Title: Genetic costs of domestication and improvement

Authors: Brook T. Moyers[1], Peter L. Morrell[2], and John K. McKay[1]

Affiliations: [1] Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, CO, USA 80521. [2] Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, USA 55108.

Address for correspondence:
Brook T. Moyers
Bioagricultural Sciences and Pest Management
Colorado State University
Fort Collins, CO, USA 80521
Phone: 707-235-1237
Email: brook.moyers@gmail.com

Running title: Cost of domestication

ABSTRACT

The ‘cost of domestication’ hypothesis posits that the process of domesticating wild species can result in an increase in the number, frequency, and/or proportion of deleterious genetic variants that are fixed or segregating in the genomes of domesticated species. This cost may limit the efficacy of selection and thus reduce genetic gains in breeding programs for these species. Understanding when and how deleterious mutations accumulate can also provide insight into fundamental questions about the interplay of demography and selection. Here we describe the evolutionary processes that may contribute to deleterious variation accrued during domestication and improvement, and review the available evidence for ‘the cost of domestication’ in animal and plant genomes. We identify gaps and explore opportunities in this emerging field, and finally offer suggestions for researchers and breeders interested in understanding or avoiding the consequences of an increased number or frequency of deleterious variants in domesticated species.

Keywords: deleterious variants, crops, domesticated animals

INTRODUCTION

Recently, we have seen a resurgence of evolutionary concepts applied to the domestication and improvement of plants and animals (e.g. Walsh 2007; Wang *et al*. 2014; Gaut, Díez, and Morrell 2015; Kono *et al*. 2016). One particular wave of this resurgence proposes a general 'cost of domestication': that the evolutionary processes experienced by lineages during domestication are likely to increase the number of deleterious variants in the genome. This 'cost' was first hypothesized by Lu *et al*. (2006), who found an increase in nonsynonymous substitutions, particularly radical amino acid changes, in domesticated compared to wild lineages of rice. These putatively deleterious variants were negatively correlated with recombination rate, which the authors interpreted as evidence that they hitchhiked along with the targets of artificial selection (Figure 1B). Lu *et al*. (2006) conclude that: "The reduction in fitness, or the genetic cost of domestication, is a general phenomenon." Here we address this claim by examining the evidence that has emerged in the last decade on deleterious variants in domesticated species.

The processes of domestication and subsequent breeding potentially impose a number of evolutionary effects on populations (Box 1). New mutations can have a range of effects on fitness, from lethal to beneficial. Deleterious variants constitute the mostly directly observable, and likely the most important, source of mutational and therefore genetic load in a population (Box 2). The shape of the distribution of fitness effects of new mutations is difficult to estimate, and estimates suggest it varies across populations and species (Keightley and Eyre-Walker 2007). However, theory predicts that a large proportion of new mutations, particularly those that occur in coding portions of the genome, will be deleterious at least in some proportion of the environments that a species occupies (Ohta 1972; 1992; Gillespie 1994). These predictions are supported by experimental responses to artificial selection and mutation accumulation experiments, as well as molecular genetic studies of variation in populations (Keightley and Lynch 2003; Eyre-Walker, Woolfit, and Phelps 2006; Boyko *et al*. 2008; Kim, Huber,

and Lohmueller 2017).

The rate of new mutations in eukaryotes varies, but is likely at least 1 x 10^{-8} / base pair / generation (Baer, Miyamoto, and Denver 2007). For the average eukaryotic genome, individuals should thereby be expected to carry a small number of new mutations not present in the parent genome(s) (Agrawal and Whitlock 2012). The realized distribution of fitness effects for these mutations in a population will be influenced by inbreeding and by effective population size (Gillespie 1999; Whitlock 2000; Arunkumar *et al*. 2015; Keightley and Eyre-Walker 2007). Generally, for a given distribution of fitness effects for segregating variants, we expect to observe relatively fewer strongly deleterious variants and more weakly deleterious variants in smaller populations and in populations with higher rates of inbreeding (Figure 2A).

Domestication and improvement often involve increased inbreeding. Allard (1999) pointed out that many important early cultigens were produced by inbreeding, with inbred lines offering agronomic and morphological consistency in the crop. In some cases, domestication involved a switch in mating system from outcrossing to highly selfing (e.g. in rice; Kovach, Sweeney, and McCouch 2007). The practice of producing inbred lines, often through single-seed descent, with the goal of reducing heterozygosity and creating genetically ‘stable’ varieties, remains a major activity in plant breeding programs. Reduced effective population sizes (N_e) during domestication and artificial selection on favorable traits also constitute forms of inbreeding (Figure 1). Even in species without the capacity to self-fertilize, inbreeding that results from selective breeding can dramatically change patterns of genomic variation. For example, the first sequenced dog genome, from a female boxer, has runs of homozygosity that span 62% of genome, with an N50 length of 6.9 Mb (versus heterozygous region N50 length 1.1 Mb; Lindblad-Tor *et al*. 2005). This level of homozygosity drastically reduces the effective recombination rate, as crossover events often exchange identical chromosomal segments. The likelihood of a beneficial allele

moving into a genomic background with fewer linked deleterious alleles is thus reduced.

Linked selection, or interference among mutations, has much the same effect. For a trait influenced by multiple loci, interference between linked loci can limit the response to selection (Hill and Robertson 1966). Deleterious variants are more numerous than those that have a positive effect on a trait and thus should constitute a major limitation in responding to selection (Felsenstein 1974). This forces selection to act on the net effect of favorable and unfavorable mutations in linkage (Figure 1B). As homozygosity and LD increase, N_e decreases, and so selection is less effective at purging moderately deleterious mutations and new slightly beneficial mutations are more likely to be lost to genetic drift (sometimes termed 'inbreeding depression'). These dynamics will shift the distribution of fitness effects for segregating variants and result in greater accumulation of deleterious variation over time (Figure 2A–C).

Overall, the cost of domestication hypothesis posits that compared to their wild relatives domesticated lineages will have:

1. Deleterious variants at higher number, frequency, and/or proportion
2. Enrichment of deleterious variants in linkage disequilibrium with loci subject to strong, positive, artificial selection

Among domesticated species, these effects may differ between lineages that experienced domestication only (e.g. landraces and non-commercial populations) and those that were subject to modern improvement (e.g. 'elite' varieties and commercial breeds). Domestication typically involves a genetic bottleneck followed by a long period of relatively weak and possibly varying selection, while during the process of improvement, intense selection over short time periods is coupled with limited recombination and an additional reduction in N_e, often followed by rapid population expansion (Figure 1A; Yamasaki, Wright, and McMullen 2007). Gaut, Díez, and Morrell (2015) propose that elite crop lines

136 could harbor a lower proportion of deleterious variants relative to landraces due
137 to strong selection for yield during improvement, but the opposing pattern could
138 be driven by lower N_e and thus increased genetic drift, limited effective
139 recombination, and rapid population expansion. At least one study, in sunflower,
140 shows little difference in the composition of deleterious variants between
141 landrace and elite lines (Renaut and Rieseberg 2015). It is likely that the relative
142 influence of these factors varies dramatically across domesticated systems.
143
144

---

145 Box 1. Domesticated lineages may experience:

146 - **Increased number of deleterious variants:**
147 - Deleterious mutations may accumulate at higher rates in
148 domesticated lineages versus their wild relatives due to the
149 reduced efficacy of selection relative to genetic drift. Mutations that
150 would be purged with a larger N_e or with higher effective
151 recombination rates are instead retained. This is reflected in a shift
152 in the distribution of fitness effects for segregating variants towards
153 more moderately deleterious alleles (Figure 2A).
154 - **However**: any significant increase in inbreeding, particularly the
155 transition from outcrossing to selfing, can result in the purging of
156 recessive, highly deleterious alleles, as these alleles are exposed in
157 homozygous genotypes more frequently (Arunkumar *et al*. 2015).
158 This shifts the more deleterious end of the distribution of fitness
159 effects for segregating variants towards neutrality (Figure 2A).
160

161 - **Increased frequency of deleterious variants:**
162 - As above, reduced N_e (due to inbreeding, genetic bottlenecks, or
163 strong selection) or reduced effective recombination rate will
164 increase the strength of genetic drift relative to selection. Stronger
165 genetic drift can allow deleterious variants to reach higher
166 frequencies. In the case of inbreeding, this pattern is called
167 inbreeding depression, but it can occur whenever N_e decreases.

- Deleterious variants that are in linkage disequilibrium with a target of artificial selection can also increase in frequency through genetic hitchhiking, as long as their fitness effects are smaller than the strength of selection on the targeted variant (Figure 1B; Hartfield and Otto 2011; Assaf, Petrov, and Blundell 2015).
- The rapid population expansion, coupled with long-distance migration common to the demographic history of many domesticated lineages, can result in the accumulation of deleterious genetic variation known as ‘expansion load’ (Peischl *et al*. 2013; Lohmueller 2014). This occurs when serial bottlenecks are followed by large population expansions (e.g. large local carrying capacities, large selection coefficients, long distance dispersal; Peischl *et al*. 2013). This phenomenon is due to the accumulation of new deleterious mutations at the ‘wave front’ of expanding populations, which then rise to high frequency via drift regardless of their fitness effects (also known as ‘gene surfing’, or ‘allelic surfing’ Klopfstein, Currat, and Excoffier 2006; Travis *et al*. 2007).

<u>Box 2: What is the link between deleterious variants and the long-standing concept of genetic load?</u>

The classic concept of genetic load identifies a reduction in fitness from a mean or optimal genotype (Haldane 1937). Deleterious variants are probably the largest single contributor to genetic load via ‘mutational load’ (Muller 1950; Felsenstein 1974; Agrawal and Whitlock 2012). Due to this, the term ‘genetic load’ has often been used colloquially to indicate mutational load, and attempts to estimate genetic load have involved identifying and quantifying deleterious genetic variants. However, the genetic load of an individual is not merely a count of deleterious variants, but is dependent on other factors, including the distributions of fitness effects and dominance coefficients for deleterious variants (see Henn *et al*. 2015). These factors are difficult to quantify and estimates of

genetic load are sensitive to their values (Boyko *et al*. 2008; Lohmueller *et al*. 2008; Lohmueller *et al*. 2014; Simons *et al*. 2014). For this reason, much of the modern literature has eschewed discussion of genetic load and focused on the various means of counting the number of deleterious variants (see Chun and Fay 2009; Marth *et al*. 2011). Similarly, this review primarily addresses patterns of deleterious variation in domesticated species, with the view that deleterious variants likely contribute significantly to genetic load in many cases.

Measures used to quantify the difference in mutational load among individuals and populations include:

Measure	**Individuals**	**Populations**
Absolute number of derived variants at otherwise conserved sites	X	X
Ratio of deleterious to synonymous variants	X	X
Average frequency of deleterious variants	–	X

Box 3: **Definitions**

allelic (or gene) surfing: the increase in frequency of a genetic variant at the wave front of an expanding population

ancient DNA: DNA samples from archaeological or historical remains

deleterious variant: a genetic variant that reduces fitness

derived mutation: a genetic variant that involves a change from the ancestral state as inferred from a related, outgroup taxon or from sampling of the ancestral genome (e.g. using ancient DNA)

effective population size (N_e): the number of individuals contributing offspring to the next generation of a population, versus the census population size

genetic bottleneck: a large reduction in census population size

genetic hitchhiking: when a genetic variant increases in frequency owing to linkage disequilibrium with a variant subject to positive selection

haplotype: set of genetic variants in LD with each other

linkage disequilibrium (LD): nonrandom association of alleles at two or more loci

LD N50: the approximate physical or genetic distance over which LD decays to half of its maximum value

linked selection: the tendency for selection on a variant to affect the frequency of nearby variants due to LD

private alleles: genetic variants unique to a population or set of populations

runs of homozygosity: regions of the genome where both of a pair of chromosomes are highly similar, thus increasing LD due to reduced effective recombination

site frequency spectrum: the distribution of variant frequencies in a population

---

<u>Deleterious variation in Domesticated Species</u>

The rapidly expanding number of genome-wide datasets has made the study of deleterious variants increasingly feasible. Since Lu *et al*. (2006) first hypothesized the accumulation of deleterious variants as a 'cost of domestication' in rice, similar studies have examined evidence for such costs in other crops and domesticated animals.

A number of approaches have been used to quantify the composition and effect of deleterious variants in domesticated genomes. These approaches largely parallel those used in studies of the effects of demography on human populations (Lohmueller 2014; Henn *et al*. 2015). Some studies have inferred a cost of domestication indirectly from the accumulation of a larger proportion or number of nonsynonymous variants in domesticated lineages versus their wild relatives, assuming that nonsynonymous changes are on average deleterious (e.g. Lu *et*

263 *al.* 2006; Cruz, Vilà, and Webster 2008). Other studies have looked for reduced
264 genetic diversity or longer distance linkage disequilibrium (e.g. Lindblad-Toh *et*
265 *al*. 2005; Lam *et al.* 2010), and therefore reduced effective recombination and
266 presumably reduced efficacy of selection in the same comparison. These
267 approaches assume that wild relatives of domesticated species have not
268 themselves experienced population bottlenecks, shifts in mating system, or any
269 of the other processes that could affect patterns of deleterious variation relative
270 to their shared ancestors.
271
272 A more direct approach used to estimate the cost of domestication is to assess
273 the number and proportion of putatively deleterious variants present in
274 populations of domesticated species. The measures used include 1) the absolute
275 number of variants at derived sites, 2) the ratio of deleterious to synonymous
276 variants, and 3) an increase in the frequency of deleterious variants within a
277 population (Table S1; Box 2; also reviewed in Lohmueller 2014). We discuss
278 approaches to identifying deleterious variants in more depth below.
279
280 Several approaches attempt to translate the number of observed deleterious
281 variants into an estimate of genetic load. The first approach is to assume that the
282 degree of phylogenetic constraint on a variant provides an estimate of the
283 strength of purifying selection. Summing over the constraint scores (typically
284 using the GERP++ software; Cooper et al. 2005) relative to the number of
285 deleterious variants provides a means of estimating mutational load (e.g. Wang
286 *et al*. *bioRxiv*; Mardsen *et al*. 2016). A more general translation of deleterious
287 variants into genetic load depends on three factors: the distribution of fitness
288 effects of deleterious variants, a model of either additive or multiplicative effects
289 on fitness, and an estimate of dominance of deleterious variants (Arunkumar *et*
290 *al*. 2015; Henn *et al*. 2015; Henn *et al*. 2016; Brandvain and Wright 2016). This
291 approach directly addresses the question of genetic load, but does not tell us
292 whether the load is the result of domestication or other evolutionary processes
293 (unless wild relatives are also assayed, again assuming that their own

evolutionary trajectory has not simultaneously been affected). It also requires accurate, unbiased algorithms for identifying deleterious variants from neutral variants (see below).

Studies taking these approaches in domesticated species are presented qualitatively in Table 1 (and quantitatively in Supplementary Table 1). We searched the literature using Google Scholar with the terms (“genetic load” or “deleterious”) and “domesticat*”, and in many cases followed references from one study to the next. To the best of our knowledge, the studies in Table 1 represent the majority of the extant literature on this topic. We excluded studies examining only the mitochondrial or other non-recombining portions of the genome. In a few cases where numeric values were not reported, we extracted values from published figures using relative distances as measured in the image analysis program ImageJ (Schneider, Rasband, and Eliceiri 2012), or otherwise extrapolated values from the provided data. Where exact values were not available, we provide estimates. Please note that no one study (or set of genotypes) contributed values to all columns for a particular domestication event, and that in many cases methods used or statistics examined varied across species.

Recombination and linkage disequilibrium

The process of domestication may increase recombination rate, as measured in the number of chiasmata per bivalent. Theory predicts that recombination rate should increase during periods of rapid evolutionary change (Otto and Barton 1997), and in domesticated species this may be driven by strong artificial selection or small N_es (Otto and Barton 2001). This is supported by the observation that recombination rates are higher in many domesticated species compared to wild relatives (e.g. chicken, Groenen *et al*. 2009; honey bee, Wilfert *et al.* 2007; and a number of cultivated plant species, Ross-Ibarra 2004; but not mammals, Muñoz-Fuentes *et al*. 2014). In contrast, recombination in some domesticated species is limited by genomic structure (e.g. in barley, where 50%

325 of physical length and many functional genes are in (peri-)centromeric regions
326 with extremely low recombination rates; The International Barley Genome
327 Sequencing Consortium 2012), although this structure may be shared with wild
328 relatives. Even when actual recombination rate (chiasmata per bivalent)
329 increases, effective recombination may be reduced in many domesticated
330 species due to an increase in linkage disequilibrium (LD) due to runs of
331 homozygousity. In other words, chromosomes may physically recombine, but if
332 the homologous chromosomes contain identical sequences then no 'effective
333 recombination' occurs and the outcome is no different than no recombination. In
334 a striking example, the effective recombination rate in maize populations has
335 decreased by an estimated 83% compared to that in the wild relative teosinte
336 (Wright *et al*. 2005; with comparable estimates in Hufford *et al.* 2012). We note
337 that Mezmouk and Ross-Ibarra (2014) found that deleterious variants are not
338 enriched in areas of low recombination in maize (although see McMullen *et al*.
339 2009; Rodgers-Melnick *et al*. 2015).
340
341 Strong directional selection like that imposed during domestication can reduce
342 genetic diversity in chromosomal regions linked to the selected locus, creating
343 runs of homozygousity (Sved 1971; Maynard Smith and Haigh 1974). These
344 regions of extended LD are among the signals used to identify targets of
345 selection and reconstruct the evolutionary history of domesticated species (e.g.
346 Tian, Stevens, and Buckler 2009). Any increase in inbreeding (e.g. after a
347 population bottleneck or a shift in mating system from outcrossing to selfing) will
348 have the same effect across the entire recombining portion of the genome, as
349 heterozygosity decreases with each generation (Pritchard and Przeworski 2001;
350 Charlesworth 2003). Extended LD has consequences for the efficacy of
351 selection: deleterious variants linked to larger-effect beneficial alleles can no
352 longer recombine away, and beneficial variants that are not in LD with larger-
353 effect beneficial alleles may be lost by genetic drift (Hartfield and Otto 2011;
354 Assaf, Petrov, and Blundell 2015). In our review of the literature, LD decays most
355 rapidly in outcrossing plant species (maize and sunflower), and extends much

356 further in self-fertilizing plants and in domesticated animals (Table S1). In all
357 cases where we have data for wild relatives, LD decays more rapidly in wild
358 lineages than in domesticated lineages (Table S1).
359
360 While we primarily report mean LD in Table S1, linkage disequilibrium also varies
361 among varieties and breeds of the same species. For example, in domesticated
362 pig breeds the length of the genome covered by runs of homozygosity, which
363 extend LD, ranges from 13.4 to 173.3 Mb, or 0.5–6.5% (Traspov *et al*. 2016; Box
364 3). Similarly, among dog breeds average LD decay (to $r^2 \leq 0.2$) ranges from 20
365 kb to 4.2 Mb (Gray *et al*. 2009). In both species, the decay of LD for wild
366 individuals occurs over shorter distances (Table S1). This suggests that patterns
367 of LD in these species have been strongly impacted by breed-specific
368 demographic history (i.e. the process of improvement), in addition to the shared
369 process of domestication.
370
371 *Genetic diversity*
372 We see consistent loss of genetic diversity when 'improved' or 'breed' genomes
373 are compared to domesticated 'landrace' or 'non-commercial' genomes, and
374 again when domesticated genomes are compared to the genomes of wild
375 relatives (Table 1; Table S1). This ranges from ~5% nucleotide diversity lost
376 between wolf populations and domesticated dogs (Gray *et al*. 2009) to 77% lost
377 between wild and improved tomato populations (Lin *et al*. 2014). The only case
378 where we see a gain in genetic diversity is in the Andean domestication of the
379 common bean, where gene flow with the more genetically diverse Mesoamerican
380 common bean is likely an explanatory factor (Schmutz *et al*. 2013). This pattern
381 is consistent with a broader review of genetic diversity in crop species: Miller and
382 Gross (2011) found that annual crops had lost an average of ~40% of the
383 diversity found in their wild relatives. This same study found that perennial fruit
384 trees had lost an average of ~5% genetic diversity, suggesting that the impact of
385 domestication on genetic diversity is strongly influenced by life history (see also
386 Gaut, Díez, and Morrell 2015). Given a relatively steady evolutionary trajectory

387 for wild populations, loss of genetic diversity in domesticated populations can be
388 attributed to artificial selection or reduced N_e due to increased inbreeding and
389 genetic bottlenecks. On an evolutionary timescales, even the oldest
390 domestications occurred recently relative to the rate at which new mutations can
391 recover the loss. This is especially true for non-recombining portions of the
392 genome like the mitochondrial genome or sex chromosomes. Most modern
393 animal breeding programs are strongly sex-biased, with few male individuals
394 contributing to each generation. In horses, for example, this has likely led to
395 almost complete loss of polymorphism on the Y chromosome via genetic drift
396 (Lippold *et al*. 2011). Loss of allelic diversity can reduce the efficacy of selection
397 by reducing additive genetic variance within species (Eyre-Walker, Woolfit, and
398 Phelps 2006). However, a loss of genetic diversity alone does not necessarily
399 signal a corresponding increase in the frequency or proportion of deleterious
400 variants, and so is not sufficient evidence of a cost of domestication.
401
402 *Synonymous versus nonsynonymous variation*
403 In six species, the domesticated lineage shows an increase in genome-wide
404 nonsynonymous to synonymous substitution rate or number compared to a wild
405 lineage (Table 1). This is also true across the domesticated lineages of cattle
406 (especially the domesticated cow; MacEachern *et al.* 2009). An exception is in
407 soybean, where the domesticated *Glycine max* and wild *G. soja* genomes
408 contain approximately the same proportion of nonsynonymous to synonymous
409 single nucleotide polymorphisms (Lam *et al*. 2010), which are not directly
410 comparable to substitutions but should exhibit similar patterns under the cost of
411 domestication hypothesis. These differences in nonsynonymous substitution rate
412 are likely driven by differences in N_e (Eyre-Walker and Keightley 2007; Woolfit
413 2009). If nonsynonymous mutations are on average deleterious, as theory and
414 empirical data suggest (Keightley and Lynch 2003; Eyre-Walker, Woolfit, and
415 Phelps 2006; Boyko *et al*. 2008; Kim, Huber, and Lohmueller 2017), then an
416 increase in nonsynonymous substitutions will reduce mean fitness (Figure 2B-C).
417 The comparisons in Table 1 suggest that this has occurred in domesticated

species. This result differs from Moray, Lanfear, and Bromham (2014), who examined rates of mitochondrial genome sequence evolution in domesticated animals and their wild relatives and found no such consistent pattern. This difference may be attributable to the focus of each review (genome-wide versus mitochondria) or, as the authors speculate, to genetic bottlenecks in some of the wild relatives included in their study.

The ratio of nonsynonymous to synonymous substitutions may not be a good estimate for mutational load. For one, nonsynonymous substitutions are particularly likely to have phenotypic effects (Kono *et al*. *bioRxiv*). Variants annotated as deleterious based on sequence conservation (see below) can in some cases contribute to agronomically important phenotypes (e.g. Nie *et al*. 2015; see also Albalat and Cañestro 2016), and artificial selection during domestication and improvement is likely to drive a portion of these variants associated with favorable phenotypes to higher frequencies (or fixation) in domesticated lineages (Kono *et al*. 2016). In addition, estimates of the proportion of nonsynonymous sites with deleterious effects range from 0.03 (in bacteria; Hughes 2005) to 0.80 (in humans; Fay *et al*. 2001; The Chimpanzee Sequencing and Analysis Consortium 2005). It follows that nonsynonymous substitution rate may have a poor correlation with patterns of deleterious variation and mutational load, at least at larger taxonomic scales. However, the consensus proportion of deleterious variants in Table S1 is between 0.05 and 0.25, which spans a smaller range. Finally, dN/dS and related ratios may be flawed estimates of functional divergence because they rely on the assumption that synonymous mutations are neutral and thus can control for substitution rate variation. This assumption may not hold true, especially for closely related taxa (e.g. Wolf *et al*. 2009; see also Kryazhimskiy and Plotkin 2008), and at least one study shows elevated rates of synonymous relative to non-coding substitutions in domesticated lineages (MacEachern *et al.* 2009).

Deleterious variants

449 An increase in mutational load can come from an increased frequency or number
450 of deleterious variants. The first can be assessed by examining shared
451 deleterious variants between wild and domesticated lineages, and both by
452 comparing shared and private deleterious variants across lineages. Looking at
453 deleterious variants that are at high frequency across all domesticated varieties
454 may provide insight into the early processes of domestication, while looking at
455 deleterious variants with varying frequencies among domesticated varieties may
456 provide insight into the processes of improvement. We recommend that
457 researchers studying deleterious variants report results per genome, and then
458 compare across genomes and among lineages. Many, but not all, of the studies
459 in Table 1 take this approach. The total number or frequency of deleterious
460 variants within a population will necessarily depend on the size of that population,
461 and so sufficient sampling is important before values can be compared across
462 populations. Values per genome are easier to compare across most available
463 genomic datasets.
464
465 Renaut and Rieseberg (2015) found a significant increase in both shared and
466 private deleterious mutations in domesticated relative to wild lines of sunflower,
467 and similar patterns in two additional closely-related species: cardoon and globe
468 artichoke (Table 1). This pattern also holds true for *japonica* and *indica*
469 domesticated rice (Liu *et al*. 2017) and for the domesticated dog (Marsden *et al*.
470 2016) compared to their wild relatives (Table 1). In horses, deleterious mutation
471 load as estimated using Genomic Evolutionary Rate Profiling (GERP) appears to
472 be higher in both domesticated genomes and the extant wild relative
473 (Przewalski's horse, which went through a severe genetic bottleneck in the last
474 century) compared to an ancient wild horse genome (Schubert *et al*. 2014).
475 These four studies, spanning a wide taxonomic range, suggest that an increase
476 in the number and proportion of deleterious variants may be a general
477 consequence of domestication. However, the other studies we present that
478 examined deleterious variants in domesticated species did not include any
479 sampling of wild lineages. Without sufficient sampling of a parallel lineage that

480 did not undergo the process of domestication, it is difficult to assess whether the
481 'cost of domestication' is indeed general.
482
483 *Identifying dangerous hitchhikers*
484 The effect size of a deleterious mutation is negatively correlated with its
485 likelihood of increasing in frequency through any of the mechanisms we discuss
486 here. That is, variants with a strongly deleterious effect are more likely to be
487 purged by selection than mildly deleterious variants. Similarly, mutations that
488 have a consistent, environmentally-independent deleterious effect are more likely
489 to be purged than mutations with environmentally-plastic effects. In the extreme
490 case, we would never expect mutations that have a consistently lethal effect in a
491 heterozygous state to contribute to a persistent cost of domestication, as these
492 would be lost in the first generation of their appearance in a population. However,
493 mutations with consistent, highly deleterious effects are likely rare relative to
494 those with smaller or environment-dependent effects, especially in inbred
495 populations (Figure 2A; Arunkumar *et al*. 2015). When thinking about patterns of
496 deleterious variation, it is therefore important to recognize that the effect of any
497 particular mutation can depend on context, including genomic background and
498 developmental environment. This complexity likely makes these classes of
499 deleterious variants more difficult to identify, and we might not expect these
500 variants to show up in bioinformatic screens (discussed below). Furthermore, all
501 of these expectations are modified by linkage: when deleterious variants are in
502 LD with targets of artificial selection, they are more likely to evade purging even
503 with consistent, large, deleterious effects (Figure 1B).
504
505 We searched the literature for examples of specific deleterious variants that
506 hitchhiked along with targets of selection during domestication and improvement.
507 We did not find many such cases, so we describe each in detail here. The best-
508 characterized example comes from rice, where an allele that negatively affects
509 yield under drought (*qDTY1.1*) is tightly linked to the major green revolution
510 dwarfing allele *sd1* (Vikram *et al*. 2015). Vikram *et al*. (2015) found that the

511 *qDTY1.1* allele explained up to 31% of the variance in yield under drought across
512 three RIL populations and two growing seasons. Almost all modern elite rice
513 varieties carry the *sd1* allele (which increases plant investment in grain yield),
514 and as a consequence these varieties are highly sensitive to drought. The
515 discovery of the *qDTY1.1* allele has enabled rice breeders to finally break the
516 linkage and create drought tolerant, dwarfed lines.
517
518 In sunflower, the *B* locus affects branching and was a likely target of selection
519 during domestication (Bachlava *et al*. 2010). This locus has pleiotropic effects on
520 plant and seed morphology that, in branched male restorer lines, mask the effect
521 of linked loci with both 'positive' (increased seed weight) and 'negative' (reduced
522 seed oil content) effects (Bachlava *et al*. 2010). To properly understand these
523 effects required a complex experimental design, where these linked loci were
524 segregated in unbranched (*b*) and branched (*B*) backgrounds. Managing these
525 effects in the heterotic sunflower breeding groups has likely also been
526 challenging.
527
528 A similarly complex narrative has emerged in maize. The gene *TGA1* was key to
529 evolution of 'naked kernels' in domesticated maize from the encased kernels of
530 teosinte (Wang *et al*. 2015). This locus has pleiotropic effects on kernel features
531 and plant architecture, and is in linkage disequilibrium with the gene *SU1*, which
532 encodes a starch debranching enzyme (Brandenburg *et al*. 2017). *SU1* was
533 targeted by artificial selection during domestication (Whitt *et al*. 2002), but also
534 appears to be under divergent selection between Northern Flints and Corn Belt
535 Dents, two maize populations (Brandenburg *et al*. 2017). This is likely because
536 breeders of these groups are targeting different starch qualities, and this work
537 may have been made more difficult by the genetic linkage of *SU1* with *TGA1*.
538
539 In the above cases, the linked allele(s) with negative agronomic effects are
540 unlikely to be picked up in a genome-wide screen for deleterious variation, as
541 they are segregating in wild or landrace populations and are not necessarily

542 disadvantageous in other contexts. One putative ‘truly’ deleterious case is in
543 domesticated chickens, where a missense mutation in the thyroid stimulating
544 hormone receptor (*TSHR*) locus sits within a shared selective sweep haplotype
545 (Rubin *et al*. 2010). However, Rubin *et al*. (2010) argue this is more likely a case
546 where a ‘deleterious’ (i.e. non-conserved) allele was actually the target of artificial
547 selection and potentially contributed to the trait of year-round egg laying in
548 chickens.
549
550 In the Roundup Ready (event 40-3-2) soybean varieties released in 1996, tight
551 linkage between the transgene insertion event and another allele (or possibly an
552 allele created by the insertion event itself) reduced yield by 5-10% (Elmore *et al*.
553 2001). This is not quite genetic hitchhiking in the traditional sense, but the yield
554 drag effect persisted through backcrossing of the transgene into hundreds of
555 varieties (Benbrook 1999). This effect likely explains, at least in part, why
556 transgenic soybean has failed to increase realized yields (Xu *et al*. 2013).
557 Consistent with this prediction, a second, independent insertion event in
558 Roundup Ready 2 Yield® does not suffer from the same yield drag effect (Horak
559 *et al*. 2015).
560
561 It is likely that ‘dangerous hitchhiker’ examples exist that have either gone
562 undetected by previous studies (possibly due to low genomic resolution, limited
563 phenotyping, or limited screening environments), have been detected but not
564 publicized, or are buried among other results in, for example, large QTL studies.
565 It is also possible that the role of genetic hitchhiking has not been as important in
566 shaping genome-wide patterns of deleterious variation as previously assumed.
567
568 <u>Gaps and Opportunities</u>
569 *How much of the genome is in LD with major targets of artificial selection?*
570 Artificial selection during domestication targets a clear change in the optimal
571 multivariate phenotype. This likely affects a significant portions of the genome:
572 available estimates include 2–4% of genes in maize (Wright *et al*. 2005) and 16%

573 of the genome in common bean (Papa *et al*. 2007) targeted by selection during
574 domestication. For crops, traits such as seed dormancy, branching,
575 indeterminate flowering, stress tolerance, and shattering are known to be
576 selected for different optima between artificial versus natural selection (Takeda
577 and Matsuoka 2008; Gross and Olsen 2010). In some cases, we know the loci
578 that underlie these domestication traits. One well-studied example is the green
579 revolution dwarfing gene, *sd1* in rice. *sd1* is surrounded by a 500kb region (~13
580 genes) with reduced allelic diversity in *japonica* rice (Asano *et al*. 2011). Another
581 example in rice is the *waxy* locus, where a 250 kb region shows reduced diversity
582 consistent with a selective sweep in temperate *japonica* glutinous varieties
583 (Olsen *et al*. 2006). The difference in the size of the region affected by these two
584 selective sweeps may be because the strength of selection on these two traits
585 varied, with weaker selection at *waxy* then *sd1*. Unfortunately, the relative
586 strength of selection on domestication traits is largely unknown, and other factors
587 can also influence the size of the genomic region affected by artificial selection.
588 The physical position of selected mutation can have large effect on this via gene
589 density and local recombination rate (e.g. in rice, Flowers *et al*. 2012). This
590 explanation has been invoked in maize (Wright *et al*. 2005), where the extent of
591 LD surrounding domestication loci is highly variable. For example a 1.1 Mb
592 region (~15 genes) lost diversity during a selective sweep on chromosome 10 in
593 maize (Tian, Stevens, and Buckler 2009), but only a 60–90 kb extended
594 haplotype came with the *tb1* domestication allele (Clark *et al*. 2004). While these
595 case studies provide examples of sweeps resulting from domestication, they also
596 show that the size of the affected region is highly variable, and we don't yet know
597 how this might impact patterns of deleterious variation. This is compounded by
598 the fact that even in highly researched species we don't always know the number
599 or location of genomic targets of selection during domestication (e.g. in maize;
600 Hufford *et al*. 2012), especially if any extended LD driven by artificial selection
601 has eroded or the intensity or mode of artificial selection has changed over time.
602
603 *What's "worse" in domesticated species: hitchhikers, drifters, or inbreds?*

604 We do not have a clear sense of which evolutionary processes contribute most to
605 the putative cost of domestication, and sometimes see contrasting patterns
606 across species. In maize, relatively few putatively deleterious alleles are shared
607 across all domesticated lines (hundreds vs. thousands; Mezmouk and Ross-
608 Ibarra 2014), which points to a larger role for the process of improvement than for
609 domestication in driving patterns of deleterious variation. In contrast, the increase
610 in dog dN/dS relative to wolf populations appears to not be driven by recent
611 inbreeding (i.e. improvement) but by the ancient domestication bottleneck
612 common to all dogs (Marsden *et al*. 2016).
613
614 Of the deleterious alleles segregating in more than 80% of maize lines, only 9.4%
615 show any signal of positive selection (Mezmouk and Ross-Ibarra 2014). This
616 suggests that hitchhiking during domestication played a relatively small role in
617 the increased frequency of deleterious variants in maize. The same study found
618 little support for enrichment of deleterious SNPs in areas of reduced
619 recombination (Mezmouk and Ross-Ibarra 2014). However, Rodgers-Melnick *et*
620 *al*. (2015) present contrasting evidence supporting enrichment of deleterious
621 variants in regions of low recombination, and the authors argue that this
622 difference is due to the use of a tool that does not rely on genome annotation
623 (Genomic Evolutionary Rate Profiling, or GERP; Cooper *et al*. 2005). This
624 complex narrative has emerged from just one well-studied domesticated species,
625 and it is likely that each species will present new and different complexities.
626
627 *Specific differences*
628 Examining general differences between wild and domesticated lineages ignores
629 species-specific demographic histories and changes in life history, which may be
630 important contributors to patterns of deleterious variation. Although we found
631 some general patterns (e.g. loss of genetic diversity, Table 1), we also see clear
632 exceptions (e.g. the Andean common bean). We can attribute these exceptions
633 to particular demographic scenarios (e.g. gene flow with Mesoamerican common
634 bean populations), assuming we have sufficient archeological, historical, or

genetic data. One clear problem is our inability to sample ancestral (pre-domestication) lineages, and our subsequent reliance on sampling of current wild relative lineages that have their own unique evolutionary trajectories. Sequencing ancient DNA can provide some insight into the history of these lineages and their ancestral states (e.g. in horses; Schubert *et al*. 2014). Currently, we know very little about most domesticated species' histories. We are still working towards understanding dynamics since domestication: even in highly-researched species like rice, the number of domestication events and subsequent demographic dynamics are hotly contested (Kovach, Sweeney, and McCouch 2007; Gao and Innan 2008; He *et al*. 2011; Molina *et al*. 2011; Huang *et al*. 2012; Gross and Zhao 2014; Civáň *et al*. 2015; Chen, Huang, and Han 2016; Choi *et al*. 2017).

A second challenge, briefly mentioned above, is understanding the relative importance of any one factor in any particular domestication event. Freedman, Lohmueller, and Wayne (2016) provide an in-depth review on this question in the domesticated dog, but it is unclear how general the relative contributions of selection and demography in dogs may be to other species. For example, in rice the shift to selfing from outcrossing during domestication appears to have played a larger role than the domestication bottleneck in shaping deleterious variation (Liu *et al*. 2017). This is useful in understanding rice domestication and its impact, but similar studies would need to be conducted across domesticated species to understand the generality of this dynamic. It is possible that general patterns may be drawn from subsets of domesticated species (e.g. vertebrates versus vascular plants, short-lived versus long-lived, outcrossing versus selfing versus clonally propagated etc.). For one, there may be general differences between annual and perennial crops, including less severe domestication bottlenecks and higher levels of gene flow from wild populations in perennials (Miller and Gross 2011; Gaut, Díez, and Morrell 2015).

Predictive algorithms

The identification of individual deleterious variants typically relies on sequence

666 conservation. If a variant occurs at a particular nucleotide site or encoded amino
667 acid is invariant across a phylogenetic comparison, then it is putatively
668 deleterious. More advanced approaches use estimates of synonymous
669 substitution rates at a locus to improve estimates of constraint on a nucleotide
670 site (Chun and Fay 2009). The majority of 'SNP annotation' approaches are
671 intended for the annotation of amino acid changing variants, although at least two
672 approaches (GERP++: Davydov *et al*. 2010; PHAST: Hubisz, Pollard, and Siepel
673 2010) can be applied to noncoding sequences when nucleotide sequences can
674 be aligned across species. This estimation of phylogenetic constraint is heavily
675 dependent on the sequence alignment; new annotation approaches have sought
676 to use more consistent sets of alignments across loci. Both GERP++ and MAPP
677 permit users to provide alignments for SNP annotation (Davydov *et al*. 2010;
678 Stone and Sidow 2005). The recently reported tool BAD_Mutations (Kono *et al*.
679 2016; Kono *et al*. bioRxiv) permits the use of a consistent set of alignments for
680 the annotation of deleterious variants by automating the download and alignment
681 of the coding portion of plant genomes from Phytozome and Ensembl Plants.
682 This currently includes 50+ sequenced angiosperm genomes (Goodstein *et al*.
683 2010; Kersey *et al*. 2016). Currently, the tool is configured for use with
684 angiosperms, but could be applied to other organisms.
685
686 Using phylogenetic conservation may be problematic in domesticated species, as
687 relaxed, balancing, or diversifying selection in agricultural environments could lift
688 constraint on sites under purifying or stabilizing selection in wild environments
689 (assuming some commonalities within these environment types). In one such
690 case, two AGPase small subunit paralogs in maize appear to be under
691 diversifying and balancing selection, respectively, even though these subunits
692 are likely under selective constraint across flowering plants (Georgelis, Shaw,
693 and Hannah 2009; Corbi *et al*. 2010). Similarly, broadly 'deleterious' traits may
694 have been under positive artificial selection in domesticated species. The loci
695 underlying these traits could be flagged as deleterious in bioinformatic screens
696 despite increasing fitness in the context of domestication. For example, the *fgf4*

697 retrogene insertion that causes chrondrodysplasia (short-leggedness) in dogs
698 would likely have a strongly deleterious effect in wolves, but has been positively
699 selected in some breeds of dog (Parker *et al*. 2009). Finally, bioinformatic
700 approaches that rely on phylogenetic conservation are likely to miss variants with
701 effects that are only deleterious in specific environmental or genomic contexts
702 (plastic or epistatic effects), or which reduce fitness specifically in agronomic or
703 breeding contexts. Specific knowledge of the phenotypic effects of putatively
704 deleterious mutations is necessary to address these issues, but as we discuss
705 below these data are challenging to obtain.
706
707 Information from the site frequency spectrum, or the number of times individual
708 variants are observed in a sample, can provide additional information about
709 which variants are most likely to be deleterious. In resequencing data from many
710 species, nonsynonymous variants typically occur at lower average frequencies
711 than synonymous variants (see Nordborg *et al*. 2005; Ross-Ibarra *et al*. 2009;
712 Günther and Schmid 2010). Mutations that are annotated as deleterious are
713 particularly likely to occur at lower frequencies than other classes of variants
714 (Marth *et al*. 2011; Kono *et al*. 2016; Liu *et al*. 2017) and may be less likely to be
715 shared among populations (Marth *et al*. 2011).
716
717 Tools for the annotation of potentially deleterious variants continue to be
718 developed rapidly (see Grimm *et al*. 2015 for a recent comparison). This includes
719 many tools that attempt to make use of information beyond sequence
720 conservation, including potential effects of variants on protein structure or
721 function (Adzhubei *et al*. 2010) or a diversity of genomic information intended to
722 improve prediction of pathogenicity in humans (Kircher *et al*. 2014). The majority
723 of SNP annotation tools are designed to work on human data and may not be
724 applicable to other organisms (Kono *et al*. *bioRxiv*). There is the potential for
725 circularity when an annotation tool is trained on the basis of pathogenic variants
726 in humans and then evaluated on the basis of a potentially overlapping set of
727 variants (Grimm *et al*. 2015). Even given these limitations, validation outside of

728 humans is more challenging because of a paucity of known phenotype-changing
729 variants. To address this issue, Kono *et al*. (*bioRxiv*) report a comparison of
730 seven annotation tools applied to a set of 2,910 phenotype-changing variants in
731 the model plant species *Arabidopsis thaliana*. The authors find that all seven
732 tools more accurately identify phenotype-changing variants likely to be
733 deleterious in *Arabidopsis* (Kono *et al*. *bioRxiv*) than in humans (Grimm *et al.*
734 2015; Dong *et al.* 2015). Diversity estimates in *Arabidopsis thaliana* suggest a
735 slightly larger estimated N_e in *Arabidopsis* than humans (Cao *et al.* 2011). Given
736 the general principle that for variants with selective coefficients $s < 1 / (2N_e)$,
737 genetic drift will dominate over the action of selection on a variant, making
738 purifying selection less effective.
739
740 *No one is perfect, not even the reference*
741 Bioinformatic approaches can suffer from reference bias (Simons *et al*. 2014;
742 Kono *et al*. 2016; Liu *et al*. 2017). With reference-based read mapping, variants
743 are typically identified as differences from reference, then passed through a
744 series of filters to identify putatively deleterious variants. Most annotation
745 approaches focus on nonsynonymous differences from references. Because a
746 reference genome, particularly when based on an inbred, has no differences
747 from itself, the reference has no nonsynonymous variants to annotate and thus
748 appears free of deleterious variants. This issue can be addressed by identifying
749 all variants, including those where the reference is different from all other
750 samples, and defining the mutation as the change relative to an inferred
751 ancestral state (see Kono *et al*. 2016). A more concerning type of reference bias
752 is that individuals that are genetically more similar to the reference genome will
753 have fewer differences from reference and thus fewer variants that annotate as
754 deleterious. This pattern is observed by Mezmouk and Ross-Ibarra (2014) who
755 find fewer deleterious variants in the stiff-stalk population of maize (to which the
756 reference genome B73 belongs) than in other elite maize populations. Along
757 similar lines, because gene models are derived from the reference genome,
758 more closely related lines with more similar coding portions of genes will appear

759 to have fewer disruptions of coding sequence (Gan *et al*. 2011). Finally,
760 reference bias can contribute to under-calling of deleterious variants, either when
761 divergent haplotypes fail to properly align to the reference or if the reference is
762 included in a phylogenetic comparison. Variants that are detected as a difference
763 from reference may then be compared against the reference in an alignment
764 testing for conservation at a nucleotide. This conflates diversity within a species
765 with the phylogenetic divergence that is being tested in the alignment. In cases
766 where the reference genome contains the novel (or derived) variant compared to
767 the state in the related species, the presence of the reference variant in the
768 alignment will cause the site to appear less constrained. This last problem is
769 easily resolved by leaving the species being tested out of the phylogenetic
770 alignment used to annotate deleterious variants (as in Schubert *et al*. 2014; Henn
771 *et al*. 2016; Kono *et al*. 2016; Marsden *et al*. 2016).
772
773 *Effect size*
774 The evidence we present supports the cost of domestication hypothesis, namely
775 that domesticated lineages carry more or higher frequency deleterious variants
776 than their wild relatives (Table 1). However, the distribution of fitness effects of
777 variants is important with the regard to the total load within an individual or
778 population (Henn *et al*. 2015). In other words, it is the cumulative effect of the
779 variants carried by an individual that make up its mutational load, not simply what
780 proportion of those variants is 'deleterious'. As we describe above, current
781 bioinformatic approaches that rely on phylogenetic conservation may identify a
782 number of false positives (driven by new fitness optima under artificial selection)
783 or false negatives (with specific epistatic, dominance, or environmentally-
784 dependent effects). Functional and quantitative genetics approaches provide
785 means of assessing the phenotypic effect of genetic variants, but there are
786 practical considerations that limit the quantity of evaluations that can be
787 conducted.
788
789 One potential issue is that the phenotypic effect of a genetic variant often

790 depends on genomic background, through both dominance (interaction between
791 alleles at the same locus) and epistasis (interaction among alleles at different
792 loci). Evaluating the effect of these variants is consequently a complex task,
793 requiring the creation and evaluation of multiple classes of recombinant
794 genomes. Similarly, the environment is an important consideration in thinking
795 about effect size. As we saw with the yield under drought *qDTY1.1* allele in rice
796 (*Identifying dangerous hitchhikers*, above; Vikram *et al*. 2015), the effect of a
797 genetic variant can depend on developmental or assessment environment.
798 These kinds of variants might be expected to be involved in local adaptation in
799 wild populations, and so would not show up in a screen based on phylogenetic
800 conservation. Nevertheless, these variants may be important in breeding
801 programs that target general-purpose genotypes with, for example, high mean
802 and low variance yields. Identifying these alleles requires assessment in the
803 appropriate environment(s) and large assessment populations (as the power for
804 detecting genotype-phenotype correlations generally scales with the number of
805 genotypes assessed). Addressing effect size in the appropriate context(s)
806 therefore involves challenges of scale. Fortunately, new types of mapping
807 populations (e.g. MAGIC) and high-throughput phenotyping platforms that can
808 enable this work are increasingly available across domesticated systems. Given
809 these tools and the resultant data, we should soon be able to parameterize
810 genomic selection and similar models with putatively deleterious variants and test
811 their cumulative effects.
812
813 *Heterosis*
814 In systems with hybrid production, complementation of deleterious variants
815 between heterotic breeding pools may contribute substantially to heterosis (e.g.
816 in maize; Hufford *et al*. 2012; Mezmouk and Ross-Ibarra 2014; Yang *et al*.
817 bioRxiv). If this is broadly true, hybrid production may be an interesting solution
818 to the cost of domestication, as long as deleterious variants are private to
819 heterotic groups rather than fixed in domesticated species. Theory suggests that
820 the deleterious variants that contribute to heterosis between populations with low

821 levels of gene flow are likely to be of intermediate effect, and may not play a
822 large role in inbreeding depression (Whitlock, Ingvarsson, and Hatfield 2000).
823 Since fitness in these contexts is largely evaluated in hybrid individuals rather
824 than inbred parents, parental populations are likely to retain a higher proportion
825 of slightly and moderately deleterious variants than even selfing populations
826 (Figure 2A). As long as hybrid crosses are the primary mode of seed production,
827 these alleles may not be of high importance to breeders even though they
828 contribute to deleterious variation more broadly. Evaluating effect size in this
829 context requires assessing both parental and hybrid populations, and is therefore
830 that much more difficult. However, assuming that complementation of deleterious
831 variants is a substantial component of heterosis, quantifying these variants
832 should improve the ability of breeding programs to predict trait values in hybrid
833 crosses. Further, the question of whether deleterious variation limits genetic
834 gains in hybrid production systems remains open.
835
836 CONCLUSIONS
837 Research on the putative costs of domestication is still relatively new, and there
838 remain many open questions. However, our review of the literature suggests that
839 deleterious variants are generally more numerous or frequent in domesticated
840 species compared to their wild relatives. This pattern is likely driven by a number
841 of processes that collectively act to reduce the efficacy of selection relative to
842 drift in domesticated populations, resulting in increased frequency of deleterious
843 variants linked to selected loci and greater accumulation of deleterious variants
844 genome-wide. We encourage further research across domesticated species on
845 these processes, and recommend that researchers: (1) Sample domesticated
846 and wild lineages sufficiently to assess diversity within as well as between these
847 groups and (2) present deleterious variant data per genome and as proportional
848 as well as absolute values. We also strongly encourage researchers and
849 breeders to think about deleterious variation in context, both genomic and
850 environmental. Finally, we think collecting empirical data on the effect sizes and
851 conditional dependence of putatively deleterious variants is increasingly feasible,

852 and would contribute greatly to this field.

853

854 FUNDING
855 This work was supported by the US National Science Foundation (award
856 numbers 1523752 to B.T.M.; and DBI 1339393 to P.L.M.).
857
858 ACKNOWLEDGEMENTS
859 We thank Greg Baute, Thomas Kono, Kathryn Turner, Jeffrey Ross-Ibarra, and
860 two anonymous and thoughtful reviewers for comments on an earlier version of
861 this manuscript, and Brandon Gaut, Loren Rieseberg, and Greg Owens for
862 helpful discussion.
863

864 REFERENCES

865 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, *et al.*
866 (2010). A method and server for predicting damaging missense mutations.
867 *Nat Methods* **7**: 248–249.

868 Agrawal AF, Whitlock MC (2012). Mutation load: the fitness of individuals in
869 populations where deleterious alleles are abundant. *Annu Rev Ecol Evol Syst*
870 **43**: 115–135.

871 Albalat R, Cañestro C (2016). Evolution by gene loss. *Nat Rev Genet* **17**: 379–
872 391.

873 Allard RW (1999). History of plant population genetics. *Annu Rev Genet* **33**: 1–
874 27.

875 Amaral AJ, Megens HJ, Crooijmans RPMA, Heuven HCM, Groenen MAM
876 (2008). Linkage disequilibrium decay and haplotype block structure in the pig.
877 *Genetics* **179**: 569–579.

878 Arunkumar R, Ness RW, Wright SI, Barrett SCH (2015). The evolution of selfing
879 is accompanied by reduced efficacy of selection and purging of deleterious
880 mutations. *Genetics* **199**: 817–829.

881 Asano K, Yamasaki M, Takuno S, Miura K, Katagiri S, Ito T, *et al.* (2011).
882 Artificial selection for a green revolution gene during japonica rice
883 domestication. *PNAS* **108**: 11034–11039.

884 Assaf ZJ, Petrov DA, Blundell JR (2015). Obstruction of adaptation in diploids by
885 recessive, strongly deleterious alleles. *PNAS* **112**: E2658–E2666.

886 Bachlava E, Tang S, Pizarro G, Schuppert GF, Brunick RK, Draeger D, *et al.*
887 (2009). Pleiotropy of the branching locus (B) masks linked and unlinked
888 quantitative trait loci affecting seed traits in sunflower. *Theor Appl Genet* **120**:
889 829–842.

890 Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, *et al.* (2017). The
891 sunflower genome provides insights into oil metabolism, flowering and
892 Asterid evolution. *Nature* **546**: 148–152.

893 Baer CF, Miyamoto MM, Denver DR (2007). Mutation rate variation in
894 multicellular eukaryotes: causes and consequences. *Nat Rev Genet* **8**: 619–
895 631.

896 Benbrook C (1999). *Evidence of the magnitude and consequences of the*
897 *Roundup Ready soybean yield drag from university-based varietal trials in*
898 *1998*. Ag BioTech InfoNet.

899 Bosse M, Megens H-J, Madsen O, Crooijmans RPMA, Ryder OA, Austerlitz F, *et*
900 *al.* (2015). Using genome-wide measures of coancestry to maintain diversity
901 and fitness in endangered and domestic pig populations. *Genome Research*
902 **25**: 970–981.

903 Bosse M, Megens H-J, Madsen O, Paudel Y, Frantz LAF, Schook LB, *et al.*
904 (2012). Regions of homozygosity in the porcine genome: consequence of
905 demography and the recombination landscape. *PLoS Genet* **8**: e1003100.

906 Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD,
907 Lohmueller KE, *et al.* (2008). Assessing the Evolutionary Impact of Amino
908 Acid Mutations in the Human Genome (MH Schierup, Ed.). *PLoS Genet* **4**:
909 e1000083.

910 Brandenburg J-T, Mary-Huard T, Rigaill G, Hearne SJ, Corti H, Joets J, *et al.*
911 (2017). Independent introductions and admixtures have contributed to
912 adaptation of European maize and its American counterparts. *PLoS Genet*
913 **13**: e1006666.

914 Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, *et al.* (2011).
915 Whole-genome sequencing of multiple *Arabidopsis* thaliana populations. *Nat*
916 *Genet* **43**: 956–963.

917 Charlesworth B (2009). Fundamental concepts in genetics: Effective population
918 size and patterns of molecular evolution and variation. *Nat Rev Genet* **10**:
919 195–205.

920 Charlesworth D (2003). Effects of inbreeding on the genetic diversity of
921 populations. *Phil Trans Roy Soc B* **358**: 1051–1070.

922 Chen E, Huang X, Han B (2016). How can rice genetics benefit from rice-
923 domestication study? *National Science Review* **3**: 278–280.

924 Choi JY, Platts AE, Fuller DQ, Hsing Y-I, Wing RA, Purugganan MD (2017). The
925 rice paradox: Multiple origins but single domestication in Asian rice. *Mol Biol*
926 *Evol*: msx049.

927 Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012). Predicting the
928 Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE* **7**:
929 e46688.

930 Chun S, Fay JC (2009). Identification of deleterious mutations within three
931 human genomes. *Genome Research* **19**: 1553–1561.

932 Civáň P, Craig H, Cox CJ, Brown TA (2015). Three geographically separate
933 domestications of Asian rice. *NPLANTS* **1**: 15164.

934 Clark RM, Linton E, Messing J, Doebley JF (2004). Pattern of diversity in the
935 genomic region near the maize domestication gene *tb1*. *PNAS* **101**: 700–707.

936 Consortium TCSAA (2005). Initial sequence of the chimpanzee genome and
937 comparison with the human genome. *Nature* **437**: 69–87.

938 Consortium TIBGS (2013). A physical, genetic and functional sequence
939 assembly of the barley genome. *Nature* **491**: 711–716.

940 Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A (2005).
941 Distribution and intensity of constraint in mammalian genomic sequence.
942 *Genome Research* **15**: 901–913.

943 Corbi J, Debieu M, Rousselet A, Montalent P, Le Guilloux M, Manicacci D, *et al.*
944 (2010). Contrasted patterns of selection since maize domestication on
945 duplicated genes encoding a starch pathway enzyme. *Theor Appl Genet* **122**:
946 705–722.

947 Cruz F, Vila C, Webster MT (2008). The legacy of domestication: accumulation of
948 deleterious mutations in the dog genome. *Mol Biol Evol* **25**: 2331–2336.

949 Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S (2010).
950 Identifying a high fraction of the human genome to be under selective
951 constraint using GERP. *PLoS Comput Biol* **6**: e1001025.

952 Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, *et al.* (2015).
953 Comparison and integration of deleteriousness prediction methods for
954 nonsynonymous SNVs in whole exome sequencing studies. *Human*
955 *Molecular Genetics* **24**: 2125–2137.

956 Elmore RW, Roeth FW, Nelson LA (2001). Glyphosate-resistant soybean cultivar
957 yields compared with sister lines. *Agron J* **93**: 408–412.

958 Eyre-Walker A, Keightley PD (2007). The distribution of fitness effects of new
959 mutations. *Nat Rev Genet* **8**: 610–618.

960 Eyre-Walker A, Woolfit M, Phelps T (2006). The distribution of fitness effects of
961 new deleterious amino acid mutations in humans. *Genetics* **173**: 891–900.

962 Fay JC, Wyckoff GJ, Wu C-I (2001). Positive and Negative Selection on the
963 Human Genome. *Genetics* **158**: 1227–1234.

964 Felsenstein J (1974). The evolutionary advantage of recombination. *Genetics* **78**:
965 737–756.

966 Flowers JM, Molina J, Rubinstein S, Huang P, Schaal BA, Purugganan MD
967 (2012). Natural selection in gene-dense regions shapes the genomic pattern
968 of polymorphism in wild and domesticated rice. *Mol Biol Evol* **29**: 675–687.

969 Freedman AH, Lohmueller KE, Wayne RK (2016). Evolutionary History, Selective
970 Sweeps, and Deleterious Variation in the Dog. *Annu Rev Ecol Evol Syst* **47**:
971 73–96.

972 Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, *et al.*
973 (2002). The structure of haplotype blocks in the human genome. *Science*
974 **296**: 2225–2229.

975 Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, *et al.* (2011).
976 Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*.
977 *Nature* **477**: 419–423.

978 Gao L-Z, Innan H (2008). Nonindependent Domestication of the Two Rice
979 Subspecies, Oryza sativassp. indicaand ssp. japonica, Demonstrated by
980 Multilocus Microsatellites. *Genetics* **179**: 965–976.

981 Gaut BS, Díez CM, Morrell PL (2015). Genomics and the contrasting dynamics of
982 annual and perennial domestication. *TIG* **31**: 709–719.

983 Georgelis N, Shaw JR, Hannah LC (2009). Phylogenetic analysis of ADP-
984 glucose pyrophosphorylase subunits reveals a role of subunit interfaces in
985 the allosteric properties of the enzyme. *Plant Physiol* **151**: 67–77.

986 Gheyas AA, Boschiero C, Eory L, Ralph H, Kuo R, Woolliams JA, *et al.* (2015).
987 Functional classification of 15 million SNPs detected from diverse chicken
988 populations. *DNA Research* **22**: 205–217.

989 Gillespie JH (1994). Substitution processes in molecular evolution. II.
990 Exchangeable models from population genetics. *Genetics* **138**: 943–952.

991 Gillespie JH (1999). The role of population size in molecular evolution. *Theor Pop*
992 *Biol* **55**: 145–156.

993 Giorgi D, Pandozy G, Farina A, Grosso V, Lucretti S, Gennaro A, *et al.* (2016).
994 First detailed karyo-morphological analysis and molecular cytological study of
995 leafy cardoon and globe artichoke, two multi-use Asteraceae crops. *CCG* **10**:
996 447–463.

997 Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, *et al.* (2011).
998 Phytozome: a comparative platform for green plant genomics. *Nucleic Acids*
999 *Res* **40**: D1178–D1186.

1000 Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, *et al.* (2009). A
1001 First-Generation Haplotype Map of Maize. *Science* **326**: 1115–1117.

1002 Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, Zhu L, *et al.*
1003 (2009). Linkage Disequilibrium and Demographic History of Wild and
1004 Domestic Canids. *Genetics* **181**: 1493–1505.

1005 Gregory TR (2017). Animal Genome Size Database.
1006 http://www.genomesize.com.

1007 Grimm DG, Azencott C-A, Aicheler F, Gieraths U, MacArthur DG, Samocha KE,
1008 *et al.* (2015). The Evaluation of Tools Used to Predict the Impact of Missense
1009 Variants Is Hindered by Two Types of Circularity. *Human Mutation* **36**: 513–
1010 523.

1011 Groenen MAM, Wahlberg P, Foglio M, Cheng HH, Megens HJ, Crooijmans
1012 RPMA, *et al.* (2008). A high-density SNP-based linkage map of the chicken
1013 genome reveals sequence features correlated with recombination rate.
1014 *Genome Research* **19**: 510–519.

1015 Gross BL, Olsen KM (2010). Genetic perspectives on crop domestication. *Trends*
1016 *Plant Sci* **15**: 529–537.

1017 Gross BL, Zhao Z (2014). Archaeological and genetic insights into the origins of
1018 domesticated rice. *PNAS* **111**: 6190–6197.

1019 Günther T, Schmid KJ (2010). Deleterious amino acid polymorphisms in
1020 *Arabidopsis thaliana* and rice. *Theor Appl Genet* **121**: 157–168.

1021 Haldane J (1937). The effect of variation of fitness. *Am Nat* **71**: 337–349.

1022 Hartfield M, Otto SP (2011). Recombination and Hitchiking of Deleterious Alleles.
1023 *Evolution* **65**: 2421–2434.

1024 He Z, Zhai W, Wen H, Tang T, Wang Y, Lu X, *et al.* (2011). Two Evolutionary
1025 Histories in the Genome of Rice: the Roles of Domestication Genes (R
1026 Mauricio, Ed.). *PLoS Genet* **7**: e1002100.

1027 Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S (2015). Estimating
1028 the mutation load in human genomes. *Nat Rev Genet* **16**: 333–343.

1029 Henn BM, Botigué LR, Peischl S, Dupanloup I, Lipatov M, Maples BK, *et al.*
1030 (2016). Distance from sub-Saharan Africa predicts mutational load in diverse
1031 human genomes. *PNAS* **113**: E440–E449.

1032 Hill WG, Robertson A (1966). The effect of linkage on limits to artificial selection.
1033 *Genetics Research* **8**: 269–294.

1034 Horak MJ, Rosenbaum EW, Kendrick DL, Sammons B, Phillips SL, Nickson TE,
1035 *et al.* (2015). Plant characterization of Roundup Ready 2 Yield® soybean,
1036 MON 89788, for use in ecological risk assessment. *Transgenic Res* **24**: 213–
1037 225.

1038 Huang X, Kurata N, Wei X, Wang Z-X, Wang A, Zhao Q, *et al.* (2012). A map of
1039 rice genome variation reveals the origin of cultivated rice. *Nature* **490**: 497–
1040 501.

1041 Hubisz MJ, Pollard KS, Siepel A (2011). PHAST and RPHAST: phylogenetic
1042 analysis with space/time models. *Briefings in Bioinformatics* **12**: 41–51.

1043 Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia J-M, Cartwright RA, *et*
1044 *al.* (2012). Comparative population genomics of maize domestication and
1045 improvement. *Nat Genet* **44**: 808–811.

1046 Hughes AL (2005). Evidence for Abundant Slightly Deleterious Polymorphisms in
1047 Bacterial Populations. *Genetics* **169**: 533–538.

1048 Keightley PD, Eyre-Walker A (2007). Joint Inference of the Distribution of Fitness
1049 Effects of Deleterious Mutations and Population Demography Based on
1050 Nucleotide Polymorphism Frequencies. *Genetics* **177**: 2251–2261.

1051 Keightley PD, Lynch M (2003). Toward a realistic model of mutations affecting
1052 fitness. *Evolution* **57**: 683–685.

1053 Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, *et al.* (2016).
1054 Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids*
1055 *Res* **44**: D574–D580.

1056 Kim BY, Huber CD, Lohmueller KE (2017). Inference of the Distribution of
1057 Selection Coefficients for New Nonsynonymous Mutations Using Large
1058 Samples. *Genetics* **206**: 345–361.

1059 Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, *et al.* (2007).
1060 Recombination and linkage disequilibrium in Arabidopsis thaliana. *Nat Genet*
1061 **39**: 1151–1155.

1062 Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J (2014). A
1063 general framework for estimating the relative pathogenicity of human genetic
1064 variants. *Nat Genet* **46**: 310–315.

1065 Klopfstein S, Currat M, Excoffier L (2005). The Fate of Mutations Surfing on the
1066 Wave of a Range Expansion. *Mol Biol Evol* **23**: 482–490.

1067 Koenig D, Jiménez-Gómez JM, Kimura S, Fulop D, Chitwood DH, Headland LR,
1068 *et al.* (2013). Comparative transcriptomics reveals patterns of selection in
1069 domesticated and wild tomato. *PNAS* **110**: E2655–E2662.

1070 Kono TJY, Fu F, Mohammadi M, Hoffman PJ, Liu C, Stupar RM, *et al.* (2016).
1071 The Role of Deleterious Substitutions in Crop Genomes. *Mol Biol Evol* **33**:
1072 2307–2317.

1073 Kono TJY, Lei L, Shih C-H, Hoffman PJ, Morrell PL, Fay JC Comparative
1074 genomics approaches accurately predict deleterious variants in plants.
1075 *bioRxiv*.

1076 Kovach MJ, Sweeney MT, McCouch SR (2007). New insights into the history of
1077 rice domestication. *TIG* **23**: 578–587.

1078 Kryazhimskiy S, Plotkin JB (2008). The Population Genetics of dN/dS. *PLoS*
1079 *Genet* **4**: e1000304.

1080 Lam H-M, Xu X, Liu X, Chen W, Yang G, Wong F-L, *et al.* (2010). Resequencing
1081 of 31 wild and cultivated soybean genomes identifies patterns of genetic
1082 diversity and selection. *Nat Genet* **42**: 1053–1059.

1083 Lau AN, Peng L, Goto H, Chemnick L, Ryder OA, Makova KD (2009). Horse
1084 Domestication and Conservation Genetics of Przewalski's Horse Inferred
1085 from Sex Chromosomal and Autosomal Sequences. *Mol Biol Evol* **26**: 199–
1086 208.

1087 Lin T, Zhu G, Zhang J, Xu X, Yu Q, Zheng Z, *et al.* (2014). Genomic analyses
1088 provide insights into the history of tomato breeding. *Nat Genet* **46**: 1220–
1089 1226.

1090 Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, *et*
1091 *al.* (2005). Genome sequence, comparative analysis and haplotype structure
1092 of the domestic dog. *Nature* **438**: 803–819.

1093 Lippold S, Knapp M, Kuznetsova T, Leonard JA, Benecke N, Ludwig A, *et al.*
1094 (2011). Discovery of lost diversity of paternal horse lineages using ancient
1095 DNA. *Nat Commun* **2**: 450–6.

1096 Liu A, Burke JM (2006). Patterns of Nucleotide Diversity in Wild and Cultivated
1097 Sunflower. *Genetics* **173**: 321–330.

1098 Liu Q, Zhou Y, Morrell PL, Gaut BS (2017). Deleterious variants in Asian rice and
1099 the potential cost of domestication. *Mol Biol Evol*: msw296.

1100 Lohmueller KE (2014). The Impact of Population Demography and Selection on
1101 the Genetic Architecture of Complex Traits. *PLoS Genet* **10**: e1004379.

1102 Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, *et*
1103 *al.* (2008). Proportionally more deleterious genetic variation in European than
1104 in African populations. *Nature* **451**: 994–997.

1105 Lu J, Tang T, Tang H, Huang J, Shi S, Wu C-I (2006). The accumulation of
1106 deleterious mutations in rice genomes: a hypothesis on the cost of
1107 domestication. *TIG* **22**: 126–131.

1108 MacEachern S, McEwan J, McCulloch A, Mather A, Savin K, Goddard M (2009).
1109 Molecular evolution of the Bovini tribe (Bovidae, Bovinae): Is there evidence
1110 of rapid evolution or reduced selective constraint in Domestic cattle? *BMC*
1111 *Genomics* **10**: 179.

1112 Mandel JR, Dechaine JM, Marek LF, Burke JM (2011). Genetic diversity and
1113 population structure in cultivated sunflower and a comparison to its wild
1114 progenitor, *Helianthus annuus* L. *Theor Appl Genet* **123**: 693–704.

1115 Marsden CD, Ortega-Del Vecchyo D, O'Brien DP, Taylor JF, Ramirez O, Vilà C,
1116 *et al.* (2016). Bottlenecks and selective sweeps during domestication have
1117 increased deleterious genetic variation in dogs. *PNAS* **113**: 152–157.

1118 Marth GT, Yu F, Indap AR, Garimella K, Gravel S, Leong WF, *et al.* (2011). The
1119 functional spectrum of low-frequency coding variation. *Genome Biology* **12**:
1120 1–17.

1121 Maynard Smith J, Haigh J (1974). The hitch-hiking effect of a favourable gene.
1122 *Genetical research* **23**: 1–13.

1123 Megens H-J, Crooijmans RP, Bastiaansen JW, Kerstens HH, Coster A, Jalving
1124 R, *et al.* (2009). Comparison of linkage disequilibrium and haplotype diversity
1125 on macro- and microchromosomes in chicken. *BMC Genet* **10**: 86.

1126 Mezmouk S, Ross-Ibarra J (2014). The Pattern and Distribution of Deleterious
1127 Mutations in Maize. *G3* **4**: 163–171.

1128 Michaelson MJ, Price HJ, Ellison JR, Johnston JS (1991). Comparison of plant
1129 DNA contents determined by Feulgen microspectrophotometry and laser flow
1130 cytometry. *Am J Bot* **78**: 183–188.

1131 Miller AJ, Gross BL (2011). From forest to field: Perennial fruit crop
1132 domestication. *Am J Bot* **98**: 1389–1414.

1133 Miyata T, Miyazawa S, Yasunaga T (1979). Two types of amino acid
1134 substitutions in protein evolution. *J Molec Evol* **12**: 219–236.

1135 Molina J, Sikora M, Garud N, Flowers JM, Rubinstein S, Reynolds A, *et al.*
1136 (2011). Molecular evidence for a single evolutionary origin of domesticated
1137 rice. *PNAS* **108**: 8351–8356.

1138 Monroe JG, McGovern C, Lasky JR, Grogan K, Beck J, McKay JK (2016).
1139 Adaptation to warmer climates by parallel functional evolution of *CBF* genes
1140 in *Arabidopsis thaliana*. *Mol Ecol* **25**: 3632–3644.

1141 Moray C, Lanfear R, Bromham L (2014). Domestication and the mitochondrial
1142 genome: comparing patterns and rates of molecular evolution in
1143 domesticated mammals and birds and their wild relatives. *Genome Biol Evol*
1144 **6**: 161–169.

1145 Morrell PL, Buckler ES, Ross-Ibarra J (2011). Crop genomics: advances and
1146 applications. *Nat Rev Genet* **13**: 85–96.

1147 Muir WM, Wong G, Zhang Y, Wang J, Groenen MAM, Crooijmans RPMA, *et al.*
1148 (2008). Genome-wide assessment of worldwide chicken SNP genetic
1149 diversity indicates significant absence of rare alleles in commercial breeds.
1150 *PNAS* **105**: 17312–17317.

1151 Nabholz B, Sarah G, Sabot F, Ruiz M, Adam H, Nidelet S, *et al.* (2014).
1152 Transcriptome population genomics reveals severe bottleneck and
1153 domestication cost in the African rice (*Oryza glaberrima*). *Mol Ecol* **23**: 2210–
1154 2227.

1155 Ng PC, Henikoff S (2003). SIFT: predicting amino acid changes that affect
1156 protein function. *Nucleic Acids Res* **31**: 3812–3814.

1157 Nie J, Wang Y, He H, Guo C, Zhu W, Pan J, *et al.* (2015). Loss-of-Function
1158 Mutations in *CsMLO1* Confer Durable Powdery Mildew Resistance in
1159 Cucumber (*Cucumis sativus* L.). *Front Plant Sci* **6**: 30.

1160 Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, *et al.* (2005).
1161 The Pattern of Polymorphism in *Arabidopsis thaliana*. *PLoS Biol* **3**: e196.

1162 Ohta T (1972). Population size and rate of evolution. *J Molec Evol* **1**: 305–314.

1163 Ohta T (1992). The nearly neutral theory of molecular evolution. *Annu Rev Ecol*
1164 *Syst* **23**.

1165 Olsen KM (2006). Selection Under Domestication: Evidence for a Sweep in the
1166 Rice *Waxy* Genomic Region. *Genetics* **173**: 975–983.

1167 Otto SP, Barton NH (2001). Selection for recombination in small populations.
1168 *Evolution* **55**: 1921–1931.

1169 Otto SP, Whitlock MC (1997). The probability of fixation in populations of
1170 changing size. *Genetics* **146**: 723–733.

1171 Papa R, Bellucci E, Rossi M, Leonardi S, Rau D, Gepts P, *et al.* (2007). Tagging
1172 the Signatures of Domestication in Common Bean (*Phaseolus vulgaris*) by

1173 Means of Pooled DNA Samples. *Ann Bot–London* **100**: 1039–1051.

1174 Parker HG, vonHoldt BM, Quignon P, Margulies EH, Shao S, Mosher DS, *et al.*
1175 (2009). An expressed *Fgf4* retrogene is associated with breed-defining
1176 chondrodysplasia in domestic dogs. *Science* **325**: 995–998.

1177 Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L (2013). On the accumulation of
1178 deleterious mutations during range expansions. *Mol Ecol* **22**: 5972–5982.

1179 Pritchard JK, Przeworski M (2001). Linkage disequilibrium in humans: models
1180 and data. *The American Journal of Human Genetics* **69**: 1–14.

1181 Renaut S, Rieseberg LH (2015). The Accumulation of Deleterious Mutations as a
1182 Consequence of Domestication and Improvement in Sunflowers and Other
1183 Compositae Crops. *Mol Biol Evol* **32**: 2273–2283.

1184 Rice A, Glick L, Abadi S, Einhorn M (2015). The Chromosome Counts Database
1185 (CCDB)–a community resource of plant chromosome numbers. *New Phytol*
1186 **206**: 19–26.

1187 Rodgers-Melnick E, Bradbury PJ, Elshire RJ, Glaubitz JC, Acharya CB, Mitchell
1188 SE, *et al.* (2015). Recombination in diverse maize is stable, predictable, and
1189 associated with genetic load. *PNAS* **112**: 3823–3828.

1190 Ross-Ibarra J (2004). The evolution of recombination under domestication: a test
1191 of two hypotheses. *Am Nat* **163**: 105–112.

1192 Ross-Ibarra J, Tenaillon M, Gaut BS (2009). Historical Divergence and Gene
1193 Flow in the Genus *Zea*. *Genetics* **181**: 1399–1413.

1194 Rubin C-J, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, *et al.*
1195 (2010). Whole-genome resequencing reveals loci under selection during
1196 chicken domestication. *Nature* **464**: 587–591.

1197 Scaglione D, Reyes-Chin-Wo S, Acquadro A, Froenicke L, Portis E, Beitel C, *et*
1198 *al.* (2015). The genome sequence of the outbreeding globe artichoke
1199 constructed. *Sci Rep*: 1–17.

1200 Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, *et al.* (2012).
1201 Insights into hominid evolution from the gorilla genome sequence. *Nature*
1202 **483**: 169–175.

1203 Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, *et al.*
1204 (2014). A reference genome for common bean and genome-wide analysis of
1205 dual domestications. *Nat Genet* **46**: 707–713.

1206 Schneider CA, Rasband WS, Eliceiri KW (2012). NIH Image to ImageJ: 25 years
1207 of image analysis. *Nat Methods* **9**: 671–675.

1208 Schubert M, Jónsson H, Chang D, Sarkissian Der C, Ermini L, Ginolhac A, *et al.*
1209 (2014). Prehistoric genomes reveal the genetic foundation and cost of horse
1210 domestication. *PNAS* **111**: E5661–E5669.

1211 Semon M, Nielsen R, Jones MP, McCouch SR (2005). The Population Structure
1212 of African Cultivated Rice *Oryza glaberrima* (Steud.): Evidence for Elevated
1213 Levels of Linkage Disequilibrium Caused by Admixture with *O. sativa* and
1214 Ecological Adaptation. *Genetics* **169**: 1639–1647.

1215 Shi T, Dimitrov I, Zhang Y, Tax FE, Yi J, Gou X, *et al.* (2015). Accelerated rates
1216 of protein evolution in barley grain and pistil biased genes might be legacy of
1217 domestication. *Plant Molecular Biology* **89**: 253–261.

1218 Simons YB, Turchin MC, Pritchard JK, Sella G (2014). The deleterious mutation
1219 load is insensitive to recent population history. *Nat Genet* **46**: 220–224.

1220 Stone EA, Sidow A (2005). Physicochemical constraint violation by missense
1221 substitutions mediates impairment of protein function and disease severity.
1222 *Genome Research* **15**: 978–986.

1223 Sved JA (1971). Linkage disequilibrium and homozygosity of chromosome
1224 segments in finite populations. *Theor Pop Biol* **2**: 125–141.

1225 Takeda S, Matsuoka M (2008). Genetic approaches to crop improvement:
1226 responding to environmental and population changes. *Nat Rev Genet* **9**: 444–
1227 457.

1228 Tian F, Stevens NM, Buckler ES (2009). Tracking footprints of maize
1229 domestication and evidence for a massive selective sweep on chromosome
1230 10. *PNAS* **106**: 9979–9986.

1231 Traspov A, Deng W, Kostyunina O, Ji J, Shatokhin K, Lugovoy S, *et al.* (2016).
1232 Population structure and genome characterization of local pig breeds in
1233 Russia, Belorussia, Kazakhstan and Ukraine. *Genetics Selection Evolution*
1234 **48**: 1–9.

1235 Travis JMJ, Munkemuller T, Burton OJ, Best A, Dytham C, Johst K (2007).
1236 Deleterious Mutations Can Surf to High Densities on the Wave Front of an
1237 Expanding Population. *Mol Biol Evol* **24**: 2334–2343.

1238 Vikram P, Swamy BPM, Dixit S, Singh R, Singh BP, Miro B, *et al.* (2015).
1239 Drought susceptibility of modern rice varieties: an effect of linkage of drought
1240 tolerance with undesirable traits. *Sci Rep* **5**.

1241 Wade CM, Giulotto E, Sigurdsson S, Zoli M (2009). Genome Sequence,
1242 Comparative Analysis, and Population Genetics of the Domestic Horse.
1243 *Science* **326**: 865–867.

1244 Walsh B (2007). Using molecular markers for detecting domestication,
1245 improvement, and adaptation genes. *Euphytica* **161**: 1–17.

1246 Wang G-D, Xie H-B, Peng M-S, Irwin D, Zhang Y-P (2014). Domestication
1247 Genomics: Evidence from Animals. *Annu Rev Anim Biosci* **2**: 65–84.

1248 Wang H, Studer AJ, Zhao Q, Meeley R (2015). Evidence that the origin of naked
1249 kernels during maize domestication was caused by a single amino acid
1250 substitution in *tga1*. *Genetics* **200**: 965–974.

1251 Wang L, Beissinger TM, Lorant A, Ross-Ibarra C, Ross-Ibarra J, Hufford M The
1252 interplay of demography and selection during maize domestication and
1253 expansion. *bioRxiv*.

1254 Whitlock MC (2000). Fixation of new alleles and the extinction of small
1255 populations: drift load, beneficial alleles, and sexual selection. *Evolution* **54**:
1256 1855–1861.

1257 Whitlock MC, Ingvarsson PK, Hatfield T (2000). Local drift load and the heterosis
1258 of interconnected populations. *Heredity* **84**: 452–457.

1259 Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES (2002). Genetic
1260 diversity and selection in the maize starch pathway. *PNAS* **99**: 12959–12962.

1261 Wilfert L, Gadau J, Schmid-Hempel P (2007). Variation in genomic recombination
1262 rates among animal taxa and the case of social insects. *Heredity* **98**: 189–
1263 197.

1264 Wolf JBW, Kunstner A, Nam K, Jakobsson M, Ellegren H (2009). Nonlinear
1265 Dynamics of Nonsynonymous (dN) and Synonymous (dS) Substitution Rates
1266 Affects Inference of Selection. *Genome Biol Evol* **1**: 308–319.

1267 Woolfit M (2009). Effective population size and the rate and pattern of nucleotide
1268 substitutions. *Biology Letters* **5**: 417–420.

1269 Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, *et al.*
1270 (2005). The Effects of Artificial Selection on the Maize Genome. *Science* **308**:
1271 1310–1314.

1272 Xu Z, Hennessy DA, Sardana K, Moschini G (2013). The Realized Yield Effect of
1273 Genetically Engineered Crops: U.S. Maize and Soybean. *Crop Sci* **53**: 735–
1274 745.

1275 Yamasaki M, Wright SI, McMullen MD (2007). Genomic Screening for Artificial
1276 Selection during Domestication and Improvement in Maize. *Ann Bot–London*
1277 **100**: 967–973.

1278 Yang J, Mezmouk S, Baumgarten A, Buckler ES, Guill KE, McMullen MD, *et al.*

1279 Incomplete dominance of deleterious alleles contributes substantially to trait
1280 variation and heterosis in maize. *bioRxiv*.

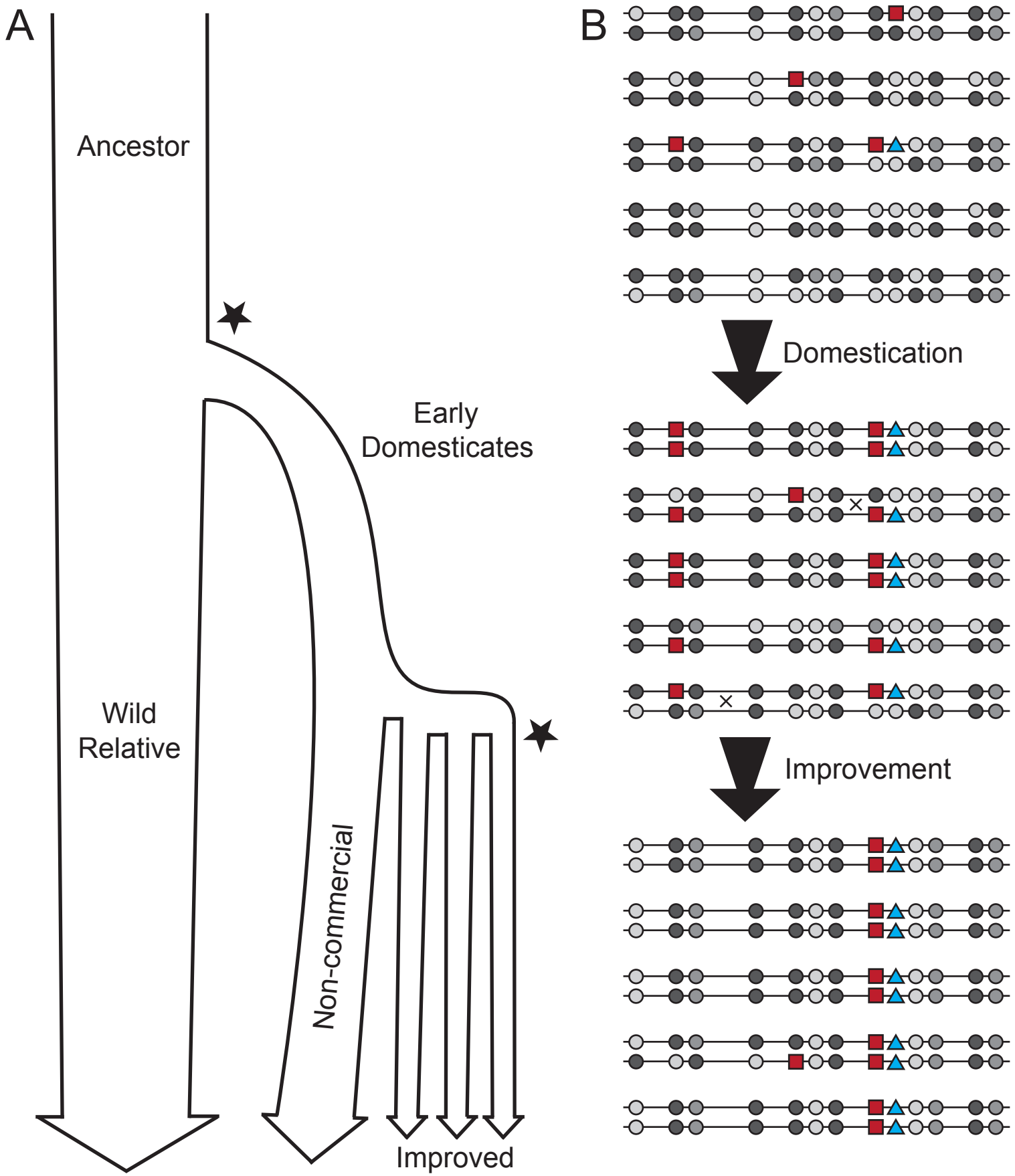
1281 Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, *et al.* (2015). Resequencing 302
1282 wild and cultivated accessions identifies genes related to domestication and
1283 improvement in soybean. *Nat Biotechnol* **33**: 408–414.

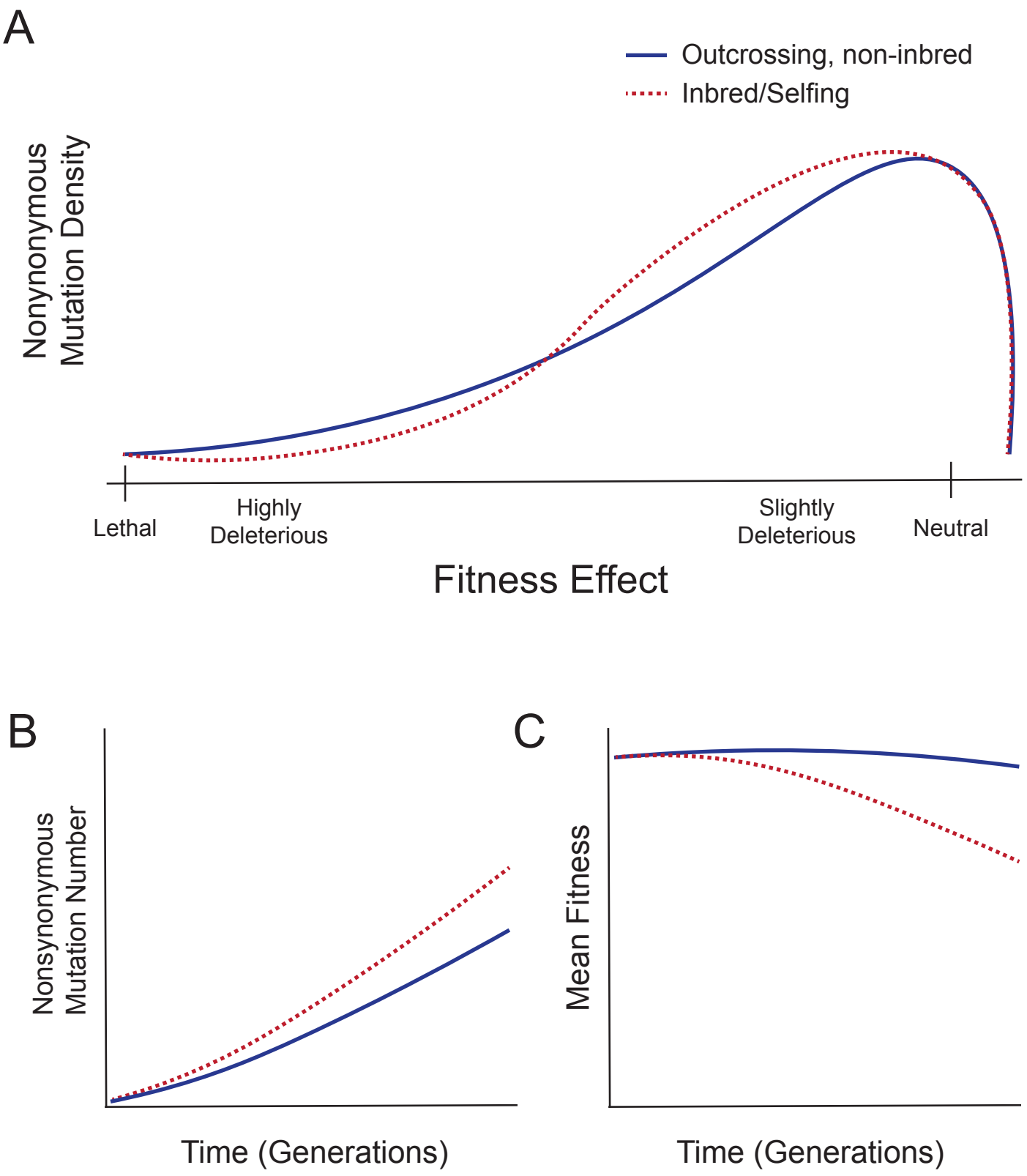
1284 Zonneveld BJM, Leitch IJ, Bennett MD (2005). First Nuclear DNA Amounts in
1285 more than 300 Angiosperms. *Ann Bot–London* **96**: 229–244.

1286
1287 TABLES AND FIGURES
1288
1289 Table 1. Evidence for a cost of domestication across domesticated plants and
1290 animals, with *Homo sapiens* included for comparison. The first three columns
1291 identify the taxa and major form of propagation and reproduction. Following
1292 columns indicate whether data from the domesticated lineage versus wild
1293 relatives or ancestors fits (+), is equivocal (=) or goes against (-) expectations for
1294 the 'cost of domestication' hypothesis, with those expectations being: extended
1295 linkage disequilibrium (LD), reduced genetic diversity (Diversity), higher rate or
1296 number of nonsynonymous to synonymous substitutions or variants (dN/dS,
1297 Ka/Ks, counts), more or higher frequency deleterious variants (Deleterious
1298 variants), or higher estimated genetic load (GERP score). The values for each of
1299 these and additional data are available in Supplementary Table 1.
1300
1301 Figure 1. Processes of domestication and improvement. (A) Typical changes in
1302 effective population size through domestication and improvement. Stars indicate
1303 genetic bottlenecks. These dynamics can be reconstructed by examining
1304 patterns of genetic diversity in contemporary wild relative, domesticated non-
1305 commercial, and improved populations. (B) Effects of artificial selection (targeting
1306 the blue triangle variant) and linkage disequilibrium on deleterious (red squares)
1307 and neutral variants (grey circles, shades represent different alleles). In the
1308 ancestral wild population, four deleterious alleles are at relatively low frequency
1309 (mean = 0.10) and heterozygosity is high (H_O = 0.51). After domestication, the
1310 selected blue triangle and linked variants increase in frequency (three remaining
1311 deleterious alleles, mean frequency = 0.46), heterozygosity decreases (H_O =
1312 0.35), and allelic diversity is lost at two sites. Recombination may change
1313 haplotypes, especially at sites less closely linked to the selected allele (Xs). After
1314 improvement, further selection for the blue triangle allele has: lowered
1315 heterozygosity (H_O = 0.08), increased deleterious variant frequency (two
1316 remaining deleterious alleles, mean = 0.55), and lost allelic diversity at six
1317 additional sites.
1318
1319 Figure 2. Effects of inbreeding on mutations and fitness. (A) Theoretical density
1320 plot of fitness effects for segregating variants. In outcrossing, non-inbred
1321 populations (solid navy line), more variants with highly deleterious, recessive
1322 effects persist that are rapidly exposed to selection and purged in inbred

1323 populations (dotted red line), shifting the left side of the distribution towards more
1324 neutral effects. At the same time, the reduced effective population size created
1325 through inbreeding causes on average higher loss of slightly advantageous
1326 mutations and retention of slightly deleterious mutations, shifting the right side of
1327 the distribution towards more deleterious effects. (B) Accumulation of
1328 nonsynonymous mutations is accelerated in selfing individuals (dotted red line)
1329 relative to outcrossing individuals (solid navy line). Synonymous mutations
1330 accumulate at similar rates in both populations. Adapted from simulation results
1331 in Arunkumar *et al*. 2015. (C) As a consequence of (B), mean individual fitness
1332 drops in selfing relative to outcrossing populations. Adapted from Arunkumar *et*
1333 *al*. 2015.
1334
1335 Supplementary Table 1. Quantitative version of Table 1: evidence for a cost of
1336 domestication across domesticated plants and animals, with *Arabidopsis thaliana*
1337 and *Homo sapiens* included for comparison. Plant 1C genome sizes from the
1338 RBG Kew database (http://data.kew.org/cvalues/), except tomato (Michaelson *et*
1339 *al*. 1991) and *Cynara* spp. (Giorgi *et al*. 2016). Plant chromosome counts from
1340 the Chromosome Counts Database (http://ccdb.tau.ac.il/). Animal 1C genome
1341 sizes and chromosome counts from the Animal Genome Size Database
1342 (http://www.genomesize.com/, mean value when multiple records were
1343 available). Gene numbers are high-confidence (if available) estimates from the
1344 vertebrate and plant Ensembl databases (http://uswest.ensembl.org/;
1345 http://plants.ensembl.org/), except common bean (Schmutz et al. 2013),
1346 sunflower (Compositate Genome Project, unpublished), and *Cynara* spp.
1347 (Scaglione et al. 2016). LD N50 is the approximate distance over which LD
1348 decays to half of maximum value. Loss of genetic diversity is calculated as 1 –
1349 (ratio of $p_{domesticated}$ to p_{wild}), or q if p not available.

Species	Common name	Primary reproduction	LD	Diversity	dN/dS or Ka/Ks	Deleterious variants	GERP score	Studies
Oryza sativa var. *japonica*	japonica rice	selfing	+	+	+	+	NA	Lu *et al*. 2006; Huang *et al*. 2012; Liu *et al*. 2017
Oryza sativa var. *indica*	indica rice	selfing	+	+	+	+	NA	Lu *et al*. 2006; Huang *et al*. 2012; Liu *et al*. 2017
Oryza glaberrima	African rice	selfing	NA	+	+	NA	NA	Semon *et al*. 2005; Nabholz *et al*. 2014
Zea mays	maize	outcrossing	NA	+	NA	NA	NA	Gore *et al*. 2009; Hufford *et al*. 2012; Mezmouk and Ross-Ibarra 2014
Hordeum vulgare	barley	selfing	+	+	+	NA	NA	The International Barley Genome Sequencing Consortium 2012; Morrell *et al*. 2014; Shi *et al*. 2015; Kono *et al*. 2016
Glycine max	soybean	selfing	+	+	=	NA	NA	Lam *et a*l. 2010; Zhou *et al*. 2015; Kono *et al*. 2016
Phaseolus vulgaris, Mesoamerican domestication	common bean	selfing	NA	+	NA	NA	NA	Schmutz *et al*. 2013
Phaseolus vulgaris, Andean domestication	common bean	selfing	NA	-	NA	NA	NA	Schmutz *et al*. 2013
Solanum lycopersicum	tomato	selfing	+	+	+	NA	NA	Koenig *et al*. 2013; Lin *et al*. 2014
Helianthus annuus	sunflower	outcrossing	+	+	NA	+	NA	Liu and Burke 2006; Mandel *et al*. 2011; Renaut and Rieseberg 2015
Cynara cardunculus var. *scolymus*	globe artichoke	clonal/outcrossing	NA	NA	NA	+	NA	Renaut and Rieseberg 2015
Cynara cardunculus var. *altilis*	cardoon	outcrossing	NA	NA	NA	+	NA	Renaut and Rieseberg 2015
Gallus gallus domesticus	chicken	outcrossing	+	+	NA	NA	NA	Muir *et al*. 2008; Megens *et al*. 2009; Gheyas *et al*. 2015
Canis familiaris	dog	outcrossing	+	+	+	+	+	Lindblad-Tor *et al.* 2005; Cruz, Vila & Webster 2008; Gray *et al*. 2009; Marsden *et al*. 2016
Equus caballus	horse	outcrossing	NA	+	NA	NA	+	Lau *et al*. 2009; Wade *et al*. 2009; Schubert *et al*. 2014
Sus scrofa	pig	outcrossing	+	+	NA	NA	NA	Amaral *et al*. 2008; Bosse *et al*. 2012; Bosse *et al*. 2015
Homo sapiens	human	outcrossing	NA	NA	=	NA	NA	Gabriel *et al*. 2002; Chun and Fay 2009; Scally *et al.* 2012


A
Ancestor
Early
Domesticates
Wild
Relative
Non-commercial
Improved
B
Domestication
Improvement


A
Outcrossing, non-inbred
Inbred/Selfing
Nonynonymous Mutation Density
Lethal
Highly Deleterious
Slightly Deleterious
Neutral
Fitness Effect
B
Nonsynonymous Mutation Number
Time (Generations)
C
Mean Fitness
Time (Generations)