# **CURRICULUM AND EDUCATION**

**Open Access** 

# Comparing learning outcomes of two collaborative activities on random genetic drift in an upper-division genetics course

John R. True<sup>1\*</sup> and Evan Abreu<sup>1</sup>

# Abstract

**Background** Random genetic drift is a difficult concept for biology undergraduates to understand. Active learning activities in a collaborative setting have the potential to improve student learning outcomes compared to traditional lectures alone and have been shown to help foster success for underrepresented students. However, few activities in this content area have been evaluated for effectiveness in improving student outcomes using peer-reviewed instruments backed by evidence of their validity and reliability. Our aim in this study was to use the Genetic Drift Instrument (GeDI) to evaluate and compare student learning gains in an upper division genetics course in which two different genetic drift activities, a faculty-developed collaborative exercise and a commercially published lab tutorial, were administered in an active-learning classroom with students working in small groups.

**Methods** The GeDI was administered in both pre- and post-testing in two semesters (n = 95 and 98 students), with the semesters differing in which activity was assigned. Instrument dimensionality, person and item fit, and reliability were evaluated using Rasch analysis. Hierarchical Linear Models (HLMs) with two-way interactions were fitted to assess whether being in a certain Intervention Type, Race/ethnicity, Gender, or First Generation Status affected learning gains. HLMs with three-way interactions were used to assess whether the activities benefited students of all backgrounds equivalently.

**Results** We found that the GeDI demonstrated unidimensionality, with high item reliability and relatively low person reliability, consistent with previous studies. Both the faculty-developed activity and the commercially available lab tutorial were associated with significant learning gains on genetic drift concepts. Students in the SimBio group had higher learning gains but the difference in effect size was small. No significant differences in learning gains were found between students from different demographic groups, and both activities appeared to benefit students of different backgrounds equivalently.

**Conclusions** The GeDI instrument could be improved by adding items that more consistently differentiate students of different ability levels, especially at high ability levels. The greater impact on learning gains in the SimBio group while statistically significant does not translate into actual meaningful differences in student understanding. While students of different background variables in the sample have equivalent learning gains and are benefitted equivalently by the different interventions, our interventions did not ameliorate inequities in genetic drift understanding as measured by the GeDI that were uncovered in pre-testing.

\*Correspondence: John R. True john.true@stonybrook.edu Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/A.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Keywords** Random genetic drift, Genetic Drift Inventory (GeDI), Evolution, Genetics, Rasch analysis, Concept inventory, Active learning

# Background

Evolution at the population level is defined by change in allele frequencies across one or more generations. Of the ways in which this can happen, (mutation, gene flow, natural or sexual selection, and random genetic drift), genetic drift is the most difficult concept for undergraduate biology students to understand (Andrews et al. 2012). Genetic drift has been identified by the National Research Council (2003) and AAAS (AAAS 2010) as one of the core concepts of biological literacy for undergraduates. Several learning objectives involving genetic drift are among a recently assembled set of concepts fundamental to undergraduate biology that draws on contributions from over 800 faculty (Hennessey and Freeman 2023; https://www.codon learning.com/). Understanding the concept of genetic drift involves accurate thinking related to randomness and sampling error. Students majoring in the natural sciences can often struggle with this type of thinking (Garvin-Doxas and Klymkowsky 2008; Masel 2012).

Misconceptions involving genetic drift are common among undergraduates. They include the notion that all change in populations involves natural (or sexual) selection (Beggrow and Nehm 2012), the idea that genetic drift is not evolution, the notion that since genetic drift is random, that it is an unpredictable phenomenon, and several others (Andrews et al. 2012, Price et al. 2014). Introductory biology students have been found to have various incorrect notions about what genetic drift is, confusing drift with mutation, natural selection, or gene flow (Nehm and Reilly 2007; Andrews et al 2012).

A deluge of information about human genomic variation and health constantly flows from the media. Students advancing to professions in biomedicine, as well as biology education, require a rigorous conceptual framework for making sense of these patterns of variation and how they got there. Genetic drift is strongly associated with the concept of neutral alleles, which represent a large proportion of the DNA sequence variation within populations. The neutral theory of molecular evolution is fundamentally based on the process of genetic drift and is situated at the core of modern evolutionary theory (Nei et al. 2010). An understanding of speciation and species differences also requires proficiency in the concept of genetic drift in that isolated populations will inevitably accumulate differences in allele frequencies due to genetic drift, some of which may eventually contribute to phenotypic differences between populations and incipient species.

The structure of genetic variation in humans, in which there is more variation within than between geographically defined populations, also ties into the concept of genetic drift in that frequency differences between variants (e.g. single nucleotide polymorphisms) that do exist between geographic populations are most likely due to genetic drift related to demographic history and not selection. In medical genetics, explanations of individual and population-level genetic differences that inform genomic medicine rely in part on this understanding (see e.g. Dudley et al 2012; Visscher et al. 2017).

In pioneering work on undergraduate student understanding of genetic drift concepts, involving open-ended questions, written surveys, and student interviews, Andrews et al (2012) found that students tended to follow a three stage progression in development of ideas about genetic drift. Stage I was characterized by undeveloped or novice ideas about both genetics and evolution. Stage II was characterized by confused conceptions, such as confusing genetic drift with other processes such as mutation and natural selection. Stage III was characterized by some development of correct genetic drift concepts but with some new misconceptions forming, such as the idea that genetic drift only occurs when certain events occur (e.g. population bottlenecks). These patterns parallel to some extent the progression of evolutionary thinking in undergraduates in general, with adaptationist thinking preceding more complex models involving both selection and drift (Beggrow and Nehm 2012).

#### **GeDI** instrument

Students are thought to struggle more with concepts related to genetic drift than concepts related to natural selection, but the latter has received more attention by researchers (Price et al. 2014). To address this gap, Price et al. (2014) developed the Genetic Drift Inventory (GeDI) as an instrument to assess student understanding of genetic drift before and after instructional intervention. The instrument consists of a set of 22 Agree/Disagree statements centered on several key concepts and several misconceptions about genetic drift (Andrews et al. 2012, Price et al. 2014). One of the first published studies utilizing the GeDI following its publication was by Price et al. (2016), which used the GeDI to assess the effectiveness of a group of introductory and upper-division courses across several undergraduate institutions

that used the SimBio lab tutorial "Genetic Drift and Bottlenecked Ferrets" (Herron et al. 2014) (n=19) compared to control courses that did not use it (n=5). Their study found that early stage students typically lacked any understanding of genetic drift, consistent with Stage I of Andrews et al. (2012). But following instruction, students had some recognition that distinct evolutionary processes exist, of which genetic drift is one (consistent with Stage II of Andrews et al. 2012), or students had some correct understanding of genetic drift concepts (consistent with Stage III of Andrews et al. 2012). Overall, Price et al. (2016) found that students in courses using the Sim-Bio lab tutorial showed substantial evidence for increased understanding of genetic drift concepts compared to the control courses.

Tornabene et al. (2018) administered the GeDI to 336 students in an upper division genetics class in a northeastern university, using Rasch analysis and a design to test whether item order impacted outcomes. Their results indicated that the GeDI was unidimensional, reflecting that it measures a single construct. They also found that the instrument showed high item reliability and that the items fit the abilities of the students. Person reliability, however, was low, suggesting that the instrument items are not able to differentiate between persons of different ability levels. In particular, Tornabene et al. (2018) recommended the development of additional items in order to better distinguish among persons of relatively high ability level. So far no revisions of the instrument have addressed this limitation.

#### Aims of this study

In order to improve genetic drift learning outcomes in an upper-division genetics course, we are interested in developing student activities centering on the process of genetic drift. In particular, active learning courses, and those that also focus on misconceptions, have been shown to improve course outcomes for underrepresented students in STEM courses (Theobald et al. 2020; Nehm et al. 2022). Here, we report a comparison of two student activities used in the course in consecutive spring semesters, with learning gains assessed in pre- and post-testing using the GeDI instrument. One activity was developed by the faculty (J.R.T.) and the other was the SimBio Bottlenecked Ferrets lab tutorial (Herron et al. 2014). Both activities were implemented in a group active-learning format, which has been demonstrated to improve learning gains in undergraduate STEM courses (Theobald et al. 2020; Nehm et al. 2022) and the SimBio tutorial has been shown in a separate study to increase student learning gains on genetic drift (Price et al. 2016). Given that the faculty-developed activity can be delivered at no cost to students, we are particularly interested in whether this activity can achieve equivalent or better learning gains than the student-purchased tutorial.

Specifically, our study explores the following research questions: (1) Does the GeDI instrument display robust or acceptable measurement properties in our sample of students? (2) Do upper-level undergraduate biology students gain more understanding of genetic drift concepts as measured by the GeDI from completing a SimBio activity compared to an instructor-developed activity? (3) Are student demographics associated with different learning gains on the GeDI independently of which activity they encountered? (4) Do both activities benefit students of different background variables equivalently?

This project to improve learning outcomes on genetic drift has been reviewed by IRB and has been given the classification: Not Human Research. (Details available upon request to authors).

#### **Course information**

We report the results of pre- and post-activity administration of the GeDI in an upper division ecological genetics course, one of three upper-division courses that can be used to satisfy the genetics requirements for the biology and biochemistry majors at a public R1 university in the northeast US. One of the pre-requisites for the course is an organismal/evolution/ecology lower division introductory biology course. Students entering the focal genetics course are expected to have had basic exposure to genetic drift in this introductory course or its equivalent at another institution.

The focal genetics course enrolls approximately 100 students each year (mostly third and fourth year biology or biochemistry majors). The course is delivered in hybrid format with asynchronous online lectures and 9–10 inperson activities, one hour and twenty minutes in length, that students work together in person to complete in permanent groups of 3–4. Course unit structure and relevant learning objectives and misconceptions (provided as a study aid in the syllabus), are listed in Appendix 1.

# Activities

In the two consecutive spring semesters, the GeDI was administered both before and immediately after the unit on Population Genetics, which is the third unit of the course and follows units on molecular genetics and Mendelian genetics.

In the Faculty-Developed Activity group, students were assigned to read Chapter 2 of Conner and Hartl (2004), which covers genetic drift. Students then completed a pre-activity homework assignment in which they constructed a concept map on metapopulations. Then the class met for one session of one hour and twenty minutes and students worked in groups of three or four to

complete the faculty-developed activity, which used a metapopulation framework with Wright's F-statistics (Wright 1965) to illustrate genetic drift. The complete pre-activity homework and in-class group activity instructions are provided in Appendix 2 and was newly developed for the semester in which it was administered. This activity was entirely devised by the faculty (J.R.T.). It had not been piloted previously and has undergone no refinement or peer-review. In this activity, each student group is given two regular decks of playing cards with the red and black cards representing two different alleles at a single locus. Each group's combined deck of 104 cards represents the allele pool of a subpopulation with the entire class representing a metapopulation. Groups initially shuffle their decks and then draw two-allele genotypes, tallying the genotype frequencies. Groups then report the allele and genotype frequencies on a large whiteboard and the starting F<sub>ST</sub> is computed for the metapopulation. Then the number of cards in each deck is drastically reduced, representing a population bottleneck in each subpopulation. Following this, groups then shuffle and draw genotypes from the reduced card deck and report out their allele and genotype frequencies and the post-bottleneck F<sub>ST</sub> is calculated for the metapopulation, with the expectation that  $F_{ST}$  will have increased due to genetic drift.

A slight increase in  $F_{ST}$  was found in the activity (Initial population  $F_{ST} = -0.008$ , Post-bottleneck population  $F_{ST} = 0.013$ ; the negative  $F_{ST}$  calculation was likely due to data recording error). The instructor briefly discussed the result with the class at the end of the activity. Students were graded on attendance and participation in the activity and the post-intervention GeDI questions were administered during the unit test about two weeks later.

In the SimBio Activity group, the genetic drift activity occurred at the same time in the course sequence. The Conner and Hartl (2004) textbook was not required in the second spring semester and the reading was not assigned. Students were instead required to purchase the online tutorial lab "Genetic Drift and Bottlenecked Ferrets" by SimBio (Herron et al. 2014; https://simbio.com/ content/genetic-drift/) at a price of about \$6.00 per student (for Spring 2024, the price has increased to \$7.50).

The SimBio Genetic Drift and Bottlenecked Ferrets lab tutorial uses simulations to illustrate the properties and dynamics of genetic drift. The endangered North American population of black-footed ferrets is the model example, with a neutral locus determining fur color as a marker representing genetic diversity. Part 1 of the tutorial introduces the example and shows how allele frequencies can change due to sampling error with different reductions of population size. Part 2 formally introduces the concept of genetic drift, as well as random mating and shows how allele frequencies among gametes and zygotes are related. Part 3 contrasts small and large populations and how population size is related to the likelihood of loss and fixation of alleles. Part 4 introduces the concepts of heterozygosity and effective population size and how different phenomena such as inbreeding and unequal sex ratio influence genetic drift. Part 5 allows students to experiment with different reserve designs (metapopulation structures) to determine which has the greatest likelihood of conserving genetic diversity. Finally, part 6 is a series of summative assessment questions.

Students were assigned to complete parts 1 and 2 individually prior to class, and during the class they worked together in groups of three to four students to complete parts 3–5. For the activity grade, students were assessed as a group on the questions in part 6 of the tutorial. The post-intervention GeDI questions were administered during the unit test about three weeks later.

In both semesters, due to a typographical error in the question file in the learning management system (Blackboard in the Faculty-Developed Activity group, Brightspace in the SimBio Activity group), one of the 22 questions was omitted from the analysis. This question was: "An increase in the proportion of beetles with short legs occurred because natural selection favored individuals with shorter legs." in Stem 7 (see Appendix 3). For the post test for both semesters, minor changes were made to the organisms and traits in two of the GeDI question stems in order to present students with some novel but parallel question versions that they had not seen before. In Stem 4, flies/plain or striped wings was changed to ladybugs/brown or red heads and in Stems 8-9, human nearsightedness was changed to human partial hearing loss (see Appendix 3). All other parts of these questions remained the same as the items published in Price et al. (2014).

Table 1 compares the concepts addressed by the two different activities and the GeDI. There is only a low amount of overlap between the two activities in the concepts they cover. The SimBio lab tutorial overlaps fairly strongly with the GeDI (see Price et al. 2016) whereas the Faculty-developed activity overlaps substantially less with the GeDI. The SimBio lab tutorial also overlaps with the GeDI in coverage of a number of misconceptions (see Price et al. 2016 Table 1) whereas the Faculty-developed activity does not address genetic drift misconceptions.

#### Methods

# Participants and sample

The study included 95 and 98 undergraduate students enrolled in an upper division genetics course in consecutive spring semesters (see above). Participants' demographics included individuals of different races/

# Table 1 Comparison of concepts covered by the two lab activities and the GeDI

Concept	Covered by faculty-developed activity	Covered by SimBio "Genetic Drift and Bottlenecked Ferrets"	Covered by GeDI
Genetic drift is random sampling error that occurs in every generation	No	Yes	Yes
Genetic drift causes loss of genetic variation within populations and hence increases differentiation between populations	Yes	Yes	Yes
Relationship between population size and strength of genetic drift	Yes	Yes	Yes
Effects of genetic drift can overwhelm effects of natural selection in small populations	No	Yes	Yes
Effect of population bottleneck on genetic drift	Yes	Yes	Yes
Definition of metapopulation	Yes	No	No
Computation of F statistics	Yes	No	No
Interpretation of F statistics	Yes	No	No

See also Price et al. 2016 Table 1

ethnicities, genders, and first generation college status. Due to the removal of some students for reasons noted later (see below), our sampling frame contains 72.5% of the total sample.

Race/ethnicity categories include American Indian or Alaskan Native, Asian, Black and African American, Hispanic or Latino, Race and ethnicity unknown, two or more races, US non-resident, and white. Gender categories included male and female. First generation Status was coded as yes or no. As the class had a small sample of many race categories, for ease of analysis the following categories were used: (1) White, (2) Asian, and (3) Under-represented Minority (URM). The URM group combined American Indian or Alaskan Native, Black and African American, and Hispanic or Latino categories. We acknowledge the problematic nature of combining groups from different backgrounds, but also sought to investigate outcome disparities based on interventions.

Within the combined sample of 193 students, to maximize statistical power by minimizing the number of variables in the race/ethnicity category, students indicating race and ethnicity unknown, two or more races, and US non-resident, were removed from the sample (n=38). Furthermore, students with missing data for first generation status (n=15) were also removed to eliminate the issues that arise from having a null category in the analysis.

# Data collection procedure

During each semester, the GeDI was administered twice: once immediately prior to the population genetics unit (pre-test) and once at the end of the unit (posttest). This allowed for the measurement of learning gains within each semester. Race/ethnicity, gender, and first generation status were provided from the institution subsequently. In the faculty-developed activity group, both pre-and post-test applications of the GeDI were given online on Blackboard with students only able to view one question at a time, with backtracking disabled. In the SimBio activity group, the pre-test was given online on Brightspace with students only able to view one question at a time, and the post-test was given on paper, with the entire test visible. The GeDI questions are organized into a series of stems, each with 1 to 4 questions (see Appendix 3). In the online versions, a question bank was made for each stem and questions within each stem were presented in a random order to each student. In both semesters, all students were given the GeDI question stems in the same order.

#### **Research Question 1**

Raw scores obtained from the GeDI were subjected to Rasch analysis using the program Winsteps version 3.68 (Linacre 2008) in order to achieve several objectives. First, Rasch analysis was used to transform raw scores into Rasch measures, providing interval-level measurements. This transformation ensured that the GeDI scores were more suitable for parametric statistical analyses. The Rasch measures were taken from the Person Scores output from Winsteps. More importantly, in order to answer Research Question 1 we used Rasch analysis to evaluate the GeDI as an instrument in terms of its ability to display robust or acceptable measurement properties in our sample of students.

#### Dimensionality analysis

Although the GeDI has been rigorously tested by its developers, and more recently using Rasch analysis (Tornabene et al. 2018), to answer Research Question 1 Rasch analysis was used to determine if the inferences generated by the GeDI measures accurately reflected student understanding of genetic drift concepts in our sample of students. This included testing for multi-dimensionality to ensure that the instrument measured a single underlying construct. Multidimensionality would indicate that the GeDI does not measure genetic drift as a single topic, but as multiple topics.

# Person and item outfit

Rasch analysis was used to measure item and person outfit MSQ (mean-square). These are measures that evaluate how well each item (question) in the GeDI and person (student) in the sample fit the Rasch model, which can allow for identification of questions or students that might be problematic and require removal. In general, item outfit MSQ values between 0.7 and 1.3 are good indicators of fitting the Rasch model (Boone et al. 2014). In the interest of preserving as much of the sample as possible we used criteria for person outfit established by Wright and Linacre specifying that while measures less than 0.5 and more than 1.5 may not be productive to measurement, only person measures greater than 2.0 are degrading to measurement (Boone et al. 2014). Two students were shown to have abnormally high misfit to a point where they were degrading the measurement and were removed from the sample.

# Wright map analysis

In a similar vein to item outfit, we used Rasch analysis to measure the breadth of student abilities that can be accurately measured by the GeDI items and visualized these measures using a Wright map. The Wright map depicts item difficulties and student abilities on the same scale. If the range of item difficulty measures captures all of the student abilities, we can be confident that the GeDI can measure students having different genetic drift knowledge. Additionally, any significant skew in the distribution of the items may indicate that the GeDI is too difficult or too easy for our sample. If there are many students below or above the range of the items, this may indicate that the GeDI items cannot precisely discern differences in student abilities above or below a certain score. By examining the distribution of the items on the Wright map, we can also see if the GeDI has a high concentration of difficult or easy questions, or ideally, a mix of both with most questions residing somewhere in the middle of the scale, indicating a well balanced spread of questions.

*Item and Person Reliability*Rasch analysis was used to test item reliability and person reliability. High item reliability ensures that all the items together effectively measure the construct, while high person reliability ensures that the instrument consistently measures the overall sample. Person reliability values below 0.8 indicate that the items may not fully distinguish student "abilities" and item reliability values below 0.9 often indicate that the sample is not large enough to determine the structure of the items and their difficulty or that there is a lack of items that measure the sample.

## **Research Questions 2 and 3**

To investigate Research Questions 2 and 3, hierarchical linear regression models (HLMs) were employed, controlling for demographic variables or intervention type depending on the question. Hierarchical linear models can allow us to isolate different variable interactions with changes in GeDI scores in order to investigate multiple interaction effects.

The analysis was performed via the R package "lme4" (Bates et al. 2021). In R, reference values were set for both the time and race/ethnicity variables. The reference time was set to pre-test to simplify interpretation of the model as a post-test reference value would depict a negative slope in the regression if students had improved over time. The reference value for the race/ethnicity category was set to white.

#### Two way interaction HLMs

To answer Research Questions 2 and 3, the primary models used raw or Rasch measures (to examine similarities between the measures) as the dependent variable with an interaction between improvement on the GeDI (Pre- or Post-test) and the intervention type the students received being modeled as the fixed effect. Race/ethnicity, gender, and first generation status were modeled as random effects in the analysis.

To answer Research Question 3, three additional models were made in which the interaction effect was changed. In each model instead of modeling the interaction between intervention and time, intervention was replaced with race/ethnicity, gender, or first generation status. In these models the intervention type was changed to random effects. The time variable in interactions allows us to model the change in GeDI scores (Preto Post-test) and going forward it will be referred to as the learning gains variable. Any significant p values for the interactions between learning gains and demographic variables would indicate that there were differences in learning gains for that background variable.

An important consideration in answering Research Question 3 rests on whether there were differences in student performances on the GeDI on the pre-test. If there were statistically significant differences in any of these background variables on the pre-test, but we found no significant interactions between the background variables with learning gains, we could infer that our interventions did not negatively affect students of different background variables but nevertheless we would not be mitigating differences in genetic drift knowledge that may exist between students with different background variables. We can detect if differences between GeDI scores exist at the pre-test for students of different background variables by referring back to the first model discussed. For the HLM that modeled the interaction between improvement on the GeDI with intervention, background variables with intercepts with significant p values indicate that there was variability in the outcome variables that was clustered within these groups. If statistically significant results were found for the interaction variables in any of the models, the R package "sJplot" was used to plot the model.

# **Research Question 4**

#### Three way interactions

To answer Research Question 4, a separate hierarchical linear model was conducted with a three way interaction. This model included interactions between race/ethnicity, learning gains, and intervention, while controlling for gender and first generation status.

# Effect size

Partial omega squared values were computed to determine the effect sizes of the interactions. This was done using the R package "effectsize" (Ben-Shachar et al. 2020). To determine the strength of the effect size, we used the standard cutoffs of a partial omega square greater than 0.60 as being a high effect size, greater than 0.30 as being at least a medium effect size, and less than 0.30 as being a small effect size.

# Results

# **Research Question 1**

To address Research Question 1 (Does the GeDI instrument display robust or acceptable measurement properties in our sample of students?), Rasch analysis was employed.

# Item and person reliability

Fitting the data to the Rasch model yielded an item reliability score of 0.95 and a person reliability score of 0.69. Based on these results, the accuracy in item difficulty was high and the person measures, while within acceptable range, suggest that the items may have some limitations in distinguishing among students at different ability levels.

#### Dimensionality analysis

Research Question 1 also involves determining whether the transformed data fit the Rasch model assumption of unidimensionality. Typically we can measure this by looking at the output for the amount of unexplained variance explained by the first contrast in the Winsteps multidimensionality output. An empirical value of 2 or greater indicates 2 or more dimensions in the items. The value for our data was 1.8, which is evidence of a lack of multidimensionality. This also lends strength to our Rasch transformed scores. The plot of the standardized residual contrast 1 contains structure. While two items regarding human genetic drift, two out of five of human genetic drift questions, deviate from the structure (A and B or Items 18 and 19), the eigenvalue for the amount of unexplained variance in the first contrast does not support these two items being a separate dimension (Fig. 1).

#### Person and item outfit

All items were within the acceptable outfit range of 0.7-1.3.

#### Wright map analysis

The Wright map containing both pre- and post-tests exhibited an approximately normal distribution of item difficulties, although most items fell into a midrange difficulty and there were fewer items at the extremes of the difficulty ranges. Overall, observation of the Wright Map (Fig. 2) shows that there was a lack of balance between less difficult and more difficult items. We found that many students were located above the most difficult item measures, indicating that the GeDI Rasch measures cannot strongly distinguish between students with the highest GeDI scores. Overall, we have limited power in what we can say about students with high genetic drift knowledge.

# **Research Question 2 and 3**

Research Question 2: Do upper-level undergraduate biology students gain more understanding of genetic drift concepts as measured by the GeDI from completing a SimBio activity compared to a faculty-developed activity? and Research Question 3: Are student demographics associated with different learning gains on the GeDI independently of which activity they encountered? To answer these questions, we fit several hierarchical linear models to analyze the interactions of several different variables on the improvement of GeDI scores. To verify that the Rasch analysis did not significantly change the model, separate models were made which substituted Rasch measures with Raw scores.

# Two way interaction HLMs: learning gains and intervention type

Modeling the interaction between the intervention type and learning gains measured in either Rasch or Raw measure produced statistically significant results for the STANDARDIZED RESIDUAL CONTRAST 1 PLOT



**Fig. 1** Contrast in the residuals for GeDI items. Letters on the plot represent different items. Counts along the bottom of the plot refer to the number of items at that position of item measure. The x-axis indicates item Rasch Measures in logits. The y-axis indicates loadings for the first principle component in the correlation matrix of the residuals. The clustering pattern of the loadings is an indication of dimensionality. The key on the right of the plot aligns the letters with their corresponding item

interactions (p=0.000619 and p=0.00695). Rasch measures are used in all subsequent models (Table 2).

We calculated partial omega squared values for the significant interaction between learning gains and the activity type and found a value of 0.03. This constitutes a small effect size. The plot of this regression shows similar scores on pre-test between the two activity groups and different measures on the post-test although the error bars overlap (Fig. 3).

# Two way interaction HLMs: learning gains and background variables

The modeling of the interaction between being Asian or URM and the improvement on the GeDI was not statistically significant (p=0.8994, p=0.8486) (Table 2). Modeling the interaction between gender and the learning gains on the GeDI was also not statistically significant (p=0.1899). Finally, the interaction between first generation status and the improvement on the GeDI was again not statistically significant (p=0.8908). In summary, there were no significant differences in learning gains for any of the background variables we examined (Table 2). However it should be noted that in the first HLM, in which all background variables were modeled as random

effect variables, the intercepts for gender (p=0.0244) and first generation status (p=0.0237) were significant for a critical value of 0.05 but the intercepts for race/ethnicity variables were not significant (Asian: p=0.1535, URM: p=0.1822). This indicates that variability in learning gains between the two activities is not explained by race/ ethnicity.

#### **Research Question 4**

# Three way interaction HLMs: learning gains, intervention type, and background variables

The final Research Question of the study was: Do any background demographic variables interact with intervention type to impact learning gains? To answer this question, a model was built to assess the presence of a three way interaction between activity type, learning gains, and URM status. This was done to verify that the statistically significant interaction between the improvement and activity type did not affect students of different race/ethnicity differently. Again, these interactions were shown to not be statistically significant (p=0.4761 and p=0.7053) for Asian and URM students respectively (Table 3). Similarly the three way interactions that replaced race/ethnicity with gender or first generation



**Fig. 2** Wright map indicating item and person measures on logit scale. Person (student) measures are on the left. Each "#" represents 3 persons. Each "". represents 1 person. Each person is represented twice on the map, once for their pre-test submission and once for their post-test submission. Each item is placed on the right (For example item 1 is labeled as Q1). Persons higher up on the scale are inferred to have higher ability. Items higher on the scale are inferred to be more difficult. The Rasch model predicts that a student should answer items correctly that are at the same logit scale value and below them. Wright maps of pre- and post-test results from both activity groups are shown in Appendix 4. A comparison table of item difficulties in pre- versus post-tests is given in Appendix 5

Status were shown to be not statistically significant (p=0.8510, and p=0.9317) (Table 3). No demographic variables were shown to be associated with how students improve on the activity types.

# Discussion

# **Research Question 1**

The Rasch analysis conducted in this study enables evaluation of the GeDI as a measurement instrument and builds upon previous research, offering additional insights into the GeDI's capabilities and limitations.

# Strengths of the GeDI

Item and Person Outfit: The analysis revealed a key strength of the GeDI: strong fit of items and persons as evaluated by Outfit MSQ. None of the items were answered so easily correct or unpredictably that they deviated from the Rasch model which would be reflected by an Outfit MSQ value outside of the acceptable 0.7–1.3 range while we used a looser range of less than 2.0 for person Outfit MSQ measures. This consistency is underscored by the fact that only two students, representing a mere 1% of respondents, showed outfit measures outside this acceptable range. This finding is consistent with that of Tornabene et al. (2018), who found that no GeDI item had an Outfit MSQ outside the previously established range, reinforcing the reliability of the instrument's items.

*Unidimensionality:* Another significant strength of the GeDI is adherence to the unidimensionality assumption of the Rasch model. Our data, corroborated by good item and person fit, indicates unidimensionality with an eigenvalue of 1.8 for unexplained variation in the first contrast. This result closely matches the findings of Tornabene et al. (2018), where 1.8 was also the eigenvalue for unexplained variation in the first contrast, further validating the GeDI's capability to measure a single construct.

#### Weaknesses of the GeDI

*Potential Multidimensionality:* A potential concern arises with the clustering of items related to genetic drift

	Intervention type	Race/ethnicity	Gender	First generation status
Rasch measure learning gains	p=000619*	p=0.8994 (Asian), 0.8486 (URM)	p=0.1889	p=0.8908
Raw measure learning gains	p=0.00695*	p=0.6875 (Asian), 0.3724 (URM)	p=0.4012	p=0.2549

Learning gains were measured by both Rasch measures and raw scores to determine whether results were consistent. Separate regressions were used to measure the impact of intervention type, race/ethnicity, gender, or first-generation status on the learning gains and p values are reported for each interaction Interactions with a p value less than the critical value of 0.01 are indicated by an asterisk



Fig. 3 Rasch measures in the two activities for both pre-(red) and post-tests (blue). Rasch measures in logits are plotted for HLM model with all background variables modeled as random effects. The interaction between learning gains and the activity type is statistically significant (p < 0.001). Vertical lines indicate 95% confidence intervals

in humans. This clustering, evident in the Standardized Residual Contrast 1 plot (Fig. 1), suggests a possible separate dimension. While the existence of multidimensionality is not supported by the eigenvalue for the unexplained variation in the first contrast, to explore this further, integrating additional items on human genetic drift into the GeDI could be beneficial.

Disparity in Reliability: The GeDI exhibited high item reliability (0.95) but low person reliability (0.69), a dichotomy that presents a major limitation in discerning between student abilities in understanding genetic drift. This pattern echoes the findings of Tornabene et al. (2018), who also reported low person reliability (0.62) and high item reliability (0.97). The root cause in the current study, evident in the Wright Map, is the misalignment of person and item measures, primarily due to a lack of challenging questions.

Comparison to Previous Studies: When compared to previous studies, our analysis revealed a more pronounced misalignment in person and item measures, with the instrument failing to differentiate among 44.6% of high-ability respondents. This issue was less severe in Tornabene et al. (2018), where only 12.5% of responses were inadequately distinguished due to the absence of high difficulty items. However, our inflated number compared to past results is likely due to our inclusion of the post-test data which included higher scores, on average, than pre-test only data. Indeed 76.2% of student's with person measures greater than Item 8, the most difficult item, were person measures from the post-test. To enhance the effectiveness of the GeDI in both measuring initial understanding and tracking learning gains over time, a larger question bank containing questions of more varied and increased difficulty would likely be helpful.

In summary, while the GeDI demonstrates robustness in certain aspects, it also reveals areas in need of improvement to better serve as a comprehensive tool for assessing and tracking student understanding of genetic drift and related concepts.

#### Table 3 Results of three-way HLMs

	Race/ethnicity	Gender	First generation status
Improvement on the GeDI and intervention type interaction	p=0.47608 (Asian), 0.70533 (URM)	p=0.851012	p=0.931759

Three regressions were used which modeled the interaction between learning gains on the GeDI and intervention type with each of the background variables. Significant p values would indicate that the interventions did not benefit students of different background variables similarly (Research Question 2). None of the three-way interaction regressions resulted in p values lower than the critical value of 0.01

#### **Research Question 2**

Our findings regarding Research Question 2 (Do upperlevel undergraduate biology students gain a better understanding of genetic drift concepts as measured by the GeDI after completing a SimBio activity compared to an instructor-developed activity?) have several dimensions.

#### Effect size of learning gains

The analysis revealed that students who participated in the SimBio activity, on average, answered 1.63 more questions correctly than those who engaged in the faculty-developed activity. However, the calculated omega square, an effect size measure, was only 0.03 for the significant interaction between learning gains and the activity type. This low partial omega square suggests that although the difference in intervention effectiveness on the GeDI score improvement is statistically significant, it does not translate into a meaningful enhancement in student understanding. However, there is a caveat to this conclusion. This study provided additional evidence for a lack of questions difficult enough to distinguish between the higher ability students. Therefore the Rasch model treats all students with measures above the highest difficulty question similarly. Given that 44.6% of the students fell into this category, it is possible that there were meaningful learning gains in very high ability students that were not accounted for in our omega square value calculation. More work is needed on this topic with a GeDI that has more difficult items.

Content that appears on the GeDI and content covered in both the faculty-developed activity and SimBio activity overlap in three concepts (Table 1). Interestingly, analysis of the person measures derived from the four items that Page 11 of 21

cover these overlapping topics alone (Items, 1,3,11, and 14), yields similar results for the two activities. Although an unpaired two-tailed t-test indicates that there were statistically significant differences (p=0.00033) in the mean learning gains between students in the facultydeveloped activity group (mean=1.09 logits) and the SimBio activity group (mean=1.71 logits), the effect sizes as measured by Cohen's D are remarkably similar for the faculty-developed activity (Cohen's D=1.13) and the SimBio activity (Cohen's D=1.19). This is consistent with the analysis of the entire set of 21 items, which showed that while there are statistically significant differences in learning gains, the differences in effect sizes are very small. Future research into how students improve on different GeDI concepts should account for and seek to confirm and contrast learning gains associated with components of interventions with differential overlaps in GeDI concepts. Additionally as learning gains were seen on items that were not covered in the lectures consistent with Price et al. (2016), further research is also needed on identifying the causes of these learning gains and whether they are the result of lecture materials and intervention alone or whether there are some other ways that students process information about genetic drift concepts.

#### Contextualizing with previous studies

We observed larger Cohen's D values (Table 4) for our learning gains than were reported for GeDI scores in other studies using SimBio or the Genie as an intervention (Price et al. 2016, Castillo et al. 2022). This difference between our results and the previous two studies may be underestimated because past studies did not analyze

#### Table 4 Results from this study and past studies using the GeDI

Study group	Pre-test mean (standard deviation)	Post-test mean (standard deviation)	Cohen's d (Lower 95% Cl Upper 95% Cl)
Faculty-developed activity	12.54 (1.01)*	16.48 (1.27)	1.05 (0.719, 1.38)
SimBio activity	12.75 (1.03)*	18.11 (1.29)	1.55 (1.18, 1.92)
(Price et al. 2016) control	11.66 (1.64)	11.17 (1.45)	- 0.04 (SD=0.2)
(Price et al. 2016) SimBio activity	13.13 (2.02)	15.38 (2.69)	0.63 (SD=0.59)
(Tornabene et al. 2018) Northeastern University	12.35 (3.59)	NA	NA
(Castillo et al. 2022) 2016 Genie Activity	14.18 (3.62)	16.46 (3.88)	0.608 (0.408, 0.807)
(Castillo et al. 2022) 2017 Genie Activity	17.18 (3.84)	19.42 (3.23)	0.632 (0.410, 0.855)
(Castillo et al. 2022) 2017 No Genie Activity	16.09 (4.16)	18.61 (3.47)	0.658 (0.430, 0.886)

Pre- and Post-Test average scores are reported in raw scores and standard deviations are reported parentheses next to the scores. Learning gains are reported in Cohen's D measures with 95% confidence intervals in parentheses. Tornabene et al. (2018) administered the GeDI only once to students as measuring learning gains was not an objective of that study so Post-Test average score and Cohen's D value are omitted due to absence. Price et al. (2016) reported standard deviation for the Cohen's D rather than confidence interval so standard deviation is reported in the table instead

\*21/22 items were used in the current study because of a clerical error during entry of one item into the learning management system

GeDI Rasch measures with HLMs and thus might have overestimated learning gains measured by the GeDI.

#### Activity features

The two student activities involved substantially nonoverlapping subsets of genetic drift concepts (Table 1). The faculty-developed activity was not designed as an exercise in using all genetic drift concepts, but instead was designed to illustrate specifically how a bottleneck event in a metapopulation causes genetic drift and how this can be detected by using the  $F_{ST}$  statistic. The use of a bottleneck example in this activity could have the effect of enhancing the tendency of students to adopt or maintain the misconception that drift is only caused by major demographic events and is not a constantly occurring process. This corresponds to misconception #11 in Table 1 of Price et al. (2016) and is covered in the SimBio lab tutorial, but is not covered by the GeDI. Therefore, additional questions would need to be added in order to determine whether students who were administered the two different activities showed learning gains in this area.

The importance to evolution of subdivided populations, what are now called metapopulations, was first highlighted by Sewall Wright (1931) and the concept has increased in importance in evolutionary and conservation biology (Hanski 1998; Akcakaya et al. 2006). However, studies on student learning in this area appear to be lacking. Both of the activities in this study involve metapopulation concepts related to genetic drift but they are mostly non-overlapping and can be considered complementary. Therefore, it might be beneficial to students to administer both activities in the same course.

# Research Question 3 and 4 Exploring demographic variables

Our study also investigated the impact of race/ethnicity, gender, and first generation status on genetic drift knowledge as evaluated by GeDI scores, a topic previously explored by Castillo et al. (2022) using two-way ANOVA. Castillo et al. found that first generation status was a statistically significant predictor (p=0.037) for GeDI posttest scores in one semester, and gender was a statistically significant predictor (p=0.030) for GeDI pre-test scores in a different semester.

We found similar results using HLMs. It is worth noting that for several reasons HLM regression is a more robust method for detecting these differences in GeDI scores among background variables. HLM is specifically designed for analyzing data with nested structure, while two way ANOVA does not inherently account for nested data structures and therefore may not accurately model the variability due to these hierarchical relationships. Additionally, as HLM accounts for the hierarchical structure of the data, it can provide more accurate estimates and inferences about the effects of the predic-

can be more generalizable. Our HLM regression found that there was variability in the GeDI scores for gender (p=0.0244) and first generation status (p=0.0237) for the same critical value used by Castillo et al. (2022). In our study, in the faculty-developed activity, males scored on average 2.09 points higher than females on the pre-test, and in the SimBio group, first generation students scored on average 2.84 points lower than non-first-generation students on the pre-test. This adds evidence to inequalities in prior knowledge for gender and first generation status variables for students.

tors on the outcome variable. Finally, due to its ability to

account for variability at multiple levels, HLM models

The prior results do not answer the question of whether the interventions differentially affected learning gains for different groups of students. Our study, the first to use Rasch analysis and HLMs for this purpose, revealed that these background variables did not significantly affect student learning gains, thus answering Research Question 3 (Are student demographics associated with different learning gains on the GeDI independently of which activity they encountered?). Additionally, the improvement in learning gains observed with the SimBio activity over the faculty-developed one was consistent across different demographic groups, allowing us to answer Research Question 4 (Do both activities benefit students of all background variables equivalently?).

However, due to the imbalance in our sample in terms of ethnicity (White: n=32, URM: n=43, Asian: n=82), gender (female: n = 103, Male: n = 54), and first generation status (first Generation: n = 50, non-first-generation: n = 107), further research in this area is warranted especially because our results do not necessarily conflict with previous studies (Castillo et al. 2022) due to the fact that we measured background variables as a predictor of learning gains rather than as predictors of scores at any one time point. Additionally it should be noted that since this study was conducted in an upper division course, it is possible that much of the student diversity that may have contributed to different results, may have already been lost in earlier biology classes. Most importantly, however, while our interventions did not affect students of different background variables differently, the activities did not mitigate pretest differences. More effective interventions that allow for these disparities to be ameliorated are needed.

# Broader perspectives on student learning about genetic drift

Lastly, we sought to compare our pre and post-test scores and gains on the GeDI with past studies (Table 4).

Consistent with previous findings, our sample scored similarly on the pre-test to other studies utilizing the GeDI (Price et al. 2016, Tornabene et al. 2018), but less so compared to Castillo et al. (2022). Performance similarities allow comparisons to previous work. Specifically, a critical aspect of this comparison involves the progression model of genetic drift understanding proposed by Andrews et al. (2012). This model delineates a threestage development of ideas about genetic drift, in which students typically progress from Stage I (broad misconceptions) to Stage II (conflating genetic drift with other evolutionary mechanisms) and potentially to Stage III (characterized by some development of correct genetic drift concepts but with some new misconceptions forming, such as the idea that genetic drift only occurs when certain events occur).

#### Pre-test findings: stage I misconceptions

Our analysis of the pre-test results revealed that a significant portion of students harbored Stage I misconceptions about genetic drift and its role in evolution. For instance, approximately half of the students incorrectly answered item 4, which probes the understanding of genetic drift in relation to evolution (47.3% in the faculty-developed activity group and 50.0% in the SimBio group). This pattern aligns with findings from Price et al. (2016), indicating this common initial misunderstanding among upper-division biology students.

#### Post-test developments: transition to stage II

The post-test results, however, indicated a notable shift. A majority of students moved beyond Stage I misconceptions, as evidenced by improved performance on item 4 (81.1% in the faculty-developed activity group and 87.8% in the SimBio group). This suggests a progression to Stage II, consistent with the previous results mentioned by Price et al. (2016).

#### Challenges in advancing to stage III

The transition from Stage II to Stage III, however, was less evident. item 8, identified as the most challenging in the GeDI by the Rasch analysis, was still answered incorrectly by about half of the students in both on both the pre- and post-tests for the faculty-developed activity (Pre: 41.1%, Post: 52.6%) and for the SimBio tutorial (Pre: 44.4%, Post: 66.6%). While the SimBio group showed more significant gains in moving past Stage II, these were not as pronounced as the gains observed in items associated with Stage I to II progression. It should be noted that the format of the GeDI is not well suited for making strong or detailed inferences about Stage III (Price et al. 2016).

# Questioning the linearity of progression of stages

The difficulty students had with Stage III misconceptions, particularly evident in their responses to item 9, suggests that the progression of understanding genetic drift may not be as linear as initially proposed by Andrews et al. (2012). For example, a higher than expected number of correct responses were observed on this Stage III item in the pre-test (68.4% in the faculty-developed activity group and 64.4% in the SimBio group), compared to items targeting Stage I and II misconceptions with lower frequencies in the sample of correct answers, indicating that many students grasp some ideas associated Stage III level understandings before they fully grasp Stage I and II concepts. This divergence from the expected linear progression supports the notion that the existing model by Andrews et al. (2012) might require revisions to accommodate this nonlinear progression in understanding, a finding also consistent with Price et al. (2016).

# Limitations

Overall, given their lack of overlap and the lack of overlap of the metapopulation activity administered in the faculty-developed activity group with the coverage of the GeDI, the two activities should not necessarily be considered alternatives to each other. Thus, the current study is limited to those concepts and misconceptions that the GeDI encompasses. With respect to the GeDI, the group that completed the faculty-developed activity can be considered a sort of baseline with which to compare the results of administration of the SimBio lab tutorial in the SimBio Activity group, since it has a substantial overlap with the GeDI. In this perspective, the greater learning gains in the SimBio Activity group, while significant, appear to be modest, and most of the total learning gain exhibited in both semesters may be attributable to treatment of genetic drift in the class lectures. Alternatively, there may have been a general benefit of applying genetic drift concepts to various types of problems in a group active learning environment. These two potential explanations are not mutually exclusive.

In the group that completed the faculty-developed activity, the Conner and Hartl (2004) textbook was listed as required for the course and chapter 2 from this book, which covers genetic drift, was assigned as reading for the population genetics unit. In the group that completed the SimBio activity, this book was not listed as required for the course and the reading was not assigned. Textbook material was incorporated into the lectures, which were identical in the two semesters, and no other assignments were given from the text in the faculty-developed activity group. Typically, many students do not complete the textbook readings and some students complete the course without buying the textbook. The number of students using the textbook in the faculty-developed activity group is unknown, but given that genetic drift learning gains were slightly better in the SimBio group than in the faculty-developed activity group, it is unlikely that the textbook substantially affected genetic drift learning in the semester in which it was assigned.

In the faculty-developed activity group, the GeDI preand post-test were both administered online in the Blackboard learning management system, with the post-test as part of the online unit test. Questions were presented one at a time and students were not able to backtrack. The Brightspace learning management system was used in the SimBio Activity group for the GeDI pre-test administration, again with one at a time presentation without backtracking. However, in the SimBio Activity group the unit test, including the GeDI post-test items, was administered in person, which allows students to see the whole test at once and backtrack. Backtracking may be used by students to attempt to use information from one part of an exam to answer a question in another part (Budhai 2020). Backtracking is considered a way to reduce student stress on exams (Novick et al. 2022) and has been associated with increases in student time management efficiency and online exam performance (Matea and Weidenhofer 2021). The ability to backtrack on the SimBio Activity group GeDI post-test may thus have contributed to the higher mean score on that administration of the instrument.

It is possible that the inadvertent omission of item 17 may have impacted our results. This item focuses on the specific misconception that "Natural selection is always the most powerful mechanism of evolution, and it is the primary agent of evolutionary change" (See Appendix 3; Price et al. 2016). Three other items in the GeDI (10, 13, and 20; see Appendix 3) also address this misconception. The exclusion of item 17, therefore, is not expected to substantially reduce coverage of this misconception in our administration of the GeDI. Moreover, while previous Rasch analysis performed by Tornabene et al. (2018) ranked this question as one of the more challenging ones in the inventory, it has not been shown to be the most difficult. Thus it is unlikely that its inclusion would have allowed our Rasch analysis to distinguish between the highest ability students.

Our implementation of the SimBio "Genetic Drift and Bottlenecked Ferrets" lab tutorial differed from a published implementation by Whitely et al. (2016), who used the tutorial in a Team-Based Learning (TBL; Michaelsen and Sweet 2011) flipped classroom setting. In the TBL framework, students were assigned an outside reading assignment before class, and then took a "Readiness Assurance Test" first as individuals, and then in groups, prior to undertaking the tutorial. At the end of the tutorial in the TBL framework, students took the graded questions at the end of the tutorial (Part 6) as individuals. In our framework, only one semester involved outside reading (see above), which was not specific to the activity, and students individually completed the first two sections of the tutorial on their own before class. At the end of class, students took the graded questions as a group, not as individuals. The TBL framework has generally been found to be associated with significant learning gains (Liu 2016) but there are no published studies on learning gains by students completing the SimBio genetic drift tutorial in the TBL framework.

# Conclusions

This study contributes insights into the understanding of genetic drift and the efficacy of the Genetic Drift Inventory (GeDI) as a measurement tool. It corroborates previous findings while also offering new perspectives on the learning process of genetic drift concepts.

Our analysis through Rasch modeling underscores the robustness of the GeDI as an assessment instrument. We found no indications of items being too predictable or easy, nor did we detect any clear signs of multidimensionality in the instrument. These results affirm that the GeDI effectively measures a single construct and demonstrates high item reliability. However, a noted limitation of the GeDI is its lack of sufficiently challenging questions, which hampers its ability to differentiate among the highest-ability students. A large proportion of students scored above the most difficult item in the instrument, indicating a need for more challenging items.

In comparing the effectiveness of the SimBio activity with the faculty-developed activity, our study found that students generally performed better on the GeDI after engaging in the SimBio activity. Despite this, the small effect size of this difference suggests that the enhanced learning gains with SimBio did not greatly surpass those achieved through the faculty-developed activity. Our overall findings lend support to the notion that active learning strategies can enhance student understanding of genetic drift. It is possible that the addition of more difficult items to the GeDI could widen the difference in performance between students exposed to the two activities, depending on which concept or misconceptions the new items addressed. For example, items covering effective population size would align with the SimBio activity but not the faculty-developed activity whereas items covering metapopulations could potentially align with both activities.

Our study also explored the influence of background variables, such as gender and first generation status, on genetic drift learning. We observed that while both the SimBio and faculty-developed activities yielded similar learning gains across different student demographics, initial disparities in understanding based on gender or first generation status that existed in undergraduate students persisted in the post-test phase. This indicates that, despite the effectiveness of these interventions, equity gaps in understanding genetic drift remain unaddressed.

Revisiting the three-stage framework proposed by Andrews et al. (2012), our findings reinforce the concept that the progression of understanding genetic drift is not a straightforward, linear process, which is consistent with the findings of Price et al. (2016). Students often exhibit a mix of misconceptions from earlier stages while simultaneously grasping concepts from more advanced stages. This suggests that gaining an understanding of genetic drift is a complex, multifaceted process.

In light of these findings, future research in genetic drift education should approach the topic as a nuanced and layered learning pathway, seeking strategies that address the diversity in student understanding of genetic drift concepts. Implementing approaches beyond in-class activities may assist in promoting equitable learning in this content area.

# **Appendix 1**

Ecological genetics course information relevant to this report.

A. Course Unit structure, Faculty-Developed Activity group

Unit 1: Basic Genetics (included both molecular genetics and Mendelian Genetics/Inheritance) Unit 2: Population Genetics (included Faculty-developed activity with GeDI pre- and post-testing) Unit 3: Natural Selection Unit 4: Quantitative Genetics Unit 5: Special Topics

B. Course Unit structure, SimBio Activity group.

Unit 1: Basic Molecular Genetics Unit 2: Mendelian Genetics/Inheritance Unit 3: Population Genetics (included SimBio lab tutorial with GeDI pre- and post-testing) Unit 4: Natural Selection Unit 5: Quantitative Genetics Unit 6: Special Topics C. Population Genetics Unit Learning Objectives (both groups).

Lecture 1: The Hardy-Weinberg Principle

- Be able to describe the assumptions of the Hardy-Weinberg model for a non-evolving population
- Be able to calculate allele frequencies from genotype frequencies given genotype frequency data from a population.
- Be able to describe how genotype frequencies reach Hardy-Weinberg expectations after one generation of random mating.
- Be able to describe why allele frequencies do not change given the assumed characteristics of a population in Hardy-Weinberg equilibrium.

Lecture 2: Linkage Disequilibrium

- Be able to define linkage disequilibrium and describe two distinct causes of linkage disequilibrium in natural populations.
- Be able to calculate the expected genotype frequencies given the allele frequencies of two unlinked loci each with two alleles.
- Be able to calculate the linkage disequilibrium between two marker loci.
- Be able to define haplotype and describe the importance of haplotypes in the study of human ancestry.

Lecture 3: Inbreeding and Relatedness

- Be able to calculate and interpret the inbreeding coefficient of a population.
- Be able to read use a pedigree to calculate the inbreeding coefficient of a focal individual.
- Be able to interpret coefficient of relatedness values between close relatives and distinguish between the coefficient of relatedness and the coefficient of kinship.
- Be able to describe the effect of inbreeding on genotype frequencies and allele frequencies.

Lecture 4: Genetic Drift

- Be able to define random genetic drift and explain how random genetic drift causes evolution of a population.
- Be able to explain the relationship between population size and the strength of genetic drift.

- Be able to explain the terms 'bottleneck effect' and 'founder effect' and provide examples of these phenomena.
- Be able to define effective population size and to use equations to calculate effective population size in cases of: (1) unequal numbers of males and females, (2) differential variance in male and female reproductive success, and (3) time series data.
- Be able to infer by comparisons of graphs of frequency through time which of two or more neutral alleles is in the smallest/largest population.

Lecture 5: Metapopulations and F-statistics

- Be able to describe what a metapopulation is and why this concept is important to biologists.
- Be able to define what population structure is.
- Be able to compute expected heterozygosity if there were no population structure in a metapopulation given allele frequencies in the subpopulations.
- Be able to use equations to calculate the three F-statistics in Sewall Wright's metapopulation model.
- Be able to interpret  $F_{ST}$  and  $F_{IS}$  values from metapopulation data and explain why they are important statistics.

D. Population Genetics Unit Misconceptions (provided in syllabus as a study aid) (both semesters).

- 1. Genotype frequencies will change due to segregation of alleles each generation.
- 2. Calculating whether a population is in Hardy-Weinberg equilibrium at a gene locus is not useful for anything
- 3. A calculation at a single gene locus is all that is necessary to determine whether a population is in Hardy-Weinberg equilibrium
- 4. If an allele of one locus is more likely to be found with a specific allele at another locus than expected by chance, this must be due to close linkage between the two loci
- 5. Inbreeding is rare in nature
- 6. All cases of higher than expected homozygote frequencies are die to mating with relatives
- 7. All cases of higher than expected heterozygote frequencies are due to the mating system
- 8. Genetic drift does not cause evolution
- 9. Very large populations do not undergo genetic drift

10. Any member of a population has an equal chance of mating with any other member of the population

# **Appendix 2**

Instructions for Faculty-developed pre-activity homework and in-class activity on metapopulations and genetic drift (bottlenecks), using F-statistics.

Pre-Activity Homework (students work individually) Instructions (on PowerPoint slide):

In the space to the right, construct a concept map explaining metapopulations using the following concepts: subpopulation, metapopulation, migration, population size, F[ST], F[IS], allele frequencies, heterozygosity, genetic drift, inbreeding, random mating.

For an example of a concept map, see (https://cdn1. byjus.com/wp-content/uploads/2021/09/evolutionconcept-map.png).

RULES:

You must use all the concepts

Each concept must be connected to at least one other concept by an arrow.

Each arrow must have at least brief text explaining the connection.

You may use concepts more than once and you may bring in additional concepts, if they help you explain metapopulations.

When you are finished, save this slide as a pdf and submit it at the Blackboard submission link.

# In-class activity (students work in groups of 3–4) Instructions:

STEP 1: Obtain a whiteboard, markers, and eraser for your group. Each group should assign one group leader. Your group is a part of a subpopulation with 2–3 other groups that are led by your TA. Your subpopulation is part of the greater course metapopulation which consists of 8 subpopulations. Once you obtain a deck of 200 cards (alleles), IMMEDIATELY begin to shuffle the deck by laying the cards face down on the table and mixing them around. Then, have each group member draw 25 individual diploid genotypes from the card pile for a total of 100 individuals by taking two cards at a time. There are two alleles, red and black. Tally the genotypes that your group draws (RR, RB, BB). You will share this number with the other groups in your subpopulation (your TA will let you know which groups to work with). STEP 2: Each group leader will collaborate with the other group leaders in the subpopulation to communicate the total genotype numbers. Group leaders will then communicate this information to groups, and working together as a group, calculate the following information: (a) total genotype numbers (RR, RB, BB) (b) allele frequencies (p=frequency of R, q=frequency of B) (c) observed heterozygosity (#RB/total) (d) expected heterozygosity (2pq).

STEP 3: Show your work to your TA, who will write the data for your subpopulation on the large whiteboard.

STEP 4: Once all subpopulation data has been written on the large whiteboard, working together as a group, spend 10–15 min to calculate the F[IS] and F[ST] of the initial generation on your whiteboards. Check your work with your TA and then take a picture of your whiteboard for submission to Blackboard.

STEP 5: Immediately begin to thoroughly shuffle the cards in the middle face down once more. A hurricane is going through the island archipelago where this metapopulation is and all subpopulations will now undergo a population bottleneck. After the deck is thoroughly shuffled in the same way as previously (representing random mating), each group member will draw 3 genotype pairs (for a total of 12 individuals per group). Record the genotypes and share this information with other groups in the subpopulation as you did before.

STEP 6: Each group leader will collaborate with the other group leaders in the subpopulation to communicate the total genotype numbers. Group leaders will then communicate this information to groups, and working together as a group, calculate the following information the same way as in the previous generation: (a) total genotype numbers (b) allele frequencies (c) observed heterozygosity (d) expected heterozygosity (2pq).

STEP 7: Show your work to your TA, who will write the data for your subpopulation on the large whiteboard.

STEP 8: Once all subpopulation data have been written on the whiteboard, working together as a group, calculate the F[IS] and F[ST]. Check your work with your TA and then take a picture of your whiteboard for submission to Blackboard.

STEP 9: Designate one member to submit the pictures showing your work on Blackboard, and working together, answer the questions.

# Appendix 3

#### Key to item numbers in the GeDI.

Stem 1 (used for Questions 1-4):

A small island is home to a unique population of land snails. This population was founded by 10 individuals that floated to the island on a log, and has been isolated from the large mainland population ever since. The mainland population has consistently had about 10,000 individuals in it through time. The island population reached 1000 individuals after several generations, and then stayed at this size through time.

A team of researchers compared the genetic variation of the mainland and the isolated island populations for a few generations after colonization. Would a biologist agree or disagree with the following statement?

1. The island population likely has fewer alleles, that is, versions of genes, than the mainland population.

2. Some harmful traits may have become more common in the island population than the mainland population.

3. Genetic drift is more pronounced in the island population than the mainland population in these first few generations.

4. The biologists observed genetic drift but not evolution.

Stem 2 (used for Questions 5-7):

(Same first paragraph as Stem 1)

After forty generations, biologists measured the genetic variation of the isolated island snail population again. They calculated that the population of snails on the island had remained isolated and that genetic drift had occurred.

Would a biologist agree or disagree with the following statement about *the processes that contributed, at least in part, to genetic drift in the population of land snails?* 

5. The fact that snails needed to adjust to the environment contributed to genetic drift.

6. The fact that individuals that were best suited to the environment had a higher rate of survival contributed to genetic drift.

7. Random survival of different individuals could not have contributed to genetic drift because random processes are unpredictable.

Stem 3 (used for Question 8)

(Same first two paragraphs as Stem 2)

Would a biologist agree or disagree with the following statement about what occurred after forty generations in the isolated population?

8. The island population may have adapted to conditions on the island through genetic drift.

Stem 4 (used for Questions 9-11)

A biologist raised 100 populations of flies in a lab. At the beginning of the experiment, each population had 16 flies: 8 with plain wings and 8 with striped wings. These 16 flies reproduced to form the first generation of offspring. In each of the 100 populations, the biologist randomly chose 16 of the offspring as breeders for the next generation. She repeated this process for 20 generations. At the end of the experiment, half of the populations contained only plain-winged flies, and the other half contained only striped-wing flies. Wing pattern is a genetically controlled trait that does not affect how well flies survive of reproduce.

Would a biologist agree or disagree with the following statement about the experimental results?

9. The populations were isolated from each other so genetic drift could not have caused the results.

10. The experiment did not control for all the variables, so the environments were different enough that natural selection contributed to the changes in the frequency of the two wing types in these populations.

11. The small number of individuals reproducing each generation contributed to the rapid changes in the frequency of the two wing types in these populations.

Stem 5: (used for Questions 12–14)

A population of 1000 dung beetles was split into five populations when irrigation canals were build through their habitat. The five new populations were called the Northern, Southern, Eastern, Western, and Central populations. Each new population consisted of about 200 individuals. The five populations continued to evolve, and no migration occurred among populations. One hundred generations later, each population still has about 200 individuals, and a biologists investigates them.

What would a biologist expect to see in the five populations after 100 generations if the environment did not change for any of the populations?

12. Each population would probably gain new mutations through genetic drift.

13. Differences among the five populations probably occurred when populations adapted to specific environments because most evolution results in adaptation.

14. Each population would probably have fewer alleles, that is, versions of genes, than it would have had 100 generations ago.

Stem 6: (used for questions 15–16)

(Same first paragraph as Stem 5)

Would a biologist agree or disagree with the following statement?

15. Chance survival of some individuals occurred in some generations, but not every generation.

16. In the smaller populations of 200 individuals, the processes leading to genetic drift could have a stronger influence on a gene than natural selection.

Stem 7: (used for questions 17–18)

(Same first paragraph as Stem 5)

All species of dung beetles lay their eggs in balls of dung. Long legs allow the beetles to create better dung balls, which improves reproductive success. Long legs were common before the canals were built. However, after the populations were separated, long legs became less common in the most southern population.

Would a biologist agree or disagree with the following statement?

17. Since there was no migration there could be no genetic drift.

Additional stem 7 question not analyzed due to typographical error in question bank: An increase in the proportion of beetles with short legs occurred because natural selection favored individuals with shorter legs.

Stem 8: (used for Question 18)

A disorder that causes nearsightedness is caused by a genetic mutation. Nearsightedness was harmful to people living on a small island because they relied on sight to interpret their surroundings. In the 1600 s, a huge storm killed many of the people on the island. Before the storm, 0.1% of the people had this disorder. Of the 50 people who survived the storm, 2% were nearsighted. Within a few generations, 10% of the islanders were nearsighted.

Would a biologist agree or disagree with the following statement about the high rate of nearsightedness in the islanders after the storm?

18. The high rate was caused by new mutations that resulted in genetic drift.

Stem 9: (used for Questions 19–21)

(Same first paragraph as Stem 8)

A biologist concludes that the change in frequency of nearsighted individuals could be evidence of genetic drift. Would she agree or disagree with the following statement?

19. Nearsightedness must have become more common through new mutations.

20. Nearsightedness must have become more common through natural selection.

21. Nearsightedness must have become more common through people migrating from neighboring islands.

# Appendix 4

**Individual Wright maps of pre- and post-test data** See Fig. 4.



**Fig. 4** Individual Wright maps of results of **A** Faculty developed activity group pre-test, **B** SimBio activity group pre-test, **C** Faculty-developed activity group post-test, **D** SimBio activity group post-test. See Fig. 2 legend for Wright map details. Each "#" represents 2 persons. Each "" represents 1 person. While some of the logit scales are unaligned this is inconsequential as measures from separate Rasch Analysis cannot be directly compared. Instead interpretation must be restricted to the relative positioning of items compared to each other on different Wright maps

# **Appendix 5**

Comparison of pre- and post-test difficulty ranks of the 21 GeDI items used. Ranks were determined by doing separate Rasch analyses in Winsteps and extracting item measures. Items with the highest measures in the

analysis are the most difficult items and those with the lowest measures are the least difficult items. A rank of 1 indicates that item was the most difficult in that iteration of the GeDI while a rank of 21 indicates the item was the least difficult in the iteration of the GeDI.

ltem	Pre test difficulty rank	Post test difficulty rank
1	14	17
2	17	18
3	21	21
4	5	14
5	8	12
6	1	6
7	18	13
8	2	1
9	16	10
10	13	8
11	11	15
12	4	5
13	6	2
14	3	3
15	7	4
16	19	20
17	20	19
18	12	9
19	9	7
20	10	11
21	15	16

#### Abbreviations

- GeDI Genetic Drift Inventory instrument
- HLM Hierarchical linear regression model
- MSQ Mean-squared
- TBL Team based learning
- URM Under-represented minority

#### Acknowledgements

We thank R. Nehm for many helpful ideas, suggestions, and criticisms during this project and the writing of this manuscript. We also thank G. Sbeglia for valuable assistance and advice on coding and analysis for this project. We are also grateful for comments and suggestions from two anonymous reviewers of an earlier version of this manuscript that helped to greatly improve it. E.A. is supported by a Dr. W. Burghardt Turner Fellowship from the Stony Brook University Graduate School.

#### Author contributions

JRT administered the instrument, collected and collated the data, and cowrote the manuscript. EA analyzed the data and co-wrote the manuscript.

#### Funding

This study was funded by Stony Brook University.

#### Availability of data and materials

Data (anonymized) are available upon request to the authors.

# Declarations

#### **Competing interests**

The authors have no competing interests.

#### Author details

<sup>1</sup>Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, USA.

Received: 28 November 2023 Accepted: 29 January 2024 Published online: 19 February 2024

#### References

AAAS. (2010). Vision and change: a call to action. Washington: AAAS.

- Akcakaya HR, Mills G, Doncaster CP. The role of metapopulations in conservation. In: Macdonald D, Service K, editors. Key topics in conservation biology. Oxford: Blackwell Publishing; 2006. p. 64–84.
- Andrews TM, Price RM, Mead LS, McElhinny TL, Thanukos A, et al. Biology undergraduates' misconceptions about genetic drift. CBE Life Sci Educ. 2012;11:248–59.
- Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models usinglme4. J Stat Softw. 2015;67:1–48.
- Beggrow EP, Nehm RH. Students' mental models of evolutionary causation: natural selection and genetic drift. Evolution: Education and Outreach. 2012;5:429–444.
- Ben-Shachar MS, Makowski D, Ludecke D, Patil I, Wiernik BM, Theriault R. Effectsize (Version 0.8.6) [RPackage]. Cran. 2020.
- Boone WJ, Staver JR, Yale MS. Rasch analysis in the human sciences. Dordrecht: Springer; 2014.
- Budhai SS. Fourteen simple strategies to reduce cheating on online examinations. Faculty Focus, May 11, 2020. https://www.facultyfocus.com/artic les/educational-assessment/fourteen-simple-strategies-to-reduce-cheat ing-on-online-examinations/.

Castillo AI, Roos BH, Rosenberg MS, Cartwright RA, Wilson MA. Genie: an interactive real-time simulation for teaching genetic drift. Evol Educ Outreach. 2022;15:3. https://doi.org/10.1186/s12052-022-00161-7.

- Conner JK, Hartl DL. A primer of ecological genetics. Sunderland: Sinauer Associates; 2004.
- Dudley JT, Kim Y, Liu L, Markov GJ, Gerold K, Chen R, Butte AJ, Kumar S. Human genomic disease variants: a neutral evolutionary explanation. Genome Res. 2012;22:1383–94.
- Garvin-Doxas K, Klymkowsky MW. Understanding randomness and its impact on student learning: lessons learned from building the biology concept inventory (BCI). CBE Life Sci Educ. 2008;7:227–33.
- Hanski I. Metapopulation dynamics. Nature. 1998;396:41-9.
- Hennessey KM, Freeman S. Nationally endorsed learning objectives to improve course design in introductory biology. bioRxiv. 2023. https://doi.org/10. 1101/2023.10.25.563732.
- Herron JC, Maruca S, Meir E. Genetic Drift and Bottlenecked Ferrets. Missoula: SimBio; 2014.
- Linacre MA. Users guide to winsteps/ministep Rasch model computer programs. Program Manual 4.0.0. 2008; http://www.winsteps.com/a/Winst eps-ManualPDF.zip.
- Liu S-NC. The effectiveness of team-based learning: a meta-analysis. Dissertation Baylor University Masters Thesis; 2016. https://baylor-ir.tdl.org/items/ a78be00b-1bb6-4909-925a-3d7591654a9c.
- Masel J. Rethinking Hardy-Weinberg and genetic drift in undergraduate biology. BioEssays. 2012;34:701–10.
- Matea K, Weidenhofer J. Alternative to paper-based examinations for assessment of human physiology? Proc Aust Conf Sci Math Edu. 2021;29:78–83.
- Michaelsen L, Sweet M. Team-based learning. In: Michaelsen LK, Sweet M, editors. New directions for teaching and learning. Hoboken: Wiley Periodicals, Inc.; 2011.
- National Research Council. BIO2010: transforming undergraduate education for future research biologists. Washington: National Academies Press; 2003.
- Nehm RH, Reilly L. Biology majors' knowledge and misconceptions of natural selection. Bioscience. 2007;57:263–72.
- Nehm RH, Finch SJ, Sbeglia GC. Is active learning enough? The contributions of misconception-focused instruction and active-learning dosage on student learning of evolution. Bioscience. 2022;72:1105–17.
- Nei M, Suzuki Y, Nozawa M. The neutral theory of molecular evolution in the genomic era. Annu Rev Genomics Hum Genet. 2010;11:265–89.
- Novick PA, Lee J, Wei S, Mundorff EC, Santangelo JR, Sonbuchner TM. Maximizing academic integrity while minimizing stress in the virtual classroom. J Microbiol Biol Educ. 2022;23(1):e00292-e321. https://doi.org/10.1128/ jmbe.00292-21.

- Price RM, Andrews TC, McElhinny TL, Mead LS, Abraham JK, Thanukos A, Perez KE. The Genetic Drift Inventory: a tool for measuring what advanced undergraduates have mastered about genetic drift. CBE Life Sci Educ. 2014;13:65–75.
- Price RM, Pope DS, Abraham JK, et al. Observing populations and testing predictions about genetic drift in a computer simulation improves college students' conceptual understanding. Evo Edu Outreach. 2016;9:8. https:// doi.org/10.1186/s12052-016-0059-6.
- Theobald EJ, Hill MJ, Tran E, et al. Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. Proc Natl Acad Sci USA. 2020;117:6476–83. https://doi.org/10.1073/pnas.1916903117.
- Tornabene RE, Lavington E, Nehm RH. Testing validity inferences for Genetic Drift inventory scores using Rasch modeling and item order analyses. Evol Educ Outreach. 2018;11:6. https://doi.org/10.1186/ s12052-018-0082-x.
- Visscher PM, Wray NR, Zhang Q, et al. 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet. 2017;101:5–22.
- Whitely AR, Steinberg E, Meir E. Flipping virtual labs into team-based learning tools. Proc Assoc Biol Lab Educ. 2016;37:22.
- Wright S. Evolution in Mendelian populations. Genetics. 1931;16:97–159.
- Wright S. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution. 1965;19:395–420.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.