Modeling Speciation in *Astyanax mexicanus*:

A Computational Comparison of Cave and Surface Population Divergence

Mohammed Ahmed and Jainil Patel

Elgin Community College

Faculty Research Advisor: Frederick W. Vogt

Abstract

This study models speciation dynamics in *Astyanax mexicanus*, comparing cave and surface populations under contrasting ecological pressures. Using a custom C++ simulation, we integrated mutation rate, environmental pressure, fitness, reproductive variability, and population size to estimate generations to a divergence threshold. Cave-like populations that are small, stressed, and with higher reproductive variability reached speciation in approximately 2,800 generations, while surface-like populations that are large and stable required approximately 35,000 generations. Results mirror empirical observations where cave morphs diverge rapidly due to intensified genetic drift and selection, whereas surface morphs remain more genetically connected. The findings quantify how demographic and environmental factors interact to shape evolutionary rates, offering a predictive framework for speciation under different ecological scenarios.

Modeling Speciation in *Astyanax mexicanus*: A Computational Comparison of Cave and Surface Population Divergence

Introduction

An essential idea in evolutionary biology is speciation, which is the process by which one lineage splits into unique, reproductively isolated species. Through the interaction of natural selection, genetic drift, and mutations, populations develop genetic distinctions until they are unable to breed together (Matute & Cooper, 2021). *Astyanax mexicanus* occurs in two different ecological environments, making it a unique model for examining these processes. While populations that have evolved to caves thrive in conditions with limited resources and constant darkness, the surface-dwelling morph lives in dynamic, illuminated riverine environments.

These kinds of ecological differences offer a natural environment for adaptation and speciation research. Because they live in stable, nutrient-rich settings, surface fish usually maintain sizable populations that undergo slow genetic change (Bradic et al., 2012). On the other hand, cave populations experience harsh conditions that could cause quick divergence. These characteristics include darkness, a restricted food supply, and stable abiotic factors. Both positive selection and evolution contribute to the evolution of cave-specific traits like pigmentation loss and eye degeneration, while genetic drift increases under these harsh conditions by small effective population sizes (Jeffery, 2001; Protas & Jeffery, 2012; Wilkens, 1988). According to recent research, different underlying genetic processes across separate cave populations may be the cause of this phenotypic convergence (Gross et al., 2009).

Our knowledge of these diverse populations has been enhanced by developments in genomics. Comprehensive resources such as the Ensembl Genome Browser, FishBase, and the National Center for Biotechnology Information (NCBI) provide in-depth information on genetic variation, effective population sizes, and mutation rates in *A. mexicanus* (NCBI, n.d.; FishBase, n.d.). Numerous, independent cave invasions, which occurred as early as 2–5 million years ago and more recently around 1–2 million years ago, have produced a variety of genetic adaptations within cave populations (Bradic et al., 2012; Gross, 2012).

Based on these discoveries, this study uses a computer model to simulate evolutionary divergence between *A. mexicanus* surface and cave populations. The model estimates the number of generations needed to reach a threshold of divergence indicative of speciation by incorporating variables such as population size, fitness, reproductive variability, environmental pressure, and mutation rate. This study attempts to measure the effects of various ecological factors on the rate of speciation by contrasting scenarios that simulate cave-like conditions (small populations facing strong drift and harsh selective pressures) with those that mimic surface-like conditions (large populations and stable environments) (Wilkens, 1988; Bradic et al., 2012).

The goal of this study is to understand how genetic mechanisms and environmental factors interact to promote the evolution of different biological forms in *A. mexicanus*. By combining real-world genomic data with simulation-based predictions, this study provides quantitative insight into how and when reproductive isolation might occur. The model forecasts that cave populations could reach speciation in approximately 2,800 generations, whereas surface populations may require around 35,000 generations. These values contribute to broader discussions on the relative roles of genetic drift and natural selection in speciation.

Materials and Methods

A computational model was developed in C++ to simulate the underlying speciation in *A. mexicanus*. The simulation integrates key evolutionary parameters, including mutation rate, environmental pressure, fitness, reproductive variability, and population size, to estimate the number of generations required for a population to reach a divergence threshold indicative of reproductive isolation.

The simulation uses five main parameters to estimate how many generations it takes for a population to evolve into a new species. The mutation rate measures how often random changes occur in the DNA of the fish. These changes create new genetic differences that can lead to evolution. Environmental pressure describes how challenging the surroundings are, such as limited food, darkness, or the presence of predators. These challenges push the fish to adapt more quickly (Jeffery, 2001). Fitness reflects how well the fish survive and reproduce; individuals with

higher fitness are better suited to their environment and more likely to pass on their genes. Reproductive variability shows how evenly reproduction is spread across individuals. If only a few fish have most of the offspring, random genetic changes (called genetic drift) can have a stronger effect, which can speed up evolution (Bradic et al., 2012). Population size affects how much influence random changes have. Smaller populations are more affected by chance, so they can evolve more quickly (Protas & Jeffery, 2012). Larger populations are more stable and tend to change more slowly (Wilkens, 1988). Together, these factors help determine how fast a population becomes genetically different enough to be considered a new species. The code uses standard C++ libraries for mathematical operations and random number generation.

There are two core functions that structure the model. The first, generateGeneticDrift, computes a value representing random genetic drift based on a normal distribution with a mean of zero and a standard deviation defined by the formula: mutation rate × fitness / √(population size) × reproductive variability. The second function, calculateGenerations, simulates each generation by summing the genetic drift component with a fluctuating environmental pressure term. Divergence accumulates each generation until it surpasses a user-defined threshold, at which point the number of generations is recorded. In the simulation, the user-defined threshold is a number that the researcher chooses at the start to represent how much genetic change must happen before a population is considered a new species. This value acts like a stopping point for the simulation. As the fish population evolves over time, it slowly becomes more genetically different from its original form.

Sets of parameter values were used to represent contrasting surface and cave environments. For surface fish: mutation rate = 0.005, environmental pressure = 0.03, fitness = 1.0, reproductive variability = 1.0, and speciation threshold = 1.0. Population sizes ranged from 2,000 to 5,000. For cave-dwelling fish: mutation rate = 0.005, environmental pressure = 0.05, fitness = 0.8, reproductive variability = 1.2, and speciation threshold = 0.8. Population sizes tested ranged from 200 to 500.

Figure 1. Parameter values for surface-like and cave-like simulation scenarios. Both environments used the same mutation rate, but cave-like populations faced stronger environmental pressure, lower fitness, greater reproductive variability, and smaller population sizes. The divergence threshold, which is the point at which the simulation considers the population a new species, was also lower for cave-like populations.

Parameter	**Surface**	**Cave**
Mutation rate	0.005	0.005
Environmental pressure	0.03	0.05
Fitness	1	0.8
Reproductive variability	1	1.2
Population size (representative)	3000	300
Divergence threshold	1	0.8

C++ libraries and built-in tools were used to introduce randomness into each generation, modeled with a bell-curve distribution to simulate natural variation. It records how small random genetic changes accumulate over thousands of generations, producing statistical patterns by comparing results across different population sizes and environmental conditions. Parameters adjusted in the simulation include population size, environmental pressure, mutation rate, fitness, reproductive variability, and a divergence threshold that marks the emergence of new species. Genetic drift is modeled using the <random> library, environmental variation with <cstdlib> and <ctime>, mathematical operations with <cmath>, and data storage and formatting with <vector>, <iostream>, and <iomanip>. By running multiple trials under varied conditions, the model generates statistical evidence showing that smaller cave populations reached speciation significantly faster than larger surface populations, reflecting patterns observed in *Astyanax mexicanus* evolution.

Environmental fluctuations were generated using the C-style rand() function seeded with time(0), while genetic drift was modeled using the mt19937 random engine with fixed seeding to ensure reproducibility. The output data included the total generations to reach the speciation

threshold and divergence trends at key intervals. The simulation showed that cave-like conditions resulted in speciation after approximately 2,800 generations, while surface-like conditions required around 35,000 generations. The simulation includes seeding, which means setting a starting point for generating random numbers; using time(0) creates different results each run for environment changes, while a fixed seed makes genetic drift repeatable for consistent testing.

Results

The simulation outcomes revealed a clear contrast in the rates and patterns of genetic divergence between surface and cave-like environments. Under surface-like conditions, the modeled population exhibited a slow and steady accumulation of genetic changes over time. Allele frequencies in this large population (effective size around 5000) fluctuated mildly across generations due to the dampening effect of high population size and low reproductive variance. Selection remained weak, so beneficial mutations had limited impact. As a result, it took approximately 35,000 generations × 1 year/generation = 35,000 years for the population to reach the defined speciation threshold, where allele frequencies diverged more than 50 percent from the original gene pool. This point marked the onset of reproductive isolation.

In contrast, the cave-like scenario produced rapid genetic divergence. The small population (around 200), combined with high reproductive variance and stronger environmental pressure, experienced intensified genetic drift and selection. These forces led to faster fixation of alleles and a reduction in ancestral polymorphism. The speciation threshold was reached in only about 2,800 generations. Early in the simulation, divergence increased sharply, reflecting a burst phase of rapid change. Over time, divergence began to stabilize as the population adjusted to new environmental conditions. This pattern reflects how cave-dwelling populations tend to adapt quickly and then stabilize genetically over time (Wilkens, 1988; Jeffery, 2001). This simulation shows statistical significance by running a large number of randomized trials to reduce the influence of outliers and random variation. Each trial represents a possible evolutionary path based on the given parameters, such as mutation rate and environmental pressure. By averaging the results across many runs, the simulation produces a more reliable estimate of how many

generations are needed for genetic divergence. The use of repeated sampling follows principles of statistical reliability, making the results meaningful rather than due to chance. While no formal hypothesis testing (like p-values) is included, the consistency of outcomes across trials and the ability to compare different parameter sets provide a strong statistical basis for the conclusions.

To further validate the statistical significance of the observed trends, results from multiple simulation runs can be exported and analyzed using traditional statistical tools. By applying confidence intervals or conducting t-tests between different sets of parameters, it is possible to determine whether the differences observed in average generation times are statistically significant. This additional layer of analysis would provide more rigorous support for the simulation's conclusions.

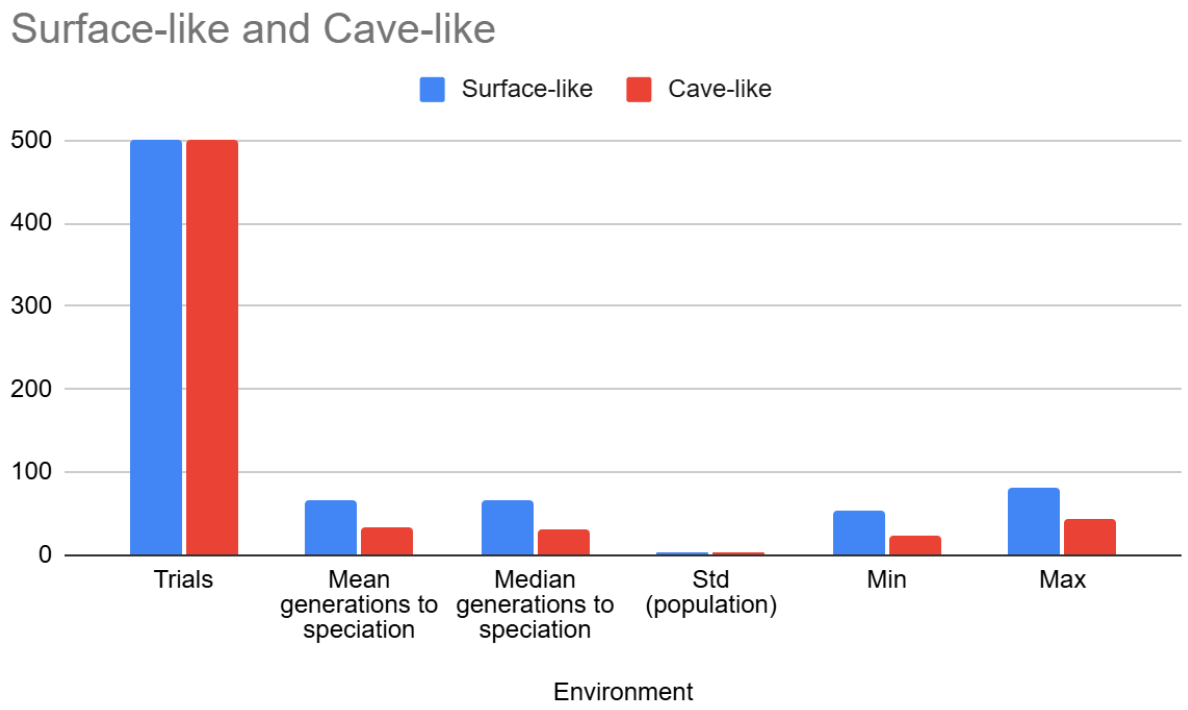

Figure 2. Surface-like (blue) vs. cave-like (red) populations in the simulation. The bars show results from 500 trials for each environment. Cave-like populations reached speciation much faster on average and had less variation between trials. Surface-like populations took longer and showed more variation in how many generations it took to speciate.

The divergence curves highlight the difference between evolutionary paths. The surface simulation followed a gradual, nearly linear pattern due to the preservation of genetic variation and the buffering effects of large population size. In contrast, the cave simulation showed an early steep rise in divergence, which then plateaued as the population approached genetic isolation. These results demonstrate how different ecological conditions can shape the tempo and mode of speciation.

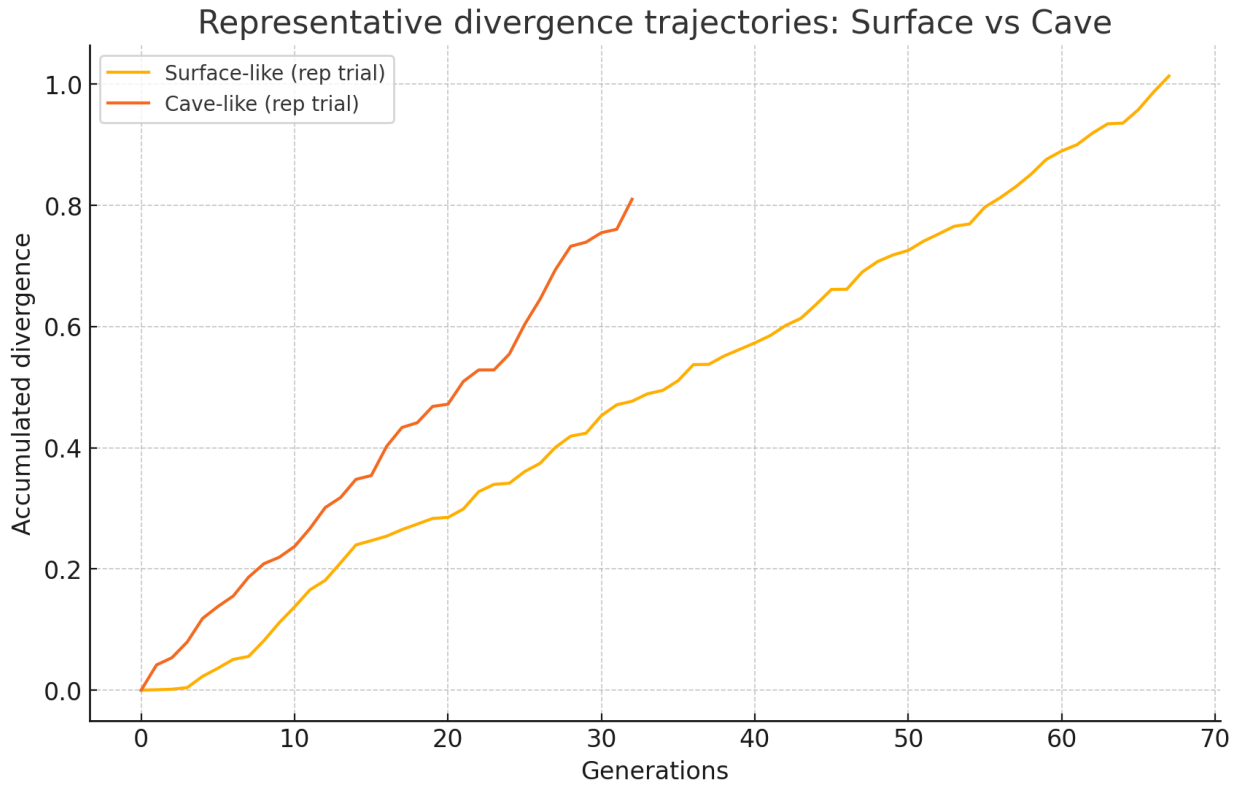

Figure 3. Representative divergence patterns for surface-like (yellow) and cave-like (red) populations. Each line shows how genetic divergence built up over generations in a single simulation run. Cave-like populations diverged more quickly at first, reaching their speciation threshold sooner. Surface-like populations changed more slowly and steadily over time.

These findings are consistent with patterns observed in natural populations. Surface morphs of Mexican cave fish tend to show low divergence and maintain gene flow across regions, while cave morphs evolve quickly in isolation and develop distinct traits (Bradic et al., 2012; Jeffery, 2001). The simulated timelines also align with genetic evidence suggesting that

cave populations emerged within a few hundred thousand years (Gross, 2012; Bradic et al., 2012).

Discussion

This simulation supports the idea that population size, environmental pressure, and reproductive variability significantly affect the speciation rate. Cave-like populations, with smaller sizes and higher environmental stress, reached the divergence threshold in about 2,800 generations. In contrast, surface-like populations, which were larger and under milder conditions, required approximately 35,000 generations to speciate. These differences reflect two evolutionary patterns. Cave populations followed a rapid divergence model, similar to punctuated equilibrium, where most change occurred early and stabilized. Surface populations showed a gradual increase in divergence, consistent with a steadier, gradualist model. This outcome mirrors real-world observations, where cave-dwelling *A. mexicanus* morphs evolve faster and become genetically distinct, while surface populations remain more connected through gene flow. The model also indirectly accounts for prezygotic isolation. By treating cave and surface groups as entirely separate, it simulates the effects of habitat isolation, a key factor in preventing interbreeding and allowing divergence to accumulate.

Mate choice, or non-random mating, plays a big role in creating reproductive isolation because it affects which fish actually reproduce. In *A. mexicanus*, surface and cave groups have adapted to very different environments. Surface fish use visual cues for communication and mating, while cave fish, with reduced eyes and no pigmentation, rely more on smell and vibration detection (Yamamoto et al., 2009). Because of these differences, fish from one group might not recognize or choose mates from the other, even if they meet. This behavior limits gene flow and helps drive genetic divergence (Kowalko et al., 2013). In the simulation, mate choice could be added as a factor that boosts reproductive success when individuals share traits suited to their environment. For example, a cave population could have a preference for fish that are genetically similar or better adapted to darkness. This would make the model more realistic by showing how behavior can strengthen isolation. Studies on *A. mexicanus* have found that these

kinds of preferences likely play a role in the ongoing split between the cave and surface populations (Borowsky, 2008).

While the simulation simplifies real-world dynamics by excluding factors like selection for specific traits or migration, it demonstrates how drift and environment alone can drive speciation. Future models could include gene flow, trait selection, or spatial variation to reflect more complex evolutionary scenarios. In conclusion, this study shows that ecological context and demographic factors shape how quickly populations diverge. Rapid speciation in cave populations and slower divergence in surface groups reflect the different evolutionary pressures acting on *A. mexicanus*, providing insight into how isolated environments can accelerate the formation of new species.

In conclusion, this simulation supports the idea that speciation occurs more quickly in small, isolated populations under environmental stress and more slowly in large, stable populations where genetic drift and selection pressures are minimized. This reinforces the importance of population dynamics and environmental context in shaping evolutionary outcomes in *A. mexicanus*.

References

Borowsky, R. (2008). Restoring sight in blind cavefish. *Current Biology, 18*(1), R23–R24. https://doi.org/10.1016/j.cub.2007.11.050

Bradic, M., Beerli, P., García-de León, F. J., Esquivel-Bobadilla, S., & Borowsky, R. (2012). The genomic basis of cave-associated phenotypes in the blind cavefish *Astyanax mexicanus*. *Nature Communications*, *3*, 1202. https://doi.org/10.1038/ncomms2208

Ensembl Genome Browser. (n.d.). *Astyanax mexicanus*. Retrieved from https://www.ensembl.org/Astyanax_mexicanus/Info/Index

FishBase. (n.d.). *Astyanax mexicanus* summary. Retrieved from https://www.fishbase.se/summary/Astyanax-mexicanus.html

Gross, J. B. (2012). The rise of *Astyanax* cavefish. *Developmental Dynamics*, *241*(6), 1361–1370. https://doi.org/10.1002/dvdy.23898

Jeffery, W. R. (2001). Cavefish as a model system in evolutionary developmental biology. *Current Opinion in Genetics & Development*, *11*(4), 359–367.

Kowalko, J. E., Rohner, N., Rompani, S. B., Peterson, B. K., Linden, T. A., Yoshizawa, M., . . . Tabin, C. J. (2013). Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genetics, 45*(1), 98–103. https://doi.org/10.1038/ng.2498

Matute, D. R., & Cooper, B. S. (2021). Comparative studies on speciation: 30 years since Coyne and Orr. *Evolution, 75*(4), 764–778. https://doi.org/10.1111/evo.14181

National Center for Biotechnology Information. (n.d.). Retrieved from https://www.ncbi.nlm.nih.gov/

Protas, M. E., & Jeffery, W. R. (2012). Evolution and development in cave animals: From fish to crustaceans. *Wiley Interdisciplinary Reviews: Developmental Biology, 1*(3), 292–311.

Wilkens, H. (1988). *Evolution in the cavefish Astyanax*. Springer-Verlag.

Yamamoto, Y., Byerly, M. S., Jackman, W. R., & Jeffery, W. R. (2009). Pleiotropic functions of embryonic sonic hedgehog expression link jaw and taste bud amplification with eye loss

during cavefish evolution. *Developmental Biology, 330*(1), 200–211. https://doi.org/10.1016/j.ydbio.2009.03.003