

REVIEW

Open Access



Genomic approaches for studying crop evolution

Mona Schreiber¹, Nils Stein¹ and Martin Mascher^{1,2*} 

Abstract

Understanding how crop plants evolved from their wild relatives and spread around the world can inform about the origins of agriculture. Here, we review how the rapid development of genomic resources and tools has made it possible to conduct genetic mapping and population genetic studies to unravel the molecular underpinnings of domestication and crop evolution in diverse crop species. We propose three future avenues for the study of crop evolution: establishment of high-quality reference genomes for crops and their wild relatives; genomic characterization of germplasm collections; and the adoption of novel methodologies such as archaeogenetics, epigenomics, and genome editing.

Introduction

Since the Neolithic, humans have domesticated a large number of different plant species to create a reliable source of nutrition for themselves and their domestic animals. Crop plants comprise a large variety of species from diverse taxa that differ in habitat, growth habit, and life cycle, such as annual grasses, perennial trees, and medicinal herbs (Table 1, Fig. 1). However, world-wide crop production is dominated by a few major crops, such as wheat, rice, maize, potato, sugar cane, and soybean [1], that serve globally as staples for human and animal nutrition. By contrast, minor crops can be broadly defined as a non-homogeneous group comprising staple crops traditionally only of regional importance, such as quinoa, teff, and African rice; or crops of world-wide importance but comparatively little contribution to human food consumption such as nuts or small fruits. Active research and breeding communities exist for almost every crop plant; however, research into

the molecular genetics of domestication has focused mainly on the major crops [2].

Crop domestication has been studied for more than a century and benefited recently from technological innovations in genomics. Comparative analysis of population genomic data of large samples of current and past varieties of crops together with their wild progenitors provides insight into the domestication history of species [3, 4], for example, (i) when and where domestication occurred, (ii) how the domesticates spread to new habitats and which genetic changes accompanied this transition, and (iii) whether gene flow has occurred between the crops and wild relatives. A synthesis of archaeological and population genetic data evidence indicated that the initial stages of domestication in Southwest Asia should be considered a protracted process [5, 6] rather than a rapid evolution of cultivated plants as presumed previously [7–9]. The “democratization” of genomics [10, 11] has now opened new avenues for understanding the genetic consequences of domestication in a much wider range of species from different centers of origin such as Mesoamerica and Africa.

Much has been written on plant domestication. Recent review articles have focused on convergent phenotypic evolution [12], causative mutations affecting phenotypic variation [13, 14], the effect of gene functions on crop adaptation and selection mechanisms [15], the reduction of genetic diversity and the influence of epigenetic modifications [16], the impact of genomic methods in future crop improvement [17], the value of crop wild relatives [18], sequencing ancient plant DNA [19, 20], and general concepts in plant domestication research [21–23]. Here, we focus on the assembly of reference genome sequences for domesticated plants and their wild relatives; surveying sequence diversity in large diversity panels; and the application of novel approaches such as epigenomics, archaeogenetics, and genome editing to plant domestication research.

* Correspondence: mascher@ipk-gatersleben.de

¹Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstraße 3, 06466 Seeland, Germany

²German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany



Table 1 Examples of domesticated crops with domestication origin, available reference genomes and sequencing strategies.

Crop	Botanical name	Lifecycle	Ploidy level	Time of domestication	Geographical origin	Reference genome	Genome size	Sequencing strategy
African rice	<i>Oryza glaberrima</i>	Annual	Diploid	~3 000 BP	Upper Niger River?	Wang et al. [200]	~860 Mb	Sanger, Roche/454
Amaranth	<i>Amaranthus</i> spp.	Annual	Tetraploid	Aztecs? ~13 th – 15 th century	Mesoamerica	Clouse et al. [46]	~456 Mb	Illumina, physical map
Apple	<i>Malus x domestica</i>	Perennial	Diploid; polyploid karyotypes exist	Late Bronze Age? ~2 000 – 1 500 BC	West-Asia	Daccord et al. [157]	~651 Mb	Illumina, PacBio, optical map
Asian rice	<i>Oryza sativa</i>	Annual	Diploid	>6 200 BC	China	Kawahara et al. [201]	~500 Mb	Sanger, Illumina, Roche/454, optical mapping
Barley	<i>Hordeum vulgare</i>	Annual	Diploid	~10 000 BC	Southwest Asia	Mascher et al. [52]	~5 Gb	Illumina, optical mapping, genetic map, HiC
Beet	<i>Beta vulgaris</i>	Biennial	Diploid	Before 800 BC	Middle East	Dohm et al. [202]	~750 Mb	Roche/454, Illumina, Sanger
Bread wheat	<i>Triticum aestivum</i>	Annual	Hexaploid	~10 000 BC	Southwest Asia	Zimin et al. [65]	~17 Gb	Illumina, PacBio
Carrot	<i>Daucus carota</i>	Annual/biennial	Diploid	Classical antiquity	Central Asia (southwest Asia)	Iorizzo et al. [49]	~473 Mb	Illumina, Roche/454, linkage map
Cassava	<i>Manihot esculenta</i>	Annual	Diploid	10 000 – 5 000 BC in South America	Amazon basin	Wang et al. [203]	~770 Mb	Illumina, Roche/454
Chickpea	<i>Cicer arietinum</i>	Annual	Diploid	~10 000 BC	Fertile Crescent	Varshney et al. [204]	~740 Mb	Illumina
Cotton	<i>Gossypium hirsutum</i>	Perennial	Tetraploid	~6 000 – 5 000 BC	India/Mexico	Li et al. [205]	~2.4 Gb	Illumina, genetic map
Cowpea	<i>Vigna unguiculata</i>	Annual	Diploid	~4 000 BC	West Africa	Muñoz-Amatriain et al. [206]	~620 Mb	Illumina, physical map, genetic map
Emmer wheat	<i>Triticum dicoccon</i>	Annual	Tetraploid	~10 000 BC	Southwest Asia	Avni et al. [58]	~12 Gb	Illumina, genetic map, HiC
Ginseng	<i>Panax ginseng</i>	Perennial	Tetraploid	Cultivated since ~3000 BC	China	Jayakodi et al. [207]	~3.5 Gb	Illumina
Intermediate wheatgrass	<i>Thinopyrum intermedium</i>	Perennial	Hexaploid	~6 000 BC	Central and south-eastern Europe to Anatolia	No reference available	~12.75 Gb	Genetic map [208]
Maize	<i>Zea mays</i>	Annual	Diploid	~6 700 BC	Mexico	Jiao et al. [63]	~2.4 Gb	PacBio, optical mapping, Illumina
Oat	<i>Avena sativa</i>	Annual	Hexaploid	European Bronze Age ~1 800 BC	Fertile Crescent	In progress: The Oat Genome Project [209]	~12 Gb	
Pepper	<i>Capsicum</i> spp.	Annual	Diploid	~4 000 BC	South and Middle America	Qin et al. [99]	~3.5 Gb	Illumina, 10X, genetic map
Potato	<i>Solanum tuberosum</i>	Annual	Autotetraploid	~7 000 – 6 000 BC	Andean region	The Potato Genome Sequencing Consortium 2011 [210]	~844 Mb	Illumina, Roche/454, Sanger
Pumpkin	<i>Cucurbita maxima</i>	Annual	Tetraploid	~8 000 BC	South America	Sun et al. [47]	~378 Mb	Illumina, genet map
Quinoa	<i>Chenopodium quinoa</i>	Annual	Tetraploid	> 5 000 BC	Mesoamerica	Jarvis et al. [32]	~ 1.5 Gb	PacBio, Illumina, optical mapping
Rapeseed	<i>Brassica napus</i>	Annual	Tetraploid	~400 – 500 BP	Europe	Yang et al. [211]	~1.13 Gb	Illumina, PacBio, optical mapping
Raspberry	<i>Rubus idaeus</i>	Perennial	Tetraploid	~450 BP	Europe and northern Asia	Van Buren et al. [48]	~293 Mb (black raspberry)	Illumina, genetic map
Rye	<i>Secale cereale</i>	Annual	Diploid	European Bronze Age ~1 500 – 1 000 BC	Europe	Bauer et al. [212]	~7.9 Gb	Illumina, genetic map
Sorghum	<i>Sorghum bicolor</i>	Annual	Diploid	~7 100 – 6 900 BC	Sahel	McCormick et al. [213]	~730 Mb	Illumina, Roche/454, Sanger
Sugar cane	<i>Saccharum</i> spp.	Perennial	Allopolyploid	~8 000 BC	Southeast Asia	Riaño-Pachón et al. [214]	~10 Gb	Illumina

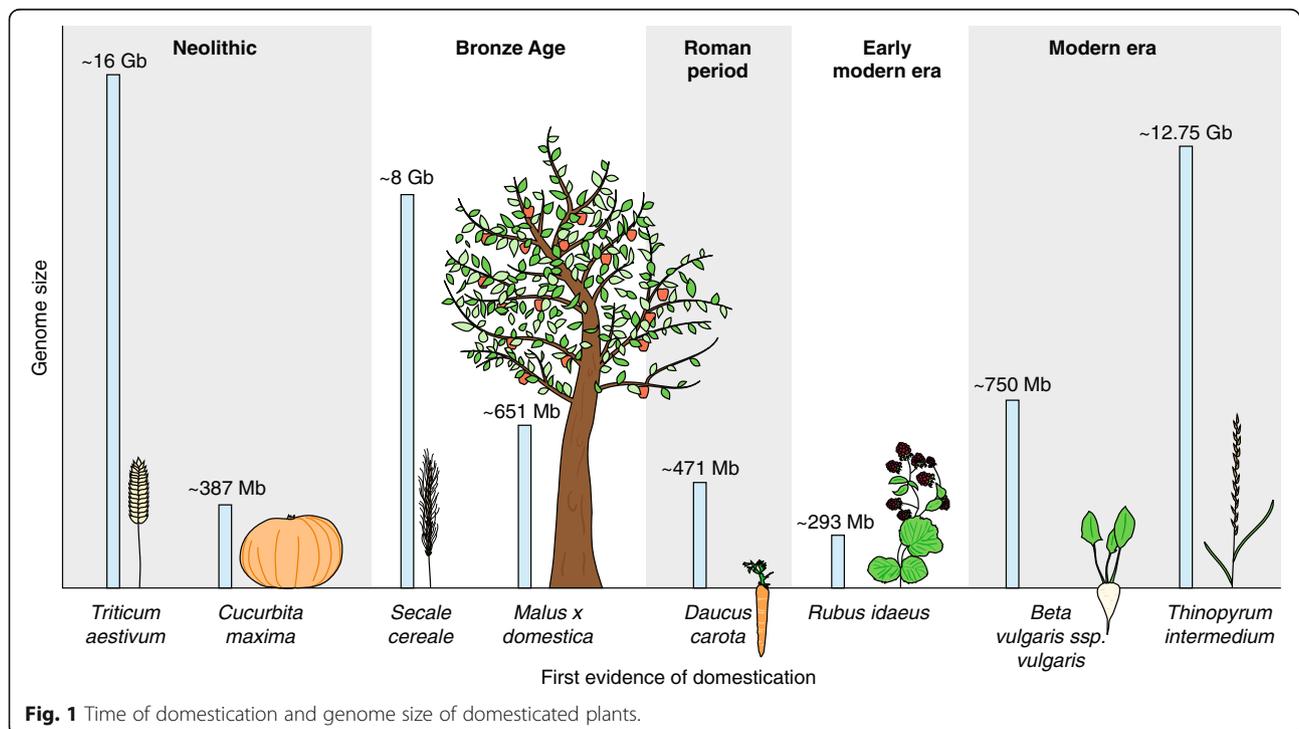
Table 1 Examples of domesticated crops with domestication origin, available reference genomes and sequencing strategies. (Continued)

Crop	Botanical name	Lifecycle	Ploidy level	Time of domestication	Geographical origin	Reference genome	Genome size	Sequencing strategy
Sunflower	<i>Helianthus annuus</i>	Annual	Diploid	~6 000 BC	Eastern North America	Badouin et al. [215]	~3.6 Gb	PacBio, genetic map
Tomato	<i>Solanum lycopersicum</i>	Annual	Diploid	Before 15 th century	Andean region	The Tomato Genome Sequencing Consortium 2012 [216]	~900 Mb	Roche/454, Illumina

High-quality reference sequence assemblies for crops and their wild relatives

Extant crop wild relatives are excellent tools to understand crop evolution and as sources of novel allelic diversity for future crop improvement [24, 25]. The wild gene pool of a crop includes its wild progenitor species, with which it is often fully interfertile, and species from the same or closely related genera that can differ greatly in divergence times and interfertility with the crop [26]. An unbiased assessment of genetic diversity in distantly related taxa using the single reference genome sequence of the domesticate is complicated by sequence divergence, which prevents the alignment of short reads, especially in non-coding regions. Comparison of short reads to a single reference will also not reveal structural variants such as chromosomal inversions and translocations. Thus, reference genome sequence assemblies of crop wild relatives are important tools to understand domestication history.

Genome sequencing and assembly have been applied to many different species for decades. In the context of plant domestication research, genomic resources such as high-quality reference genome sequences of crops and their wild relatives, but also dense genetic and physical maps, have provided the infrastructure for the genetic mapping of loci underlying key domestication traits [27, 28] and their subsequent isolation by map-based cloning [29–32]. Genome sequence assemblies serve as common references for the alignment of resequencing data from diversity panels comprising crops and their wild progenitors [33, 34], thus underpinning genome scans for phenotypic associations and for targets of selection under domestication [3, 35]. In the past, the large size, repeat-rich structure and polyploid nature of many crop genomes have been major impediments to the construction of contiguous sequence assemblies [36]. Here, we summarize recent developments in sequencing technology and computational methods that have contributed



to overcoming these long-standing obstacles; provide recent examples for the construction of high-quality reference for crops and their wild relatives; and outline future directions.

The traditional approach of laborious and time-consuming Sanger sequencing along a minimum tiling path of bacterial artificial chromosomes (BACs) [37] has been attempted for only a few crops with small genomes such as rice [38] or crops of the highest economic importance—and commensurate research funding—such as maize [39], wheat [40], and barley [41]. In the latter two species, progress in sequencing and genome mapping technologies has prompted the respective international sequencing consortia that set out years ago with physical map construction to revise their strategy by adopting short-read sequencing [42, 43]. Wet-lab protocols and computational methods for genome sequence assembly using short Illumina reads were established about a decade ago [44, 45]. The general approach of genome sequence assembly is shown in Fig. 2 and can be summarized as follows: (i) contig assembly from deep-coverage paired-end reads; (ii) scaffolding with mate-pair information; (iii) filling of sequence gaps introduced in this step; and (iv) high-level ordering of sequence scaffolds into so-called pseudomolecules as representatives of entire chromosomes ('superscaffolding'). A large number of crop plant genomes have been sequenced using this method, including minor crops such as amaranth [46],

pumpkin [47], raspberry [48], and carrot [49]. Genome assemblies for 12 species in the genus *Oryza*, i.e., wild and domesticated rice, and an outgroup species (*Leersia perrieri*) provided a comprehensive overview of structural genome evolution, thus contributing to the overarching aim of the International Oryza Map Alignment Project to establish genus-wide comparative genomics to discover genes for crop improvement [50].

In the large-genome cereals wheat, barley, and rye, short-read sequencing has been widely adopted to compile complementary datasets for higher-order scaffolding such as linkage maps [51, 52], physical maps [53], and chromosome-specific sequences [54]. However, strong reservations persisted against performing initial contig assembly with short reads only. This skepticism was vindicated by the mixed success of initial efforts in wheat and barley. The gene space was reasonably complete and approximately ordered along the genome with the help of genetic and physical maps; however, sequence contiguity was on the kilobase-scale and the repetitive portion of the genome was severely underrepresented [55–57]. Until recently, researchers were unwilling to take the risk of amassing the large amount of sequences required for short-read assembly of multi-gigabase crop genomes without a proven strategy for assembling them.

In an eye-opening study, Avni et al. [58] constructed a chromosome-scale assembly of the genome of tetraploid wild emmer (genome size 10 Gb) from very deep Illumina sequencing data from multiple paired-end, mate-pair libraries and chromosome-conformation capture sequencing data, thus establishing a precedent for the construction of a high-quality sequence assembly of a repeat-rich polyploid plant genome [59]. Several factors contributed to the feasibility and success of their approach, such as improved library construction methods ensuring uniform genome representation [60] and increases in throughput and read length of the Illumina platform (2 × 250 bp) accompanied by decreases in sequencing costs. One important caveat of the work by Avni et al. [58] is that the software used to construct sequence scaffolds with megabase-scale contiguity is the trade secret of a commercial service provider, NRGene. Open-source alternatives for Illumina-only sequence assembly exist and their results exceed assembly metrics of previous efforts by an order of magnitude [61], but they have not yet achieved the contiguity of the wild emmer assembly.

Long reads (> 10 kb) from the PacBio or Oxford Nanopore platforms are two orders of magnitude longer than Illumina reads (100–300 bp) but have reduced sequence accuracy [62]. Recently, PacBio sequencing was used to construct a new version of the maize reference genome [63], which achieved a higher contiguity and better genome representation than the previous

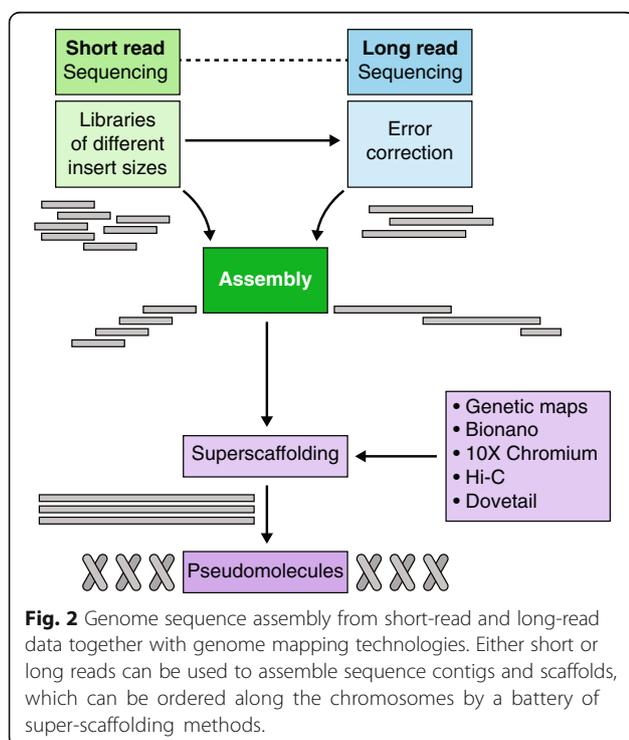


Fig. 2 Genome sequence assembly from short-read and long-read data together with genome mapping technologies. Either short or long reads can be used to assemble sequence contigs and scaffolds, which can be ordered along the chromosomes by a battery of super-scaffolding methods.

BAC-by-BAC assembly [39] and corrected many errors in the order and orientation of sequence contigs of its predecessor. Schmidt et al. [64] used Oxford Nanopore data to assemble the genome sequence of the tomato wild relative *Solanum pennellii* (genome size ~1 Gb). The resultant sequence scaffolds were highly contiguous (N50 2.5 Mb) but required correction with supplementary Illumina reads to increase accuracy at the single nucleotide level. Similarly, Zimin et al. [65] used a combination of PacBio and Illumina data to reconstruct the genome sequences of hexaploid bread wheat and its diploid progenitor *Aegilops tauschii* [66]. The N50 values of these assemblies are in the range of several hundred kilobases, which confirms that long-read sequences can produce better assembly than short-read technology even in the most complex genomes, but also highlights the necessity of obtaining short-read sequences for error correction and complementary mid- and long-range information to achieve chromosome-scale contiguity. Another important drawback of using inaccurate long reads are the immense computational requirements of the assembly process (>100,000 CPU hours for bread wheat [65]).

Until recently, there was a ‘contiguity gap’ in the assembly of complex plant genomes. The contiguity of sequence assemblies was limited to kilobase-sized contigs, but conventional methods for higher-order scaffolding and assembly validation such as genetic mapping [67, 68], cytogenetic methods [69, 70], BAC-based physical mapping [53, 71], or radiation hybrid mapping [72] are effective only at the scale of megabases. Moreover, these techniques are time-consuming and require resources that are difficult or sometimes impossible to generate (e.g., genetic maps in asexual organisms). Methodology has improved dramatically over the past 3 years. Deep-coverage short-read sequencing of paired-end and mate-pair can now deliver contigs of megabase-scale contiguity in complex plant genomes. Furthermore, a plethora of methods—many of them employing Illumina sequencing—has been developed to validate, correct, and improve initial sequence assembly from either long or short reads (Fig. 2). Among the first of these was optical mapping, a method that measures the distances of fluorescently labeled nicking sites in linearized long DNA molecules to construct megabase-scale contigs composed of molecules with matching labeling patterns [73]. Optical mapping on the Bionano platform has been used for scaffolding the PacBio assembly of the maize genome [63] and for higher-order scaffolding of BAC-based or chromosome-specific assemblies in wheat, barley, and tetraploid finger millet [42, 74, 75].

One of the limitations of the current Bionano platform is the need for input assemblies of high contiguity so that sequence scaffolds comprise a sufficient number of

label sites to confidently align sequence assembly and optical map. Technologies based on Illumina short-read sequencing can better control data density by increasing sequencing depth. One of these, the Chromium 10X platform, employs a microfluidic device to create so-called ‘linked reads’, which incorporate short barcode sequences that are shared by reads originating from the same high-molecular weight DNA fragment [76]. Chromium reads serve as mid-range linkage information in a similar manner as mate-pair reads or BAC end sequences, but with the added value of longer span size (>50 kb) and multiple linked short reads that support connections between sequence scaffolds to effectively bridge repetitive regions [77]. Intriguingly, deeply sequenced 10X libraries can be used on their own for genome assembly. After stripping of the barcode from the reads, 10X chromium data can play the role of short-read sequences for assembling by contigs, which are subsequently scaffolded by leveraging information on the molecular origin of reads [78]. Hulse-Kemp et al. [79] used this approach to assemble the genome sequence of bell pepper (*Capsicum annuum*) and achieved a scaffold N50 of 3.7 Mb.

Arguably, the most disruptive technology in recent assembly for genome mapping has been chromosome conformation capture sequencing (Hi-C). Hi-C was originally developed as a method to assay genome-wide chromatin contact probabilities [80], but it was soon realized that the linkage information afforded by chromatin proximity can be effectively used for scaffolding fragmented genome assemblies to chromosome-scale contiguity [81, 82]. In addition to ordering and orienting scaffolds, Hi-C can also effectively detect misassemblies. Chromatin contact probabilities between pairs of loci are strongly correlated to their distance in the linear genome [42, 80, 83]. Thus, the likely cause for any strong deviations from the expected rate of distance-dependent decay of contact probabilities are misassemblies [84]. Hi-C mapping made it possible to order sequence scaffolds in the large (>300 Mb) non-recombining proximal regions of the barley [42] and tetraploid wheat [58] genomes, for which the construction of high-resolution molecular marker maps of high density had remained elusive [85]. Lightfoot et al. [86] used a combination of Hi-C and PacBio long reads to increase the contiguity of the genome assembly of amaranth to chromosome-scale scaffolds. The principle of proximity ligation also underlies the Chicago method, which derives linkage information from Hi-C libraries constructed from chromatin reconstituted in vitro from high molecular weight DNA [87] and is offered commercially by Dovetail Genomics. Dovetail scaffolding was used to improve the assemblies of model plants [88, 89], such as lettuce [90], quinoa [32], and an individual chromosome of hexaploid wheat [91].

We believe it is possible to obtain for any plant taxon—wild or domesticated—a high-quality reference genome sequence assembly within a year’s time frame without prior resources. But, which is the most cost-effective combination of sequencing methods and genome mapping approaches [88, 89]? A contiguous, complete, and correctly annotated reference sequence will support research into the contribution of ancestral diversity in the wild progenitors, the footprints of artificial selection in the domesticate, and gene flow between wild and cultivated taxa.

Genomic characterization of germplasm collections

The wild progenitors of most crop plants remain extant [92] and can be collected from the wild. Furthermore, traditional landraces have been collected and stored in germplasm collections (so called ‘genebanks’ or ‘seed banks’) for more than a century. Germplasm collections can provide the raw material for population genomic studies to unravel the origin of crops, their demographic history, as well as past and present selection pressures. Several strategies based on high-throughput sequencing are available to catalogue and analyze genetic variation in crop diversity panels, namely whole-genome sequencing, exome capture, RNA sequencing, and reduced representation resequencing. The key difference between these approaches lies in the fraction of the genomes targeted for sequencing, which determines the requirements for prior resources and the per sample cost (Table 2).

The most straightforward method for assessing genetic diversity in species with a reference genome is shotgun sequencing of short-insert libraries on the Illumina platform and alignment of the sequence reads to the reference assembly [93]. Whole-genome shotgun sequencing has been used in major crops such as maize [3, 94], rice [33, 95, 96], soybean [97], and Solanaceous species [98, 99] to study genome-wide sequence diversity in the crop and its wild relatives with a focus on domestication history and the genetic basis of crop improvement. In rice, whole-genome sequencing and phenotyping of diversity panels, together with efficient transgenic methods for testing candidate gene function, has emerged as a powerful approach for isolating genes of agronomic importance [95, 100]. Deep-coverage whole-genome sequencing data of multiple individuals is a prerequisite for advanced population genetic methods to infer demographic history such as the Pairwise-Sequentially Markovian Coalescent model [101] and its derivatives [102, 103], which were used to study historic changes of population size in maize [104], grapevine [105], African rice [106], and their respective wild progenitors.

An important drawback of whole-genome sequencing is the financial burden associated with amassing sufficient sequence data for large diversity panels in minor crops or those with large genomes. Several techniques have been developed to reduce genome complexity prior to sequencing so as to increase read depth in certain genomic regions. Sequence capture with oligonucleotide baits can be applied at the scale of whole exome [107] or gene families [108]. For example, whole-exome capture assays have been developed in wheat [109, 110] and barley [111] and applied in population genetic studies. Russell et al. [112] analyzed exome sequences from >250 wild and domesticated barleys to understand contrasting patterns of diversity in both taxa and to study changes in haplotype structure of flowering time genes during range expansion. Similarly, Avni et al. [58] used exome capture data from 65 accessions of wild and domesticated emmer to detect genomic regions under selection. Resistance gene enrichment sequencing (RenSeq) was originally developed to enable rapid mapping of nucleotide binding-site leucine-rich repeat resistance genes in mutant collections and segregation populations [108, 113], but has been recently adopted by Arora et al. [114] for association genetics and applied to gene isolation in *Aegilops tauschii*, a wild diploid progenitor of bread wheat. Witek et al. [115] combined resistance gene enrichment sequencing with PacBio sequencing to clone a resistance gene against potato late blight disease in *Solanum americanum*, a diploid, non-tuber bearing wild relative of potato.

High-throughput RNA sequencing (RNA-seq) [116] is an alternative to whole-exome capture that does not require the design of oligonucleotide baits and can assess sequence variation for a large portion of the transcribed part of the genome. For example, Nabholz et al. [117] used RNA-seq to study domestication bottlenecks in African rice. As a method to quantify transcript abundance, RNA sequencing affords information on gene expression levels in addition to genetic variation. Koenig et al. [34] performed RNA-seq on tomato and related *Solanum* species to detect footprints of selection based on genetic variation at the DNA sequence level, and also found selection pressure on gene expression level in the domesticate. Lemmon et al. [118] analyzed allele-specific expression in F₁ crosses between maize and teosinte to understand the changes in the regulatory architecture of gene expression as a consequence of domestication.

Reduced representation sequencing [119] is known by many names such as RAD-seq [120], genotyping-by-sequencing (GBS) [121], or SLAF-seq [122]. The common denominator of all these methods is genomic complexity reduction by digestion with restriction enzymes and subsequent short-read sequencing of fragments bordering restriction sites. We will now use the

Table 2 Comparison of different resequencing strategies.

	Reference genome needed?	Complexity reduction	Access to non-coding regions	Cost per sample	Advantage	Disadvantages	Examples
Whole genome resequencing	+++		++++	++++	<ul style="list-style-type: none"> - Complete representation of the genome - Analysis of structure possible 	Expensive	Date palm [217], Maize [218], Pepper [99], Rice [208], Sorghum [219], Soybean [220], Tomato [221]
Exome capture	+++	++		+++	<ul style="list-style-type: none"> - Sequence of protein coding regions & functional elements - Known target regions 	<ul style="list-style-type: none"> - Only exons - Prior probe design required 	Barley [112], Bread wheat [222], Cotton [223], Pine [224], Rapeseed [225], Soybean [226], Sugarcane [227]
RNAseq	++	++++		++	<ul style="list-style-type: none"> - Gene expression analysis possible - Only protein coding regions - Reference-free analysis methods exist 	<ul style="list-style-type: none"> - Only exons - Only protein coding regions - Coverage depends on transcript abundance 	African rice [117], Carrot [228] Common bean [229], Cotton [230], Pepper [99], Sunflower [140], Tomato [34]
Genotyping-by-sequencing	++	++++	+++	+	<ul style="list-style-type: none"> - High-throughput identification of whole-genome markers - Reference-free analysis methods exist 	<ul style="list-style-type: none"> - Sparse marker data - Handling of missing data required 	Cassava [138], Chickpea [231], Cotton [69], Cowpea [232], Oat [233], Soybean [234], Watermelon [235]

term ‘genotyping-by-sequencing (GBS)’ as it captures the essence of these methods: they do not produce resequencing data for a gene set, but apply high-throughput sequencing to obtain dense genome-wide marker data. As such, GBS is similar to genotyping with SNP chips. In contrast to SNP chips, however, no prior knowledge and expenses are required to develop genotyping assays and ascertainment bias [123] is reduced. Challenges in the analysis of GBS data include allelic dropout [124] and handling of missing data [125, 126]. In principle, GBS data can be analyzed without a reference genome [127, 128], but care needs to be taken in the interpretation of estimates of population genetic parameters [129].

GBS is a versatile and inexpensive method for natural and experimental populations of crops and their wild relatives. Linkage maps for wild relatives of Triticeae crops [130–132] and crop–wild cross in tomato [133] have been constructed in recent years. Moreover, GBS has been used to characterize crop–wild introgression lines in barley [134] and tomato [135]. In addition to the analysis of experimental populations, GBS has been widely adopted for population genetic studies. For example, it has been used to compare diversity between crops and their wild progenitors in chickpea [136], amaranth [137], and cassava [138]; to study gene flow between wild and domesticated sunflower taxa [139, 140]; and to elucidate the demographic history of wild *Phaseolus vulgaris*, the progenitor of common bean [141].

Moreover, GBS is a cost-effective method to screen large germplasm collections. For example, Romay et al. [142] used GBS to study genetic variation in 2815 maize inbred lines maintained at the US national seedbank. If used at the scale of entire collections, GBS holds the potential of developing molecular passport data for gene bank management, complementing traditional morphological markers and field evaluation. This can help resolve issues regarding (i) duplication within and between gene banks around the world [143], (ii) the genetic integrity of accessions after decades of ex situ maintenance [144, 145], and (iii) the development of truly representative core collections [146] to be subjected to whole-genome sequencing [147] and in-depth phenotyping. The power of this approach was exemplified by a recent study in maize. Navarro et al. [148] genotyped and phenotyped a panel of > 4000 F₁ hybrids between elite breeding material and diverse landraces sampled from the collection of the International Maize and Wheat Improvement Center (CIMMYT). Genome-wide association scans revealed a co-association of genes to both flowering time and altitude. Similar approaches are underway in other cereal crops such as wheat [149], barley [150], common bean [151], Solanaceous crops [152], and rice [153]. As more genotypic and phenotypic data for

large germ plasm collections accrue, strategies for the effective utilization of plant genetic resources into breeding without inflicting yield losses are needed [154].

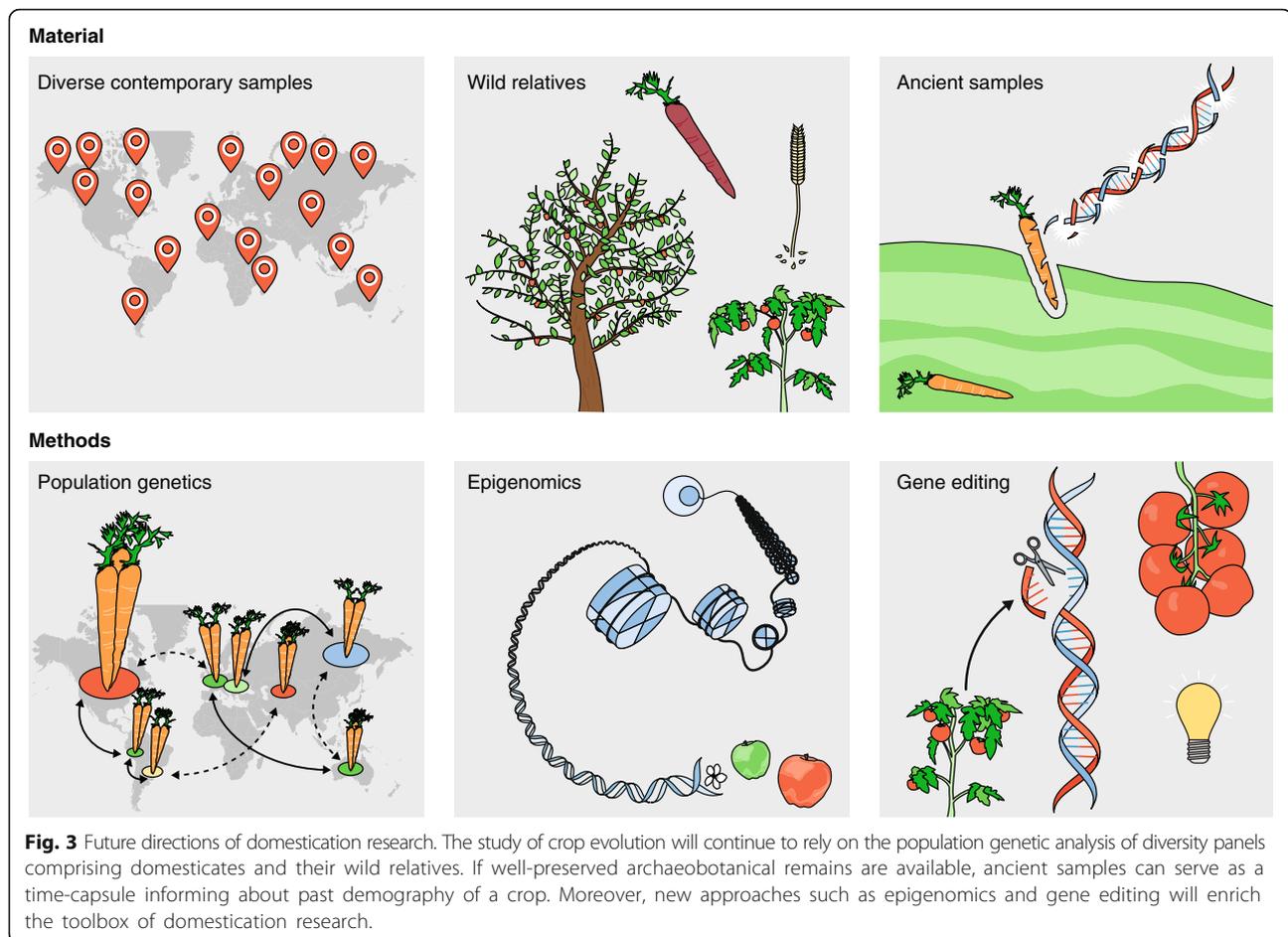
Novel approaches to domestication research

High-throughput sequencing has also been used to obtain measurements on aspects of the genome other than nucleotide sequence variation. Here, we describe future directions in domestication research that will benefit from these technological innovations, such as epigenomics, archaeogenetics, genome editing, domestication of novel crops, and new computational methods for analysis of population genetic data (Fig. 3).

Epigenomics refers to the study of the entirety of heritable changes other than changes in DNA sequence, such as DNA methylation or histone modifications. The interaction between euchromatin and heterochromatin likely plays a role in silencing of transposable elements and influences gene expression [155]. Compared to genetic diversity, little is known about epigenetic diversity in crops, which may prove an untapped reservoir of useful variation for crop improvement [156]. Some important initial results have been published recently. Daccord et al. [157] generated an improved genome assembly of domesticated apple and created a genome-wide map of DNA methylation. Their results hint at a potential role of epigenetic marks in the expression of agronomic traits in perennial fruit trees. Epigenetic regulation is of particular importance to study the relationship of subgenomes in polyploid crops such as wheat [158], cotton [159], and rapeseed [160].

Archaeogenetics—the extraction, sequencing, and analysis of ancient DNA fragments—has transformed our concepts of the history of human and animal species [161–163]. Until recently, plant genetics has not had access to this window into the past, mainly owing to a paucity of archaeobotanical remains containing sufficient amounts of well-preserved DNA [164, 165]. Recent studies in maize and barley retrieved DNA sequences from samples preserved under arid conditions and analyzed them together with sequences from extant individuals. da Fonseca et al. [166] thus reconstructed the past demography of maize in the southwest United States. Mascher et al. [167] reported a close genetic affinity of 6000-year-old barley grains from the Judean desert to present-day landraces from the Southern Levant. Swarts et al. [168] employed genomic prediction models trained on modern data to understand the temporal dynamics of adaptation to temperate climates as maize cultivation spread northwards.

Genome editing with CRISPR-Cas9 technology [169] has enriched the plant geneticist’s toolkit [170]. The rapid induction of targeted mutations will be instrumental in validating putative domestication genes in the wild



individuals and creating novel useful variation in the domesticate. For example, Soyk et al. [171] isolated two regulators of inflorescence architecture in tomato by means of map-based cloning and validation through classic mutants and CRISPR-Cas9 knock-out. Naturally occurring mutations in both genes had been independently selected either to increase fruit size or to facilitate mechanical harvesting. However, due to negative epistatic interaction between both genes, combinations of alleles that are beneficial on their own resulted in loss of fertility and excessive branching. Soyk et al. [171] then evaluated allelic series of both loci for epistatic interactions and found new beneficial allelic combinations that overcame negative epistasis. Braatz et al. [172] used CRISPR-Cas9 in tetraploid oilseed rape (*Brassica napus*) to knockout simultaneously both homeologs of *ALCA-TRAZ*, a known regulator of seed shattering in *Arabidopsis* [173]. Siliques of double mutants were partially resistant to shattering, an important trait to avoid seed loss during mechanical harvesting.

The domestication of new plant species has been proposed as an important future contribution to sustainable agriculture. For instance, the development of perennial

grain crops has received considerable attention [174, 175]. Progress has been made in bringing intermediate wheatgrass (*Thinopyrum intermedium*; Fig. 1), a perennial relative of wheat, into cultivation with the establishment of dense linkage maps [131] and the implementation of genomic selection [176]. The domestication of bioenergy crops has been put forward to meet the growing demands for biofuel. Proposed targets include the aquatic fern *Azolla* [177], *Miscanthus* species [178], and the duckweeds [179].

As large population genomic datasets accrue in more species, analysis methods need to keep pace with the growing amount of input data. Efficient data structures have been devised to structure and handle large marker matrices [180, 181]. Imputation strategies to infer missing genotypes in low-coverage sequence data have been adopted in human genetics [182, 183]. Moreover, imputation methods, which take into account inbreeding or are geared towards experimental populations, have been developed specifically for plant genetics [184]. We expect genotype imputation to be widely used in plant genetic studies as comprehensive haplotype reference panels become available [185]. As an alternative or

complement to the imputation of discrete allelic states, statistical uncertainties in genotype calling from shallow sequencing data can be recorded and considered during population genetic analyses [186, 187].

New algorithms have been developed to speed up traditional analysis methods such as principal component analysis [188, 189] and statistical inference of population structure [190]. These include flashpca [191], FastPCA [192], fastSTRUCTURE [193], ADMIXTURE [194], and sNMF [195]. Methods for understanding past demographic processes such as bottlenecks and migration events include genome-wide comparisons of allele frequencies from dense genomic marker datasets [196, 197], fitting coalescent models to whole-genome sequence data [100, 101], and computational environments for demographic simulations [198, 199].

In summary, progress in sequencing technology and analysis methods will make it possible to study the genetics and genomics of domestication in a wider range of crop species. In the coming years, chromosome-scale reference sequence assemblies and resequencing studies of large diversity panels will contribute to understanding the past and present diversity of domesticated plants and their wild relatives.

Abbreviations

BAC: Bacterial artificial chromosome; GBS: Genotyping-by-sequencing; Hi-C: Chromosome conformation capture (quantifies interactions between all possible pairs of fragments simultaneously); RNA-seq: RNA sequencing; SNP, Single-nucleotide polymorphism

Acknowledgments

None.

Funding

This work was supported by a grants from the Leibniz Association to N.S. (Pakt für Forschung und Innovation: SAW-2015-IPK-1) and from the German Federal Ministry of Education and Research (BMBF; grant 031B0190A: ú SHAPE) to N.S. and M.M.

Authors' contributions

MS, NS, and MM wrote the article. MS designed the figures. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published online: 21 September 2018

References

- Food and Agriculture Organisation of the United Nations. FAOSTAT statistics database. 2016. <http://www.fao.org/faostat/en/#home>. Accessed 9 Aug 2018.
- Doebley JF, Gaut BS, Smith BD. The molecular genetics of crop domestication. *Cell*. 2006;127:1309–21.
- Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia JM, Cartwright RA, et al. Comparative population genomics of maize domestication and improvement. *Nat Genet*. 2012;44:808–11.
- Cao K, Zheng Z, Wang L, Liu X, Zhu G, Fang W, et al. Comparative population genomics reveals the domestication history of the peach, *Prunus persica*, and human influences on perennial fruit crops. *Genome Biol*. 2014; 15:415.
- Purugganan MD, Fuller DQ. Archaeological data reveal slow rates of evolution during plant domestication. *Evolution*. 2011;65:171–83.
- Fuller DQ, Denham T, Arroyo-Kalin M, Lucas L, Stevens CJ, Qin L, et al. Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proc Natl Acad Sci U S A*. 2014;111:6147–52.
- Abbo S, Lev-Yadun S, Gopher A. Agricultural origins: centers and noncenters – a Near Eastern reappraisal. *Crit Rev Plant Sci*. 2010;29:317–28.
- Innan H, Kim Y. Pattern of polymorphism after strong artificial selection in a domestication event. *Proc Natl Acad Sci U S A*. 2004;101:10667–72.
- Hillman GC, Davies MS. Domestication rates in wild-type wheats and barley under primitive cultivation. *Biol J Linn Soc*. 1990;39:39–78.
- Eklund R, Galindo J. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity (Edinb)*. 2011;107:1–15.
- Cannon SB, May GD, Jackson SA. Three sequenced legume genomes and many crop species: rich opportunities for translational genomics. *Plant Physiol*. 2009;151:970–7.
- Lenser T, Theißen G. Molecular mechanisms involved in convergent crop domestication. *Trends Plant Sci*. 2013;18:704–14.
- Meyer RS, Purugganan MD. Evolution of crop species: genetics of domestication and diversification. *Nat Rev Genet*. 2013;14:840–52.
- Huang X, Han B. Natural variations and genome-wide association studies in crop plants. *Annu Rev Plant Biol*. 2014;65:531–51.
- Olsen KM, Wendel JF. A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu Rev Plant Biol*. 2013;64:47–70.
- Shi J, Lai J. Patterns of genomic changes with crop domestication and breeding. *Curr Opin Plant Biol*. 2015;24:47–53.
- Bevan MW, Uauy C, Wulff BB, Zhou J, Krasileva K, Clark MD. Genomic innovation for crop improvement. *Nature*. 2017;543:346–54.
- Turcotte MM, Araki H, Karp DS, Poveda K, Whitehead SR. The eco-evolutionary impacts of domestication and agricultural practices on wild species. *Philos Trans R Soc Lond Ser B Biol Sci*. 2017;372:20160033.
- Gutaker RM, Burbano HA. Reinforcing plant evolutionary genomics using ancient DNA. *Curr Opin Plant Biol*. 2017;36:38–45.
- Palmer SA, Smith O, Allaby RG. The blossoming of plant archaeogenetics. *Ann Anat*. 2012;194:146–56.
- Abbo S, Gopher A. Near eastern plant domestication: A history of thought. *Trends Plant Sci*. 2017;22:491–511.
- Kantar MB, Nashoba AR, Anderson JE, Blackman BK, Rieseberg LH. The genetics and genomics of plant domestication. *Bioscience*. 2017;67:971–82.
- Pankin A, von Korff M. Co-evolution of methods and thoughts in cereal domestication studies: a tale of barley (*Hordeum vulgare*). *Curr Opin Plant Biol*. 2017;36:15–21.
- Harlan JR, Zohary D. Distribution of wild wheats and barley. *Science*. 1966; 153:1074–80.
- 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. *Agroecol Sust Food*. 2014;38:369–77.
- Harlan JR, de Wet JM. Toward a rational classification of cultivated plants. *Taxon*. 1971;20:509–17.
- Briggs WH, McMullen MD, Gaut BS, Doebley J. Linkage mapping of domestication loci in a large maize–teosinte backcross resource. *Genetics*. 2007;177:1915–28.
- Peng J, Ronin Y, Fahima T, Röder MS, Li Y, Nevo E, et al. Domestication quantitative trait loci in *Triticum dicoccoides*, the progenitor of wheat. *Proc Natl Acad Sci U S A*. 2003;100:2489–94.
- Jin J, Huang W, Gao JP, Yang J, Shi M, Zhu M-Z, et al. Genetic control of rice plant architecture under domestication. *Nat Genet*. 2008;40:1365–9.
- Pourkheirandish M, Hensel G, Kilian B, Senthil N, Chen G, Sameri M, et al. Evolution of the grain dispersal system in barley. *Cell*. 2015;162:527–39.
- Clark RM, Wagler TN, Quijada P, Doebley J. A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat Genet*. 2006;38:594–7.
- Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJ, et al. The genome of *Chenopodium quinoa*. *Nature*. 2017;542:307–12.
- Huang X, Kurata N, Wang Z-X, Wang A, Zhao Q, Zhao Y, et al. A map of rice genome variation reveals the origin of cultivated rice. *Nature*. 2012; 490:497–501.

34. Koenig D, Jiménez-Gómez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, et al. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc Natl Acad Sci U S A*. 2013;110:E2655–62.
35. Shang Y, Ma Y, Zhou Y, Zhang H, Duan L, Chen H, et al. Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science*. 2014;346:1084–8.
36. Feuillet C, Leach JE, Rogers J, Schnable PS, Eversole K. Crop genome sequencing: lessons and rationales. *Trends Plant Sci*. 2011;16:77–88.
37. Lander ES, Linton LM, Birren B, Nussbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860–921.
38. International Rice Genome Sequencing Project. The map-based sequence of the rice genome. *Nature*. 2005;436:793–800.
39. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: complexity, diversity, and dynamics. *Science*. 2009;326:1112–5.
40. Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, et al. Structural and functional partitioning of bread wheat chromosome 3B. *Science*. 2014;345:1249721.
41. Schulte D, Close TJ, Graner A, Langridge P, Matsumoto T, Muehlbauer G, et al. The international barley sequencing consortium—at the threshold of efficient access to the barley genome. *Plant Physiol*. 2009;149:142–7.
42. Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, et al. A chromosome conformation capture ordered sequence of the barley genome. *Nature*. 2017;544:427–33.
43. Beier S, Himmelbach A, Schmutzer T, Felder M, Taudien S, Mayer KF, et al. Multiplex sequencing of bacterial artificial chromosomes for assembling complex plant genomes. *Plant Biotechnol J*. 2016;14:1511–22.
44. Li R, Fan W, Tian G, Zhu H, He L, Cai J, et al. The sequence and de novo assembly of the giant panda genome. *Nature*. 2010;463:311–7.
45. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res*. 2010;20:265–72.
46. Clouse JW, Adhikary D, Page JT, Ramaraj T, Deyholos MK, Udall JA, et al. The amaranth genome: genome, transcriptome, and physical map assembly. *Plant Genome*. 2016;9. <https://doi.org/10.3835/plantgenome2015.07.0062>.
47. Sun H, Wu S, Zhang G, Jiao C, Guo S, Ren Y, et al. Karyotype stability and unbiased fractionation in the paleo-allotetraploid cucurbita genomes. *Mol Plant*. 2017;10:1293–306.
48. VanBuren R, Bryant D, Bushakra JM, Vining KJ, Edger PP, Rowley ER, et al. The genome of black raspberry (*Rubus occidentalis*). *Plant J*. 2016;87:535–47.
49. Iorizzo M, Ellison S, Senalik D, Zeng P, Satapoomin P, Huang J, et al. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat Genet*. 2016;48:657–66.
50. Jacquemin J, Bhatia D, Singh K, Wing RA. The International *Oryza* Map Alignment Project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. *Curr Opin Plant Biol*. 2013;16:147–56.
51. Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, et al. High-throughput genotyping by whole-genome resequencing. *Genome Res*. 2009;19:1068–76.
52. Mascher M, Muehlbauer GJ, Rokhsar DS, Chapman J, Schmutz J, Barry K, et al. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J*. 2013;76:718–27.
53. van Oeveren J, de Rooter M, Jesse T, van der Poel H, Tang J, Yalcin F, et al. Sequence-based physical mapping of complex genomes by whole genome profiling. *Genome Res*. 2011;21:618–25.
54. Mayer KF, Taudien S, Martis M, Simkova H, Suchankova P, Gundlach H, et al. Gene content and virtual gene order of barley chromosome 1H. *Plant Physiol*. 2009;151:496–505.
55. International Wheat Genome Sequencing Consortium. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*. 2014;345:1251788.
56. Chapman JA, Mascher M, Buluc AN, Barry K, Georganas E, Session A, et al. A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biol*. 2015;16:26.
57. International Barley Genome Sequencing Consortium. A physical, genetic and functional sequence assembly of the barley genome. *Nature*. 2012;491:711–6.
58. Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, et al. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science*. 2017;357:93–7.
59. Lu F-H, McKenzie N, Kettleborough G, Heavens D, Clark MD, Bevan MW. Independent assessment and improvement of wheat genome assemblies using Fosill jumping libraries. *Gigascience*. 2018;7. <https://doi.org/10.1093/gigascience/giy053>.
60. Aird D, Ross MG, Chen W-S, Danielsson M, Fennell T, Russ C, et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol*. 2011;12:R18.
61. Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, Wright J, et al. An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Res*. 2017;27:885–96.
62. Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, et al. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat Biotechnol*. 2012;30:693–700.
63. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, et al. Improved maize reference genome with single-molecule technologies. *Nature*. 2017;546:524–7.
64. Schmidt MH-W, Vogel A, Denton AK, Istace B, Wormit A, van de Geest H, et al. De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. *Plant Cell*. 2017;29:2336–48.
65. Zimin AV, Puiu D, Hall R, Kingan S, Clavijo BJ, Salzberg SL. The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. *Gigascience*. 2017;6:1–7.
66. Zimin AV, Puiu D, Luo M-C, Zhu T, Koren S, Marçais G, et al. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res*. 2017;27:787–92.
67. Fierst JL. Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools. *Front Genet*. 2015;6:220.
68. Mascher M, Stein N. Genetic anchoring of whole-genome shotgun assemblies. *Front Genet*. 2014;5:208.
69. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotechnol*. 2015;33:531–7.
70. Aliyeva-Schnorr L, Beier S, Karafiátová M, Schmutzer T, Scholz U, Doležel J, et al. Cytogenetic mapping with centromeric bacterial artificial chromosomes contigs shows that this recombination-poor region comprises more than half of barley chromosome 3H. *Plant J*. 2015;84:385–94.
71. Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, et al. High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. *Genomics*. 2003;82:378–89.
72. Hukriede NA, Joly L, Tsang M, Miles J, Tellis P, Epstein JA, et al. Radiation hybrid mapping of the zebrafish genome. *Proc Natl Acad Sci U S A*. 1999;96:9745–50.
73. Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, Austin MD, et al. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat Biotechnol*. 2012;30:771–6.
74. Hatakeyama M, Aluri S, Balachadran MT, Sivarajan SR, Patrignani A, Grüter S, et al. Multiple hybrid de novo genome assembly of finger millet, an orphan allotetraploid crop. *DNA Res*. 2017. <https://doi.org/10.1093/dnares/dsx036>.
75. Stankova H, Hastie AR, Chan S, Vrana J, Tulpova Z, Kubalaková M, et al. BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes. *Plant Biotechnol J*. 2016;14:1523–31.
76. Zheng GX, Lau BT, Schnall-Levin M, Jarosz M, Bell JM, Hindson CM, et al. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat Biotechnol*. 2016;34:303–11.
77. Yeo S, Coombe L, Chu J, Warren RL, Biról I. ARCS: assembly roundup by chromium scaffolding. *bioRxiv*. 2017:100750. <https://doi.org/10.1101/100750>.
78. Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. Direct determination of diploid genome sequences. *Genome Res*. 2017;27:757–67.
79. Hulse-Kemp AM, Maheshwari S, Stoffel K, Hill TA, Jaffe D, Williams SR, et al. Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. *Hortic Res*. 2018;5:4.
80. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozcy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326:289–93.

81. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat Biotechnol.* 2013;31:1119–25.
82. Kaplan N, Dekker J. High-throughput genome scaffolding from in vivo DNA interaction frequency. *Nat Biotechnol.* 2013;31:1143–7.
83. Grob S, Schmid MW, Grossniklaus U. Hi-C analysis in *Arabidopsis* identifies the KNOT, a structure with similarities to the flamenco locus of *Drosophila*. *Mol Cell.* 2014;55:678–93.
84. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science.* 2017;356:92–5.
85. Ariyadasa R, Mascher M, Nussbaumer T, Schulte D, Frenkel Z, Poursarebani N, et al. A sequence-ready physical map of barley anchored genetically by two million single-nucleotide polymorphisms. *Plant Physiol.* 2014;164:412–23.
86. Lightfoot D, Jarvis DE, Ramaraj T, Lee R, Jellen E, Maughan P. Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. *BMC Biol.* 2017;15:74.
87. Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* 2016;26:342–50.
88. Jiao W-B, Accinelli GG, Hartwig B, Kiefer C, Baker D, Severing E, et al. Improving and correcting the contiguity of long-read genome assemblies of three plant species using optical mapping and chromosome conformation capture data. *Genome Res.* 2017;27:778–86.
89. Moll KM, Zhou P, Ramaraj T, Fajardo D, Devitt NP, Sadowsky MJ, et al. Strategies for optimizing BioNano and Dovetail explored through a second reference quality assembly for the legume model, *Medicago truncatula*. *BMC Genomics.* 2017;18:578.
90. Reyes-Chin-Wo S, Wang Z, Yang X, Kozik A, Arikrit S, Song C, et al. Genome assembly with *in vitro* proximity ligation data and whole-genome triplication in lettuce. *Nat Commun.* 2017;8:14953.
91. Thind AK, Wicker T, Šimková H, Fossati D, Moullet O, Brabant C, et al. Rapid cloning of genes in hexaploid wheat using cultivar-specific long-range chromosome assembly. *Nat Biotechnol.* 2017;35:793–6.
92. Ford-Lloyd BV, Schmidt M, Armstrong SJ, Barazani O, Engels J, Hadas R, et al. Crop wild relatives—undervalued, underutilized and under threat? *Bioscience.* 2011;61:559–65.
93. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* 2016;17:333–51.
94. Chia J-M, Song C, Bradbury PJ, Costich D, De Leon N, Doebley J, et al. Maize HapMap2 identifies extant variation from a genome in flux. *Nat Genet.* 2012;44:803–7.
95. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet.* 2010;42:961–7.
96. Huang X, Zhao Y, Li C, Wang A, Zhao Q, Li W, et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat Genet.* 2012;44:32–9.
97. Fang C, Ma Y, Wu S, Liu Z, Wang Z, Yang R, et al. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* 2017;18:161.
98. Hardigan MA, Laimbeer FPE, Newton L, Crisovan E, Hamilton JP, Vaillancourt B, et al. Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc Natl Acad Sci U S A.* 2017;114:E9999–E10008.
99. Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, et al. Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc Natl Acad Sci U S A.* 2014;111:5135–40.
100. Yano K, Yamamoto E, Aya K, Takeuchi H, Lo P-C, Hu L, et al. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat Genet.* 2016;48:927–34.
101. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature.* 2011;475:493–6.
102. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet.* 2014;46:919–25.
103. Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Mol Ecol.* 2016;25:1058–72.
104. 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.
105. Zhou Y, Massonnet M, Sanjak JS, Cantu D, Gaut BS. Evolutionary genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *Proc Natl Acad Sci U S A.* 2017;114:11715–20.
106. Meyer RS, Choi JY, Sanches M, Plessis A, Flowers JM, Amas J, et al. Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nat Genet.* 2016;48:1083–8.
107. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature.* 2009;461:272–6.
108. Jupe F, Witek K, Verweij W, Śliwka J, Pritchard L, Etherington GJ, et al. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* 2013;76:530–44.
109. Jordan KW, Wang S, Lun Y, Gardiner LJ, MacLachlan R, Hucl P, et al. A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biol.* 2015;16:48.
110. Krasileva KV, Vasquez-Gross HA, Howell T, Bailey P, Paraiso F, Clissold L, et al. Uncovering hidden variation in polyploid wheat. *Proc Natl Acad Sci U S A.* 2017;114:E913–21.
111. Mascher M, Richmond TA, Gerhardt DJ, Himmelbach A, Clissold L, Sampath D, et al. Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J.* 2013;76:494–505.
112. Russell J, Mascher M, Dawson IK, Kyriakidis S, Calixto C, Freund F, et al. Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. *Nat Genet.* 2016;48:1024–30.
113. Steuernagel B, Periyannan SK, Hernández-Pinzón I, Witek K, Rouse MN, Yu G, et al. Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat Biotechnol.* 2016;34:652–5.
114. Arora S, Steuernagel B, Chandramohan S, Long Y, Matry O, Johnson R, et al. Resistance gene discovery and cloning by sequence capture and association genetics. *bioRxiv.* 2018:248146. <https://doi.org/10.1101/248146>.
115. Witek K, Jupe F, Witek AI, Baker D, Clark MD, Jones JD. Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT sequencing. *Nat Biotechnol.* 2016;34:656–60.
116. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008;5:621–8.
117. Nabholz B, Sarah G, Sabot F, Ruiz M, Adam H, Nidelet S, et al. Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (*Oryza glaberrima*). *Mol Ecol.* 2014;23:2210–27.
118. Lemmon ZH, Bukowski R, Sun Q, Doebley JF. The role of cis regulatory evolution in maize domestication. *PLoS Genet.* 2014;10:e1004745.
119. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet.* 2011;12:499–510.
120. Davey JW, Blaxter ML. RADSeq: next-generation population genetics. *Brief Funct Genomics.* 2010;9:416–23.
121. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One.* 2011;6:e19379.
122. Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, et al. SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS One.* 2013;8:e58700.
123. Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res.* 2005;15:1496–502.
124. Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhué C, Pudlo P, et al. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Mol Ecol.* 2013;22:3165–78.
125. Arnold B, Corbett-Detig RB, Hartl D, Bomblies K. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Mol Ecol.* 2013;22:3179–90.
126. Manching H, Sengupta S, Hopper KR, Polson SW, Ji Y, Wisser RJ. Phased genotyping-by-sequencing enhances analysis of genetic diversity and reveals divergent copy number variants in maize. *G3 (Bethesda).* 2017;7:2161–70.

127. Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, et al. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet.* 2013;9:e1003215.
128. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. *Mol Ecol.* 2013;22:3124–40.
129. Shafer ABA, Peart CR, Tusso S, Maayan I, Brelford A, Wheat CW, Wolf JBW. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods Ecol Evol.* 2017;8:907–17.
130. Wendler N, Mascher M, Himmelbach A, Bini F, Kumlehn J, Stein N. A high-density, sequence-enriched genetic map of *Hordeum bulbosum* and its collinearity to *H. vulgare*. *Plant Genome.* 2017;10. <https://doi.org/10.3835/plantgenome2017.06.0049>.
131. Kantarski T, Larson S, Zhang X, DeHaan L, Borevitz J, Anderson J, et al. Development of the first consensus genetic map of intermediate wheatgrass (*Thinopyrum intermedium*) using genotyping-by-sequencing. *Theor Appl Genet.* 2017;130:137–50.
132. Zhang Y, Zhang J, Huang L, Gao A, Zhang J, Yang X, et al. A high-density genetic map for P genome of *Agropyron Gaertn.* based on specific-locus amplified fragment sequencing (SLAF-seq). *Planta.* 2015;242:1335–47.
133. Bao Z, Meng F, Strickler SR, Dunham DM, Munkvold KR, Martin GB. Identification of a candidate gene in *Solanum habrochaites* for resistance to a race 1 strain of *Pseudomonas syringae pv. tomato*. *Plant Genome.* 2015;8. <https://doi.org/10.3835/plantgenome2015.02.0006>.
134. Wendler N, Mascher M, Himmelbach A, Johnston P, Pickering R, Stein N. Bulbosum to go: a toolbox to utilize *Hordeum vulgare/bulosum* introgressions for breeding and beyond. *Mol Plant.* 2015;8:1507–19.
135. Celik I, Gurbuz N, Uncu AT, Fray A, Doganlar S. Genome-wide SNP discovery and QTL mapping for fruit quality traits in inbred backcross lines (IBLs) of *Solanum pimpinellifolium* using genotyping by sequencing. *BMC Genomics.* 2017;18:1. <https://doi.org/10.1186/s12864-016-3406-7>.
136. Bajaj D, Das S, Badoni S, Kumar V, Singh M, Bansal KC, et al. Genome-wide high-throughput SNP discovery and genotyping for understanding natural (functional) allelic diversity and domestication patterns in wild chickpea. *Sci Rep.* 2015;5:12468.
137. Stetter MG, Müller T, Schmid KJ. Genomic and phenotypic evidence for an incomplete domestication of South American grain amaranth (*Amaranthus caudatus*). *Mol Ecol.* 2017;26:871–86.
138. Bredeson JV, Lyons JB, Prochnik SE, Wu GA, Ha CM, Edsinger-Gonzales E, et al. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat Biotechnol.* 2016;34:562–70.
139. Mondon A, Owens GL, Poverene M, Cantamutto M, Rieseberg LH. Gene flow in Argentinian sunflowers as revealed by genotyping by sequencing data. *Evol Appl.* 2017;11:193–204.
140. Baute GJ, Owens GL, Bock DG, Rieseberg LH. Genome-wide genotyping-by-sequencing data provide a high-resolution view of wild *Helianthus* diversity, genetic structure, and interspecies gene flow. *Am J Bot.* 2016;103:2170–7.
141. Ariani A, Bery Mier y Teran JC, Gepts P. Spatial and temporal scales of range expansion in wild *Phaseolus vulgaris*. *Mol Biol Evol.* 2018;35:119–31.
142. Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, et al. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* 2013;14:R55.
143. van Hintum TJ, Visser DL. Duplication within and between germplasm collections. *Genet Resour Crop Evol.* 1995;42:135–45.
144. Jakob SS, Rodder D, Engler JO, Shaaf S, Ozkan H, Blattner FR, et al. Evolutionary history of wild barley (*Hordeum vulgare* subsp. *spontaneum*) analyzed using multilocus sequence data and paleodistribution modeling. *Genome Biol Evol.* 2014;6:685–702.
145. Parzies H, Spoor W, Ennos R. Genetic diversity of barley landrace accessions (*Hordeum vulgare* ssp. *vulgare*) conserved for different lengths of time in ex situ gene banks. *Heredity.* 2000;84:476–86.
146. Odong T, Jansen J, Van Eeuwijk F, van Hintum TJ. Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. *Theor Appl Genet.* 2013;126:289–305.
147. 3,000 Rice Genomes project. The 3,000 Rice Genomes project. *GigaScience.* 2014;3:7.
148. Navarro JAR, Willcox M, Burgueño J, Romay C, Swarts K, Trachsel S, et al. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat Genet.* 2017;49:476–80.
149. Seeds of Discovery. <http://www.cimmyt.org/seeds-of-discovery/>. Accessed 10 Aug 2018.
150. IPK Gatersleben BRIDGE Web Portal. <http://bridge.ipk-gatersleben.de>. Accessed 10 Aug 2018.
151. Bean Adapt <https://www.facebook.com/BeanAdapt/>. Accessed 10 Aug 2018.
152. G2P-SOL Linking genetic resources, genomes and phenotypes of Solanaceous crops. <http://www.g2p-sol.eu/>. Accessed 10 Aug 2018.
153. Li JY, Wang J, Zeigler RS. The 3,000 rice genomes project: new opportunities and challenges for future rice research. *Gigascience.* 2014;3:8.
154. Longin CFH, Reif JC. Redesigning the exploitation of wheat genetic resources. *Trends Plant Sci.* 2014;19:631–6.
155. Springer NM, Lisch D, Li Q. Creating order from chaos: epigenome dynamics in plants with complex genomes. *Plant Cell.* 2016;28:314–25.
156. Springer NM, Schmitz RJ. Exploiting induced and natural epigenetic variation for crop improvement. *Nat Rev Genet.* 2017;18:563–75.
157. Daccord N, Celton J-M, Linsmith G, Becker C, Choise N, Schijlen E, et al. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nat Genet.* 2017;49:1099–106.
158. Guo X, Han F. Asymmetric epigenetic modification and elimination of rDNA sequences by polyploidization in wheat. *Plant Cell.* 2014;26:4311–27.
159. Song Q, Zhang T, Stelly DM, Chen ZJ. Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol.* 2017;18:99.
160. Chen ZJ, Pikaard CS. Epigenetic silencing of RNA polymerase II transcription: a role for DNA methylation and histone modification in nucleolar dominance. *Genes Dev.* 1997;11:2124–36.
161. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. *Science.* 2010;328:710–22.
162. Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, Tresset A, et al. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proc Natl Acad Sci U S A.* 2007;104:15276–81.
163. Larson G, Karlsson EK, Perri A, Webster MT, Ho SY, Peters J, et al. Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc Natl Acad Sci U S A.* 2012;109:8878–83.
164. Hofreiter U, Serre D, Poinar HN, Kuch M, Pääbo S. Ancient DNA. *Nat Rev Genet.* 2001;2:353–9.
165. Schlumbaum A, Tensen M, Jaenicke-Després V. Ancient plant DNA in archaeobotany. *Veg Hist Archaeobotany.* 2008;17:233–44.
166. da Fonseca RR, Smith BD, Wales N, Cappellini E, Skoglund P, Fumagalli M, et al. The origin and evolution of maize in the Southwestern United States. *Nat Plants.* 2015;1:14003.
167. Mascher M, Schuenemann VJ, Davidovich U, Marom N, Himmelbach A, Hubner S, et al. Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nat Genet.* 2016;48:1089–93.
168. Swarts K, Gutaker RM, Benz B, Blake M, Bukowski R, Holland J, et al. Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. *Science.* 2017;357:512–5.
169. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science.* 2014;346:1258096.
170. Jung C, Capistrano-Gossmann G, Braatz J, Sashidhar N, Melzer S. Recent developments in genome editing and applications in plant breeding. *Plant Breed.* 2018;137:1–9.
171. Soyk S, Lemmon ZH, Oved M, Fisher J, Liberatore KL, Park SJ, et al. Bypassing negative epistasis on yield in tomato imposed by a domestication gene. *Cell.* 2017;169:1142–55.
172. Braatz J, Harloff H-J, Mascher M, Stein N, Himmelbach A, Jung C. CRISPR-Cas9 targeted mutagenesis leads to simultaneous modification of different homoeologous gene copies in polyploid oilseed rape (*Brassica napus*). *Plant Physiol.* 2017;174:935–42.
173. Rajani S, Sundaresan V. The *Arabidopsis* myc/bHLH gene *ALCATRAZ* enables cell separation in fruit dehiscence. *Curr Biol.* 2001;11:1914–22.
174. Cox TS, Glover JD, Van Tassel DL, Cox CM, DeHaan LR. Prospects for developing perennial grain crops. *Bioscience.* 2006;56:649–59.
175. Hayes R, Newell M, DeHaan L, Murphy K, Crane S, Norton M, et al. Perennial cereal crops: an initial evaluation of wheat derivatives. *Field Crop Res.* 2012; 133:68–89.
176. Zhang X, Sallam A, Gao L, Kantarski T, Poland J, DeHaan LR, et al. Establishment and optimization of genomic selection to accelerate the domestication and improvement of intermediate wheatgrass. *Plant Genome.* 2016;9. <https://doi.org/10.3835/plantgenome2015.07.0059>.
177. Brouwer P, Bräutigam A, Külahoglu C, Tazelaar AO, Kurz S, Nierop KG, et al. *Azolla* domestication towards a biobased economy? *New Phytol.* 2014;202: 1069–82.

178. Robson P, Jensen E, Hawkins S, White SR, Kenobi K, Clifton-Brown J, et al. Accelerating the domestication of a bioenergy crop: identifying and modelling morphological targets for sustainable yield increase in *Miscanthus*. *J Exp Bot*. 2013;64:4143–55.
179. Cheng JJ, Stomp AM. Growing duckweed to recover nutrients from wastewaters and for production of fuel ethanol and animal feed. *Clean Soil Air Water*. 2009;37:17–26.
180. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*. 2012;28:3326–8.
181. Zheng X, Gogarten SM, Lawrence M, Stulp A, Conomos MP, Weir BS, et al. SeqArray—a storage-efficient high-performance data format for WGS variant calls. *Bioinformatics*. 2017;33:2251–7.
182. Browning BL, Browning SR. Genotype imputation with millions of reference samples. *Am J Hum Genet*. 2016;98:116–26.
183. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods*. 2013;10:5–6.
184. Swarts K, Li H, Romero Navarro JA, An D, Romay MC, Hearne S, et al. Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. *Plant Genome*. 2014;7. <https://doi.org/10.3835/plantgenome2014.05.0023>.
185. Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*. 2018; 557:43–9.
186. Fumagalli M, Vieira FG, Linderoth T, Nielsen R. ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics*. 2014;30:1486–7.
187. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*. 2014;15:356.
188. Cavalli-Sforza LL, Menozzi P, Piazza A. The history and geography of human genes. 2nd ed. Princeton: Princeton University Press; 1994.
189. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2:e190.
190. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–59.
191. Abraham G, Inouye M. Fast principal component analysis of large-scale genome-wide data. *PLoS One*. 2014;9:e93766.
192. Galinsky KJ, Bhatia G, Loh P-R, Georgiev S, Mukherjee S, Patterson NJ, et al. Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia. *Am J Hum Genet*. 2016;98:456–72.
193. Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics*. 2014;197:573–89.
194. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19:1655–64.
195. Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. Fast and efficient estimation of individual ancestry coefficients. *Genetics*. 2014;196:973–83.
196. Soraggi S, Wiuf C, Albrechtsen A. Powerful inference with the D-statistic on low-coverage whole genome data. *G3 (Bethesda)*. 2018;8:551–66.
197. Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting FST: the impact of rare variants. *Genome Res*. 2013;23:1514–21.
198. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet*. 2009;5:e1000695.
199. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. Robust demographic inference from genomic and SNP data. *PLoS Genet*. 2013;9: e1003905.
200. Wang M, Yu Y, Haberger G, Marri PR, Fan C, Goicoechea JL, et al. The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat Genet*. 2014;46:982–8.
201. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*. 2013;6:4.
202. Dohm JC, Minoche AE, Holtgräwe D, Capella-Gutiérrez S, Zakrzewski F, Tafer H, et al. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature*. 2014;505:546–9.
203. Wang W, Feng B, Xiao J, Xia Z, Zhou X, Li P, et al. Cassava genome from a wild ancestor to cultivated varieties. *Nat Commun*. 2014;5:1110.
204. Varshtey RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, et al. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat Biotechnol*. 2013;31:240–6.
205. Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, et al. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat Genet*. 2014;46:567–72.
206. Muñoz-Amatrián M, Mirebrahim H, Xu P, Wanamaker SI, Luo M, Alhakami H, et al. Genome resources for climate-resilient cowpea, an essential crop for food security. *Plant J*. 2016;89:1042–54.
207. Jayakodi M, Choi B-S, Lee S-C, Kim N-H, Park JY, Jang W, et al. Ginseng Genome Database: an open-access platform for genomics of Panax ginseng. *BMC Plant Biol*. 2018;18:62.
208. Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, et al. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat Biotechnol*. 2011;30:105–11.
209. Avena Genome. The Oat Genome project. <http://avenagenome.org>. Accessed 9 Aug 2018.
210. The Potato Genome Sequencing Consortium. Genome sequence and analysis of the tuber crop potato. *Nature*. 2011;475:189–95.
211. Yang J, Liu D, Wang X, Ji C, Cheng F, Liu B, et al. The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nat Genet*. 2016;48:1225–32.
212. Bauer E, Schmutzer T, Barilar I, Mascher M, Gundlach H, Martis MM, et al. Towards a whole-genome sequence for rye (*Secale cereale* L.). *Plant J*. 2017; 89:853–69.
213. McCormic RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, et al. The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J*. 2018; 93:338–54.
214. Riaño-Pachón DM, Mattiello L. Draft genome sequencing of the sugarcane hybrid SP80-3280. *F1000Research*. 2017;6:861.
215. Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, et al. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature*. 2017;546:148–52.
216. The Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*. 2012;485:635–41.
217. Hazzouri KM, Flowers JM, Visser HJ, Khierallah HSM, Rosas U, Pham GM, et al. Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. *Nat Commun*. 2015;6:8824.
218. Xu J, Yuan Y, Xu Y, Zhang G, Guo X, Wu F, et al. Identification of candidate genes for drought tolerance by whole-genome resequencing in maize. *BMC Plant Biol*. 2014;14:83.
219. Mace ES, Tai S, Gilding EK, Li Y, Prentis PJ, Bian L, et al. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat Commun*. 2013;4:2320.
220. Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotechnol*. 2015;33:408–14.
221. Causse M, Desplat N, Pascual L, Le Paslier M-C, Sauvage C, Bauchet G, et al. Whole genome resequencing in tomato reveals variation associated with introgression and breeding events. *BMC Genomics*. 2013;14:791.
222. Allen AM, Barker GLA, Wilkinson P, Burrige A, Winfield M, Coghill J, et al. Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). *Plant Biotechnol J*. 2013;11:279–95.
223. Salmon A, Udall JA, Jeddalo JA, Wendel J. Targeted capture of homoeologous coding and noncoding sequence in polyploid cotton. *G3 (Bethesda)*. 2012;2:921–30.
224. Neves LG, Davis JM, Barbazuk WB, Matias K. Whole-exome targeted sequencing of the uncharacterized pine genome. *Plant J*. 2013;75:146–56.
225. Clarke WE, Parkin IA, Gajardo HA, Gerhardt DJ, Higgins E, Sidebottom C, et al. Genomic DNA enrichment using sequence capture microarrays: a novel approach to discover sequence nucleotide polymorphisms (SNP) in *Brassica napus* L. *PLoS One*. 2013;8:e81992.
226. Bolon Y-T, Haun WJ, Xu WW, Grant D, Stacey MG, Nelson RT, et al. Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. *Plant Physiol*. 2011;156:240–53.
227. Song J, Yang X, Resende MFR, Neves LG, Todd J, Zhang J, et al. Natural allelic variations in highly polyploidy saccharum complex. *Front Plant Sci*. 2016;7:804.
228. Rong J, Lammers Y, Strasburg JL, Schidlo NS, Ariyurek Y, de Jong TJ, et al. New insights into domestication of carrot from root transcriptome analyses. *BMC Genomics*. 2014;15:895.
229. Bellucci E, Bitocchi E, Ferrarini A, Benazzo A, Biagetti E, Klie S, et al. Decreased nucleotide and expression diversity and modified coexpression

- patterns characterize domestication in the common bean. *Plant Cell*. 2014; 26:1901–12.
230. Bao Y, Hu G, Fligel LE, Salmon A, Bezanilla M, et al. Parallel up-regulation of the profilin gene family following independent domestication of diploid and allopolyploid cotton (*Gossypium*). *Proc Natl Acad Sci U S A*. 2011;108: 21152–7.
231. Verma S, Gupta S, Bandhiwal N, Kumar T, Bharadwaj C, Bhatia S. High-density linkage map construction and mapping of seed trait QTLs in chickpea (*Cicer arietinum* L.) using Genotyping-by-Sequencing (GBS). *Sci Rep*. 2015;5:17512.
232. Xiong H, Shi A, Mou B, Qin J, Motes D, Lu W, et al. Genetic diversity and population structure of cowpea (*Vigna unguiculata* L. Walp). *PLoS One*. 2016;11:e0160941.
233. Huang Y-F, Poland JA, Wight CP, Jackson EW, Tinker NA. Using Genotyping-By-Sequencing (GBS) for genomic discovery in cultivated oat. *PLoS One*. 2014;9:e102448.
234. Jarquin D, Kocak K, Posadas L, Hyma K, Jedlicka J, Graef G, Lorenz A. Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics*. 2014;15:740.
235. Nimmakayala P, Levi A, Abburi L, Abburi VL, Tomason YR, Saminathan T, et al. Single nucleotide polymorphisms generated by genotyping by sequencing to characterize genome-wide diversity, linkage disequilibrium, and selective sweeps in cultivated watermelon. *BMC Genomics*. 2014;15:767.