

Article

# A Procedure for Modeling Genetic Diversity Distortions in Populations of Organisms with Mixed Reproductive Strategies

Anastasiya Poroshina \*  and Dmitry Sherbakov 

Laboratory of Molecular Systematics, Limnological Institute, Siberian Branch of Russian Academy of Sciences, 664033 Irkutsk, Russia; sherb@lin.irk.ru

\* Correspondence: a.poroshina@lin.irk.ru

**Abstract:** We propose an approach for modeling the pattern of the genetic diversity of microsatellite markers in a population with a mixed breeding strategy. Part of the population is reproduced sexually, and part is produced asexually. The method of the proposed simulation is different from others in that it produces a set of microsatellite markers as the outcome of a computer simulation of processes in a fixed-size population. These markers can be utilized with the assistance of available software to calculate various metrics of genetic diversity. Our approach is implemented in Python 3.10 and is accompanied by additional scripts that ensure result compatibility with programs that calculate different population characteristics.

**Keywords:** individual-oriented modeling; reproductive mode; genetic diversity; population genetics

**MSC:** 92D15



**Citation:** Poroshina, A.; Sherbakov, D. A Procedure for Modeling Genetic Diversity Distortions in Populations of Organisms with Mixed Reproductive Strategies. *Mathematics* **2023**, *11*, 2985. <https://doi.org/10.3390/math11132985>

Academic Editors: Sergey A. Lashin and Takashi Suzuki

Received: 2 May 2023

Revised: 22 June 2023

Accepted: 27 June 2023

Published: 4 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The prevailing models of genetic processes in a population typically describe the dependence of the physical number of organisms  $N(x, t, \omega)$  on factors such as geographical coordinates ( $x$ ), time ( $t$ ), as well as population processes including reproductive strategies or changes in environmental capacity ( $\omega$ ). These models are very useful for understanding population processes, but their common problem lies in the difficulty of verifying and comparing modeling results to real-life data. Therefore, modeling results are of limited use for predicting the most probable population scenarios, resulting in the current population's genetic makeup. To overcome this problem, an alternative micro-evolutionary modeling approach was proposed. It involves modeling genetic processes as a collection of "nucleotide" sequences ( $D = D(x, t, \omega)$ ) that evolve during the simulation through random non-adaptive mutations (base substitutions) [1,2]. This approach enables the observation of the changes in molecular diversity patterns resulting from the modeled evolutionary process [3]. Consequently, the sets of "nucleotide" sequences  $D$  may be used to calculate population parameters through any metrics of diversity, and thus, compare them to existing population models. Particularly, population parameter  $\theta$  may be obtained from  $D$ , and thus, compared to the  $N(x, t, \omega)$  models.

Recent advancements in molecular population studies have made microsatellite repeats [4] and single nucleotide polymorphisms [5,6] even more important than the molecular sequences for which the previously proposed evolutionary modeling approach was initially developed. Therefore, there is a need to expand this approach to include new genetic markers and consider their unique properties beyond nucleotide sequences [7]. Specifically, we aim to model distortions of genetic diversity patterns in response to the biological properties of a population using microsatellite markers. This requires their distinctive features to be accounted for, such as the mostly independent segregation and the significantly larger and varying number of alleles in each locus, unlike the only four possible states in each position of a nucleotide sequence.

With the emergence of several powerful applications, obtaining an ever-expanding range of population descriptive statistics has become possible [8–11]. When combined with efficient models of various and complex scenarios that shape the current genetic diversity in populations, these applications allow for the comparison of different explanations for recent genetic polymorphisms and allow us to choose the most probable explanation among them using approximate Bayesian calculations [12] as implemented in software such as R [13].

Therefore, this paper describes a novel set of algorithms implementing  $M(x, t, \omega)$  models of populations of organisms that are capable of switching between sexual and asexual reproduction. We expand the previously developed functionality to include microsatellite markers and demonstrate that these models can address new questions about the genetic consequences of such processes.

The fine differences between genetic consequences of the reproductive mode are numerous, complex and very deep [14,15]. Naturally, many efforts were made to describe the genetic diversity of asexual, sexual and mixed populations [16–19].

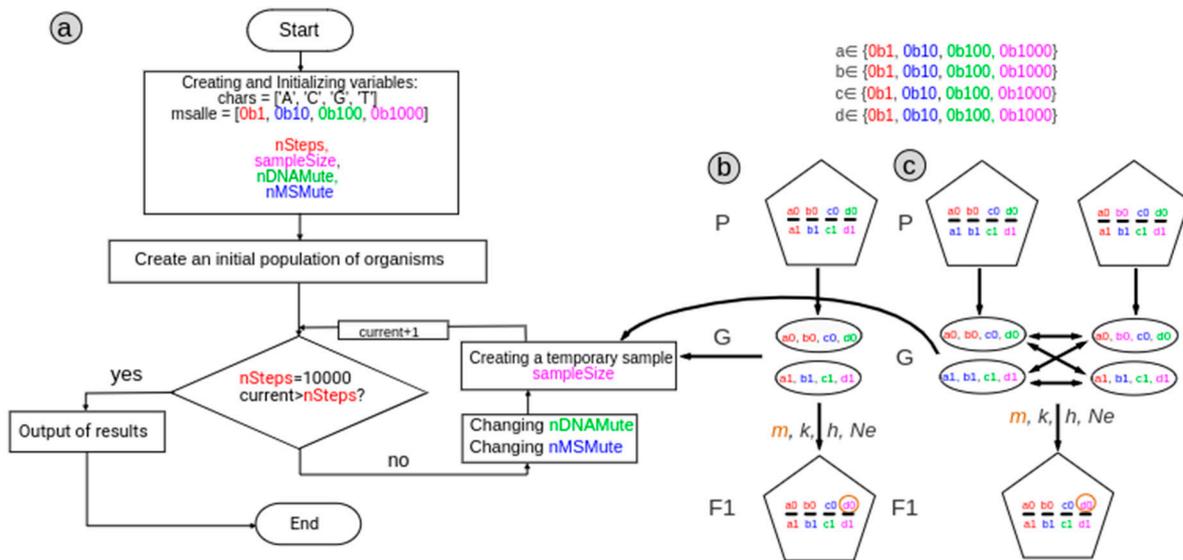
## 2. Materials and Methods

Individual-oriented modeling (IOM) [20] is a software development methodology based on the concept of objects. It allows for the modeling of real-world objects and processes using classes and their interactions. The main concepts of IOM are objects, classes, inheritance, encapsulation and polymorphism [21]. Objects are instances of classes that contain data and methods for working with those data. Classes are templates for creating objects that define a set of attributes and methods that will be present in all objects of that class. Inheritance is a mechanism that allows for new classes to be created based on existing ones by inheriting their properties and methods. Encapsulation is the principle that an object's data and methods are hidden from other objects and can only be accessed through specific methods. Polymorphism is the ability of objects from different classes to have methods with the same name but a different behavior [22].

IOM simplifies the design and development of complex systems, increases their flexibility, and facilitates maintenance. IOM is used in various fields such as software development, database design and business process modeling. There are different programming languages that support the individual-oriented paradigm, such as Java, C++, Python and Ruby. A common feature of these languages is the support for inheritance, encapsulation and polymorphism. In our work, we use the Python programming language and packages such as NumPy and Pandas.

IOM methods can differentiate reproduction strategies by representing them as separate classes or objects, each with their own set of attributes and behaviors. This method allows for the system to be represented as a set of objects. In the context of the reproduction of living organisms, this approach allows for different strategies to be defined, namely, the following:

1. Asexual reproduction strategy (Figure 1b) (without mixing genetic material). Within this strategy, new individuals are inherited from their parents and are genetically identical. These organisms do not possess a great adaptability to changing environmental conditions.
2. Reproduction strategy (Figure 1c) with mixing of genetic material (sexual reproduction). This strategy allows for the creation of offspring with a unique genetic code, increasing the adaptability and survival of organisms in changing environmental conditions. For example, a species that reproduces sexually can be modeled as an object with attributes such as gender, genetic information and behavior during mating.



**Figure 1.** Figure 1. (a) Flow chart of the base model of population. (b,c) Dive details on the marker inheritance differences between asexual (b) and sexual reproduction (c).  $N_e$ —effective population size,  $m$ —mutation probability,  $k$ —number of offspring,  $h$ —probability of sexual reproduction.

In general, IOM methods can provide a flexible and powerful way to differentiate between different reproduction strategies in a population, allowing for a more detailed and accurate representation of the modeled biological system.

IOMs used in this paper are written in Python ver. 3.10. Auxiliary scripts used for data analysis and for data transformation into different formats are written in Python ver. 3.10 and R version 4.2.1. All scripts are made generally available at <http://github.com/AnastasiyaPoroshina/apomix> (accessed on 29 April 2023).

The flow chart of the base model used here is given in Figure 1a, and the variants of the scheme for asexually- and sexually-reproducing organisms are given in Figure 1b,c.

### 3. Results and Discussion

In this study, we expand on a previously described modeling approach to enable the use of microsatellite markers for tracing the pattern of population genetic diversity. The main novelty of this model is that it produces a pattern of microsatellite haplotypes at any moment of the simulation, rather than simulating the effect of different circumstances on the physical number of organisms at a given time. The obtained output is in a format that is suitable for minor transformations to any software used to estimate various population parameters such as the Essential Biodiversity Variables [23]. Accompanying scripts may then transform the results into formats that are directly suitable for different software used to calculate the EBV parameters, including genetic diversity, genetic differentiation  $F_{ST}$ , inbreeding  $F_I$ , effective population size  $N_e$  and more sophisticated metrics like average linkage disequilibrium  $D$ .

One such metric is the frequency of recombination events, which can be used as an indicator of the relative contribution of sexual and asexual reproduction to the gene pool of the population [16,19,24,25]. For example, if recombination events are relatively rare, this suggests that asexual reproduction is dominant and genetic diversity is primarily generated through mutation [26]. Conversely, if recombination events are frequent, this suggests that sexual reproduction is dominant and genetic diversity is primarily generated through recombination.

Another metric is the level of heterozygosity in the population. Heterozygosity is a measure of genetic diversity that reflects the presence of different alleles at a given gene locus. A population that reproduces primarily through sexual reproduction is expected to

have higher levels of heterozygosity, as a result of the shuffling of genetic material during meiosis and fertilization [3,26].

Finally, the presence of clonal lineages can be used as an indicator of the dominance of asexual reproduction in the population. Clonal lineages are groups of individuals that are genetically identical or very similar and are generated through asexual reproduction. The proportion of clonal lineages in the population can be used as a rough estimate of the frequency of asexual reproduction [27].

In summary, the coexistence of sexual and asexual reproduction in a population can be assessed through a range of genetic metrics, including recombination frequency, heterozygosity levels and the presence of clonal lineages. Such assessments can help us understand the role of the reproductive mode in shaping genetic diversity and adaptive potential in natural populations.  $D$  is the one of simplest metrics.

In order to compare the simulations of  $M(x, t, \omega)$  to our previous simulations (Semovsky et al., 2004) every object (organism) along with “microsatellites” bear one mitochondrial DNA marker (maternally inherited), and in the case of a diploid organism, two “nuclear markers” are also present. This information, naturally, may be used to independently calculate the same metrics of the genetic diversity of the population (except for  $D$ ).

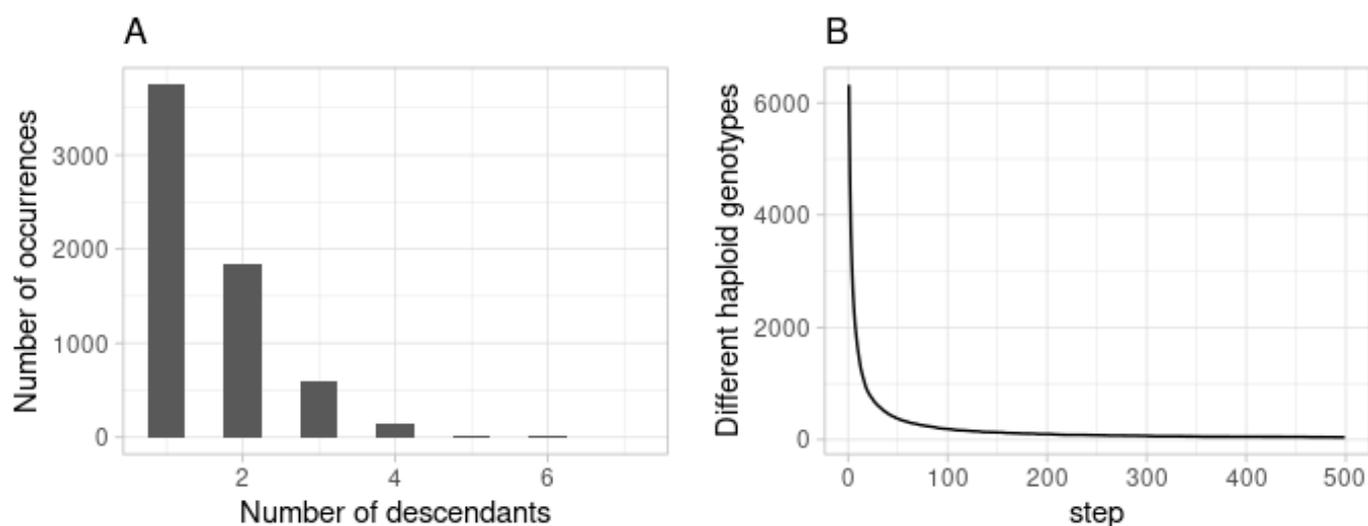
Microsatellite haplotypes (single-copy genotypes) are represented as binary numbers, where the number of positions is defined by the number of alleles (variants) in each marker. For example, for a marker of 3 possible alleles, the alleles will be 100, 010 and 001. Since we aim mostly at the diploid organisms bearing two copies of each marker, their genotypes may be represented as  $x_1 + x_2 * 2^n$ , where  $x_1$  and  $x_2$  are haplotypes and  $n$  is a number of haplotypes possible for this marker. The ordinary genetic notation for this would be  $\frac{x_1}{x_2}$ . This enables us to use binary arithmetic to count haplotypes and genotypes for each marker and for any of their combinations. This substantially increases the speed of calculations.

The base model consists of a constant number of diploids, potentially hermaphrodite organisms, which may choose their mode of reproduction randomly. The behavior of DNA markers is as described by Semovski, Sherbakov et al. [1,2]. The microsatellite extension was added as follows:

- A single population has a default starting size of 10,000 objects (organisms);
- Each organism reproduces once in its life, producing a random number of descendants by following one of the two reproductive strategies (asexual and sexual);
- Asexual reproduction is a simple copying of the parent object;
- Sexual reproduction is a combination of haploid genomes of randomly drawn parents. In each parent, each marker (locus) is represented by an independently and randomly drawn copy (Figure 1c);
- The mutation of a microsatellite marker is a random transition between different possible states (alleles) of a single copy of the marker. In other words, there is no linkage between microsatellite loci.

The size of the population remains the same in the base model, though it can be changed by using “fatefile” consisting of one column of numbers (population sizes) changing in a linear fashion. The length of the column should be equal to the length of a simulation.

The essential properties of the base model are illustrated in Figure 2. The two graphs illustrate the properties of an example model on which the speed of the model response to the conditions depends. The most important property is the diversity of the unspoiled population. Since, according to both reproductive strategies, the parental genotypes are defined by a random draw with replacement, the exact distribution of the number of siblings depends on the population size. The default population consisting of 104 individuals is shown in Figure 2A. Indeed, the genetic diversity, as measured by the number of different haplotypes regardless of the number of each allele, rapidly drops (Figure 2B). Therefore, for a population of 104 organisms, the haplotype diversity after approximately 100th is mostly defined by the mutation process and, indeed, by the mode of reproduction.



**Figure 2.** (A) Distribution of number of descendants in a fully asexual population of  $10^4$  organisms. (B) Diversity depletion in a fully asexual population of  $10^4$  organisms.

The mutation process abides by the following rules:

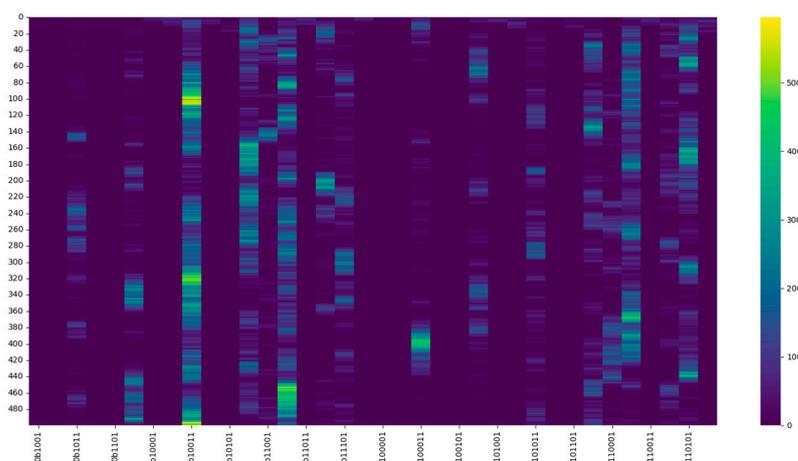
- All variation is restricted to the initially defined set of possible alleles;
- All markers mutate independently;
- All possible genotypes are equally viable (no selection).

The current version of the model assumes that the reproductive mode is not inherited. Instead, in each generation, a subset of individuals that will engage in sexual reproduction is randomly drawn from the population, while the remaining organisms are allowed to reproduce asexually. This allows for the ratio of sexual/asexual strategies to be changed during a simulation and provides a more realistic simulation of natural populations. This feature is similar to the “clonal correction” method proposed by Kamvar et al. [10], which helps to correct the sampling bias in asexual populations. However, in the current version of the model, it also allows for the study of the effects of varying proportions of sexual versus asexual reproduction on genetic diversity and evolution.

A result of a simulation performed for a population of 10,000 organisms with a 0.33 probability to take part in a sexual reproduction for 500 steps is given in Figure 3. Each step is plotted as a horizontal array of bars, and the time flows along the vertical axis. The intensity of a bar is proportional to the number of copies of a given genotype, which is indicated on the horizontal axis following the binary notation as described above. It gives an idea about the changes of dominating genotypes and the succession of the dominants. It can be seen that at each generation, all the base characteristics of the population as well as a number of other metrics may be estimated. This base model may be modified by changing it according to the assumed scenario and/or the by defining the changes in the sexual/asexual ratio, which models seasonal shifts that are typical to many organisms adapting to seasonal fluctuations, and may be used to examine if this causes significant changes in their genetic diversity.

This base model can be modified by changing the assumed scenario or by defining the changes in the sexual/asexual ratio. For example, seasonal shifts in reproductive strategies are common in many organisms adapting to environmental fluctuations. Modifying the model to reflect such shifts could help to examine whether they lead to significant changes in genetic diversity over time. The choice of the metrics used for the calculation as well as the application of the approach to model organisms will be published elsewhere.

Overall, the flexibility of the model allows for a wide range of scenarios to be tested and can help to provide insights into the basic mechanisms driving evolution and genetic diversity in populations of animals that are able to regulate the prevalence of sexual or asexual reproduction.



**Figure 3.** Graphic presentation of the base model performance at 33% of sexually reproducing organisms. All organisms were sorted by their genotypes. Time goes from top to bottom. Color is proportional to the number of occurrences of haplotype (single set of markers). The haplotypes are given at the bottom in binary notation as described in text.

**Author Contributions:** Conceptualization, A.P. and D.S.; Methodology, A.P. and D.S.; Software, D.S.; Validation, A.P.; Formal analysis visualization, A.P.; Investigation, A.P.; Resources, A.P.; Data curation, D.S.; Writing—original draft, D.S.; Writing—review & editing, A.P.; Supervision, D.S.; Project administration, D.S. and A.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Russian Science Foundation grant number [22-24-00791].

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data sharing not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Semovski, S.; Verheyen, E.; Sherbakov, D. Simulating the evolution of neutrally evolving sequences in a population under environmental changes. *Ecol. Model.* **2004**, *176*, 99–107. [\[CrossRef\]](#)
2. Semovski, S.V.; Sherbakov, D.Y.; Verheyen, E. Genetic traces of environmental variations in ancient lakes. In *Dying and Dead Seas Climatic versus Anthropic Causes*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 341–357.
3. Fujita, M.K.; Singhal, S.; Brunes, T.O.; Maldonado, J.A. Evolutionary Dynamics and Consequences of Parthenogenesis in Vertebrates. *Annu. Rev. Ecol. Evol. Syst.* **2020**, *51*, 191–214. [\[CrossRef\]](#)
4. Aylward, M.; Sagar, V.; Natesh, M.; Ramakrishnan, U. How methodological changes have influenced our understanding of population structure in threatened species: Insights from tiger populations across India. *Philos. Trans. R. Soc. B Biol. Sci.* **2022**, *377*, 20200418. [\[CrossRef\]](#)
5. Speed, D.; Kaphle, A.; Balding, D.J. SNP-based heritability and selection analyses: Improved models and new results. *Bioessays* **2022**, *44*, 2100170. [\[CrossRef\]](#)
6. Tang, M.; Wang, T.; Zhang, X. A review of SNP heritability estimation methods. *Briefings Bioinform.* **2022**, *23*, bbac067. [\[CrossRef\]](#)
7. Rengefors, K.; Kremp, A.; Reusch, T.B.; Wood, A.M. Genetic diversity and evolution in eukaryotic phytoplankton: Revelations from population genetic studies. *J. Plankton Res.* **2017**, *39*, 165–179. [\[CrossRef\]](#)
8. Hejase, H.A.; Dukler, N.; Siepel, A. From Summary Statistics to Gene Trees: Methods for Inferring Positive Selection. *Trends Genet.* **2020**, *36*, 243–258. [\[CrossRef\]](#)
9. De Jong, M.J.; de Jong, J.F.; Hoelzel, A.R.; Janke, A. SambaR: An r package for fast, easy and reproducible population-genetic analyses of biallelic SNP data sets. *Mol. Ecol. Resour.* **2021**, *21*, 1369–1379. [\[CrossRef\]](#)
10. Kamvar, Z.N.; Tabima, J.F.; Grünwald, N.J. Poppr: An r package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2014**, *2*, e281. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Serrote, C.M.L.; Reiniger, L.R.S.; Silva, K.B.; Rabaiolli, S.M.D.S.; Stefanel, C.M. Determining the Polymorphism Information Content of a molecular marker. *Gene* **2020**, *726*, 144175. [\[CrossRef\]](#)

12. Csilléry, K.; Blum, M.G.; Gaggiotti, O.E.; François, O. Approximate Bayesian Computation (ABC) in practice. *Trends Ecol. Evol.* **2010**, *25*, 410–418. [[CrossRef](#)]
13. Csilléry, K.; François, O.; Blum, M.G.B. abc: An R package for approximate Bayesian computation (ABC). *Methods Ecol. Evol.* **2012**, *3*, 475–479. [[CrossRef](#)]
14. Schön, I.; Martens, K.; van Dijk, P. Lost sex. In *The Evolutionary Biology of Parthenogenesis*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009; pp. 1–615.
15. Yang, Y.Y.; Kim, J.G. The optimal balance between sexual and asexual reproduction in variable environments: A systematic review. *J. Ecol. Environ.* **2016**, *40*, 89–106. [[CrossRef](#)]
16. Decaestecker, E.; De Meester, L.; Mergeay, J. Cyclical Parthenogenesis in Daphnia: Sexual Versus Asexual Reproduction. In *Lost Sex: The Evolutionary Biology of Parthenogenesis*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009; pp. 295–316. [[CrossRef](#)]
17. Meilland, J.; Ezat, M.M.; Westgård, A.; Manno, C.; Morard, R.; Siccha, M.; Kucera, M. Rare but persistent asexual reproduction explains the success of planktonic foraminifera in polar oceans. *J. Plankton Res.* **2022**, *45*, 15–32. [[CrossRef](#)]
18. Miller, S.M.; Stuart, K.C.; Burke, N.W.; Rollins, L.A.; Bonduriansky, R. Genetic and phenotypic consequences of local transitions between sexual and parthenogenetic reproduction in the wild. *bioRxiv* **2022**. [[CrossRef](#)]
19. Schaible, R.; Ringelhan, F.; Kramer, B.H.; Miethe, T. Environmental challenges improve resource utilization for asexual reproduction and maintenance in hydra. *Exp. Gerontol.* **2011**, *46*, 794–802. [[CrossRef](#)]
20. DeAngelis, D.L.; Rose, K.A.; Huston, M.A. Individual-Oriented Approaches to Modeling Ecological Populations and Communities. In *Frontiers in Mathematical Biology*; Springer: Berlin/Heidelberg, Germany, 1994; pp. 390–410. [[CrossRef](#)]
21. Lorek, H.; Sonnenschein, M. Object-oriented support for modelling and simulation of individual-oriented ecological models. *Ecol. Model.* **1998**, *108*, 77–96. [[CrossRef](#)]
22. Fazalova, V.; Semovski, S.V.; Sherbakov, D.Y.; Meunier, G. The effect of demographic population factors and individual biological parameters on the rate of the neutral molecular evolution. *Russ. J. Genet.* **2007**, *43*, 973–980. [[CrossRef](#)]
23. Hoban, S.; Archer, F.I.; Bertola, L.D.; Bragg, J.G.; Breed, M.F.; Bruford, M.W.; Coleman, M.A.; Ekblom, R.; Funk, W.C.; Grueber, C.E.; et al. Global genetic diversity status and trends: Towards a suite of Essential Biodiversity Variables (EBVs) for genetic composition. *Biol. Rev.* **2022**, *97*, 1511–1538. [[CrossRef](#)]
24. Innes, D.J.; Singleton, D.R. Variation in allocation to sexual and asexual reproduction among clones of cyclically parthenogenetic *Daphnia pulex* (Crustacea: Cladocera). *Biol. J. Linn. Soc.* **2000**, *71*, 771–787. [[CrossRef](#)]
25. Kaliszewicz, A. Interference of asexual and sexual reproduction in the green hydra. *Ecol. Res.* **2010**, *26*, 147–152. [[CrossRef](#)]
26. Sagorny, C.; von Döhren, J. Occasional sexual reproduction significantly affects the population structure of the widespread, predominantly asexually reproducing marine worm *Lineus sanguineus* (Nemertea: Pilidiophora). *Mar. Biol.* **2022**, *169*, 92. [[CrossRef](#)]
27. Preston, R.; Blomster, J.; Schagerström, E.; Seppä, P. Clonality, polyploidy and spatial population structure in Baltic Sea *Fucus vesiculosus*. *Ecol. Evol.* **2022**, *12*, e9336. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.