Trends in Genetics

Review

Genomics and the Contrasting Dynamics of Annual and Perennial Domestication

Brandon S. Gaut,[1,*] Concepción M. Díez,[2] and Peter L. Morrell[3]

Plant domestication modifies a wild species genetically for human use. Among thousands of domesticated plants, a major distinction is the difference between annual and perennial life cycles. The domestication of perennials is expected to follow different processes than annuals, with distinct genetic outcomes. Here we examine domestication from a population genetics perspective, with a focus on three issues: genetic bottlenecks during domestication, introgression as a source of local adaptation, and genetic load. These three issues have been studied nominally in major annual crops but even less extensively in perennials. Here we highlight lessons from annual plants, motivations to study these issues in perennial plants, and new approaches that may lead to further progress.

Trends

The study of a handful of annual crops has heavily influenced our understanding of the genetics of domestication. In contrast to annuals, the domestication of perennials has generally occurred more recently and involved clonal propagation.

Annual and perennial domesticates appear to differ in the severity of domestication bottlenecks. New approaches to studying demographic history promise to yield fresh insights into these effects, particularly for perennials.

Both annuals and perennials are understudied with respect to identifying the genetic basis of local adaptation and the contribution of introgression to modern cultigens. Recent studies suggest that wild-to-crop introgression has been a particularly important feature of the evolution of perennial crops.

The 'cost of domestication', in terms of increased genetic load, is an area of growing interest. The magnitude and effect of genetic load remains uncharacterized for perennial plants, where an important contributor to load may be somatic mutations among clonal variants.

Introduction

Plant domestication is the genetic modification of a wild species to create phenotypes altered for human needs [1]. For many major crops, domestication occurred ∼10 000 to ∼12 000 years ago, when it correlated with a shift from hunter–gatherer to agrarian societies [2]. However, the domestication of crops is an ongoing process, and crops like cranberries and pecans have been domesticated as recently as the last ∼150 to ∼200 years [3,4]. Altogether, ∼2500 plant species have been either partially or fully domesticated. These species range from the 'Big Three' – maize, rice, and wheat, which together provided over 35% of calories for human consumption in 2010 (http://faostat.fao.com/stats) – to other important global staples such as barley, soybean, sorghum, and yams, to crops of limited use and geographic distribution like cherimoya (*Annona cherimola* Mill) and starfruit (*Averrhoa carambola* L.).

Among the array of domesticated species, one major distinction is between annual and perennial life cycles. The most important staple crops, including the Big Three, are annuals, but perennial crops are also critically important; they include food and oil staples like apples, grapes, olive, palms, and tree nuts and account for approximately one-eighth of the world's total food-producing surface [5]. Importantly, it appears that the domestication process differs dramatically between annuals and long-lived perennials, especially tree crops. Compared with annuals, perennial crops tend to have been domesticated more recently, are typically outcrossing, have extended juvenile periods, and are propagated clonally [4,5]. While the dynamics of domestication in perennials differs markedly from that of annuals, the genetics of domestication of perennials has been studied in less depth. Indeed, historically these species have been considered intractable due to their long generation times and high rates of outcrossing [3].

In this review we focus on the process and genetic consequences of plant domestication, with a particular emphasis on potential differences between annual and perennial crops. To date, genomic approaches – and particularly population-genetic approaches – have been applied

[1] Department of Ecology and Evolutionary Biology, University of California, Irvine, Irvine, CA 92697, USA
[2] Departamento de Agronomía, Universidad de Córdoba – Campus de Excelencia Internacional Agroalimentario ceiA3, Edificio C4, Campus de Rabanales, 14014 Córdoba, Spain

http://dx.doi.org/10.1016/j.tig.2015.10.002

widely to annual plants but not to perennials. There is still much to learn about the genetic patterns and consequences of the domestication of long-lived perennials.

The Process of Domestication

In theory, domestication of annuals is a straightforward process [6] (Figure 1, Key Figure). It begins with an enhanced human awareness of a wild plant. Once the species is recognized as a food source, it may be gathered and brought to a camp or village and left as refuse in a midden, perhaps initiating cultivation before explicit domestication. From there, conscious and unconscious selection for agronomic traits [7] – like photoperiod sensitivity, increased yield, increased fruit or grain size, and loss of seed dormancy – leads to genetic modification. Once domesticated, a successful crop is dispersed from its center of domestication to new geographic locales. In these new locations, the crop must adapt to local conditions, through selection acting to improve local adaptation of the domesticate and sometimes by hybridization with locally adapted wild populations.

This model of domestication holds reasonably well for many annual domesticates. Maize, for example, has been domesticated only once, ~9000 years ago in lowland Mexico [8]. After domestication, it spread rapidly throughout the Americas [9] and the process of local adaptation – for example, to higher-altitude growing areas in Mexico – may have included post-domestication hybridization with local, wild populations [10–12].

Nonetheless, the domestication process can vary widely, even among seed-propagated annuals. For example, barley shows evidence of at least two domestications [13] – once in the Fertile Crescent and again east of the Zagros Mountains [14] – and the genomic composition of barley landraces reflects contributions from distinct wild barley populations [15]. Other examples of annual crops resulting from multiple domestications include common bean [16], sorghum [17], and rice. Rice has been domesticated at least twice, in Asia [18,19] and in Africa [20].

[3]Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN 55108, USA

*Correspondence: bgaut@uci.edu (B.S. Gaut).

Key Figure

The Domestication Process in Annuals and Perennials

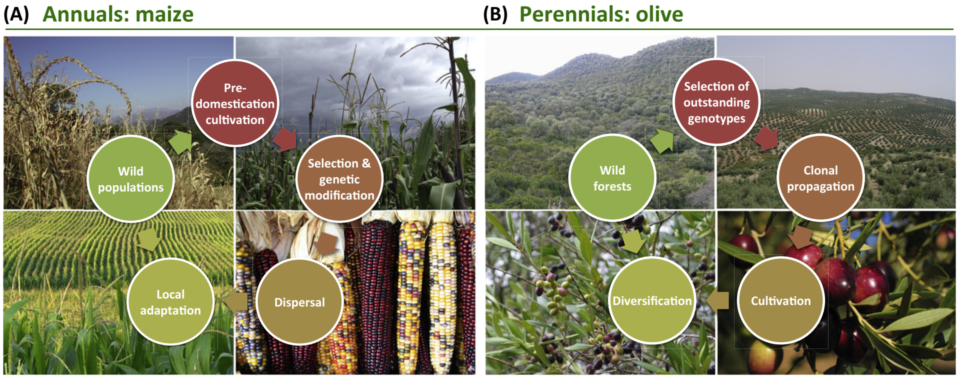
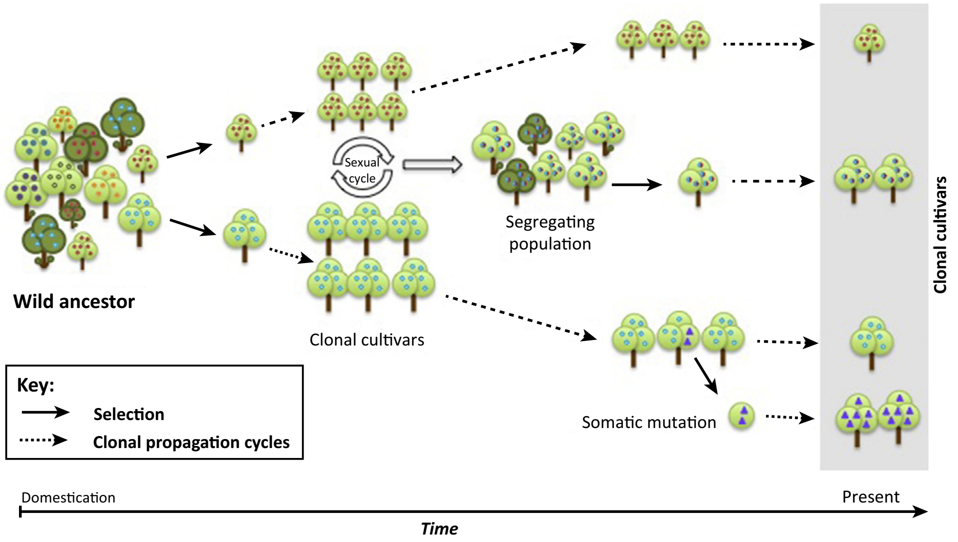


Figure 1. (A) The process of domestication for an annual like maize, modeled after [6]. (B) The process of domestication for a long-lived perennial crop like olive. In both cases, the arrow from wild populations to domesticates indicates the possibility of introgression events that facilitate local adaptation after crop dispersion. Photographs are by the authors, except the bottom photographs of corn, both of which are from Wikipedia.

The domestication of annuals – and our level of understanding of that process – contrasts dramatically with that of long-lived perennials. Like annuals, the domestication of perennials begins with the recognition of the value of a species in the wild (Figure 1). However, some perennials have long juvenile periods; for example, the juvenile period of olives lasts 15–20 years [21]. Because of the duration of the juvenile period, the domestication of perennials may have less frequently involved the movement of seeds and their subsequent propagation (Figure 1). Rather, perennial domestication has relied more on clonal propagation [3], which has at least three important consequences.

First, because most long-lived perennials are outcrossers [3,5], clonal propagation tends to capture heterozygous genotypes. Second, because these genotypes may be maintained indefinitely, many new mutations may arise somatically rather than sexually. Nonetheless, some crossing does occur in the formation, maintenance, and improvement of perennial crops (Figure 2). Third, as a consequence of extended juvenility and restricted sexual generations, perennial domesticates tend to be younger in terms of generations relative to annuals [3,22]. In addition, most perennial crops appear to have been domesticated more recently, for reasons that are uncertain but may indicate that the technology for clonal propagation may have come after the initial domestication of annual plants [4].

The number of domestication events has been poorly studied for perennial fruit crops, with a few exceptions like apple (*Malus* × *domestica* Borkh.), grape (*Vitis vinifera* ssp. *sativa*) and olive (*Olea europaea* ssp. *europaea* var. *sativa*). Both grapes and olives have been thought to have multiple domestication centers [23], with as many as nine domestication events postulated for olive [24]. However, recent studies are more consistent, with fewer domestication events for both olive and grape followed by diversification via admixture with other pre-cultivated or wild forms [25–27]. For apple, domestication occurred in Central Asia from *Malus sirvensii* M. Roem, but there is also increasing evidence for a significant genetic contribution from the wild relative *Malus orientalis* Uglitzk and the European crabapple (*Malus sylvestris* L.) [28,29]. The number of domestication events in perennial crops is a topic that merits further study.



Trends in Genetics

Figure 2. A View of the Domestication and Divergence of Perennial Tree Crops. This view encompasses the fact that clonal propagation can be coupled with sexual outcrossing to generate genetic diversity that becomes the basis for new clonal cultivars. In addition, somatic mutations introduce new variation into a clonal cultivar.

Demography and Bottlenecks

Domestication often involves a genetic bottleneck and thus major crops carry less genetic diversity than their wild progenitors. Once DNA sequencing became available, it was possible to use nucleotide variation to model the dynamics of these bottlenecks. The first use of nucleotide diversity data to infer past demographic history in any species examined the genetic bottleneck associated with the domestication of maize [30]. The basic approach was to contrast DNA sequence variability between maize and its wild ancestor to fit population genetic models to summaries of the sequence data, assuming a demographic event (i.e., the domestication bottleneck). Initial models were very simple, with the purpose of estimating the population-genetic parameters of domestication. For example, using bottleneck models Zhu *et al.* (2007) inferred that rice domestication was based on a bottleneck population of ~1500 individuals, assuming that domestication took 3000 years [31].

The comparison of crop diversity with that of its wild ancestor has become a standard (and expected) exercise in the study of crop genetic diversity. Genome-wide comparisons between wild and crop germplasm have been applied to maize [32], rice [19,33], soybean [34,35], common bean [36], and cucumber [37]. These studies suggest that the severity of bottlenecks, as measured by the loss of diversity in the domesticate, vary substantially among annual species. For example, maize is estimated to have retained 83% of the nucleotide diversity of its wild ancestor [32] while soybean, tomato, and cucumber retain 64% [34], 38%, and 51% [37], respectively. The reduction of genetic diversity has been even more drastic in wheat [38,39]. As the data have become more comprehensive, so have the methods to explore demographic history [40–42], including methods that are appropriate for application to a single heterozygous genome sequence (Box 1). Note that these inferences are subject to sampling effects; that is, if present-day populations do not represent the breadth of diversity of older populations, the effect of the genetic bottleneck is likely to be underestimated. Similarly, if the incorrect extant wild populations are sampled as representative of predomestication diversity, estimates of the bottleneck could be misleading in either direction.

Domestication bottlenecks seem to be less severe in perennials than annuals. Miller and Gross (2011) report that studies of perennial fruit crops suggest that they maintain an average of ~95% of the neutral variation found in wild populations, while annuals retain an average of ~60%. Some major fruit crops, such as apple [29,43], cherry [44], and grape [26], exhibit little to no loss of diversity (although there are exceptions, like peaches [45]). Overall, the lack of dramatic bottleneck effects may help explain why perennial fruit crops exhibit fewer phenotypic shifts during domestication than do annual crops [4].

What mechanisms favor the retention of nucleotide variation within long-lived perennial crops? The first is that cross-pollination is the dominant mating system and inbreeding depression is often severe [5], such that the retention of genetic diversity is favored. Second, as we have

Box 1. Demographic Inference

The pairwise sequentially Markovian coalescent (PSMC) model [41] is a recent innovation. Rather than assuming a simplified demographic model with few parameters, the PSMC can provide a more detailed history of population size with fewer assumptions. The PSMC requires genomic data because it is applied to the full genome sequence of a single diploid individual. The goal of the method is to find the time to most recent common ancestor (TMRCA) between the two alleles carried by an individual. Because there are thousands of loci within the genome, the distribution of the TMRCA can be inferred across loci, and this distribution provides insights into past demographic history. The PSMC has been extended to the multiple sequentially Markovian coalescent (MSMC) to use samples from multiple individuals [98]. These approaches have exciting potential because the genome of a crop of interest can be assembled *de novo* and serve as the basis for demographic inference. To our knowledge these and similar models have yet to be applied to crop domestication. However, the PSMC was used to investigate the domestication history of horses, an interesting case for which the wild progenitor species is no longer extant [78].

mentioned, perennials have generally passed through fewer generations since domestication; since the winnowing effects of genetic bottlenecks accrue over time, this also favors the retention of diversity. Third, perennials have overlapping generations, a feature that is also likely to reduce the severity of bottlenecks. Finally, hybridization has often played a central role in the origin and diversification of perennials. For example, genomic sequencing of citrus fruits has revealed that cultivated mandarins feature admixture of *Citrus maxima* into the ancestral mandarin species (*Citrus reticulata)* and that sour orange is an F_1 hybrid of *C. maxima* and *C. reticulata* parents [46]. Similarly, interspecific hybridization has played a role in the formation of apple [28], banana [47], strawberry [48], and several other species [3].

While long-lived perennial crops may not have a common history of severe genetic bottlenecks, modern breeding is currently narrowing the genetic base of many long-lived perennial crops. For example, in the USA 15 apple cultivars currently comprise 90% of apple production [49] and one cultivar (Hass) accounts for 95% of avocado production [50]. Additionally, olive cultivars are being developed for mechanical harvesting systems that require dwarfed trees and large yields under high-density planting. Because only a small percentage of existing cultivars meet these requirements, future cultivation may capture only a fraction of olive diversity. Thus, even if perennial crops did not experience a strong domestication bottleneck, many are undergoing an ‘improvement bottleneck’ associated with modern breeding [43,44].

Introgression and Local Adaptation

Once domesticated, a successful crop is often dispersed from its domestication center to new geographic locales (Figure 1). This dispersion necessitates adaptation to new environments. Adaptation can be achieved by selection on new or existing genetic diversity within the crop, but it is becoming apparent that adaptation is often achieved with the help of introgression.

Introgression may occur between a crop and its wild relatives or even between independent domestication events of a single crop. As examples of crop-to-crop introgression, it is thought that domestication genes have migrated between varieties of Asian rice [51]. African rice may also have some features of introgression from Asian rice [52], although more recent analyses seem to refute this possibility [20]. Similarly, flowering time alleles in cultivated European barley appear to be introgressed from an independent domestication event in Central Asia [53] and barley landraces exhibit evidence of genetic contributions from multiple locally proximate wild barley populations in a manner that suggests that some genetic exchange among barley domestications was adaptive [15]. Crop-to-crop introgression also occurs in long-lived perennials; in avocado, for example, many modern cultivars are recent hybrids between trees derived from three independent domestication events [54].

Wild-to-crop introgression may be another potent source for crop adaptation. At least within its native range, the progenitor has a longer history of adaptation to local environments and as a consequence should be better adapted (and include more adaptive variants) than newly introduced crops [55]. For example, the adaptation of maize to the Mexican highlands may have been facilitated by introgression with wild populations [11,12] and resistance to a fungal pathogen in wheat was likely derived from introgression with wild populations [39].

In general, the history of admixture between crops and their wild progenitors has not been sufficiently studied, particularly given the potential for the discovery of agronomically important alleles in wild populations. There are several emerging analytical methods that can help determine whether a genomic segment has been introgressed. For example, the D statistic [56] relies on the identification of an ancestral state (Box 2) to identify introgression between populations (Figure 3). The computation of D requires data from three populations plus an outgroup that is used to determine the ancestral state of variants. When it is applied to markers across the

Box 2. Inferring the Ancestral State

The inference of the ancestral state of mutations plays an important role in the study of domestication. For agronomically important phenotypic traits, the phenotypic state in the domesticate will either be largely absent (e.g., seed shattering) or segregating within the wild progenitor (e.g., reduced branching). For SNPs, the vast majority of variants segregate in both the crop and the wild progenitor. That is, they are older than the domestication event. These variants can be quantified in terms of the descriptive statistics used in population genetics, by dividing SNPs into classes based on their occurrence in either wild or domesticated samples. In most cases, the majority of SNPs will be 'shared,' another portion will be 'private' to either group (with more private SNPs in wild samples), and few if any 'fixed' differences will be observed.

The inference of the ancestral state of mutations requires comparison with a related species that can be used as an outgroup (Figure I). The species should be closely related, such that there are few repeated mutations at a nucleotide site, but distantly enough related that few polymorphisms are shared between the outgroup and the ingroup. In practice, the best outgroup may not be the most closely related species, as the time to 'reciprocal monophyly' for most loci is ~1.7*N* generations, where *N* is the population size, with 99% reciprocal monophyly at ~5.3*N* generations [99]. In practical terms this means that, for a species like maize, it may be more appropriate to sequence a genome from the related genus *Tripsacum* as an outgroup than other species in the genus *Zea* [95].

The inference of the ancestral state of SNPs has important applications to the study of domestication. Most directly, it makes it possible to determine the absolute frequency of a variant in a population. This information can be used to test for departures from models of neutral evolution, using the site frequency spectrum, including tests like Fay and Wu's *H* [100], where an excess of high-frequency derived variants can be indicative of selection acting on a locus. Ancestral state inference is also important for the identification of deleterious mutations and is especially important for reducing reference bias [101]. Tests for introgression such as the *D* statistic described in Figure 3 also make use of the ancestral state of mutations. Finally, it is crucial for the calculation of divergence times between a domesticate and its wild ancestor.

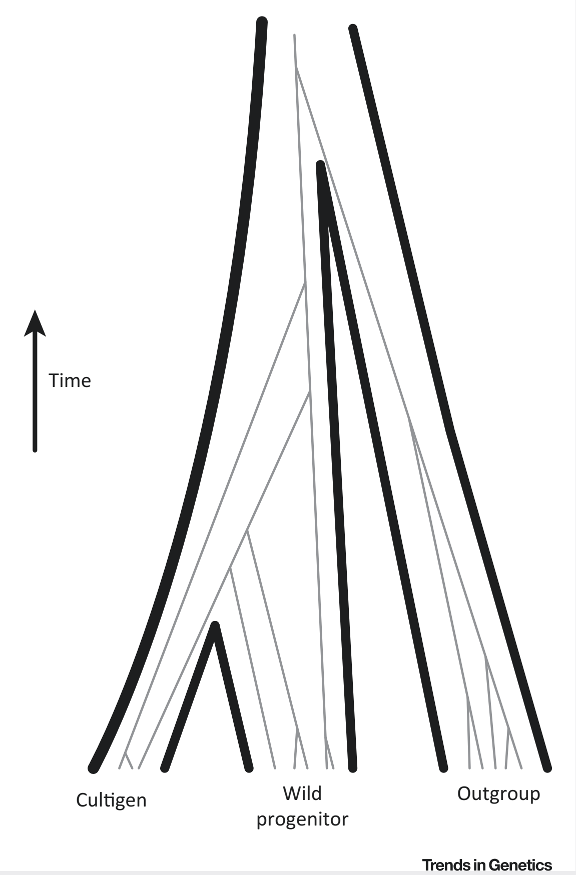


Figure I. Allelic History in a Crop and its Progenitor. The bold lines represent the genetic pool of a cultigen, its wild ancestor, and an outgroup. The lighter lines represent allelic lineages within species. As drawn, the cultivar and wild ancestor share several allelic lineages, representing segregating variation. Hence, an outgroup is necessary to infer the ancestral state of mutations.

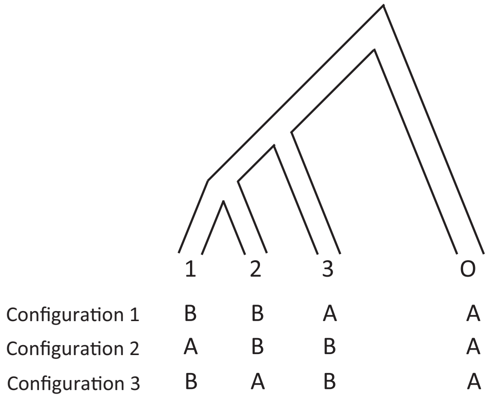


Figure 3. A Schematic of the *D* Statistic, Which Requires Data from Three Populations (1, 2, and 3) and an Outgroup (O). The outgroup is used to infer the ancestral state (A) of a genetic variant. The derived state (B) of the genetic variant may be segregating among populations 1, 2, and 3, resulting in the three configurations in the figure. Among the three configurations, Configuration 1 (BBAA) is expected to be most common because populations 1 and 2 are sister populations sharing the most recent common ancestor. However, Configuration 2 (ABBA) and Configuration 3 (BABA) are expected to be equally common in the absence of introgression. Summed over many genetic variants, gene flow either between populations 1 and 3 or between populations 1 and 2 will skew the frequency of Configuration 2 (ABBA) versus Configuration 3 (BABA), thus providing evidence of introgression.

genome, it may identify specific genomic regions with a history of introgression. The *D* test is also known as the ABBA–BABA test, owing to the pattern of shared polymorphisms among the sampled populations [57] (Figure 3).

Our impression is that wild-to-crop introgression may be even more important in perennials than annuals, for two reasons. First, genetic similarity between wild populations and local cultivars is found commonly in some long-lived perennials. In olives, for example, wild olives (*O. europaea* ssp. *europaea* var. *sylvestris*) survived in glacial refugia throughout Europe [58] and thus are likely to contain locally adapted alleles. It may be unsurprising, therefore, that studies of wild and cultivated forms suggest admixture on a local scale in Spain, Italy, and Turkey [59–61]. Second, many crops, like olives and grapes, abound as local cultivars, raising the possibility that local introgression contributed to the formation of these cultivars. It can be difficult, however, to discriminate between local adaptation due to wild-to-crop introgression and the possibility of sexual crossing between domesticated varieties and subsequent selection among seedlings (Figure 2).

In general, while there have been many surveys of genetic diversity in long-lived perennial crops, most have not utilized enough markers to identify the historical processes that produce locally adapted cultivars. Increased emphasis on genomics – and genotyping methods like RAD-seq [62] – will greatly aid these studies within perennial crops [5]. The increasing availability of genome-wide markers is complemented by analytical approaches for studying demographic history (Box 1) and introgression (Figure 3). For example, recent studies of apples and olives have used approximate Bayesian computation (ABC) methods to test alternative models of population history, including the inference of introgression between cultivars and wild relatives [25,28]. Both of these papers found evidence of modest but significant introgression between cultivars and their local wild relatives. A complementary approach to ABC models will be tests of association between SNPs and ecological data [63]; these will be particularly revealing about adaptive introgression if SNPs that associate with ecological data are shared more often than expected between nearby wild and cultivated populations.

We close this section by noting that the ancestors of perennial crops and their wild relatives have contributed genetic diversity in another important way, through rootstock. A wide variety of rootstock confers adaptation to soil conditions and soil pathogens. This has been central to overcoming epidemics of *Phylloxera* in grape and the *Citrus* tristeza virus (CTV) in *Citrus* [64,65]. Historically, the use of rootstocks has also facilitated the introduction of a cultivar into formally 'wild' areas. For example, the grafting of centennial olives of southern Spain onto wild oleasters may have instantaneously converted a wild forest into an olive orchard [66].

Genetic Load

As discussed above, domestication is often accompanied by a genetic bottleneck, particularly for annual crops. One consequence of this demographic event is a shift in the efficacy of selection, because efficacy scales with population size [67]. Because population sizes are smaller during a bottleneck, selection is expected to be less effective with respect to both the fixation of positive alleles and the removal of deleterious mutations. Because of the latter, some variants that reduce fitness but are nonlethal will continue to segregate as deleterious variants.

The manner in which deleterious variants affect yield and productivity in crop species remains poorly understood [68] and is expected to vary with the mating system and dominance coefficients. We can, however, posit three general predictions: (i) the relative proportion of deleterious variants will be higher in domesticates than in wild relatives, due to genetic bottlenecks; (ii) deleterious variants will be enriched near loci subjected to strong positive selection during domestication and improvement, due to hitchhiking effects [69]; and (iii) the relative proportion of deleterious variants will be reduced in elite cultivars compared with landraces, because strong selection for yield will purge deleterious mutations from the gene pool.

These predictions regarding the demography and the load of deleterious mutations are being tested with empirical data, especially in humans (who, like annual crops, have experienced bottlenecks at various times in their history). Generally these papers proceed by identifying nucleotide variants throughout the entire genome and parsing variants to identify those predicted to have deleterious effects. There are multiple computational methods to predict whether variants are deleterious – such as SIFT [70], Polyphen2 [71], PROVEAN [72], and others [73] – but all tend to identify deleterious variants as variants that occur in highly conserved genic regions and/or that confer radical changes on an amino acid sequence. Despite clear expectations, debate rages in the human literature about whether the accumulation of deleterious mutations can be, or is expected to be, detected [74].

The data remain limited in crops, but the available evidence suggests that the accumulation of deleterious alleles has occurred in rice [75], maize [76], and sunflower [77]; this effect has been called the 'cost of domestication' [78]. These observations have potential implications for breeding. Because selection is forced to act on the net fitness across all variants in a genomic region [69], deleterious mutations linked to regions of agronomic value will reduce the efficacy of selection [79]. This suggests the need to balance the strength of selection applied in breeding against the effective rate of recombination within a population, which is a factor largely determined by the effective number of breeding individuals [69]. One pragmatic approach may be to incorporate deleterious variants into models used to make predictions for genomic selection [68].

Are perennial crops likely to be subject to the same cost of domestication through the accumulation of deleterious mutations? Empirically we do not know the answer, but the predictions are not entirely clear either. On the one hand, three features of perennial cultivation argue against the accumulation of deleterious mutations: (i) less-stringent domestication bottlenecks; (ii) the possibility that some perennial crops cross with wild relatives (thus potentially purging deleterious mutations through recombination and assortment); and (iii) lower mutation rates than annual plants [80], so that the effect of new mutations may not be as substantial as for annual crops.

On the other hand, perennial crops are often outcrossing, and recessive deleterious variants are more rarely exposed to selection under an outcrossing mating system, resulting in the potential to harbor a greater number of deleterious variants [81]. Perennial crops are also often propagated clonally, with only occasional cycles of sexual selection (Figure 2). Clonality is equivalent to a long-term population size of one, so selection against deleterious mutations will not be effective and deleterious somatic mutations have the potential to accumulate (Figure 2). There is evidence that varieties often contain somatic mutations, because somatic mutations have been described in apple, peach, pear, plum [82], and olives [66]. Somatic mutations have also generated phenotypic variability that has become the basis for new varieties of *Citrus* and grape, among others [65,83,84]. An unknown but appreciable number of these mutations may be caused by transposable elements (TEs) [84–86] (see Outstanding Questions).

Somatic mutations have contributed to the formation of new fruit crop varieties, but we know little about the mechanism and rates of their generation or about the proportion that are deleterious. Genomic approaches may help address these questions and hence promote an understanding of the trade-offs between clonal and sexual reproduction for these species [22].

Concluding Remarks

We have emphasized features of domestication that may differ between annuals and perennials, but also emphasized shortcomings in our knowledge of both systems. For a few annuals, there is a reasonable amount of information available about domestication in terms of the number of domestication events, the effects of genetic bottlenecks, and even the targets of positive selection during domestication [87]. However, even in these highly studied crops there are outstanding questions about introgression, local adaptation, genetic load and a myriad of other questions that can be addressed with population-genetic approaches (see Outstanding Questions).

By contrast, the study of perennial domestication is at best nascent from a population-genetic perspective. Many questions about the domestication of perennials and the subsequent steps of dispersion and cultivar improvement remain to be addressed. Part of the lag in the study of perennials has been slower genomics progress, but that obstacle is being overcome rapidly. With an increase in genomic data, the study of perennials is poised to take advantage of approaches and methodologies previously applied primarily to annual plants. The outcomes will bring new insights about the process of domestication, the extent to which domestication has influenced genomic architecture, and the relative importance of clonal versus sexual propagation both within a crop and between the crop and its wild relatives.

Acknowledgments

B.S.G. is supported by National Science Foundation (NSF) grant IOS-1542703, P.L.M. is supported by NSF grant IOS-1339393, and the US Department of Agriculture National Institute for Food and Agriculture (USDA NIFA) 2011-68002-30029. C.M.D. is supported by a postdoctoral fellowship from the University of Córdoba.

References

1. Doebley, J.F. *et al.* (2006) The molecular genetics of crop domestication. *Cell* 127, 1309–1321
2. Purugganan, M.D. and Fuller, D.Q. (2009) The nature of selection during plant domestication. *Nature* 457, 843–848
3. Miller, A.J. and Gross, B.L. (2011) From forest to field: perennial fruit crop domestication. *Am. J. Bot.* 98, 1389–1414
4. Meyer, R.S. *et al.* (2012) Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48
5. McClure, K.A. *et al.* (2014) Genomics: a potential panacea for the perennial problem. *Am. J. Bot.* 101, 1780–1790

Outstanding Questions

There are numerous outstanding questions regarding genetic diversity in both annual and perennial domesticates.

Epigenetics. Although it remains somewhat open to debate, epigenetic variation may contribute to phenotypic variation without underlying genetic changes [88]. Thus, epigenetic shifts may play a role in domestication, but the interaction between epigenetics and domestication remains unstudied. Epigenetic mechanisms may be particularly important in clonal propagation because clones are propagated from adult tissue and because the shift from juvenile to adult tissue is mediated in part by miRNAs and other epigenetic mechanisms [89].

TEs. TEs may play a major role in domestication, for at least three reasons. First, TEs are likely to be a primary source of the somatic mutations that generate clonal diversity in perennials (e.g., [90]). Second, there may be enhanced TE activity in domesticates as a consequence of demographic shifts and selection. Third, TEs regulate gene expression, both via direct genetic effects and through interactions with epigenetic mechanisms [91]. TEs are known to have generated mutations that have been crucial in the phenotypic evolution of perennial crops [92–94]. Unfortunately, TEs are typically disregarded in studies of crop genomic diversity, which in our view is akin to 'throwing out the baby with the bath water'.

Genetic inversions. Recent papers have found large chromosomal inversions that contribute disproportionately to ecological divergence among wild populations of the ancestors of domesticated crops (e.g., [55,95]). These observations raise at least three sets of questions. First, are inversions important contributors to local adaptation? Second, if so, do they contribute disproportionately relative to the remaining genome or are they simply easier to detect because they constitute a large block of extended linkage disequilibrium? Finally, if they are important in the wild ancestors of crops, do they also contribute to the local adaptation of crops?

Genetic convergence. It has long been hypothesized that domestication is mediated by similar sets of genes in different taxa. To that end, Wang *et al.*

6. Gepts, P. (2014) The contribution of genetic and genomic approaches to plant domestication studies. *Curr. Opin. Plant Biol.* 18, 51–59
7. Ross-Ibarra, J. *et al.* (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 104 (Suppl. 1), 8641–8648
8. Matsuoka, Y. *et al.* (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6080–6084
9. Fukunaga, K. *et al.* (2005) Genetic diversity and population structure of teosinte. *Genetics* 169, 2241–2254
10. van Heerwaarden, J. *et al.* (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1088–1092
11. Hufford, M.B. *et al.* (2013) The genomic signature of crop–wild introgression in maize. *PLoS Genet.* 9, e1003477
12. Takuno, S. *et al.* (2015) Independent molecular basis of convergent highland adaptation in maize. *Genetics* 200, 1297–1312
13. Pourkheirandish, M. *et al.* (2015) Evolution of the grain dispersal system in barley. *Cell* 162, 527–539
14. Morrell, P.L. and Clegg, M.T. (2007) Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. *Proc. Natl. Acad. Sci. U.S.A.* 104, 3289–3294
15. Poets, A.M. *et al.* (2015) Barley landraces are characterized by geographically heterogeneous genomic origins. *Genome Biol.* 16, 173
16. Bitocchi, E. *et al.* (2012) Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc. Natl. Acad. Sci. U.S.A.* 109, E788–E796
17. Lin, Z. *et al.* (2012) Parallel domestication of the *Shattering1* genes in cereals. *Nat. Genet.* 44, 720–724
18. Molina, J. *et al.* (2011) Molecular evidence for a single evolutionary origin of domesticated rice. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8351–8356
19. Huang, X. *et al.* (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490, 497–501
20. Wang, M. *et al.* (2014) The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat. Genet.* 46, 982–988
21. Fontanazza, G. and Baldoni, L. (1990) Proposed programme for the genetic improvement of the olive. *Olive* 34, 1–20
22. McKey, D. *et al.* (2010) The evolutionary ecology of clonally propagated domesticated plants. *New Phytol.* 186, 318–332
23. Arroyo-Garcia, R. *et al.* (2006) Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. *Mol. Ecol.* 15, 3707–3714
24. Breton, C. *et al.* (2009) The origins of the domestication of the olive tree. *C. R. Biol.* 332, 1059–1064
25. Diez, C.M. *et al.* (2015) Olive domestication and diversification in the Mediterranean Basin. *New Phytol.* 206, 436–447
26. Myles, S. *et al.* (2011) Genetic structure and domestication history of the grape. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3530–3535
27. Besnard, G. *et al.* (2013) The complex history of the olive tree: from Late Quaternary diversification of Mediterranean lineages to primary domestication in the northern Levant. *Proc. Biol. Sci.* 280, 20122833
28. Cornille, A. *et al.* (2012) New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genet.* 8, e1002703
29. Cornille, A. *et al.* (2014) The domestication and evolutionary ecology of apples. *Trends Genet.* 30, 57–65
30. Eyre-Walker, A. *et al.* (1998) Investigation of the bottleneck leading to the domestication of maize. *Proc. Natl. Acad. Sci. U.S.A.* 95, 4441–4446
31. Zhu, Q. *et al.* (2007) Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Mol. Biol. Evol.* 24, 875–888
32. Hufford, M.B. *et al.* (2012) Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44, 808–811
33. He, Z. *et al.* (2011) Two evolutionary histories in the genome of rice: the roles of domestication genes. *PLoS Genet.* 7, e1002100
34. Lam, H.M. *et al.* (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42, 1053–1059
35. Zhou, Z. *et al.* (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33, 408–414
36. Schmutz, J. *et al.* (2014) A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713
37. Qi, J. *et al.* (2013) A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. *Nat. Genet.* 45, 1510–1515
38. Haudry, A. *et al.* (2007) Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Mol. Biol. Evol.* 24, 1506–1517
39. Cavanagh, C.R. *et al.* (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Natl. Acad. Sci. U.S.A.* 110, 8057–8062
40. Gutenkunst, R.N. *et al.* (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5, e1000695
41. Li, H. and Durbin, R. (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475, 493–496
42. Harris, K. *et al.* (2014) Decoding coalescent hidden Markov models in linear time. *Res. Comput. Mol. Biol.* 8394, 100–114
43. Gross, B.L. *et al.* (2014) Genetic diversity in *Malus × domestica* (Rosaceae) through time in response to domestication. *Am. J. Bot.* 101, 1770–1779
44. Mariette, S. *et al.* (2010) Population structure and genetic bottleneck in sweet cherry estimated with SSRs and the gametophytic self-incompatibility locus. *BMC Genet.* 11, 77
45. Verde, I. *et al.* (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat. Genet.* 45, 487–494
46. Wu, G.A. *et al.* (2014) Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat. Biotechnol.* 32, 656–662
47. Heslop-Harrison, J.S. and Schwarzacher, T. (2007) Domestication, genomics and the future for banana. *Ann. Bot.* 100, 1073–1084
48. Hancock, J.F. *et al.* (2008) Strawberries. In *Temperate Fruit Crop Breeding* (Hancock, J., ed.), pp. 393–437, Springer
49. United States Apple Association (2015) usapple.org
50. California Avocado Commission (2015) www.californiaavocado.com
51. Hua, L. *et al.* (2015) *LABA1*, a domestication gene associated with long, barbed awns in wild rice. *Plant Cell* 27, 1875–1888
52. Semon, M. *et al.* (2005) The population structure of African cultivated rice *Oryza glaberrima* (Steud.): evidence for elevated levels of linkage disequilibrium caused by admixture with *O. sativa* and ecological adaptation. *Genetics* 169, 1639–1647
53. Jones, H. *et al.* (2008) Population-based resequencing reveals that the flowering time adaptation of cultivated barley originated east of the Fertile Crescent. *Mol. Biol. Evol.* 25, 2211–2219
54. Chen, H. *et al.* (2009) Tracing the geographic origins of major avocado cultivars. *J. Hered.* 100, 56–65
55. Fang, Z. *et al.* (2014) Two genomic regions contribute disproportionately to geographic differentiation in wild barley. *G3 (Bethesda)* 4, 1193–1203
56. Durand, E.Y. *et al.* (2011) Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28, 2239–2252
57. Sousa, V. and Hey, J. (2013) Understanding the origin of species with genome-scale data: modelling gene flow. *Nat. Rev. Genet.* 14, 404–414
58. Carrion, Y. *et al.* (2010) *Olea europaea* L. in the north Mediterranean Basin during the Pleniglacial and the Early-Middle Holocene. *Quat. Sci. Rev.* 29, 952–968

[20] found evidence for genetic convergence among rice domestication events. There is, however, little evidence to date that domestication enriches for convergent events on a genome-wide scale [96]. If convergence proves to be the rule rather than the exception, it will facilitate the transfer of genetic information among crops. Further inferences are likely to require analyses at the level of networks [96].

New variants or segregating variation. Does domestication typically rely on genetic diversity that is segregating in wild populations or on novel mutations? The answer to this question has important implications for inferring process [97] and also for generating the expected distribution of linkage disequilibrium around sites that contribute to domestication phenotypes. Addressing this question requires additional genome-wide studies of genetic diversity in the wild ancestors of annual and perennial crops.

59. Baldoni, L. *et al.* (2006) Genetic structure of wild and cultivated olives in the central Mediterranean Basin. *Ann. Bot.* 98, 935–942
60. Belaj, A. *et al.* (2010) Genetic diversity and relationships of wild and cultivated olives at regional level in Spain. *Sci. Hortic. (Amsterdam)* 124, 323–330
61. Yoruk, B. and Taskin, V. (2014) Genetic diversity and relationships of wild and cultivated olives in Turkey. *Plant Syst. Evol.* 300, 1247–1258
62. Baird, N.A. *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3, e3376
63. Eckert, A.J. *et al.* (2010) Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* 185, 969–982
64. Moreno, P. *et al.* (2008) *Citrus* tristeza virus: a pathogen that changed the course of the citrus industry. *Mol. Plant Pathol.* 9, 251–268
65. This, P. *et al.* (2006) Historical origins and genetic diversity of wine grapes. *Trends Genet.* 22, 511–519
66. Diez, C.M. *et al.* (2011) Centennial olive trees as a reservoir of genetic diversity. *Ann. Bot.* 108, 797–807
67. Kimura, M. (1983) The neutral theory of molecular evolution. In *Evolution of Genes and Proteins* (Nei, M. and Koehn, R.K., eds), pp. 208–233, Sinauer
68. Morrell, P.L. *et al.* (2011) Crop genomics: advances and applications. *Nat. Rev. Genet.* 13, 85–96
69. Hill, W.G. and Robertson, A. (1966) The effect of linkage on limits to artificial selection. *Genet. Res.* 8, 269–294
70. Ng, P.C. and Henikoff, S. (2003) SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31, 3812–3814
71. Adzhubei, I.A. *et al.* (2010) A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249
72. Choi, Y. *et al.* (2012) Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* 7, e46688
73. Chun, S. and Fay, J.C. (2009) Identification of deleterious mutations within three human genomes. *Genome Res.* 19, 1553–1561
74. Lohmueller, K.E. (2014) The distribution of deleterious genetic variation in human populations. *Curr. Opin. Genet. Dev.* 29, 139–146
75. Lu, J. *et al.* (2006) The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends Genet.* 22, 126–131
76. Mezmouk, S. and Ross-Ibarra, J. (2014) The pattern and distribution of deleterious mutations in maize. *G3 (Bethesda)* 4, 163–171
77. Renaut, S. and Rieseberg, L.H. (2015) The accumulation of deleterious mutations as a consequence of domestication and improvement in sunflowers and other Compositae crops. *Mol. Biol. Evol.* 32, 2273–2283
78. Schubert, M. *et al.* (2014) Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci. U.S.A.* 111, E5661–E5669
79. Felsenstein, J. (1974) The evolutionary advantage of recombination. *Genetics* 78, 737–756
80. Gaut, B. *et al.* (2011) The patterns and causes of variation in plant nucleotide substitution rates. *Annu. Rev. Ecol. Evol. Syst.* 42, 245–266
81. Charlesworth, B. and Charlesworth, D. (1998) Some evolutionary consequences of deleterious mutations. *Genetica* 102–103, 3–19
82. Badenes, M.L. and Byrne, D.H. (2012) Fruit breeding. *Handb. Plant Breed.* 8, 3–37
83. Moore, G.A. (2001) Oranges and lemons: clues to the taxonomy of *Citrus* from molecular markers. *Trends Genet.* 17, 536–540
84. Pelsy, F. *et al.* (2010) An extensive study of the genetic diversity within seven French wine grape variety collections. *Theor. Appl. Genet.* 120, 1219–1231
85. Oliver, K.R. *et al.* (2013) Transposable elements: powerful contributors to angiosperm evolution and diversity. *Genome Biol. Evol.* 5, 1886–1901
86. Breto, M.P. *et al.* (2001) The diversification of *Citrus clementina* Hort. ex Tan., a vegetatively propagated crop species. *Mol. Phylogenet. Evol.* 21, 285–293
87. Glemin, S. and Bataillon, T. (2009) A comparative view of the evolution of grasses under domestication. *New Phytol.* 183, 273–290
88. Cortijo, S. *et al.* (2014) Mapping the epigenetic basis of complex traits. *Science* 343, 1145–1148
89. Wang, J.W. *et al.* (2011) miRNA control of vegetative phase change in trees. *PLoS Genet.* 7, e1002012
90. Carrier, G. *et al.* (2012) Transposable elements are a major cause of somatic polymorphism in *Vitis vinifera* L. *PLoS ONE* 7, e32973
91. Hollister, J.D. and Gaut, B.S. (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* 19, 1419–1428
92. Kobayashi, S. *et al.* (2004) Retrotransposon-induced mutations in grape skin color. *Science* 304, 982
93. Butelli, E. *et al.* (2012) Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* 24, 1242–1255
94. Otto, D. *et al.* (2013) The columnar mutation ("*Co* gene") of apple (*Malus* × *domestica*) is associated with an integration of a *Gypsy*-like retrotransposon. *Mol. Breed.* 33, 863–880
95. Fang, Z. *et al.* (2012) Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics* 191, 883–894
96. Gaut, B.S. (2015) Evolution is an experiment: assessing parallelism in crop domestication and experimental evolution (Nei Lecture, SMBE 2014, Puerto Rico). *Mol. Biol. Evol.* 32, 1661–1671
97. Innan, H. and Kim, Y. (2008) Detecting local adaptation using the joint sampling of polymorphism data in the parental and derived populations. *Genetics* 179, 1713–1720
98. Schiffels, S. and Durbin, R. (2014) Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* 46, 919–925
99. Rosenberg, N.A. (2003) The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. *Evolution* 57, 1465–1477
100. Fay, J.C. and Wu, C-I. (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155, 1405–1413
101. Simons, Y.B. *et al.* (2014) The deleterious mutation load is insensitive to recent population history. *Nat. Genet.* 46, 220–224