337–346. DOI: 10.1093/molbev/msr211

Owusu-Ansah, E., & Banerjee, U. (2009). Reactive Oxygen Species prime *Drosophila* haematopoietic progenitors for differentiation. *Nature*, 461(7263), 537–541. DOI: 10.1038/nature08313

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. DOI: 10.1093/bioinformatics/bty633

Patananan, A. N., Wu, T.-H., Chiou, P.-Y., & Teitell, M. A. (2016). Modifying the mitochondrial genome. *Cell Metabolism*, 23(5), 785–796. DOI: 10.1016/j.cmet.2016.04.004

Pavlova, A., Gan, H. M., Lee, Y. P., Austin, C. M., Gilligan, D. M., Lintermans, M., & Sunnucks, P. (2017). Purifying selection and genetic drift shaped Pleistocene evolution of the mitochondrial genome in an endangered Australian freshwater fish. *Heredity*, 118(5), 466–476. DOI: 10.1038/hdy.2016.120

Picard, M., Hepple, R. T., & Burelle, Y. (2012). Mitochondrial functional specialization in glycolytic and oxidative muscle fibers: Tailoring the organelle for optimal function. *American Journal of Physiology-Cell Physiology*, 302(4), C629–C641. DOI: 10.1152/ajpcell.00368.2011

Pond, S. L. K., Frost, S. D. W., & Muse, S. V. (2005). HyPhy: Hypothesis testing using phylogenies. *Bioinformatics*, 21(5), 676–679. DOI: 10.1093/bioinformatics/bti079

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., & Korobeynikov, A. (2020). Using SPAdes de novo assembler. *Current Protocols in Bioinformatics*, 70(1), e102. DOI: 10.1002/cpbi.102

R Core Team. (2023). R: A language and environment for statistical computing [Software].

Rahn, J. J., Bestman, J. E., Stackley, K. D., & Chan, S. S. L. (2015). Zebrafish lacking functional DNA polymerase gamma survive to juvenile stage, despite rapid and sustained mitochondrial DNA depletion, altered energetics and growth. *Nucleic Acids Research*, 43(21), 10338–10352. DOI: 10.1093/nar/gkv1139

Raimondi, D., Orlando, G., Vranken, W. F., & Moreau, Y. (2019). Exploring the limitations of biophysical propensity scales coupled with machine learning for protein sequence analysis. *Scientific Reports*, 9(1), 16932. DOI: 10.1038/s41598-019-53324-w

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, 67(5), 901–904. DOI: 10.1093/sysbio/syy032

Rand, D. M. (2017). Fishing for adaptive epistasis using mitonuclear interactions. *PLOS Genetics*, 13(3), e1006662. DOI: 10.1371/journal.pgen.1006662

Rand, D. M., Clark, A. G., & Kann, L. M. (2001). Sexually antagonistic cytonuclear fitness interactions in Drosophila melanogaster. *Genetics*, 159(1), 173–187. DOI: 10.1093/genetics/159.1.173

Rawson, P. D., & Burton, R. S. (2002). Functional coadaptation between cytochrome c and cytochrome c oxidase within allopatric populations of a marine copepod. *Proceedings of the National Academy of Sciences*, 99(20), 12955–12958. DOI: 10.1073/pnas.202335899

Rawson, P. D., & Burton, R. S. (2006). Molecular evolution at the cytochrome oxidase subunit 2 gene among divergent populations of the intertidal copepod, *Tigriopus californicus*. *Journal of Molecular Evolution*, 62(6), 753–764. DOI: 10.1007/s00239-005-0074-7

Rego, A. C., Vesce, S., & Nicholls, D. G. (2001). The mechanism of mitochondrial membrane potential retention following release of cytochrome c in apoptotic GT1-7 neural cells. *Cell Death & Differentiation*, 8(10), 995–1003. DOI: 10.1038/sj.cdd.4400916

Ren, L., Tu, X., Luo, M., Liu, Q., Cui, J., Gao, X., Zhang, H., Tai, Y., Zeng, Y., Li, M., Wu, C., Li, W., Wang, J., Wu, D., & Liu, S. (2025). Genomes reveal pervasive distant hybridization in nature among cyprinid fishes. *GigaScience*, 14, giae117. DOI: 10.1093/gigascience/giae117

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, 16(6), 276–277. DOI: 10.1016/S0168-9525(00)02024-2

Roberts, V. A., & Pique, M. E. (1999). Definition of the interaction domain for cytochrome c on cytochrome c oxidase. *Journal of Biological Chemistry*, 274(53), 38051–38060. DOI: 10.1074/jbc.274.53.38051

Rollwitz, E., & Jastroch, M. (2021). Plate-based respirometry to assess thermal sensitivity of zebrafish embryo bioenergetics *in situ*. *Frontiers in Physiology*, 12, 1544. DOI: 10.3389/FPHYS.2021.746367/BIBTEX

Rooney, J. P., Ryde, I. T., Sanders, L. H., Howlett, E. H., Colton, M. D., Germ, K. E., Mayer, G. D., Greenamyre, J. T., & Meyer, J. N. (2015). PCR based determination of mitochondrial DNA copy number in multiple species. *Methods in Molecular Biology*, 1241, 1–14. DOI: 10.1007/978-1-4939-1875-1

Rüber, L., Kottelat, M., Tan, H. H., Ng, P. K. L., & Britz, R. (2007). Evolution of miniaturization and the phylogenetic position of *Paedocypris*, comprising the world's smallest vertebrate. *BMC Evolutionary Biology*, 7, 38. DOI: 10.1186/1471-2148-7-38

Saraste, M. (1999). Oxidative Phosphorylation at the fin de siècle. *Science*, 283(5407), 1488–1493. DOI: 10.1126/science.283.5407.1488

Sato, W., Hitaoka, S., Inoue, K., Imai, M., Saio, T., Uchida, T., Shinzawa-Itoh, K., Yoshikawa, S., Yoshizawa, K., & Ishimori, K. (2016). Energetic mechanism of cytochrome c-cytochrome c oxidase electron transfer complex formation under turnover conditions revealed by mutational effects and docking simulation. *Journal of Biological Chemistry*, 291(29), 15320–15331. DOI: 10.1074/jbc.M115.708065

Sayers, E. W., Beck, J., Bolton, E. E., Brister, J. R., Chan, J., Connor, R., Feldgarden, M., Fine, A. M., Funk, K., Hoffman, J., Kannan, S., Kelly, C., Klimke, W., Kim, S., Lathrop, S., Marchler-Bauer, A., Murphy, T. D., O'Sullivan, C., Schmieder, E., … Pruitt, K. D. (2024). Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Research*, 53(D1), D20–D29. DOI: 10.1093/nar/gkae979

Scharlau, M., Geren, L., Zhen, E. Y., Ma, L., Rajagukguk, R., Ferguson-Miller, S., Durham, B., & Millett, F. (2019). Definition of the interaction domain and electron transfer route between cytochrome c and cytochrome oxidase. *Biochemistry*, 58(40), 4125–4135. DOI: 10.1021/acs.biochem.9b00646

Schmidt, T. R., Wildman, D. E., Uddin, M., Opazo, J. C., Goodman, M., & Grossman, L. I. (2005). Rapid electrostatic evolution at the binding site for cytochrome c on cytochrome c oxidase in anthropoid primates. *Proceedings of the National Academy of Sciences*, 102(18), 6379–6384. DOI: 10.1073/pnas.0409714102

Schrödinger, LLC. (2025). *The PyMOL Molecular Graphics System* (Version 3.1.4.1) [Computer software].

Sharma, L. K., Lu, J., & Bai, Y. (2009). Mitochondrial respiratory complex I: Structure, function and implication in human diseases. *Current Medicinal Chemistry*, 16(10), 1266–1277.

Sholihah, A., Delrieu-Trottin, E., Condamine, F. L., Wowor, D., Rüber, L., Pouyaud, L., Agnèse, J.-F., & Hubert, N. (2021). Impact of Pleistocene eustatic fluctuations on evolutionary dynamics in Southeast Asian biodiversity hotspots. *Systematic Biology*, 70(5), 940–960. DOI: 10.1093/sysbio/syab006

Shtolz, N., & Mishmar, D. (2019). The Mitochondrial genome – on selective constraints and signatures at the organism, cell, and single mitochondrion levels. *Frontiers in Ecology and Evolution*, 7, 342. DOI: 10.3389/fevo.2019.00342

Signes, A., & Fernandez-Vizarra, E. (2018). Assembly of mammalian oxidative phosphorylation complexes I–V and supercomplexes. *Essays in Biochemistry*, 62(3), 255–270. DOI: 10.1042/EBC20170098

Simakova, M. N., & Simakov, N. N. (2014). Topography prediction of helical transmembrane proteins by a new modification of the sliding window method. *BioMed Research International*, 2014, 921218. DOI: 10.1155/2014/921218

Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, 6(1), 31. DOI: 10.1186/1471-2105-6-31

Spikings, E. C., Alderson, J., & John, J. C. St. (2007). Regulated mitochondrial DNA replication during oocyte maturation is essential for successful porcine embryonic development. *Biology of Reproduction*, 76(2), 327–335. DOI: 10.1095/biolreprod.106.054536

SRA Toolkit Development Team, National Center for Biotechnology Information. (n.d.). SRA Toolkit. Available from: https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software

Steinke, S., Chiu, H.-Y., Yu, P.-S., Shen, C.-C., Erlenkeuser, H., Löwemark, L., & Chen, M.-T. (2006). On the influence of sea level and monsoon climate on the southern South China Sea freshwater budget over the last 22,000 years. *Quaternary Science Reviews*, 25(13), 1475–1488. DOI: 10.1016/j.quascirev.2005.12.008

Stewart, J. B., Freyer, C., Elson, J. L., Wredenberg, A., Cansu, Z., Trifunovic, A., & Larsson, N.-G. (2008). Strong purifying selection in transmission of mammalian mitochondrial DNA. *PLOS Biology*, 6(1), e10. DOI: 10.1371/journal.pbio.0060010

Streisinger, G., Walker, C., Dower, N., Knauber, D., & Singer, F. (1981). Production of clones of homozygous diploid zebra fish (*Brachydanio rerio*). *Nature*, 291(5813), 293–296. DOI: 10.1038/291293a0

Sukumar, M., Liu, J., Mehta, G. U., Patel, S. J., Roychoudhuri, R., Crompton, J. G., Klebanoff, C. A., Ji, Y., Li, P., Yu, Z., Whitehill, G. D., Clever, D., Eil, R. L., Palmer, D. C., Mitra, S., Rao, M., Keyvanfar, K., Schrump, D. S., Wang, E., … Restifo, N. P. (2016). Mitochondrial membrane potential identifies cells with enhanced stemness for cellular therapy. *Cell Metabolism*, 23(1), 63–76. DOI: 10.1016/j.cmet.2015.11.002

Sunnucks, P., Morales, H. E., Lamb, A. M., Pavlova, A., & Greening, C. (2017). Integrative approaches for studying mitochondrial and nuclear genome co-evolution in oxidative phosphorylation. *Frontiers in Genetics*, 8, 25. DOI: 10.3389/fgene.2017.00025

Suurväli, J., Whiteley, A. R., Zheng, Y., Gharbi, K., Leptin, M., & Wiehe, T. (2020). The laboratory domestication of zebrafish: From diverse populations to inbred substrains. *Molecular Biology and Evolution*, 37(4), 1056–1069. DOI: 10.1093/molbev/msz289

Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, 34(Web Server issue), W609–W612. DOI: 10.1093/nar/gkl315

Tait, S. W. G., & Green, D. R. (2010). Mitochondria and cell death: Outer membrane permeabilization and beyond. *Nature Reviews Molecular Cell Biology*, 11(9), 621–632. DOI: 10.1038/nrm2952

The UniProt Consortium. (2025). UniProt: The Universal Protein Knowledgebase in 2025. *Nucleic Acids Research*, 53(D1), D609–D617. DOI: 10.1093/nar/gkae1010

Tretter, L., Patocs, A., & Chinopoulos, C. (2016). Succinate, an intermediate in metabolism, signal transduction, ROS, hypoxia, and tumorigenesis. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1857(8), 1086–1101. DOI: 10.1016/j.bbabio.2016.03.012

Trevena, R. L., Veire, B. M., Chamberlain, T. J., Moravec, C. E., & Pelegri, F. (2025). Embryonic lethality, juvenile growth variation, and adult sterility correlate with phylogenetic distance of Danionin hybrids. *Evolution & Development*, 27(1), e12495. DOI: 10.1111/ede.12495

Turelli, M., & Moyle, L. C. (2007). Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics*, 176(2), 1059–1088. DOI: 10.1534/genetics.106.065979

Urbina, M. A., & Glover, C. N. (2013). Relationship between fish size and metabolic rate in the xxyconforming Inanga *Galaxias maculatus* reveals size-dependent strategies to withstand hypoxia. *Physiological and Biochemical Zoology*, 86(6), 740–749. DOI: 10.1086/673727

van der Windt, G. J. W., Chang, C. H., & Pearce, E. L. (2016). Measuring bioenergetics in T cells using a seahorse extracellular flux analyzer. *Current Protocols in Immunology*, 2016, 3.16B.1-3.16B.14. DOI: 10.1002/0471142735.im0316bs113

van Gisbergen, M. W., Voets, A. M., Starmans, M. H. W., de Coo, I. F. M., Yadak, R., Hoffmann, R. F., Boutros, P. C., Smeets, H. J. M., Dubois, L., & Lambin, P. (2015). How do changes in the mtDNA and mitochondrial dysfunction influence cancer and cancer therapy? Challenges, opportunities and models. *Mutation Research*, 764, 16–30. DOI: 10.1016/J.MRREV.2015.01.001

Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov, M., Nair, S., Mirdita, M., Yeo, J., Kovalevskiy, O., Tunyasuvunakool, K., Laydon, A., Žídek, A., Tomlinson, H., Hariharan, D., Abrahamson, J., Green, T., Jumper, J., … Velankar, S. (2024). AlphaFold Protein Structure Database in 2024: Providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research*, 52(D1), D368–D375. DOI: 10.1093/nar/gkad1011

Vincent, A. E., Ng, Y. S., White, K., Davey, T., Mannella, C., Falkous, G., Feeney, C., Schaefer, A. M., McFarland, R., Gorman, G. S., Taylor, R. W., Turnbull, D. M., & Picard, M. (2016). The spectrum of mitochondrial ultrastructural defects in mitochondrial myopathy. *Scientific Reports*, 6(1), 30610. DOI: 10.1038/srep30610

Wani, N. A., Vettical, B. S., & Hong, S. B. (2017). First cloned Bactrian camel (Camelus bactrianus) calf produced by interspecies somatic cell nuclear transfer: A step towards preserving the critically endangered wild Bactrian camels. *PloS One*, 12(5), e0177800. DOI: 10.1371/journal.pone.0177800

Weaver, R. J., Rabinowitz, S., Thueson, K., & Havird, J. C. (2022). Genomic Signatures of Mitonuclear Coevolution in Mammals. *Molecular Biology and Evolution*, 39(11), msac233. DOI: 10.1093/molbev/msac233

Welch J. J. (2004). Accumulating Dobzhansky-Muller incompatibilities: Reconciling theory and data. *Evolution*, 58(6), 1145–1156. DOI: 10.1111/j.0014-3820.2004.tb01695.x

Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L., & Scheffler, K. (2015). RELAX: Detecting relaxed selection in a phylogenetic framework. *Molecular Biology and Evolution*, 32(3), 820–832. DOI: 10.1093/molbev/msu400

West, A. P., Shadel, G. S., & Ghosh, S. (2011). Mitochondria in innate immune responses. *Nature Reviews Immunology*, 11(6), 389–402. DOI: 10.1038/nri2975

Westerfield, M. (2007) *The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio)*, (5th Ed). University of Oregon Press, Eugene.

Wick, R. R., Schultz, M. B., Zobel, J., & Holt, K. E. (2015). Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics*, 31(20), 3350–3352. DOI: 10.1093/bioinformatics/btv383

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.

Wolters, J. F., Charron, G., Gaspary, A., Landry, C. R., Fiumera, A. C., & Fiumera, H. L. (2018). Mitochondrial recombination reveals mito-mito epistasis in yeast. *Genetics*, 209(1), 307–319. DOI: 10.1534/genetics.117.300660

Woolfit, M., & Bromham, L. (2003). Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. *Molecular Biology and Evolution*, 20(9), 1545–1555. DOI: 10.1093/molbev/msg167

Wu, W., Schmidt, T. R., Goodman, M., & Grossman, L. I. (2000). Molecular evolution of cytochrome c oxidase subunit I in primates: Is there coevolution between mitochondrial and nuclear genomes? *Molecular Phylogenetics and Evolution*, 17(2), 294–304. DOI: 10.1006/mpev.2000.0833

Xie, D., Chen, G., Meng, X., Wang, H., Bi, X., Fang, M., Yang, C., Zhou, Y., Long, E., & Feng, S. (2021). Comparable number of genes having experienced positive selection among great ape species. *Animals: An Open Access Journal from MDPI*, 11(11), 3264. DOI: 10.3390/ani11113264

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28–36. DOI: 10.1111/2041-210X.12628

Zhao, L., Zhou, W., He, J., Li, D.-Z., & Li, H.-T. (2024). Positive selection and relaxed purifying selection contribute to rapid evolution of male-biased genes in a dioecious flowering plant. *eLife*, 12. DOI: 10.7554/eLife.89941.3

Zhao, R.-Z., Jiang, S., Zhang, L., & Yu, Z.-B. (2019). Mitochondrial electron transport chain, ROS generation and uncoupling (Review). *International Journal of Molecular Medicine*, 44(1), 3–15. DOI: 10.3892/ijmm.2019.4188

Zheng, H.-X., Yan, S., Zhang, M., Gu, Z., Wang, J., & Jin, L. (2024). Mitochondrial DNA genomes reveal relaxed purifying selection during human population expansion after the last glacial maximum. *Molecular Biology and Evolution*, 41(9), msae175. DOI: 10.1093/molbev/msae175

Zhu, T., Sato, Y., Sado, T., Miya, M., & Iwasaki, W. (2023). MitoFish, MitoAnnotator, and MiFish Pipeline: Updates in 10 Years. *Molecular Biology and Evolution*, 40(3), msad035. DOI: 10.1093/molbev/msad035