1   **Genie: An interactive real-time simulation for teaching genetic drift**

2   Andreina I. Castillo [1], Ben H. Roos [2], Michael S. Rosenberg [3], Reed A. Cartwright [2,4*], and

3   Melissa A. Wilson [2,4,5*]

4

5   [1] Department of Environmental Science, Policy and Management. University of California,

6   Berkeley.

7   [2] The Biodesign Institute, Arizona State University.

8   [3] Center for the Study of Biological Complexity. Virginia Commonwealth University.

9   [4] School of Life Sciences. Arizona State University.

10   [5] Center for Evolution and Medicine. Arizona State University.

11

12

13

14

15

16

17

18   * To whom correspondence may be addressed.

19   Email: cartwright@asu.edu and mwilsons@asu.edu

20   **Abstract**

21   Neutral evolution is a fundamental concept in evolutionary biology but teaching this and

22   other non-adaptive concepts is specially challenging. Here we present Genie, a browser-based

23   educational tool that facilitates demonstration of concepts such as genetic drift, population

24   isolation, gene flow, and genetic mutation. Because it does not need to be downloaded and

25   installed, Genie can scale to large groups of students and is useful for both in-person and online

26   instruction. Genie was used to teach genetic drift to Evolution students at Arizona State

27   University during Spring 2016 and Spring 2017. The effectiveness of Genie to teach key genetic

28   drift concepts and misconceptions was assessed with the Genetic Drift Inventory developed by

29   Price et al. (2014). Overall, Genie performed comparably to that of traditional static methods

30   across all evaluated classes. We have empirically demonstrated that Genie can be successfully

31   integrated with traditional instruction to reduce misconceptions about genetic drift.

32

33

34

35

36

37

38

39

40   Keywords: Genetic Drift, Simulations, Evolution, Education.

1. Introduction

A well-recognized challenge in biological science education is successfully teaching evolutionary concepts (Alters and Nelson 2002). However, even within the same discipline, some topics remain more challenging to teach than others, and the number and efficacy of tools available for instruction varies (Shulman 1987; Ziadie and Andrews 2018). For instance, multiple strategies have been developed to improve the teaching of concepts like natural selection (Ziadie and Andrews 2018). On the other hand, best practices to teach equally important topics such as non-adaptive evolution remain largely understudied (Kalinowski et al. 2013). This is particularly problematic for topics like genetic drift because concepts of adaptive and non-adaptive evolution form independent elements in evolutionary thinking, and better understanding of one does not necessarily implies better comprehension of the other (Beggrow and Nehm 2012). To address this, studies devoted to developing, improving, and testing teaching strategies for non-adaptive evolutionary concepts are needed.

Previous studies have created approaches aimed to identify student misconceptions regarding genetic drift (Andrews et al. 2012; Price et al. 2014), and study activities and software have been developed, tested, and made publicly available (Price et al. 2016; Revell 2019; Staub 2002). These serve as indicators that the knowledge gap regarding genetic drift instruction is being addressed. Nonetheless, diverse class environments, student cohorts, and even teaching styles require distinct sets of tools; therefore, new tools are important for improving evolutionary instruction. Furthermore, there is an academic push for improving the teaching strategies currently set in place and to utilize alternative instruction methods (Lee et al. 2017; Nelson 2008; Tanner and Allen 2005). In particular, teaching strategies that favor discussion and testing of evolutionary concepts among student have been shown to be most effective (George M. Slavich

3

64   and Zimbardo 2012). As a result, tools that can be used to facilitate free in-class exploration of

65   evolutionary concepts, are especially useful since they allow students to both learn these

66   concepts and develop critical thinking skills.

67   Here, we developed a web application (Genie) designed to demonstrate several

68   population genetics and evolutionary notions including genetic drift, gene flow, and random

69   mutation. This application conducts a real time simulation of the change in allele frequencies in a

70   finite population of spatially isolated individuals. Using colors, the application allows students to

71   visualize changes in a population over time and understand how those visual changes translate to

72   fluctuations in allele frequency, and eventually, fixation or loss of an allele. This web-based

73   software is accessible to students and leads to increased knowledge of genetic drift concepts, as

74   tested using a Genetic Drift Inventory (Price et al. 2014). These types of assessments have

75   proven to be useful in capturing student's understanding of other complex evolutionary concepts

76   in the past (Perez et al. 2013). The Genie software requires no startup other than navigating to a

77   web page, thus making the use of programmed stochastic simulations to demonstrate the concept

78   of genetic drift practical to both educators and students in face-to-face and online instruction.

79

80   **2. Methods**

81   *2.1 Genie simulation program*

82   Genie (https://cartwrig.ht/apps/genie/) is a web-based, stochastic simulation app written

83   in JavaScript. The simulation uses a spatially explicit Moran Model (Nei et al. 1976) to describe

84   a finite population of 1,024 individuals on a 32 by 32 grid. Individuals are haploid with a single

4

85   locus. The locus mutates according to the infinite alleles model (Nei et al. 1976). Genie works as

86   follows:

87   • *Population Initialization.* The simulation begins when a population is randomly initialized

88   according to Hoppe's Urn (Perez et al. 2013). Briefly, the population is created one

89   individual at a time, and each individual either carries a new, unique allele or is a copy of a

90   previously created individual. The probability that individual $i + 1$ has a new allele is $\theta/(\theta+i)$

91   and the probability that the individual copies an existing allele is $(i)/(\theta+i)$, where $\theta = 2N\mu$, N

92   is the population size, and μ is the mutation rate. If an individual copies an allele, it randomly

93   chooses a previously initialized individual uniformly. At initialization μ is $= 0.001$ to ensure

94   diversity within the initial population, but the mutation rate of each generation can be

95   specified by the user, defaulting to 0.

96   • *Algorithm.* At each step of the simulation, a randomly selected individual dies, leaving its

97   corresponding cell momentarily empty. A parent allele is then randomly selected from the

98   eight immediate neighboring cells (including adjacent and diagonal). Cells on the edges and

99   corners of the simulation have fewer neighbors than internal cells, causing a small edge

100   effect. The probability that a new individual will have the same allele as its parent its 1-μ,

101   and the probability that an individual has a new, unique allele is μ. Each 'generation' consists

102   of 2000 death/birth steps after which the population is redrawn in the visualization window.

103   • *Running.* The application contains four components: a grid, where the population is displayed

104   (Supplementary file 1a); a control panel, where users can manipulate the simulation's

105   mutation parameter (Supplementary file 1b); an upper graph, where users can see the number

106   of alleles in the population at any given time (Supplementary file 1c); and a lower graph,

107   where users can see the frequency of different alleles at any given time (Supplementary file

108   1d). Both graphs update in real time as the simulation runs. Each initial allele is assigned one

109   of 18 basic colors, while each mutant allele is assigned one of six neon colors. A single

110   button allows users to toggle between starting the simulation or pausing it. A reset button

111   allows users to restart and reinitialize the simulation at any point.

112   • *Barriers*. Users can create a barrier in the population grid. To do so, users alter a cell (by

113   clicking on it) or alter a set of cells (by clicking and dragging the cursor to select multiple

114   cells). When a barrier is created, the color associated with the cell changes to black. Barriers

115   act neither as parent cells (they are never replicated) nor die. Thus, for each created barrier

116   cell the total population size declines by one. By building barriers, users can construct

117   physical constraints that restrict the movement of alleles between subpopulations. Barriers

118   can be used to create subpopulations of different sizes and shapes, as well as to study the

119   effects of corridors on gene flow. Barriers can be removed by clicking on the chosen cell(s) a

120   second time; this will set the cell color to white and designate the cell as unoccupied.

121   Neighboring cells will replicate into unoccupied cells; unoccupied cells cannot serve as a

122   parent of a neighboring cell.

123   • *Forced Mutation*. Users can force a mutation to occur in a manner similar to creating

124   barriers. Cells can be mutated by holding the SHIFT button while clicking the cell, or while

125   clicking and dragging the cursor across several cells. Forcing a mutation immediately creates

126   a new, unique allele in each of the chosen cell(s).

127

128   *2.2. Data collection*

129   Genie's efficacy as a tool for teaching Genetic Drift concepts was tested in the Evolution

130   (BIO345) class at Arizona State University (ASU). Genie was used during the practical portion

6

131   (recitation) of the BIO345 course in the Spring 2016 and Spring 2017 classes. All participants in

132   the Spring 2016 class used Genie during practical class sessions. In the Spring 2017 class, half of

133   the participants used the dynamic visualization of Genie while the other half used static

134   illustrations. Participants in both the Spring 2016 and Spring 2017 classes were given the option

135   to opt-in to the study at the end of the semester. In addition, participants were given the option to

136   provide their demographic information: reported gender, reported ethnicity, and first-generation

137   college student status. All research was reviewed and approved by Arizona State University's

138   IRB protocol STUDY00003707.

139   The impact of Genie as a tool for teaching concepts of genetic drift was evaluated using

140   the Genetic Drift Inventory (Price et al. 2014). The inventory was used without changes (22

141   questions assessing different aspects of genetic drift) in pre- and post-recitation assessments. The

142   pre- and post-recitation assessments (considered as homework for the entire class) were

143   individually answered by each participant. The pre-recitation assessment was posted online on

144   Blackboard two days before recitation. Participants were asked to answer all questions by 3:00

145   pm of the day of the recitation. The post-recitation assessment was posted on Blackboard at 9:00

146   pm after the last recitation session ended. Participants had two days to individually complete the

147   post-recitation assessment. All participants were allowed the same amount of time to complete

148   both the pre- and post-recitation assessments. Participant's answers were recorded, and their

149   individual pre- and post-recitation scores were calculated by summing the number of correctly

150   answered questions (value 1 point) out of the 22 questions in the Genetic Drift Inventory.

151   *2.3. Genie assessment*

152   The complete dataset was divided into two major groups based on instruction year. These

153   groups were: the entire Spring 2016 class (henceforth referred to as Genie 2016) and the entire

7

154   Spring 2017 class. The 2017 class was further subdivided into groups based on the instruction

155   method used during the practical class session. These groups were: participants that used Genie

156   during the recitation session in 2017 (henceforth referred to as Genie 2017) and the participants

157   who did not use Genie during the recitation session in 2017 (henceforth referred to as Non-Genie

158   2017). The Genie 2016 class was subsequently divided into eight in-class groups of roughly

159   equal size, while each 2017 class was divided into four in-class groups of roughly equal size

160   (two Genie and two Non-Genie). The groups were designated based on recitation start times,

161   Graduate Teaching Assistants (TA) pairs; and in the case of 2017, on the use of dynamic (Genie)

162   vs. static (Non-Genie) instruction methods. No more than 48 participants participated in each

163   recitation session. All analyses and figures were developed using R v3.2. The code and datasets

164   used are available (Supplementary files 2-12,

165   https://github.com/AndreinaCastillo/Genie_manuscript_data_analysis).

166   The putative relationship between participants' demographics and the pre- and post-

167   recitation scores was evaluated using a two-way ANOVA. The following demographic

168   parameters were used as explanatory variables: reported gender, reported ethnicity, and first-

169   generation college student status. In the case of 2017, the use of Genie as an instruction tool was

170   also considered as an explanatory variable. The two-way ANOVA was performed independently

171   for Genie 2016, Genie 2017, and Non-Genie 2017. Next, we assessed if the pre- and post-

172   recitation performance varied between the three class groups or among subgroups within each

173   class. To conduct this analysis, the distribution of pre- and post- recitation scores was assessed

174   using the 'fitdistrplus' (Delignette-Muller and Dutang 2015) and 'betareg' (Cribari-Neto and

175   Zeileis 2010) R packages. Potential differences between pre- and post-recitation scores were

176   evaluated both between classes and within each in-class group. In addition, a Cohen's d was used

8

177 to measure the effect size between pre- and post-recitation scores within each class, and to

178 estimate differences in pre- and post- recitation scores between Genie 2017 and Non-Genie 2017.

179   In addition, potential differences in groups of participants based on their initial

180 performance levels following instruction were assessed. Participant's scores within Genie 2016,

181 Genie 2017, and Non-Genie 2017 were divided into four quantiles based on their pre-recitation

182 scores. The first quantile included participants with scores ranging from 0 to 0.25, the second

183 quantile included participants with scores between 0.26-0.5, the third quantile included

184 participants with scores of 0.51-0.75, and the fourth quantile included participants with scores of

185 0.76-1. For each quantile within each class, a paired Student's t-test was performed in order to

186 evaluate if participants with different performance levels (i.e. within each quantile) benefited

187 differently from the use of Genie. Furthermore, a paired Student's t-test was performed between

188 individual participants' pre- and post-recitation scores within each class.

189   Finally, question-specific performance was evaluated to determine how Genie aided

190 participants in addressing the specific genetic drift concepts and misconceptions listed in the

191 Genetic Drift Inventory (Price et al. 2014). The number of correct answers in pre- and post-

192 recitation sessions associated with each question were calculated from participants' individual

193 answers, and the totals where then compiled by class. Differences between pre- and post-

194 recitation scores for each question were assessed using a McNemars χ^2 test. In addition, the

195 difference in the number of correct answers per question in Genie 2017 vs. Non-Genie pre- and

196 post-recitation sessions was assessed using a Fisher's exact test.

197

198

3. Results

Demographic representation varied among cohorts (Table 1). Participants identifying as People of Color ('POC') (N = 168) were less represented compared to participants identifying as 'White' (N = 238) in Genie 2016. Both groups were roughly equal in Genie 2017 ('POC' = 136 and 'White' = 144) and Non-Genie 2017 ('POC' = 120 and 'White' = 112). On the other hand, participants identifying as 'Female' (N = 230 in Genie 2016, N = 190 in Genie 2017, and N = 140 in Non-Genie 2017) were more numerous than participants identifying as 'Male' (N = 176 in Genie 2016, N = 90 in Genie 2017, and N = 92 in Non-Genie 2017). Likewise, 'First-generation' college students were less numerous (N = 126 in Genie 2016, N = 46 in Genie 2017, and N = 36 in Non-Genie 2017) than 'Not First-generation' college students (N = 280 in Genie 2016, N = 234 in Genie 2017, and N = 196 in Non-Genie 2017). Regardless of these differences, pre- and post-recitation performance levels were similar in participants from different demographic backgrounds across the three evaluated class groups (Figure 1).

A two-way ANOVA found that most demographic explanatory variables did not affect pre- and post-recitation scores (Table 2). For instance, the variable 'Reported gender' did not have a clear effect on pre- or post-recitation scores in either year (Genie 2016 pre-recitation $F = 1.004$, p-value = 0.318; Genie 2016 post- recitation $F = 2.388$, p-value = 0.124; 2017 pre-recitation $F = 0.037$, p-value = 0.848; 2017 post- recitation $F = 0.010$, p-value = 0.921). Likewise, the variable 'Reported ethnicity' did not have a clear effect in pre- ($F = 1.046$, p-value = 0.397) or post-recitation scores ($F = 0.372$, p-value = 0.896) in Genie 2016, or the post-recitation scores ($F = 0.895$, p-value = 0.511) in 2017. However, it did have a significant effect ($F = 2.286$, p-value = 0.029) in pre-recitation scores in 2017.

221   The variable 'First-generation' college had a statistical effect in pre- (F = 4.142, p-value

222   = 0.043) and post-recitation scores (F = 7.955, p-value = 0.005) in Genie 2016, but not in 2017

223   (pre- recitation F = 0.466, p-value = 0.496; post- recitation F = 0.260, p-value = 0.610). 'First-

224   generation' college students had lower pre-recitations scores than 'Not First-generation' students

225   in Genie 2016 and Genie 2017, while the opposite trend was observed in Non-Genie 2017

226   (Figure 1). While post-recitation scores were higher than pre-recitation scores in both groups,

227   'First-generation' college students showed slightly less improvement than 'Not First-generation'

228   college students. For instance, 'First-generation' college students saw an increase of 0.09

229   compared to 0.13 for 'Not First-generation' college students in Genie 2016. The same trend was

230   observed in Non-Genie 2017 (0.13 vs. 0.18 for 'First-generation' college students and 'Not First-

231   generation' college students, respectively) and Genie 2017 (0.13 vs. 0.14 for 'First-generation'

232   college students and 'Not First-generation' college students, respectively). It should be noted that

233   the number of 'First-generation' college students was higher in Genie 2016 (N = 126) as well as

234   Genie 2017 (N = 46), compared to Non-Genie 2017 (N = 36). The 'Genie used' variable showed

235   a significant effect (F = 6.131, p-value = 0.014) in pre-recitation scores in 2017 but not in post-

236   recitation scores (F = 2.350, p-value = 0.127). The difference between mean pre-recitation scores

237   between the Genie and Non-Genie 2017 classes was small (0.0495, or approximately 1 out of 22

238   questions), with Genie 2017 having a higher mean pre-recitation score (0.7808) than Non-Genie

239   2017 (0.7313). The difference in post-recitation scores was 0.0368, with Genie 2017 still having

240   a higher mean (0.8828) than Non-Genie 2017 (0.8460).

241   Overall, pre- and post-recitation scores were different among the classes analyzed. The

242   mean pre-recitation score for Genie 2016 (0.6444) was lower than in either Genie 2017 or Non-

243   Genie 2017 (see above). This same trend was maintained for post-recitation scores in Genie 2016

244   (0.7481) compared to Genie 2017 or Non-Genie 2017 (see above). Differences in post-recitation

245   scores could be largely explained by the initial differences in pre-recitations scores, or in other

246   words, the initial class performance (Table 3). Furthermore, in-class groups showed similar

247   performance levels among Genie 2016, Genie 2017, and Non-Genie 2017. An outlier to this

248   observation was the 'TA Pair1 7:30pm' in-class group (p-value=0.017) during Genie 2016,

249   which was composed exclusively of honor students. Overall, the density curve of post-recitation

250   scores showed that participant performance improved in all classes regardless of the instruction

251   method used (Figure 2). Furthermore, Cohen's d values (Table 4) showed a moderate

252   improvement in post-recitation scores compared to the pre-recitation scores in Genie 2016

253   (0.608, CI: 0.408-0.807), Genie 2017 (0.632, CI: 0.410-0.855), and Non-Genie 2017 (0.658, CI:

254   0.430-0.886). In addition, Cohen's d values also showed a small difference in pre- (0.272, CI:

255   0.051-0.493) and post-recitation (0.242, CI: 0.021-0.462) scores between Genie 2017 and Non-

256   Genie 2017.

257   Post-recitation scores where higher than pre-recitation scores for most individual

258   participants within all classes (Table 5, Supplementary file 13). Furthermore, participants across

259   three out of four quantiles of initial pre-recitation scores showed improvement in their post-

260   recitation scores (Table 6). There were too few participants (1-2 individuals across all groups)

261   with pre-recitation scores between 0-0.25 (the lowest quantile), making unfeasible to statistically

262   evaluate this group. In Genie 2016, the highest improvement was observed in participants with

263   pre-recitation scores between 0.26-50 (T = -7.855, p-value = 4.294×10^{-11}). In Genie 2017 (T = -

264   7.118, p-value =1.191×10^{-09}) and Non-Genie 2017 (T = -7.714, p-value = 4.412×10^{-11}), the

265   highest improvement was seen in participants with pre-recitation scores between 0.51-0.75.

266 Participants from the top quantile (pre-recitation scores between 0.76-1), had the smallest

267 improvement in their post-recitation scores, particularly in Non-Genie 2017 (Table 6).

268   Participants understanding of key genetic drift concepts and misconceptions statistically

269 improved after instruction with or without Genie (Table 7, Figure 3). Post-recitation scores were

270 generally higher in Genie 2017 than in Non-Genie 2017 except for two questions (Q10 and Q15,

271 Figure 4). When the difference between pre- and post-recitation scores by question was plotted

272 (Figure 5), there was variation in which questions had a higher score improvement in Non-Genie

273 2017 or Genie 2017. Non-Genie 2017 showed higher improvements in questions related to key

274 concepts (Q1, Q10, and Q13) and misconceptions associated to genetic drift as natural

275 selection/adaptation/acclimation to the environment that may result from a need to survive (M2),

276 genetic drift as random mutation (Q9), and genetic drift as gene flow or migration (Q11). On the

277 other hand, Genie 2017 showed higher improvements in misconceptions related to natural

278 selection being always the most powerful mechanism of evolution, and as the primary agent of

279 evolutionary change (M4). However, a Fisher's exact test showed that instruction method (Genie

280 vs. Non-Genie) was not associated with student's switching answers from correct to incorrect or

281 incorrect to correct between pre- and post-recitation (Table 8). Results were comparable with or

282 without including students within honor sections (Supplementary file 14).

283

284 **4. Discussion**

285   There are numerous other software capable of generating genetic drift simulations. Some

286 of them can be easily downloaded and installed (Kliman et al. 2008; Revell 2019), others include

287 an ample array of parameters to be modified by the user

288   (http://evolution.gs.washington.edu/popgen/popg.html), and others can be found publicly

289   available online (e.g. the Genetic Drift Simulator

290   (http://www.biology.arizona.edu/evolution/act/drift/drift.html or Phyletica

291   (http://phyletica.org/teaching/drift-simulator/). Some of this software even have a dynamic

292   interface similar to that developed by Genie

293   (http://virtualbiologylab.org/NetWebHTML_FilesJan2016/RandomEffectsModel.html). One

294   notable software (Avida-ED) has a dynamic presentation and accessibility (https://avida-

295   ed.msu.edu/app/AvidaED.html) alike Genie, but has additional parameters that permit for the

296   evaluation of complex evolutionary hypotheses. While this list is not exhaustive, it provides a

297   glimpse on how computational tools, and especially those found freely in web-interfaces, are

298   becoming predominantly used in science teaching. The objective of this paper is not to compare

299   Genie's performance to all these tools, instead, the authors aim to present an additional teaching

300   tool that can be added to an instructor's repertoire. As such, we endeavor to show that Genie can

301   be efficiently used alongside teacher-centered class instruction. A comparison was made between

302   Genie-based instruction and instruction using static images (henceforth referred as teacher-

303   centered instruction). The comparison was chosen since teacher-centered methods still are

304   commonly used in science teaching (Tanner and Allen 2004) and have been traditionally used

305   when teaching evolutionary topics in ASU.

306   There were no significant differences in the performance levels among participants from

307   distinct demographic backgrounds. Despite differences in levels of representation across groups

308   (e.g. more 'Female' than 'Male' participants in both Genie 2016 and 2017), pre- and post-

309   recitation scores were similar. This suggests that Genie was as an effective teaching tool

310   regardless of participant's demographics. However, it should be mentioned that the variable

311   'First-generation' college students had statistical effect on participant's performance. While

312   participant's performance increased in all methods of instruction, 'First-generation' college

313   students showed slightly lower improvement than 'Non-first generation' college students.

314   Multiple studies have attempted to address the social class gap among undergraduate students

315   and explain why 'First-generation' college students, in occasion, perform more poorly than

316   'Non-first generation' college students (Grineski et al. 2018; Tibbetts et al. 2018). One finding

317   pertinent to our assessment is that 'First-generation' college students tend to underperform when

318   they know that their performance is going to be compared to that of other students in the class

319   (Jury et al. 2015). This might be an unintended consequence of the in-class methods used here,

320   which favored in-class discussion and student participation. However, while not possible to

321   address here, is possible that these results point towards the unique disadvantages and social-

322   related pressures that 'First-generation' college students face within ASU. These results should

323   be evaluated in more detail in future studies.

324   We were unable to control for previous classes that BIO345 students took. Although, all

325   students in BIO345 are required to have passed BIO340 (General Genetics), which typically

326   includes instruction in evolutionary genetics; BIO340 is taught my multiple instructors, who do

327   not teach evolutionary genetics equally. However, participants' performance in all classes

328   increased following instruction. Interestingly, mean scores showed that the increase in

329   performance between pre- and post-recitation was ~0.1 regardless of the teaching method used.

330   The main distinction were the pre-recitation scores, with some classes initially performing better

331   than others. Taken together, these results are indicative that both teacher-centered and Genie-

332   based teaching strategies led to a comparable improvement in participant's scores, regardless of

333   the initial performance level of the class. Thus, it is possible to conclude that Genie can perform

334   as efficiently as traditionally teacher-centered instruction. In addition, when participant groups

335   were divided based on the initial performance within their class (pre-recitation performance

336   quantiles) both instruction methods lead to an increase in post-recitation scores. However, some

337   quantiles saw more improvement than others. The highest improvement was observed in

338   participants from the second pre-recitation score quantile in Genie 2016 and the third pre-

339   recitation score quantile in 2017. On the other hand, participants with the highest pre-recitation

340   scores (fourth quantile), showed the smallest improvement. The latter trend was more evident

341   between Genie 2017 and Non-Genie 2017, with fourth quantile participants taking Genie-based

342   classes having a slightly larger improvement than participants taking teacher-centered classes.

343   This suggest that participants with initial lower understanding of the material similarly benefited

344   from instruction with Genie as with other methods, while participants with initial higher

345   understanding of the material benefited a little more when Genie was used.

346   In addition, participants' performance was not affected by the instructor or the participant

347   populations within the group, except for the 'TA Pair1 7:30pm' group during Genie 2016. The

348   'TA Pair1 7:30pm' group was formed by a small number of honors students; therefore, it is

349   possible that this group simply performed better compared to the general class population in

350   Genie 2016. Previous studies have found that instructors' mastery of the content, as well as their

351   overall teaching style play a critical role in students' learning process (Alsharif and Qi 2014;

352   Maleki et al. 2017). Thus, our results are indicative that Genie performs similarly well even with

353   teachers using diverse teaching styles and having variable levels of expertise.

354   Overall, understanding of genetic drift key concepts and misconceptions improved

355   following instruction with all teaching strategies. Nonetheless, it should be noted that there were

356   small variations in how well participant's performed in individual questions when using

16

357   teaching-centered vs. Genie-based methods. Within 2017, teaching-centered methods performed

358   better in questions related to key concepts on genetic drift as well as in understanding the

359   difference between genetic drift and other non-adaptive or adaptive evolutionary mechanisms.

360   On the other hand, Genie-based teaching aid the participants' understanding on the importance of

361   genetic drift as a source for evolutionary change and its relationship with natural selection. These

362   results suggest that a combined teaching strategy using Genie alongside with traditional teacher-

363   centered methods might help participants in gaining a more rounded comprehension of genetic

364   drift concepts. In agreement with this conclusion, previous analyses have shown that a

365   combination of traditional teaching-centered methods, with student-centered methods, and active

366   learning strategies results in superior student performance (Dolan and Collins 2015; Shir et al.

367   2016; Wieman 2014).

368   In the case of evolution teaching, strategies that favor student's development of critical

369   thinking skills are especially useful. For instance, tools and methods that aid in creating and

370   testing hypotheses have been effective in improving student's understanding and acceptance of

371   evolutionary theory (Lark et al. 2018; Smith et al. 2016). Likewise, instruction using computer

372   simulation programs has proven to be valuable in facilitating student's recognition of the breadth

373   of evolutionary mechanisms that can act in a population (Kliman 2001). In that regard, flexible

374   software that can be used to illustrate how diverse evolutionary forces are intrinsically linked can

375   be particularly useful in teaching. For example, the award-winning software Avida-ED, has been

376   effectively used to introduce evolutionary ideas to freshmen students via hypothesis testing (Abi

377   Abdallah et al. 2020), understand the role of low-impact mutations in evolution (Nelson and

378   Sanford 2011), and to test the robustness of genetic drift in small populations (Labar and Adami

379   2017). Nonetheless, it should not be forgotten than different students can master the same topic

17

380   using different paths (Price et al. 2016), and that different classroom settings might be more

381   suitable for distinct teaching methods. Simply put, there is no 'fit all' teaching strategy that can

382   be universally implemented. Therefore, providing instructors with broad repertoire of teaching

383   tools can aid them in finding those that better work for the topic being instructed, the specific

384   class needs, and the instructor style. In this regard, we expect Genie can be added to the

385   repertoire of higher education tools to be used for teaching genetic drift and other non-adaptive

386   evolution concepts.

387

388   **5. Conclusion**

389   The present study shows that Genie can be successfully used for teaching undergraduate

390   students concepts related with genetic drift and non-adaptive evolution. Genie performed

391   comparably to traditional teacher-centered methods across all evaluated groups. Moreover,

392   Genie-based and teacher-centered approaches led to participants understanding distinct key

393   concepts and misconceptions of genetic drift. This indicates that Genie can be effectively used

394   alongside other teaching strategies to provide a rounded view of non-adaptive evolution. In a

395   related note, Genie provides a mean for participants to develop and test their own hypotheses,

396   which can be useful in practicing critical thinking skills.

397

398 **Tables**

399 **Table 1. Demographic breakdowns of participants in each year and section shows variable**

400 **representation of different groups**. The breakdown of participants in each year of the class

401 who participated in the assessment, including those who self-identified as people of color

402 ('POC') or 'white', 'female' or 'male', and 'first-generation' college students or not.

Categories	Genie 2016	Genie 2017	Non-Genie 2017
POC	168	136	120
White	238	144	112
Female	230	190	140
Male	176	90	92
Not first-generation college	280	234	196
First-generation college	126	46	36

403

404

405

406

407

408

409

410

411

412

413

414

415

19

416 **Table 2. Most demographic predictors did not affect performance on the evaluation of**

417 **genetic drift knowledge.** The general linear regression of pre- and post-recitation scores with

418 demographic predictors, the degrees of freedom (Df), and the summary statistics.

Query	Predictor	Df	Sum sq	Mean Sq	F-value	Pr(>F)
Are pre-recitation scores dependent on demographic variables in 2016?	First generation college	1	0.106	0.106	4.142	0.043±
	Ethnicity	6	0.160	0.027	1.046	0.397
	Gender	1	0.026	0.026	1.004	0.318
	First generation college: Ethnicity	5	0.334	0.067	2.613	0.026±
	First generation college: Gender	1	0.006	0.006	0.242	0.623
	Ethnicity: Gender	4	0.140	0.035	1.373	0.245
	First generation college: Ethnicity: Gender	3	0.070	0.023	0.912	0.436
	Residuals	181	4.624	0.026		
Are post-recitation scores dependent on demographic variables in 2016?	First generation college	1	0.232	0.232	7.955	0.005±
	Ethnicity	6	0.065	0.011	0.372	0.896
	Gender	1	0.070	0.070	2.388	0.124
	First generation college: Ethnicity	5	0.430	0.086	2.949	0.014±
	First generation college: Gender	1	0.023	0.023	0.800	0.372
	Ethnicity: Gender	4	0.135	0.034	1.156	0.332
	First generation college: Ethnicity: Gender	3	0.054	0.018	0.616	0.606
	Residuals	181	5.281	0.029		
Are pre-recitation scores dependent on demographic variables in 2017?	First generation college	1	0.014	0.014	0.466	0.496
	Ethnicity	7	0.478	0.068	2.286	0.029±
	Gender	1	0.001	0.001	0.037	0.848
	Genie used	1	0.183	0.183	6.131	0.014±
	First generation college: Ethnicity	5	0.084	0.017	0.563	0.728
	First generation college: Gender	1	0.012	0.012	0.397	0.529
	Ethnicity: Gender	6	0.380	0.063	2.120	0.052
	First generation college: Genie used	1	0.122	0.122	4.096	0.044±
	Ethnicity: Genie used	5	0.157	0.031	1.053	0.388
	Gender: Genie used	1	0.284	0.284	9.495	0.002±
	First generation college: Ethnicity: Gender	3	0.148	0.049	1.653	0.178
	First generation college: Ethnicity: Genie used	3	0.165	0.055	1.846	0.140
	Ethnicity: Gender: Genie used	3	0.104	0.035	1.155	0.328
	Residuals	217	6.484	0.030		
Are post-recitation scores dependent on demographic variables in 2017?	First generation college	1	0.006	0.006	0.26	0.610
	Ethnicity	7	0.141	0.020	0.895	0.511
	Gender	1	0.0002	0.0002	0.010	0.921
	Genie used	1	0.053	0.053	2.350	0.127
	First generation college: Ethnicity	5	0.046	0.009	0.409	0.842
	First generation college: Gender	1	0.011	0.011	0.485	0.487
	Ethnicity: Gender	6	0.182	0.030	1.346	0.238
	First generation college: Genie used	1	0.017	0.017	0.757	0.385
	Ethnicity: Genie used	5	0.233	0.047	2.062	0.071
	Gender: Genie used	1	0.036	0.036	1.580	0.210
	First generation college: Ethnicity: Gender	3	0.033	0.011	0.486	0.692
	First generation college: Ethnicity: Genie used	3	0.129	0.043	1.902	0.130
	Ethnicity: Gender: Genie used	3	0.031	0.010	0.454	0.715
	Residuals	217	4.90	0.023		

419 ± Significant p-values

420

421 **Table 3. Post-recitation scores were mostly influenced by pre-recitation scores, but not**

422 **class section.** The Beta regression tests of pre- and post-recitation scores for specific queries,

423 including the predictors, estimates, standard errors (SE), z-scores, and p-values.

Query	Predictor	Estimate	SE	z-score	p-value
Are the pre-recitation scores different across classes (Genie 2016, Genie 2017, and Non-Genie 2017)?	Genie 2017	0.961	0.104	9.241	$< 2 \times 10^{-16\pm}$
	Non-Genie 2017	0.521	0.102	5.081	$3.75 \times 10^{-07\pm}$
Are the post-recitation scores dependent on the pre-recitation scores and/or class (Genie 2016, Genie 2017, and Non-Genie 2017)?	Pre-recitation scores	3.453	0.237	14.55	$< 2 \times 10^{-16\pm}$
	Genie 2017	0.509	0.103	4.948	$7.51 \times 10^{-07\pm}$
	Non-Genie 2017	0.281	0.099	2.847	$0.004\pm$
Are the post-recitation scores of 2016 dependent on the pre-recitation scores and/or the class section?	Pre-recitation score	3.198	0.391	8.17	$3.08 \times 10^{-16\pm}$
	TA Pair1 4:30pm	-0.384	0.217	-1.764	0.078
	TA Pair1 6:00pm	-0.164	0.233	-0.702	0.483
	TA Pair1 7:30pm	-0.68	0.285	-2.387	$0.017\pm$
	TA Pair2 3:00pm	0.309	0.226	1.366	0.172
	TA Pair2 4:30pm	-0.096	0.205	-0.468	0.64
	TA Pair2 6:00pm	-0.336	0.234	-1.433	0.152
	TA Pair2 7:30pm	0.052	0.242	0.213	0.831
Are the post-recitation scores of 2017 dependent on the pre-recitation scores and/or the class section when using Genie?	Pre-recitation score	3.52	0.452	7.797	$6.64 \times 10^{-15\pm}$
	TA Pair1 7:00pm	-0.209	0.21	-0.996	0.319
	TA Pair2 1:30pm	-0.032	0.209	-0.152	0.879
	TA Pair2 7:00pm	0.006	0.223	0.026	0.98
Are the post-recitation scores of 2017 dependent on the pre-recitation scores and/or the class section when not using Genie?	Pre-recitation score	3.583	0.37	9.695	$< 2 \times 10^{-16\pm}$
	TA Pair1 4:30pm	-0.023	0.194	-0.116	0.907
	TA Pair2 3:00pm	0.33	0.188	1.75	0.08
	TA Pair2 4:30pm	-0.126	0.19	-0.659	0.509

424 $\pm$ Significant p-values

425

426

427

428

429

430

431

432

433   **Table 4. Effect of Genie on learning outcomes.** The size-effect analysis of recitation scores per

434   year.

Query	Cohen's d (Lower 95% CI, Upper 95% CI)
Are the scores different before and after instruction in 2016 (Genie)?	0.608 (0.408, 0.807)
Are the scores different before and after instruction in 2017 (Genie)?	0.632 (0.410, 0.855)
Are the scores different before and after instruction in 2017 (No-Genie)?	0.658 (0.430, 0.886)
Are the pre-recitation scores different between 2017's Genie and Non-Genie classes?	0.272 (0.051, 0.493)
Are the post-recitation scores different between 2017's Genie and Non-Genie classes?	0.242 (0.021, 0.462)
Is the difference in pre- and post-recitation scores different between 2017's Genie and Non-Genie classes?	-0.087 (-0.307, 0.133)

435

436

437

438

439

440

441

442

443

444

445

446

447

448 **Table 5. Individual participants in all classes showed higher post-recitation scores**

449 **compared to their pre-recitation scores.** Paired Student's t-test for individual participants in

450 pre- and post-recitation scores per class.

Query	t -value (p-value)	Mean of the difference	Lower 95% CI, Upper 95% CI
Are paired post- and pre-recitation scores different in Genie 2016?	$-9.747 \ (2.2\text{x}10^{-16})^{\pm}$	-0.104	-0.125, -0.083
Are paired post- and pre-recitation scores different in Genie 2017?	$-8.966 \ (6.913\text{x}10^{-16})^{\pm}$	-0.102	-0.124, -0.079
Are paired post- and pre-recitation scores different in Non-Genie 2017?	$-9.816 \ (2.2\text{x}10^{-16})^{\pm}$	-0.115	-0.138, -0.092

451 $^{\pm}$ Significant p-values

452

453

454

455

456

457

458

459

460

461

462

463

464

Table 6. Participants from different pre-recitation scores quantiles showed improvement in their post-recitation scores. Student's t-test for pre- to post-recitation scores depending on pre-recitation performance quantile.

Query	Quantile (pre-recitation scores range)	t -value (p-value)	Lower 95% CI, Upper 95% CI	Mean pre-score	Mean post-score
Do different groups of participants have differences performances in Genie 2016?	1st Quantile (0-0.25)	Not enough observations			
	2nd Quantile (0.26-0.5)	-7.855 (4.294×10^{-11}) ±	-0.230, -0.137	0.448	0.631
	3rd Quantile (0.51-0.75)	-4.897 (2.968×10^{-06}) ±	-0.120, -0.051	0.638	0.724
	4th Quantile (0.76-1)	-3.630 (0.0004) ±	-0.082, -0.024	0.859	0.913
Do different groups of participants have differences performances in Genie 2017?	1st Quantile (0-0.25)	Not enough observations			
	2nd Quantile (0.26-0.5)	-5.186 (0.0014) ±	-0.341, -0.142	0.448	0.689
	3rd Quantile (0.51-0.75)	-7.118 (1.191×10^{-09}) ±	-0.222, -0.125	0.636	0.810
	4th Quantile (0.76-1)	-3.835 (0.0002) ±	-0.069, -0.022	0.903	0.948
Do different groups of participants have differences performances in Non-Genie 2017?	1st Quantile (0-0.25)	Not enough observations			
	2nd Quantile (0.26-0.5)	-6.128 (1.203×10^{-06}) ±	-0.341, -0.170	0.421	0.676
	3rd Quantile (0.51-0.75)	-7.714 (4.412×10^{-11}) ±	-0.214, -0.126	0.642	0.812
	4th Quantile (0.76-1)	-2.337 (0.021) ±	-0.059, -0.005	0.890	0.922

± Significant p-values

Table 7. For most individual questions, participant post-recitation performance improved across classes following instruction either with or without Genie. McNemar's test performed on (in)correct to (in)correct pre- and post-recitation answers per question. McNemar statistic and p-value are provided. We show this for (a) Genie 2016, (b) Genie 2017, and (c) Non-Genie 2017.

Evaluation		Question	Genie 2016					
			Incorrect to incorrect switches	Incorrect to correct switches	Correct to incorrect switches	Correct to correct switches	McNemar stat.	p-value
Key concepts		Q1	8	22	14	159	1.778	0.182
		Q3	10	53	7	133	35.267	$2.88 \times 10^{-09\pm}$
		Q15	9	26	15	153	2.951	0.086
		Q4	28	44	13	118	16.86	$4.02 \times 10^{-05\pm}$
		Q10	21	28	18	136	2.174	0.14
		Q13	18	86	9	90	62.411	$2.79 \times 10^{-15\pm}$
		Q16	58	28	43	74	3.169	0.075
Misconceptions*	1	Q7	10	32	19	142	3.314	0.069
	2	Q5	38	51	15	99	19.636	$9.37 \times 10^{-06\pm}$
		Q6	69	53	19	62	16.056	$6.15 \times 10^{-05\pm}$
		Q8	71	46	22	64	8.471	$0.004^{\pm}$
	3	Q2	39	41	27	96	2.882	0.09
	4	Q9	8	28	21	146	1	0.317
		Q12	36	60	22	85	17.610	$2.71 \times 10^{-05\pm}$
		Q17	49	58	12	84	30.229	$3.84 \times 10^{-08\pm}$
		Q20	22	40	18	123	8.345	$0.004^{\pm}$
	5	Q14	61	46	32	64	2.513	0.113
		Q19	28	36	39	100	0.12	0.729
		Q22	45	30	30	98	0	1
	6	Q11	17	50	17	119	16.254	$5.54 \times 10^{-05\pm}$
		Q18	12	29	8	154	11.919	$5.56 \times 10^{-04\pm}$
		Q21	20	24	28	131	0.308	0.579

$\pm$ Significant p-values
* Misconceptions (Price et al. 2014)
1. Genetic drift is unpredictable because it has a random component.
2. Genetic drift is natural selection/adaptation/acclimation to the environment that may result from a need to survive.
3. Genetic drift is not evolution because it does not lead to directional change that increases fitness.
4. Natural selection is always the most powerful mechanism of evolution, and it is the primary agent of evolutionary change.
5. Genetic drift is random mutation.
6. Genetic drift is gene flow or migration.

486   Table 7. For most individual questions, participant post-recitation performance improved

487   across classes following instruction either with or without Genie (cont.).

Evaluation		Genie 2017						
		Question	Incorrect to incorrect switches	Incorrect to correct switches	Correct to incorrect switches	Correct to correct switches	McNemar stat.	p-value
Key concepts		Q1	0	6	3	155	1	0.317
		Q3	1	13	3	147	6.25	$0.012^{\pm}$
		Q15	5	17	8	134	3.24	0.072
		Q4	7	17	12	128	0.862	0.353
		Q10	9	22	6	127	9.143	$0.002^{\pm}$
		Q13	5	32	5	122	19.703	$9.05 \times 10^{-06\pm}$
		Q16	31	29	17	87	3.13	0.077
Misconceptions*	1	Q7	6	23	11	124	4.235	$0.04^{\pm}$
	2	Q5	7	31	6	120	16.892	$3.96 \times 10^{-05\pm}$
		Q6	23	47	5	89	33.923	$5.73 \times 10^{-09\pm}$
		Q8	28	35	8	93	16.953	$3.83 \times 10^{-05\pm}$
	3	Q2	28	23	12	101	3.457	0.063
	4	Q9	7	19	5	133	8.167	$0.004^{\pm}$
		Q12	21	31	6	106	16.892	$3.96 \times 10^{-05\pm}$
		Q17	28	44	7	85	26.843	$2.21 \times 10^{-07\pm}$
		Q20	8	23	4	129	13.37	$2.56 \times 10^{-04\pm}$
	5	Q14	17	30	8	109	12.737	$3.59 \times 10^{-04\pm}$
		Q19	9	15	14	126	0.034	0.853
		Q22	10	23	9	122	6.125	$0.013^{\pm}$
	6	Q11	4	23	7	130	8.533	$0.003^{\pm}$
		Q18	2	14	4	144	5.556	$0.018^{\pm}$
		Q21	5	13	2	144	8.067	$0.005^{\pm}$

$^{\pm}$ Significant p-values

* Misconceptions (Price et al. 2014)

1. Genetic drift is unpredictable because it has a random component.
2. Genetic drift is natural selection/adaptation/acclimation to the environment that may result from a need to survive.
3. Genetic drift is not evolution because it does not lead to directional change that increases fitness.
4. Natural selection is always the most powerful mechanism of evolution, and it is the primary agent of evolutionary change.
5. Genetic drift is random mutation.
6. Genetic drift is gene flow or migration.

488

489

490

491

492

493

Table 7. For most individual questions, participant post-recitation performance improved across classes following instruction either with or without Genie (cont.).

Evaluation		Question	Non-Genie 2017					
			Incorrect to incorrect switches	Incorrect to correct switches	Correct to incorrect switches	Correct to correct switches	McNemar stat.	p-value
Key concepts		Q1	0	13	5	139	3.556	0.059
		Q3	4	16	8	129	2.667	0.102
		Q15	2	18	5	132	7.348	0.007$^\pm$
		Q4	13	14	11	119	0.36	0.549
		Q10	6	18	4	129	8.909	0.003$^\pm$
		Q13	8	39	3	107	30.857	$2.78\times10^{-08\pm}$
		Q16	27	30	27	73	0.158	0.691
Misconceptions*	1	Q7	11	23	10	113	5.121	0.024$^\pm$
	2	Q5	16	39	4	98	28.488	$9.43\times10^{-08\pm}$
		Q6	24	60	5	68	46.538	$8.98\times10^{-12\pm}$
		Q8	36	44	8	69	24.923	$5.97\times10^{-07\pm}$
	3	Q2	43	20	20	74	0	1
	4	Q9	10	20	6	121	7.538	0.006$^\pm$
		Q12	32	34	11	80	11.756	$6.07\times10^{-04\pm}$
		Q17	34	33	4	86	22.730	$1.86\times10^{-06\pm}$
		Q20	9	15	4	129	6.368	0.012$^\pm$
	5	Q14	21	32	12	92	9.091	0.003$^\pm$
		Q19	16	26	11	104	6.081	0.014$^\pm$
		Q22	16	23	6	112	9.966	0.002$^\pm$
	6	Q11	6	32	5	114	19.703	$9.05\times10^{-06\pm}$
		Q18	1	16	6	134	4.545	0.033$^\pm$
		Q21	10	18	12	117	1.200	0.273

$^\pm$ Significant p-values
* Misconceptions (Price et al. 2014)
1. Genetic drift is unpredictable because it has a random component.
2. Genetic drift is natural selection/adaptation/acclimation to the environment that may result from a need to survive.
3. Genetic drift is not evolution because it does not lead to directional change that increases fitness.
4. Natural selection is always the most powerful mechanism of evolution, and it is the primary agent of evolutionary change.
5. Genetic drift is random mutation.
6. Genetic drift is gene flow or migration.

496

497

498

499

500

501

502 **Table 8. Comparison of performance between Genie 2017 and Non-Genie 2017, controlling**

503 **by question.** Fisher's exact test testing the association between switches from 'Incorrect to

504 Correct' and 'Correct to Incorrect' answers per question and by method of instruction (Genie

505 2017 and Non-Genie 2017).

Evaluation		Question	Genie 2017		Non-Genie 2017		OR.est	p-value
			Incorrect to correct switches	Correct to correct switches	Incorrect to correct switches	Correct to correct switches		
Key concepts		Q1	6	3	13	5	0.777	1.000
		Q3	13	3	16	8	2.126	0.473
		Q15	17	8	18	5	0.597	0.523
		Q4	17	12	14	11	1.111	1.000
		Q10	22	6	18	4	0.818	1.000
		Q13	32	5	39	3	0.497	0.463
		Q16	29	17	30	27	1.529	0.321
Misconceptions*	1	Q7	23	11	23	10	0.910	1.000
	2	Q5	31	6	39	4	0.534	0.501
		Q6	47	5	60	5	0.785	0.749
		Q8	35	8	44	8	0.797	0.785
	3	Q2	23	12	20	20	1.900	0.242
	4	Q9	19	5	20	6	1.137	1.000
		Q12	31	6	34	11	1.661	0.421
		Q17	44	7	33	4	0.764	0.755
		Q20	23	4	15	4	1.519	0.700
	5	Q14	30	8	32	12	1.400	0.610
		Q19	15	14	26	11	0.459	0.136
		Q22	23	9	23	6	0.671	0.562
	6	Q11	23	7	32	5	0.519	0.349
		Q18	14	4	16	6	1.304	1.000
		Q21	13	2	18	12	4.207	0.094

$^{\pm}$ Significant p-values

* Misconceptions (Price et al. 2014)

1. Genetic drift is unpredictable because it has a random component.
2. Genetic drift is natural selection/adaptation/acclimation to the environment that may result from a need to survive.
3. Genetic drift is not evolution because it does not lead to directional change that increases fitness.
4. Natural selection is always the most powerful mechanism of evolution, and it is the primary agent of evolutionary change.
5. Genetic drift is random mutation.
6. Genetic drift is gene flow or migration.

506

507

508

509   **Figures**
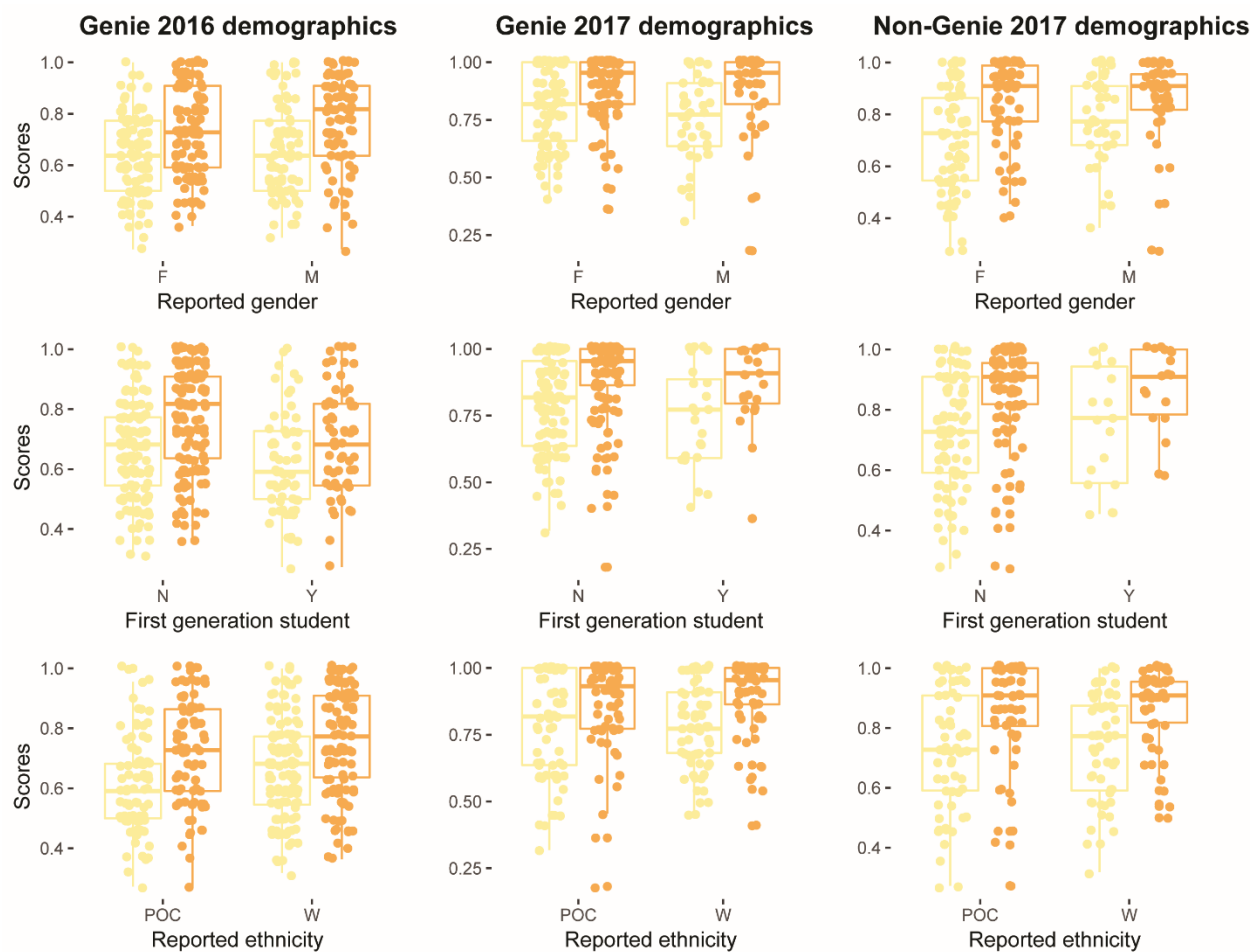
511   **Figure 1. Pre- and post-recitation performance was comparable in participants from**

512   **different demographic groups.** Pre-recitation (pale yellow) and post-recitation (dark yellow)

513   scores are plotted. Three demographic variables are plotted: 'Reported gender', 'First-
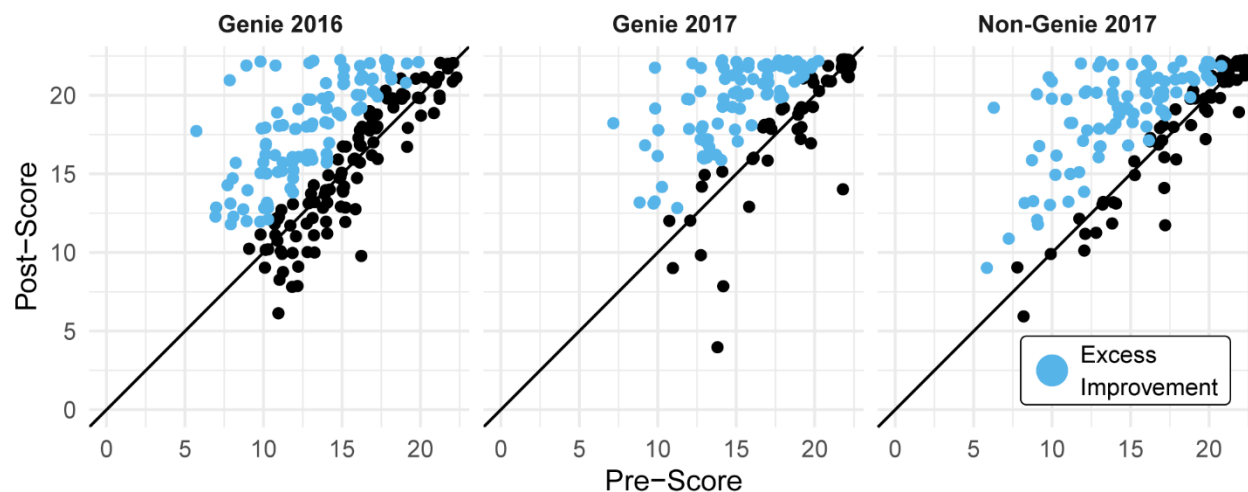
514   generation' college, and 'Reported ethnicity'.

515
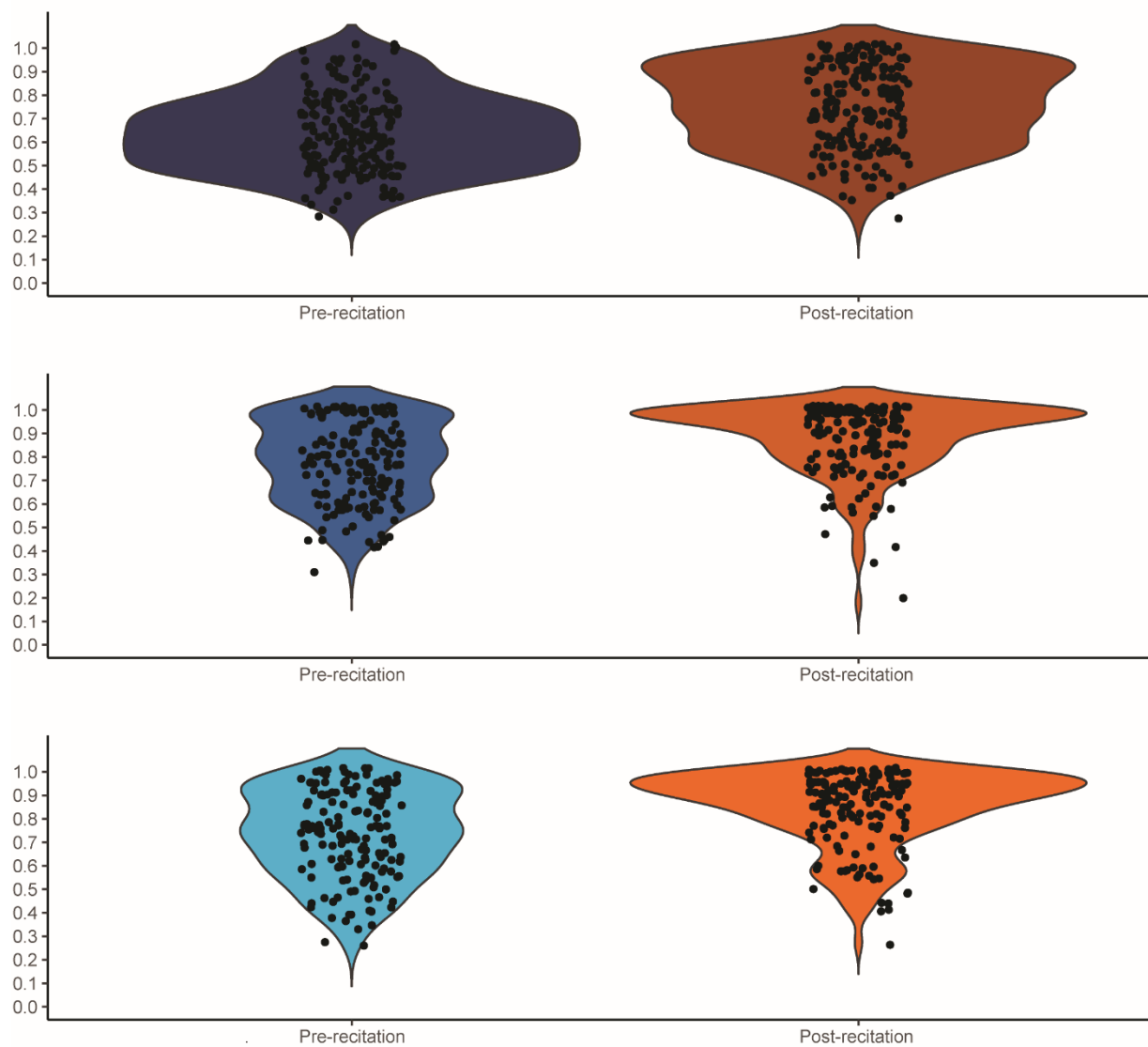
Figure 2. Post-recitation performance improved regardless of the instruction method. Pre-recitation (blue) and post-recitation (orange) scores are plotted for each class.

521

Figure 3. Students' test scores generally improved after instruction. Blue dots represent excess improvement in class performance. The presence of blue points in a graph indicates that there were more students whose post-test score was better than their pre-test score. The number of blue points indicates how many more students improved their scores than students whose scores decreased.

522
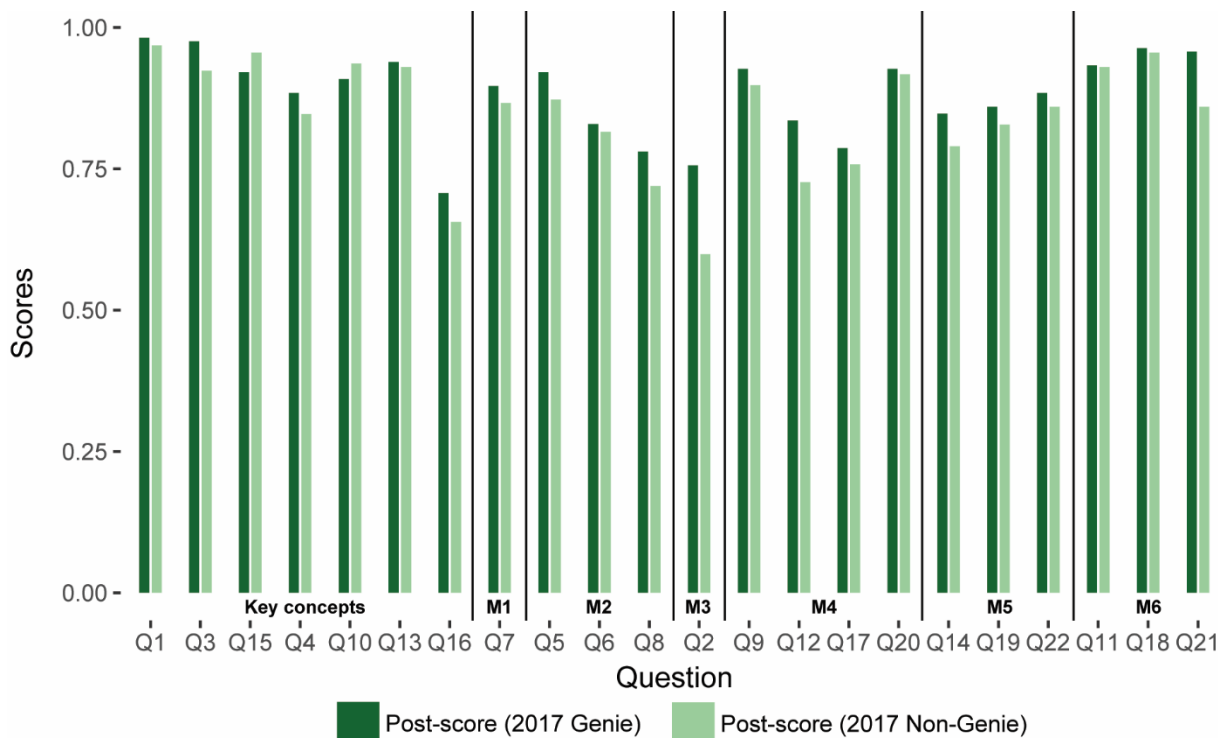
523

524

525

526

527

528

* Misconceptions (Price et al. 2014)
1. Genetic drift is unpredictable because it has a random component.
2. Genetic drift is natural selection/adaptation/acclimation to the environment that may result from a need to survive.
3. Genetic drift is not evolution because it does not lead to directional change that increases fitness.
4. Natural selection is always the most powerful mechanism of evolution, and it is the primary agent of evolutionary change.
5. Genetic drift is random mutation.
6. Genetic drift is gene flow or migration.

Figure 4. Post-recitation scores by question (Price et al. 2014) were generally higher in Genie 2017 compared to Non-Genie 2017. A bar plot comparing Genie 2017 (dark green) and Non-Genie 2017 (pale green) is shown. Questions have been grouped according to the classification provided by Price et al. (2014), with questions pertaining to Key concepts and misconceptions (M1-M6) separated by horizontal bars.
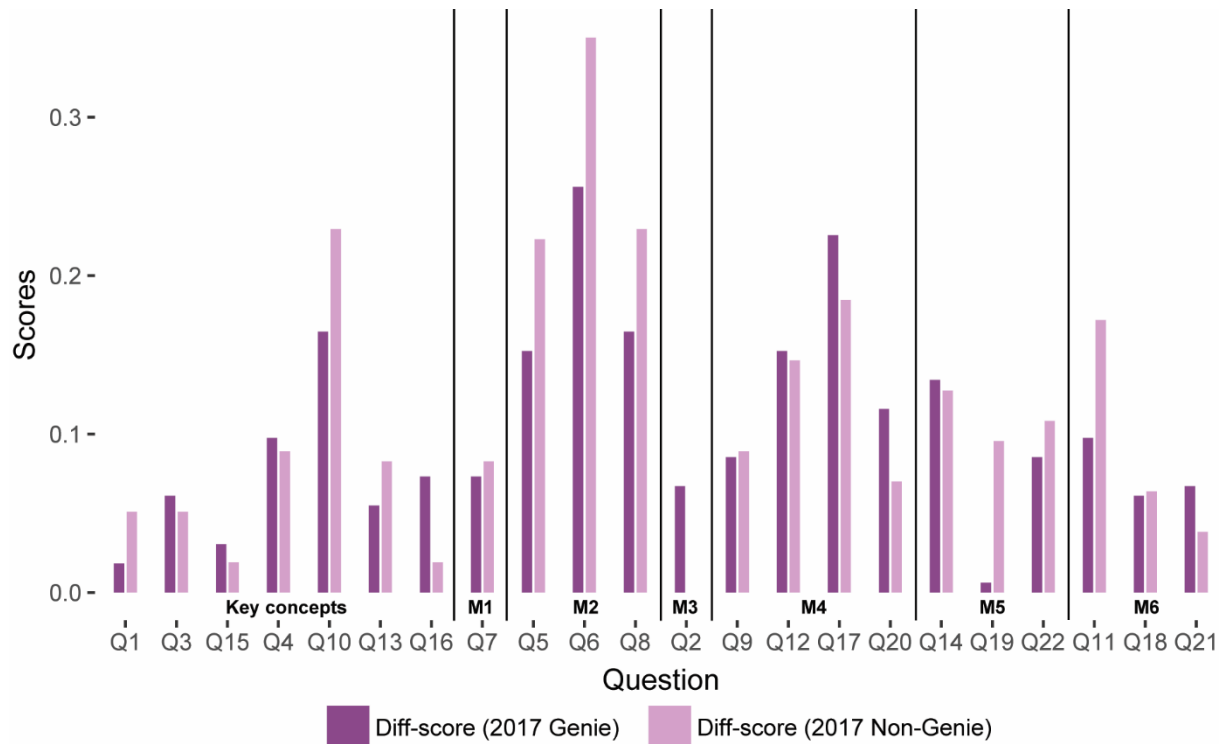
537

* Misconceptions (Price et al. 2014)
1. Genetic drift is unpredictable because it has a random component.
2. Genetic drift is natural selection/adaptation/acclimation to the environment that may result from a need to survive.
3. Genetic drift is not evolution because it does not lead to directional change that increases fitness.
4. Natural selection is always the most powerful mechanism of evolution, and it is the primary agent of evolutionary change.
5. Genetic drift is random mutation.
6. Genetic drift is gene flow or migration.

538

539 **Figure 5. The difference between pre- and post-recitation scores by question (Price et al.**

540 **2014) shows that some questions saw higher improvement in Non-Genie 2017 while other**

541 **showed higher improvement in Genie 2017.** A bar plot comparing Genie 2017 (dark purple)

542 and Non-Genie 2017 (pale purple) is shown. Questions have been grouped according to the

543 classification provided by Price et al. (2014), with questions pertaining to Key concepts and

544 misconceptions (M1-M6) separated by horizontal bars.

545

546

33

547 **Ethics approval and consent to participate**

548 The study was approved by IRB protocol: STUDY00003707.

549 **Consent for publication**

550 Not applicable.

551 **Availability of data and materials**

552 A previous version of this manuscript is available as preprint (https://doi.org/10.1101/268672).

553 All data and code used has been made available as supplementary materials. Genie is publicly

554 available at https://cartwrig.ht/apps/genie/

555 **Competing interests**

556 The authors declare to competing interests.

557 **Funding**

558 This study was supported by the National Science Foundation award DBI-1356548 to RAC and

559 the National Institute of General Medical Sciences of the National Institutes of Health under

560 Award Number R35GM124827 to MAW. The content is solely the responsibility of the authors

561 and does not necessarily represent the official views of the National Institutes of Health.

562 **Authors' contributions**

563 AIC wrote and edited the manuscript, performed the statistical analyses, and participated in in-

564 class instruction. BHR designed and wrote Genie. MSR revised the manuscript, designed the

565 class study, and participated in in-class instruction. RAC designed and wrote Genie, edited the

566   manuscript, and the performed statistical analyses. MAW edited the manuscript, designed the

567   class study, and participated in in-class instruction prior to recitations using Genie.

568   **Acknowledgements**

569   The authors would like to thank the students of Spring 2016 and Spring 2016, BIO345 course at

570   ASU for their participation on this study**.**

References

571

572 Abi Abdallah DS, Fonner CW, Lax NC, Babeji MR, Palé FA. Evaluating the Use of Avida-ED
573 Digital Organisms to Teach Evolution & Natural Selection. Am. Biol. Teach. 2020;82(2):114–9.

574 Alsharif NZ, Qi Y. A three-year study of the impact of instructor attitude, enthusiasm, and
575 teaching style on student learning in a medicinal chemistry course. Am. J. Pharm. Educ.
576 2014;78(7).

577 Alters BJ, Nelson CE. Perspective: Teaching evolution in higher education. Evolution (N. Y).
578 2002;56(10):1891–901.

579 Andrews TM, Price RM, Mead LS, McElhinny TL, Thanukos A, Perez KE, et al. Biology
580 undergraduates' misconceptions about genetic drift. CBE Life Sci. Educ. 2012;11(3):258–9.

581 Beggrow EP, Nehm RH. Students' Mental Models of Evolutionary Causation: Natural Selection
582 and Genetic Drift. Evol. Educ. Outreach. 2012;5(3):429–44.

583 Cribari-Neto F, Zeileis A. Journal of Statistical SoftwareBeta Regression in R. J. Stat. Softw.
584 2010;34(2):1–24.

585 Delignette-Muller ML, Dutang C. fitdistrplus: An R package for fitting distributions. J. Stat.
586 Softw. 2015;64(4):1–34.

587 Dolan EL, Collins JP. We must teach more effectively: Here are four ways to get started. Mol.
588 Biol. Cell. 2015;26(12):2151–5.

589 George M. Slavich, Zimbardo PG. Transformational Teaching: Theoretical Underpinnings, Basic
590 Principles, and Core Methods. Educ Psychol Rev. 2012;24(4):569–608.

591 Grineski S, Daniels H, Collins T, Morales DX, Frederick A, Garcia M. The conundrum of social
592 class: Disparities in publishing among STEM students in undergraduate research programs at a
593 Hispanic majority institution. Sci Educ. 2018;102(2):283–303.

594 Jury M, Smeding A, Darnon C. First-generation students underperformance at university: the
595 impact of the function of selection. Front. Psychol. 2015;6(May):1–11.

596 Kalinowski ST, Leonard MJ, Andrews TM, Litt AR. Six classroom exercises to teach natural
597 selection to undergraduate biology students. CBE Life Sci. Educ. 2013;12(3):483–93.

598 Kliman BR, Cedar PD, College C, Education N. The EvolGenius Population Genetics Computer
599 Simulation : How it Works. Nat. Educ. 2008;1(3):1–9.

600 Kliman RM. A Project-Based Approach to Teaching Complex Population Genetics to
601 Undergraduates . Teach. Popul. Genet. 2001;27(4):13–20.

602 Labar T, Adami C. Evolution of drift robustness in small populations. Nat. Commun. [Internet].
603 Springer US; 2017;8(1):1–12. Available from: http://dx.doi.org/10.1038/s41467-017-01003-7

604 Lark A, Richmond G, Mead LS, Smith JJ, Pennock RT. Exploring the Relationship between
605 Experiences with Digital Evolution and Students' Scientific Understanding and Acceptance of
606 Evolution. Am. Biol. Teach. 2018;80(2):74–86.

607 Lee TW, Grogan KE, Liepkalns JS. Making evolution stick: Using sticky notes to teach the

608   mechanisms of evolutionary change. Evol. Educ. Outreach [Internet]. Springer US; 2017;10(1).
609   Available from: https://doi.org/10.1186/s12052-017-0074-2

610   Maleki F, Talaei MH, Moghadam SRM, Shadigo S, Taghinejad H, Mirzaei A. Investigating the
611   influence of teachers' characteristics on the teacher-student relations from students' perspective
612   at ilam university of medical sciences. J. Clin. Diagnostic Res. 2017;11(6):JC04–8.

613   Nei M, Chakraborty R, Fuerst PA. Infinite allele model with varying mutation rate. Genetics.
614   1976;73(11):4164–8.

615   Nelson CE. Teaching evolution (and all of biology) more effectively: Strategies for engagement,
616   critical reasoning, and confronting misconceptions. Integr. Comp. Biol. 2008;48(2):213–25.

617   Nelson CW, Sanford JC. The effects of low-impact mutations in digital organisms. Theor. Biol.
618   Med. Model. [Internet]. BioMed Central Ltd; 2011;8(1):9. Available from:
619   http://www.tbiomed.com/content/8/1/9

620   Perez KE, Hiatt A, Davis GK, Trujillo C, French DP, Terry M, et al. The evodevoci: A concept
621   inventory for gauging students' understanding of evolutionary developmental biology. CBE Life
622   Sci. Educ. 2013;12(4):665–75.

623   Price RM, Andrews TC, McElhinny TL, Mead LS, Abraham JK, Thanukos A, et al. The genetic
624   drift inventory: A tool formeasuring what advanced undergraduates havemastered about genetic
625   drift. CBE Life Sci. Educ. 2014;13(1):65–75.

626   Price RM, Pope DS, Abraham JK, Maruca S, Meir E. Observing populations and testing
627   predictions about genetic drift in a computer simulation improves college students' conceptual
628   understanding. Evol. Educ. Outreach. Springer US; 2016;9(1).

629   Revell LJ. learnPopGen: An R package for population genetic simulation and numerical analysis.
630   Ecol. Evol. 2019;9(14):7896–902.

631   Shir N, Bidabadi A, Isfahani AN, Rouhollahi A, Khalili R, Bidabadi NS. Effective teaching
632   method in higher education Effective teaching methods in higher education: requirements and
633   barriers. J Adv Med Educ Prof. Oct. 2016;4(4):170–8.

634   Shulman LS. Knowledge and Teaching: Foundations of the new reform. Harv. Educ. Rev.
635   1987;57(1):1–21.

636   Smith JJ, Johnson WR, Lark AM, Mead LS, Wiser MJ, Pennock RT. An Avida-ED digital
637   evolution curriculum for undergraduate biology. Evol. Educ. Outreach. Springer US; 2016;9(1).

638   Staub NL. Teaching Evolutionary Mechanisms: Genetic Drift and M&M's®. Bioscience.
639   2002;52(4):373.

640   Tanner K, Allen D. Approaches to biology teaching and learning: Learning styles and the
641   problem of instructional selection - Engaging all students in science courses. Cell Biol. Educ.
642   2004;3(4):197–201.

643   Tanner K, Allen D. Approaches to biology teaching and learning: Understanding the wrong
644   answers-teaching toward conceptual change. Cell Biol. Educ. 2005;4(SUMMER):112–7.

645   Tibbetts Y, Priniski SJ, Hecht CA, Borman GD, Harackiewicz JM. Different institutions and

646 different values: Exploring first-generation student fit at 2-year colleges. Front. Psychol.
647 2018;9(APR):1–19.

648 Wieman CE. Large-scale comparison of science teaching methods sends clear message. Proc.
649 Natl. Acad. Sci. U. S. A. 2014;111(23):8319–20.

650 Ziadie MA, Andrews TC. Moving evolution education forward: A systematic analysis of
651 literature to identify gaps in collective knowledge for teaching. CBE Life Sci. Educ. 2018;17(1).

652