

1

2

3

4 **The adaptive function of sexual reproduction: resampling the**
5 **genotype pool.**

6

7

8

Donal A. Hickey

9

Department of Biology, Concordia University

10

7141 Sherbrooke West,

11

Montreal, Quebec H4B 1R6

12

Canada

13

14

and

15

16

G. Brian Golding

17

Department of Biology, McMaster University

18

1280 Main St. West,

19

Hamilton, Ontario L8S 4K1

20

Canada

21

22

23

Corresponding author: Donal A. Hickey, Department of Biology,
24 Concordia University, 7141 Sherbrooke West,
25 Montreal, Quebec H4B 1R6, Canada.

26

E-mail: donal.hickey@concordia.ca

27

28

29

30

31 **Abstract.**

32

33 Recombination allows a finite population to resample the genotype pool, i.e., the
34 universe of all possible genotypic combinations. This is important in populations that
35 contain abundant genetic variation because, in such populations, the number of potential
36 genotypes is much larger than the number of individuals in the population. Here, we show
37 how recombination, in combination with natural selection, enables an evolving sexual
38 population to replace existing genotypes with new, higher-fitness genotypic combinations.
39 In contrast to this, an asexual population is limited to selection among existing genotypes.
40 Since it has been shown that most eukaryotic species are genetically polymorphic, our
41 model can explain the ubiquity of sex among such species. The model also indicates that
42 classic population genetics theory is applicable to ecological studies of natural selection
43 acting on standing genetic variation.

44

45

46 Key words: recombination, natural selection, evolution of sex,
47 standing genetic variation.

48 **Introduction.**

49

50 The function of sexual reproduction has been the subject of much debate among
51 biologists (see Otto 2009, Meirmans and Strand 2010, Colegrave 2012, for recent reviews).
52 Here, we illustrate the effects of sexual reproduction in a finite population where many
53 genes are undergoing selection. An example of this situation would be where adaptation is
54 based on the standing genetic variation in a local population (Barrett and Schluter 2007,
55 Burke et al 2010, Kosheleva and Desai 2017, McDonald et al 2016, Teotonio et al 2009).
56 One of the earliest explanations for the function of sex was that it combines different
57 beneficial mutations in a single individual, resulting in fitter genotypic combinations
58 (Fisher 1930, Muller 1932). Those authors focused on new beneficial mutations rather than
59 on the standing genetic variation. Maynard Smith (1978) pointed out that, given a low
60 mutation rate, small populations might not contain multiple mutations at the same time. By
61 focusing on the abundant standing genetic variation rather than only new mutations, we
62 avoid Maynard Smith's concern. Of course, these new mutations will eventually contribute
63 to the standing genetic variation once they reach significant frequencies.

64 Biologists are well aware that Mendelian genetics in an outbreeding sexual
65 population is essentially a game of chance. In this study, we use a lottery analogy in order
66 to explain, in simple terms, the biological function of genetic recombination in a sexually
67 reproducing population. We point out that in the case of inheritance, the lottery is biased
68 by selectively-induced, heritable changes in the frequencies of individual genetic variants.
69 We argue that it is the interaction between random recombination and deterministic natural
70 selection that allows sexual populations to evolve much more efficiently than asexual
71 populations. This provides a consistent, short-term selective advantage for sexual
72 reproduction.

73 The genetic lottery is special in a number of ways, and by taking these special
74 features into account we can better understand its function. The first feature to be taken
75 into account is that the winning combination, or optimal genotype, does not change
76 drastically and unpredictably from one generation to the next. Environments do change,
77 but this change usually happens gradually over many generations. Thus, the current
78 environment is a reasonable estimator of environmental conditions in the following
79 generations. The second point to consider is that the number of possible genotypic

80 combinations can be many orders of magnitude larger than the total population size (see
81 discussion below). This means that the chance that even a single individual within the entire
82 population will have the optimal genotype is extremely remote. The third point is that both
83 genetic lotteries and many conventional lotteries can have partial winners even in the
84 absence of an overall jackpot winner. In both cases there is an advantage in having at least
85 some of the winning numbers. Finally, the crucial difference between a conventional lottery
86 and the genetic lottery is that inheritance provides an informational feedback each
87 generation regarding the “winning” genotypic combination. We explain this point in more
88 detail below.

89

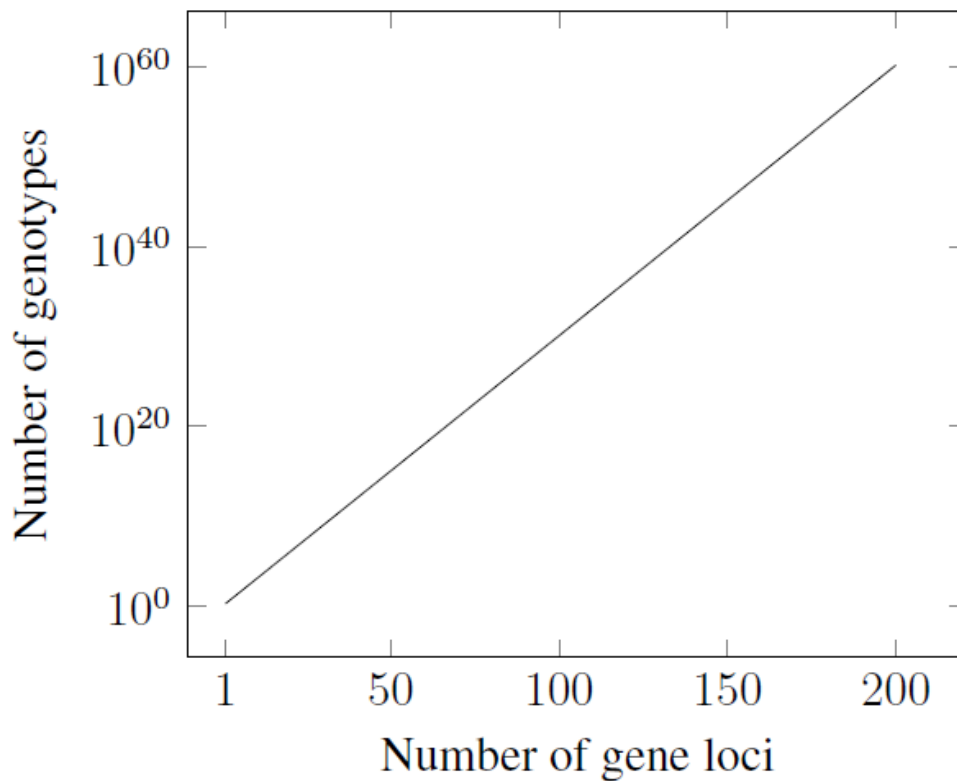
90 **The model.**

91 Students of genetics are familiar with the process of generating various genotypic
92 combinations of alleles at two, three or four variable loci. But we rarely stop to think how
93 this plays out at the genomic level. Eukaryotic genomes are estimated to contain
94 approximately 20,000 genetic loci and many of these loci are genetically polymorphic
95 (Lewontin and Hubby 1966). This discovery of abundant genetic variability in natural
96 populations means that the impact of recombination is much greater than what was
97 previously envisaged (Stearns 1985). If we take the overly conservative estimate that only
98 1% of the gene loci within the genome are genetically variable, we are still left with 200
99 genes. If we make a further conservative assumption that none of the genes has more than
100 two allelic variants, we can calculate that there are 2^{200} possible genotypic combinations.
101 This number is larger than 10^{60} which in turn is several orders of magnitude larger than the
102 estimated number of atoms in the solar system. Obviously, no biological population is large
103 enough to contain even a single individual representing each of these combinations. In fact,
104 real populations, because of their finite size, contain only a tiny fraction of all possible
105 genotypic combinations.

106 The relationship between the number of genetically variable genes and the total
107 number of possible genotypic combinations was first noted by East (1918). He pointed out
108 that ten variable genes could generate 1,024 different genotypes and that twenty variable
109 genes could generate 1,032,576 genotypic combinations. Here, we extend this calculation
110 up to 200 genes. The relationship between the number of variable genes and the

111 corresponding number of genotypes is shown in Figure 1. From the figure it can be seen
112 that once we reach more than thirty genetic loci, the number of possible genetic
113 combinations already exceeds most real population sizes, especially those of macroscopic
114 plants and animals. For example, 30 biallelic loci can generate more than one billion
115 different genotypes. The number of genotypes corresponding to 100 loci is approximately
116 10^{30} , which is larger than the number of atoms in the solar system; the number
117 corresponding to 200 loci is many orders of magnitude larger than this again.

118



119

120

121 **Figure 1.** The relationship between the number of genetically variable genes and the
122 number of possible genotypic combinations.

123 The graph shows the results for up to 200 biallelic loci. The Y axis is on a log scale.

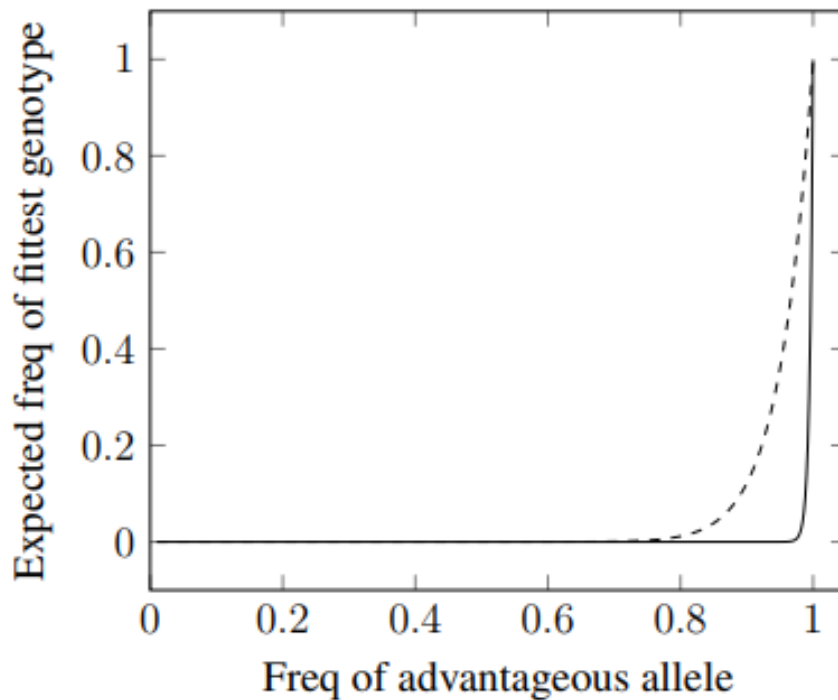
124

125

126 Using the lottery analogy, we can think of an individual genotype containing 200
127 variable genes as a “ticket” with a list of 200 numbers. In the simple case where we consider
128 only two variants per gene, we can represent the variants as a series of “zeros” and “ones”,
129 where zero represents the more common allelic variant and the number one represents a
130 rarer advantageous variant. Again, for simplification let’s assume that the initial frequency
131 of the advantageous variant at each genetic locus is 0.1 or ten percent. If the tickets are
132 chosen randomly there is a ten percent chance of observing a “1” at any one of the 200
133 positions. A lottery jackpot winner would have a “1” at all two hundred positions. But the
134 chance of finding such a ticket in the initial population is only 1 out of 10^{200} . In other words,
135 we would need to have a number of tickets that is many orders of magnitude larger than
136 the number of atoms in the observable universe in order to have a reasonable chance of
137 observing one such ticket. This means that, in practice, such a combination would not be
138 observed. In fact, all of the tickets would have relatively few ones and many zeros. For any
139 given ticket, we expected to see an average of 20 ones and 180 zeros; this expectation is
140 based on the initial frequencies of ones and zeros. But if the numbers are randomly assigned
141 based on these frequencies, not every ticket will have exactly 20 ones. The numbers will
142 be binomially distributed around the average of 20 and the majority of the values will fall
143 between 12 and 28. This means that there is not only an essentially zero chance of observing
144 a ticket with 200 ones (the jackpot winner), there is also virtually zero chance of observing
145 a ticket with even half that number of ones. So, at first glance, this genetic lottery seems
146 entirely hopeless.

147 But, as stated above, the genetic lottery is special in that there is a differential
148 survival of tickets depending on their genetic fitness, i.e., the number of “ones” on the
149 ticket. This means that, after selection, tickets containing a higher number of ones will be
150 more represented than those tickets containing less ones. This is the process of natural
151 selection. But natural selection alone can only result in a population containing tickets with
152 the highest number of ones that already exist in the initial population. In our example, this
153 would be about 30 out of the maximum of 200. But if the rounds of selection alternate with
154 rounds of sexual outbreeding and recombination, we observe a further effect. Specifically,
155 as the frequency of “ones” increases due to selection, recombination will randomly
156 recombine them to former a higher average number of “ones” per ticket. Thus, the power

157 of random recombination lies in its ability to adjust the number of “ones” per ticket to
158 match the current frequencies at individual positions. For example, when the frequency of
159 “ones” reaches 0.15, then the average number per ticket will be 30 rather than the initial
160 value of 20. The relationship between the rising allele frequencies at each position and the
161 frequency of the “jackpot combination” (that contains a “one” at all 200 positions) is shown
162 in Figure 2. As can be seen from the figure, it is only toward the end of the process that the
163 winning combination appears, where it then rises dramatically in frequency. This pattern
164 can be explained mathematically by saying that the frequency of the winning combination
165 is an extreme power function of the 200 individual frequencies.



166

167 **Figure 2.**

168 The relationship between the increase in the frequency of favorable alleles at
169 individual genetic loci (X axis) and the expected frequency of the optimal genotypic
170 combination.

171 As can be seen from the figure, the frequency of the optimal genotype stays close to zero
172 over most of the frequency range at the individual loci. This effect is most pronounced for
173 200 gene loci (solid line) and less so for 20 loci (dashed line).

174

175

176 In successive rounds of a conventional lottery, new tickets are printed using the
177 same expected frequencies of numbers as in the previous round. But in the genetic lottery
178 these frequencies are modified based on the genetic composition of the selected parents.
179 For example, early in the process the frequency of ones at individual positions may have
180 risen from 0.1 to 0.105 due to selection. In this case random recombination will generate
181 new “tickets” with an average of 21 “ones” per ticket, a slight increase from the original
182 value of 20. As the process continues, both the frequency at individual positions and the
183 total number of ones per ticket will increase concomitantly until, eventually, all tickets will
184 contain the winning combination. In a conventional lottery, such a continual adjustment of
185 the frequencies would be considered a blatant case of cheating. But it is precisely this
186 “trick” that nature uses to arrive at a genotypic combination that appears to be
187 mathematically unattainable at the outset. This process is similar to the multiplicative
188 weights update algorithm, MWUA, that has been used as an optimisation method in game
189 theory and machine learning (Arora et al 2012, Chastain et al 2014). Specifically, random
190 recombination updates the distribution of “tickets” each generation based on the post-
191 selection frequencies of the zeros and ones at each of the 200 positions.

192 If we restate the problem in more formal genetic terms, we can consider the
193 “tickets” as genotypes and the zeros and ones as allelic variants. At the start of this process,
194 the frequency of the favorable allelic variant at each locus is 0.1. Consequently, the
195 probability of finding one individual with the favorable allele at every locus in the first
196 generation is only 10^{-200} . Although the prospect of finding an individual with the optimal
197 genetic combination (i.e., a favorable allele at all 200 loci) is essentially zero at the start of
198 the process, inheritance provides a feedback loop between generations that continuously
199 improves the odds. All genotypes in the population will have a relatively low fitness
200 initially, but natural selection will tend to favor “the best of the bad lot”. These early rounds
201 of selection will result in incremental increases in the frequency of favourable alleles at
202 each of the 200 loci. Recombination will then generate new random combinations of these
203 alleles. And since the individual allele frequencies have increased slightly, the expected
204 frequency of genotypes with higher numbers of favorable alleles per chromosome will also

205 increase automatically. In this way, despite its random nature, recombination facilitates the
206 production of genotypic combinations that would be hopelessly improbable in its absence.

207

208

209

210

211 **Discussion**

212 The exponentially increasing number of genotypic combinations as the number of
213 genetically variable loci increases was first pointed out by East (1918). Subsequently, Iles
214 et al (2003) used this relationship to point out that multi-locus genotypes could be very rare
215 even in very large finite populations. Our focus here is on the situation where the number
216 of loci is of the order of 100 or greater. As illustrated in Figure 1, once the number of loci
217 under selection exceeds twenty or thirty, the number of possible genotypic combinations
218 becomes much greater than most biological population sizes. This means that natural
219 selection cannot act on the full range of genotypes. Rather it is limited to a narrow slice of
220 such genotypes in any given generation. It is tempting to think of natural selection as
221 always favoring the fittest possible genotype. But when many loci are under simultaneous
222 natural selection, the theoretically fittest genotype is often absent from the population.
223 Thus, selection cannot act directly on this genotype, although we tend to think of the
224 process in such teleological terms. In fact, the production of the optimal genotypic
225 combination is simply the automatic consequence of recurrent rounds of natural selection
226 and recombination. Most of the selection occurs among genotypes of varying degrees of
227 intermediate fitness. And the effects of recombination are purely random. Yet the
228 interaction between selection and random recombination provides an efficient mechanism
229 for the eventual production of the optimal genotype. The power of recombination is that it
230 allows a finite population to effectively resample genotypes from a virtual infinite
231 population with the current allele frequencies (Ewens 1972, Edhan et al 2017, Hickey and
232 Golding 2018). It is in this way that it produces higher fitness genotypes that were not
233 previously present in the population.

234 Our argument is based on the assumption that the frequency of any given genotypic
235 combination after a round of recombination can be estimated by multiplying the

236 frequencies at the individual loci. This assumption requires that approximate linkage
237 equilibrium is maintained during the course of the selection. Nagylaki (1993) showed that
238 this is true when the recombination fraction is equal to, or greater than the selection
239 coefficient. For example, two adjacent genes that are one centimorgan apart would remain
240 in approximate linkage equilibrium if the selection coefficient were 0.01. Non-adjacent
241 genes would maintain linkage equilibrium in the presence of higher intensities of selection.
242 This explains the somewhat paradoxical situation that, whereas a given gene can remain in
243 linkage disequilibrium with its nearest neighbors for many generations, it becomes
244 randomized with regard to the majority of other genes rather quickly. Under such
245 conditions, and when the fitness interactions between loci are multiplicative, the
246 multiplicative weights update algorithm can be used (Arora et al 2012). Here, we point out
247 that this process provides an efficient mechanism for the production of higher fitness
248 genotypic combinations that were not previously present in the population.

249 Previous studies of recombination in finite populations have focused on the
250 adjustment of existing genotype frequencies (Otto and Barton 2001, Iles et al 2003, Otto
251 and Gerstein 2006). Here, our focus is not on the frequencies of existing genotypic
252 combinations, but rather on the production of new genotypes that did not previously exist
253 in the population. As fitter haplotypes increase in frequency, recombination can act to
254 produce still fitter haplotypes (Iles et al 2003). Recombination enables the population to
255 resample the universe of all possible genotypic combinations based on the current allele
256 frequencies (Ewens 1972, Hickey and Golding 2018). Thus, a finite sexual population is
257 not limited to selection among those genotypes that already exist within it. Previously,
258 adaptive evolution based on selection at many genetic loci was seen as potentially very
259 costly in terms of selective deaths (Haldane 1957). However, Haldane's calculations were
260 based on a model of selection among fixed genotypes. Once we include the possibility of
261 recurrent disassembly and reassembly of genotypes through random recombination, the
262 problem envisaged by Haldane disappears (Hickey and Golding 2019).

263 This model of selection on the standing genetic variation also points to a possible
264 solution for the problem of the cost of sex that was described by Maynard Smith (1978).
265 This reproductive cost comes from the fact that an asexual female that produced only
266 asexual daughters would be twice as efficient in passing along her genes to future

267 generations compared to a sexual female. This should provide an immediate twofold
268 advantage for a newly arising asexual clone. Such a prediction is indeed true in a non-
269 evolving, monomorphic population. But in a population that is actively evolving based on
270 multi-locus selection, other factors come into play. Most important, the average fitness of
271 the sexual population continues to increase each generation while the asexual clone is
272 genetically frozen. This means that the initial twofold relative advantage is continually
273 discounted as the fitness of the sexual population increases. Simulations have shown that
274 a new asexual clone initially increases in frequency as predicted by Maynard Smith (1978),
275 but this increase is followed by a subsequent decrease in frequency as the fitness of the
276 sexual population "catches up" and overtakes the asexual clone. The asexual clone is
277 eventually eliminated from the population (Hickey and Golding 2018). Thus, the advantage
278 of an asexual clone is short-lived in an actively evolving population.

279 Previous authors have also compared Mendelian inheritance to a lottery (Williams
280 1975, Maynard Smith 1978). But the model discussed by those authors was significantly
281 different from the one proposed here. For example, it was assumed that the "winning"
282 genotypic combination would change randomly from generation to generation, thus
283 eliminating the possibility of informational feedback between generations. Neither did they
284 consider the possibility that the number of possible genotypes could be many orders of
285 magnitude greater than the population size. It is worth noting that, according to our model,
286 the production of the optimal genotypic combination is the result of an iterative process
287 that occurs over many generations; it is not the result of a single "lucky toss" as envisaged
288 by Williams (1975). This iterative process can explain why sexual reproduction needs to
289 occur continuously. Peck and Waxman (2000) wondered why intermittent sexual
290 reproduction is not more common in nature and a number of other authors have sought
291 explanations for obligate sex (Kleiman and Hadany 2015, Crouch 2017).

292 One of the earliest proposals for the advantage of sex was that it generates genetic
293 variants (Weissman 1887). While this is true, sexual reproduction does not usually increase
294 the level of variation relative to that in the previous generation. Rather, it replenishes the
295 genotypic variation that would have otherwise been eroded by selection and random
296 genetic drift. Even more important, it continually shifts the genotypic distribution towards
297 a greater average fitness, as we have shown above. Without sex and recombination, the

298 adaptive evolutionary process would be greatly impeded. Classic population genetics
299 studies (e.g. Haldane 1957) envisaged selection as performing an exhaustive search
300 among all possible genotypic combinations. As the number of combinations explodes in
301 size, however, such an exhaustive search becomes prohibitively difficult. But when we
302 include recurrent rounds of recombination, then natural selection performs an essentially
303 heuristic search that is much more efficient (Hickey and Golding 2018). Through
304 recombination, nature exploits its own form of the multiple weights update algorithm (see
305 Arora et al 2012, Chastain et al 2014). It does this by favoring the best available sub-
306 optimal genotypes and then recombining them. This solution removes the doubts that have
307 been expressed about the application of classic population genetics theory to adaptive
308 evolutionary change (see Messer et al 2016, Reznick 2016). Since recombination allows
309 selection to act independently at each locus, many genes can be selected simultaneously
310 without incurring a prohibitively large reproductive cost.

311 Our conclusion is that the adaptive function of sexual reproduction lies in its ability
312 to translate increasing allele frequencies at many individual loci into increasing numbers
313 of favorable alleles per chromosome. In the short term, natural selection acts on whole
314 genotypes but recombination then disassembles these selected genotypes and recombines
315 them into new, higher-fitness combinations. In the long term, this two-step iterative process
316 of selection and recombination can produce genotypes containing potentially thousands of
317 advantageous genetic variants. Without recombination, the efficient production of such
318 genotypes would be mathematically unattainable in a finite population. Looking at the
319 problem from another perspective, we could say that recombination is the process that
320 translates short-term genotypic selection into long-term genic evolution.

321

322

323 **Acknowledgements.**

324 This work was supported by grant number RGPIN-2020-05733 from NSERC Canada to
325 GBG.

326

327

328

329 **References**

- 330 Barrett, R.D.H., Schluter, D. 2007. Adaptation from standing genetic variation.
331 Trends Ecol. Evol., 23: 38-44.
- 332 Burke, M.K., Dunham, J.P., Shahrestani, P., Thornton, K.R., Rose, M.R., Long, A.D.
333 2010. Genome-wide analysis of a long-term evolution experiment in *Drosophila*.
334 Nature 467, 587-590.
- 335 Chastain, E., Livnat, A., Papadimitriou, C., Vaziranian U. 2014. Algorithms, games and
336 evolution. Proc. Nat. Acad. Sci. USA, 111, 10620-10623.
- 337 Colegrave, N. 2012. The evolutionary success of sex. EMBO Rep. 13, 774-778.
- 338 Crouch, D.J.M., 2017. Statistical aspects of evolution under natural selection, with
339 implications for the advantage of sexual reproduction.
340 J. Theoret. Biol. 431, 79 -86.
- 341 East, E. M. 1918. The role of reproduction in evolution. Amer. Nat. 52, 273-289.
- 342 Edhan, O., Hellman, Z. and Sherill-Rofe, D. 2017. Sex with no regrets: How sexual
343 reproduction uses a no regret learning algorithm for evolutionary advantage.
344 J Theor Biol. 426, 67-81.
- 345 Ewens, W. J. 1972. The substitutional load in a finite population.
346 Amer. Natur. 106, 273-282.
- 347 Fisher, R.A., 1930. The Genetical Theory of Natural Selection. Oxford: Clarendon.
- 348 Haldane, J.B.S. 1957. The cost of natural selection. J. Genet. 55, 511-524.
- 349 Hickey, D.A., Golding, G.B. 2018. The advantage of recombination when selection is
350 acting at many loci. J. Theoret. Biol. 442, 123-128.
- 351 Hickey, D.A., Golding, G.B. 2019. Sex solves Haldane's dilemma. Genome 62, 761-768.
- 352 Iles, M.M., Walters, K., Cannings, C., 2003. Recombination can evolve in large finite
353 populations given selection on sufficient loci. Genetics 165, 2249-2258.
- 354 Kleiman, M., Hadany, L. 2015. The evolution of obligate sex: the roles of sexual
355 selection and recombination. Ecol. Evol. 5, 2572-2583.
- 356 Kosheleva, K., Desai, M.M., 2017. Recombination alters the dynamics of adaptation on
357 standing variation in laboratory yeast populations. Mol. Biol. Evol. 35, 180-201.
- 358 Lewontin, R.C., Hubby, J.L., 1966. A molecular approach to the study of genic

- 359 Heterozygosity in natural populations. II. Amount of variation and degree of
360 heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* 54,
361 595-609.
- 362 Maynard Smith, J., 1978. *The Evolution of Sex*. (Cambridge University Press).
- 363 McDonald, M. J., Rice, D. P., Desai, M. M., 2016. Sex speeds adaptation by altering the
364 dynamics of molecular evolution. *Nature* 531, 233-236.
- 365 Meirmans, S., Strand, R., 2010. Why are there so many theories for sex, and what do we
366 do with them? *J. Hered.* 101(Supplement 1), S3–S12
- 367 Messer, P.W., Ellner, S.P., Hairston, N.G. 2016. Can population genetics adapt to rapid
368 evolution? *Trends Genet.* 32, 408-418.
- 369 Muller, H.J., 1932. Some genetic aspects of sex. *Amer. Natur.* 66, 118-138.
- 370 Otto, S.P., 2009 The evolutionary enigma of sex. *Am Nat* 174(Suppl 1), S1–S14.
- 371 Otto, S.P., Barton, N.H. 2001. Selection for recombination in small populations.
372 *Evolution* 55, 1921-1931.
- 373 Otto, S.P., Gerstein, A. C. 2006. Why have sex? The population genetics of sex
374 and recombination. *Biochem. Soc. Transact.* 34, 519-522.
- 375 Peck, J.R., Waxman, D. 2000. What's wrong with a little sex? *J. Evol. Biol.* 13, 63-69.
- 376 Reznick, D. 2016. Hard and soft selection revisited: how evolution by natural selection
377 works in the real world. *J. Hered.* 107: 3-14.
- 378 Stearns, S.C. 1985. *The evolution of sex and the role of sex in evolution*.
379 *Experientia* 41, 1231-1235.
- 380 Teotónio, H., Chelo, I.M., Bradic, M., Rose, M.R., Long, A.D., 2009. Experimental
381 evolution reveals natural selection on standing genetic variation.
382 *Nat Genet.* 41, 251-257.
- 383 Weismann, A. 1887. On the significance of the polar globules. *Nature*, 36, 607-609.
- 384 Williams, G. C., 1975. *Sex and evolution*. Princeton University Press, Princeton,
385 New Jersey.
- 386