A novel cost framework reveals evidence for competitive selection in the evolution of complex traits during plant domestication

Robin G Allaby[1*], Chris J Stevens[2,3], Dorian Q Fuller[2,4]

1.School of Life Sciences, University of Warwick, Coventry, UK.
2.Institute of Archaeology, UCL, London
3.School of Archaeology and Museology, Peking University, Beijing, China
4.School of Cultural Heritage, Northwest University, Xi'an, Shaanxi, China

*email r.g.allaby@warwick.ac.uk

Abstract

Most models of selection incorporate some notion of an environmental degradation in which the majority of the population becomes less fit with respect to a character resulting in a pressure to adapt. Such models have been variously associated with an adaptation cost, the substitution load. Conversely, adaptative mutations that represent an improvement in fitness in the absence of environmental change have generally been assumed to be associated with negligible cost. However, such adaptations could represent a competitive advantage that diminishes resource availability for others and so induce a cost. This type of adaptation in the form of seedling competition has been suggested as a mechanism for increases in seed size during domestication. Here we present a novel cost framework for competitive selection that demonstrates significant differences in behaviour to environmental based selection in typical initial selection intensity and intensity over time. We show that selection intensity over time in grain size metrics of nine archaeological crops increases in one to several episodes fitting closely to the competitive selection model of single large effect alleles, but surprisingly in direct contrast to the expectations of the standard model of stabilizing selection. While size trait changes ultimately slow down in crops over time as expected from pleiotropic constraints expressed in the standard model, the mechanism outlined here shows possible complexities within the environmental based mode of shifting optimums in the standard model and a fundamental insight into the factors driving domestication.

Significance statement

We present here a new model framework for selection based on direct competition between individuals rather than the more conventional approach of individual's fitness being measured against an environmental gradient. The model explains patterns of increasing selection intensity seen in archaeological grain sizes of nine domesticated crops that otherwise contradict the expectations of shifting stabilising selection of complex traits. We show that grain size increases seen across domesticated crops are consistent with spontaneous competition between seedlings under cultivation for resources and so reveal a fundamental insight into the mechanism of plant adaptation to the human environment.

Natural selection that gives rise to the differential reproductive success of competing individuals of varying fitness leads to the notion first pointed out by Haldane (1957)[1] of a cost in terms of a declining population size during adaptation. Such costs lead to theoretical selection limits which have been extensively debated in terms of their severity, depending in part on pleiotropic effects between alleles, alternative adaptive solutions and modes selection such as truncation selection in selective breeding[2-4]. Regardless of the incorporation or not of cost penalties, most models of selection work on a similar basis to that originally described by Haldane in which individuals of a population are universally disadvantaged in regard to some parameter, from which fitter individuals evolve. In other words a universal genetic load is bestowed on the population because of some change in their environment. A second category of criticism to the original Haldane model was that such universal genetic load may not represent the majority of adaptive evolution. Instead, selection may favour 'advantageous' alleles in the absence of any environmental degradation or punishment of the wild type[5,6], and could represent improvements such as in size[7]. Such selective episodes were thought unlikely to carry a cost largely because of the absence of a genetic load[5-8], if the population were permitted to expand. However, if a population has reached the current carrying capacity of its environment, as is expected under a niche concept, and therefore has no capacity to expand, then a slight cost is expected to be shared among the wild type population as a competitive pressure is applied from the adapted phenotype. Intuitively, it is implied that as the adapted phenotype increases in frequency, the relative disadvantage experienced by the wild type population will increase and an intensification of selection is predicted. Conceptually, this contrasts the environmental degradation model of selection, as rather than a sudden appearance of a large genetic load across the population determined by some environmental factor a small genetic load is generated by the appearance of an

innovative mutation, which then will grow in effect over time as the mutation gains frequency.

This category of selection may apply, for instance, in the case of increased seed size during plant domestication where it has been argued that seedling competition may be responsible selecting for more robust early growth[9,10]. An increased ability for resource acquisition, in this case nutrient and light, suggests a cost to the remaining population when resource is finite. Seed size is a complex trait influenced by many loci of varying effect which are well described under the standard stabilizing selection model[11]. This latter represents a category of environmental selection models in which the strength of selection decreases as the optimum is approached and pleiotropic constraints apply. While much is understood about the potential interactions of alleles of varying effect from numerous different loci in the process of adaptation through simulations[12-15], here we consider the fate of a single allele of varying effect on a competitive trait advantage and then attempt to place that within the wider context of complex trait adaptation.

Results

Basic model and rationale. In this model we assume that fitness is associated with a probability of acquiring resource. Broadly, this may be any organism that first encounters a resource and then does, or does not, acquire it. We assume that individuals have equal encounter probabilities but differ principally in their ability to acquire the resource once encountered. The probability of an adaptive mutant and wild type acquiring a resource on encounter are given is p_m and p_w respectively. Consequently, the probability of a resource being acquired ($P(aq)$) on encounter can be given as:

$$P(aq) = \left(\frac{N_m}{N} p_m\right) + \left(\frac{N_w}{N} p_w\right) \qquad (1)$$

Where *Nm, Nw* and *N* are the number of mutant phenotypes, wild types and total population respectively. If we assume that the resource is finite and is exploited completely, then the expected number of encounters, or trials (*tr*), required for each unit of resource (*r*) can be given as:

$$tr/r = 1/P(aq) \qquad (2)$$

The parameter *tr* will follow a Poisson distribution, but here we take the λ value as a simplifying adequate estimate for the purposes of the model. Consequently, the expected resource (*E(m)*) a single mutant phenotype will obtain can be given as:

$$E(m) = \frac{1}{N} p_m \frac{tr}{r} r \qquad (3)$$

We can measure the selective advantage of a mutant by considering the increased resource acquisition relative to the equal share that would be received expected under no advantage. The excess, or advantage (*a*) that the mutant will obtain relative to the wild type can be given as the increased proportion of resource relative to equal share:

$$a = \frac{E(m) - \frac{r}{N}}{\frac{r}{N}} \qquad (4)$$

The remaining wild type population will be relatively disadvantaged by an amount that is shared equally amongst them, which will reduce the fitness of the wild type relative to the adaptive mutant. If we take an equal share of the resource to represent the sufficient amount required for survival, then the reduced proportion represents the probability of obtaining sufficient resource. Under this assumption we can take the reduction in probability of sufficient require acquisition as the selection coefficient (*s*). Assuming the expected resource acquired for each mutant given in equation (3) will remain constant the effect of increasing numbers of adaptive mutant will have an additive effect on the remaining wild type population which share the burden of reduced resource such that the selection coefficient experienced by any single wild type individual can be expressed as:

$$s = \frac{a \times Nm}{Nw} \quad (5)$$

Equation (5) expanded using equations (1-4) and simplified, with *r* assigned a value of 1, express the selection coefficient as:

$$s = \left(\frac{Np_m}{N_m p_m + N_w p_w} - 1\right)\frac{N_m}{N_w} \quad (6)$$

Equation (6) is validated by an agent-based simulation in which mutant and wild type agents are assigned respective values of p_m and p_w, and the resulting proportional loss of resource relative to an equal share measured, Fig S1. This validation demonstrates that the potential effects of increasing competition between mutants as their frequency increases does not change the outcome. We can therefore apply the rules of equation (6) to ascribe idealized selection coefficients within a competitive selection model.

A remaining apparent problem is that we do not necessarily know the values of the variables p_m and p_w. Equation (6) can be rearranged for the ratio of p_m and p_w for a single mutant:

$$\frac{p_m}{p_w} = \frac{N-1}{\frac{N}{s(N-1)+1} - 1} \quad (7)$$

Given that the intrinsic success probabilities on encounter are constant, the ratio holds true for any number of N_m and N_w, and consequently the ratio itself is dependent only on the population size and the selection coefficient. A corollary is that any two probabilities for p_m and p_w that satisfy the ratio defined by *s* and *N* can be used to calculate *s* in equation (6). Therefore, any probability can be assigned to one variable (p_m or p_w) and equation (7) can be used to calculate the corresponding variable and we do not need to know the true values of p_m and p_w to find *s*, only a pair of values that satisfy equation (6) for a particular value of *s*.

Intensity of competitive selection. The additive nature of the strength of competitive selection as mutants increase in frequency as described in equation (6) suggests that the initial selection coefficients involved may be very small and considerably weaker than genetic drift. In such cases selection could not effectively act until mutant allele frequencies had been elevated sufficiently through drift processes. The interplay between drift and selection is largely dependent on the parameter δp, the expected change in frequency of an adaptive mutant due to selection over a single generation. The value of δp, the change in frequency of allele p under selection for a given level of inbreeding derived from the mating system (F) is given as:

$$\delta p = \frac{p^2+pqF+0.5(2pq-2pqF)(1-h)}{(1-s(q^2+pqF)-h(2pq-2pqF)} - p \qquad (8)$$

Where h is the selection coefficient against the heterozygote, and s can be calculated from equation (6) after determining the number of advantageous phenotype individuals (N_m) for a given value of p in the case of dominant alleles using:

$$N_m = N - N(q^2(1-F) + qF) \qquad (9)$$

And in the case of recessive advantageous alleles:

$$N_m = N(p^2(1-F) + pF) \qquad (10)$$

In this model for simplicity we consider only completely dominant or recessive alleles. When δp is less than 1/2N then no increment in p is expected in terms of individuals in the population, and the behaviour of p will be close to neutral. Consequently, at frequencies of p at which δp exceeds 1/2N, it is expected that there should be a shift from more drift-like behaviour to more selection like behaviour to an extent that is determined by the probability that $p'>p$, where p' is the value p becomes in the subsequent generation. Simulations under the competitive model broadly support this threshold behaviour showing a biphasic process in which advantageous mutants wander in a drift like fashion at lower frequencies, but switch into strong selection with increasing frequency, Fig S2. Under an environmental selection

188 model, it is well known that intensities (Ns) >> 1 are required for selection to overcome drift.
189 However, despite an initial section intensity of only 0.5 competitive selection is still
190 effective, and remains so at initial intensities ten-fold lower, Fig S3. Similar biphasic
191 behaviour is of course expected under the standard environmental selection model, although
192 at such low intensity the influence of drift remains great because selection does not intensify
193 with increasing frequency of adaptive mutants, Fig S4.

194 Once an adaptive mutant has broken through a drift threshold, the strength of selection is
195 determined by the fitness effect. While in the case of environmental selection the fitness
196 effect is described by the selection coefficient s, in competitive selection the fitness effect is
197 described by the p_w/p_m ratio but s is a function of the ratio and the frequency of the mutant.
198 Unlike s, the p_w/p_m ratio has the property of being independent of N for a given selection
199 intensity (Ns), Fig S5. Therefore, a single p_w/p_m ratio describes all population sizes
200 undergoing competitive selection for a particular intensity. Consequently, a probable fate of a
201 competitive mutant can be described as an interplay between the p_w/p_m ratio and the neutral
202 probability of reaching threshold frequencies at which selection is sufficient to increase
203 mutant frequencies by an individual or more each generation, Figure S6. This demonstrates
204 that in order for a high probability of fixation p_m must be considerably higher than p_w to
205 achieve the necessary intensity of selection, and the larger the population size, the more
206 pronounced must be the difference between the p_m and p_w probabilities. An unsurprising
207 corollary is that competitive mutants are generally more likely to be successful in small
208 populations.

209 The likelihood of various ratios of p_w and p_m occurring is expected to be described in
210 the distribution of fitness effects, which are well described by an exponential distribution[16-18].

211 The probability of the magnitude in difference between p_w and p_m can therefore take the
212 form:

213 $$p\left(\frac{p_w}{p_m}\right) = e^{-\lambda\frac{p_m}{p_w}} \quad (11)$$

214 In which the parameter λ takes a value of 1 as a general solution[18]. Using equation 2 of
215 Orr (2003)[18] for the fitness effect distribution of *s* for environmental selection, we compared
216 the profiles of mutants of various effects surpassing drift threshold occurring a rate of $2N\mu$,
217 accounting for a decrease in the probability of population survival[19] for values of s > 0.5, Fig.
218 1. Our first major finding is a clear separation of the most likely initial selection intensities
219 under competitive versus environmental based selection. Whereas we find that environmental
220 based selection is expected to occur at intensities >> 1 as expected, and the optimal value for
221 *s* falls around 0.05 in agreement with estimates from variance analysis based on empirical
222 data[20], competitive selection intensities are most likely in this model at values around 0.5.
223 These results suggest that the most likely p_w/p_m values to successfully undergo selection are
224 around 0.75 or that mutants that are 50% more likely to acquire a resource than the wild type
225 are most likely to occur. This suggests alleles of typically large effect are most likely to
226 undergo selection in this way, although effects could commonly range from 1% to 200% in
227 increased likelihood of resource acquisition.
228
229 **Evidence of competitive selection in grain size evolution during domestication.** We used
230 the framework described here to investigate the patterns of selection intensity over time
231 observed in size traits of grains in plant domestication. Using the archaeometric
232 measurements of length, breadth and thickness of grains from nine plant species over time[21],
233 we calculated selection coefficients using Haldanes based on shifting means and standard
234 deviations over time for a given trait heritability[22], green diamonds in Fig 2. Previously, we

have found the Haldane metric approach has yielded estimates which closely match selection coefficients calculated from a Hardy-Weinberg model for shattering traits[23], suggesting it is a robust approach for calculating selection strength from this data type. A number of trends of increasing selection coefficients are apparent within the data over time in all nine species, often associated with the initiation of increases in trait dimensions. In some cases upticks in selection intensity appeared to be followed by downward trends, as in the cases of breadth in barley, emmer thickness, length in lentil, pea, soybean and sunflower. To investigate whether the upward trends could be explained by episodes of competitive selection we found best fitting theoretical profiles of s by iterative regression for s and N variables (see methods) over time for selected bins of selection coefficients that displayed a trend from low to high value ranges. We found 28 instances of upticks in empirical selection intensity that correlated with modelled competitive selection episodes, over 50% these with r values > 0.9 and over 85% with r values > 0.75, grey lines in Fig 2 and Table S1. Over half the correlations showed a significantly better fit than could be achieved from fits to random upward trends (p_{null}, Table S1), and 80% of trends that showed a strong correlation with competitive selection ($r > 0.65$) had a probability of 0.1 or less of achieving as good a fit with random trends. We therefore conclude that the majority of curves identified as a good fit to competitive selection are not over fitted, while in a few cases such as sunflower length the fit could be spurious. The latter may be resolved in the future with more archaeological data points.

The close correlation between the selection dynamics of complex size traits and the competitive model which considers the fate of single adaptive alleles, suggests that in most cases during the intensification of selection, single alleles are initially largely responsible for shifts in size traits. In some cases concurrent competitive episodes were identified for different traits in the same species, such as in the case of barley breadth and thickness. It is possible that multiple alleles hitchhiked together each affecting one trait, or that a single

allele of large effect was pleiotropic in its trait effects. In other cases, such as the first two episodes in both sunflower breadth and length appear more staggered, suggesting a complex sequence of episodes. The frequency of *p* at the height of the competitive episode trend (*p* height, Table S1) was generally found to be quite low, usually between 1-5%, Table S1. It is unlikely this represents the final frequency that a putative adaptive allele reached. Since the log phase of frequency increase is short under this model at these intensities, typically less than 100 generations between 0.02 and 0.95, the density of archaeological sampling is unlikely to recover the latter stages of a fixation episode. We cannot therefore conclude from these data whether these signals are associated with allele fixations under selection.

In most cases a model is supported of multiple competitive episodes for each crop. The fitted selection intensities on archaeometric data match remarkably closely to the theoretical expected intensities described in the model, Fig. 3, indicating adaptive advantages that range from a 0.9-116% improvement in resource acquisition with a mode of 65%. Given the close correlations of the pattern of selection intensities to the empirical archaeometric data, and the close match to the expected theoretical intensities, we conclude that these nine crop species most likely underwent competitive selection episodes which appear to have been associated with the initiation of increases in trait dimensions. The subsequent episodes of downward trends in selection intensity after competitive episodes may represent the more complex activity of numerous loci that are more recalcitrant to model.

Discussion

Complex traits such as size are well known to be under stabilizing selection, which has a well established model framework[11]. A prediction of this model is that as traits approach their optimal value the selection pressure becomes less, as selection strength is a function of the difference between the trait value and the trait optimum. In part, this system, is derived from

285 the observation that under selective breeding the rate of crop trait change diminishes over
286 time as improvements become constrained by the pleiotropic effects of other traits[20,24].
287 Consequently, the observation of increasing selection intensity over time of complex size
288 traits from archaeometric data appear anomalous to stabilizing selection as it is currently
289 understood.

290 Cultivation led in almost all cases of crops to an increase in grain size dimensions over
291 several millenia[21]. The new environment of cultivation may have heralded a new set of
292 optimums for trait dimensions which encouraged a shift in size. In particular it is argued that
293 by provision of stable resources through cultivation the optimal strategy for plants shifted
294 from nutrient conservation to nutrient acquisition[9,25]. Models of stabilizing selection suggest
295 that it is alleles of small effect which are most likely to respond first to shifts in optima, partly
296 because they are expected to occur at intermediate frequencies rather than either rare or fixed
297 in the case of large effect alleles[26]. Furthermore, selective sweeps are expected to be rare[15],
298 and where alleles of large effect are involved, small effect alleles are likely to replace them in
299 the course of adaptation[13]. Under these circumstances, we would expect trait adjustment to
300 follow a pattern of decelerating selection pressure. The pattern we observe in the
301 archaeological record of increasing intensity repeatedly in many crops is suggestive that an
302 alternative process to the standard model of stabilizing selection is involved.

303 If a new trait optimum is greatly removed from the current trait mean, outside the range of
304 trait effects in the current genetic variation, it may be that large jumps may not be within
305 evolutionary reach[14]. Escape from such local optima traps can be achieved through new
306 mutations of large effect which can launch the phenotype to a new part of the adaptive
307 landscape[14]. The patterns here suggest that such leaps may have occurred repeatedly across
308 crops which had been moved to a cultivated environment, initiated by a mechanism of
309 competitive selection. This fits the notion of alleles of large effect appearing which enabled

crops to switch to capitalizing on resource availability through larger grain size, which in turn could have led to robust growth and seedling competition, for instance shading wild type plants. The fit with the competitive model suggests initially simple single gene driven changes. It may be the case that such alleles became fixed or numerous other loci of lesser effect could have come into play in later stages[13]. However, the latter seems less likely if the genetic variation had been previously insufficient and trait values had been stuck on local optima. After competitive episodes, trait values appear to slow down in their rate of increase, and in some instances we see reducing trends in selection intensities, which appears to be reflective of a standard stabilizing selection model as traits appear to reach pleiotropically constrained new optima.

The cost consequences of adaptive mutations in the absence of environmental deterioration have largely been overlooked since the conception of the substitution load. As was initially suggested[5], we find that the initial impact on the population is very low, but the cost consequence for the wild type becomes significant with time and its properties as described here are detectable in time series data from which selection coefficients can be derived. Our study suggests that in crop domestication competitive selection was a widespread occurrence and provides a new complexity to modelling the evolution of complex traits. It remains to be seen how widespread such selection might have been in wild populations.

Materials and Methods

Model verification The behaviour of parameters as described in equation (6) was verified using a simple agent based simulation in which individuals were randomly selected and then deemed successful in resource acquisition with a probability of p_m or p_w depending on

335 whether they were wild type or mutant individuals. Sampling continued until all resource had
336 been allocated. Values of 0.21 and 0.2 were assigned to p_m and p_w respectively in a
337 population size of 100 individuals with resource size 1000 and 10,000 replications. Sampling
338 was repeated for increasing numbers of mutants from 1-99. The perl script used for this
339 (benign_profile_of_s.pl) is available from
340 (https://warwick.ac.uk/fac/sci/lifesci/research/archaeobotany/downloads/competitive_selectio
341 n).
342 ***Competitive selection simulations*** A previously described agent-based simulator[19] was
343 developed to include a competitive selection regime using the equations presented here. The
344 modification was that each generation frequency dependent selection coefficients were
345 calculated for a given number of advantageous phenotypes using equation 6, deriving values
346 of p_m or p_w using equation 7. The simulator is available from
347 (https://warwick.ac.uk/fac/sci/lifesci/research/archaeobotany/downloads/competitive_selectio
348 n).
349 ***Threshold frequency determination*** Probabilities of reaching threshold frequencies through
350 drift were calculated by first determining the number of individuals at which $\delta p > 1/2N$ using
351 equations 8-10 and equation 6 to determine the frequency dependent values of s. The neutral
352 probability of an allele reaching that frequency is given as $1/f_{(threshold)}$., where $f_{(threshold)}$ is the
353 frequency of p at which $\delta p > 1/2N$. Similar thresholds were calculated for environmental
354 selection in which s is constant rather than frequency dependent.
355 ***Likelihood profiles of selection intensity*** Likelihoods were calculated as the product of
356 probabilities of a p_w/p_m ratio for an associated selection intensity under an exponential
357 distribution model of parameter $\lambda=1$ (equation 11), the neutral probability of mutant
358 frequencies becoming sufficiently high such that $\delta p>1/2N$, calculated as in figure S6, the
359 probability of population survival for a given magnitude of s as previously calculated[19], and

360 the rate of mutation generation 2Nμ. Mutation rate is notionally included here as μ=0.5 for
361 convenience of calculation and display.

362 ***Calculation of s from archaeometric data*** Selection coefficients were determined by first
363 calculating haldanes using the method of Kinnison and Hendry (2001)[22] in which:

364 $$haldanes = \frac{\left(\frac{x_2}{s_p}\right) - \left(\frac{x_1}{s_p}\right)}{g}$$

365 where x is the mean trait value at time 1 and 2, s_p is the pooled standard deviation of trait
366 values across time, and g is the number of generations which is assumed to equate to
367 radiocarbon date years in the case of annual crops. Haldanes were converted to s following

368 $$s = \frac{haldanes}{h^2}$$

369 where h^2 is the trait heritability, assumed here to be approximately 0.9 for size traits[21]. All
370 measurements were taken from the supplementary data of Fuller *et al.* (2014)[21]. Using this
371 approach, values of s were calculated for all data point pairs which were less than 1000 years
372 and more than 100 years apart in age. The exception was in the case of sunflowers, where the
373 low number of useful data points was mitigated by considering data points up to 2000 years
374 apart.

375 ***Competitive selection track generation and fitting*** Selection coefficients determined from
376 archaeometric data were binned out into categories that showed a trend of low to high
377 selection coefficients, terminating in the highest selection coefficient. Multiple bin
378 combinations were attempted, results are shown for best fits found. Theoretical selection
379 tracks were calculated for given values of initial s and N where values of p increment each
380 generation by δp as calculated in equation (8), values of s for each value of p were calculated
381 using equation (6), where values of N_m and N_w were determined using equations (9) and
382 (10).Theoretical selection tracks were then generated for a given population size, iterating
383 from a lowest value of s determined as s_{max}/4N where s_{max} is the highest selection coefficient

in the bin until the quality of the fit as determined by pmcc reached its maxima. The best fit for each track was achieved using custom scripts by sliding the track along the x axis and finding the least squares difference in y values at real data points in time. The fit was then evaluated by calculating r as a linear regression of $y_{theoretical}$ against $y_{observed}$ coordinates. To check for spurious correlation or over fitting we calculated p_{null}, the probability that observed r could occur for a random upward trend. To do this we compared r values against null distributions of r values generated as follows. Using constraints of s bounded by values of 0.0001 and 0.02 and time bounded by archaeological sample ages (t), we randomly generated data sets by first generating two values of s and t to give upper and lower s values of a random upward trend. Intervening data points were then randomly generated which were constrained by the upper and lower values of s and t initially generated. The number of data points generated matched the observed number of data points. We then fitted a competitive selection curve to the data and calculated the r value as described above. For each observed r value we generated 1000 such replicates to produce a non-parametric null r distribution. The proportion of replicate r values above the observed value gave our estimate of p_{null}.

Acknowledgements

The authors would like to thank Logan Kistler, Becky Cribdon and Hsiao-Lei Liu for useful discussions in developing this manuscript.

1.Haldane JBS. (1957) The cost of selection. *J Genet*. 55:511–24.

2.Maynard-Smith J.(1968) "Haldane's dilemma" and the rate of evolution. *Nature* 219:1114–6

3.Barton NH (1995) Linkage and the limits to natural selection. *Genetics* 140:821-841.

4.Sved, JA (1968) Possible rates of gene substitution in evolution. *Amer. Natur*. 102:283-293.

5.Brues AM (1964) The cost of evolution versus the cost of not evolving. *Evolution* 18:379-383.

6.Brues AM (1969) Genetic load and its varieties. *Science* 164:1130-1136.

7.Van Valen L (1963) Haldane's dilemma, evolutionary rates, and heterosis. *Amer. Natur.* 97:185-190.

8.Felsenstein J. (1971) On the biological significance of the cost of gene substitution. *Amer. Nat.* 105, 1-11.

9.Fuller, DQ and Stevens, CJ (2017) Open for Competition: Domesticates, Parasitic Domesticoids and the Agricultural Niche. Archaeology International 20: 110–121.

10.Fuller, D. Q. and Chris J. Stevens (2019) The making of the botanical battleground: domestication and the origins of the World's weed floras. In *Far from the Hearth: Essays in Honour of Martin K. Jones* (edited by E. Lightfoot, X. Liu, and D. Q. Fuller). Cambridge: McDonald Institute of Archaeology. Pp. 9-21.

11.Barton NH (1986) The maintenance of polygenic variation through a balance between mutation and stabilizing selection *Genet. Res. Camb.* 47:209-216.

12.Pavlidis P, Metzler D, Stephan W (2012) Selective sweeps in multilocus models of quantitative traits. *Genetics* 192:225-239.

13.Chevin L.M., Hospital F (2008) Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics* 180:1645-1660.

14.Jain K, Stephan W (2017) Rapid adaptation of a polygenic trait after a sudden environmental shift. *Genetics* 206:389-406.

15.Thornton K (2019) Polygenic adaptation to an environmental shift: temporal dynamics of variation under Gaussian stabilizing selection and additive effects on a single trait. *Genetics* 213:1513-1530.

16.Gillespie, JH (1983) A simple stochastic gene substitution model. Theor. Popul. Biol. **23:** 202–215.

17.Gillespie, JH (1984) Molecular evolution over the mutational landscape. Evolution **38:** 1116–1129.

18.Orr, HA (2003) The distribution of fitness effects among beneficial mutations. *Genetics* 163: 1519-1526.

19.Allaby RG, Kitchen JL, Fuller DQ (2015) Surprisingly low limits of selection in plant domestication. *Evolutionary Bioinformatics* 11:(S2) 41-51.

20.Johnson T, Barton N (2005) Theoretical models of selection and mutation on quantitative traits. *Phil. Trans R. Soc. B.* 360:1411-1425.

21.Fuller DQ, Denham T, Arroyo-Kalin M, Lucas L, Stephens C, Qin L, Allaby RG, Purugganan MD (2014) Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proc. Natl. Acad. Sci. U.S.A.* 111:6147-6152.

22.Kinnison MT, Hendry AP (2001) The pace of modern life II: From rates of contemporary microevolution to pattern and process. Genetica 112-113(1):145–164.

23.Allaby, R. G., Stevens, S., Lucas, L., Maeda, O., & Fuller, D. Q. (2017). Geographic mosaics and changing rates of cereal domestication. *Philosophical Transactions of the Royal Society B*, *372*, 20160429.

24.Dudley JW, Lambert RJ (2004) 100 generations of selection for oil and protein in corn. *Plant Breeding Rev.* 24:79-110.

25.Milla R, Osborne CP, Turcotte MM, Violle C (2015) Plant domestication seen through an ecological lens. *Trends Ecol. Evol.* 30(8):463-469.

26.De Vladar HP, Barton N (2014) Stability and response of polygenic traits to stabilizing selection and mutation. *Genetics* 197:749-767.

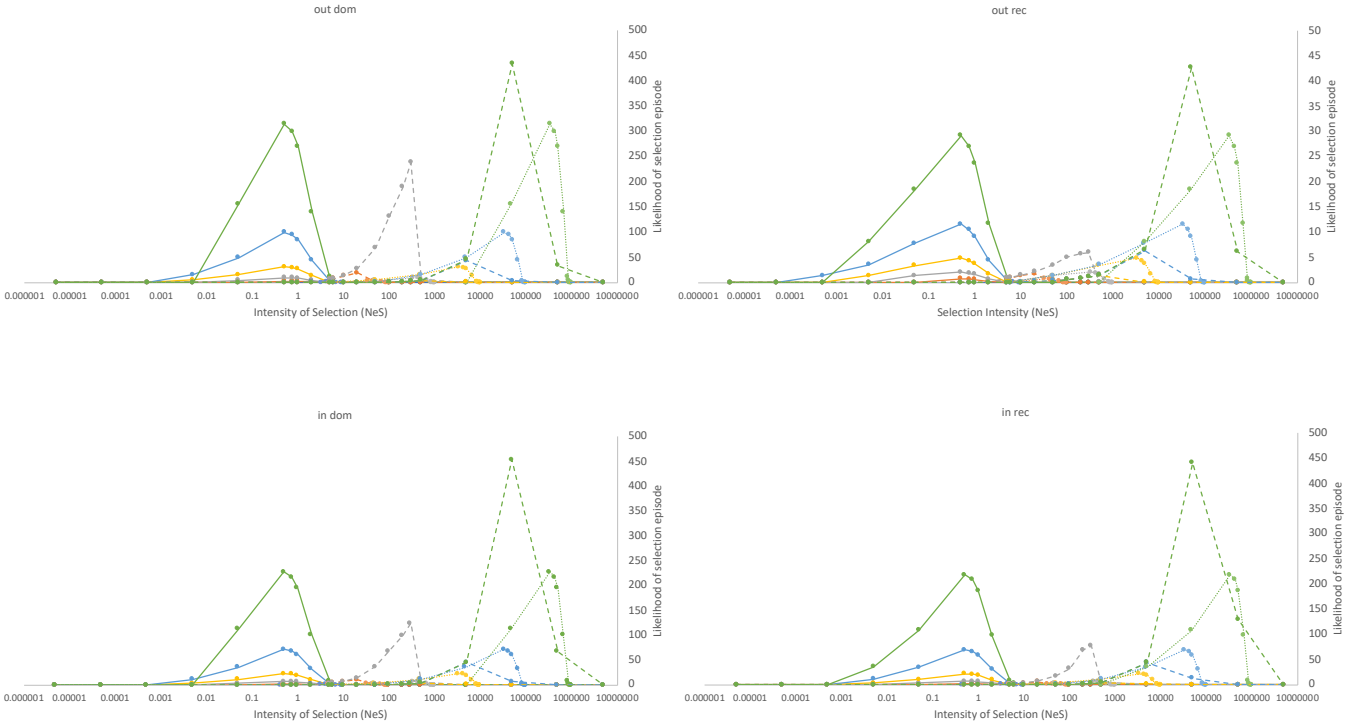

*Figure 1. Likelihood profiles of competitive and environmental selection intensity. Likelihoods were calculated as p(threshold)*p(magnitude)*2Nμ*p(survival): the product of probabilities of a p_w/p_m ratio for an associated selection intensity under an exponential distribution model of parameter $\lambda=1$, equation (11), the neutral probability of mutant frequencies becoming sufficiently high such that $\delta p>1/2N$, calculated as in figure S6, the probability of population survival for a given magnitude of s as calculated in (Allaby et al 2015), and the rate of mutation generation 2Nμ. Mutation rate is notionally included here as μ=0.5. Likelihoods were calculated for different population sizes as follows: blue (10), orange (100), grey (1000), yellow (10,000), light blue (100,000) and green (1000,000). Solid lines indicate competitive selection at initial intensities, dotted lines represent competitive selection at final intensities (p = 0.99), dashed lines represent environmental selection. For N >= 10,000, values for environmental selection are scaled for graphing visualization. Values are calculated under outbreeding (2% selfing) dominant (out dom), inbreeding (98% selfing) dominant (in dom), outbreeding recessive (out rec) and inbreeding recessive (in rec) conditions.*

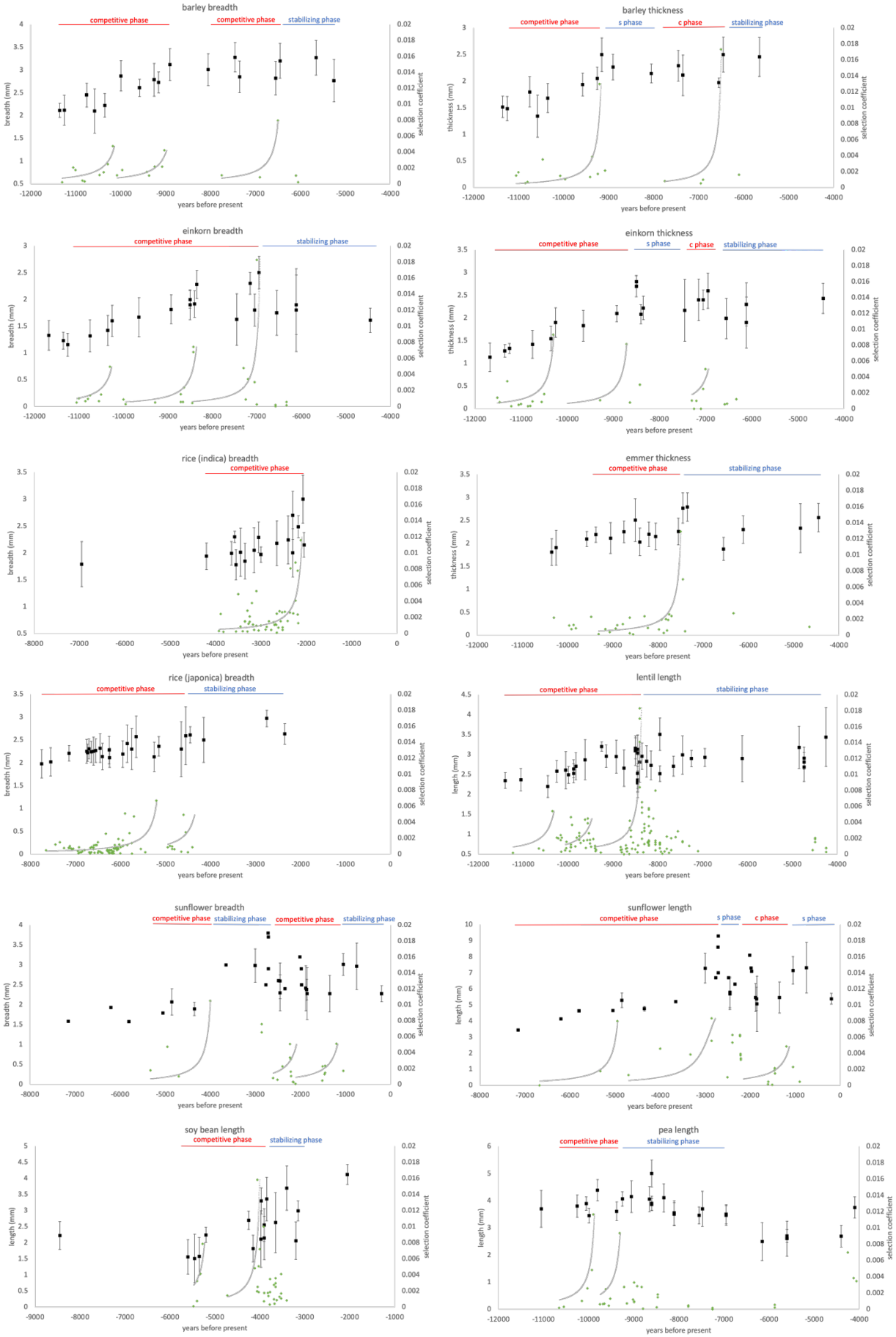
barley breadth
competitive phase
competitive phase
stabilizing phase
breadth (mm)
selection coefficient
years before present
barley thickness
competitive phase
s phase
c phase
stabilizing phase
thickness (mm)
selection coefficient
years before present
einkorn breadth
competitive phase
stabilizing phase
breadth (mm)
selection coefficient
years before present
einkorn thickness
competitive phase
s phase
c phase
stabilizing phase
thickness (mm)
selection coefficient
years before present
rice (indica) breadth
competitive phase
breadth (mm)
selection coefficient
years before present
emmer thickness
competitive phase
stabilizing phase
thickness (mm)
selection coefficient
years before present
rice (japonica) breadth
competitive phase
stabilizing phase
breadth (mm)
selection coefficient
years before present
lentil length
competitive phase
stabilizing phase
length (mm)
selection coefficient
years before present
sunflower breadth
competitive phase
stabilizing phase
competitive phase
stabilizing phase
breadth (mm)
selection coefficient
years before present
sunflower length
competitive phase
s phase
c phase
s phase
length (mm)
selection coefficient
years before present
soy bean length
competitive phase
stabilizing phase
length (mm)
selection coefficient
years before present
pea length
competitive phase
stabilizing phase
length (mm)
selection coefficient
years before present

Figure 2. Competitive selection tracks fitted to archaeobotanical grain metric data. Grain metrics shown in black with standard deviations (axis left), data from Fuller et al 2014. Selection coefficients (green diamond, axis right) were determined by first calculating haldanes using the method of Kinnison and Hendry 2001 and converting to s following s = haldanes/h^2 where h^2 is the trait heritability, assumed here to be approximately 0.9 for size traits (Fuller et al 2014). Selection coefficients were binned out from high points to low points going back in time. Theoretical profiles of selection coefficients (grey lines, axis right) were determined beginning at a strength equal to s/4N of the highest selection coefficient, see methods. Profiles were fitted to archaeometric data by least squares regression of the y coordinates, r values were determined by pmcc on y theoretical and observed coordinates.

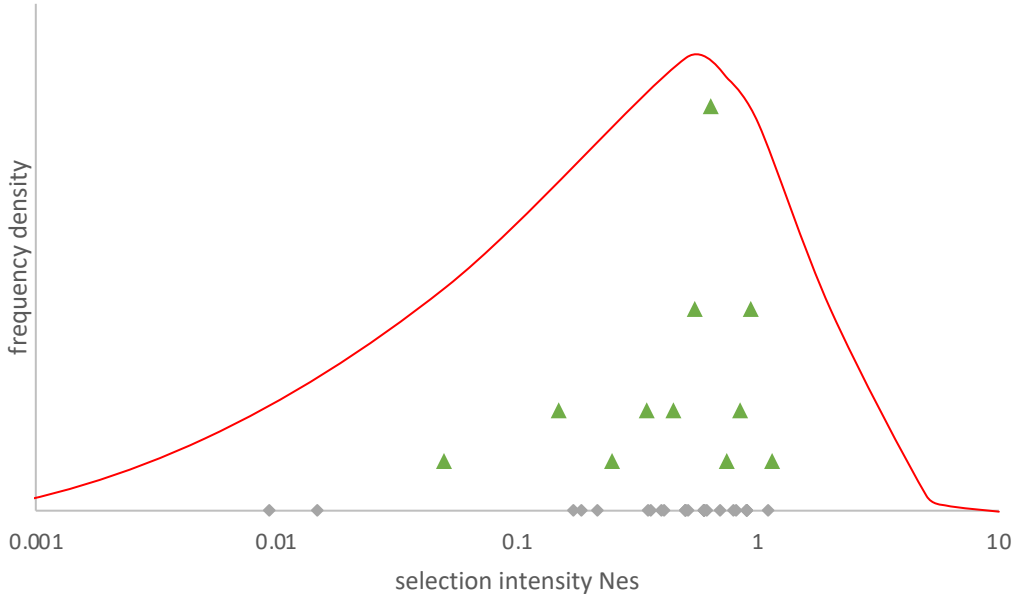

Figure 3 Best fit competitive selection intensities on archaeometric grain data and theoretical intensity likelihood profile for inbreeding dominant crops. Red line depicts the frequency density based on N=100,000 likelihood profile for inbreeding dominant plants as calculated in Figure 2 (note that inbreeding recessive, out breeding dominant and recessive do not differ greatly when scaled). Grey diamonds indicate selection intensity values determined in Table S1, green triangles indicate interval frequencies of determined selection intensities from Table S1.

Supplementary Information

A novel cost framework reveals competitive selection in the evolution of complex traits during plant domestication

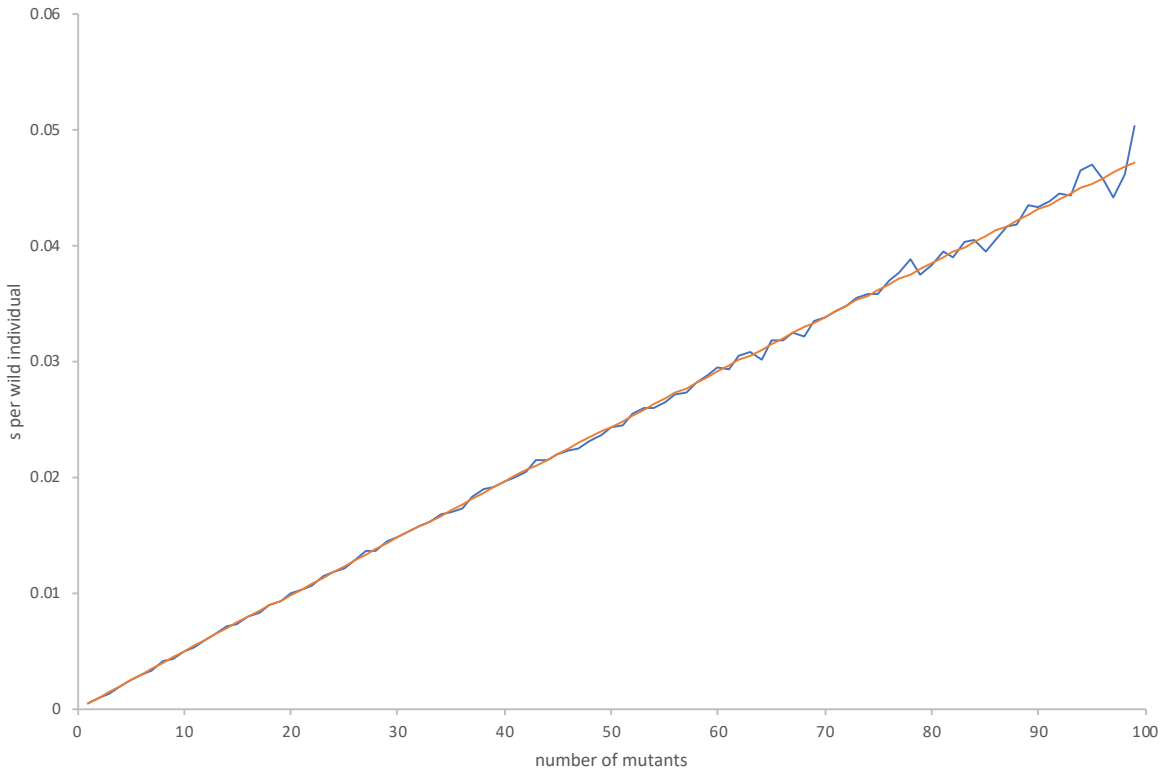

Figure S1. Observed and theoretical proportional loss of resource in wild type under conditions of benign mutant selection. Observed values (blue) generated assigning values of 0.21 and 0.2 to p_m and p_w respectively in a population size of 100 individuals with resource size 1000 and 10,000 replications. Individuals were randomly selected and then deemed successful in resource acquisition with a probability of p_m or p_w depending on whether they were wild type or mutant individuals. Sampling continued until all resource had been allocated. Model available as script benign_profile_of_s.pl from (website). Theoretical values (orange) calculated from equation (6) using the same values of p_m and p_w.

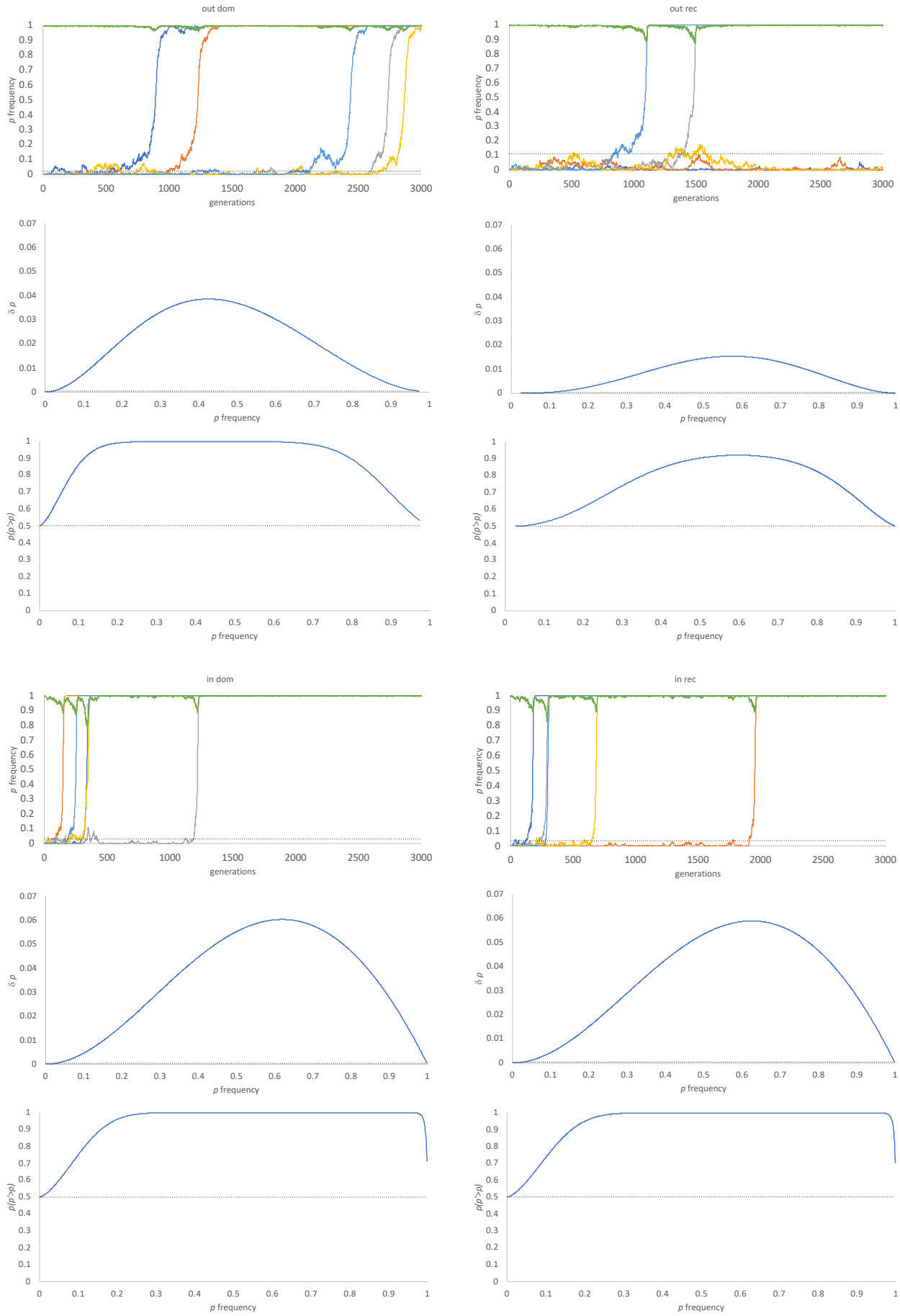
out dom
p frequency
generations
δ p
p frequency
p(p'>p)
p frequency
out rec
p frequency
generations
δ p
p frequency
p(p'>p)
p frequency
in dom
p frequency
generations
δ p
p frequency
p(p'>p)
p frequency
in rec
p frequency
generations
δ p
p frequency
p(p'>p)
p frequency

Figure S2. Competitive selection frequency dependent thresholds. Top panels: competitive selection simulations at 5 loci, for s= 0.0005 and N = 1000. Frequency of selected mutant (p) denoted by colours (dark blue, orange, grey, yellow and light blue respectively), population size N shown in green. Threshold frequencies of p at which $\delta p > 1/2N$ shown by dotted line. Middle panels: Values of δp at given frequencies of p calculated using equation (8) using p dependent values of s calculated in equation (6), for which values of N_m and N_w were calculated in equations (9) and (10). Threshold at which $\partial p = 1/2N$ shown by dotted line. Bottom panels: Probabilities of $p' > p$. Values of $p' = p + \delta p$. Normal approximation to the binomial distribution used in which $\mu = 2Np'$ and $\sigma = (2Np'(1-p')^{0.5}$ and cumulative probabilities calculated for $x > p$. Pure genetic drift levels are shown by dotted line where $\mu = p$. Simulations carried out under outbreeding (2% selfing) dominant (out dom), inbreeding (98% selfing) dominant (in dom), outbreeding recessive (out rec) and inbreeding recessive (in rec) conditions.

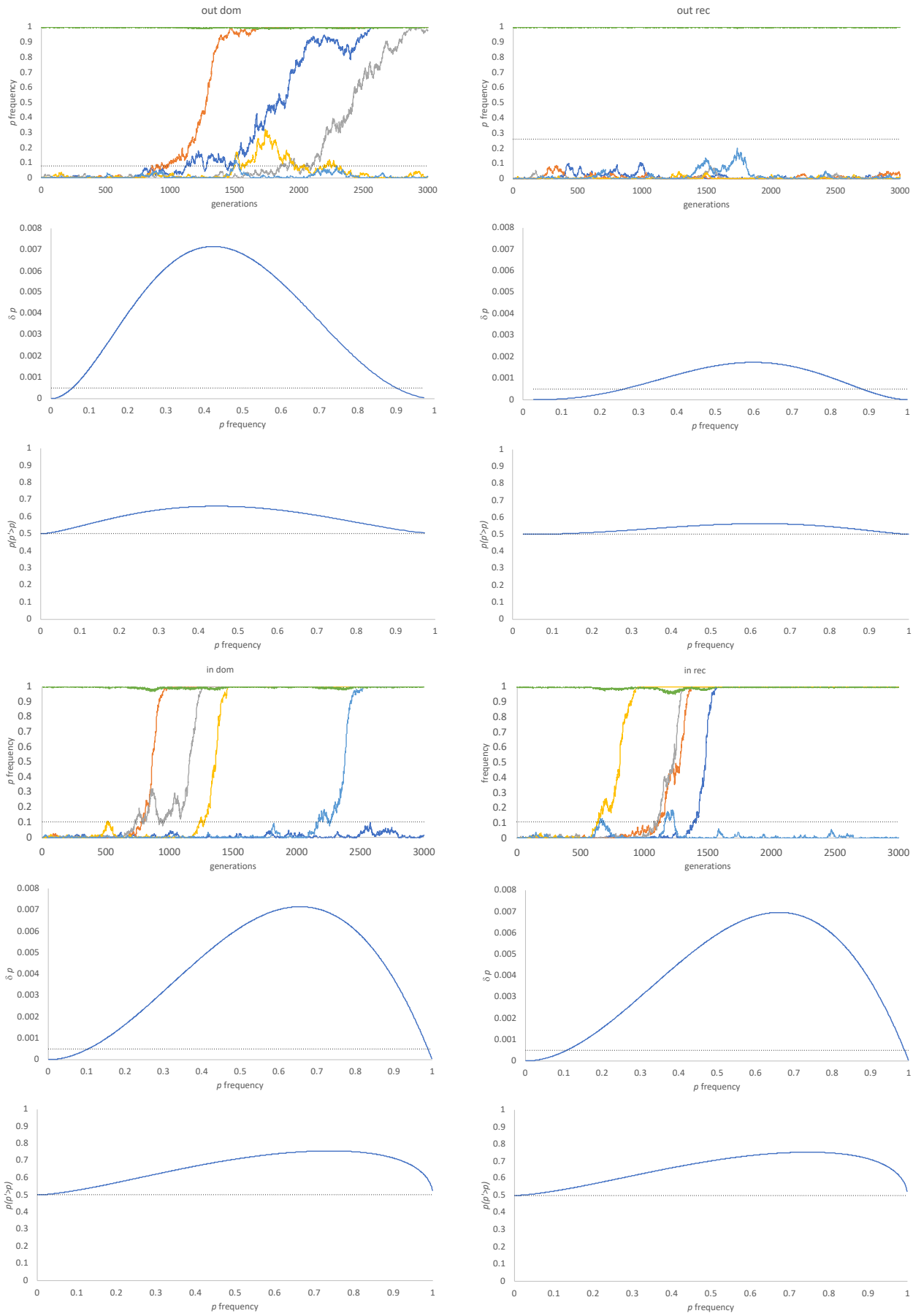
out dom
p frequency
generations
δ p
p frequency
p(p'>p)
p frequency
out rec
p frequency
generations
δ p
p frequency
p(p'>p)
p frequency
in dom
p frequency
generations
δ p
p frequency
p(p'>p)
p frequency
in rec
frequency
generations
δ p
p frequency
p(p'>p)
p frequency

Figure S3. Competitive selection frequency dependent thresholds. Top panels: competitive selection simulations at 5 loci, for s= 0.00005 and N = 1000. Frequency of selected mutant (p) denoted by colours (dark blue, orange, grey, yellow and light blue respectively), population size N shown in green. Threshold frequencies of p at which $\delta p > 1/2N$ shown by dotted line. Middle panels: Values of δp at given frequencies of p calculated using supplementary equation (4) using p dependent values of s calculated in equation (6), for which values of N_m and N_w were calculated in supplementary equation (5). Threshold at which $\delta p = 1/2N$ shown by dotted line. Bottom panels: Probabilities of $p' > p$. Values of $p' = p + \delta p$. Normal approximation to the binomial distribution used in which $\mu = 2Np'$ and $s = (2Np'(1-p')^{0.5}$ and cumulative probabilities calculated for $x > p$. Pure genetic drift levels are shown by dotted line where $\mu = p$. Simulations carried out under outbreeding (2% selfing) dominant (out dom), inbreeding (98% selfing) dominant (in dom), outbreeding recessive (out rec) and inbreeding recessive (in rec) conditions.

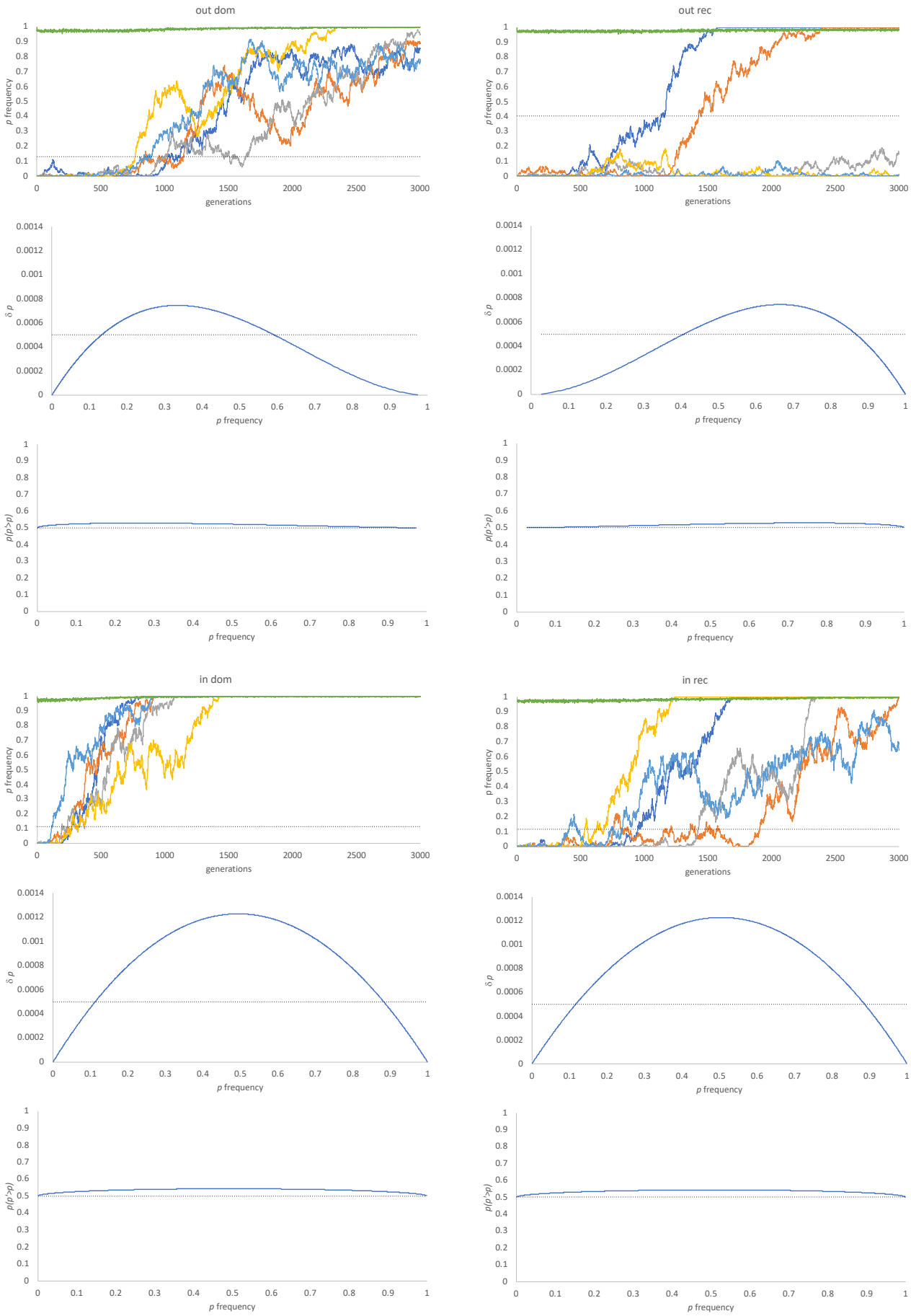
out dom
out rec
in dom
in rec
p frequency
generations
δ p
p(p'>p)
p frequency

Figure 7. Environmental selection frequency dependent thresholds. Top panels: environmental selection simulations at 5 loci, for s= 0.005 and N = 1000. Frequency of selected mutant (p) denoted by colours (dark blue, orange, grey, yellow and light blue respectively), population size N shown in green. Threshold frequencies of p at which $\delta p > 1/2N$ shown by dotted line. Middle panels: Values of δp at given frequencies of p calculated using supplementary equation (4) using p dependent values of s calculated in equation (6). Threshold at which $\delta p = 1/2N$ shown by dotted line. Bottom panels: Probabilities of $p' > p$. Values of $p' = p + \delta p$. Normal approximation to the binomial distribution used in which $\mu = 2Np'$ and $\sigma = (2Np'(1-p'))^{0.5}$ and cumulative probabilities calculated for $x > p$. Pure genetic drift levels are shown by dotted line where $\mu = p$. Simulations carried out under outbreeding (2% selfing) dominant (out dom), inbreeding (98% selfing) dominant (in dom), outbreeding recessive (out rec) and inbreeding recessive (in rec) conditions.

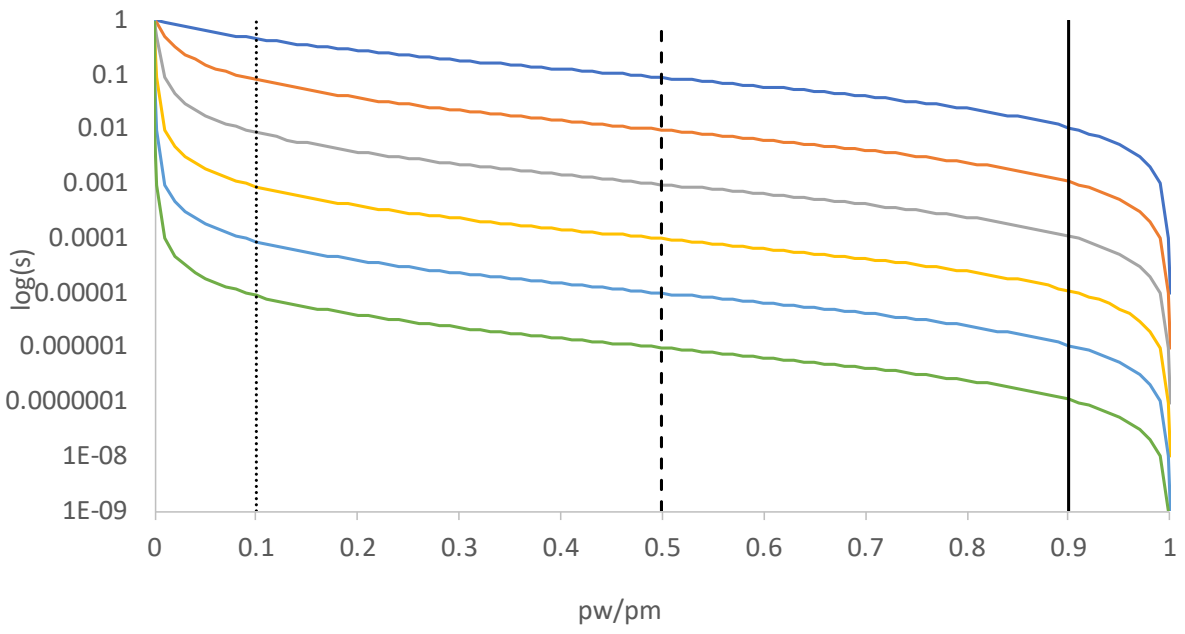

Figure S5 The relationship between pw/pm and s for given population sizes. Values of s were derived from equation (6) for given pw/pm values for the following population sizes: blue (10), orange (100), grey (1000), yellow (10,000), light blue (100,000) and green (1000,000). Vertical bars denote selection intensities of 0.1 (solid), 1 (dashed) and 10 (dotted).

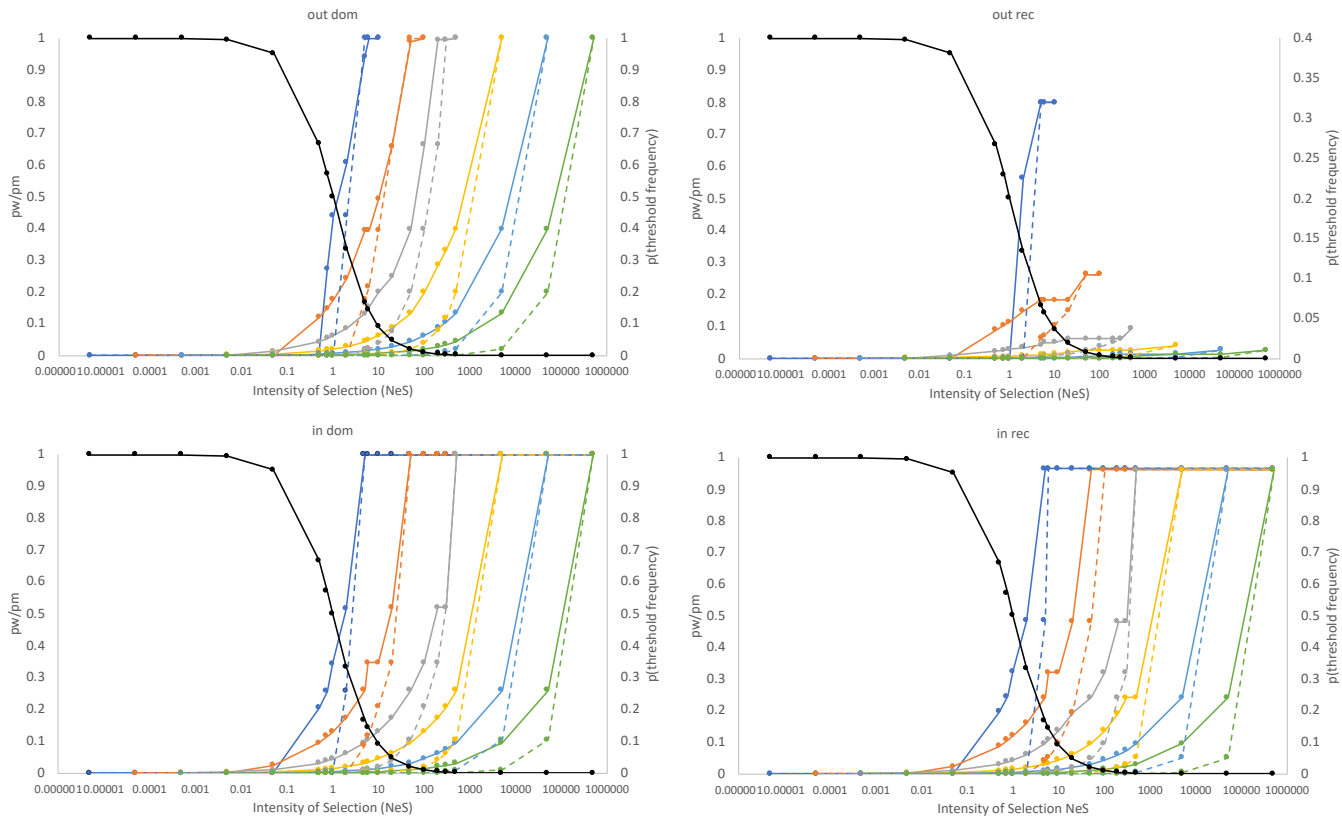

Figure S6. Probability of reaching drift/selection threshold frequencies for given selection intensities and their associated p_w/p_m ratios. The p_w/p_m ratios shown in dark black. Probabilities of reaching threshold frequencies through drift were calculated by first determining the number of individuals at which $\delta p > 1/2N$ using equations 8-10 and equation 6 to determine the frequency dependent values of s. The neutral probability of an allele reaching that frequency is given as $1/f_{(threshold)}$. Probabilities were calculated for population sizes of blue (10), orange (100), grey (1000), yellow (10,000), light blue (100,000) and green (1000,000), solid lines represent competitive selection probabilities and dashed lines represent the corresponding environmental selection probabilities. Values are calculated under outbreeding (2% selfing) dominant (out dom), inbreeding (98% selfing) dominant (in dom), outbreeding recessive (out rec) and inbreeding recessive (in rec) conditions.

					Bin Dates (BP)					
	r	*s*	*Nes*	p_m/p_w	first	last	*N*	timeframe	*p height*	p_{null}
barley breadth 1	0.83274258	8.04733E-06	0.8047328	1.80473928	11300	10170	100000	1130	0.006	0.0392
barley breadth 2	0.82593431	9.04202E-06	0.90420174	1.90420991	10075	9025	100000	1050	0.005	0.0704
barley breadth 3	0.98325575	8.15863E-06	0.40793151	1.40793484	7750	6500	50000	1250	0.019	0.0589
barley thickness 1	0.76449516	6.12979E-06	0.61297929	1.61298304	11050	9200	100000	1850	0.021	0.0361
barley thickness 2	0.99658765	7.34625E-06	0.36731245	1.36731514	7750	6500	50000	1250	0.047	<0.001
einkorn breadth 1	0.92669622	1.00989E-05	0.50494372	1.50494881	11050	10300	50000	750	0.009	0.0101
einkorn breadth 2	0.75952857	6.07423E-06	0.60742346	1.60742715	9950	8425	100000	1525	0.012	0.0762
einkorn breadth 3	0.29709117	5.18241E-06	0.51824057	1.51824325	8450	7000	100000	1450	0.035	0.7295
einkorn thickness 1	0.88384627	9.09309E-06	0.90930918	1.90931744	11512	10300	100000	1212	0.01	0.0060
einkorn thickness 2	0.99631575	7.1627E-06	0.35813512	1.35813768	10000	8712	50000	1288	0.022	<0.001
einkorn thickness 3	0.94067196	6.05017E-06	0.60501711	1.60502077	7300	7000	100000	300	0.008	0.0251
emmer thickness 1	0.94465145	8.12901E-06	0.81290079	1.8129074	9312	7500	100000	1812	0.015	<0.001
rice (indica) breadth 1	0.80466945	5.11564E-06	0.51156442	1.51156704	3925	2125	100000	1800	0.023	<0.001
rice (japonica) breadth 1	0.67239719	2.16677E-05	0.21667677	1.21668147	7650	5200	10000	2450	0.029	<0.001
rice (japonica) breadth 2	0.65032522	1.84918E-05	0.18491785	1.18492127	4950	4600	10000	350	0.026	0.0941
lentil length 1	0.93984652	6.05421E-06	0.60542145	1.60542512	11225	10350	100000	875	0.009	0.1093
lentil length 2	0.45014131	9.04643E-06	0.90464278	1.90465096	10062	9562	100000	500	0.005	0.2052
lentil length 3	0.84050712	6.18286E-06	0.61828609	1.61828991	9385	8400	100000	985	0.029	0.0010
pea length 1	0.94233761	5.11646E-06	0.51164627	1.51164889	10650	9895	100000	755	0.022	0.0262
pea length 2	0.96610421	6.09332E-06	0.6093318	1.60933551	9750	9312	100000	438	0.015	0.0020
sunflower length 1	0.98754599	6.07981E-06	0.6079807	1.60798439	6675	4950	100000	1725	0.013	0.1597
sunflower length 2	0.82637007	9.33817E-06	0.00933817	1.00933825	4700	2855	1000	1845	0.896	0.2298
sunflower length 3	0.92894554	1.7484E-05	0.17484025	1.17484331	2155	1200	10000	955	0.027	0.1081
sunflower breadth 1	0.92332224	6.10484E-06	0.61048379	1.61048752	5325	4000	100000	1325	0.017	0.2515
sunflower breadth 2	0.81764058	1.50505E-05	0.01505051	1.01505073	2607	2215	1000	392	0.328	0.0270
sunflower breadth 3	0.93326734	9.05106E-06	0.90510587	1.90511406	2110	1200	100000	910	0.005	0.0461
soybean length 1	0.96449128	7.07858E-06	0.70785833	1.70786334	5475	5275	100000	200	0.011	0.0531
soybean length 2	0.54628783	1.11584E-05	1.11583602	2.11584847	4725	4060	100000	665	0.014	0.7487

Table 1. Best fit parameters of theoretical competitive selection to archaeometric grain data