Evolutionary insights into plant breeding

Sarah D. Turner-Hissong[a,b]*, Makenzie E. Mabry[c], Timothy M. Beissinger[d,e], Jeffrey Ross-Ibarra[a,b], J. Chris Pires[c]

[a] Center for Population Biology, University of California, Davis, CA, USA

[b] Department of Evolution and Ecology, University of California, Davis, CA, USA

[c] Bond Life Science Center and Division of Biological Sciences, University of Missouri, Columbia, MO, USA

[d] Division of Plant Breeding Methodology, Department of Crop Science, Georg-August-Universtät, Göttingen, Germany

[e] Center for Integrated Breeding Research, Georg-August-Universtät, Göttingen, Germany

* Corresponding author

Sarah Turner-Hissong
2320 Storer Hall
University of California, Davis
One Shields Avenue
Davis, CA 95616
turnersarahd@gmail.com

Highlights:

- Population genetics can inform breeding strategies to increase genetic gain
- Genomic regions with selective sweeps are promising targets to introgress diversity
- Demography determines genetic architecture and deleterious load
- Deleterious alleles can either be purged via selection or masked via hybridization
- Gene editing enables rapid, precise introgression and *de novo* domestication

Abstract

Crop domestication is a fascinating area of study, as evidenced by a multitude of recent reviews. Coupled with the increasing availability of genomic and phenomic resources in numerous crop species, insights from evolutionary biology will enable a deeper understanding of the genetic architecture and short-term evolution of complex traits, which can be used to inform selection strategies. Future advances in crop improvement will rely on the integration of population genetics with plant breeding methodology, and the development of community resources to support research in a variety of crop life histories and reproductive strategies. We highlight recent advances in the role of selective sweeps and demographic history in shaping genetic architecture, how these breakthroughs can inform selection strategies, and the application of precision gene editing to leverage these connections.
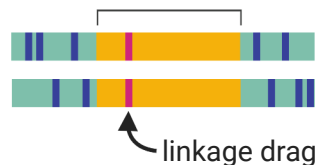
Key words: domestication, population genetics, plant breeding, demography, deleterious load, selective sweeps, *de novo* domestication, wild relatives

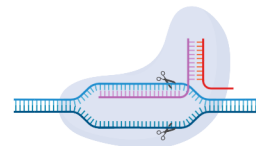Graphical Abstract

Introduction [205 words]

As access to and the quantity of genomic data increases, there are new opportunities for collaborations among the fields of population genetics and applied breeding. Investigations into crop origins and domestication both enhance our understanding of evolutionary biology (reviewed in [1]) and improve the efficiency of strategies for crop improvement. However, our understanding of crop origins, domestication, and genome structure remains limited in most species, ultimately delaying the impact this knowledge can have on applied breeding. Continued advances in the areas of crop domestication and improvement will benefit from collaborations between population geneticists and plant breeders. On one hand, breeders develop highly structured, pedigreed populations and empirical data which can be of use to population geneticists. On the other hand, knowledge of how evolutionary forces have and continue to shape existing diversity can support applied breeding efforts. The expansion of these collaborations across a wide range of crop species and disciplines will require development and availability of community infrastructure (**Box 1**).

In this review, we integrate recent advances and applications of crop domestication research and how these findings inform strategies for applied crop improvement. Specifically, we discuss the evolutionary forces shaping the domestication process and their influence on both genetic architecture and response to selection.

Factors influencing the processes of plant domestication and improvement [561 words]

The process of crop domestication can be broadly defined as selection to improve the adaptation of plants for human cultivation and preferences [2–5]. During this process, selection acts primarily on a suite of traits that distinguishes domesticated types from wild progenitor(s), referred to as the domestication syndrome [6]. In crop species, these are typically traits that increase yield and productivity, harmonize with crop management practices, promote ease of harvest, and improve palatability [5,7]. This initial period of domestication is followed by diversification, an extension of the domestication process that includes ongoing selection and improvement. Traits related to improvement vary among populations of a crop species for attributes like regional adaptation, amenability to modern agricultural practices, and quality [7–9]. At all stages, the process of domestication is shaped by complex forces such as human selection, environmental variation, genotype by environment interactions, and polyploidy [2,5]. These forces

in turn influence qualities important for applied crop improvement, such as phenotypic change, heritability, and the efficiency of selection and genetic gain [2].

Two emerging areas of recent interest include (i) expanding demographic inference to diverse crops and (ii) adapting existing methods to account for current and historical polyploidy. To date, the bulk of domestication research has focused on model crop systems (e.g. tomato) or highly domesticated, economically important crops (e.g. maize, wheat, and rice), with limited attention to the broad diversity of other crops [10]. Several recent studies have highlighted and worked to improve this insufficient representation of diverse life histories in domestication research [10–13]. Similarly, the effects of polyploidy on crop domestication are not yet well understood. All crop species are at least ancestrally polyploid [14], with many staple crops (e.g. wheat, cassava, and rapeseed) experiencing more recent rounds of additional polyploid events. However, the effect of multiple genomes doubling and merging (if allopolyploid) and slowly returning to a diploid state is rarely considered when studying existing genetic variation in domesticated populations (reviewed in [15]). The relevance of ancestral whole genome duplication was demonstrated for maize in [16], with evidence showing that the subgenomes of maize do not contribute equally to phenotypic variation. Advances in our ability to estimate the relationship between subgenomes and trait variation will be an exciting frontier for crop improvement research. These gaps and many others are in part being addressed by advances in the fields of paleogenomics (reviewed in [17]) and population genetics methodology. Together, these advances will enable deeper investigations into a larger range of evolutionary trajectories, population changes in the recent past, and the role of polyploidy in shaping crop variation. This knowledge will, in turn, be helpful to understand how modern selection is influencing crop diversity and potential for improvement.

Many additional topics within population genetics have relevance for applied crop breeding. These include the role of crop wild relatives in hybridization and crop evolution (**Box 2,** reviewed in [1,18]), the propensity for local adaptation (e.g. the 'home field advantage' observed in maize [19]), and opportunities to detect and leverage polygenic selection (reviewed in [20,21]). For the purposes of this review, we highlight in more detail three areas where plant breeding can benefit from evolutionary insights: the impact of selective sweeps on genetic diversity, demographic history in relation to the accumulation of genetic load, and applications of gene editing technologies for *de novo* domestication and precision breeding.

Selective sweeps shape genetic diversity [466 words]

Because domestication is inherently an adaptive process, selection is expected to play a dominant role in shaping diversity in plant genomes. Selection acts by increasing the frequency of beneficial mutations, which can result in selective sweeps that reduce genetic diversity in regions surrounding initially rare adaptive mutations [22–24]. Identifying these regions is of interest to understand how domestication has shaped genome-wide diversity, to determine which genes influence agronomically relevant traits, and to facilitate introgression of novel alleles. Typically, this also requires knowledge of the wild progenitor(s) for a given species (**Box 2**).

At the species level, the mode of selective sweeps in a population has important implications for crop evolution and breeding. Selective sweeps can be categorized as hard, where a single mutation rapidly rises in frequency (e.g. naked kernels in maize [25]), or as soft, where more than one allele contributes to adaptive substitution via selection on standing variation or multiple *de novo* mutations (e.g. determinate growth habit in soybean [26]). Traits influenced by hard sweeps have likely reached fixation in many elite breeding populations, bringing with them valleys of reduced diversity. Identifying these regions will enable opportunities for targeted introgression of beneficial alleles from wild or landrace materials to elite germplasm. In contrast, soft sweeps occur on two or more genetic backgrounds and represent regions with a less dramatic reduction in diversity, providing source material that is amenable for rapid response to selection [27]. This has proven to be relevant in maize, with soft sweeps selected during initial domestication experiencing additional selection during improvement [8].

Understanding the predominant mechanism of selective sweeps in a species can also provide guidance on mapping approaches. For instance, genome wide association mapping is expected to work well for traits controlled by hard sweeps, but biparental populations may be more effective for identifying loci that have evolved via soft sweeps (**Figure 1A**). Evidence also suggests that soft sweeps likely contribute to quantitative trait variation [28], and thus may be more important when targeting improvement of complex traits such as yield. In these cases, even a small effect on allele frequency may confer a shift in phenotype [29].

In theory, identifying the primary mechanism of sweeps in crop species will help inform the selection of germplasm pools and parental lines when establishing breeding populations. However, in most crop species it is still unknown if selection has been driven primarily by hard or

soft sweeps. This is in part due to challenges in the identification of selective sweeps ([30], methods reviewed in [31]), which can be confounded by population structure (see discussion in [32]) and gene flow from wild or feral populations (**Box 2**). Moving forward, it will be interesting to tease apart the relative influence of hard and soft sweeps on genetic architecture and continued crop evolution.

Demographic history influences deleterious load and the cost of domestication [663 words]

Breeding occurs on a relatively short time scale, such that *de novo* mutation is unlikely to play a prominent role in patterning beneficial genetic diversity (but see notable exceptions for naked kernels in maize [25] and inflorescence complexity in tomato [33]). Instead, breeders rely predominantly on existing genetic variation, with potential sources including wild relatives [34,35] and landraces [36]. As domesticated populations are by definition a subset of wild diversity, most crops have undergone a population bottleneck of sorts associated with domestication [37]. Such changes in population size can have dramatic impacts on patterns of existing genetic variation [29,38,39]. Recent advances in technology and computation have led to an improved understanding of demographic history, providing insights into the history of reproductive strategies and generational fluctuations in population size (reviewed in [38]). Demographic parameters such as the size and duration of previous bottlenecks [40], the extent and scope of population expansion [41], and reproductive techniques (i.e. inbreeding, outcrossing, or clonal propagation) [42] vary greatly among crops and have a profound impact on which modern breeding strategies are expected to be most effective. Furthermore, evidence suggests that demography influences the mode of selective sweeps in a population [29], the accuracy of genomic prediction [43,44], and the ability to identify associations with causal variants (discussed in [21]), particularly when effects on fitness and the trait of interest are correlated [45] (**Figure 1B**). In the case of genetic architecture, demographic changes can explain up to 55% of additive genetic variation, with recent population growth increasing the number of low frequency mutations [29].

A consequence of past fluctuations in population size is the accumulation of deleterious alleles. Because population bottlenecks increase genetic drift and decrease the effectiveness of selection, weakly deleterious alleles may persist in populations and even increase in frequency during such events. A number of studies have now provided evidence for the "cost of domestication" hypothesis, in which bottlenecks during domestication led to a concomitant

increase in the genetic load of deleterious alleles [46–49]. Although most deleterious alleles remain rare individually, as a group they may contribute substantially to inbreeding depression and variation in agronomic performance [50]. For instance, purging deleterious alleles in elite maize germplasm has led to an increase in both inbred and hybrid performance over time [51]. Alternatively, in potato, development of elite inbred lines is limited by line-specific deleterious mutations, which, combined with self-incompatibility, result in severe inbreeding depression [52]. In this case, deleterious load can either be purged via selection for recombinants or masked via hybridization of complementary lineages [52] (**Figure 1C**).

Recent work has demonstrated a correlation between recombination rate and the presence of deleterious alleles [53–55], with regions of low recombination expected to harbor a greater number of deleterious variants. Taken together, these examples suggest that the breeding scheme may impact recombination and patterns of deleterious load at the population level (**Figure 1C**). For crops with low recombination rates and an accumulation of deleterious mutations (e.g. cassava [53]), strategies to improve response to selection may include maximizing recombination by using larger population sizes and weaker selection, incorporating additional generations of outcrossing, or increasing the frequency of crossover events (reviewed in [56]).

An ongoing concern in applied breeding is the removal of potentially adaptive genetic variation that is linked to sites under strong selective pressure, although recent evidence in modern wheat suggests strong selection does not always purge adaptive diversity [57]. Whether to minimize drift, maximize recombination, or purge deleterious alleles, a takeaway from evolutionary theory that is well supported by recent sequence-based studies of crop populations is that large population sizes are important for breeding success. However, the long-established approach of breeding is to increase selection intensity in order to increase response to selection, a tactic that is consistent with the breeder's equation. All else being equal, increasing the strength of selection amounts to decreasing population size. In the future, marrying these two fundamentally opposing principles will be crucial for continued breeding success.

Gene editing enables *de novo* domestication and precision breeding [320 words]

As we begin to uncover more about the structure of crop genomes and loci underlying traits of agronomic quality, there are opportunities to introgress valuable traits into existing crop species and to rapidly domesticate new crops. This process can be accomplished efficiently using

revolutionary gene editing technologies like CRISPR-Cas9, especially in concert with doubled haploid production [58], to introduce beneficial alleles/genes without linkage drag [59], to generate novel quantitative variation [60], to directly remove deleterious alleles [61], and/or to increase recombination rates [62]. Recently, the value of gene editing was demonstrated for rapid improvement of plant architecture, flower production, and fruit size in groundcherry (*Physalis pruinosa*) [63]. Notable caveats of this approach include the requirement for a reference genome assembly, either for the target species or a closely related species, the potential to introduce negative epistatic effects [33], challenges with plant regeneration, and the incorporation of multiple gene targets [64].

Gene editing is also a promising tool to generate new diversity and to offset the effects of genetic hitchhiking in breeding germplasm. For instance, linked selection on traits such as fruit weight and disease resistance altered metabolome of tomato, presenting an opportunity for precision breeding to modify nutritional and flavor attributes [65]. These hitchhiking effects and others, such as those observed in rice [66] and maize [67], represent promising targets for gene editing to ameliorate linkage drag. For example, gene editing was successfully applied to rapidly domesticate tomato while maintaining disease resistance and salt tolerance [59].

To date, *de novo* domestication and precision breeding via gene editing has only been demonstrated in the Solanaceae, which has been recommended as a model system for the design and implementation of these methods [64]. The success of initial studies suggests this will be a promising tool for the domestication of new crops, such as perennial grains [68], and will be especially useful for traits with identified selective sweeps in related species.

Conclusions [108 words]

The increased accessibility and quantity of genomic sequence data has certainly been a boon for all genetics-related fields. However, increased accessibility of data alone will not accelerate breeding progress. Continued advances in breeding and population genetics will require coordinated efforts within crop-based communities and interdisciplinary cross-talk. There are still numerous complex and outstanding questions, including the relationship between crops and wild relatives, how population genetics parameters vary across species, and how these forces play off of each other and work together to influence crop adaptation. Future work will require both

theoretical advances in population genetics and breeding, as well as empirical studies to validate predictions from these theories.

Box 1. Interdisciplinary admixture: community resources and broader impacts

As we continue to generate large, multi-omic data, there is an increasing need for scalable, systematic community efforts to organize and curate these massive resources. Several crop based community resources are available as part of the AgBioData consortium [69], which provides guidelines for developing and maintaining agricultural databases. Projects within the umbrella of AgBioData include the germplasm resource information network (GRIN, https://www.ars-grin.gov), crop-based platforms like MaizeGDB [70], and ontology databases like Planteome (http://www.planteome.org). These resources not only provide researchers with access to centralized information and data, but also act as a medium to facilitate coordination among research groups (e.g. the Project Portal in MaizeGDB [71]). Other initiatives, such as *Genomes to Fields* [72], focus on translating research findings for public benefit by facilitating public-private partnerships, interdisciplinary collaborations, and communication with government agencies.

The absence of these centralized databases for many crop species places a burden on researchers to identify relevant work and to generate data essentially *de novo*, resulting in redundant efforts and a substantial bottleneck in crop research. Increased access to community resources with shared data repositories and best practices will be necessary to accelerate progress in domestication and breeding efforts. These databases and the corresponding metadata are essential to fully leverage existing data, to avoid duplicated efforts (e.g. sequencing the same genotype), and to reliably share data across research groups. Primary challenges developing these resources include funding, communication, and standardization, with specifics varying based on the breeding system and research goals for a given crop species. Below, we summarize standard community germplasm, data, and outreach resources related to crop domestication and breeding.

Germplasm

1. **Reliable stock sources** - Germplasm, or live genetic material, is the primary source of diversity for breeding programs and genetic research. The integrity, quality, and availability of these resources is curated by genebanks, such as the CGIAR Genebank Platform (https://www.genebanks.org/), GRIN (https://www.ars-grin.gov/), the Maize Genetics COOP (http://maizecoop.cropsci.uiuc.edu/), and the Tomato Genetics Resource Center (https://tgrc.ucdavis.edu/). Researchers can both submit samples for preservation

and request germplasm, which is often available to researchers under a Standard Material Transfer Agreement.

2. **Diversity panels** - These are collections of germplasm that represent a chosen range of diversity present in a crop species and can be tailored to specific goals of an investigation. For instance, evolutionary biologists may be interested in a diversity panel which represents the scope of diversity within a species and its wild relatives, while breeders may prefer a panel that includes diverse, but elite, breeding materials where alleles at major domestication loci have already reached fixation.

3. **Linkage mapping populations -** These populations are developed by crossing two parental lines and selfing the resultant hybrid to produce segregating populations or recombinant inbred lines (RILs), or by crossing many inbred lines in a directed fashion to generate a multi-parent advanced generation intercross (MAGIC) population [73]. The resultant lines can then be used to produce linkage maps and to identify quantitative trait loci (QTL).

Data resources

1. **Genome assemblies -** Platforms such as PacBio and Oxford Nanopore now enable high quality genome assemblies for a wide range of crop species (see, for example, [74]). This also opens the opportunity for multiple references in a given species, with greater potential for the assembly of pangenomes and genome graphs [75] that will provide a more accurate representation of crop diversity.

2. **Resequencing data -** It is now financially feasible to generate low coverage whole genome sequencing (WGS), transcriptome data, and reduced representation data (e.g. genotyping-by-sequencing, restriction site associated DNA sequencing) at the population level for a wide range of species. This data can then be paired with an ever-increasing suite of software tools.

3. **Phenotype data** - The type of data collected will depend on project goals, with those focused on crop breeding more concerned with yield, abiotic/biotic stress, and quality. Data types can include manual measurements, images, and physiological traits.

4. **Standards for data collection and curation** - Optimal use of data across disciplines requires adherence to a set of best practices for collecting data across field sites, lab protocols, formatting code, reporting metadata, and maintaining data repositories (ideally open access) for future use. A community standard will also enhance reproducibility and ensure that other researchers can reuse data with sufficient knowledge of how it was produced. Helpful resources include guidelines on organizing and minimizing error in spreadsheets [76] and the best practices outlined by AgBioData [69].
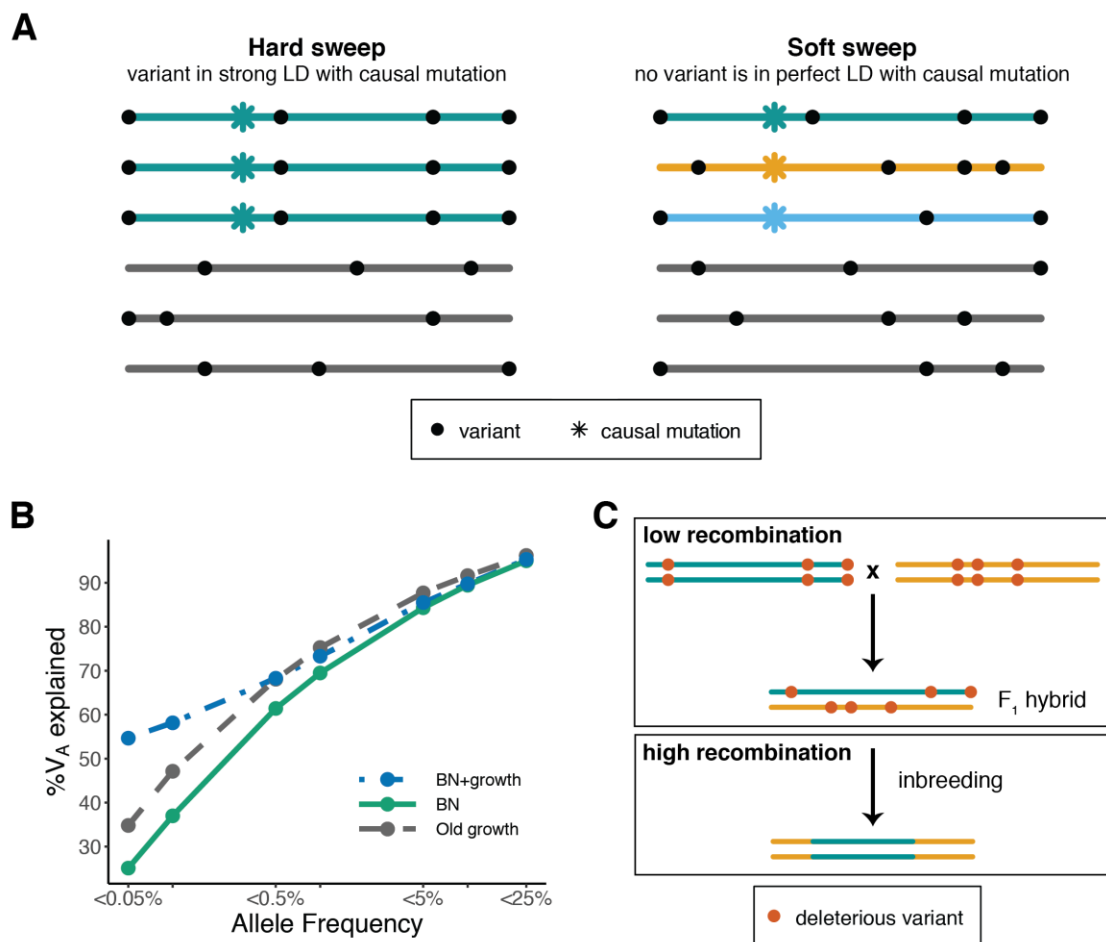
Outreach and broader impacts

1. **Recruitment of diverse scientists -** The interdisciplinary nature of plant domestication research is an incredible opportunity to recruit new and diverse talent to the plant sciences [77] and to generate interest in the applied science of plant breeding. The generation of public resources as advocated here is also an excellent platform to promote equity, especially for early career scientists, by providing rapid access to data repositories. For instance, access to existing data will help to accelerate preliminary analyses for granting cycles and to facilitate contributions from researchers without access to large financial resources and infrastructure.

2. **Classroom engagement** - The topic of crop domestication has a natural and interesting connection to human culture and history, which can help to engage students from a wide variety of backgrounds. For instance, in an evolutionary context, crop domestication is a non-controversial example of evolution, albeit by artificial rather than natural selection. In a more applied setting, crop improvement has a direct impact on food systems, which are a shared connection of the human experience. Development of lesson plans and storyboards related to crop domestication and breeding can be accomplished through partnerships between researchers and experts in science education. These resources can then be publicly disseminated for use by local institutions.

3. **Public dialogue -** Evolution and crop origins are a fascinating area of study with the potential to engage public interest in science and food systems, especially regarding the source of modern crops. Increasing transparency between scientists and the public via historical and ongoing domestication can facilitate public dialogue on gene editing.

Community repositories could make outreach resources available to multiple institutions and a wider audience by including guidelines for successful approaches, e.g. through the Plant Genome Research Outreach Portal, http://www.plantgdb.org/PGROP [78]. Providing access to these resources will reduce the need for researchers to develop outreach activities from the ground up.

Box 2. Wild relatives keep plant origins in the weeds

Crop wild relatives represent both the raw material for breeding and a valuable source of diversity. This diversity can be harnessed to introgress novel alleles for disease resistance and other quality traits into improved germplasm [4,79–81] (see review in [82] on the conservation and utility of crop wild relatives for breeding). As such, these wild populations are a critical component to understand how selection for traits important to agriculture has shaped the genome of crop species during domestication and improvement [79,83,84]. Key challenges in studying wild relatives include the limited availability of germplasm and genetic data in many crop species [82] and the possibility that extant wild populations are not truly ancestral, but instead feral or admixed with cultivated germplasm [85]. This relationship is further complicated by the potential of gene flow with wild, feral, or wild-domesticate hybrid plants at multiple points over time, which has been reported for many crop species [86–90]. Recent work has also demonstrated that feralization of cultivated crops can occur through different mechanisms, such as adaptive introgression from wild relatives (e.g. eggplant [87]) or by selection on a small number of genomic regions at various stages of the domestication process (e.g. weedy rice [91]). Moving forward, an understanding of crop wild relative diversity and the extent of introgression with domesticated germplasm will be necessary to identify genomic regions for additional crop improvement, as well as to prevent the establishment and spread of weedy populations. This will require increasing the representation of wild relatives in crop germplasm collections and the development of tools to understand the extent of hybridization between wild and domesticated populations.

Figure 1. (**A**) Implications of hard and soft sweeps for genetic mapping. On the left, a trait is controlled by a hard sweep and a marker is in strong linkage disequilibrium (LD) with the causal variant. On the right is a soft sweep scenario with multiple origins, where the causal variant occurs on different haplotypes (represented by different colors). In this case, it is more difficult to reliably identify a marker that is in perfect LD with the causal variant. (**B**) Example demonstrating the influence of demographic history on the proportion of additive genetic variance explained by SNPs at different allele frequencies (figure reproduced from [45]). When a SNP effect is strongly correlated with fitness, rare alleles explain more additive genomic variance for populations that experience a bottleneck followed by recent growth (BN+growth) compared to populations with a recent bottleneck (BN) or no recent growth (old growth). (**C**) The relationship of deleterious variants and recombination. In this example, when recombination is low, deleterious variants persist in the population. Assuming most deleterious variants are on average partially recessive [50], their effects can be masked via hybridization if the same deleterious variants do not occur in each haplotype (complementation). In contrast, when recombination is high deleterious variants can be quickly purged during the inbreeding process.

ANNOTATED REFERENCES: The majority of the references (please aim to cite approximately 50) should come from the period under review (i.e. the past two years) and, in general, at least 10% of these should be selected and annotated as being papers of special interest (*) or outstanding interest (**). Annotated references MUST be from the past two years, and the annotation should provide a brief description of the major findings and the importance of the study. This is an essential part of each review and is very popular with our readers.

Papers of outstanding interest:

** 29. Stetter MG, Thornton K, Ross-Ibarra J: **Genetic architecture and selective sweeps after polygenic adaptation to distant trait optima**. *PLoS Genet* 2018, **14**:e1007794.

- Simulations across a wide parameter space demonstrate the effects of demographic history on the speed polygenic adaptation, selective sweeps, and present genetic architecture.

** 32. Hermisson J, Pennings PS: **Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation**. *Methods in Ecology and Evolution* 2017, **8**:700–716.

- An overview of theoretical expectations for selective sweeps and challenges in their detection. Observations in species with drastically different population histories (*Drosophila*, humans, and microbes) demonstrate that soft sweeps are widespread, but also exemplify the challenges in distinguishing between 'soft' and 'hard' patterns.

** 59. Li T, Yang X, Yu Y, Si X, Zhai X, Zhang H, Dong W, Gao C, Xu C: **Domestication of wild tomato is accelerated by genome editing**. *Nat Biotechnol* 2018, doi:10.1038/nbt.4273.

- Gene editing enabled the introduction of desirable domestication traits into four wild accessions of tomato while maintaining valuable traits like salt and disease tolerance.

** 62.   Mieulet D, Aubert G, Bres C, Klein A, Droc G, Vieille E, Rond-Coissieux C, Sanchez M, Dalmais M, Mauxion J-P, et al.: **Unleashing meiotic crossovers in crops**. *Nat Plants* 2018, **4**:1010–1016.

- Mutagenesis of genes that limit meiotic recombination, specifically *RECQ4* orthologs, produced a three-fold increase in crossovers in rice, pea, and tomato, suggesting that similar approaches can be applied in other crop species to promote recombination.

Papers of special interest:

* 52.   Zhang C, Wang P, Tang D, Yang Z, Lu F, Qi J, Tawari NR, Shang Y, Li C, Huang S: **The genetic basis of inbreeding depression in potato**. *Nat Genet* 2019, **51**:374–378.

● Analysis of mutational load in 151 diploid potato lines reveals that large effect deleterious variants are line-specific, with the potential to purge deleterious load through recombination and hybridization.

* 61.   Johnsson M, Gaynor RC, Jenko J, Gorjanc G, de Koning D-J, Hickey JM: **Removal of alleles by genome editing (RAGE) against deleterious load**. *Genet Sel Evol* 2019, **51**:14.

● Performed simulations of animal breeding populations to test multiple scenarios whereby deleterious variants could be removed via gene editing. Findings suggest that, while removal of deleterious load via gene editing has potential, simultaneous editing within a single parent is required when deleterious variants are recessive.

* 85.   Wang H, Vieira FG, Crawford JE, Chu C, Nielsen R: **Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice**. *Genome Res* 2017, **27**:1029–1038.

● Analysis of admixture in 203 domesticated and 446 wild rice accessions reveals that modern wild rice experienced continuous gene flow from domesticated rice, with the possibility that the true ancestral form has been replaced by feral rice.

REFERENCES

1. Purugganan MD: **Evolutionary insights into the nature of plant domestication**. *Curr Biol* 2019, **29**:R705–R714.

2. Larson G, Piperno DR, Allaby RG, Purugganan MD, Andersson L, Arroyo-Kalin M, Barton L, Climer Vigueira C, Denham T, Dobney K, et al.: **Current perspectives and the future of domestication studies**. *Proc Natl Acad Sci U S A* 2014, **111**:6139–6146.

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

4. Hufford MB, Berny Mier Y Teran JC, Gepts P: **Crop biodiversity: An unfinished magnum opus of nature**. *Annu Rev Plant Biol* 2019, **70**:727–751.

5. Smýkal P, Nelson MN, Berger JD, Von Wettberg EJB: **The impact of genetic changes during crop domestication**. *Agronomy* 2018, **8**:119.

6. Hammer K: **The domestication syndrome**. *Kulturpflanze* 1984, **32**:11–34.

7. Olsen KM, Wendel JF: **A bountiful harvest: genomic insights into crop domestication phenotypes**. *Annu Rev Plant Biol* 2013, **64**:47–70.

8. Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia J-M, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, et al.: **Comparative population genomics of maize domestication and improvement**. *Nat Genet* 2012, **44**:808–811.

9. Hufford KM, Canaran P, Ware DH, McMullen MD, Gaut BS: **Patterns of selection and tissue-specific expression among maize domestication and crop improvement loci**. *Plant Physiol* 2007, **144**:1642–1653.

10. Meyer RS, DuVal AE, Jensen HR: **Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops: Tansley review**. *New Phytol* 10/2012, **196**:29–48.

11. Warschefsky EJ, von Wettberg EJB: **Population genomic analysis of mango (*Mangifera indica*) suggests a complex history of domestication**. *New Phytol* 2019, **222**:2023–2037.

12. Miller AJ, Gross BL: **From forest to field: perennial fruit crop domestication**. *Am J Bot* 2011, **98**:1389–1414.

13. Gaut BS, Díez CM, Morrell PL: **Genomics and the contrasting dynamics of annual and perennial domestication**. *Trends Genet* 2015, **31**:709–719.

14. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al.: **Ancestral polyploidy in seed plants and angiosperms**. *Nature* 2011, **473**:97–100.

15. Blischak PD, Mabry ME, Conant GC, Pires JC: **Integrating networks, phylogenomics, and population genomics for the study of polyploidy**. *Annu Rev Ecol Evol Syst* 2018, **49**:253–278.

16. Renny-Byfield S, Rodgers-Melnick E, Ross-Ibarra J: **Gene fractionation and function in the ancient subgenomes of maize**. *Molecular Biology and Evolution* 2017, **34**:1825–1832.

17. Pont C, Wagner S, Kremer A, Orlando L, Plomion C, Salse J: **Paleogenomics: reconstruction of plant evolutionary trajectories from modern and ancient DNA**. *Genome Biol* 2019, **20**:29.

18. Janzen GM, Wang L, Hufford MB: **The extent of adaptive wild introgression in crops**. *New Phytol* 2019, **221**:1279–1288.

19. Ewing PM, Runck BC, Kono TYJ, Kantar MB: **The home field advantage of modern plant breeding**. *PLoS One* 2019, **14**:e0227079.

20. Pritchard JK, Di Rienzo A: **Adaptation - not by sweeps alone**. *Nat Rev Genet* 2010, **11**:665–667.

21. Uricchio LH: **Evolutionary perspectives on polygenic selection, missing heritability, and GWAS**. *Hum Genet* 2019, doi:10.1007/s00439-019-02040-6.

22. Maynard Smith J, Haigh J: **The hitch-hiking effect of a favourable gene**. *Genet Res* 1974/002/001, **23**:23–35.

23. Begun DJ, Aquadro CF: **Molecular population genetics of the distal portion of the X chromosome in Drosophila: evidence for genetic hitchhiking of the yellow-achaete region**. *Genetics* 1991, **129**:1147.

24. Kim Y, Stephan W: **Selective sweeps in the presence of interference among partially linked loci**. *Genetics* 2003, **164**:389–398.

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

26. Zhong L, Yang Q, Yan X, Yu C, Su L, Zhang X, Zhu Y: **Signatures of soft sweeps across the Dt1 locus underlying determinate growth habit in soya bean [*Glycine max* (L.) Merr.]**. *Mol Ecol* 2017, **26**:4686–4699.

27. Pennings PS, Hermisson J: **Soft Sweeps II—molecular population genetics of adaptation from recurrent mutation or migration**. *Mol Biol Evol* 2006, **23**:1076–1084.

28. Beissinger TM, Hirsch CN, Vaillancourt B, Deshpande S, Barry K, Buell CR, Kaeppler SM, Gianola D, de Leon N: **A genome-wide scan for evidence of selection in a maize population under long-term artificial selection for ear number**. *Genetics* 2014, **196**:829–840.

29. Stetter MG, Thornton K, Ross-Ibarra J: **Genetic architecture and selective sweeps after polygenic adaptation to distant trait optima**. *PLoS Genet* 2018, **14**:e1007794.

30. Garud NR, Rosenberg NA: **Enhancing the mathematical properties of new haplotype homozygosity statistics for the detection of selective sweeps**. *Theor Popul Biol* 2015, **102**:94–101.

31. Pavlidis P, Alachiotis N: **A survey of methods and tools to detect recent and strong positive selection**. *J Biol Res* 2017, **24**:7.

32. Hermisson J, Pennings PS: **Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation**. *Methods in Ecology and Evolution* 2017, **8**:700–716.

33. Soyk S, Lemmon ZH, Oved M, Fisher J, Liberatore KL, Park SJ, Goren A, Jiang K, Ramos A, van der Knaap E, et al.: **Bypassing negative epistasis on yield in tomato imposed by a domestication gene**. *Cell* 2017, **169**:1142–1155.e12.

34. Dempewolf H, Baute G, Anderson J, Kilian B, Smith C, Guarino L: **Past and future use of wild relatives in crop breeding**. *Crop Sci* 2017, **57**:1070–1082.

35. Brozynska M, Furtado A, Henry RJ: **Genomics of crop wild relatives: expanding the gene pool for crop improvement**. *Plant Biotechnol J* 2016, **14**:1070–1085.

36. Lopes MS, El-Basyoni I, Baenziger PS, Singh S, Royo C, Ozbek K, Aktas H, Ozer E, Ozdemir F, Manickavelu A, et al.: **Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change**. *J Exp Bot* 2015, **66**:3477–3486.

37. Doebley J: **Isozymic evidence and the evolution of crop plants**. In *Isozymes in Plant Biology*. Edited by Soltis DE, Soltis PS, Dudley TR. Springer Netherlands; 1989:165–191.

38. Gaut BS, Seymour DK, Liu Q, Zhou Y: **Demography and its effects on genomic variation in crop domestication**. *Nat Plants* 2018, **4**:512–520.

39. Coop G: **Does linked selection explain the narrow range of genetic diversity across species?** *bioRxiv* 2016, doi:10.1101/042598.

40. Doebley JF, Gaut BS, Smith BD: **The molecular genetics of crop domestication**. *Cell* 2006, **127**:1309–1321.

41. Beissinger TM, Wang L, Crosby K, Durvasula A, Hufford MB, Ross-Ibarra J: **Recent demography drives changes in linked selection across the maize genome**. *Nature Plants* 2016, **2**:16084.

42. Fu Y-B: **Understanding crop genetic diversity under modern plant breeding**. *Theor Appl Genet* 2015, **128**:2131–2142.

43. MacLeod IM, Hayes BJ, Goddard ME: **The effects of demography and long-term**

selection on the accuracy of genomic prediction with sequence data. *Genetics* 2014, **198**:1671–1684.

44. Durvasula A, Lohmueller KE: **Negative selection on complex traits limits genetic risk prediction accuracy between populations**. *bioRxiv* 2019, doi:10.1101/721936.

45. Lohmueller KE: **The impact of population demography and selection on the genetic architecture of complex traits**. *PLoS Genet* 2014, **10**:e1004379.

46. Moyers BT, Morrell PL, McKay JK: **Genetic costs of domestication and improvement**. *J Hered* 2018, **109**:103–116.

47. Wang L, Beissinger TM, Lorant A, Ross-Ibarra C, Ross-Ibarra J, Hufford MB: **The interplay of demography and selection during maize domestication and expansion**. *Genome Biol* 2017, **18**:215.

48. Kono TJY, Fu F, Mohammadi M, Hoffman PJ, Liu C, Stupar RM, Smith KP, Tiffin P, Fay JC, Morrell PL: **The role of deleterious substitutions in crop genomes**. *Mol Biol Evol* 2016, **33**:2307–2317.

49. Liu Q, Zhou Y, Morrell PL, Gaut BS: **Deleterious variants in Asian rice and the potential cost of domestication**. *Mol Biol Evol* 2017, **34**:908–924.

50. Yang J, Mezmouk S, Baumgarten A, Buckler ES, Guill KE, McMullen MD, Mumm RH, Ross-Ibarra J: **Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize**. *PLoS Genet* 2017, **13**:e1007019.

51. Troyer AF, Wellin EJ: **Heterosis decreasing in hybrids: yield test inbreds**. *Crop Sci* 2009, **49**:1969–1976.

52. Zhang C, Wang P, Tang D, Yang Z, Lu F, Qi J, Tawari NR, Shang Y, Li C, Huang S: **The genetic basis of inbreeding depression in potato**. *Nat Genet* 2019, **51**:374–378.

53. Ramu P, Esuma W, Kawuki R, Rabbi IY, Egesi C, Bredeson JV, Bart RS, Verma J, Buckler ES, Lu F: **Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation**. *Nat Genet* 2017, **49**:959–963.

54. Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES: **Open chromatin reveals the functional maize genome**. *Proc Natl Acad Sci U S A* 2016, **113**:E3177–84.

55. Zhang M, Zhou L, Bawa R, Suren H, Holliday JA: **Recombination rate variation, hitchhiking, and demographic history shape deleterious load in poplar**. *Mol Biol Evol* 2016, **33**:2899–2910.

56. Blary A, Jenczewski E: **Manipulation of crossover frequency and distribution for plant breeding**. *Theor Appl Genet* 2019, **132**:575–592.

57. Voss-Fels KP, Stahl A, Wittkop B, Lichthardt C, Nagler S, Rose T, Chen T-W, Zetzsche H,

Seddig S, Majid Baig M, et al.: **Breeding improves wheat productivity under contrasting agrochemical input levels**. *Nat Plants* 2019, **5**:706–714.

58. Wang B, Zhu L, Zhao B, Zhao Y, Xie Y, Zheng Z, Li Y, Sun J, Wang H: **Development of a haploid-inducer mediated genome editing system for accelerating maize breeding**. *Mol Plant* 2019, **12**:597–602.

59. Li T, Yang X, Yu Y, Si X, Zhai X, Zhang H, Dong W, Gao C, Xu C: **Domestication of wild tomato is accelerated by genome editing**. *Nat Biotechnol* 2018, doi:10.1038/nbt.4273.

60. Rodríguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB: **Engineering quantitative trait variation for crop improvement by genome editing**. *Cell* 2017, **171**:470–480.e8.

61. Johnsson M, Gaynor RC, Jenko J, Gorjanc G, de Koning D-J, Hickey JM: **Removal of alleles by genome editing (RAGE) against deleterious load**. *Genet Sel Evol* 2019, **51**:14.

62. Mieulet D, Aubert G, Bres C, Klein A, Droc G, Vieille E, Rond-Coissieux C, Sanchez M, Dalmais M, Mauxion J-P, et al.: **Unleashing meiotic crossovers in crops**. *Nat Plants* 2018, **4**:1010–1016.

63. Lemmon ZH, Reem NT, Dalrymple J, Soyk S, Swartwood KE, Rodriguez-Leal D, Van Eck J, Lippman ZB: **Rapid improvement of domestication traits in an orphan crop by genome editing**. *Nat Plants* 2018, **4**:766–770.

64. Zsögön A, Cermak T, Voytas D, Peres LEP: **Genome editing as a tool to achieve the crop ideotype and de novo domestication of wild relatives: Case study in tomato**. *Plant Sci* 2017, **256**:120–130.

65. Zhu G, Wang S, Huang Z, Zhang S, Liao Q, Zhang C, Lin T, Qin M, Peng M, Yang C, et al.: **Rewiring of the fruit metabolome in tomato breeding**. *Cell* 2018, **172**:249–261.e12.

66. Olsen KM, Caicedo AL, Polato N, McClung A, McCouch S, Purugganan MD: **Selection under domestication: evidence for a sweep in the rice waxy genomic region**. *Genetics* 2006, **173**:975–983.

67. Palaisa K, Morgante M, Tingey S, Rafalski A: **Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep**. *Proc Natl Acad Sci U S A* 2004, **101**:9885–9890.

68. Crews TE, Cattani DJ: **Strategies, advances, and challenges in breeding perennial grain crops**. *Sustain Sci Pract Policy* 2018, **10**:2192.

69. Harper L, Campbell J, Cannon EKS, Jung S, Poelchau M, Walls R, Andorf C, Arnaud E, Berardini TZ, Birkett C, et al.: **AgBioData consortium recommendations for sustainable genomics and genetics databases for agriculture**. *Database* 2018, **2018**.

70. Lawrence CJ, Dong Q, Polacco ML, Seigfried TE, Brendel V: **MaizeGDB, the community database for maize genetics and genomics**. *Nucleic Acids Res* 2004, **32**:D393–7.

71. Cannon EKS, Birkett SM, Braun BL, Kodavali S, Jennewein DM, Yilmaz A, Antonescu V, Antonescu C, Harper LC, Gardiner JM, et al.: **POPcorn: an online resource providing access to distributed and diverse maize project data**. *Int J Plant Genomics* 2011, **2011**:923035.

72. Lawrence-Dill CJ, Schnable PS, Springer NM: **Idea factory: the maize genomes to fields initiative**. *Crop Sci* 2019, **59**:1406–1410.

73. Huang BE, Verbyla KL, Verbyla AP, Raghavan C, Singh VK, Gaur P, Leung H, Varshney RK, Cavanagh CR: **MAGIC populations in crops: current status and future prospects**. *Theor Appl Genet* 2015, **128**:999–1017.

74. Belser C, Istace B, Denis E, Dubarry M, Baurens F-C, Falentin C, Genete M, Berrabah W, Chèvre A-M, Delourme R, et al.: **Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps**. *Nat Plants* 2018, **4**:879–887.

75. Rakocevic G, Semenyuk V, Lee W-P, Spencer J, Browning J, Johnson IJ, Arsenijevic V, Nadj J, Ghose K, Suciu MC, et al.: **Fast and accurate genomic analyses using genome graphs**. *Nat Genet* 2019, **51**:354–362.

76. Broman KW, Woo KH: **Data organization in spreadsheets**. *Am Stat* 2018, **72**:2–10.

77. Gepts P, Harlan JR, Famula TR, Bettinger RL, Brush SB, Damania AB, McGuire PE, Qualset CO: *Biodiversity in Agriculture: Domestication, Evolution, and Sustainability*. Cambridge University Press; 2012.

78. Baran SB, Lawrence CJ, Brendel V: **Plant genome research outreach portal. A gateway to plant genome research "outreach" programs and activities**. *Plant Physiol* 2004, **134**:889.

79. Schreiber M, Stein N, Mascher M: **Genomic approaches for studying crop evolution**. *Genome Biol* 2018, **19**:140.

80. Tanksley SD, McCouch SR: **Seed banks and molecular maps: unlocking genetic potential from the wild**. *Science* 1997, **277**:1063–1066.

81. McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E, Burke JM, Charest D, Cloutier S, Cole G, et al.: **Agriculture: Feeding the future**. *Nature* 2013, **499**:23–24.

82. Zhang H, Mittal N, Leamy LJ, Barazani O, Song B-H: **Back into the wild-Apply untapped genetic diversity of wild relatives for crop improvement**. *Evol Appl* 2017, **10**:5–24.

83. Dempewolf H, Eastwood RJ, Guarino L, Khoury CK, Müller JV, Toll J: **Adapting agriculture to climate change: a global initiative to collect, conserve, and use crop wild**

relatives. *Agroecology and Sustainable Food Systems* 2014, **38**:369–377.

84. Meyer RS, Purugganan MD: **Evolution of crop species: genetics of domestication and diversification**. *Nat Rev Genet* 2013, **14**:840–852.

85. Wang H, Vieira FG, Crawford JE, Chu C, Nielsen R: **Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice**. *Genome Res* 2017, **27**:1029–1038.

86. Choi JY, Platts AE, Fuller DQ, Hsing Y-I, Wing RA, Purugganan MD: **The rice paradox: multiple origins but single domestication in asian rice**. *Mol Biol Evol* 2017, **34**:969–979.

87. Page A, Gibson J, Meyer RS, Chapman MA: **Eggplant domestication: pervasive gene flow, feralization, and transcriptomic divergence**. *Mol Biol Evol* 2019, **36**:1359–1372.

88. Velázquez-López R, Wegier A, Alavez V, Pérez-López J, Vázquez-Barrios V, Arroyo-Lambaer D, Ponce-Mendoza A, Kunin WE: **The mating system of the wild-to-domesticated complex of *Gossypium hirsutum* l. is mixed**. *Front Plant Sci* 2018, **9**:574.

89. Flowers JM, Hazzouri KM, Gros-Balthazard M, Mo Z, Koutroumpa K, Perrakis A, Ferrand S, Khierallah HSM, Fuller DQ, Aberlenc F, et al.: **Cross-species hybridization and the origin of North African date palms**. *Proc Natl Acad Sci U S A* 2019, **116**:1651–1658.

90. Remigereau M-S, Lakis G, Rekima S, Leveugle M, Fontaine MC, Langin T, Sarr A, Robert T: **Cereal domestication and evolution of branching: evidence for soft selection in the Tb1 orthologue of pearl millet (*Pennisetum glaucum* [L.] R. Br.)**. *PLoS One* 2011, **6**:e22404.

91. Li L-F, Li Y-L, Jia Y, Caicedo AL, Olsen KM: **Signatures of adaptation in the weedy rice genome**. *Nat Genet* 2017, **49**:811–814.