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.

Corresponding author: Ross-Ibarra, J. (rossibarra@ucdavis.edu).
Keywords: population genomics; crop wild relatives; domestication; local adaptation; introgression; *Zea*.
[*] *Current address*: Department of Biology, University of Oulu, 90014 University of Oulu, Finland.

0168-9525/$ – see front matter © 2012 Elsevier Ltd. All rights reserved. http://dx.doi.org/10.1016/j.tig.2012.08.004 Trends in Genetics, December 2012, Vol. 28, No. 12

Table 1. Attributes of the teosinte taxa

Taxa	Genomic			Environmental			Life history	
	Chromosomes	Mean number of knobs[a]	Genome size 1*C* (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