

Enhancing understanding of population genetics through automated modeling and simulation

Elena V. Komarova

Independent researcher, Moscow, Russia

Abstract. Understanding population genetics concepts, particularly the Hardy-Weinberg equilibrium law, presents significant challenges to students at various educational levels. Our research addresses methodological and practical limitations in teaching these concepts through educational model experiments. We document the evolution of an instructional approach for studying the Hardy-Weinberg law, progressing from material models with manual calculations to a fully automated simulation system. The enhanced methodology allows for substantial increases in model population size and facilitates both technical and biological replication in the educational context. Through iterative design and testing with undergraduate students between 2015-2023, we demonstrate that automated modeling significantly enhances students' conceptual understanding by visualizing abstract genetic-evolutionary processes. Our approach overcomes traditional experimental limitations of mass sampling and replication while making complex population-level phenomena accessible to students. Survey results indicate improved comprehension of genetic equilibrium concepts, with increasing student engagement in exploring evolutionary mechanisms. This work contributes to the theoretical integration of computational modeling and science education by establishing a pedagogical framework that connects abstract genetic concepts to their practical application, promoting deeper scientific literacy through guided simulation experiences.

Keywords: population genetics education, computational modeling, Hardy-Weinberg equilibrium, technical replication, biological replication, science education, simulation-based learning, automated experimentation, student misconceptions, genetic literacy

1. Introduction

The subject of biology in educational contexts represents a didactically adapted system of scientific biological knowledge. As with other natural sciences, experimentation plays a fundamental role in biology as a research method, providing the factual basis upon which broader generalizations and patterns can be established. While extensive research exists regarding experimental methods in biological sciences – including types, methods, requirements, limitations, and historical development – our focus centers specifically on adapting experimental approaches for educational purposes at the secondary and undergraduate levels.

Within educational contexts, model experiments hold particular significance for conveying complex biological concepts that students might otherwise find inaccessible. Previous research has examined the implementation of model experiments in biology education [13, 14], addressing both pedagogical approaches and their effectiveness. The present study extends this work by specifically addressing the challenges of teaching population genetics concepts through experimental approaches.

Modern biology curricula at secondary and undergraduate levels are structured around fundamental theoretical principles of biological science. While Mendelian laws of heredity have traditionally formed the cornerstone of genetics education,

📞 0000-0002-3476-3351 (E. V. Komarova)

✉ komarova1978@mail.ru (E. V. Komarova)



© Copyright for this article by its authors, published by the *Academy of Cognitive and Natural Sciences*. This is an Open Access article distributed under the terms of the Creative Commons License Attribution 4.0 International (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

contemporary developments in theoretical biology – particularly in genetics, ecology, and evolution – necessitate greater emphasis on population-level genetic concepts. The Hardy-Weinberg equilibrium law, which describes the stability of genotype frequencies in a population under specific conditions, has emerged as a key principle for understanding evolutionary mechanisms. However, research consistently indicates that students struggle to master these abstract population-level concepts [10, 20, 23, 24, 26, 30].

1.1. Current state of student understanding

To assess the current level of understanding regarding the Hardy-Weinberg equilibrium law, we conducted a survey among 52 high school students. The survey evaluated students' comprehension of genetic equilibrium concepts, their understanding of the law's mathematical representation, and their ability to apply these principles to practical problems. The questions focused on:

1. Identification of the correct mathematical equation representing the Hardy-Weinberg law.
2. Recognition of equations describing genotypic and allelic population structures.
3. Knowledge of the conditions necessary for the law's validity.
4. Application of the law to solve population genetics problems.

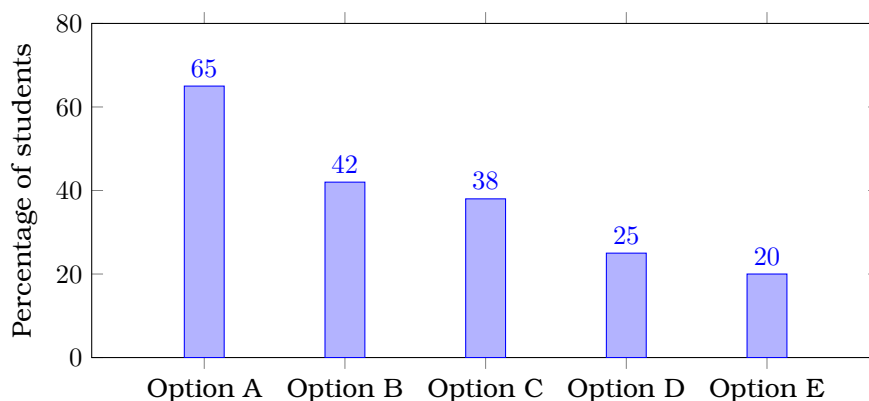


Figure 1: Student responses to the first task: identifying the correct mathematical equation of the Hardy-Weinberg law. Option A: $p + q = 1$; Option B: $(p + q)^2 = 1$; Option C: $p^2 + 2pq + q^2 = 1$; Option D: $p^2 + pq + q^2 = 1$; Option E: $p + 2pq + q = 1$.

Survey results revealed significant gaps in students' understanding of population genetics concepts. Regarding the conditions necessary for the Hardy-Weinberg law to apply, students demonstrated fragmentary knowledge: 29% recognized the requirement for large population sizes, 24% understood the need for random mating, 18% acknowledged the absence of new mutations, 12% identified the necessity of equal fertility across genotypes, 12% recognized non-overlapping generations, 18% noted the absence of gene exchange with other populations, 18% understood the requirement for genes to be on autosomes rather than sex chromosomes, and 12% recognized the need for equal viability across genotypes.

More concerning was students' inability to solve genetic problems, with only three students attempting one problem and all producing incorrect solutions. These findings align with previous research suggesting that students often adopt a formal, memorization-based approach to population genetics concepts without developing genuine understanding. This formalistic learning approach contributes to difficulties in applying these concepts to evolutionary processes or human population genetics.

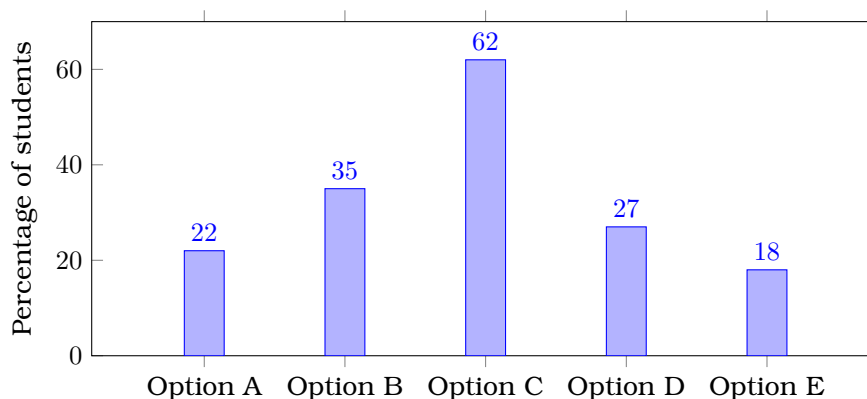


Figure 2: Student responses to the second task: identifying the equation describing genotypic structure of the population. Option A: $p + q = 1$; Option B: $(p + q)^2 = 1$; Option C: $p^2 + 2pq + q^2 = 1$; Option D: $p^2 + pq + q^2 = 1$; Option E: $p + 2pq + q = 1$.

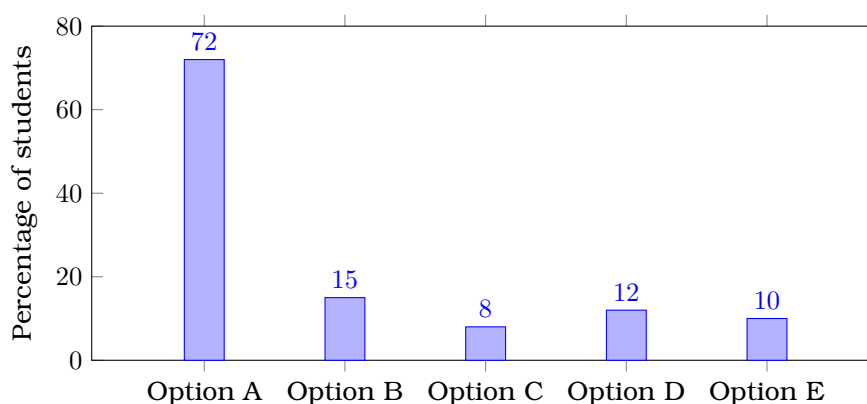


Figure 3: Student responses to the third task: identifying the equation describing allelic population structure. Option A: $p + q = 1$; Option B: $(p + q)^2 = 1$; Option C: $p^2 + 2pq + q^2 = 1$; Option D: $p^2 + pq + q^2 = 1$; Option E: $p + 2pq + q = 1$.

1.2. Challenges in educational experimentation

The challenges in teaching population genetics concepts through educational experimentation arise from several factors. Simulation and modeling offer particularly effective approaches for demonstrating complex biological processes, including genetic and evolutionary dynamics in populations. However, implementing meaningful educational experiments in this domain presents specific challenges related to the fundamental requirements of scientific experimentation.

A scientific biological experiment should ideally fulfill several key requirements: reliability, rule of single difference, replication, and mass sampling. In an educational context, replication and mass sampling present particular difficulties for the following reasons:

1. Temporal limitations within the educational process.
2. Limited access to sufficient experimental objects.
3. Inaccessibility of certain biological phenomena for direct observation due to scale (molecules, cells, viral particles) or complexity (population-level processes).

Recent methodological discussions in biological science have highlighted the issue of pseudoreplication in research [6, 18]. This phenomenon occurs when observations are not statistically independent but are treated as if they are, leading to potentially

misleading results. While initially viewed negatively [11], attitudes toward pseudoreplication have evolved, with ongoing debate regarding its status as a genuine methodological problem versus a contrived concern [16, 34, 35].

Reish, Dewey and Kirschman [33] define pseudoreplication as a statistical error occurring when treatments are not replicated or when replicates are not statistically independent. In educational contexts, understanding the difference between technical replication (repeating measurements on the same sample) and biological replication (measurements from different biological sources) becomes crucial. Technical replication provides information about measurement precision and accuracy for a specific object, while biological replication helps determine whether observations generalize across a group [17, 38].

Our research addresses these challenges by developing an enhanced methodology for teaching population genetics through model experimentation. We specifically focus on creating a pedagogical approach that balances scientific rigor with educational practicality, allowing students to meaningfully engage with population genetics concepts while developing proper understanding of experimental methodology.

2. Methodology and research design

To address the challenges of teaching population genetics concepts through educational experimentation, we developed and refined a model experiment methodology over a five-year period (2015-2020). Our approach evolved through several stages to overcome limitations in mass sampling and replication while providing students with meaningful learning experiences.

2.1. Research questions

We began by formulating several key questions to guide our methodological development:

1. Is a model experiment necessary for teaching genetic equilibrium concepts and conditions?
2. Can other educational methods effectively replace model experimentation?
3. How can we overcome the limitations of mass sampling and replication in an educational setting?
4. How can we help students distinguish between technical and biological replication?
5. What is the optimal approach for visualizing genetic changes across generations in a population?

Through careful consideration of these questions, we determined that while alternative teaching approaches exist (e.g., multimedia presentations, theoretical material, problem-solving), model experimentation offers unique advantages for conceptualizing abstract genetic principles. These advantages include:

1. Providing concrete manifestations of abstract biological categories.
2. Visualizing ideal population processes not directly observable in nature.
3. Simulating changes in real populations across multiple generations in compressed timeframes.
4. Allowing for variation in replication with minimal material costs.
5. Teaching proper replication concepts in experimental design.
6. Demonstrating the effects of various factors (including stochastic effects) on population structure.

2.2. Evolution of the experimental methodology

Our methodological approach progressed through three distinct stages of development, each addressing specific limitations of the previous stage.

Stage 1: Material models with manual calculation

Initially, we implemented a physical model experiment using material objects to represent gene alleles. Students manually created genotype models, calculated frequencies, and constructed models of parental and daughter populations across generations. This approach allowed direct manipulation of model elements but was limited by physical constraints and calculation time.

Stage 2: Hybrid material-digital approach

In the second stage, we combined material modeling with digital tools. Students still created physical models and performed random combinations of alleles, but entered results into web-based tables that automatically calculated frequencies and generated graphical representations. This hybrid approach reduced calculation burdens while maintaining hands-on engagement with model construction.

We developed several web pages to support different modeling scenarios:

- Model experiment 1: Study of genetic structure in ideal populations (two variants).
- Model experiment 2: Study of genetic structure under natural selection.
- Model experiment 3: Modeling gene flow effects on genetic structure.
- Model experiment 4: Modeling random processes and gene drift effects.

This system allowed students to enter experimental results from physical modeling activities and automatically process the data to visualize genetic changes across generations. However, significant limitations remained, particularly regarding sample size and replication.

Stage 3: Fully automated modeling

In the third stage, we addressed the limitations of previous approaches by developing a fully automated modeling system. This approach eliminated physical constraints on population size and allowed for extensive replication, both technical and biological. The enhanced system enabled simulation of much larger populations (potentially millions of individuals) and increased the number of observable generations, thereby providing more statistically robust results that better aligned with theoretical expectations.

This automated approach was implemented through a new web page (<https://web.archive.org/web/20221221225840/http://mybio.education/mod/exp6/en/>) that allowed users to specify initial population parameters and observe multiple generations, with automatic visualization of results. The system's enhancements included:

1. Ability to model populations of hundreds, thousands, or millions of individuals.
2. User-specified initial allele frequencies.
3. Increased generational scope (five generations plus parental population).
4. Automatic calculation and visualization of genetic structure changes.
5. Support for both technical replication (multiple generations within one session) and biological replication (different populations across sessions).

2.3. Experimental design and validation

To evaluate the effectiveness of our enhanced methodology, we conducted a validation study during the 2019-2020 academic year. The study involved 6 undergraduate biology students from the Institute of Living Systems at Immanuel Kant Baltic Federal University and 12 eleventh-grade students from the “School of the Future” in Kaliningrad Region, Russia.

Model experiment 1. Study of the genetic structure of the ideal population (third option)

1. In column 2 for the parent generation P, we introduce the number of pairs of two-body gene alleles (in other words, the number of individuals).
2. We are determined by the ratio of the dominant (A) and recessive (a) alleles, and we introduce their values in columns 9 and 10 for the parent generation P.
3. Click the "Calculate" button.
4. Click the "Calculate" button opposite the lines F1, F2, F3, F4, F5.
5. Click on the "Show Graphs" button.
6. Based on the analysis of the obtained graphs and diagrams, formulate the conclusions of the plan:
 - Change in the frequency of genotypes in generations;
 - Change in the ratio of gene frequencies in generations;
 - The direction of evolutionary changes in the population.

Table 6. Genetic structure of the ideal population

Generation	Number of individuals	Distribution of genotypes						Gene frequencies		
		AA	Aa	aa	A(p)	a(q)				
1	2	3	4	5	6	7	8	9	10	
P	20000	10910	0.545	6180	0.309	2910	0.145	0.7	0.3	Calculate
F1	20000	10893	0.545	6214	0.311	2893	0.145	0.738	0.380	Calculate
F2	20000	10883	0.544	6234	0.312	2883	0.144	0.738	0.380	Calculate
F3	20000	10890	0.544	6220	0.311	2890	0.144	0.738	0.380	Calculate
F4	20000	10823	0.541	6354	0.318	2823	0.141	0.736	0.376	Calculate
F5	20000	10869	0.543	6262	0.313	2869	0.143	0.737	0.379	Calculate

[Show Graphs](#)

Figure 4: Results of a model experiment with a population of 20,000 individuals, starting with allele frequencies $p = 0.7$ and $q = 0.3$.

Model experiment 1. Study of the genetic structure of the ideal population (third option)

1. In column 2 for the parent generation P, we introduce the number of pairs of two-body gene alleles (in other words, the number of individuals).
2. We are determined by the ratio of the dominant (A) and recessive (a) alleles, and we introduce their values in columns 9 and 10 for the parent generation P.
3. Click the "Calculate" button.
4. Click the "Calculate" button opposite the lines F1, F2, F3, F4, F5.
5. Click on the "Show Graphs" button.
6. Based on the analysis of the obtained graphs and diagrams, formulate the conclusions of the plan:
 - Change in the frequency of genotypes in generations;
 - Change in the ratio of gene frequencies in generations;
 - The direction of evolutionary changes in the population.

Table 6. Genetic structure of the ideal population

Generation	Number of individuals	Distribution of genotypes						Gene frequencies		
		AA	Aa	aa	A(p)	a(q)				
1	2	3	4	5	6	7	8	9	10	
P	20000	2295	0.115	3410	0.171	14295	0.715	0.2	0.8	Calculate
F1	20000	2261	0.113	3478	0.174	14261	0.713	0.336	0.844	Calculate
F2	20000	2310	0.116	3380	0.169	14310	0.716	0.340	0.846	Calculate
F3	20000	2274	0.114	3452	0.173	14274	0.714	0.337	0.845	Calculate
F4	20000	2221	0.111	3558	0.178	14221	0.711	0.333	0.843	Calculate
F5	20000	2285	0.114	3430	0.172	14285	0.714	0.338	0.845	Calculate

[Show Graphs](#)

Figure 5: Results of a model experiment with a population of 200,000 individuals with different initial allele frequencies ($p = 0.2$ and $q = 0.8$), demonstrating how the system can be used for biological replication.

The validation occurred within the framework of the “Effective High School” conference organized in collaboration with the National Research University Higher School of Economics. During practical sessions on modeling genetic-evolutionary processes, participants were divided into two groups with different instructional sequences:

1. Group 1: Began with the original manual model experiment followed by the fully automated approach.
2. Group 2: Used only the fully automated modeling approach.

Both groups performed multiple experimental trials with different initial parameters (population size and allele frequencies). Data collection included:

1. Observation of participant interaction with the modeling systems.
2. Documentation of student difficulties and questions.
3. Post-activity interviews regarding comprehension and learning experience.

This design allowed us to assess both the technical functionality of the modeling system and its pedagogical effectiveness in developing student understanding of population genetics concepts, particularly regarding technical and biological replication.

3. Results

The implementation and validation of our enhanced methodology for teaching population genetics concepts yielded significant insights into both the technical aspects of the modeling system and its educational effectiveness. We present these findings in relation to the key research questions and methodological goals.

3.1. Technical implementation results

The fully automated modeling system successfully addressed the limitations identified in previous approaches, enabling:

1. Simulation of much larger populations, ranging from thousands to millions of individuals.
2. User-controlled specification of initial allele frequencies.
3. Visualization of genetic structure across five generations plus the parental population.
4. Support for both technical replication (multiple generations within a session) and biological replication (different population parameters across sessions).

This technical implementation significantly expanded the scope and capability of the educational model experiment, bringing it closer to the conditions of real population genetics research while maintaining accessibility for educational purposes.

3.2. Validation study results

The validation study with undergraduate biology students and high school students revealed important patterns in user experience and learning outcomes.

3.2.1. User experience with different approaches

Participants in Group 1, who began with the manual approach before transitioning to the automated system, reported several challenges with the initial methodology:

1. Difficulty understanding the purpose of repetitive manual calculations at the beginning of the activity.
2. Fatigue and frustration with routine data handling tasks.
3. Increased error rates in calculations due to fatigue, requiring repeated experimentation.
4. Strong preference for the automated system once introduced.

In contrast, Group 2 participants, who used only the automated system, completed the assigned tasks more efficiently. However, both groups expressed uncertainty about the purpose of performing multiple experimental replicates with different population parameters (biological replication), while understanding the value of observing multiple generations of the same population (technical replication).

3.2.2. Understanding of replication concepts

A key finding was the differential understanding of technical versus biological replication concepts:

1. Participants readily grasped the concept of technical replication, understanding why observing multiple generations of the same population was scientifically valuable.
2. The concept of biological replication – using different starting parameters to explore the generalizability of findings – was less intuitively understood.

This observation aligns with contemporary discussions in scientific methodology regarding pseudoreplication [16, 25, 34, 35]. The lack of clear understanding of biological replication among participants suggests a need for more explicit instruction regarding experimental design principles.

3.2.3. Learning outcomes

The automated modeling system demonstrated significant advantages for learning outcomes:

1. Reduced cognitive load from mechanical calculation tasks allowed greater focus on conceptual understanding.
2. Visual representation of population changes across generations enhanced comprehension of the Hardy-Weinberg equilibrium.
3. The ability to observe large populations made statistical patterns more evident, reinforcing theoretical principles.
4. Opportunity to explore different initial conditions facilitated deeper understanding of the factors influencing population genetics.

These benefits align with recent research on computational modeling in science education, which emphasizes the value of interactive simulation for developing conceptual understanding of complex systems [9, 27].

3.3. Analysis of student understanding

The evaluation revealed three key patterns in student understanding of population genetics concepts through the model experiment approach:

1. Students became more adept at connecting mathematical formulations with biological processes when using the automated system, demonstrating improved integration of quantitative and qualitative aspects of population genetics.
2. The automated approach helped students overcome key threshold concepts in population genetics, particularly regarding the stability of genotype frequencies in ideal populations and the factors that disturb this equilibrium.
3. Students developed greater awareness of the importance of sample size and replication in scientific research, though more support was needed to fully appreciate the distinction between technical and biological replication.

These findings are consistent with research by Cannady et al. [5] on scientific sensemaking in students, which demonstrates how structured engagement with scientific practices can enhance content learning across disciplines.

3.4. Comparative analysis

To place these findings in context, table 1 presents a comparative analysis of the three methodological approaches developed throughout this research.

This comparison demonstrates the evolution of our approach toward greater efficiency, scale, and conceptual focus, while maintaining the essential educational benefits of model experimentation.

Table 1

Comparative analysis of three methodological approaches

Characteristic	Material models with manual calculation	Hybrid material-digital approach	Fully automated modeling
Maximum population size	50-100 individuals	50-100 individuals	Unlimited (potentially millions)
Number of observable generations	3 (parental + 2 daughter)	3 (parental + 2 daughter)	6 (parental + 5 daughter)
Technical replication	Limited by calculation time	Limited by material constraints	Extensive and efficient
Biological replication	Difficult due to resource constraints	Difficult due to resource constraints	Easily implemented through parameter variation
Data visualization	Manual graphing	Automatic generation of limited graphs	Comprehensive automated visualization
Student cognitive focus	Mechanical calculation processes	Data entry and interpretation	Conceptual understanding and analysis
Implementation time	2-3 class periods	1-2 class periods	Single class period with deeper exploration
Pedagogical emphasis	Process understanding	Hybrid process-concept learning	Conceptual understanding with methodological awareness

4. Discussion

Our research addresses fundamental challenges in teaching complex population genetics concepts through the development and refinement of an enhanced modeling methodology. The findings contribute to several key areas of biology education research, with implications for both theoretical understanding and practical implementation.

4.1. Addressing educational challenges in population genetics

Population genetics concepts, particularly the Hardy-Weinberg equilibrium law, present significant educational challenges due to their abstract nature, mathematical complexity, and disconnect from students' everyday experience. Our study confirms previous findings regarding students' difficulties with these concepts [10, 20, 23, 24, 26, 30], but offers a promising approach to address these challenges through computational modeling.

The results of our validation study suggest that an automated modeling approach significantly enhances students' ability to engage with population genetics concepts by:

1. Providing concrete visualization of abstract genetic processes.
2. Demonstrating the mathematical relationships between allele and genotype frequencies.
3. Allowing exploration of equilibrium conditions and disturbances.
4. Reducing cognitive load associated with mechanical calculations.
5. Facilitating understanding of statistical patterns through large sample sizes.

These benefits align with theoretical perspectives on the value of simulation and modeling in science education [1, 4]. As Tutwiler and Grotzer [39] argue, immersive, interactive simulations are particularly valuable for helping students understand complex causal dynamics that are otherwise difficult to observe directly.

4.2. Bridging technical and conceptual understanding

A significant contribution of our approach is its potential to bridge technical and conceptual understanding in population genetics education. Traditional approaches often treat mathematical formulations and biological concepts separately, leading to fragmented understanding. Our methodology integrates these aspects by connecting the mathematical relationships of the Hardy-Weinberg law with visual representations of population changes across generations.

This integration reflects recent theoretical perspectives on the relationship between scientific methodological knowledge and concept mastery [15, 41]. As Papadouris and Constantinou [29] suggest, elevating attention to the epistemic and ontological aspects of content knowledge, rather than focusing solely on substantive content, enhances conceptual understanding and epistemological stance toward science learning.

4.3. Replication concepts in biology education

Our findings regarding students' differential understanding of technical versus biological replication highlight an important area for improvement in science education. While participants readily grasped the value of technical replication (observing multiple generations of the same population), they struggled to appreciate the importance of biological replication (testing with different population parameters).

This observation connects to broader discussions about pseudoreplication in scientific research [11, 16, 34, 35] and suggests that explicit instruction regarding replication concepts could enhance students' methodological understanding. As Naegele, Gough and Yaffe [28] note, understanding the distinction between technical and biological replication is crucial for proper experimental design and data interpretation in biological sciences.

The challenge of teaching replication concepts effectively relates to deeper issues in scientific literacy and methodological understanding. Recent work by Priest [31] emphasizes the importance of "critical science literacy", which involves understanding the social character of science and the nature of scientific consensus. Our findings suggest that computational modeling approaches can provide a foundation for developing this deeper literacy by engaging students in authentic scientific practices and methodological considerations.

4.4. Technological affordances for science education

The evolution of our methodology from manual modeling to fully automated simulation reflects broader trends in educational technology and computational modeling in science education. Our findings regarding the effectiveness of the automated approach align with research on the role of web-based tools in making abstract concepts concrete for students [2, 12].

The advantages of our automated modeling system include:

1. Accessibility and ease of use for students with varying technical backgrounds.
2. Scalability to accommodate large virtual populations that better demonstrate statistical principles.
3. Visualization capabilities that make abstract genetic processes observable.
4. Flexibility to modify parameters and explore different scenarios.
5. Efficiency in generating results quickly, allowing for more experimental iterations.

These affordances address key challenges in science education identified by Ruzman and Rosli [36], who emphasize the importance of inquiry-based education that awakens learners' curiosity and wonder. By providing an interactive platform for exploring population genetics concepts, our approach supports the development of higher-order thinking skills through meaningful inquiry.

The implementation of our web-based system also responds to recent calls for more effective integration of computational modeling in biology education [19, 27]. As Voit et al. [40] argue, computational approaches are becoming central to biological and medical research, necessitating greater emphasis on these skills in educational contexts.

4.5. Limitations and areas for future development

Despite the demonstrated benefits of our enhanced methodology, several limitations should be acknowledged:

1. Current implementation focuses only on ideal populations without evolutionary factors.
2. Student understanding of biological replication concepts requires additional support.
3. The system could benefit from more explicit connections to real-world population genetics research.
4. Long-term retention of concepts was not assessed in the current study.

These limitations suggest several directions for future development:

1. Expansion of the automated modeling system to include evolutionary factors such as natural selection, gene flow, genetic drift, and mutation.
2. Development of more explicit instructional scaffolding regarding experimental design and replication concepts.
3. Integration of case studies from real population genetics research to enhance contextual relevance.
4. Longitudinal assessment of conceptual understanding and transfer to other biological domains.

5. Educational implications and theoretical framework

Our research contributes to the development of a more robust theoretical framework for teaching complex biological concepts through computational modeling. Building on existing work in science education, we propose an integrated framework that connects scientific methodological knowledge with conceptual understanding through simulation-based learning experiences.

5.1. Enhancing conceptual understanding through modeling

The results of our study support the value of computational modeling for enhancing students' conceptual understanding of abstract genetic concepts. This aligns with theoretical perspectives on conceptual change in science education, which emphasize the importance of making abstract ideas concrete and observable [3].

Our findings suggest that effective modeling approaches:

1. Provide visual representations of otherwise abstract processes.
2. Allow manipulation of variables to observe effects on outcomes.
3. Demonstrate relationships between mathematical formulations and biological phenomena.

4. Support exploration of multiple scenarios to identify patterns and principles.
5. Bridge theoretical knowledge and practical application.

These elements contribute to what Cannady et al. [5] describe as “scientific sense-making”, which supports science content learning across disciplines and instructional contexts.

5.2. Pedagogical content knowledge for population genetics

The development and implementation of our enhanced methodology contributes to expanding pedagogical content knowledge (PCK) for teaching population genetics. As defined by Shulman [37], PCK encompasses the specialized knowledge that teachers need to make subject matter comprehensible to students.

Our approach enhances PCK for population genetics by:

1. Identifying key threshold concepts and potential misconceptions.
2. Providing concrete strategies for visualizing abstract genetic processes.
3. Offering tools for navigating between mathematical and biological aspects of population genetics.
4. Developing awareness of methodological considerations in genetic research.
5. Establishing connections between theoretical principles and observable outcomes.

This contribution addresses the need identified by Fikriyah, Yasir and Qomaria [8] for enhanced PCK among science educators, particularly regarding complex biological concepts.

5.3. Integrated framework for computational modeling in biology education

Drawing on our findings and existing theoretical perspectives, we propose an integrated framework for computational modeling in biology education that encompasses four key dimensions:

1. **Conceptual understanding** – developing robust mental models of biological processes through simulation-based exploration.
2. **Methodological awareness** – building understanding of scientific methods, including experimental design, replication, and data interpretation.
3. **Technological integration** – using digital tools to visualize, manipulate, and analyze complex biological systems.
4. **Contextual relevance** – connecting abstract principles to real-world biological phenomena and research applications.

This framework aligns with recent perspectives on the integration of computational thinking in science education [21, 32] and supports the development of what Dorfner, Förtsch and Neuhaus [7] describes as conceptual rather than content-focused teaching.

Figure 6 illustrates the relationships between these dimensions and their contribution to enhanced biological literacy.

This framework provides a foundation for designing and implementing computational modeling approaches across various biological domains, extending beyond population genetics to other complex biological systems.

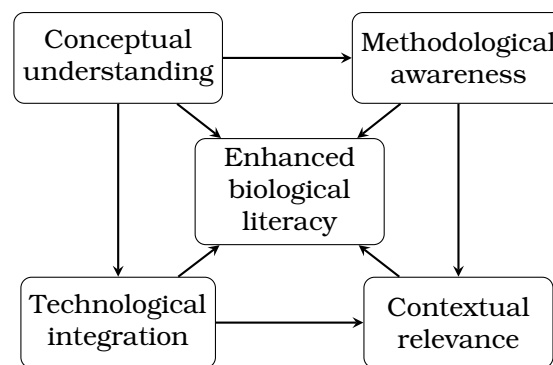


Figure 6: Integrated framework for computational modeling in biology education, showing relationships between key dimensions and their contribution to enhanced biological literacy.

5.4. Practical implications for biology education

Our research offers several practical implications for biology education at secondary and undergraduate levels:

1. Integration of computational modeling approaches into standard biology curricula, particularly for abstract concepts like population genetics, can enhance conceptual understanding and engagement.
2. Professional development for biology educators should include training in computational modeling approaches and their pedagogical implementation, addressing both technical skills and conceptual knowledge.
3. Development of accessible, web-based modeling tools for various biological concepts can support inquiry-based learning and conceptual exploration.
4. Evaluation of student understanding should incorporate modeling-based assessments that reveal conceptual integration rather than relying solely on recall of facts or formulae.
5. Computational modeling in biology education offers opportunities for interdisciplinary connections with mathematics, computer science, and statistics, enhancing students' cross-disciplinary literacy.

These implications respond to calls for more innovative approaches to biology education that prepare students for the increasingly computational nature of biological research [22, 40].

6. Conclusion

Our research has documented the development and validation of an enhanced methodology for teaching complex population genetics concepts through automated modeling. By addressing the limitations of traditional educational experiments – particularly regarding mass sampling and replication – we have created an approach that balances scientific rigor with educational accessibility.

The evolution of our methodology from manual modeling to fully automated simulation reflects a broader trend toward computational approaches in science education. Our findings demonstrate that this approach enhances students' conceptual understanding of population genetics principles, particularly the Hardy-Weinberg equilibrium law, while also developing awareness of methodological considerations in biological research.

The proposed integrated framework for computational modeling in biology education offers a theoretical foundation for extending these approaches to other biological

domains. By connecting conceptual understanding, methodological awareness, technological integration, and contextual relevance, this framework supports the development of enhanced biological literacy that prepares students for contemporary scientific practice.

Future research should focus on expanding the automated modeling system to incorporate evolutionary factors, developing more explicit scaffolding for experimental design concepts, and assessing long-term retention and transfer of learning. Additionally, exploration of similar approaches for other complex biological systems could further validate the effectiveness of computational modeling as a pedagogical strategy.

References

- [1] Alam, A., 2023. Leveraging the Power of 'Modeling and Computer Simulation' for Education: An Exploration of its Potential for Improved Learning Outcomes and Enhanced Student Engagement. *Proceedings - IEEE International Conference on Device Intelligence, Computing and Communication Technologies, DICCT 2023*. pp.445 – 450. Available from: <https://doi.org/10.1109/DICCT56244.2023.10110159>.
- [2] Bobić, A., Cheong, C., Filippou, J., Cheong, F. and Guetl, C., 2020. Infret: Enhancing a Tool for Explorative Learning of Information Retrieval Concepts. In: M.E. Auer, H. Hortsch and P. Sethakul, eds. *The Impact of the 4th Industrial Revolution on Engineering Education*. Cham: Springer International Publishing, *Advances in Intelligent Systems and Computing*, vol. 1134, pp.67–78. Available from: https://doi.org/10.1007/978-3-030-40274-7_7.
- [3] Braithwaite, D.W. and Goldstone, R.L., 2015. Effects of Variation and Prior Knowledge on Abstract Concept Learning. *Cognition and Instruction*, 33(3), pp.226–256. Available from: <https://doi.org/10.1080/07370008.2015.1067215>.
- [4] Bredeweg, B. and Salles, P., 2011. International Workshop on Learning by Modelling in Science Education. In: G. Biswas, S. Bull, J. Kay and A. Mitrovic, eds. *Artificial Intelligence in Education*. Berlin, Heidelberg: Springer Berlin Heidelberg, *Lecture Notes in Computer Science*, vol. 6738, pp.633–633. Available from: https://doi.org/10.1007/978-3-642-21869-9_133.
- [5] Cannady, M.A., Vincent-Ruz, P., Chung, J.M. and Schunn, C.D., 2019. Scientific sensemaking supports science content learning across disciplines and instructional contexts. *Contemporary Educational Psychology*, 59, p.101802. Available from: <https://doi.org/10.1016/j.cedpsych.2019.101802>.
- [6] Colegrave, N. and Ruxton, G.D., 2018. Using Biological Insight and Pragmatism When Thinking about Pseudoreplication. *Trends in Ecology and Evolution*, 33(1), pp.28–35. Available from: <https://doi.org/10.1016/j.tree.2017.10.007>.
- [7] Dorfner, T., Förtsch, C. and Neuhaus, B.J., 2020. Use of technical terms in German biology lessons and its effects on students' conceptual learning. *Research in Science and Technological Education*, 38(2), pp.227–251. Available from: <https://doi.org/10.1080/02635143.2019.1609436>.
- [8] Fikriyah, A., Yasir, M. and Qomaria, N., 2021. Analysis of pedagogical content knowledge on students of science education as pre-service teachers in Madura secondary school. *IOP Conference Series: Earth and Environmental Science*, 747(1), p.012108. Available from: <https://doi.org/10.1088/1755-1315/747/1/012108>.
- [9] Guo, Q., He, Y., Chen, Y. and Qiao, C., 2024. Design and application of computational modeling in science education research: a systematic review. *Interactive Learning Environments*. Available from: <https://doi.org/10.1080/10494820.2024.2372832>.
- [10] Hammersmith, R.L. and Mertens, T.R., 1990. Teaching the Concept of Genetic

- Drift Using a Simulation. *The American Biology Teacher*, 52(8), pp.497–499. Available from: <http://www.jstor.org/stable/4449186>.
- [11] Hurlbert, S.H., 1984. Pseudoreplication and the Design of Ecological Field Experiments. *Ecological Monographs*, 54(2), pp.187–211. Available from: <https://doi.org/10.2307/1942661>.
- [12] Hwang, G.J., Kuo, F.R., Chen, N.S. and Ho, H.J., 2014. Effects of an integrated concept mapping and web-based problem-solving approach on students' learning achievements, perceptions and cognitive loads. *Computers and Education*, 71, pp.77–86. Available from: <https://doi.org/10.1016/j.compedu.2013.09.013>.
- [13] Komarova, E. and Starova, T., 2020. Majority values of school biological education in the context of education for sustainable development. *E3S Web of Conferences*, 166, p.10029. Available from: <https://doi.org/10.1051/e3sconf/202016610029>.
- [14] Komarova, O.V. and Azaryan, A.A., 2018. Computer simulation of biological processes at the high school. *CEUR Workshop Proceedings*, 2257, pp.24–32.
- [15] Koponen, I.T. and Pehkonen, M., 2010. Coherent knowledge structures of physics represented as concept networks in teacher education. *Science and Education*, 19(3), pp.259–282. Available from: <https://doi.org/10.1007/s11191-009-9200-z>.
- [16] Kozlov, M.V. and Hulbert, S.H., 2006. Pseudoreplication, chatter, and the international nature of science: a response to D.V. Tatarnikov. *Zhurnal obshchei biologii*, (67), p.145–152.
- [17] Lazic, S.E., Mellor, J.R., Ashby, M.C. and Munafo, M.R., 2020. A Bayesian predictive approach for dealing with pseudoreplication. *Scientific Reports*, 10(1), p.2366. Available from: <https://doi.org/10.1038/s41598-020-59384-7>.
- [18] Li-Gao, R., Hughes, D.A., Cessie, S. le, De Mutsert, R., Den Heijer, M., Rosendaal, F.R., Dijk, K.W. van, Timpson, N.J. and Mook-Kanamori, D.O., 2019. Assessment of reproducibility and biological variability of fasting and postprandial plasma metabolite concentrations using 1H NMR spectroscopy. *PLoS ONE*, 14(6), p.e0218549. Available from: <https://doi.org/10.1371/journal.pone.0218549>.
- [19] Madamanchi, A., Cardella, M.E., Glazier, J.A. and Umulis, D.M., 2018. Factors Mediating Learning and Application of Computational Modeling by Life Scientists. *Proceedings - Frontiers in Education Conference, FIE*. vol. 2018-October. Available from: <https://doi.org/10.1109/FIE.2018.8659328>.
- [20] Maret, T.J. and Rissing, S.W., 1998. Exploring Genetic Drift & Natural Selection through a Simulation Activity. *The American Biology Teacher*, 60(9), pp.681–683. Available from: <http://www.jstor.org/stable/4450580>.
- [21] Martinovic, D. and Milner-Bolotin, M., 2021. Examination of modelling in K-12 STEM teacher education: Connecting theory with practice. *STEM Education*, 1(4), pp.279–298. Available from: <https://doi.org/10.3934/steme.2021018>.
- [22] McKlin, T., Lee, T., Wanzer, D., Magerko, B., Edwards, D., Grossman, S., Bryans, E. and Freeman, J., 2019. Exploring the correlation between teacher pedagogical content knowledge and content knowledge in computer science classrooms. *Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE*. p.315. Available from: <https://doi.org/10.1145/3304221.3325556>.
- [23] Mertens, T.R., 1992. Introducing Students to Population Genetics & the Hardy-Weinberg Principle. *The American Biology Teacher*, 54(2), pp.103–107. Available from: <http://www.jstor.org/stable/4449417>.
- [24] Moore, R., ed., 1994. *Biology Labs That Work: The Best of How-To-Do-Its*. Reston, VA: National Association of Biology Teachers. Available from: <https://files.eric.ed.gov/fulltext/ED411149.pdf>.
- [25] Moreau, D. and Wiebels, K., 2023. Ten simple rules for designing and conducting undergraduate replication projects. *PLoS Computational Biology*, 19(3), p.e1010957. Available from: <https://doi.org/10.1371/journal.pcbi.1010957>.

- [26] Mukhopadhyay, C., Henze, R. and Moses, Y.T., 2014. *Genetic Drift and Gene Flow Illustration. How Real is Race? A Sourcebook on Race, Culture, and Biology*. 2nd ed. Lanham: AltaMira Press. Available from: <http://www.sjsu.edu/people/carol.mukhopadhyay/race/Gene-Flow-and-Genetic-Drift-Illustration-2014.pdf>.
- [27] Musaeus, L.H., Tatar, D. and Musaeus, P., 2024. Computational modelling in high school biology: A teaching intervention. *Journal of Biological Education*, 58(4), pp.812–828. Available from: <https://doi.org/10.1080/00219266.2022.2118353>.
- [28] Naegle, K., Gough, N.R. and Yaffe, M.B., 2015. Criteria for biological reproducibility: What does “n” mean? *Science Signaling*, 8(371), pp.fs7–fs7. Available from: <https://doi.org/10.1126/scisignal.aab1125>.
- [29] Papadouris, N. and Constantinou, C.P., 2017. Integrating the epistemic and ontological aspects of content knowledge in science teaching and learning. *International Journal of Science Education*, 39(6), pp.663–682. Available from: <https://doi.org/10.1080/09500693.2017.1299950>.
- [30] Pongsophon, P., Roadrangka, V. and Campbell, A., 2007. Counting Buttons: demonstrating the Hardy-Weinberg principle. *Science in School*, (6), pp.30–35.
- [31] Priest, S., 2016. Critical Science Literacy: Making Sense of Science. *Communicating Climate Change: The Path Forward*. London: Palgrave Macmillan UK, Palgrave Studies in Media and Environmental Communication, pp.115–135. Available from: https://doi.org/10.1057/978-1-137-58579-0_6.
- [32] Psycharis, S., 2013. Examining the effect of the computational models on learning performance, scientific reasoning, epistemic beliefs and argumentation. *Computers and Education*, 68, pp.253–265. Available from: <https://doi.org/10.1016/j.compedu.2013.05.015>.
- [33] Reish, H.M., Dewey, L. and Kirschman, L.J., 2024. A host of issues: pseudoreplication in host–microbiota studies. *Applied and Environmental Microbiology*, 90(8), pp.e01033–24. Available from: <https://doi.org/10.1128/aem.01033-24>.
- [34] Rosenberg, G.S., 2019. Several comments of the translator about “truth, lies and statistics”. *Samarskaya Luka: problems of regional and global ecology*, (28), p.53–38. Available from: <https://doi.org/10.24411/2073-1035-2019-10274>.
- [35] Rosenberg, G.S. and Gelashvili, D.B., eds, 2008. *Problemy ekologicheskogo eksperimenta (Planirovanie i analiz nabludenii) [Problems of ecological experiment (Planning and analysis of observations)]*. Togliatti: Cassandra.
- [36] Ruzaman, N.K. and Rosli, D.I., 2020. Inquiry-based education: Innovation in participatory inquiry paradigm. *International Journal of Emerging Technologies in Learning*, 15(10), pp.4–15. Available from: <https://doi.org/10.3991/ijet.v15i10.11460>.
- [37] Shulman, L.S., 2019. Those who understand: Knowledge growth in teaching [Aquellos que entienden: Desarrollo del conocimiento en la enseñanza]. *Profesorado*, 23(3), pp.269–295. Available from: <https://doi.org/10.30827/profesorado.v23i3.11230>.
- [38] Starmer, J., 2017. StatQuest: Technical and Biological replicates, clearly explained! Available from: <https://statquest.org/statquest-technical-and-biological-replicates-clearly-explained/>.
- [39] Tutwiler, M.S. and Grotzer, T., 2014. Why Immersive, Interactive Simulation Belongs in the Pedagogical Toolkit of “Next Generation” Science: Facilitating Student Understanding of Complex Causal Dynamics. In: Information Resources Management Association, ed. *STEM Education: Concepts, Methodologies, Tools, and Applications*. Hershey, PA: IGI Global, chap. 83, pp.1578–1597. Available from: <https://doi.org/10.4018/978-1-4666-7363-2.ch083>.
- [40] Voit, E.O., Shah, A.M., Olivença, D. and Vodovotz, Y., 2023. What’s next for computational systems biology? *Frontiers in Systems Biology*, 3. Available from:

<https://doi.org/10.3389/fsysb.2023.1250228>.

- [41] Wells, G., 2008. Learning to use scientific concepts. *Cultural Studies of Science Education*, 3(2), pp.329–350. Available from: <https://doi.org/10.1007/s11422-008-9100-6>.