

# Surprisingly Low Limits of Selection in Plant Domestication

Robin G. Allaby<sup>1</sup>, James L. Kitchen<sup>1</sup> and Dorian Q. Fuller<sup>2</sup>

<sup>1</sup>School of Life Sciences, Gibbet Hill Campus, University of Warwick, Coventry, UK. <sup>2</sup>Institute of Archaeology, University College London, London, UK.

## Supplementary Issue: Evolutionary Genomics

**ABSTRACT:** Current debate concerns the pace at which domesticated plants emerged from cultivated wild populations and how many genes were involved. Using an individual-based model, based on the assumptions of Haldane and Maynard Smith, respectively, we estimate that a surprisingly low number of 50–100 loci are the most that could be under selection in a cultivation regime at the selection strengths observed in the archaeological record. This finding is robust to attempts to rescue populations from extinction through selection from high standing genetic variation, gene flow, and the Maynard Smith-based model of threshold selection. Selective sweeps come at a cost, reducing the capacity of plants to adapt to new environments, which may contribute to the explanation of why selective sweeps have not been detected more frequently and why expansion of the agrarian package during the Neolithic was so frequently associated with collapse.

**KEYWORDS:** domestication, cost of selection, Haldane, Maynard-Smith

**SUPPLEMENT:** Evolutionary Genomics

**CITATION:** Allaby et al. Surprisingly Low Limits of Selection in Plant Domestication. *Evolutionary Bioinformatics* 2015:11(S2) 41–51 doi: 10.4137/EBO.S33495.

**TYPE:** Original Research

**RECEIVED:** October 19, 2015. **RESUBMITTED:** December 10, 2015. **ACCEPTED FOR PUBLICATION:** December 13, 2015.

**ACADEMIC EDITOR:** Jake Cui, Associate Editor

**PEER REVIEW:** Five peer reviewers contributed to the peer review report. Reviewers' reports totaled 2665 words, excluding any confidential comments to the academic editor.

**FUNDING:** This work was kindly supported by the Leverhulme Trust (grant number F/00 215/BC). DQF's research on Comparing Pathways to Agriculture is supported by a grant from European Research Council (no. 323842). The authors confirm that the funders had no influence over the study design, content of the article, or selection of this journal.

**COMPETING INTERESTS:** Authors disclose no potential conflicts of interest.

**CORRESPONDENCE:** r.g.allaby@warwick.ac.uk

**COPYRIGHT:** © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

Paper subject to independent expert blind peer review. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE).

Published by Libertas Academica. Learn more about this journal.

## Introduction

Domestication is an evolutionary process that provides a cornerstone to understand the mechanism of selection.<sup>1</sup> In the case of plants, the evolution of domestication involves the selection of a characteristic group of traits that are collectively termed as the domestication syndrome.<sup>2,3</sup> These traits include the loss of shattering, changes in seed size, loss of photo-period sensitivity, and changes in plant and spikelet architecture.<sup>4</sup> It is interesting that among crops of a similar type such as cereals, the number of obvious syndrome traits is also similar.<sup>4</sup> It is not known whether this represents the maximum number of traits that could have been selected within the time period of syndrome fixation spanning several thousand years or whether further traits could have been selected in each case but were not. There has been much debate about how these traits were selected, the pace and strength of selection, and the extent to which evolution under domestication continues today.<sup>4–13</sup>

Classic field trials of experimental harvesting of wild progenitors of wheat suggested that in the case of cereals, selection coefficients ( $s$ ) as high as 0.6 could have been in operation during domestication, resulting in fixation of loss of shattering traits within a few decades.<sup>14</sup> These findings have been taken in support of a rapid transition model of agricultural

origin<sup>15</sup> in which domesticated forms of crops appeared over a very short time period in a “Neolithic Revolution”.<sup>16</sup> However, an emergent feature of the archaeological record in recent years has been the protracted appearance of domesticated crops,<sup>17–24</sup> and estimates of selection coefficients made directly from the archaeobotanical record have been as low as 0.003.<sup>13,25</sup> Under a protracted scenario of domestication, the expected patterns of genetic diversity need to be re-evaluated in order to interpret the evolutionary history of domestication.<sup>26,27</sup> For instance, under protraction, traits may have been selected more slowly in the face of gene flow between cultivated and wild populations, resulting in the appearance of relatively weak selection coefficients. This extended time period would increase the opportunities for parallelisms in syndrome traits, as similar traits are independently selected in distinct geographic locations and genomic mosaicism associated with phylogeography could occur.<sup>27</sup> It could also be the case that the protracted process allowed more traits to be selected than would have been possible under the restricted time period of a rapid transition of a few tens to hundreds of years. It is therefore useful to establish the amount of selection that would have been possible to drive the evolution of domestication syndrome under protracted and rapid transition scenarios.



Selection comes at a cost in that some organisms must die before reproducing each generation for their genes to be selected against the wild-type allele, therefore causing a reduction in the overall population size.<sup>28–31</sup> A consequence of this cost is that the amount of selection that a population can withstand, without dropping to unsustainably low demographic levels, is limited. Haldane noted that for this reason, there is a limit to the number of traits that plant breeders are able to select at a given time, and that the pace at which evolution can be driven by natural selection is limited because of the cost of selection.<sup>32</sup> If the size of a population is reduced below a critical level, it becomes highly susceptible to extinction by demographic stochasticity, even if the genetic capacity to adapt to new environmental conditions is present in the population.<sup>33</sup> Haldane<sup>32</sup> deduced that the cost of selection would tend to make adaptive evolutionary rates slow to medium, with selection with single-gene substitutions taking several hundred generations.

The Haldane model was criticized by Maynard Smith for being overly simplistic, thereby failing to take account of the fact that numerous adaptive solutions may apply to a single selection pressure.<sup>34</sup> Under this argument, many thousands of loci could be under selection for a single cost, and consequently the Haldane model has largely been dismissed as being too inaccurate.<sup>35</sup> The possibility of numerous adaptive solutions for any one selection pressure leads to the increased possibility of evolutionary rescue. As substitution of new adapted genotypes outpaces the decline in population due to changed environmental conditions, such adaptive genotypes come to dominate and establish an adapted population.<sup>36,37</sup> The likelihood of evolutionary rescue increases with larger initial population sizes, increased standing genetic variation from which new adaptations are drawn, and lower levels of maladaptation of older established genotypes. Thus, parameters of demography and selection strength will constrain the conditions under which a population can successfully adapt in a new environment. Selection strength, however, is partly constrained by how many genetic loci are involved in the new adaptive phenotypes.

Taken in relation to crop domestication, these principles of evolution raise questions as to how many loci could have been selected simultaneously during the domestication process and how strong the selection coefficient on any one locus would have been.

In order to understand the pace of selection of the domestication syndrome traits, it is necessary to know the number of genetic loci involved. Increasingly, the underlying genetic bases of domestication syndrome are being elucidated,<sup>38</sup> and the number of loci associated with their control has recently been estimated to be at least 27, and as much as 70 in tetraploid wheats.<sup>39,40</sup> Estimates of loci under selection from the genome analysis of other crops have yielded a range of numbers, with up to 1,200 loci being suggested in maize, but an expectation of ~40 is prevalent at the signal strength of *tb1*<sup>41</sup> and 36 loci

in sunflower.<sup>42</sup> The power to detect signatures of selection in genetic data is limited,<sup>43</sup> and whether these numbers of loci under selection represent the totality of loci or simply the most strongly selected loci that reach the threshold of detectability is thus unclear. These numbers are high relative to the typical number of syndrome traits because some traits, such as seed size, are under polygenic control,<sup>44</sup> while other traits, such as loss of shattering, are under the control of one or just a few genes.<sup>45–50</sup> Intuitively, it seems likely that traits under monogenic control may be subject to stronger selection pressures than those under polygenic control because the selection coefficients associated with each locus for a trait have an additive effect. Therefore, as loci governing a trait are progressively added, the value of *s* for each locus must progressively reduce in order to maintain the same overall selection pressure on the individual organism. Consequently, it might be expected that those traits of domestication syndrome that are under monogenic control would have been under the strongest selection and would have appeared at the earliest. Surprisingly, the reverse is observed in the archaeological record in that the tough rachis mutant appears to be selected slowly and lately when compared to other traits,<sup>4,18</sup> and increase in seed size appears very early on in the archaeological record, despite the complexity of its genetic control and a rate of selection that is not as different as one might expect to that of shattering.<sup>13,25</sup> While the timing of the first appearance of traits is attributable to a sequence of different behaviors of proto-farmers subjecting different selection pressures on plants at different times,<sup>12,51</sup> the surprising similarity in rates highlights the need to better understand the selection pressures involved in domestication.

The aim of this study was to use an individual-based model to simulate virtual plants with progressively greater numbers of loci and strengths of selection to determine the maximum number of loci that would possibly be under selection in a cultivation regime.

## Methods

**Rationale of an individual-based model to explore selection.** In this study, the limitations imposed by the cost of selection were examined through computer simulations to establish the relationship between the number of loci under selection, the strength of selection, and the ability of plant populations to recover from reduced population sizes resulting from rounds of selection. The simulations were carried out using a program written by R.G.A. available for download (<http://www2.warwick.ac.uk/fac/sci/lifesci/research/archaeobotany/downloads/limits>).

The simulations were executed under a scenario based on the archaeological record in which traits appear and are selected over a period of 3,000 years<sup>4,13,18,19</sup> in annual plants, with nonoverlapping generations. Populations of virtual plants were endowed with a number of loci, which were considered unlinked to each other and inherited through a process of random segregation. In reality, the overwhelming majority

of domestication syndrome loci have been identified to be regulatory in nature.<sup>7,52</sup> Consequently, it can be inferred that most mutations have been epistatic in their effect. One can consider the changes in gene expression caused by mutation at a distal regulatory locus to be a phenotypic consequence, and that the focus of selection would act on the regulatory locus rather than the regulated locus. Such a situation can be reasonably modeled by treating each locus under selection as independent to other loci. This model makes no assumption about the function of the loci under selection, and thus the stage of the life cycle in which the resultant phenotype is expressed, and is inclusive of epistatic mutations. Furthermore, phenotypic traits may be under monogenic or polygenic control. A model of independent selection of loci inherently describes traits under monogenic control. Traits under polygenic loci are also described if it is assumed that the overall trait, such as seed size, is contributed by independently segregating loci, which is largely true for quantitative traits. One might also consider that the selective value of a mutation at a locus is dependent on the presence of mutations at other loci. However, this would require a specific knowledge of the dependencies, which would preclude a general model. Finally, it is known that some domestication syndrome loci are located in close proximity to each other,<sup>53</sup> and that positively selected mutations may be associated with either linked deleterious loci, or have slightly deleterious pleiotropic effects.<sup>54</sup> In the case of tightly linked loci, it is reasonable to model a single locus and the associated selection coefficient, thereby representing an overall selective value.

Given the assumptions outlined, it is reasonable to use a general model of selection of adaptive variants at independently segregating loci in order to assess the amount of selection that a plant population can endure under domestication. Rather than using a model of mutation, adaptive variants were seeded into populations and reseeded if lost through drift, thereby negating the effect of population size on the early stages of selection. Each locus was associated with a selection coefficient ( $s$ ), which selected against the wild-type allele. Adaptive variants were generated for loci in the population that had a fitness value of 1, and so they were not selected against. Individuals survive with a probability equal to the product of the fitness values of alleles across all loci. This represents an implementation of the Haldane view of selection cost. The Maynard Smith threshold view of selection cost can be incorporated by a modification of this model, in which individuals survive by independent probability trials for each locus, such that only one locus needs to be “successful” during a round of selection in order for the individual to survive. The two models can be combined into a third model that accounts for a number of different selection pressures, each of which has multiple adaptive solutions across a number of loci, but in which each locus provides an adaptive solution to just one selection pressure. All three models were applied in this study.

Under this system, after a round of selection, the resulting individuals in the next generation would be fewer than the previous generation. The ability of the population to recover from such a round of selection is determined by a maximum fecundity ( $m_f$ ) parameter, where each individual was capable of having more than one progeny, causing the population to expand. However, population expansion is tempered by both environmental checks and an environmental carrying capacity, and so the number of individuals generated in the next generation is expanded from the current generation value by an amount, which is determined by the  $m_f$  parameter, up to the carrying capacity population size that could not be exceeded. For each set of simulation conditions, the probability of extinction, severity of selection bottleneck, and the rate and extent of fixation of domestication syndrome trait controlling alleles were determined.

#### **Model parameters and execution. Initialization.**

Simulations were carried out with populations of 1,000 individuals. The simulation begins with an initialization of the population in which diploid hermaphrodite individuals are assigned wild-type alleles for the defined number of loci under selection. A seeding rate of 0.001 was used to generate adaptive variants in the population, such that on average, a single advantageous allele variant per locus would appear each generation in populations that had no variants at those loci. New individuals were generated by randomly selecting two parent individuals, and then generating a gamete from each by randomly selecting one of the two possible homologs at each locus and combining the gametes. There was a probability of selecting the same parent donor twice equal to the mating strategy (0.02 for outcrossing simulations and 0.98 for inbreeding populations).

Newly generated individuals were then immediately subjected to a round of selection (to represent selection at any part of the life cycle prior to reproduction). The mode of selection followed the Haldane, Maynard Smith, or the mixed model outlined below.

*Haldane selection model.* In this model, each locus hosts an adaptive variant to a different selection pressure. The model therefore assumes that a single selection pressure is associated with a single gene. Newly generated individuals then survived with a probability, which is equal to the product of the fitness values of the alleles they carried, such that the probability of survival ( $su$ ) was defined as follows:

$$su = \prod_1^k \omega_i \quad (1)$$

For  $k$  loci, where  $\omega_i$  is the fitness of the  $i$ th locus as given by:

$$\omega_i = 1 - s_i \quad (2)$$

Where  $s_i$  is the selection coefficient of the  $i$ th locus. The selection coefficient of the  $i$ th locus was moderated by the value lambda for heterozygotes ( $S_{het}$ ) such that



$$s_{ibet} = (1 - \lambda)s_i \quad (3)$$

A value of 0 was taken for lambda in simulations in this study to represent recessive mutations, which represent the majority of known mutations associated with domestication, and a value of 1 to represent dominant mutations.<sup>38</sup>

**Maynard Smith threshold selection model.** This model differs from the Haldane model in that numerous loci host adaptive variants to a selection pressure, and only one locus needs to pass a round of selection. Therefore, this model assumes that a single selection pressure is associated with multiple genes. Only a single selection pressure is considered in this model. Therefore, the probability of survival can be summarized as follows:

$$su = 1 - s^k \quad (4)$$

for  $k$  loci, which hosts adaptive solutions to a selection pressure.

**Mixed selection model.** This model combines the Haldane and Maynard Smith models by including numerous selection pressures, each with numerous loci that host adaptive variant solutions. While each pressure has multiple adaptive variants, each adaptive variant provided a solution to only a single pressure in this model. In this case, the probability of survival can be expressed as follows:

$$su = \prod_1^j (1 - s_i^k) \quad (5)$$

where there are  $j$  selection pressures, and each pressure has  $k$  loci hosting adaptive variants. In order to survive a round of selection, each individual had to have at least one successful trial for each selection pressure.

**Fecundity and population recovery.** In the first generation, the steps described above to generate individuals were repeated for a number of times equal to the population size, which led to a number of individuals in the next generation that were fewer than this value. In subsequent generations, the number of attempts to generate new individuals was given by

$$N_{\text{attempts}} = N_{n-1}m_f \quad (6)$$

For ( $N_{\text{attempts}} < \text{initial population size}$ ), where  $N_{\text{attempts}}$  is the number of individuals created and then challenged,  $N_{n-1}$  is the number of individuals in the previous generation, and  $m_f$  is the maximum fecundity parameter. In cases where  $N_{\text{attempts}}$  was calculated to be greater than the initial population size, the latter value was used instead to represent the carrying capacity of the environment.

This whole process was iterated for the specified number of generations. Simulations were carried out for 3,000 generations. Each set of simulation conditions was repeated 100 times, unless otherwise stated, and the average frequencies

of advantageous variants for each locus and population sizes were recorded for each generation.

**Standing variation trials.** In the case of standing variation trials, the simulations were executed as described above, but adaptive variants for each locus were seeded at a frequency of 50%. During seeding, an approximation was applied, in which adaptive variants were distributed across individuals with respect to the mating strategy, such that if an individual had been selected to carry an adaptive variant, there was a probability equal to the mating strategy that they would be homozygous. If the frequency of adaptive variants decreased to zero, then they would be reseeded with a frequency of only one, as with the previous simulations, rather than re-establishing a 50% frequency.

**Gene flow trials.** Gene flow was achieved through immigration following rounds of selection on the native population. The maximum number of immigrants ( $m_i$ ) possible was calculated as follows:

$$m_i = gf \times cc \quad (7)$$

where  $gf$  is the gene flow and  $cc$  is the carrying capacity of the environment, in terms of the maximum population size possible. The immigration space ( $is$ ) was then calculated as follows:

$$is = cc - N \quad (8)$$

where  $N$  is the current population size. The lesser of the two values between  $is$  and  $m_i$  were then taken as the number of immigrants to generate and add to the current population. If immigrants were deemed to be from an area of high standing variation, then they were seeded with adaptive variants as described above.

## Results and Discussion

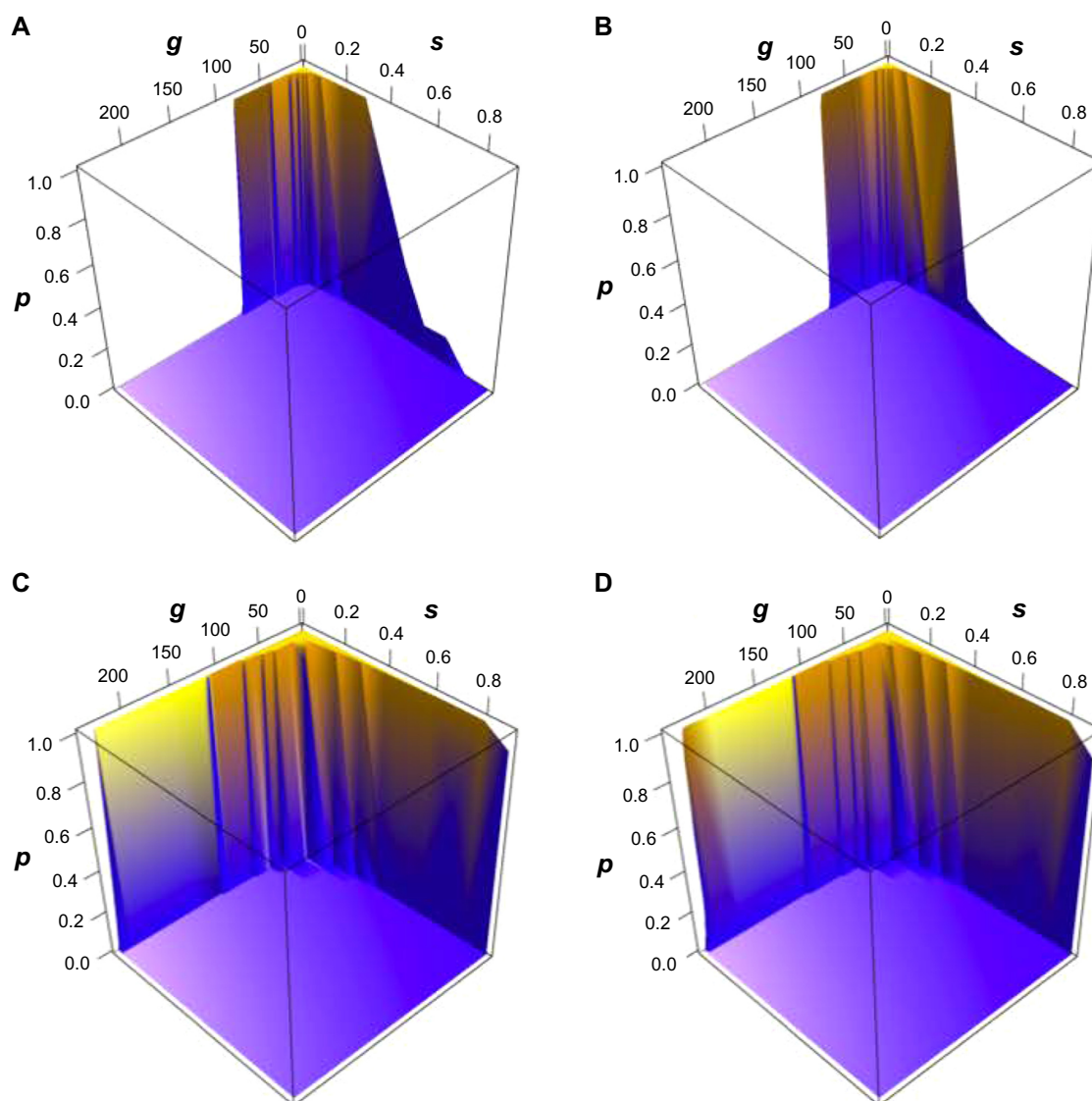
In the first instance, simulations were carried out using the simplest, Haldane-based, model of cost to provide a baseline expectation. After this, we explored parameters that could rescue populations from extinction and increase the amount of selection that could be endured. Rescuing parameters considered were increased fecundity, selection from high standing variation, gene flow through immigration, and threshold selection, in which multiple adaptive solutions across loci are available in a population.

**A baseline simulation of the maximum number of loci underselection.** In the first set of simulation experiments, a model in which increasing numbers of loci subject to the same selection coefficient was applied. A conservative reproductive regime was applied in which there was an underlying assumption that the organisms were held in close check by organisms of different species in the environment, close to Darwin's original insight that the typically geometric potential for species to reproduce is held back by complex interspecific competition. Under this regime, the  $m_f$  parameter was set to 1.5, such that a population was capable of expanding 50% at most per generation.

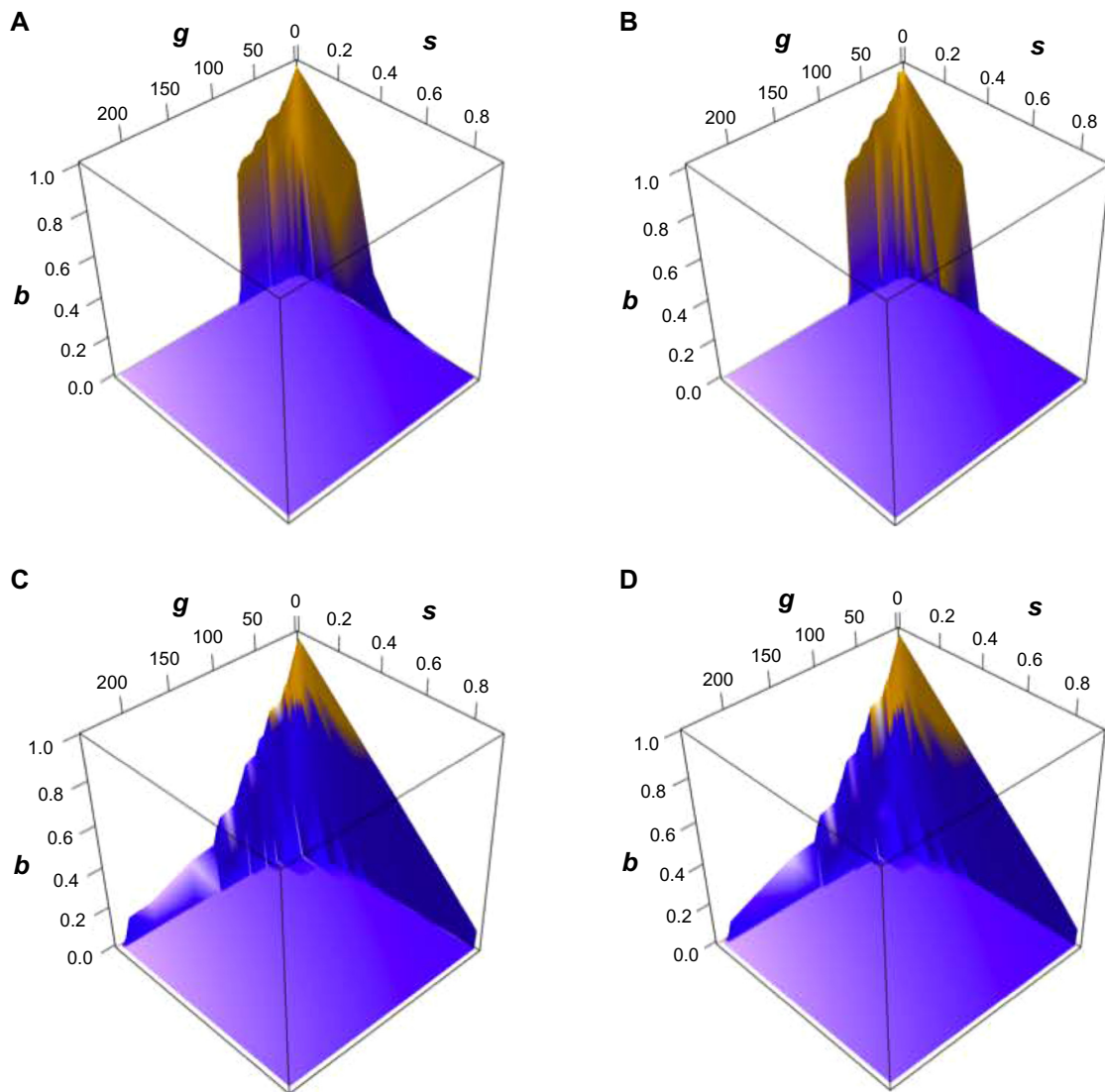
Simulations were carried out on populations that began with 1,000 individuals to reflect a reasonable size based on genetic diversity studies,<sup>55,56</sup> but it should also be remembered that the pace of selection for a given allele under a given selection coefficient ( $s$ ) is expected to be independent of population size.<sup>57</sup> This should not be confused with the parameter of selection intensity ( $2N_e s$ ) often used in coalescent approaches that describes the impact of selection on genetic diversity,<sup>58</sup> which is dependent on population size.

Each set of parameters were simulated 100 times and considered as recessive adaptive variants. Two mating systems were explored, a 2% out-crossing (inbreeding) population, similar to that of wheat or barley and a 98% out-crossing (out-crossing) population to represent out-crossing crops (Supplementary Tables 1–4). These mating systems represent the normal biases

expected in plants.<sup>59</sup> Simulations were carried out with increasing numbers of loci and increasing selection coefficients, with the proportion of simulations resulting in population extinction at the end of the simulation sets being a record of how the population fared. These three simulation output parameters were used to construct the probability of survival landscapes under the two mating system regimes, as shown in Figure 1A and 1B and Supplementary Tables 1 and 2. It is notable that in all simulation parameters explored, the switch from survival to extinction with increasing loci number under selection is precipitous, with populations going from an estimated probability of survival of 1 to 0 with the addition of just one to five loci in most cases. This change is more abrupt in out-crossing populations than inbreeding ones. Generally, very few loci could be simultaneously under selection at selection coefficients of  $\geq 0.1$ ,



**Figure 1.** Probability landscapes of population survival ( $p$ ) for a given number of loci ( $g$ ) under a selection coefficient ( $s$ ). Graphs are colored to show relief only. (A) Inbreeding population with  $m_t$  value of 1.5. (B) Out-crossing population with  $m_t$  value of 1.5. (C) Inbreeding population with  $m_t$  value of 10. (D) Out-crossing population with  $m_t$  value of 10.



**Figure 2.** Landscapes of minimum population bottleneck ( $b$ ) expressed as a proportion of the original population size for a given number of loci ( $g$ ) under a selection coefficient ( $s$ ). Graphs are colored to show relief only. **(A)** Inbreeding population with  $m_f$  value of 1.5. **(B)** Out-crossing population with  $m_f$  value of 1.5. **(C)** Inbreeding population with  $m_f$  value of 10. **(D)** Out-crossing population with  $m_f$  value of 10.

and values of  $s > 0.3$  were unlikely to be survived by populations under the conservative  $m_f$  regime. We found that 80 and 83 were the maximum numbers of loci that could be selected under the conservative expansion regime before extinction occurred under the lowest explored value of  $s$  (0.005). This value is close to typical values under natural selection and that calculated from cereals in the archaeological record.<sup>13,25</sup>

The probability of survival became  $< 1$  in populations that generally experienced bottlenecks of  $< 30\%$  and  $< 10\%$  at the lower values of  $s$ . While bottlenecks of this extremity may be found in nature, it is unlikely that such a bottleneck could be tolerated under a cultivation regime since the majority of the food source would disappear. We constructed landscapes to view the effect of selection on bottleneck size, as shown in Figure 2. A precipitous drop is discernible under the conserved  $m_f$  regime, which occurs at a threshold point when the

bottleneck size is between 60% and 70% of the initial population size. We judge that the reasonable level of bottleneck that could be tolerated by cultivators is 60%–70%, before there is too great a reduction in food production for it to be worthwhile investing in cultivation.

We considered the effect of dominance by repeating simulations seeded with dominant adaptive variants using a value of  $s$  set to 0.01. In both the cases of outbreeding and inbreeding populations, there was complete extinction between 45 and 50 loci under selection, as shown in Supplementary Table 3. We conclude that the difference to the overall selection cost is marginal between recessive and dominant adaptive variants.

A baseline result of 50–100 loci using the simple Haldane model of cost and a conservative fecundity regime is particularly interesting because it closely mirrors the

number of loci under selection identified in genome studies of major crops.<sup>39–42</sup>

**Rescue from selection cost by increased fecundity.** A higher fecundity is expected to result in an increased ability for a population to recover from rounds of selection, and therefore such populations are expected to withstand greater numbers of loci under selection. The second fecundity regime we considered was considerably more liberal and the  $m_f$  value was set to 10, allowing a ten-fold expansion. The underlying rationale in this case was that a cultivated environment inherently reduces interspecific competition, excepting that which is due to human predation. The extent to which a crop population can expand is a function of the number of propagules generated per plant and the proportion of the harvest that is set aside for sowing in the following year. Previous studies have suggested that a quarter of a crop harvest may be set aside, which represents about an eighth of the propagules generated in a generation as it was estimated that approximately 50% of seeds were harvested, with the rest being lost due to dispersal and not being sown.<sup>14</sup> Under such a regime, an eight-fold increase per individual would maintain an overall constant population size, or an  $m_f$  value of 1. However, cereals such as barley have up to 20 or 30 grains per plant, which would lead to a three- to four-fold increase in population size per generation under these conditions. In reality, it is likely that a lower proportion of the grain would be sown to compensate for bringing the  $m_f$  parameter to something close to our conservative regime; however, we selected an  $m_f$  value of 10 to explore the possible effect that could be introduced by human agency. We consider an  $m_f$  value of 10 to be in excess of what is likely to be achieved.

Under the higher  $m_f$  regime, we found that 227 and 230 loci could be under selection in out-crossing and inbreeding systems, respectively, at the lowest explored value of  $s$  (0.01), as shown in Figure 1C and D and Supplementary Tables 4 and 5. The rescue effect is also apparent in the size of the population bottleneck with increasing pressure, as shown in Figure 2C and D, in which there is a steady decline in bottleneck size rather than the precipitous threshold point observed under the conserved fecundity regime. However, despite the rescue effect of a higher fecundity, the population size still reduces to below 60% of the starting population size at less than 100 loci under selection, as shown in Supplementary Figure 1. Reducing the selection coefficient to 0.005 to explore the bottleneck size at 100 loci confirms that the population is still reduced to around or just below the 60% level. We conclude that rescuing the population by increased fecundity does not increase the maximum number of loci that could be under selection in a cultivation regime to >100 loci without the invocation of unrealistically high fecundity values.

**Rescue by selection from high standing variation.** The simulations so far consider variants that are seeded in very low frequencies in wild populations that are generally selected against in the wild. A number of regulatory genes have been identified in maize that have been selected during

domestication, which occur in a wide range of frequencies (0.1–0.88) in the wild progenitor.<sup>60,61</sup> In these sorts of cases, it is not expected that the wild and cultivated populations are subject to diametrically opposed selection regimes, but instead, aspects of adaptations to the wild environment or neutral standing variation have selective values in the cultivated environment. Models have shown that even strong selections will likely not leave a detectable signature of selection from standing variation at higher frequencies,<sup>58,62</sup> because less of the adjacent genomic variation is lost during the sweep process. Consequently, this part of the selection process during domestication would be largely invisible from genome diversity scan approaches. To investigate whether our estimates of the number of loci that could be under selection were different from standing variation rather than spontaneous mutation, we carried out simulations in which the starting frequency of the adaptive variant was seeded at 0.5 to reflect the expected average level of standing variation, as shown in Supplementary Figure 2 and Supplementary Tables 6 and 7. Generally, populations could sustain a larger number of loci under selection under these conditions of standing variation. Under these conditions, there was a greater difference between the mating strategies than for selection from low frequency variants. Outbreeding populations sustained about 50% more loci before population extinction was observed, whereas inbreeding populations sustained over 100% more loci. The difference between the mating strategies is likely explained by the difference in heterozygote proportions between the two population types. Inbreeding populations hold most of the recessive mutations in homozygous individuals, which therefore confer an advantage to the individual. The difference between selection from high standing variation and low frequency variants is less pronounced when the population bottleneck is considered, as depicted in Supplementary Figure 3. The 60%–70% threshold that we consider to be a realistic pragmatic limit for cultivation is reached in the 50–110 loci range from standing variation for weak selection ( $s = 0.005$ ). We therefore conclude that although adaptation from high standing variation can rescue populations from extinction, it would need to be atypically high in order to raise the bottleneck population size above the threshold of viability in a cultivated environment.

**Rescue through gene flow from adjacent populations.** The previous simulations consider selections within an isolated population. The influx of individuals from an adjacent environment that is not subject to the same selection regime would be expected to have a rescue effect in terms of the population size and consequently an impact of selection. The domestication of plants may have occurred either within or outside the wild progenitor biogeographic range. Under conditions of high gene flow, it has been postulated that domestication may not be possible, because of swamping of the cultivated gene pool with wild alleles.<sup>27</sup> Furthermore, archaeological evidence supports the notion that cultivated populations that



were outside the wild range in the Near East were rescued by re-stocking from the wild.<sup>22</sup>

In a fourth set of experiments, we explored the impact of gene flow through immigration to see whether the rescue effect would increase the number of loci that could be under selection and the effect on the populations' ability to adapt. A selection coefficient  $s$  of 0.01 was selected, with either 45 or 50 loci under selection representing and exceeding the loci limit for isolated populations (see Supplementary Tables 1 and 2). We simulated two levels of gene flow, which represents an immigrating population per generation that represented a proportion of 0.01 or 0.1 of the original population size, as shown in Supplementary Figure 4 and Supplementary Table 8. The introduction of gene flow had the expected effect of preventing population extinction, but it slowed down or prevented the fixation of the selectively advantageous variants. We then simulated a "standing variation" population in which the variant under selection occurred at a 50% frequency in the population, but not in the surrounding populations that supplied immigrants, using 130 and 140 loci for inbreeding populations and 80 and 85 loci for out-crossing populations representing and exceeding the loci limits for those conditions, as shown in Supplementary Figure 5 and Supplementary Table 8. In this case, the gene flow again slowed down the rate of mutant fixation. Under conditions that would have caused an isolated population to become extinct, out-crossing populations did not fix the mutant above 0.1%. A very slow increase in mutant frequency under high gene flow conditions was apparent in inbreeding populations (140 loci, gene flow 0.1). However, this increase in frequency is equivalent to that which would be expected for a selection pressure of only 0.0002, which is too weak to overcome the effects of genetic drift in a population of this size. Hence, the change in allele frequency is attributable to drift rather than selection in this case. A situation in which one population has a standing variation that is absent from adjacent populations appeared to be an unlikely one, and so we also simulated gene flow to a population with standing variation (50% mutant frequency) from similar adjacent populations, as shown in Supplementary Figure 6 and Supplementary Table 8. In these simulations, the result of gene flow was in all cases to prevent an increase in the variant frequency at even relatively low levels of gene flow. These simulations lead to the counterintuitive observation that standing variation from surrounding populations, rather than aiding selection in a population through rescue by the introduction of "fit" individuals, actually serves to compete against the selected population, thereby preventing adaptation.

The gene flow simulations show that although gene flow will prevent the extinction of a population under strong selection, it does not increase the number of loci that can be under selection, even if the immigrating individuals also have a proportion of fit genotypes.

**Rescue by threshold selection.** The Haldane model is considered to be too simplistic by many because it fails to

take account of the situation in which multiple different and possibly unrelated traits may offer an adaptive solution to a single selection pressure.<sup>35</sup> Maynard-Smith<sup>34</sup> demonstrated that under such circumstances, all loci would be subject to a single cost and that many thousands of loci could be simultaneously under selection. This leads to the threshold model of selection in which individuals that have the highest combination of adaptive loci will tend to be successful. It also leads to a rescue effect because of the increased probability that any one individual in a population would survive a round of selection. We applied a "Maynard Smith" model in which all loci carried an adaptive solution to a single selection pressure. In this case, an individual survives a round of selection if any one of the loci passes a selective trial, rather than all loci being required to pass selective trials in the Haldane model. After applying the Maynard-Smith model,<sup>34</sup> we applied a value of  $s$  set to 0.5, which under the Haldane model would not be endurable by more than one locus, and applied it to five loci simultaneously, as shown in Supplementary Figure 7. Rescue of the population was absolute with no extinction occurring in these simulations. However, typically only one or two adaptive variants were fixed, while the remaining variants fluctuated in frequency in a drift-like manner. Under weaker selection, one might expect that most if not all of the alternative adaptive variants would be lost through drift, which we confirmed when simulating 40 loci under a value of  $s$  of 0.01, without reseeded adaptive variants, as shown in Supplementary Figure 8. It has been previously suggested that alternative adaptive solutions might be expected to be selectively neutral and relative to each other.<sup>27</sup> This suggests that although many loci could be responsive to the same selection pressure, as Maynard Smith described, in reality, the usual outcome would be the fixation of a single adaptive variant. We explored this further by considering a mixed model in which there were multiple selection pressures, each of which had a number of adaptive solutions as in our Maynard Smith model, but each selection pressure was independent of the others, so that no one locus could provide adaptive solutions to more than one selection pressure. In the first instance, we considered five selection pressures, each of which had 4 loci providing adaptive variants, giving a total of 20 loci under selection, at a value of  $s$  set to 0.01, as shown in Supplementary Figure 9. In this case, the number of adaptive variants fixed approximates the number of selection pressures, while the remaining adaptive variants appear to behave in a neutral fashion.

Under a mixed model, it would be expected that the effectiveness of rescue would increase with an increasing number of loci offering adaptive variants per selection pressure. Increasing effectiveness of rescue should be associated with less severe population bottlenecks in the adaptive process. We explored the number of loci that would need to be under selection in order to have significantly more selection pressures endured than what was found with the simple Haldane model. In this series of simulations, we had





90 selection pressures with  $s$  set to a value of 0.005, which would have been sufficient to cause extinction under the simple Haldane model (Supplementary Tables 1 and 2). The number of adaptive variants per selection pressure was set to 4 in the lowest instance and increased geometrically by a factor of 2, as shown in Supplementary Table 9. Surprisingly, for the population to maintain a level of >60% of its initial level, a total of 5760 loci had to offer adaptive variants to the 90 selection pressures. This would equate to around 20% of genes in the genome of a cereal, such as barley, thereby offering adaptive solutions, which would seem unlikely. Therefore, despite the potential for threshold selection to rescue a population from extinction, we conclude that it would not be sufficient to increase the number of selection pressures above the 50–100 level observed with the simple Haldane model, and despite the possibility of many hundreds of loci being potentially adaptive, we would not expect to see more than one variant fixed per pressure on average.

## Conclusion

Artificial selection associated with domestication is increasingly being considered as being similar to the process of natural selection.<sup>63,64</sup> One critical difference between the two may be the severity of the population bottleneck caused by the cost of selection that would be tolerable. In the case of cultivation, too severe a bottleneck would result in an economic collapse of the agrarian system. The threshold level of severity of the bottleneck that could be tolerated would in part be dependent on the level of dependency of an economy on cultivation. We have taken a value of 60% to represent a pragmatic limit in our simulations, but a wide range of values on either side of this value would cause little change in the maximum tolerable number of selection pressures (Fig. 2 and Supplementary Figs. 1–3 and Supplementary Tables 1–9). While there are numerous ways in which the cost of selection as outlined in the simplest form of the Haldane-based model can be ameliorated, when the severity of the population bottleneck is a critical factor, the findings of this model are remarkably robust. The criticism of Haldane's model by Maynard Smith that multiple adaptive solutions can be selected under a single cost is modified by the fact that there is interference between competing adaptive variants, which negates their universal fixation. We find that the level of rescue required through this mechanism to deviate greatly from the Haldane model would require significant portions of the entire genome to offer adaptive solutions. For simplicity, the model we applied does not overtly consider pleiotropy in which an adaptive variant may be responsive to more than one selection pressure. Extensive pleiotropy might be expected to reduce the number of loci required for any given number of selection pressures in the mixed model of selection. However, this is unlikely to change our conclusions about the maximum number of loci that could be under selection, since the effect of pleiotropy would be to reduce the number of loci under selection. In any one instance of a pleiotropic effect, the

two selection pressures acting upon a locus would have the simple effect of increasing selection at that locus, and as such can be considered as a single selection pressure. It is therefore our conclusion that the evolution of plant domestication, and subsequently under domestication, is unlikely to have endured more than 50–100 selection pressures, with a similar number of adaptive loci involved.

These findings suggest that genome-wide efforts to detect signatures of selection in crops are probably recovering most of the loci that have been selected and that these loci were most likely selected from spontaneous mutations since selection from standing variation is less likely to be detected in genomic signatures. The difference between the number of signatures of selection detected and the upper limits described here reflects the amount of selection that could have occurred from standing variation. This suggests that domesticated wheat<sup>39,40</sup> was mostly based on spontaneous mutation, while maize<sup>41</sup> and sunflower<sup>42</sup> may have had progressively more of their domestication adaptations from standing variation. Germane to this observation is that selection of recessive mutants is quicker in inbreeding populations, where the majority of individuals are homozygous through selfing. Therefore, it would have been easier for the inbreeding crops, such as wheat and barley, to select spontaneous mutation, while maize and sunflower would likely have had a greater pressure to incorporate standing variation and dominant adaptive variants.

While these simulations do not preclude the existence of selective sweeps, they do show that sweeps come at a cost of reducing the selection load that a population is capable of enduring. This could explain not only why sweeps are not observed more frequently in nature but also why agricultural expansion was repeatedly associated with collapse in new environments, shortly after arrival.<sup>65–67</sup> It is possible that the rapid pace of expansion could have forced equally rapid adaptation of plants to latitude, which would have required strong selection of a low number of loci, an adaptation of low complexity. Given the dynamic complex environment into which agriculture had advanced, it may have been the case that the plant populations were incapable of further adaptation to changing conditions as they occurred. To better understand the expansion of agriculture, further consideration is needed of the pace of movement across the latitudinal selection gradient in the context of tolerable limits of plants and whether different paces are associated with adaptations of low and high complexity, respectively.<sup>68</sup>

## Author Contributions

Conceived and designed the experiments: RGA. Analyzed the data: RGA, JLK. Wrote the first draft of the manuscript: RGA. Contributed to the writing of the manuscript: RGA, DQF. Agree with manuscript results and conclusions: RGA, JLK, DQF. Jointly developed the structure and arguments for the paper: RGA, DQF. Made critical revisions and approved



final version: RGA, DQF. All authors reviewed and approved of the final manuscript.

## Supplementary Materials

**Supplementary Table 1.** Summary of simulation outputs for out-crossing populations with  $mf = 1.5$ .

**Supplementary Table 2.** Summary of simulation outputs for inbreeding populations with  $mf = 1.5$ .

**Supplementary Table 3.** Summary of simulation outputs for dominant adaptive variants with  $mf = 1.5$ ,  $s = 0.01$ .

**Supplementary Table 4.** Summary of simulation outputs for out-crossing populations with  $mf = 10$ .

**Supplementary Table 5.** Summary of simulation outputs for inbreeding populations with  $mf = 10$ .

**Supplementary Table 6.** Summary of simulation outputs for inbreeding populations with mutants commencing at a standing variation frequency of 0.5,  $mf = 1.5$ .

**Supplementary Table 7.** Summary of simulation outputs for out-crossing populations with mutants commencing at a standing variation frequency of 0.5,  $mf = 1.5$ .

**Supplementary Table 8.** Summary of simulation outputs for populations with gene flow from adjacent populations under a 0.01  $s$  selection.

**Note:** ¶ Population standing variation – starting frequency of selected mutant. § immigrant population standing variation-starting frequency of selected mutant

**Supplementary Table 9.** Summary of simulation outputs for a mixed model with  $mf = 1.5$ ,  $s = 0.005$ .

**Supplementary Figure 1.** Minimum population bottleneck ( $b$ ) expressed as a percentage of original population size for a given number of loci ( $g$ ). Loci selected at 0.005 for  $mf = 1.5$ , selected at 0.01 for  $mf = 10$ .

**Supplementary Figure 2.** Probability of population survival for the number of loci under selection in inbreeding (blue) and out-crossing (red) systems from a standing frequency of the mutant under selection at 50%. (A) selection coefficient  $s$  equal to 0.3. (B)  $s$  equal to 0.01. (C)  $s$  equal to 0.005.  $mf$  parameter set to 1.5.

**Supplementary Figure 3.** Minimum population bottleneck expressed as a percentage of original population size for a given number of loci under selection in inbreeding (blue) and out-crossing (red) populations from a standing frequency of the mutant under selection at 50%. (A) selection coefficient  $s$  equal 0.3. (B)  $s$  equal to 0.01. (C)  $s$  equal to 0.005.  $mf$  parameter set to 1.5.

**Supplementary Figure 4.** Selection of adaptive variants in a population with gene flow from adjacent populations. (A) Frequency of adaptive variant alleles under selection over time under a selection coefficient of 0.01, for given numbers of loci ( $g$ ), levels of gene flow ( $gf$ ) and mating system ( $m$ ), either inbreeding ( $i$ ) or out-crossing ( $o$ ). (B) Population sizes over time for the same simulations as in A.

**Supplementary Figure 5.** Selection of standing variation in a population with gene flow from adjacent populations

without high standing variation. (A) Frequency of adaptive variant alleles under selection over time under a selection coefficient of 0.01, for given numbers of loci ( $g$ ), levels of gene flow ( $gf$ ) and mating system ( $m$ ), either inbreeding ( $i$ ) or out-crossing ( $o$ ). (B) Population sizes over time for the same simulations as in A.

**Supplementary Figure 6.** Selection of standing variation in a population with gene flow from adjacent populations with high standing variation. (A) Frequency of adaptive variant alleles under selection over time under a selection coefficient of 0.01, for given numbers of loci ( $g$ ), levels of gene flow ( $gf$ ) and mating system ( $m$ ), either inbreeding ( $i$ ) or out-crossing ( $o$ ). (B) Population sizes over time for the same simulations as in A.

**Supplementary Figure 7.** Selection of adaptive variants to the same strong pressure under the threshold selection model. Five independent simulations under the same conditions of 5 loci, each subject to a selection coefficient  $s$  of 0.5. Frequencies of different adaptive variants are assigned different colours arbitrarily.

**Supplementary Figure 8.** Selection of adaptive variants to the same weak pressure under the threshold selection model. Five independent simulations under the same conditions of 40 loci, each subject to a selection coefficient of 0.01. Frequencies of different adaptive variants are assigned different colours arbitrarily. For clarity, all reseedling events after 100 generations are removed, showing that in most cases all but one or two adaptive variants will fail to be selected.

**Supplementary Figure 9.** Selection of adaptive variants to multiple pressures under the mixed selection model. Five independent simulations under the same conditions of 5 selection pressures, each associated with 4 loci hosting adaptive variants. All 20 loci subject to a selection coefficient of 0.01. Frequencies of different adaptive variants are assigned different colours arbitrarily.

## REFERENCES

1. Darwin C. Variation under domestication. *Origin of species*. London: Murray; 1859:7–43.
2. Harlan JR, de Wet JM, Price EG. Comparative evolution of cereals. *Evolution*. 1973;27:311–25.
3. Hammer K. Das domestikations syndrom. *Kulturpflanze*. 1984;32:11–34.
4. Fuller DQ. Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the old world. *Ann Bot*. 2007;100:903–9.
5. Brown TA, Jones MK, Powell W, et al. The complex origins of domesticated crops in the fertile crescent. *Trends Ecol Evol*. 2009;24:103–9.
6. Honne BJ, Heun M. On the domestication genetics of self fertilizing plants. *Veg Hist Archaeobot*. 2009;18:269–72.
7. Purugganan MD, Fuller DQ. The nature of selection during plant domestication. *Nature*. 2009;457:843–8.
8. Allaby RG, Brown T, Fuller DQ. A simulation of the effect of inbreeding on crop domestication genetics with comments on the integration of archaeobotany and genetics: a reply to Honne and Heun. *Veg Hist Archaeobot*. 2010;19:151–8.
9. Abbo S, Lev-Yadun S, Gopher A. Agricultural origins: centres and non-centres: a Near Eastern reappraisal. *CRC Crit Rev Plant Sci*. 2010;29:317–28.
10. Abbo S, Lev-Yadun S, Gopher A. Origin of Near Eastern plant domestication: homage to Claude Levi-Strauss and 'La Pensée Sauvage'. *Genet Resour Crop Evol*. 2011;58:175–9.
11. Abbo S, Lev-Yadun S, Gopher A. Plant domestication and crop evolution in the Near East: on events and processes. *CRC Crit Rev Plant Sci*. 2012;31(3):241–57.

12. Fuller DQ, Willcox G, Allaby RG. Cultivation and domestication had multiple origins: arguments against the core area hypothesis for the origins of agriculture in the Near East. *World Archaeol.* 2011;43:628–52.
13. Fuller DQ, Denham T, Arroyo-Kalin M, et al. Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proc Natl Acad Sci U S A.* 2014;111(17):6147–52.
14. Hillman GC, Davies MS. Domestication rates in wild-type wheats and barley under primitive cultivation. *Biol J Linn Soc.* 1990;39:39–78.
15. Diamond J. Evolution, consequences and future of plant and animal domestication. *Nature.* 2002;418:700–7.
16. Zohary D, Hopf M. *Domestication of Plants in the Old World.* 3rd ed. Oxford: Oxford University Press; 2000.
17. Asouti E, Fuller DQ. A contextual approach to the emergence of agriculture in Southwest Asia. *Curr Anthropol.* 2013;54(3):299–345.
18. Tanno K-I, Willcox G. How fast was wild wheat domesticated? *Science.* 2006;311:1886.
19. Tanno K-I, Willcox G. Distinguishing wild and domestic wheat and barley spikelets from early Holocene sites in the Near East. *Veg Hist Archaeobot.* 2012;21:107–15.
20. Weiss E, Kislev ME, Hartmann A. Autonomous cultivation before domestication. *Science.* 2006;312:1608–10.
21. Willcox G. The distribution, natural habitats and the availability of wild cereals in relation to their domestication in the Near East: multiple events, multiple centres. *Veg Hist Archaeobot.* 2005;14:534–41.
22. Willcox G, Fornite S, Herveux LH. Early Holocene cultivation before domestication in northern Syria. *Veg Hist Archaeobot.* 2008;17:313–25.
23. Willcox G, Stordeur D. Large-scale cereal processing before domestication during the tenth millennium cal BC in northern Syria. *Antiquity.* 2012;86:99–114.
24. Hillman GC, Hedges R, Moore AMT, et al. New evidence of late glacial cereal cultivation at Abu Hureyra on the Euphrates. *Holocene.* 2001;11:383–93.
25. Purugganan MD, Fuller DQ. Archaeological data reveal slow rates of evolution during plant domestication. *Evolution.* 2011;65:171–83.
26. Allaby RG, Fuller DQ, Brown TA. The genetic expectations of a protracted model for the origins of domesticated crops. *Proc Natl Acad Sci U S A.* 2008;105:13982–6.
27. Allaby RG. Integrating the processes in the evolutionary system of domestication. *J Exp Bot.* 2010;61:935–44.
28. Haldane JBS. More precise expressions for the cost of natural selection. *J Genet.* 1960;57:351–60.
29. Flake RH, Grant V. An analysis of the cost-of-selection concept. *Proc Natl Acad Sci U S A.* 1974;71(9):3716–20.
30. Darlington PJ. The cost of evolution and the imprecision of adaptation. *Proc Natl Acad Sci U S A.* 1977;74(4):1647–51.
31. Grant V. *Organismic Evolution.* San Francisco, CA: W. H. Freeman and Co; 1977: 133.
32. Haldane JBS. The cost of selection. *J Genet.* 1957;55:511–24.
33. Gomulkiewicz R, Holt RD. When does evolution by natural selection prevent extinction? *Evolution.* 1995;49:201–7.
34. Maynard-Smith J. “Haldanes dilemma” and the rate of evolution. *Nature.* 1968;219:1114–6.
35. Barton N, Partridge L. Limits to natural selection. *Bioessays.* 2000;22:1075–84.
36. Orr HA, Unckless RL. Population extinction and the genetics of adaptation. *Am Nat.* 2008;172(2):160–9.
37. Chevin LM, Lande R. When do adaptive plasticity and genetic evolution prevent extinction of a density-regulated population? *Evolution.* 2010;64(4):1143–50.
38. Fuller DQ, Allaby RG. Seed dispersal and crop domestication: shattering, germination and seasonality in evolution under cultivation in fruit development and seed dispersal. *Annu Plant Revol.* 2009;38:238–95.
39. Peleg Z, Fahima T, Korol AB, et al. The genetic basis of wheat domestication and evolution under domestication. *J Exp Bot.* 2011;62:5051–61.
40. Peng J, Ronin Y, Fahima T, et al. Domestication quantitative loci in *Triticum dicoccoides*, the progenitor of wheat. *Proc Natl Acad Sci U S A.* 2003;100:2489–94.
41. Wright SI, Bi IV, Schroeder SG, et al. The effects of artificial selection on the maize genome. *Science.* 2005;308:1310–4.
42. Chapman MA, Pashley CH, Wenzler J, et al. A genomic scan for selection reveals candidates for genes involved in the evolution of cultivated sunflower (*Helianthus annuus*). *Plant Cell.* 2008;20:2931–45.
43. Yi X, Liang Y, Huerta-Sanchez E, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science.* 2010;329:75–8.
44. Gupta P, Rustgi S, Kumar N. Genetic and molecular basis of grain size and grain number and its relevance to grain productivity in higher plants. *Genome.* 2006;49:565–71.
45. Konishi S, Izawa T, Lin SY, et al. A SNP caused loss of seed shattering during rice domestication. *Science.* 2006;312:1392–6.
46. Li C, Zhou A, Sang T. Rice domestication by reducing shattering. *Science.* 2006;311:1936–9.
47. Li W, Gill BS. Multiple genetic pathways for seed shattering in the grasses. *Funct Integr Genomics.* 2006;6:300–9.
48. Takahashi R. The origin and evolution of cultivated barley. In: Demerec M, ed. *Advances in Genetics 7.* New York, NY: Academic Press; 1955:227–66.
49. Ishikawa R, Thanh PT, Nimura N, et al. Allelic interaction at seed-shattering loci in the genetic backgrounds of wild and cultivated rice species. *Gene Genet Syst.* 2010;85:265–71.
50. Ishii T, Numaguchi K, Miura K, et al. *OsLG1* regulates a closed panicle trait in domesticated rice. *Nat Genet.* 2013;45:462–5.
51. Fuller DQ, Allaby RG, Stevens C. Domestication as innovation: the entanglement of techniques, technology and chance in the domestication of cereal crops. *World Archaeol.* 2010;42(1):13–28.
52. Meyer RS, Purugganan MD. Evolution of crop species: genetics of domestication and diversification. *Nat Rev Genet.* 2013;14:840–52.
53. Gepts P. Crop domestication as a long term selection experiment. *Plant Breed Rev.* 2004;24:1–44.
54. Bomblied K, Doebley JF. Pleiotropic effects of the duplicate maize *FLORICAULA/LEAFY* genes *zfl1* and *zfl2* on traits under selection during maize domestication. *Genetics.* 2006;172:519–31.
55. Zhu Q, Zheng X, Luo J, et al. Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Mol Biol Evol.* 2007;24:875–88.
56. Eyre-Walker A, Gaut B. Investigation of the bottleneck leading to the domestication of maize. *Proc Natl Acad Sci U S A.* 1998;95:4441–6.
57. Haldane JBS. *A mathematical theory of natural and artificial selection. Part I.* Vol 23. Transactions of the Cambridge Philosophical Society, Cambridge; 1924:3–41.
58. Innan H, Kim Y. Pattern of polymorphism after strong artificial selection. *Proc Natl Acad Sci U S A.* 2004;101:10667–72.
59. Vogler DW, Kalisz S. Sex among the flowers: the distribution of plant mating systems. *Evolution.* 2001;55:202–4.
60. Weber A, Clark RM, Vaughn L, et al. Major regulatory genes in maize contribute to standing variation in teosinte (*Zea mays* ssp. *parviglumis*). *Genetics.* 2007;177:2349–59.
61. Studer A, Zhao Q, Ross-Ibarra J, et al. Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat Genet.* 2011;43:1160–5.
62. Teshima KM, Coop G, Przeworski M. How reliable are empirical genomic scans for selective sweeps. *Genome Res.* 2007;16:702–16.
63. Larson G, Piperno DR, Allaby RG, et al. Current perspectives and the future of domestication studies. *Proc Natl Acad Sci U S A.* 2014;111:6139–46.
64. Allaby RG, Kistler L, Gutaker RM, et al. Archaeogenomic insights into the adaptation of plants to the human environment: pushing plant–hominin co-evolution back to the Pliocene. *J Hum Evol.* 2015;79:150–7.
65. Timpson A, Colledge S, Crema E, et al. Reconstructing regional population fluctuations in the European Neolithic using radiocarbon dates: a new case-study using an improved method. *J Archaeol Sci.* 2014;52:549–57.
66. Shennan S, Downey SS, Timpson A, et al. Regional population collapse followed initial agricultural booms in mid-Holocene Europe. *Nat Commun.* 2013;4:2486.
67. Stephens CJ, Fuller DQ. Did Neolithic farming fail? The case for a Bronze Age agricultural revolution in the British Isles. *Antiquity.* 2012;86:707–22.
68. Kitchen JL, Allaby RG. Systems modeling at multiple levels of regulation: linking systems and genetic networks to spatially explicit plant populations. *Plants.* 2013;2(1):16–49.