The Genetic and Morphological Evolution of Fish through the Protein Cytochrome C Oxidase Subunit One

Sammer Marzouk, University of Chicago Laboratory Schools, 1362 E 59th St, Chicago, IL 60637, marzouk.sammer@gmail.com
Abstract

Cytochrome C Oxidase Subunit 1 (COX1) is a protein that helps to catalyze the reduction of water into oxygen in Eukaryotes. Through the analysis of COX1 from online public genetic databases in 16 species of fish, an evolutionary phylogeny of fish was derived from the data. This paper considered three hypotheses: Axolotl (*Ambystoma mexicanum*) and Tiktaalik (*Tiktaalik roseae*) share a common ancestor was determined; that *A.gueldenstaedtii* and *P.spathula* both share a common ancestor; and that *P.marinus* and *S.acanthias* are the outliers of the phylogeny. The evolutionary phylogeny used the percent ID between the two species of fish. From these differences, analysis is done to the data and the data is used to make phylogenies based on the morphological and genetic evolution of these fish. From the data derived from the phylogenies, the results demonstrates the claims that Axolotl (*Ambystoma mexicanum*) and Tiktaalik (*Tiktaalik roseae*) share a common ancestor was determined, *A.gueldenstaedtii* and *P.spathula* both share a common ancestor, and *P.marinus* and *S.acanthias* are the outliers of the phylogeny. The data gathered can be used to connect tetrapods to fish, and contribute to the theory that tetrapods came from fish. The information presented in this paper can be used to make a complete phylogeny of all organisms in the biosphere.

Introduction

Of the 30,000 species of fish, the phylogenetic trees of fish have not been heavily studied on built upon for the reason that their has not been a reliable connection between fish and tetrapods (Fiz-Palacios O et al., 2013). Charles Darwin first proposed that animals could be categorized by 2016.
the morphological traits they possessed, and supported this by categorizing finches based on their beaks. This type of phylogeny made by data from morphological traits is the morphological phylogeny. As DNA sequencing technology became in greater use, a new type of phylogeny from which protein-sequences were used was also used to categorize animals. These two types of phylogenies would show a hypothesis of the evolution of certain beings. But sometimes there would be differences in these two types of phylogenies. This demonstrates the need of how can a cladogram be inferred using genetic data be reconciled with a cladogram using only morphological data? From a sample size of 16 species of fish, a morphological phylogeny was previously derived (Fig. 1). The phylogeny has 21 traits and shows which traits evolved in which species (Fig. 1). From the morphological phylogeny, it is shown that gills, spiracles, and a cartilage skeleton are the basic traits that make this clade of species (Fig. 1). Overtime, more diverse and complex traits divided up the fish into separate taxa (Fig. 1). In the morphological phylogeny Tiktaalik (Tiktaalik roseae), and Axolotl (Ambystoma mexicanum) were both in the same clade, which lead to the hypothesis that if the morphological and protein phylogenies were compared, then it would determine that they shared a common ancestor (Fig. 1). Also, in the morphological phylogeny, it is shown that A. gueldenstaedtii and P. spathula both share a common ancestor, and P. marinus and S. acanthias are the outliers of the phylogeny (Fig. 1). In order to get the most accurate results, this morphological data was combined with the protein-sequence of Cytochrome C Oxidase Subunit 1 (COX1) (Fig. 5). Cytochrome C Oxidase Subunit 1 (COX1) is one out of the three mitochondrial DNA subunits, the others being MT-CO2, MT-CO3 that are part of respiratory complex IV (Tsukihara et al., 1996). In respiratory complex IV, this enzyme completes the electron transport system, and its function of 2016.
catalyzing the reduction of water into oxygen (Tsukihara et al., 1996). Most Eukaryotes, and all vertebrates have this protein (Tsukihara et al., 1996). The COX1 from the Spiny Dogfish (Squalus acanthias), Sea Lamprey (Petromyzon marinus), Saddled Bichir (Polypterus endlicheri), Paddlefish (Polyodon spathula), Ghost Shark (Callorhynchus milii), Coelacanth (Latimeria chalumnae), Bowfin (Amia calva), Pollen’s Cichlid (Paratilapia pollen), Tiktaalik (Tiktaalik roseae), Ropefish (Erpetoichthys calabaricus), Great White Shark (Carcharodon carcharias), Axolotl (Ambystoma mexicanum), Cuvier’s Bichir (Polypterus senegalus), Russian Sturgeon (Acipenser gueldenstaedti), Australian Lungfish (Neoceratodus forsteri), and Goldfish (Carassius auratus) were sequenced and compared in order to find the percent ID (Fig. 4).
This percent ID was used in order to make a phylogeny. The protein-sequenced phylogeny was compared with the morphological phylogeny to make a final phylogeny. This paper describes the morphological and protein similarity between the 16 species of fish, and describes the use of building a more extensive phylogenetic tree.

**Materials and Methods**
2.1 Morphological Phylogeny

The morphological tree phylogeny was constructed based off of 21 synapomorphies of a sample group of 16 species of fish (Fig. 1). A species would be compared with the other species of fish. If the species had or did not have a specific trait, they would be placed in a different clade than species that did not have another trait (Fig. 1). The continuation of this pattern would give a complete morphological phylogeny (Fig. 1). The morphological phylogeny was based off of a data matrix which was derived from the characteristics that the species of fish might or might not have (Fig. 6). For every characteristic, if it was present in that species of fish it had a “Y” otherwise it had a “N” in it (Fig. 6).

![Fig 6 Data Matrix](https://example.com/data_matrix.png)

Fig 6 Data Matrix. This is the data matrix that was used to make the morphological phylogeny. Every time a species had a certain trait, a “Y”, which represents yes was put. If the species did not have a “N”, then it represents no, and means that the species does not have that trait.
2.2 Protein Sequences and Construction of Phylogeny

The protein sequence of the fish was derived from an online public genomic database (Fig. 7). The data was then compared and analyzed by a series of specialized computer algorithms. These computer algorithms gave the specific percent identity (specific as in rounded to the nearest tenth) and the specific percent similar. These percentages were then used, and compiled on a genetic chart. From the genetic chart, the phylogenetic tree was derived (Fig. 7). In order to make the phylogeny, the data on the first species, *S. acanthias* was compared to the rest of the fish (Fig. 2). The one with the largest percent identity was put into a clade with *S. acanthias* (Fig. 2). After
this, the next species that was closest to *S.acanthias* was put in the same clade the two previous fish.

![Protein Sequence of Fish](image)

**Fig. 7 Protein Sequence of Fish.** This is the protein sequence of COX1 in the 16 species of fish. All genetic information was derived from a genetic database, and is demonstrated here in the form FASTA.
2.3 Combination of Morphological and Protein Phylogeny

In order to combine the types of phylogenies, further examination on the similarities and differences of the phylogenies would be conducted (Fig. 3). Similarities would be kept and considered a constant of both phylogenies (Fig. 3). Differences in species placement were decided by comparing the two phylogenies and deriving a spot for the species based on a
specialized computer algorithm. Any contradictions would be resolved by analysing the percent similar and determining which species had the greater percent similar (Fig. 3). This species would then be placed accordingly (Fig. 3).

2.4 Data Analysis

All sequenced protein data has been verified by the proper scientific body. Data analysis was by a specialized computer algorithm developed by William Pearson and run on the LALIGN server. All data was tested for specific outliers and the percent similar induced function of amino acids. Percent identity included exact amino acid comparisons. Identical and similar percentages were rounded to the nearest tenth.

Results

From the data, three specific conclusions were drawn. The first conclusion drawn from the data was that Axolotl (*Ambystoma mexicanum*) and Tiktaalik (*Tiktaalik roseae*) share a common ancestor (Fig. 1, Fig. 2, Fig. 3).
Morphologically, *A. mexicanum*, and *T. roseae* share the synapomorphies of tetrapod limbs, lobed fins, cosmoid scales, an air-filled swim bladder, an ossified skeleton, cycloid scales, a reduced amount of scales, jaws, paired appendages, and having a cartilage skeleton (Fig. 1). Since *T. roseae* is an extinct species, no protein sequence of COX1 from *T. roseae* was available (Fig. 2). Only the morphological phylogeny could be used to infer where *T. roseae* goes on the final phylogeny (Fig. 1). In the final phylogeny, when the data from the previous phylogenies were combined, it was shown that *A. mexicanum* and *T. roseae*, both from protein comparisons and morphological traits, shared a common ancestor (Fig. 1, Fig. 2, Fig. 3)
In the morphological phylogeny, it is shown that *A. gueldenstaedtii* and *P. spathula* are in the same clade as they both have jelly-filled swim bladders, jaws, paired appendages, placoid scales, gills, spiracles, and cartilage skeletons (Fig. 1). And in the protein-derived phylogeny, it is shown that *A. gueldenstaedtii* and *P. spathula* have a 98.5 percent identity with each other (Fig. 2). In the final phylogeny, it was shown that *A. gueldenstaedtii* and *P. spathula* with the 98.5 2016.
percent identity and the similar synapomorphies are related and have once shared a common ancestor (Fig. 5).
The third conclusion derived from the data is that *P. marinus* and *S. acanthias* are the outliers of 2016.

![Table](#)

**Fig. 5 Genetic Table of Protein Comparisons.** This table shows the protein similarity in each of the fish. The first percentage shown is the percent identical, and the second one shown is the percent similar. The reason why Tiktaalik has no percent comparisons is because there is no COX1 available for this species of fish.

The third conclusion derived from the data is that *P. marinus* and *S. acanthias* are the outliers of 2016.
the phylogeny (Fig.1). In the morphological phylogeny, *P. marinus* and *S. acanthias* were shown to share many similar synapomorphies (Fig. 1). These include gills, spiracles, and cartilage skeletons (Fig. 1). But in the protein-derived phylogeny, it was demonstrated that the COX1 comparison between the two species of fish yield an 87.3 percent identity (Fig. 5). This is a high average, but compared to the other relationships, such as *S. acanthias* and *A. gueldenstaedtii* sharing a 90.4 percent identity, and *P. marinus* and *P. pollen*, who morphologically are different, only sharing three synapomorphies by having a 90.8 percent identity (Fig.1, Fig.5). In the final phylogeny, it was decided that *S. acanthias* and *A. gueldenstaedtii* would be the outliers of the group, because the protein-derived data was more accurately studied, and the synapomorphies could be modified in a way to change the position of any of the two fish (Fig. 3). These three conclusions were the three main results. Another finding revealed that all of the 16 fish species are all related (Fig. 5). Of the 15 COX1 protein sequences that were available, all of them had an percent identity greater than 88% (Fig. 5). Morphologically, the species of fish are shown to all have similar traits, except *P. marinus*, which is shown to be the outlier of the group (Fig. 5). This information correlates to the protein-derived phylogeny in which *P. marinus* is also the outlier of the clade, and also in the final phylogeny in which *P. marinus* is the outlier of the clade (Fig.1, Fig.2, Fig.3). *S. acanthias* is also shown to be constant in its position in all three phylogenies. In the morphological, protein-derived, and final phylogenies, *S. acanthias* was shown to always be in the same clade as *C. carcharias* (Fig.1, Fig.2, Fig.3). Every species on the morphological and protein-derived phylogeny were in the same position in the final phylogeny except *P. senegalus*, *P. Endlicheri* and *A. calva* (Fig.1, Fig.2, Fig.3). In protein-derived phylogeny, it was demonstrated that *P. Endlicheri* was similar to *P. senegalus* while in the morphological
phylogeny, it was demonstrated that *C.milii* and *A.calva* were more related *P.senegalus* then *P.Endlicheri* was (Fig.1, Fig.2, Fig.3).

**Discussion**

In an effort to build a phylogeny that demonstrates a link between fish and tetrapods, COX1 from 16 different species was sequenced and made into a protein-derived phylogeny (Fig. 3). This phylogeny was combined with a morphological phylogeny with 21 synapomorphies of the 16 species of fish to make a final phylogeny (Fig. 1). From this final phylogeny, three main conclusion were derived: *A.mexicanum* and *T.roseae* share a common ancestor; *A.gueldenstaedtii* and *P.spathula* both share a common ancestor; and *P.marinus* and *S.acanthias* are the outliers of the phylogeny. The first conclusion which was Axolotl (*Ambystoma mexicanum*) and Tiktaalik (*Tiktaalik roseae*) share a common ancestor was derived from the morphological phylogeny (Fig. 1). In the morphological phylogeny, both *A.mexicanum* and *T.roseae* formed a clade, and shared 15 synapomorphies (Fig. 1). In the protein-derived phylogeny, there was no evidence of a relationship between *A.mexicanum* and *T.roseae* as *T.roseae* is extinct (Fig. 2). *A.mexicanum* and *T.roseae* have been shown to share many synapomorphies (Ericsson R et al., 2012). *A.mexicanum* and *T.roseae* both share a specific organization of neck muscles, which goes around their gills and allows them to move the head (Ericsson R et al., 2012). *A.mexicanum* and *T.roseae* also both share appendicular skeleton which connect to their neck muscles (Ericsson R et al., 2012). *A.mexicanum* and *T.roseae* have both been shown to be closely related as they have been put on the same clade on morphological 2016.
phylogenies (Shubin H et al., 2014). *A. mexicanum* and *T. roseae* share a unique pelvic girdle and fin that allows for movement, and it is demonstrated that *T. roales* pelvic girdle was a transitional phase for *A. mexicanum*’s pelvic girdle (Shubin H et al., 2014). The second conclusion was that *A. gueldenstaedtii* and *P. spathula* both share a common ancestor, which was derived from the morphological and protein-based phylogeny (Fig. 1, Fig. 2). In the morphological phylogeny, *A. gueldenstaedtii* and *P. spathula* were in the same clade and shared 8 synapomorphies (Fig. 1). In the protein-derived phylogeny, *A. gueldenstaedtii* and *P. spathula* shared a 98.5 percent identity (Fig. 5). Morphologically, it is supported that *A. gueldenstaedtii* and *P. spathula* are related as they both have eggs with funnels and clear groves (Ludwig, A, 2008). Also, *A. gueldenstaedtii* and *P. spathula* share the same sized micropylar opening (Ludwig, A, 2008). Genetically the data is supported as *A. gueldenstaedtii* and *P. spathula* have similar heteroplasmatic movements in the mtDNA control region (Wang W et al., 2010). This allows both species to have lengths of central units 78-83 bp (Wang W et al., 2010). The conclusion that *A. gueldenstaedtii* and *P. spathula* are related shown through the fact that *A. gueldenstaedtii* and *P. spathula* are in the same Kingdom, Phylum, Superclass, Class, and order (Carroll R, 1998). This demonstrates a support for the claim that *A. gueldenstaedtii* and *P. spathula* are closely related (Carroll R, 1998). The third conclusion was that *P. marinus* and *S. acanthias* are the outliers of the phylogeny, which was derived from the morphological and protein-derived phylogeny (Fig. 2, Fig. 3). In the morphological phylogeny, *P. marinus* and *S. acanthias* share three synapomorphies, and were in different clades (Fig. 1). In the protein-derived phylogeny, *P. marinus* and *S. acanthias* share an 87.3 percent identity (Fig. 2). This is the lowest percent identity of the 15 species of fish that had a COX1 protein sequence available (Fig. 5). It is
supported that *P. marinus* and *S. acanthias* are morphologically different, as *P. marinus* and *S. acanthias* have different secretory granule compositions (Wilson M et al., 2002). *P. marinus* and *S. acanthias* also have a different type of microfilament network (Wilson M et al., 2002).

Genetically, the claim that *P. marinus* and *S. acanthias* are outliers of the fish phylogeny is supported. It has been shown that *P. marinus* and *S. acanthias* have different gas transfer rates due to the difference in respiratory system IV (Randall D, 1995). And it has been demonstrated that *P. marinus* and *S. acanthias* are both in genetically separate groups, which share a 0.400 similarly plasma level rate (Steve F et al., 1998). *P. marinus* and *S. acanthias* are also shown to be put into separate clades (Grim J et al., 2011). *P. marinus* and *S. acanthias* are demonstrated to be not very related, as they are in different classes, subclasses, orders, families and genuses (Grim J et al., 2011). This finding that *P. marinus* and *S. acanthias* shown to be a support for the idea that *P. marinus* and *S. acanthias* are not very related (Grimm J et al., 2011).

From the phylogenies, it was shown that species that were in the same genus and families were more closely related morphologically and genetically (Fig. 3). It was also shown that species that were morphologically similar were also similar in their protein structures (Fig. 3). A point of controversy on the topic of combining the protein-derived phylogeny and the morphological phylogeny is that other versions have put *P. marinus* and *S. acanthias* closer on the final phylogeny as they are morphologically similar (Fig. 3). This idea of putting *P. marinus* and *S. acanthias* close to each other because they have similar synapomorphies disrupts the phylogeny as *A. gueldenstaedtii* and *P. spathula* are more morphologically similar and have similar COX1 sequences (Fig. 5, Fig. 1). Putting *P. marinus* and *S. acanthias* in place of
*A. gueldenstaedtii* and *P. spathula* would push *A. gueldenstaedtii* and *P. spathula* apart on the phylogeny, even though they are more closely related (Fig. 3). In the phylogeny, there are only 16 limited taxa for the 16 species of fish to be put on. If one fish's position in the phylogeny changes, it will impact the rest of the phylogeny. These results build upon the idea that tetrapods and fish may share a common ancestor, and tetrapods may once been fish. This idea has implications throughout biological sciences (Nancy B et al., 2002). A phylogeny connecting tetrapods to fish could build a more detailed connection of the evolution of genes (Nancy B et al., 2002). This information could be more helpful as it could be used in the medical field to find specific new nutrients and new antibiotics (Nancy B et al., 2002). A phylogeny between tetrapods to fish could also allow for new connection to be made between primates and other animals (Nancy B et al., 2002). This connection could allow for medicine to be tested on fish before the medicines use of primates (Nancy B et al., 2002). A limitation of the protein-derived phylogeny was that *T. roseae* was extinct and did not have any sequences of COX1 available. Because of this, *T. roseae* only had one source for its placement on the final phylogeny, and that was the morphological phylogeny. The morphological phylogeny was only based off of certain traits. If these traits were chosen in specific ways, the morphological phylogeny could be theoretically controlled and be made to look any way the synapomorphies allow it to be. This could have affected the final phylogeny, as it allows for variations of the phylogeny. The process of combining morphological and protein-derived phylogenies is a flawed process in itself (Gontcharova. A et al., 2003). Even though computers and mathematical formulas make the combination more precise, phylogenies made from the combination of morphological and protein-derived phylogenies suffer from inadequate, noncongruent taxon sampling, long-branch
attraction, or conflicting evolutionary models of the genes analyzed (Gontcharova. A et al., 2003). In summary, the present data suggests that as time progressed, fish started to gain more tetrapod-like synapomorphies, and supports the theory that tetrapods came from fish. Additional studies concerning the morphological and genetic similarity between fish and terrestrial organisms might help better understand the link between tetrapods and fish.

References and Notes


Carroll, R. Vertebrate paleontology and evolution. 1998. 24; 698.


2016.


