

Poeciliid fishes in the genomic era

¹George-Cătălin Muntean, ^{2,3}Marian Proorocu, ^{2,4,5}I. Valentin Petrescu-Mag

¹ Faculty of Animal Science and Biotechnologies, University of Agricultural Sciences and Veterinary Medicine, Cluj-Napoca, Romania; ² Department of Environmental Engineering and Protection, Faculty of Agriculture, University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, Romania; ³ Enviromep SRL, Colonia Faget, Cluj, Romania;

⁴ University of Oradea, Oradea, Romania; ⁵ Bioflux SRL, Cluj-Napoca, Romania.

Corresponding author: I. V. PetrescuMag, zoobiomag2004@yahoo.com

Abstract. Through this paper, we aim to report the research within the group of poeciliid fishes and compare it to the current state of the art genomic research. We also want to show the importance of combining the two fields, one zoological and the other molecular, for the development of knowledge. The sequencing of the *Saccharomyces cerevisiae* genome was a major milestone in genomics and has since led to the sequencing of many other eukaryotic genomes, including those of humans, mice, and many other species. With the genomic era, genome sequencing began in many vertebrates, including poeciliid fishes. Poeciliids are a diverse family of freshwater fish that includes popular aquarium fish such as guppies, mollies, and swordtails. There are two species of Poeciliid fish whose genomes have been sequenced: *Poecilia reticulata* Peters, 1859 and *Xiphophorus maculatus* (Günther, 1866). Moreover, we have the mitochondrial genome sequenced in several other poeciliid species: *Poecilia formosa* (Girard, 1859), *Xiphophorus variatus* (Meek, 1904), *Xiphophorus hellerii* Heckel, 1848, *Xiphophorus couchianus* (Girard, 1859), *Poecilia parae* Eigenmann, 1894, *Poecilia latipinna* (Lesueur, 1821), *Poecilia mexicana Steindachner*, 1863. These genome sequences provide valuable resources for researchers studying the genetics of poeciliid fish and their evolutionary history. By sequencing an animal's genome, we can identify the genes and genetic variants responsible for various traits such as disease susceptibility, growth rate, and behavior. This knowledge can be used to develop new therapies and treatments, improve breeding strategies, and enhance animal welfare.

Key Words: eukaryotic genome, genomics, guppy, platyfish, Poeciliidae, sequenced.

Introduction. Through this paper, we aim to report the research within the group of poeciliid fishes and compare it to the current state of the art in the genomic research. We also want to show the importance of combining the two fields, one zoological and the other molecular, for the development of knowledge.

Genomics. Genomics is the study of genomes, which are the complete set of genetic material (DNA) within an organism (Green et al 2020). Genomics deals with the structure, function, evolution, and mapping of genomes, as well as the interactions between genes and their environment. The field of genomics encompasses a wide range of research areas, including:

- Genome sequencing and assembly: This involves determining the order of nucleotides (A, C, G, and T) that make up an organism's genome and assembling them into a complete genome sequence (Luo et al 2020).
- Functional genomics: This involves studying the function of genes and other genomic elements, including how they interact with each other and with the environment (Cano-Gamez & Trynka 2020).
- Comparative genomics: This involves comparing the genomes of different organisms to identify similarities and differences in gene structure, function, and evolution (Eichler 2019).

- Epigenomics: This involves studying modifications to the genome that affect gene expression, such as DNA methylation and histone modifications (Noble & Hunter 2020).
- Metagenomics: This involves studying the genomic material of entire communities of microorganisms, such as those found in soil, water, or the human gut (Appolinario et al 2019).
- Personalized genomics: This involves using genomic information to tailor medical treatments to an individual's unique genetic makeup (Lunshof et al 2010).
- Genomics plays an important role in advancing our understanding of biology, medicine, agriculture, environment, and other fields (Breed et al 2019).

The first fully sequenced eukaryotic genome. The first fully sequenced eukaryotic genome was that of the baker's yeast, *Saccharomyces cerevisiae*. The genome sequence was published in 1996 by an international consortium of scientists led by the European Molecular Biology Laboratory (EMBL) and Stanford University (Goffeau et al 1996). The genome sequence was determined using a combination of physical mapping, chromosome walking, and sequencing of overlapping clones. The *S. cerevisiae* genome consists of approximately 12.1 million base pairs and contains around 6,000 protein-coding genes (Yazgan & Krebs 2012). The sequencing of the *S. cerevisiae* genome was a major milestone in genomics and has since led to the sequencing of many other eukaryotic genomes, including those of humans, mice, and many other species.

Genome size. The size of eukaryotic genomes can vary greatly depending on the species. The genome size can range from less than 10 megabases (Mb) in some unicellular eukaryotes to more than 100 gigabases (Gb) in certain plants, such as *Paris japonica* (Hidalgo et al 2017).

In general, the genome size of animals is smaller than that of plants, with most animal genomes falling in the range of 0.1-10 Gb. Humans, for example, have a diploid genome size of approximately 3 billion base pairs (or 3 Gb) (Venter et al 2001). However, there are exceptions to this trend, such as the axolotl (*Ambystoma mexicanum*), a type of salamander (Proorocu & Petrescu-Mag 2022), which has a genome size estimated to be 10 times larger than the human genome (Nowoshilow et al 2018).

The size of the genome does not necessarily correlate with the organism's complexity or the number of genes it contains, as some organisms have large genomes with many repetitive sequences, while others have small genomes with little or no repetitive sequences.

Poeciliid fishes in the genomic era. The genomic era refers to the period in which the sequencing of genomes has become more widespread and accessible. The genomic era began in the 1990s with the first complete sequencing of a free-living organism, *Haemophilus influenzae* (Fleischmann et al 1995), and continued with the sequencing of the human genome in 2001 (Venter et al 2001). Since then, technological advances have made genome sequencing faster, cheaper, and more accurate, leading to the sequencing of thousands of genomes across a wide range of organisms.

The genomic era has revolutionized the field of biology, allowing scientists to better understand the genetic basis of traits and diseases, track the evolution of species, and improve the breeding of plants and animals. It has also opened new opportunities in medicine, allowing for personalized treatments based on an individual's genomic information. The genomic era has enabled a deeper understanding of the complex genetic processes that underlie life and has paved the way for many important scientific discoveries and applications.

With the genomic era, genome sequencing began in many vertebrates, including poeciliid fishes. Poeciliids are a diverse family of freshwater fish that includes popular aquarium fish such as guppies, mollies, and swordtails (Petrescu-Mag et al 2013; Papuc et al 2022). There are two species of poeciliid fish whose genomes have been sequenced.

Poecilia reticulata Peters, 1859: the entire genome of *P. reticulata* has been sequenced and is available in public databases. The genome assembly was first published in 2015 and is based on a combination of short-read and long-read sequencing technologies. The genome is approximately 731 Mb in size and is composed of $n=23$ chromosomes ($2n=46$; Y chromosome not included in the study) (Künstner et al 2016). The mitochondrial genome, in addition to the nuclear genome, of *P. reticulata* has also been sequenced and is available in public databases such as GenBank. The mitochondrial genome of *P. reticulata* is approximately 16.6 kilobases (Kb) in length and consists of 13 protein-coding genes, 22 transfer RNA genes, two ribosomal RNA genes, and a non-coding control region, similar to other fish species in the Poeciliidae family (Künstner et al 2016).

Southern platyfish, *Xiphophorus maculatus* (Günther, 1866): the entire genome of *X. maculatus* has been sequenced and is available in public databases. The genome assembly was first published in 2013 and is based on a combination of short-read and long-read sequencing technologies. The genome is approximately 729 Mb in size and is composed of 24 chromosomes ($2n=46$) (Schartl et al 2013). The mitochondrial genome, in addition to the nuclear genome, of *X. maculatus* has also been sequenced and is available in public databases such as GenBank. The mitochondrial genome of *X. maculatus* is approximately 16.6 Kb in length and consists of 13 protein-coding genes, 22 transfer RNA genes, two ribosomal RNA genes, and a non-coding control region, similar to *P. reticulata* (Schartl et al 2013).

Amazon molly, *Poecilia formosa* (Girard, 1859): The entire genome of *P. formosa* has not been sequenced yet. However, the mitochondrial genome of *P. formosa* has been sequenced and is available in public databases such as GenBank. The mitochondrial genome of *P. formosa* is approximately 16.6 Kb in length and consists of 13 protein-coding genes, 22 transfer RNA genes, two ribosomal RNA genes, and a non-coding control region (Dang et al 2016).

Variable platyfish, *Xiphophorus variatus* (Meek, 1904): The mitochondrial genome of *X. variatus* has been sequenced and is available in public databases such as GenBank. The mitochondrial genome of *X. variatus* is approximately 16.6 Kb in length and consists of 13 protein-coding genes, 22 transfer RNA genes, two ribosomal RNA genes, and a non-coding control region, similar to *P. formosa*, *P. reticulata* and *X. maculatus* (Eastis et al 2021).

We already have the mitochondrial genome sequenced in several other poeciliid species: *Xiphophorus hellerii* Heckel, 1848, *Xiphophorus couchianus* (Girard, 1859) (Zhang et al 2016), *Poecilia parae* Eigenmann, 1894 (Fast et al 2023), *Poecilia latipinna* (Lesueur, 1821), *Poecilia mexicana* Steindachner, 1863 (Schoch et al 2020). These genome sequences provide valuable resources for researchers studying the genetics of poeciliid fish and their evolutionary history (see Figure 1).

The mitochondrial genome is useful for animal phylogeny studies for several reasons. In animals, the mitochondrial genome is relatively small (compared to nuclear genomes) and circular in structure. This allows for easy isolation and sequencing of the entire genome. Mitochondrial DNA (mtDNA) is maternally inherited, meaning it is passed down from the mother to her offspring. This results in a lack of recombination, which maintains the integrity of the gene order and makes it easier to trace evolutionary relationships. Mitochondria have a high mutation rate compared to nuclear DNA (Jones et al 2001), making them useful for studying recent evolutionary events and relationships. The mitochondrial genome contains a large number of conserved regions that can be used to design primers for PCR amplification and sequencing. The mitochondrial genome provides a valuable tool for reconstructing evolutionary relationships within and between animal species, as well as for understanding the evolutionary history of animal groups.

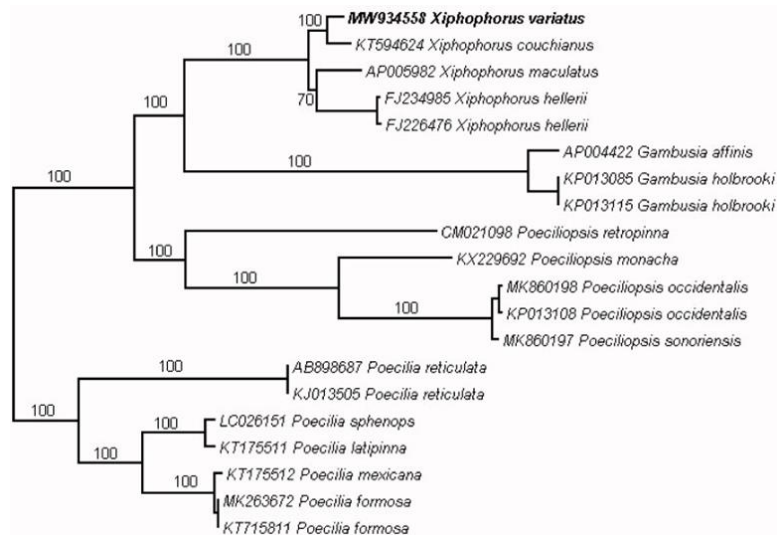


Figure 1. The evolution of the mitochondrial genome in the Poeciliidae group was reconstructed by utilizing the maximum-likelihood (MLE) method in the RAxML-HPC BlackBox. This was made through the CIPRES portal using a sequence of 16,944 bp. A resulting file was analyzed with Dendroscope. Only species with known nucleotide sequences were included in the study. Numbers on branch lengths are bootstrap support values, expressed as % (source: Eastis et al 2021).

Why is knowing the genome of animals important? Knowing the entire genome sequence of animals is important for several reasons:

- **Understanding genetic basis of traits:** By sequencing an animal's genome, we can identify the genes and genetic variants responsible for various traits such as disease susceptibility, growth rate, and behavior. This knowledge can be used to develop new therapies and treatments, improve breeding strategies, and enhance animal welfare.
- **Evolutionary studies:** Comparative genomics can reveal evolutionary relationships between species, and the genomic changes that have led to the emergence of new traits and adaptations.
- **Conservation efforts:** Genome sequencing can provide valuable information for conservation efforts by identifying genetic diversity within populations, understanding population history, and identifying potential threats to endangered species.
- **Biomedical research:** Animals are used extensively in biomedical research to study human diseases and develop treatments. Knowing the genome sequence of an animal can provide a valuable resource for biomedical research, as it can help identify genetic causes of diseases and test potential treatments.
- **Agriculture:** Genome sequencing can be used to improve the production of livestock and crops by identifying genes that control desirable traits such as yield, disease resistance, and nutrient content.
- **Genome sequencing of animals is important for advancing our understanding of biology, improving animal welfare, and promoting human health and well-being.**

When can we hope to have all the genomes of the living world sequenced? It's difficult to make an accurate estimate because the rate of technological development and the cost of sequencing genomes are constantly changing. However, based on current trends, it's possible that we could have the genomes of all extant species completely sequenced within the next few decades or even sooner.

For example, the cost of sequencing a genome has dropped significantly over the past two decades. In 2001, the cost of sequencing a single human genome was around \$100 million, while in 2021, the cost has dropped to under \$1,000 (www.ourworldindata.org). This cost reduction has made genome sequencing more

accessible to researchers and has led to an explosion in the number of sequenced genomes.

Additionally, new sequencing technologies such as long-read sequencing and nanopore sequencing have greatly improved the speed and accuracy of genome sequencing. These technologies allow researchers to sequence larger and more complex genomes, which were previously difficult to sequence using older technologies.

Given these advancements, it's possible that we could have the genomes of all extant species sequenced within the next few decades or sooner. However, this will depend on many factors such as funding, technological progress, and the willingness of researchers and governments to invest in genome sequencing efforts.

Conclusions. The sequencing of the *S. cerevisiae* genome was a major milestone in genomics and has since led to the sequencing of many other eukaryotic genomes, including those of humans, mice, and many other species. With the genomic era, genome sequencing began in many vertebrates, including poeciliid fishes. Poeciliids are a diverse family of freshwater fish that includes popular aquarium fish such as guppies, mollies, and swordtails. There are two species of poeciliid fish whose genomes have been sequenced: *Poecilia reticulata* Peters, 1859 and *Xiphophorus maculatus* (Günther, 1866). Moreover, we have the mitochondrial genome sequenced in several other poeciliid species: *Poecilia formosa* (Girard, 1859), *Xiphophorus variatus* (Meek, 1904), *Xiphophorus hellerii* Heckel, 1848, *Xiphophorus couchianus* (Girard, 1859), *Poecilia parae* Eigenmann, 1894, *Poecilia latipinna* (Lesueur, 1821), *Poecilia mexicana* Steindachner, 1863. These genome sequences provide valuable resources for researchers studying the genetics of poeciliid fish and their evolutionary history. By sequencing an animal's genome, we can identify the genes and genetic variants responsible for various traits such as disease susceptibility, growth rate, and behavior. This knowledge can be used to develop new therapies and treatments, improve breeding strategies, and enhance animal welfare.

Conflict of interest. The authors declare that there is no conflict of interest.

References

- Appolinario L. R., Tschöcke D., Paixão R. V., Venas T., Calegario G., Leomil L., Silva S. B., Thompson C. C., Thompson F. L., 2019 Metagenomics sheds light on the metabolic repertoire of oil-biodegrading microbes of the South Atlantic Ocean. *Environmental Pollution* 249:295-304.
- Breed M. F., Harrison P. A., Blyth C., Byrne M., Gaget V., Gellie N. J., Groom S. V. C., Hodgson R., Mills J. G., Prowse T. A. A., Steane D. A., Mohr J. J., 2019 The potential of genomics for restoring ecosystems and biodiversity. *Nature Reviews Genetics* 20(10):615-628.
- Cano-Gamez E., Trynka G., 2020 From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. *Frontiers in Genetics* 11:424.
- Dang X., Xia Y., Xu Q., Zhang J., 2016 The complete mitochondrial genome of the *Poecilia formosa* (Amazon molly). *Mitochondrial DNA Part A* 27(5):3523-3524.
- Eastis A. N., Fast K. M., Sandel M. W., 2021 The complete mitochondrial genome of the variable platyfish *Xiphophorus variatus*. *Mitochondrial DNA Part B* 6(9):2640-2642.
- Eichler E. E., 2019 Genetic variation, comparative genomics, and the diagnosis of disease. *New England Journal of Medicine* 381(1):64-74.
- Fast K. M., Rakestraw A. W., Sandel M. W., 2023 Complete mitochondrial genome of a livebearing freshwater fish (Cyprinodontiformes: Poeciliidae): *Poecilia parae*. *Mitochondrial DNA B Resour* 8(2):215-219 [in press].
- Fleischmann R. D., Adams M. D., White O., Clayton R. A., Kirkness E. F., Kerlavage A. R., et al, 1995 Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269(5223):496-512.

- Green E. D., Gunter C., Biesecker L. G., Di Francesco V., Easter C. L., Feingold E. A., et al, 2020 Strategic vision for improving human health at The Forefront of Genomics. *Nature* 586(7831):683-692.
- Goffeau A., Barrell B. G., Bussey H., Davis R. W., Dujon B., Feldmann H., Galibert F., Hoheisel J. D., Jacq C., Johnston M., Louis E. J., Mewes H. W., Murakami Y., Philippsen P., Tettelin H., Oliver S. G., 1996 Life with 6000 genes. *Science*. 25:274(5287):546, 563-7. doi: 10.1126/science.274.5287.546. PMID: 8849441.
- Hidalgo O., Pellicer J., Christenhusz M., Schneider H., Leitch A. R., Leitch I. J., 2017 Is there an upper limit to genome size? *Trends in Plant Science* 22(7):567-573.
- Jones J. B., Song J. J., Hempen P. M., Parmigiani G., Hruban R. H., Kern S. E., 2001 Detection of mitochondrial DNA mutations in pancreatic cancer offers a "mass"-ive advantage over detection of nuclear DNA mutations. *Cancer Research* 61(4):1299-1304.
- Künstner A., Hoffmann M., Fraser B. A., Kottler V. A., Sharma E., Weigel D., Dreyer C., 2016 The genome of the Trinidadian guppy, *Poecilia reticulata*, and variation in the Guanapo population. *PLoS One* 11(12):e0169087.
- Lunshof J. E., Bobe J., Aach J., Angrist M., Thakuria J. V., Vorhaus D. B., et al, 2010 Personal genomes in progress: from the human genome project to the personal genome project. *Dialogues in Clinical Neuroscience* 12:47-60.
- Luo X., Zhou Y., Zhang B., Zhang Y., Wang X., Feng T., et al, 2020 Understanding divergent domestication traits from the whole-genome sequencing of swamp-and river-buffalo populations. *National Science Review* 7(3):686-701.
- Noble D., Hunter P., 2020 How to link genomics to physiology through epigenomics. *Epigenomics* 12(4):285-287.
- Nowoshilow S., Schloissnig S., Fei J. F., Dahl A., Pang A. W., Pippel M., et al, 2018 The axolotl genome and the evolution of key tissue formation regulators. *Nature* 554(7690):50-55.
- Papuc T., Balint C., Petrescu-Mag R. M., 2022 Why is *Poecilia wingei* a distinct species from other species of the genus? *Poec Res* 12(1):29-33.
- Petrescu-Mag R. M., Pășărin B., Șonea C. G., Petrescu-Mag I. V., 2013 Customer preferences and trends for aquarium fish in Transylvania (Romania). *North-West J Zool* 9(1):166-171.
- Proorocu M., Petrescu-Mag I. V., 2022 Axolotls and neoteny: those who forgot to grow up. *ELBA Bioflux* 14(1):12-16.
- Schartl M., Walter R. B., Shen Y., Garcia T., Catchen J., Amores A., et al, 2013 The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nature Genetics* 45(5):567-572.
- Schoch C. L., Ciufo S., Domrachev M., et al, 2020 NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database (Oxford)*. 2020:baaa062. PubMed: 32761142 PMC: PMC7408187.
- Venter J. C., Adams M. D., Myers E. W., Li P. W., Mural R. J., Sutton G. G., et al, 2001 The sequence of the human genome. *Science* 291(5507):1304-1351.
- Yazgan O., Krebs J. E., 2012 Mitochondrial and nuclear genomic integrity after oxidative damage in *Saccharomyces cerevisiae*. *Front Biosci* 17:1079-1093.
- Zhang K., Liu S., Chen D. Y., Chen X., Zhang G. Y., 2016 The complete mitochondrial genome of the Monterrey platyfish (*Xiphophorus couchianus*). *Mitochondrial DNA A DNA Mapp Seq Anal* 27(6):4453-4454.
- *** www.ourworldindata.org [Last view: 23.10.2022].

Received: 07 December 2022. Accepted: 19 December 2022. Published online: 30 December 2022.

Authors:

George-Cătălin Muntean, Faculty of Animal Science and Biotechnologies, University of Agricultural Sciences and Veterinary Medicine, Cluj-Napoca, Romania, e-mail: muntean.gc@gmail.com

Marian Proorocu, University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, Faculty of Agriculture, Department of Environmental Engineering and Protection, Calea Mănăştur 3-5, 400372, Cluj-Napoca, Romania, e-mail: mproorocu@yahoo.com

Ioan Valentin Petrescu-Mag, University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, Faculty of Agriculture, Department of Environmental Engineering and Protection, 3-5 Calea Mănăştur Street, 400372 Cluj-Napoca, Romania, e-mail: zoobiomag2004@yahoo.com

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

How to cite this article:

Muntean G.-C., Proorocu M., Petrescu-Mag I. V., 2022 Poeciliid fishes in the genomic era. *Poec Res* 12(1):46-52.