Review

Cell PRESS

Teosinte as a model system for population and ecological genomics

Matthew B. Hufford[1], Paul Bilinski[1], Tanja Pyhäjärvi[1]*, and Jeffrey Ross-Ibarra[1,2]

[1] Plant Sciences Department, University of California, Davis, CA 95616, USA
[2] The Genome Center and Center for Population Biology, University of California, Davis, CA 95616, USA

As the cost of next-generation sequencing diminishes and genomic resources improve, crop wild relatives are well positioned to make major contributions to the field of ecological genomics via full-genome resequencing and reference-assisted *de novo* assembly of genomes of plants from natural populations. The wild relatives of maize, collectively known as teosinte, are a more varied and representative study system than many other model flowering plants. In this review of the population and ecological genomics of the teosintes we highlight recent advances in the study of maize domestication, introgressive hybridization, and local adaptation, and discuss future prospects for applying the genomic resources of maize to this intriguing group of species. The maize/teosinte study system is an excellent example of how crops and their wild relatives can bridge the model/non-model gap.

The genomic era and crop wild relatives

Over the past decade, reference genome sequences have been completed for numerous crops (e.g., Table 1 in [1]). Although the purpose of such sequencing efforts has largely been crop improvement, these newly developed resources also hold great promise for population genomic studies in related wild taxa. Genetic tools developed for crops have been successfully applied in ecological and evolutionary studies of wild relatives [2–6]. However, genomic studies of natural populations have trailed behind those of domesticated and other model species. The opportunities afforded by the taxonomic proximity of wild relatives to species with reference genomes are only beginning to be realized as the diminishing cost of next-generation sequencing technologies and availability of large-scale genotyping arrays make population-scale data feasible. Here we describe recent progress in genomic studies in the genus *Zea* (family *Poaceae*), which comprises an ecologically diverse collection of species including domesticated maize and its closest wild relatives. These taxa have large, complex genomes and have already contributed much to modern plant genetics (Box 1). The past few years have seen the publication of the maize genome [7], full-genome resequencing of hundreds of domesticated and wild *Zea* individuals [8–10], and characterization of transcriptome-wide patterns of expression

[11–14]. In describing recent advances we will: (i) present wild *Zea* (collectively known as the teosintes) as an ideal study system for ecological and population genomics; (ii) describe longstanding questions that are benefiting from the application of genome-wide data, including maize domestication from teosinte, reciprocal gene flow between teosinte and maize, and local adaptation of teosinte populations; and (iii) discuss future research prospects for this fascinating system.

The teosinte taxa

The teosintes are a group of primarily outcrossing species in the genus *Zea* endemic to a region stretching from northern Mexico to Central America (Figure 1a,b). These species occupy a wide range of habitats and have diverse ecologies (Figure 1c; Table 1). The genus *Zea* includes five species (Figure 1d). The annual species *Zea mays* consists of four subspecies: (i) ssp. *mays*, the domesticated maize, also known as corn; (ii) ssp. *parviglumis*, the direct progenitor of maize [15], found in the mesic low and middle elevations of southwest Mexico; (iii) ssp. *mexicana*, distributed across the cooler high elevations of the Mexican Central Plateau; and (iv) ssp. *huehuetenangensis*, found only in western Guatemala. The diploid *Zea diploperennis* and tetraploid *Zea perennis* are closely related perennial species narrowly distributed along the mountain slopes of western Mexico. Finally, *Zea luxurians* and *Zea nicaraguensis* are annual, flood-tolerant species that grow in southeastern Guatemala and the Pacific coast of Nicaragua [16,17]. We refer to taxa hereafter solely by species or

Glossary

Allopatry: describes geographically and reproductively isolated populations.
Chromosomal knob: heterochromatic DNA in *Zea* comprised primarily of two classes of tandem satellite repeats, 180 bp and the 350 bp TR-1. Knobs in *Zea* are found both in the middle and on the ends of chromosomal arms, and have been associated with meiotic drive.
Fine-scale/cryptic structure: non-random exchange of genetic material among individuals in close proximity that is difficult to detect based on observable characters.
Fractionation: the mutational process leading to the reduction of gene number following a whole-genome duplication event.
Inferred allele frequency: frequency of alleles in the ancestral population estimated based on knowledge of current genotypes, population structure, and the process of genetic drift.
Introgression: gene flow and persistence of genetic material from one population or species into the gene pool of another.
Purifying selection: a process of natural selection in which deleterious alleles are removed from the gene pool.
Sympatry: describes populations which inhabit overlapping geographic areas thus allowing mating.

Table 1. Attributes of the teosinte taxa

Taxa	Genomic			Environmental			Life history	
	Chromosomes	Mean number of knobs[a]	Genome size 1C (pg)[b]	Mean annual precipitation (mm)[c]	Mean annual temperature (°C)[c]	Average altitude (m)[c]	Life cycle	Hybridization with other Zea taxa[d]
Zea mays ssp. *mays*	2n=2x=20	4	2.73	1141 ± 802	18.04 ± 5.93	1328 ± 1106	Annual	Evidence for hybridization with *parviglumis, mexicana, huehuetenangensis, diploperennis*, and *luxurians*
Zea mays ssp. *parviglumis*	2n=2x=20	8.4	2.94	1141 ± 194	22.84 ± 2.49	1203 ± 387	Annual	Viable hybrids commonly formed with maize supported with genetic evidence; forms limited hybrid zone with *mexicana*
Zea mays ssp. *mexicana*	2n=2x=20	8.5	2.58	794 ± 149	16.38 ± 2.16	2179 ± 325	Annual	Viable hybrids commonly formed with maize supported with genetic evidence; forms limited hybrid zone with *parviglumis*; some genetic evidence for hybridization with *luxurians*
Zea mays ssp. *huehuetenangensis*	2n=2x=20	11.5	N/A[e]	2093 ± 275	21.22 ± 1.06	1181 ± 0.3	Annual	Capable of hybridization with maize
Zea diploperennis	2n=2x=20	7	2.65	1235 ± 76	17.94 ± 0.92	1779 ± 152	Perennial	Documented rare but viable hybridization with maize supported with genetic evidence
Zea perennis	2n=4x=40	N/A	5.28	1009 ± 236	16.82 ± 1.98	1946 ± 329	Perennial	No evidence of successful hybridization
Zea luxurians	2n=2x=20	10	4.58	1629 ± 1121	22.88 ± 1.17	858 ± 327	Annual	Limited evidence of hybridization with maize with some supporting genetic evidence
Zea nicaraguensis	2n=2x=20	N/A	N/A	1746 ± 85	27.91 ± 0.1	24 ± 0.1	Annual	N/A

[a]For ssp. *mays*, the number of knobs in B73 is as reported [84]. For all other species, mean knobs were calculated from [45].

[b]Genome sizes are reported from the Kew (http://data.kew.org) Plant DNA *C*-Values Database (*mays, diploperennis, perennis, luxurians*) and from [8] (*parviglumis and mexicana*).

[c]Environmental information was estimated from bioclim (http://www.worldclim.org/bioclim) variables of sites where populations are documented.

[d]Hybridization reports originate from [2,43–45,50,51].

[e]N/A, not available.

subspecies name. Exploration continues to uncover divergent populations of several *Zea* taxa [18], suggesting that there may be more to learn about the ecological and taxonomic diversity of the genus.

The teosintes garner much attention due to their close relationship with maize and their relevance to the study of plant domestication, but they also represent a particularly attractive study system for ecological and evolutionary genomics in their own right. Compared to other model angiosperms, the larger size, higher content of repetitive elements, and greater complexity of *Zea* genomes [8] make them more representative of flowering plants (Box 1; for a comparison of wild *Zea* to other emerging model systems see Table 1 in [19]). In addition, in particular portions of their ranges, numerous large and relatively undisturbed natural populations of teosinte species can be found spanning steep environmental gradients [20,21]. These particular characteristics make the teosintes well suited for studies of local adaptation and natural selection. The teosintes also provide opportunities for the study of evolution through hybridization due to the history of diffusion of maize from its domestication center in the Mexican lowlands into sympatry (see Glossary) with interfertile,

locally adapted teosinte taxa in novel environments [15,22]. Finally, the teosintes include examples of both ancient [23] and recent [24] polyploidy as well as species with varied life-history strategies (e.g., perennial vs annual growth habit and asexual vs sexual reproduction; Table 1). As we will describe in the following sections, the promise of genomic studies in this compelling study system is now being realized, and answers to numerous ecological and evolutionary questions will soon be within reach due to the opportunities provided by newly emerging genomic technologies.

Population genomics of teosinte

Evolutionary genomics of maize domestication

For much of the 20th century, the identity of the direct progenitor of maize was actively debated (see [25,26] for a summary). Genetic analysis of allozyme [27,28] and microsatellite data [15] largely settled the dispute, convincingly demonstrating a single domestication of maize in the Balsas River Basin of southwest Mexico from the teosinte *parviglumis*. One riddle in the geography of maize domestication remained, however: maize landraces (i.e., varieties grown by local farmers) most genetically similar to *parviglumis* were

Box 1. The pan-*Zea* genome

Early studies in maize were the first to physically map genes onto chromosomes and reveal the phenomenon of crossing over [76]. Subsequent cytological studies showed large-scale structural differences in *Zea* genomes due to varying heterochromatic knob content and chromosomal inversions (T.A. Kato Yamakake, PhD thesis, University of Massachusetts, 1975; [77]). Array- and sequence-based analyses have also demonstrated substantial genetic and structural variation across taxa [8,12,42,78]. However, the recent origin of the genus [61] and gene flow among taxa [2], when combined with evidence for structural variation [79,80], suggest that researchers will be able to describe core and dispensable fractions of a pan-*Zea* genome [81].

A large-scale effort is underway to investigate low-copy genic regions in *Zea* species. Currently, the high-confidence filtered gene set of the 2.1 Gb maize genome (maizesequence.org; release 5b.60) includes ~40 000 genes. Microarray data indicated that ~10% of genes across *Zea mays* accessions were associated with structural variants [82,83]; however, estimates from next-generation sequencing of 103 *Zea* genomes suggested structural variation in as many as 70% of genes and identified significant associations between variants and phenotype [8]. Analyses of syntenic regions between sorghum and maize have shown that genes with lower expression were more likely to be lost, suggesting a potential mechanism behind observed variation in gene content in inbred lines [70,83].

Extensive variation is not limited to the genic fraction of *Zea* genomes. Cytological studies of repetitive elements [knobs, transposable elements (TEs), centromere repeats], which comprise >85% of the genome, have shown dramatic differences across *Zea* [80,84]. Sequence-based analyses, however, indicate that whereas individual families of TEs have changed in abundance in recent history [12], the relative frequencies of most TEs appear similar between maize and its sister species *luxurians* [42]. Sequence analysis also confirmed predictions [85,86] that chromosomal knob content correlates positively with genome size within *Z. mays*, whereas TEs account for genome size differences between *Z. mays* and *Tripsacum* [8] (Figure I). Fluorescent *in situ* hybridization has previously shown visible differences in centromere repeat (CentC) abundance between *parviglumis* and maize, and these differences can now be confirmed and quantified using low-coverage, whole-genome sequencing (cf. [42]).

Analysis of shared and non-shared genomic regions across *Zea* individuals has begun to describe the core and dispensable fractions of the pan-*Zea* genome [81]. The emerging picture is much more fluid than previously expected, and challenges the notion of genome conservation even at the intraspecies level (Figure II). Because of its great diversity, a pan-*Zea* genome could serve as a model for studying genomic evolution.
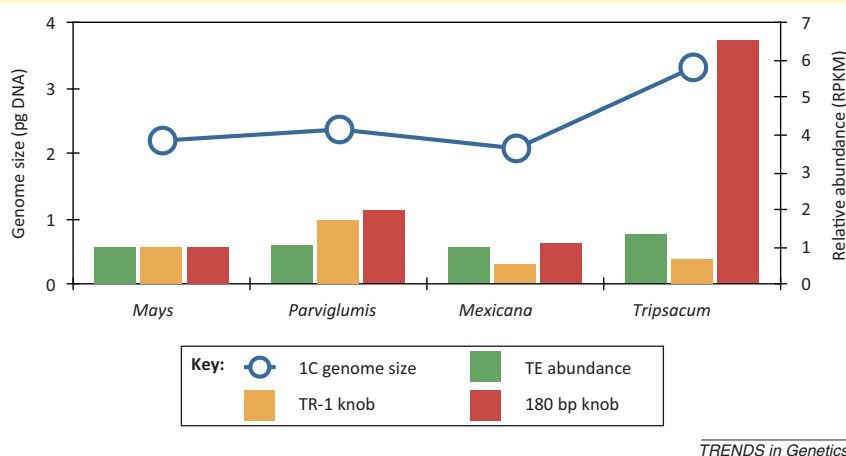
TRENDS in Genetics

Figure I. Genomic content of different classes of structural elements in maize and its relatives. Genome size, represented as a line graph, is plotted on the left axis. Repetitive element abundance was standardized against maize. B73 was used as the single reference line for maize, whereas *parviglumis* (*n* = 9) and *mexicana* (*n* = 2) are averages. A single *Tripsacum dactyloides* individual was also used. Data for the figure were obtained from [8].
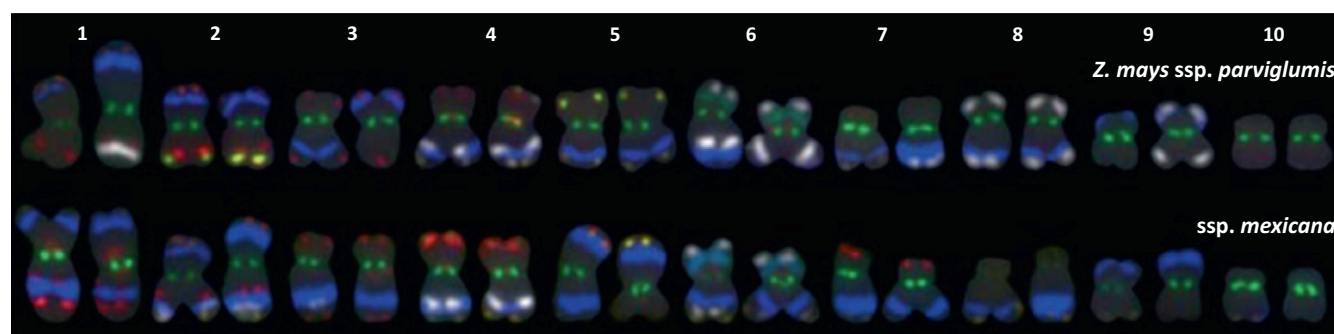
TRENDS in Genetics

Figure II. Fluorescence *in situ* hybridization in two closely related *Zea* taxa, *parviglumis* and *mexicana*. Chromosome number is indicated on the upper axis, and homologous chromosomes are shown. Karyotype probes include: 180 bp knob (blue), TR-1 knob (pseudo-colored white), CentC and subtelomere 4-12-1 (green), 5S ribosomal gene (yellow), Cent4 (orange), NOR (blue-green), and TAG microsatellite (1-26-2) and subtelomere 1.1 (red). Regions of overlapping probes exhibit a corresponding color change [84]. Reprinted, with permission, from [84] (S. Karger AG, Basel).
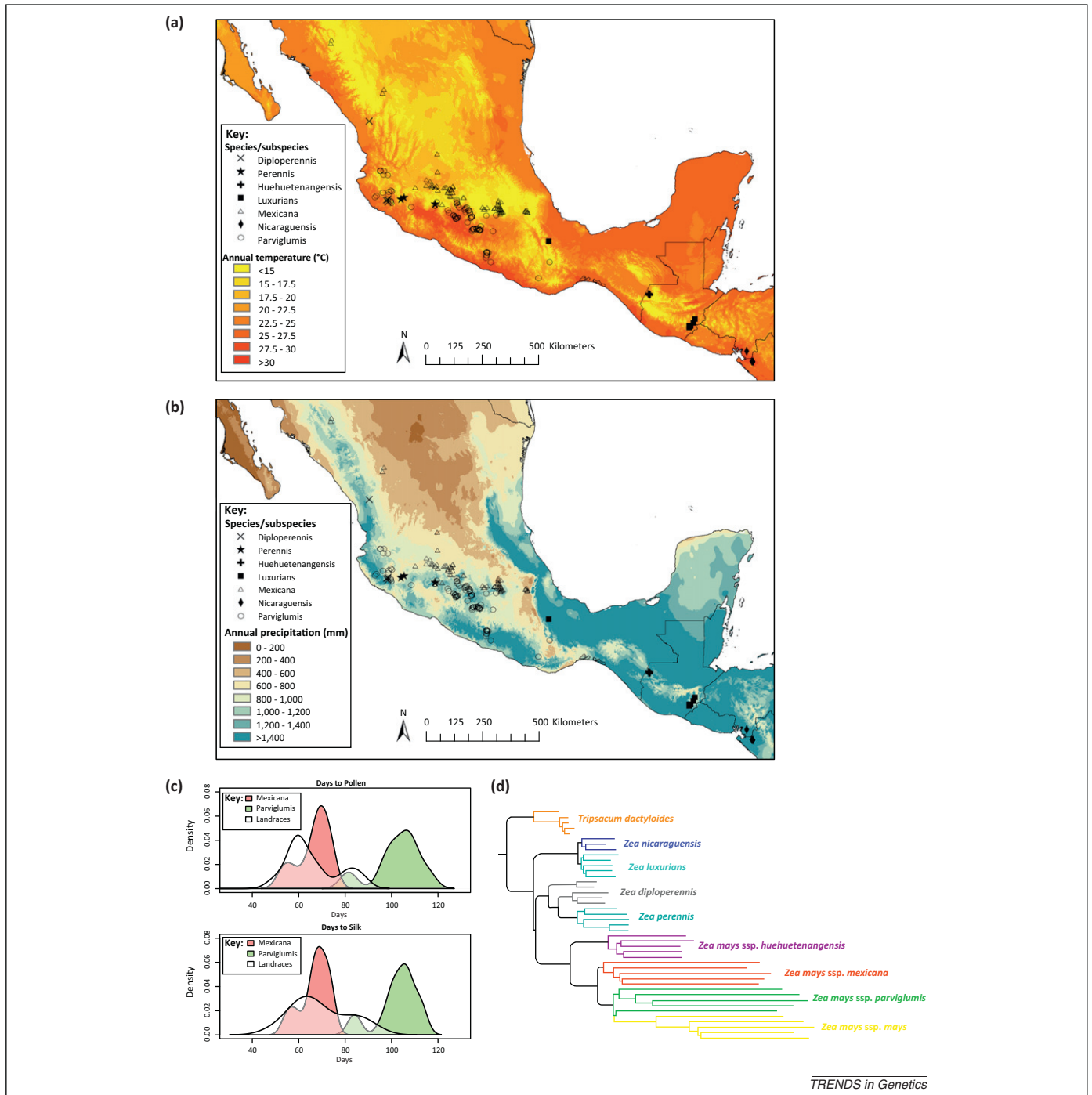
Figure 1. Ecological and taxonomic diversity of *Zea*. The wild *Zea* are distributed across a wide climatic range of **(a)** temperature and **(b)** precipitation. **(c)** Within-taxon and between-taxa variation has also been observed in traits such as flowering time (summary based on data from [62]). **(d)** Phylogeny of *Zea* based on ~1000 SNPs from across the genome (data from [65]).

from the Mexican highlands, where *parviglumis* is not found [15]. This contradiction was recently resolved using a sizeable panel of single nucleotide polymorphisms (SNPs) typed in *parviglumis*, *mexicana*, and landraces from across the Americas [29]. *Parviglumis* and *mexicana* were less differentiated from each other than either was from maize, and much more admixture was found between maize and *mexicana* than between maize and *parviglumis*. Somewhat counter-intuitively, introgression from *mexicana* was shown as the likely cause of the genetic similarity between highland maize landraces and *parviglumis*. In contrast to

comparisons with *parviglumis*, when inferred allele frequencies of ancestral maize were compared to extant maize landraces, the highest genetic similarity was found with landraces from lowland West Mexico, the putative maize center of origin [29].

In addition to these phylogeographic questions, evolutionary biologists studying maize have been keenly interested in identifying the genes responsible for the radical phenotypic transformation that occurred during maize domestication. One of the more notable differences between maize and teosinte is plant architecture: whereas

parviglumis often produces numerous lateral branches and tillers (basal branches), maize typically possesses an unbranched single stalk. Much of this disparity in plant architecture is caused by the gene *teosinte branched1* (*tb1*), encoding a transcription factor that represses the growth of axillary organs [30]. The causal polymorphism at the *tb1* locus has recently been shown to be a single transposable element insertion in the upstream regulatory region of *tb1* that acts as an enhancer of gene expression [31]. The allele that confers the maize phenotype was also found to be segregating in some teosinte populations [31], suggesting domestication has, at least in particular instances, occurred from standing variation in the wild progenitor ([32,33] for additional potential examples). Maize is also strikingly distinct from *parviglumis* and other teosintes in the structure of its female infructescence: teosinte possesses numerous spikelets of disarticulating seed that are covered in a hard protective casing (glume), whereas maize produces a small number of infructescences with naked grain bound to a cob, making it entirely dependent on humans for propagation. Variation in grain exposure has been shown to be largely due to a single locus, *teosinte glume architecture* (*tga1* [34]). Several other domestication loci controlling traits such as flowering time, kernel weight and quality, and prolificacy (i.e., number of ears) have been identified through a combination of topdown (phenotype to selected genomic region [35,36]) and bottom-up (selected genomic region to phenotype [37–40]) genetic approaches (see [41] for a contrast and discussion of these approaches).

Genomic technologies are now revealing that domestication involved more extensive changes between maize and teosinte than merely the fixation of alternative alleles at a few large-effect loci. The availability of genome-wide resequencing data from over 100 teosinte, landrace, and modern improved lines as part of the maize HapMap2 project [8] has recently allowed a population genomic analysis of maize domestication and improvement [12]. This work identified extended regions of differentiation throughout the genome, revealed patterns consistent with stronger selection during domestication (the transition from teosinte to maize landraces) than during subsequent crop improvement (the transition from open-pollinated landraces to modern inbred lines), and identified several compelling candidate genes with stronger signatures of selection than the canonical loci *tb1* and *tga1* [12]. Comparison of polymorphism data in maize landraces and teosinte confirmed a substantial loss (17%) of diversity during the domestication bottleneck, but a genome-wide excess of rare alleles provided evidence that diversity has begun to recover. A slower recovery was found within genes, likely due to the effects of purifying selection. Several genomic regions in maize landraces showed signs of introgression from the highland teosinte *mexicana*, suggesting an important role for gene flow during post-domestication maize diffusion. The magnitude and level of variation of gene expression were also shown to have changed significantly during domestication [12] with substantial re-wiring of coexpression networks in the maize transcriptome relative to *parviglumis* [14]. The maize HapMap2 project also revealed large-scale, structural changes across the genomes of *Zea* (Box 1). Pervasive structural variants (SVs) were identified as regions showing differences in the presence or copy-number of segments of the genome. Finally, genome size differences within the species *Z. mays* were found to be caused by segregation of heterochromatic regions known as chromosomal knobs [8], whereas between-species differences (*Z. mays* compared to *luxurians* [42] and the closely related taxon *Tripsacum dactyloides* [8]) have been shown to be due largely to transposable element content.

Introgression between teosinte and maize

Subsequent to domestication, teosinte populations have persisted in both allopatry and sympatry with maize. Hybridization between maize and teosinte is well-documented where they co-occur [43–45], but genetic barriers to hybridization are also well known. At least three loci are known to act as prezygotic crossing barriers between maize and teosinte: *teosinte crossing barrier1* (*tcb1*) and gametophyte factors (*ga*) *1* and *2* [46–48]. The biochemical

---

Box 2. Genetic tools for studying maize

Maize reference genome (maizesequence.org). The version AGPv2 assembly of cultivar B73 consists of 2.1×10^9 bases, defines 39 656 genes, and reports 63 540 transcripts with high confidence. The genome sequence facilitates both full-genome resequencing of teosinte lines and annotation, which are essential for evaluating the functional consequences of observed genetic variation. The reference genome can also be used in expression analyses through read-mapping of sequenced RNA.

Genotyping methods. The Illumina MaizeSNP50 single nucleotide polymorphism (SNP) array [87] and Genotyping by Sequencing (GBS) [88] can be used for high-throughput genotyping in teosinte.

Zinc-finger nuclease genome modification. Facilitates precise addition and disruption of target nucleotides [89], and could be used to transfer, for example, teosinte alleles into maize genomes.

Databases and resources

Ac/Ds Tagging and Mu Insertion Databases (plantgdb.org/prj/ AcDsTagging and teosinte.uoregon.edu/mu-illumina/). These resources list loci with potentially disruptive transposon insertions and provide seeds of the corresponding maize lines [90,91].

CoGe (genomevolution.org). Tools for plant comparative genomics, including well-known classical genes in maize, and sets of maize orthologs syntenic with other grasses.

Genomaize (genomaize.org/). Hosts multiple tools for maize biologists, including a genome browser for nucleosome occupancy and a web tool for *in silico* PCR.

Maize Transcription Atlas (plexdb.org). Describes gene expression in 60 tissues and developmental stages for tens of thousands of genes in the reference cultivar B73 [13].

MaizeCyc (pathway.gramene.org/gramene/maizecyc.shtml). Catalogs metabolic and transport pathways of maize.

MaizeGDB (maizegdb.org). Incorporates maize genetic data from various public resources in a comprehensive database and genome browser [92].

Maize Stock Center (maizecoop.cropsci.uiuc.edu). Collection of maize genetic stocks, including information on known mutants.

Maize Transposable Element (TE) Database (maizetedb.org). Describes hundreds of maize TE families and has been used to estimate abundance of individual TE families in maize and related taxa [8,42].

Panzea (panzea.org). Numerous databases of genotypic and phenotypic data for association mapping and evolutionary analyses.

QTeller (qteller.com). A tool to summarize function, expression, and synteny information from maize genes in a given genomic interval.

mechanism underlying the action of these loci remains unclear, but alleles conferring incompatibility at *ga1* and *tcb1* appear to be present at higher frequency in *mexicana* than in *parviglumis* populations [48].

Hybridization barriers may dramatically limit the rate of gene flow, but theory predicts that if particular combinations of genotypes show increased fitness, introgression can occur despite barriers to gene exchange [49]. Recent
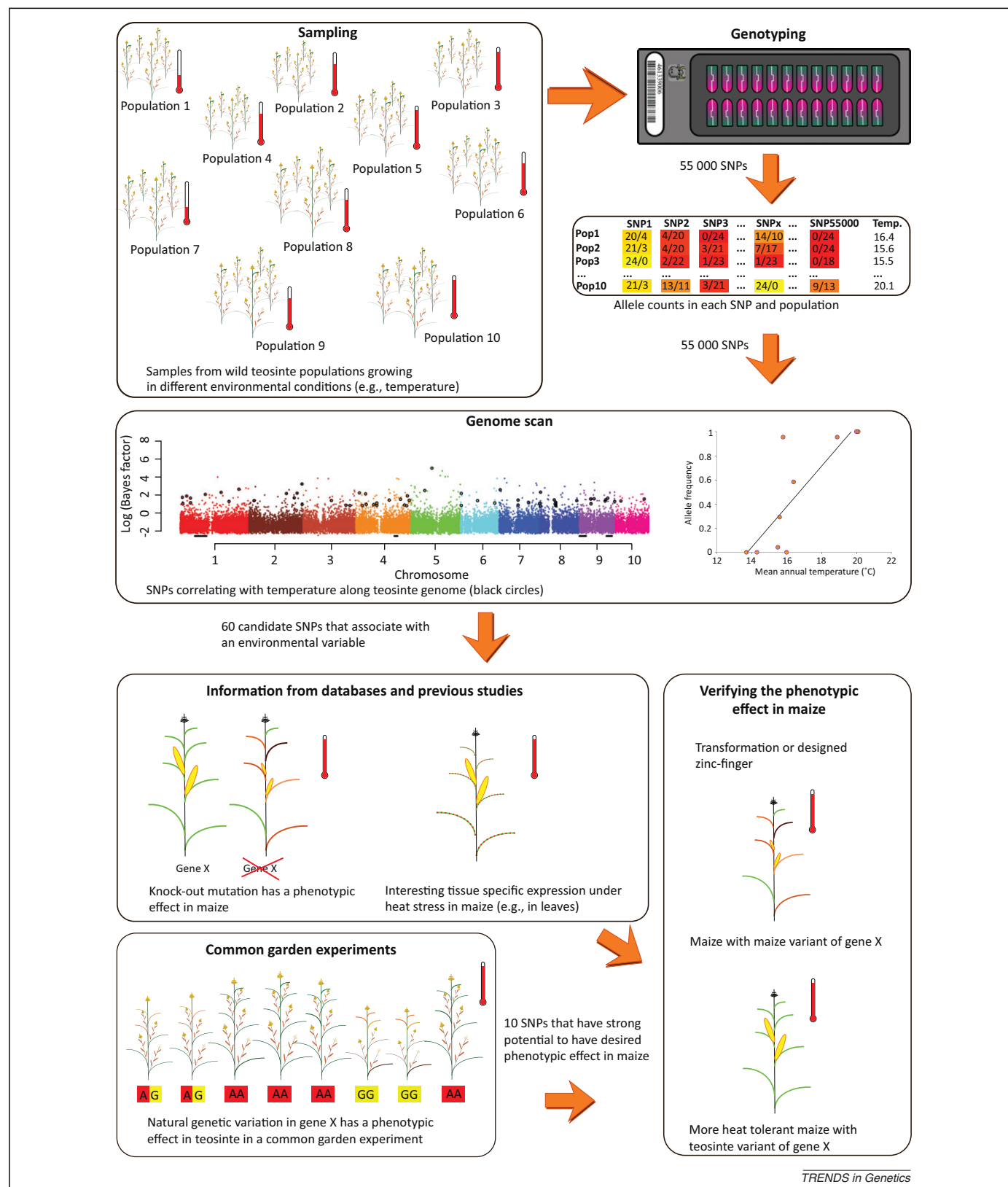

Figure 2. Discovery process from population genomics in teosinte to crop improvement in maize. Natural variation among wild teosinte populations can be used to identify genetic variation that correlates with environmental variables. Further, common garden experiments, maize mutation databases, and tissue-specific expression data can be used to identify genes potentially underlying adaptation to the environment. Finally, causative variants can be transformed into maize to verify phenotypic effects.

evidence suggests reciprocal introgression between the partially incompatible subspecies *mexicana* and maize may be a compelling example of such gene flow. Following its domestication in the lowlands of southwest Mexico from *parviglumis*, maize diffused to the highlands of the Mexican Central Plateau where it encountered dramatically different environmental conditions and came into sympatry with the highland teosinte *mexicana*. Modern highland maize and *mexicana* share multiple phenotypic traits (e.g., abundant macrohairs and leaf pigmentation [45]). This shared morphology has been attributed both to introgression from maize into *mexicana* conferring crop mimicry [50] and to *mexicana* introgression into maize conferring highland adaptation [51]. The application of molecular markers has revealed considerable evidence for gene flow in both directions, suggesting that as much as 20% of the genome of highland maize is derived from *mexicana* and approximately 10% of the *mexicana* genome is derived from maize [29].

Recent work has taken advantage of the availability of genome-wide SNP typing to evaluate the genomic imprint of introgression in sympatric populations of highland Mexican maize and *mexicana* [52]. Numerous genomic regions of introgression were identified in both directions of gene flow, but domestication loci appeared to be particularly resistant to introgression. Interestingly, regions of *mexicana*-to-maize introgression were shared across populations, suggesting either ancient gene flow or multiple independent introgression events at the same loci. These shared regions of introgression from *mexicana* into maize were found to colocalize with previously identified QTL for highland traits [53], suggesting a possible role for *mexicana* in conferring maize adaptation to highland conditions. A growth-chamber experiment simulating highland conditions, and including maize populations with and without *mexicana* introgression at these shared loci, showed significantly higher leaf sheath macrohair counts, extent of pigment, and plant height in populations with a high frequency of *mexicana*-to-maize introgression [52].

Local adaptation in teosinte

Local adaptation occurs when resident genotypes have higher fitness than genotypes from other locations [54]. Currently, there are only a handful of examples in which the genetic basis of local adaptation is well understood. Because of their sessile nature, plants are particularly well suited for studies of local adaptation [19], and genome-wide data have recently been analyzed in plant populations across wide geographic ranges [55,56].

Among the teosintes, *mexicana* and *parviglumis* are the best candidates for studies of local adaptation. Both have widespread geographic distributions (Figure 1a,b) consisting of hundreds of distinct populations [2,20,57] that display varying levels of genetic differentiation [2,58], from very fine-scale structure (F_{ST}: 1–2%) within a site [59,60] to more pronounced differentiation between subspecies (F_{ST}: ~15% [61]). Populations of *parviglumis* and *mexicana* also span a wide climatic range of temperature and precipitation (Table 1). Phenotypically, these subspecies are differentiated based on the smaller spikelets and fruitcases and more open tassels of *parviglumis* in comparison to

mexicana [16], and by the hairier leaf sheaths and higher pigment content found in *mexicana* [53]. Within subspecies, individual *parviglumis* and *mexicana* populations are known to vary in several morphological and phenological characteristics [57,62,63]. It is not yet clear whether population-level phenotypic differences are due to local adaptation or if they are instead byproducts of genetic drift.

Recent studies have investigated the molecular variation underlying local adaptation in teosinte. A study of variation in plant immunity genes across six *parviglumis* populations found little evidence for widespread geographically variable selection, but did observe diversity patterns suggesting localized natural selection in the gene *wip1* [64]. The *wip1* gene product, a protease inhibitor, is known to interfere with the dietary proteins of insects, and selection at this locus could represent local adaptation to herbivory. A second study of SNP diversity in *mexicana* and *parviglumis* identified *Inv1n*, a 50 Mb chromosomal inversion found segregating in 33 *parviglumis* populations [65]. The frequency of *Inv1n* was shown to vary along an altitudinal cline (i.e., a continuous gradient) and associate with environmental and morphological characters. Such inversions are predicted to be beneficial when they capture locally adapted alleles or when the breakpoints of the inversions themselves are beneficial [66]. Most recently, a genome-wide analysis of 36 719 SNPs typed in 10 *mexicana* and 11 *parviglumis* populations identified candidate loci underlying local adaptation [67]. Contrary to previous findings in *Arabidopsis thaliana* [56], enrichment of nongenic SNPs was found among candidates, in line with the observation that a substantial proportion of phenotypic associations in maize are found in noncoding regions [8]. In addition, several newly identified inversion polymorphisms showed an excess of differentiation among populations and subspecies, together with enrichment for association with environmental variables such as temperature and altitude. The complexity and size of the *Zea* genome (Box 1) may provide ample opportunities for such large-scale chromosomal rearrangements to occur. Finally, the study revealed that factors such as hierarchical population structure and uneven distribution of environmental variation can bias detection of locally adaptive loci based on a single statistical method, underscoring the importance of utilizing multiple approaches.

Concluding remarks and future prospects

The teosintes provide opportunities to examine many compelling questions at the genome level using the vast number of resources developed in maize (Box 2). As the cost of high-throughput sequencing continues to decrease and *de novo* reference-assisted assembly becomes commonplace in teosinte, genomic studies can move farther beyond maize, delving more deeply into the evolutionary histories of these taxa. For example, the species *luxurians* and *nicaraguensis* are uniquely adapted to the flooded and coastal conditions of Central America. Unlike other teosintes, these taxa develop aerenchyma (air channels that allow gas exchange between submerged and above-water organs of a plant) both in flooded and non-flooded conditions [68], and multiple QTL have already been identified for this trait [69]. Past studies have suggested potential hybridization between these

Review

teosintes and local maize landraces [50,51,61], and additional field and genome-wide analyses should assess evidence for adaptive introgression for flood tolerance in landraces at loci encoding these traits.

The teosintes also provide an excellent system in which to study the genetics of speciation. The genus appears to have radiated relatively recently, and shared polymorphism and ongoing gene-flow are evident across species [61]. Perhaps of greatest interest would be field-based studies of local adaptation (e.g., development of near-isogenic lines and reciprocal transplant experiments), incipient speciation, and admixture between *parviglumis* and *mexicana*. These subspecies appear to have diverged only recently, but already exhibit locally adaptive phenotypes, and gene flow between them is limited to a narrow hybrid zone [2,67]. Studies of gene flow and reinforcement within this zone could shed light on the process of divergence in these subspecies.

A further potential topic of investigation in *Zea* is the recent autopolyploidization (within the last few hundred thousand years [24]) of *perennis*. Investigation of the more ancient polyploidization event ancestral to all of *Zea* has shown that genome fractionation has been biased toward a single subgenome and is ongoing, distinguishing even modern inbred maize lines [70]. *Perennis* presents an opportunity to investigate incipient fractionation/diploidization in natural populations by utilizing newly developed resources to assess genome-wide patterns of gene expression, DNA methylation, and deletion.

Genomic resources could also be brought to bear on the evolution of the perennial life-history in *diploperennis* and *perennis*, a feature that distinguishes these species from the remaining annual *Zea* taxa. A previous QTL investigation of *diploperennis* and *parviglumis* did not find large-effect loci controlling traits associated with perennialism, and very few genes were linked to the evolution of the perennial growth habit [71]. Full genome resequencing may allow population-genetic-based assessment of selection, but it seems likely that increased marker density would be better exploited by multiparent mapping populations (e.g., [72,73]) of annual/perennial taxa that can take advantage of a large number of recombination events to provide increased resolution. Full transcriptome patterns of gene expression would also provide valuable data for identifying loci involved in the contrasting growth habits across *Zea*. Once these full-genome data are generated, and we have a better idea about the loci involved in *Zea* perennialism, many fascinating ecological and evolutionary questions can be investigated. For example: do genes associated with perennialism explain the capacity of these species to tolerate conditions outside the normal growing season of the annual *Zea* (e.g., cold-tolerance or drought-resistance loci)? To what degree is there overlap with the pathways and orthologous genes involved in perennialism in closely related taxa such as rice and sorghum?

Although several fascinating evolutionary questions can be addressed within *Zea* on a species-by-species basis, the genus as a whole can also be seen as an ideal study system for investigating broader and potentially far-reaching topics. For instance, *Zea* genomes are more representative than the genomes of most current angiosperm models in that they are comprised primarily of repetitive elements [transposable elements (TEs), microsatellites, and centromere and chromosomal knob repeats] [74]. Within *Zea*, the evolution of repetitive elements can be studied over hundreds of thousands of years, revealing epochs of TE transposition, the expansion and contraction of knob and centromere repeat arrays, and the structural and functional roles of this largely ignored fraction of the genome. *Zea* species also have a wide range of effective population sizes, allowing comparisons of the efficiency of natural selection and the effects of rare variants across closely related taxa. Association studies in the wild taxa of *Zea* will also provide greater resolution for the genetic architecture of complex traits than is currently possible in their domesticated congeners owing to the increased number of recombination events in natural populations and the rapid decay of linkage disequilibrium [8].

Finally, much of the proposed evolutionary genomic work in the teosintes could feed back into the improvement of maize (Figure 2). For example, temperature-related highland adaptation loci could be important for maize breeding in the face of climate change, flood-tolerance loci identified in introgression or continued QTL studies of *luxurians* and *nicaraguensis* could help mitigate crop losses due to flooding [75], and genomic studies of *diploperennis* and *perennis* may facilitate the development of perennial maize, a resource that could potentially reduce inputs and prove particularly useful in agroecosystems of the developing world. In this way, the model/non-model gap will be bridged not only through application of maize-derived technologies to teosinte, but also through valuable contributions from the rich evolutionary past of *Zea* to the future of maize.

References

1 Morrell, P.L. *et al.* (2012) Crop genomics: advances and applications. *Nat. Rev. Genet.* 13, 85–96
2 Fukunaga, K. *et al.* (2005) Genetic diversity and population structure of teosinte. *Genetics* 169, 2241–2254
3 Martinez-Castillo, J. *et al.* (2006) Structure and genetic diversity of wild populations of lima bean (*Phaseolus lunatus* L.) from the Yucatan Peninsula, Mexico. *Crop Sci.* 46, 1071–1080
4 Peleg, Z. *et al.* (2008) Genetic structure of wild emmer wheat populations as reflected by transcribed versus anonymous SSR markers. *Genome* 51, 187–195
5 Sapir, Y. *et al.* (2007) Patterns of genetic diversity and candidate genes for ecological divergence in a homoploid hybrid sunflower, *Helianthus anomalus*. *Mol. Ecol.* 16, 5017–5029
6 Zheng, X.M. and Ge, S. (2010) Ecological divergence in the presence of gene flow in two closely related *Oryza* species (*Oryza rufipogon* and *O. nivara*). *Mol. Ecol.* 19, 2439–2454
7 Schnable, P.S. *et al.* (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115

8 Chia, J-M. *et al.* (2012) Maize HapMap2 identifies extant variation from a genome in flux. *Nat. Genet.* 44, 803–807

9 Gore, M.A. *et al.* (2009) A first-generation haplotype map of maize. *Science* 326, 1115–1117

10 Jiao, Y. *et al.* (2012) Genome-wide genetic changes during modern breeding of maize. *Nat. Genet.* 44, 812–815

11 Li, P. *et al.* (2010) The developmental dynamics of the maize leaf transcriptome. *Nat. Genet.* 42, 1060–1067

12 Hufford, M.B. *et al.* (2012) Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44, 808–811

13 Sekhon, R.S. *et al.* (2011) Genome-wide atlas of transcription during maize development. *Plant J.* 66, 553–563

14 Swanson-Wagner, R. *et al.* (2012) Reshaping of the maize transcriptome by domestication. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11878–11883

15 Matsuoka, Y. *et al.* (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6080–6084

16 Iltis, H.H. and Doebley, J.F. (1980) Taxonomy of *Zea* (Gramineae). II. Subspecific categories in the *Zea mays* complex and a generic synopsis. *Am. J. Bot.* 67, 994–1004

17 Doebley, J. (1990) Molecular systematics of *Zea* (Gramineae). *Maydica* 35, 143–150

18 Sanchez, G.J.J. *et al.* (2011) Three new teosintes (*Zea* spp., *Poaceae*) from Mexico. *Am. J. Bot.* 98, 1537–1548

19 Anderson, J.T. *et al.* (2011) Evolutionary genetics of plant adaptation. *Trends Genet.* 27, 258–266

20 Sanchez, G.J.J. and Corral, J.A.R. (1997) Teosinte distribution in Mexico. In *Gene Flow Among Maize Landraces, Improved Maize Varieties and Teosinte: Implications for Transgenic Maize* (Serratos, J.A. *et al.*, eds), pp. 18–36, International Maize and Wheat Improvement Center (CIMMYT)

21 Hufford, M.B. *et al.* (in press) Inferences from the historical distribution of wild and domesticated maize provide ecological and evolutionary insight. *PLoS ONE*.

22 Vigouroux, Y. *et al.* (2008) Population structure and genetic diversity of new world maize races assessed by DNA microsatellites. *Am. J. Bot.* 95, 1240–1253

23 Gaut, B.S. and Doebley, J.F. (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc. Natl. Acad. Sci. U.S.A.* 94, 6809–6814

24 Tiffin, P. and Gaut, B.S. (2001) Sequence diversity in the tetraploid *Zea perennis* and the closely related diploid *Z. diploperennis*: insights from four nuclear loci. *Genetics* 158, 401–412

25 Bennetzen, J. *et al.* (2001) Genetic evidence and the origin of maize. *Lat. Am. Antiq.* 12, 84–86

26 Doebley, J. (2001) George Beadle's other hypothesis: one-gene, one-trait. *Genetics* 158, 487–493

27 Doebley, J. (1990) Molecular evidence and the evolution of maize. *Econ. Bot.* 44, 6–27

28 Doebley, J. *et al.* (1987) Patterns of isozyme variation between maize and Mexican annual teosinte. *Econ. Bot.* 41, 234–246

29 van Heerwaarden, J. *et al.* (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1088–1092

30 Doebley, J. *et al.* (1997) The evolution of apical dominance in maize. *Nature* 386, 485–488

31 Studer, A. *et al.* (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat. Genet.* 43, 1160–1163

32 Gallavotti, A. *et al.* (2004) The role of *barren stalk1* in the architecture of maize. *Nature* 432, 630–635

33 Sigmon, B. and Vollbrecht, E. (2010) Evidence of selection at the *ramosa1* locus during maize domestication. *Mol. Ecol.* 19, 1296–1311

34 Wang, H. *et al.* (2005) The origin of the naked grains of maize. *Nature* 436, 714–719

35 Briggs, W.H. *et al.* (2007) Linkage mapping of domestication loci in a large maize-teosinte backcross resource. *Genetics* 177, 1915–1928

36 Buckler, E.S. *et al.* (2009) The genetic architecture of maize flowering time. *Science* 325, 714–718

37 Tenaillon, M.I. *et al.* (2004) Selection versus demography: a multilocus investigation of the domestication process in maize. *Mol. Biol. Evol.* 21, 1214–1225

38 Vigouroux, Y. *et al.* (2002) Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9650–9655

39 Wright, S.I. *et al.* (2005) The effects of artificial selection of the maize genome. *Science* 308, 1310–1314

40 Yamasaki, M. *et al.* (2005) A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* 17, 2859–2872

41 Ross-Ibarra, J. *et al.* (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8641–8648

42 Tenaillon, M.I. *et al.* (2011) Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*. *Genome Biol. Evol.* 3, 219–229

43 Baltazar, B.M. *et al.* (2005) Pollination between maize and teosinte: an important determinant of gene flow in Mexico. *Theor. Appl. Genet.* 110, 519–526

44 Ellstrand, N.C. *et al.* (2007) Spontaneous hybridization between maize and teosinte. *J. Hered.* 98, 183–187

45 Wilkes, H.G. (1967) *Teosinte: The Closest Relative of Maize*, The Bussey Institute of Harvard University

46 Evans, M.M.S. and Kermicle, J.L. (2001) *Teosinte crossing barrier1*, a locus governing hybridization of teosinte with maize. *Theor. Appl. Genet.* 103, 259–265

47 Kermicle, J.L. and Evans, M.M.S. (2010) The *Zea mays* sexual compatibility gene *ga2*: naturally occurring alleles, their distribution, and role in reproductive isolation. *J. Hered.* 101, 737–749

48 Kermicle, J.L. *et al.* (2006) The *gametophyte-1* locus and reproductive isolation among *Zea mays* subspecies. *Maydica* 51, 219–225

49 Barton, N.H. (2001) The role of hybridization in evolution. *Mol. Ecol.* 10, 551–568

50 Wilkes, H.G. (1977) Hybridization of maize and teosinte, in Mexico and Guatemala and improvement of maize. *Econ. Bot.* 31, 254–293

51 Doebley, J. (1990) Molecular evidence for gene flow among *Zea* species – genes transformed into maize through genetic-engineering could be transferred to its wild relatives, the teosintes. *Bioscience* 40, 443–448

52 Hufford, M.B. *et al.* (2012) The genomic signature of crop-wild introgression in maize. arXiv:1208.3894v2 http://arxiv.org/abs/1208.3894

53 Lauter, N. *et al.* (2004) The inheritance and evolution of leaf pigmentation and pubescence in teosinte. *Genetics* 167, 1949

54 Kawecki, T.J. and Ebert, D. (2004) Conceptual issues in local adaptation. *Ecol. Lett.* 7, 1225–1241

55 Fournier-Level, A. *et al.* (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science* 334, 86–89

56 Hancock, A.M. *et al.* (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334, 83–86

57 Weber, A. *et al.* (2007) Major regulatory genes in maize contribute to standing variation in teosinte (*Zea mays* ssp. *parviglumis*). *Genetics* 177, 2349–2359

58 Moeller, D.A. *et al.* (2007) Population structure and its effects on patterns of nucleotide polymorphism in teosinte (*Zea mays* ssp. *parviglumis*). *Genetics* 176, 1799–1809

59 Hufford, M.B. *et al.* (2011) Influence of cryptic population structure on observed mating patterns in the wild progenitor of maize (*Zea mays* ssp. *parviglumis*). *Mol. Ecol.* 20, 46–55

60 van Heerwaarden, J. *et al.* (2010) Fine scale genetic structure in the wild ancestor of maize (*Zea mays* ssp. *parviglumis*). *Mol. Ecol.* 19, 1162–1173

61 Ross-Ibarra, J. *et al.* (2009) Historical divergence and gene flow in the genus *Zea*. *Genetics* 181, 1399–1413

62 Rodriguez, F.J.G. *et al.* (2006) Characterization of floral morphology and synchrony among *Zea* species in Mexico. *Maydica* 51, 383–398

63 Weber, A.L. *et al.* (2008) The genetic architecture of complex traits in teosinte (*Zea mays* ssp. *parviglumis*): new evidence from association mapping. *Genetics* 180, 1221–1232

64 Moeller, D.A. and Tiffin, P. (2008) Geographic variation in adaptation at the molecular level: a case study of plant immunity genes. *Evolution* 62, 3069–3081

65 Fang, Z. *et al.* (2012) Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics* 191, 883–894

66 Kirkpatrick, M. and Barton, N. (2006) Chromosome inversions, local adaptation and speciation. *Genetics* 173, 419–434

67 Pyhäjärvi, T. *et al.* (2012) Complex patterns of local adaptation in teosinte. arXiv:1208.0634v1 http://arxiv.org/abs/1208.0634

68 Mano, Y. *et al.* (2006) Variation for root aerenchyma formation in flooded and non-flooded maize and teosinte seedlings. *Plant Soil* 281, 269–279

69 Mano, Y. *et al.* (2008) A linkage map of maize × teosinte Zea luxurians and identification of QTLs controlling root aerenchyma formation. *Mol. Breeding* 21, 327–337

70 Schnable, J.C. *et al.* (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4069–4074

71 Westerbergh, A. and Doebley, J. (2004) Quantitative trait loci controlling phenotypes related to the perennial versus annual habit in wild relatives of maize. *Theor. Appl. Genet.* 109, 1544–1553

72 Churchill, G. *et al.* (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.* 36, 1133–1137

73 Macdonald, S.J. and Long, A.D. (2007) Joint estimates of quantitative trait locus effect and frequency using synthetic recombinant populations of *Drosophila melanogaster*. *Genetics* 176, 1261–1281

74 Tenaillon, M.I. *et al.* (2010) A triptych of the evolution of plant transposable elements. *Trends Plant Sci.* 15, 471–478

75 Mano, Y. and Omori, F. (2007) Breeding for flooding tolerant maize using 'teosinte' as a germplasm resource. *Plant Root* 1, 17–21

76 Creighton, H.B. and McClintock, B. (1931) A correlation of cytological and genetical crossing-over in *Zea mays*. *Proc. Natl. Acad. Sci. U.S.A.* 17, 492–497

77 Ting, Y.C. (1967) Common inversion in maize and teosinte. *Am. Nat.* 101, 87–89

78 Springer, N.M. *et al.* (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet.* 5, e1000734

79 Brunner, S. *et al.* (2005) Evolution of DNA sequence nonhomologies among maize inbreds. *Plant Cell* 17, 343–360

80 Lamb, J.C. and Birchler, J.A. (2006) Retroelement genome painting: cytological visualization of retroelement expansions in the genera *Zea* and *Tripsacum*. *Genetics* 173, 1007–1021

81 Morgante, M. *et al.* (2007) Transposable elements and the plant pan-genomes. *Curr. Opin. Plant Biol.* 10, 149–155

82 Eichten, S.R. *et al.* (2011) B73-Mo17 near-isogenic lines demonstrate dispersed structural variation in maize. *Plant Physiol.* 156, 1679–1690

83 Swanson-Wagner, R.A. *et al.* (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res.* 20, 1689–1699

84 Albert, P.S. *et al.* (2010) Diversity of chromosomal karyotypes in maize and its relatives. *Cytogenet. Genome Res.* 129, 6–16

85 Laurie, D.A. and Bennett, M.D. (1985) Nuclear-DNA content in the genera *Zea* and *Sorghum* – intergeneric, interspecific and intraspecific variation. *Heredity* 55, 307–313

86 Poggio, L. *et al.* (1998) Genome size and environmental correlations in maize (*Zea mays* ssp. *mays, Poaceae*). *Ann. Bot. (Lond.)* 82, 107–115

87 Ganal, M.W. *et al.* (2011) A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS ONE* 6, e28334

88 Elshire, R.J. *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6, e19379

89 Shukla, V.K. *et al.* (2009) Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* 459, 437–441

90 Vollbrecht, E. *et al.* (2010) Genome-wide distribution of transposed dissociation elements in maize. *Plant Cell* 22, 1667–1685

91 Williams-Carrier, R. *et al.* (2010) Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy *Mutator* lines of maize. *Plant J.* 63, 167–177

92 Lawrence, C.J. *et al.* (2004) MaizeGDB, the community database for maize genetics and genomics. *Nucleic Acids Res.* 32, D393–D397