Wild *Zea* as a Model System for Population and Ecological Genomics

Matthew B. Hufford[1], Paul Bilinski[1], Tanja Pyhäjärvi[1], 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

Corresponding author: Ross-Ibarra, J. (rossibarra@ucdavis.edu)

Keywords: population genomics; crop wild relatives; domestication; local adaptation; introgression; *Zea*

Abstract

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.*, see Table 1 of [1]). While 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 wild relatives' taxonomic proximity to reference genomes

2

37 are only beginning to be realized as the diminishing cost of next-generation

38 sequencing technologies and availability of large-scale genotyping arrays make

39 population-scale data feasible. Here we describe recent progress in genomic studies

40 in the genus *Zea* (Family *Poaceae*), which comprises an ecologically diverse

41 collection of species including domesticated maize and its closest wild relatives.

42 These taxa have large, complex genomes and have already contributed much to

43 modern plant genetics (Box 1). The past few years have seen the publication of the

44 maize genome [7], full-genome resequencing of hundreds of domesticated and wild

45 *Zea* individuals [8-10], and characterization of transcriptome-wide patterns of

46 expression [11-14]. In describing recent advances we will: 1) Present wild *Zea*

47 (collectively known as the teosintes) as an ideal study system for ecological and

48 population genomics; 2) Describe long-standing questions that are benefitting from

49 the application of genome-wide data, including maize domestication from teosinte,

50 reciprocal gene flow between teosinte and maize, and local adaptation of teosinte

51 populations; and 3) Discuss future research prospects for this fascinating research

52 system.

53

54 **The Teosinte Taxa**

55

56   The teosintes are a group of primarily outcrossing species in the genus *Zea*

57 endemic to a region stretching from northern Mexico to Central America (Figure 1A

58 & 1B). These species occupy a wide range of habitats and have diverse ecologies

59 (Figure 1C; Table 1; Box 1). The genus *Zea* includes five species (Figure 1D). The

60 annual species *Zea mays* consists of four subspecies: 1) ssp. *mays,* the domesticated

61 maize, also known as corn; 2) ssp. *parviglumis*, the direct progenitor of maize [15],

62 found in the mesic low and middle elevations of southwest Mexico; 3) ssp. *mexicana*,

63 distributed across the cooler high elevations of the Mexican Central Plateau; and 4)

64 ssp. *huehuetenangensis*, found only in western Guatemala. The diploid *Zea*

65 *diploperennis* and tetraploid *Zea perennis* are closely related perennial species

66 narrowly distributed along the mountain slopes of western Mexico. Finally, *Zea*

67 *luxurians* and *Zea nicaraguensis* are annual, flood-tolerant species that grow in

68 southeastern Guatemala and the Pacific coast of Nicaragua [16,17]. We refer to taxa

69 hereafter solely by species or subspecies name. Exploration continues to uncover

70 divergent populations of several *Zea* taxa [18], suggesting that there may be more to

71 learn about the ecological and taxonomic diversity of the genus.

72   The teosintes garner much attention due to their close relationship with

73 maize and their relevance to the study of plant domestication, but they also

74 represent a particularly attractive study system for ecological and evolutionary

75 genomics in their own right. When compared to other model angiosperms, the

76 larger size, higher content of repetitive elements, and greater complexity of *Zea*

77 genomes [8] make them more representative of flowering plants (Box 1; for a

78 comparison of wild *Zea* to other emerging model systems see Table 1 of [19]).

79 Additionally, in certain portions of their ranges, numerous large and relatively

80 undisturbed natural populations of the teosinte species can be found spanning steep

81 environmental gradients ([20], M. B. Hufford *et al.*, unpublished). These particular

82 characteristics make the teosintes well suited for studies of local adaptation and

4

83   natural selection. The teosintes also provide opportunities for the study of

84   evolution through hybridization due to the history of diffusion of maize from its

85   domestication center in the Mexican lowlands into sympatry with interfertile, locally

86   adapted teosinte taxa in novel environments [15,21]. Finally, the teosintes include

87   examples of both ancient [22] and recent [23] polyploidy as well as species with

88   varied life-history strategies (*e.g.,* perennial vs. annual growth habit and asexual vs.

89   sexual reproduction; Table 1). As we will describe in the following sections, the

90   promise of genomic studies in this compelling study system is now being realized

91   and answers to numerous ecological and evolutionary questions will soon be within

92   reach due to the opportunities provided by newly emerging genomic technologies.

93

94   **Population Genomics of Teosinte**

95

96   *Evolutionary Genomics of Maize Domestication*

97

98   For much of the twentieth century, the identity of the direct progenitor of

99   maize was actively debated (for a summary see [24,25]). Genetic analysis of

100 allozyme [26,27] and microsatellite data [15] largely settled the dispute,

101 convincingly demonstrating a single domestication of maize in the Balsas River

102 Basin of southwest Mexico from the teosinte *parviglumis*. One riddle in the

103 geography of maize domestication remained, however: maize landraces (*i.e.*,

104 varieties grown by local farmers) most genetically similar to *parviglumis* were from

105 the Mexican highlands, where *parviglumis* is not found [15]. This contradiction was

106   resolved using a sizeable panel of single nucleotide polymorphisms (SNPs) typed in

107   *parviglumis*, *mexicana*, and landraces from across the Americas [28]. *Parviglumis*

108   and *mexicana* were less differentiated from each other than either was from maize,

109   and much more admixture was found between maize and *mexicana* than between

110   maize and *parviglumis.* Somewhat counter-intuitively, introgression from *mexicana*

111   was shown as the likely cause of the genetic similarity between highland maize

112   landraces and *parviglumis.* In contrast to comparisons with *parviglumis*, when

113   inferred allele frequencies of ancestral maize were compared to extant maize

114   landraces, the highest genetic similarity between *parviglumis* and maize was found

115   with landraces from lowland West Mexico, the putative maize center of origin [28].

116   In addition to these phylogeographic questions, evolutionary biologists

117   studying maize have been keenly interested in identifying the genes responsible for

118   the radical phenotypic transformation that occurred during maize domestication.

119   One of the more notable differences between maize and teosinte is plant

120   architecture: whereas *parviglumis* often produces numerous lateral branches and

121   tillers (basal branches), maize typically possesses an unbranched single stalk. Much

122   of this disparity in plant architecture is caused by the gene *teosinte branched1* (*tb1*),

123   a transcription factor that represses the growth of axillary organs [29]. The causal

124   polymorphism at the *tb1* locus has recently been shown to be a single transposable

125   element insertion in the upstream regulatory region of *tb1* that acts as an enhancer

126   of gene expression [30]. The allele that confers the maize phenotype was also found

127   to be segregating in some teosinte populations [30], suggesting domestication has,

128   at least in certain instances, occurred from standing variation in the wild progenitor

6

129   (for additional potential examples see [31,32]). Maize is also strikingly distinct from

130   *parviglumis* and other teosintes in the structure of its female infructescence:

131   teosinte possesses numerous spikelets of disarticulating seed that are covered in a

132   hard protective casing (glume), whereas maize produces a small number of

133   infructescences with naked grain bound to a cob, making it entirely dependent on

134   humans for propagation. Variation in grain exposure has been shown to be largely

135   due to a single locus, *teosinte glume architecture* (*tga1*; [33]). Several other

136   domestication loci controlling traits such as flowering time, kernel weight and

137   quality, and prolificacy (*i.e.*, number of ears) have been identified through a

138   combination of top-down (phenotype to selected genomic region; [34,35]) and

139   bottom-up (selected genomic region to phenotype; [36-39]) genetic approaches (see

140   [40] for a contrast and discussion of these approaches).

141   Genomic technologies are now revealing that domestication involved more

142   extensive changes between maize and teosinte than merely the fixation of alternate

143   alleles at a few large-effect loci. The availability of genome-wide resequencing data

144   from over 100 teosinte, landrace, and modern improved lines as part of the maize

145   HapMap2 project [8] has recently allowed a population genomic analysis of maize

146   domestication and improvement [12]. This work identified extended regions of

147   differentiation throughout the genome, revealed patterns consistent with stronger

148   selection during domestication (the transition from teosinte to maize landraces)

149   than during subsequent crop improvement (the transition from open-pollinated

150   landraces to modern inbred lines) and identified a number of compelling candidate

151   genes with stronger signatures of selection than the canonical loci *tb1* and *tga1* [12].

152　Comparison of polymorphism data in maize landraces and teosinte confirmed a

153　substantial loss (17%) of diversity during the domestication bottleneck, but a

154　genome-wide excess of rare alleles provided evidence that diversity has begun to

155　recover. A slower recovery was found within genes, likely due to the effects of

156　purifying selection. Several genomic regions in maize landraces showed signs of

157　introgression from the highland teosinte *mexicana*, suggesting an important role for

158　gene flow during post-domestication maize diffusion. The magnitude and level of

159　variation of gene expression were also shown to have changed significantly during

160　domestication [12] with substantial re-wiring of coexpression networks in the

161　maize transcriptome relative to *parviglumis* [14]. The maize HapMap2 project also

162　revealed large-scale, structural changes across the genomes of *Zea* (Box 1).

163　Pervasive structural variants (SVs) were identified as regions showing differences in

164　the presence or copy number of segments of the genome. Finally, genome size

165　differences within the species *Z. mays* were found to be caused by segregation of

166　heterochromatic regions known as chromosomal knobs [8], whereas between-

167　species differences (*Z. mays* compared to *luxurians* [41] and the closely-related

168　taxon *Tripsacum dactyloides* [8]) have been shown to be due largely to transposable

169　element content.

170

171　*Introgression Between Teosinte and Maize*

172

173　　　Subsequent to domestication, teosinte populations have persisted in both

174　allopatry and sympatry with maize. Hybridization between maize and teosinte is

8

175 well documented where they co-occur [42-44], but genetic barriers to hybridization

176 are also well known. At least three loci are known to act as pre-zygotic crossing

177 barriers between maize and teosinte: *Teosinte crossing barrier 1* (*Tcb1*) and

178 gametophyte factors *(Ga) 1* and *2* [45-47]. The biochemical mechanism underlying

179 the action of these loci remains unclear, but alleles conferring incompatibility at *Ga1*

180 and *Tcb1* appear to be present at higher frequency in *mexicana* than in *parviglumis*

181 populations [47].

182   Hybridization barriers may dramatically limit the rate of gene flow, yet

183 theory predicts that if certain combinations of genotypes show increased fitness,

184 introgression can occur despite barriers to gene exchange [48]. Recent evidence

185 suggests reciprocal introgression between the partially incompatible subspecies

186 *mexicana* and maize may be a compelling example of such gene flow. Following its

187 domestication in the lowlands of southwest Mexico from *parviglumis*, maize diffused

188 to the highlands of the Mexican Central Plateau where it encountered dramatically

189 different environmental conditions and came into sympatry with the highland

190 teosinte *mexicana*. Modern highland maize and *mexicana* share multiple phenotypic

191 traits (*e.g.,* abundant macrohairs and leaf pigmentation; [44]). This shared

192 morphology has been attributed both to introgression from maize into *mexicana*

193 conferring crop mimicry [49] and to *mexicana* introgression into maize conferring

194 highland adaptation under the assumption that shared traits are beneficial under

195 highland conditions [50]. The application of molecular markers has revealed

196 considerable evidence for gene flow in both directions, suggesting as much as 20%

197 of highland maize genomes are derived from *mexicana* and approximately 10% of

198 *mexicana* genomes are derived from maize [28].

199   Recent work has taken advantage of the availability of genome-wide SNP

200 typing to evaluate the genomic imprint of introgression in sympatric populations of

201 highland Mexican maize and *mexicana* [51]. Numerous genomic regions of

202 introgression were identified in both directions of gene flow yet domestication loci

203 appeared particularly resistant to introgression. Interestingly, regions of *mexicana*-

204 to-maize introgression were shared across populations suggesting either ancient

205 gene flow or multiple independent introgression events at the same loci. These

206 shared regions of introgression from *mexicana* into maize were found to colocalize

207 with previously identified QTL for highland traits [52], suggesting a possible role for

208 *mexicana* in conferring maize adaptation to highland conditions. A growth chamber

209 experiment simulating highland conditions and including maize populations with

210 and without *mexicana* introgression at these shared loci showed significantly higher

211 leaf sheath macrohair counts, extent of pigment, and plant height in populations

212 with a high frequency of *mexicana*-to-maize introgression [51].

213

214 *Local Adaptation in Teosinte*

215

216   Local adaptation occurs when resident genotypes have higher fitness than

217 genotypes from other locations [53]. Currently, there are only a handful of examples

218 in which the genetic basis of local adaptation is well understood. Because of their

219 sessile nature, plants are particularly well suited for studies of local adaptation [19],

10

220   and genome-wide data have recently been analyzed in plant populations across

221   wide geographic ranges [54,55].

222   Among the teosintes, *mexicana* and *parviglumis* are the best candidates for

223   studies of local adaptation. Both have widespread geographic distributions (Figure

224   1A & 1B) consisting of hundreds of distinct populations [2,20,56] that display

225   varying levels of genetic differentiation [2,57], from very fine-scale structure (F_{ST}: 1-

226   2%) within a site [58,59] to more pronounced differentiation between subspecies

227   (F_{ST}: ~15%; [60]). Populations of *parviglumis* and *mexicana* also span a wide

228   climatic range of temperature and precipitation (Table 1). Phenotypically, these

229   subspecies are differentiated based on the smaller spikelets and fruitcases and more

230   open tassels of *parviglumis* in comparison to *mexicana* [16] and by the hairier leaf

231   sheaths and higher pigment content found in *mexicana* [52]. Within subspecies,

232   individual *parviglumis* and *mexicana* populations are known to vary in several

233   morphological and phenological characteristics [56,61,62]. It is not yet clear

234   whether population-level phenotypic differences are due to local adaptation or if

235   they are rather byproducts of genetic drift.

236   Recent studies have investigated the molecular variation underlying local

237   adaptation in teosinte. A study of variation in plant immunity genes across six

238   *parviglumis* populations found little evidence for widespread geographically

239   variable selection but did observe diversity patterns suggesting localized natural

240   selection in the gene *wip1* [63]. *wip1*, a protease inhibitor, is known to interfere

241   with the dietary proteins of insects, and selection at this locus could represent local

242   adaptation to herbivory. A second study of SNP diversity in *mexicana* and

11

243 *parviglumis* identified *Inv1n*, a 50-Mb chromosomal inversion found segregating in

244 33 *parviglumis* populations [64]. The frequency of *Inv1n* was shown to vary along an

245 altitudinal cline (*i.e.,* a continuous gradient) and associate with environmental and

246 morphological characters. Such inversions are predicted to be beneficial when they

247 capture locally adapted alleles or when the breakpoints of the inversions themselves

248 are beneficial [65]. Most recently, a genome-wide analysis of 36,719 SNPs typed in

249 10 *mexicana* and 11 *parviglumis* populations identified candidate loci underlying

250 local adaptation [66]. Contrary to previous findings from *Arabidopsis thaliana* [55],

251 enrichment of non-genic SNPs was found among candidates, in line with the

252 observation that a substantial proportion of phenotypic associations in maize are

253 found in noncoding regions [8]. In addition, several newly identified inversion

254 polymorphisms showed an excess of differentiation among populations and

255 subspecies along with enrichment for association with environmental variables such

256 as temperature and altitude. The complexity and size of the *Zea* genome (Box 1)

257 may provide ample opportunities for such large-scale chromosomal rearrangements

258 to occur. Finally, the study revealed that factors such as hierarchical population

259 structure and uneven distribution of environmental variation can bias detection of

260 locally adaptive loci based on a single statistical method, underscoring the

261 importance of utilizing multiple approaches.

262

263 **Future Prospects**

264

265   The teosintes provide opportunities to examine many compelling questions

266   at the genome level using the vast number of resources developed in maize (Box 2).

267   As the cost of high-throughput sequencing continues to decrease and *de novo*,

268   reference-assisted assembly becomes commonplace in teosinte, genomic studies can

269   move farther beyond maize, delving more deeply into the evolutionary histories of

270   these taxa. For example, the species *luxurians* and *nicaraguensis* are uniquely

271   adapted to the flooded and coastal conditions of Central America. Unlike other

272   teosintes, these taxa develop aerenchyma (air channels that allow gas exchange

273   between submerged and above-water organs of a plant) both in flooded and non-

274   flooded conditions [67], and multiple QTL have already been identified for this trait

275   [68]. Past studies have suggested potential hybridization between these teosintes

276   and local maize landraces [49,50,60], and additional field and genome-wide

277   analyses should assess evidence for adaptive introgression for flood tolerance in

278   landraces at loci encoding these traits.

279   The teosintes also provide an excellent system in which to study the genetics

280   of speciation. The genus appears to have radiated relatively recently, and shared

281   polymorphism and ongoing gene flow are evident across species [60]. Perhaps of

282   greatest interest would be field-based studies of local adaptation (*e.g.*, development

283   of near isogenic lines and reciprocal transplant experiments), incipient speciation,

284   and admixture between *parviglumis* and *mexicana*. These subspecies appear to have

285   diverged only recently, but already exhibit locally adaptive phenotypes, and gene

286   flow between them is limited to a narrow hybrid zone [2,66]. Studies of gene flow

287   and reinforcement within this zone could shed light on the process of divergence in

288   these subspecies.

289   Yet another potential topic of investigation in *Zea* is the recent

290   autopolyploidization (within the last few hundred thousand years; [23]) of *perennis*.

291   Investigation of the more ancient polyploidization event ancestral to all of *Zea* has

292   shown that genome fractionation has been biased toward a single subgenome and is

293   ongoing, distinguishing even modern inbred maize lines [69]. *Perennis* presents an

294   opportunity to investigate incipient fractionation/diploidization in natural

295   populations by utilizing newly developed resources to assess genome-wide patterns

296   of gene expression, DNA methylation, and deletion.

297   Genomic resources could also be brought to bear on the evolution of the

298   perennial life history in *diploperennis* and *perennis*, a feature that distinguishes these

299   species from the remaining annual *Zea* taxa. A previous QTL investigation of

300   *diploperennis* and *parviglumis* did not find large effect loci controlling traits

301   associated with perennialism, and very few genes were linked to the evolution of the

302   perennial growth habit [70]. Full genome resequencing may allow population-

303   genetic-based assessment of selection, but it seems likely that increased marker

304   density would be better exploited by multiparent mapping populations (*e.g.*,

305   [71,72]) of annual/perennial taxa that can take advantage of a large number of

306   recombination events to provide increased resolution. Full transcriptome patterns

307   of gene expression would also be valuable data for identifying loci involved in the

308   contrasting growth habits across *Zea*. Once these full-genome data are generated

309   and we have a better idea about the loci involved in *Zea* perennialism, many

310 fascinating ecological and evolutionary questions can be investigated. For example:

311 Do genes associated with perennialism explain these species' capacity to tolerate

312 conditions outside the normal growing season of the annual *Zea* (*e.g.*, cold-tolerance

313 or drought-resistance loci)? To what degree is there overlap with the pathways and

314 orthologous genes involved in perennialism in closely related taxa such as rice and

315 sorghum?

316   While a number of fascinating evolutionary questions can be addressed

317 within *Zea* on a species-by-species basis, the genus as a whole can also be seen as an

318 ideal study system for investigating broader and potentially far-reaching topics. For

319 instance, *Zea* genomes are more representative than the genomes of most current

320 angiosperm models in that they are comprised primarily of repetitive elements

321 (transposable elements (TE's), microsatellites, centromere and chromosomal knob

322 repeats) [73]. Within *Zea*, the evolution of repetitive elements can be studied over

323 hundreds of thousands of years, revealing epochs of transposition of TE's, the

324 expansion and contraction of knob and centromere repeat arrays, and the structural

325 and functional roles of this largely ignored fraction of the genome. *Zea* species also

326 have a wide range of effective population sizes, allowing for comparisons of the

327 efficiency of natural selection and the effects of rare variants across closely related

328 taxa. Association studies in the wild taxa of *Zea* will also provide greater resolution

329 for the genetic architecture of complex traits than is currently possible in their

330 domesticated congeners due to the increased number of recombination events in

331 natural populations and the rapid decay of linkage disequilibrium [8].

332   Finally, much of the proposed evolutionary genomic work in the teosintes

333   will likely feed back into the improvement of maize (Figure 2). For example,

334   temperature-related highland adaptation loci could be important for maize breeding

335   in the face of climate change, flood-tolerance loci identified in introgression or

336   continued QTL studies of *luxurians* and *nicaraguensis* could help mitigate crop losses

337   due to flooding [74], and genomic studies of *diploperennis* and *perennis* may

338   facilitate the development of perennial maize, a resource that could potentially

339   reduce inputs and prove particularly useful in agroecosystems of the developing

340   world. In this way, the model/non-model gap will be bridged not only through

341   application of maize-derived technologies to teosinte, but also through valuable

342   contributions from the rich evolutionary past of *Zea* to the future of maize.

343

Acknowledgements

353

Glossary

16

355

356   **Allopatry**: describes geographically and reproductively isolated populations

357   **Chromosomal knob**: heterochromatic DNA in *Zea* comprised primarily of two

358   classes of tandem satellite repeats, 180bp and the 350bp TR-1. Knobs in maize are

359   found both in the middle and on the ends of chromosomal arms, and have been

360   associated with meiotic drive

361   **Fine-scale/cryptic structure**: non-random barriers to exchange of genetic material

362   among individuals in close proximity that are difficult to detect based on observable

363   characters

364   **Fractionation**: the mutational process leading to the reduction of gene number

365   following a whole-genome duplication event

366   **Inferred allele frequency**: frequency of alleles in the ancestral population

367   estimated based on knowledge of current genotypes, population structure, and the

368   process of genetic drift

369   **Introgression**: gene flow and persistence of genetic material from one population

370   or species into the gene pool of another

371   **Purifying selection**: a process of natural selection in which deleterious alleles are

372   removed from the gene pool

373   **Sympatry**: describes populations which inhabit overlapping geographic areas

374   allowing mating

375

376   **Boxes**

377

17

378 **Box 1:** The Pan-*Zea* Genome.

379   Early studies in maize were the first to physically map genes onto

380 chromosomes and reveal the phenomenon of crossing over [75]. Subsequent

381 cytological studies showed large-scale structural differences in *Zea* genomes due to

382 varying heterochromatic knob content and chromosomal inversions [T. A. Kato

383 Yamakake, PhD thesis, University of Massachusetts, 1975; 76]. Array- and

384 sequence-based analyses have also demonstrated substantial genetic and structural

385 variation across taxa [8,12,41,77]. However, the recent origin of the genus [60] and

386 gene flow among taxa [2], when combined with evidence for structural variation

387 [78,79], suggest researchers will be able to describe core and dispensable fractions

388 of a Pan-*Zea* genome [80].

389   A large-scale effort is underway to investigate low-copy genic regions in *Zea*

390 species. Currently, the high-confidence filtered gene set of the 2.1-Gb maize genome

391 (maizesequence.org; release 5b.60) includes ~40,000 genes. Microarray data

392 indicated ~10% of genes across *Zea mays* accessions were associated with

393 structural variants [81,82]; however, estimates from next-generation sequencing of

394 103 *Zea* genomes suggested structural variation in as many as 70% of genes and

395 identified significant associations between variants and phenotype [8]. Analyses of

396 syntenic regions between sorghum and maize show that genes with lower

397 expression are more likely to be lost, suggesting a potential mechanism behind

398 observed variation in gene content in inbred lines [69,82].

399   Extensive variation is not limited to the genic fraction of *Zea* genomes.

400 Cytological studies of repetitive elements (knobs, transposable elements (TE's),

401 centromere repeats), which comprise >85% of the genome, have shown dramatic

402 differences across *Zea* [79,83]. Sequence-based analyses, however, indicate that

403 whereas individual families of TE's have changed in abundance in recent history

404 [12], the relative frequencies of most TE's appear similar between maize and its

405 sister species *luxurians* [41]. Sequence analysis also confirmed predictions [84,85]

406 that chromosomal knob content correlates positively with genome size within *Z.*

407 *mays*, whereas transposable elements account for genome size differences between

408 *Z. mays* and *Tripsacum* [8]. Fluorescent *in situ* hybridization has previously shown

409 visible differences in centromere repeat (CentC) abundance between *parviglumis*

410 and maize, and these differences can now be confirmed and quantified using low-

411 coverage, whole-genome sequencing (c.f. [41]).

412   Analysis of shared and non-shared genomic regions across *Zea* individuals

413 has begun to describe the core and dispensable fractions of the Pan-*Zea* genome

414 [80]. The emerging picture is much more fluid than previously expected and

415 challenges the notion of genome conservation even at the intraspecies level. Because

416 of its great diversity, a Pan-*Zea* genome would likely serve as a model for studying

417 genomic evolution.

418

419 **Box 1 Figure Legends**

420 **Box 1, Figure 1**. Genomic content of different classes of structural elements in

421 maize and its relatives. Genome size, represented as a line graph, is plotted on the

422 left axis. Repetitive element abundance was standardized against maize. B73 was

423 used as the single reference line for maize, while *parviglumis* (N=9) and *mexicana*

19

424 (N=2) are averages. A single *Tripsacum dactyloides* individual was also used. Data

425 for figure were obtained from [8].

426

427 **Box 1, Figure 2**. Fluorescence *in situ* hybridization in two closely related *Zea* taxa,

428 *parviglumis* and *mexicana*. Chromosome number is indicated on the upper axis, and

429 homologous chromosomes are shown. Karyotype probes include: 180-bp knob

430 (blue), TR-1 knob (pseudo-colored white), CentC and subtelomere 4-12-1 (green),

431 5S ribosomal gene (yellow), Cent4 (orange), NOR (bluegreen), and TAG

432 microsatellite (1-26-2) and subtelomere 1.1 (red). Regions of overlapping probes

433 exhibit a corresponding color change [83].

434

435 **Box 2:** Genetic tools for studying maize.

436 **Maize reference genome** (maizesequence.org) The version AGPv2 assembly of

437 cultivar B73 consists of 2.1×10^9 bases, defines 39,656 genes and reports 63,540

438 transcripts with high confidence. The genome sequence facilitates both full-genome

439 resequencing of teosinte lines and annotation, which is essential for evaluating the

440 functional consequences of observed genetic variation. The reference genome can

441 also be used in expression analyses through read mapping of sequenced RNA

442 **Genotyping methods** The Illumina MaizeSNP50 single nucleotide polymorphism

443 (SNP) array [86] and Genotyping by Sequencing (GBS) [87] can be used for high

444 throughput genotyping in teosinte.

445 **Zinc-finger nuclease genome modification** [88] facilitates precise addition and

446 disruption of target nucleotides and could be used to transfer, for example, teosinte

447 alleles into maize genomes.

448

449 **Databases and Resources:**

450 **MaizeGDB** (maizegdb.org): incorporates maize genetic data from various public

451 resources in a comprehensive database and genome browser.

452 **Panzea** (panzea.org): numerous databases of genotypic and phenotypic data for

453 association mapping and evolutionary analyses.

454 **Genomaize** (genomaize.org/): hosts multiple tools for maize biologists, including a

455 genome browser for nucleosome occupancy and a web tool for *in silico* PCR.

456 *Ac/Ds* **tagging and** *Mu* **insertion databases** (plantgdb.org/prj/AcDsTagging/)

457 (teosinte.uoregon.edu/mu-illumina/): these resources list loci with potentially

458 disruptive transposon insertions and provide seeds of the corresponding maize

459 lines [89,90].

460 **The maize transposable element (TE) database** (maizetedb.org): describes

461 hundreds of maize TE families and has been used to estimate abundance of

462 individual TE families in maize and related taxa [8,41].

463 **Maize transcription atlas** (plexdb.org): describes gene expression in 60 tissues

464 and developmental stages for tens of thousands of genes in the reference cultivar

465 B73 [13].

466 **MaizeCyc** (pathway.gramene.org/gramene/maizecyc.shtml): catalogs metabolic

467 and transport pathways of maize.

21

468   **QTeller** (qteller.com): a tool to summarize function, expression, and synteny

469   information from maize genes in a given genomic interval.

470   **CoGe** (genomevolution.org): tools for plant comparative genomics, including well-

471   known classical genes in maize, and sets of maize orthologs syntenic with other

472   grasses

473   **Maize Stock Center** (maizecoop.cropsci.uiuc.edu): collection of maize genetic

474   stocks, including information on known mutants

475

476

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. *P. Natl. Acad. Sci. USA* 109, 11878-11883

15 Matsuoka, Y. *et al.* (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *P. Natl. Acad. Sci. USA* 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*

521   *for transgenic maize* (Serratos, J.A., Willcox, M.C., and Castillo-González, F., eds),
522   pp. 18-36, CIMMYT

523 21 Vigouroux, Y. *et al*. (2008) Population structure and genetic diversity of new world
524   maize races assessed by DNA microsatellites. *Am*. *J*. *Bot*. 95, 1240–1253

525 22 Gaut, B.S. and Doebley, J.F. (1997) DNA sequence evidence for the segmental
526   allotetraploid origin of maize. *P*. *Natl*. *Acad*. *Sci*. *USA* 94, 6809–6814

527 23 Tiffin, P. and Gaut, B.S. (2001) Sequence diversity in the tetraploid *Zea perennis* and
528   the closely related diploid *Z*. *diploperennis*: Insights from four nuclear loci.
529   *Genetics* 158, 401–412

530 24 Bennetzen, J. *et al*. (2001) Genetic evidence and the origin of maize (Biology,
531   archaeology). *Lat*. *Am*. *Antiq*. 12, 84–86

532 25 Doebley, J. (2001) George Beadle's other hypothesis: One-gene, one-trait. *Genetics*
533   158, 487–493

534 26 Doebley, J. (1990) Molecular evidence and the evolution of maize. *Econ*. *Bot*. 44, 6–
535   27

536 27 Doebley, J. *et al*. (1987) Patterns of isozyme variation between maize and Mexican
537   annual teosinte. *Econ*. *Bot*. 41, 234–246

538 28 van Heerwaarden, J. *et al*. (2011) Genetic signals of origin, spread, and introgression
539   in a large sample of maize landraces. *P*. *Natl*. *Acad*. *Sci*. *USA* 108, 1088–1092

540 29 Doebley, J. *et al*. (1997) The evolution of apical dominance in maize. *Nature* 386,
541   485–488

542 30 Studer, A. *et al*. (2011) Identification of a functional transposon insertion in the maize
543   domestication gene *tb1*. *Nat*. *Genet*. 43, 1160–U164

544 31 Gallavotti, A. *et al*. (2004) The role of *barren stalk1* in the architecture of maize.
545   *Nature* 432, 630–635

546 32 Sigmon, B. and Vollbrecht, E. (2010) Evidence of selection at the *ramosa1* locus
547   during maize domestication. *Mol*. *Ecol*. 19, 1296–1311

548 33 Wang, H. *et al*. (2005) The origin of the naked grains of maize. *Nature* 436, 714–719

549 34 Briggs, W.H. *et al*. (2007) Linkage mapping of domestication loci in a large maize-
550   teosinte backcross resource. *Genetics* 177, 1915–1928

551 35 Buckler, E.S. *et al*. (2009) The genetic architecture of maize flowering time. *Science*
552   325, 714–718

553 36 Tenaillon, M.I. *et al*. (2004) Selection versus demography: A multilocus investigation
554   of the domestication process in maize. *Mol*. *Biol*. *Evol*. 21, 1214–1225

555 37 Vigouroux, Y. *et al*. (2002) Identifying genes of agronomic importance in maize by
556   screening microsatellites for evidence of selection during domestication. *P*. *Natl*.
557   *Acad*. *Sci*. *USA* 99, 9650–9655

558 38 Wright, S.I. *et al*. (2005) The effects of artificial selection of the maize genome.
559   *Science* 308, 1310–1314

560 39 Yamasaki, M. *et al*. (2005) A large-scale screen for artificial selection in maize
561   identifies candidate agronomic loci for domestication and crop improvement. *Plant*
562   *Cell* 17, 2859–2872

563 40 Ross-Ibarra, J. *et al*. (2007) Plant domestication, a unique opportunity to identify the
564   genetic basis of adaptation. *P*. *Natl*. *Acad*. *Sci*. *USA* 104, 8641–8648

565 41 Tenaillon, M.I. *et al*. (2011) Genome size and transposable element content as
566   determined by high-throughput sequencing in maize and *Zea luxurians*. *Genome*
567   *Biol*. *Evol*. 3, 219–229
568 42 Baltazar, B.M. *et al*. (2005) Pollination between maize and teosinte: an important
569   determinant of gene flow in Mexico. *Theor*. *Appl*. *Genet*. 110, 519–526
570 43 Ellstrand, N.C. *et al*. (2007) Spontaneous hybridization between maize and teosinte. *J*.
571   *Hered*. 98, 183–187
572 44 Wilkes, H.G. (1967) Teosinte: The Closest Relative of Maize. The Bussey Institute of
573   Harvard University
574 45 Evans, M.M.S. and Kermicle, J.L. (2001) *Teosinte crossing barrier1*, a locus
575   governing hybridization of teosinte with maize. *Theor*. *Appl*. *Genet*. 103, 259–265
576 46 Kermicle, J.L. and Evans, M.M.S. (2010) The *Zea mays* sexual compatibility gene
577   *ga2*: naturally occurring alleles, their distribution, and role in reproductive isolation.
578   *J*. *Hered*. 101, 737–749
579 47 Kermicle, J.L. *et al*. (2006) The *gametophyte-1* locus and reproductive isolation
580   among *Zea mays* subspecies. *Maydica* 51, 219–225
581 48 Barton, N.H. (2001) The role of hybridization in evolution. *Mol*. *Ecol*. 10, 551–568
582 49 Wilkes, H.G. (1977) Hybridization of maize and teosinte, in Mexico and Guatemala
583   and improvement of maize. *Econ*. *Bot*. 31, 254–293
584 50 Doebley, J. (1990) Molecular evidence for gene flow among *Zea* species - genes
585   transformed into maize through genetic-engineering could be transferred to its wild
586   relatives, the teosintes. *Bioscience* 40, 443–448
587 51 Hufford, M.B. *et al*. (2012) The genomic signature of crop-wild introgression in maize.
588   arXiv:1208.3894v2 URL: http://arxiv.org/abs/1208.3894
589 52 Lauter, N. *et al*. (2004) The inheritance and evolution of leaf pigmentation and
590   pubescence in teosinte. *Genetics* 167, 1949
591 53 Kawecki, T.J. and Ebert, D. (2004) Conceptual issues in local adaptation. *Ecol*. *Lett*. 7,
592   1225–1241
593 54 Fournier-Level, A. *et al*. (2011) A map of local adaptation in *Arabidopsis thaliana*.
594   *Science* 334, 86–89
595 55 Hancock, A.M. *et al*. (2011) Adaptation to climate across the Arabidopsis thaliana
596   genome. *Science* 334, 83–86
597 56 Weber, A. *et al*. (2007) Major regulatory genes in maize contribute to standing
598   variation in teosinte (*Zea mays* ssp. *parviglumis*). *Genetics* 177, 2349
599 57 Moeller, D.A. *et al*. (2007) Population structure and its effects on patterns of
600   nucleotide polymorphism in teosinte (*Zea mays* ssp. *parviglumis*). *Genetics* 176,
601   1799
602 58 Hufford, M.B. *et al*. (2011) Influence of cryptic population structure on observed
603   mating patterns in the wild progenitor of maize (*Zea mays* ssp. *parviglumis*). *Mol*.
604   *Ecol*. 20, 46–55
605 59 van Heerwaarden, J. *et al*. (2010) Fine scale genetic structure in the wild ancestor of
606   maize (*Zea mays* ssp. *parviglumis*). *Mol*. *Ecol*. 19, 1162–1173
607 60 Ross-Ibarra, J. *et al*. (2009) Historical divergence and gene flow in the genus *Zea*.
608   *Genetics* 181, 1399
609 61 Rodriguez, F.J.G. *et al*. (2006) Characterization of floral morphology and synchrony
610   among *Zea* species in Mexico. *Maydica* 51, 383–398

611 62 Weber, A.L. *et al*. (2008) The genetic architecture of complex traits in teosinte (*Zea*
612   *mays* ssp. *parviglumis*): new evidence from association mapping. *Genetics* 180,
613   1221

614 63 Moeller, D.A. and Tiffin, P. (2008) Geographic variation in adaptation at the
615   molecular level: a case study of plant immunity genes. *Evolution* 62, 3069–3081

616 64 Fang, Z. *et al*. Megabase-scale inversion polymorphism in the wild ancestor of maize.
617   *Genetics* 191, 883-894

618 65 Kirkpatrick, M. and Barton, N. (2006) Chromosome inversions, local adaptation and
619   speciation. *Genetics* 173, 419–434

620 66 Pyhäjärvi, T. *et al*. (2012) Complex patterns of local adaptation in teosinte.
621   arXiv:1208.0634v1 URL: http://arxiv.org/abs/1208.0634

622 67 Mano, Y. *et al*. (2006) Variation for root aerenchyma formation in flooded and non-
623   flooded maize and teosinte seedlings. *Plant Soil* 281, 269–279

624 68 Mano, Y. *et al*. (2008) A linkage map of maize x teosinte Zea luxurians and
625   identification of QTLs controlling root aerenchyma formation. *Mol*. *Breeding* 21,
626   327–337

627 69 Schnable, J.C. *et al*. (2011) Differentiation of the maize subgenomes by genome
628   dominance and both ancient and ongoing gene loss. *P*. *Natl*. *Acad*. *Sci*. *USA* 108,
629   4069–4074

630 70 Westerbergh, A. and Doebley, J. (2004) Quantitative trait loci controlling phenotypes
631   related to the perennial versus annual habit in wild relatives of maize. *Theor*. *Appl*.
632   *Genet*. 109, 1544–1553

633 71 Churchill, G. *et al*. (2004) The Collaborative Cross, a community resource for the
634   genetic analysis of complex traits. *Nat*. *Genet*. 36, 1133-1137

635 72 Macdonald, S.J. and Long, A.D. (2007) Joint estimates of quantitative trait locus
636   effect and frequency using synthetic recombinant populations of *Drosophila*
637   *melanogaster*. *Genetics* 176, 1261-1281

638 73 Tenaillon, M.I. *et al*. (2010) A triptych of the evolution of plant transposable elements.
639   *Trends Plant Sci*. 15, 471–478

640 74 Mano, Y. and Omori, F. Breeding for flooding tolerant maize using "teosinte" as a
641   germplasm resource. *Plant Root* 1, 17–21

642 75 Creighton, H.B. and McClintock, B. (1931) A correlation of cytological and genetical
643   crossing-over in *Zea mays*. *P*. *Natl*. *Acad*. *Sci*. *USA* 17, 492–497

644 76 Ting, Y.C. (1967) Common inversion in maize and teosinte. *Am*. *Nat*. 101, 87–89

645 77 Springer, N.M. *et al*. (2009) Maize inbreds exhibit high levels of copy number
646   variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS*
647   *Genet*. 5, e1000734

648 78 Brunner, S. *et al*. (2005) Evolution of DNA sequence nonhomologies among maize
649   inbreds. *Plant Cell* 17, 343–360

650 79 Lamb, J.C. and Birchler, J.A. (2006) Retroelement genome painting: Cytological
651   visualization of retroelement expansions in the genera *Zea* and *Tripsacum*. *Genetics*
652   173, 1007–1021

653 80 Morgante, M. *et al*. (2007) Transposable elements and the plant pan-genomes. *Curr*.
654   *Opin*. *Plant Biol*. 10, 149–155

655 81 Eichten, S.R. *et al*. (2011) B73-Mo17 near-isogenic lines demonstrate dispersed
656   structural variation in maize. *Plant Physiol*. 156, 1679–1690

657 82 Swanson-Wagner, R.A. *et al.* (2010) Pervasive gene content variation and copy
658   number variation in maize and its undomesticated progenitor. *Genome Res.* 20,
659   1689–1699
660 83 Albert, P.S. *et al.* (2010) Diversity of chromosomal karyotypes in maize and its
661   relatives. *Cytogenet. Genome Res.* 129, 6–16
662 84 Laurie, D.A. and Bennett, M.D. (1985) Nuclear-DNA content in the genera *Zea* and
663   *Sorghum* - intergeneric, interspecific and intraspecific variation. *Heredity* 55, 307–
664   313
665 85 Poggio, L. *et al.* (1998) Genome size and environmental correlations in maize (*Zea*
666   *mays* ssp. *mays, Poaceae*). *Ann. Bot.-London* 82, 107–115
667 86 Ganal, M.W. *et al.* (2011) A large maize (*Zea mays* L.) SNP genotyping array:
668   development and germplasm genotyping, and genetic mapping to compare with the
669   B73 reference genome. *PLoS One* 6, e28334
670 87 Elshire, R.J. *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach
671   for high diversity species. *PLoS One* 6, e19379
672 88 Shukla, V.K. *et al.* (2009) Precise genome modification in the crop species *Zea mays*
673   using zinc-finger nucleases. *Nature* 459, 437–U156
674 89 Vollbrecht, E. *et al.* (2010) Genome-wide distribution of transposed dissociation
675   elements in maize. *Plant Cell* 22, 1667–1685
676 90 Williams-Carrier, R. *et al.* (2010) Use of Illumina sequencing to identify transposon
677   insertions underlying mutant phenotypes in high-copy Mutator lines of maize. *Plant*
678   *J.* 63, 167–177

679 **Figure Legends:**

680

681 **Figure 1:** Ecological and Taxonomic Diversity of *Zea*

682 The wild *Zea* are distributed across a wide climatic range of temperature (a) and

683 precipitation (b). Within-taxon and between-taxa variation has also been observed

684 in traits such as flowering time (summarized based on data from [61]) (c). The

685 phylogeny of *Zea* (d) based on ∼1000 SNPs from across the genome (data are from

686 [64]).

687

688 **Figure 2:** Discovery process from population genomics in teosinte to crop

689 improvement in maize. Natural variation among wild teosinte populations can be

690 used to identify genetic variation that correlates with environmental variables.

691 Further, common garden experiments, maize mutation databases, and tissue specific

692 expression data can be used to identify genes potentially underlying adaptation to

693 environment. Finally, causative variants can be transformed into maize to verify

694 phenotypic effects.

695

Table 1

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=10	4	2.73	1141 ± 802	18.04 ± 5.93	1328 ± 1106	Annual	-
Zea mays ssp. *parviglumis*	2N=2X=10	8.375	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=10	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=10	11.5	N/A	2093 ± 275	21.22 ± 1.06	1181 ± 0.3	Annual	Capable of hybridization with maize
Zea diploperennis	2N=2X=10	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=20	N/A	5.28	1009 ± 236	16.82 ± 1.98	1946 ± 329	Perennial	No evidence of successful hybridization
Zea luxurians	2N=2X=10	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=10	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 reported from [83]. For all other species, mean knobs were calculated from [44].
b) Genome sizes are reported from the Kew (http://data.kew.org) Plant DNA *C*-values database (*mays, diploperennis, perennis, luxurians*) and [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 within [2,42-44,49,50]

Figure 1

(a)

Legend

species/subspecies
- ✕ diploperennis
- ★ perennis
- ✚ huehuetenangensis
- ■ luxurians
- △ mexicana
- ◆ nicaraguensis
- ○ parviglumis

Annual Temperature (°C)
- <15
- 15 - 17.5
- 17.5 - 20
- 20 - 22.5
- 22.5 - 25
- 25 - 27.5
- 27.5 - 30
- >30

N   0 125 250   500 Kilometers

(b)

Legend

species/subspecies
- ✕ diploperennis
- ★ perennis
- ✚ huehuetenangensis
- ■ luxurians
- △ mexicana
- ◆ nicaraguensis
- ○ parviglumis

Annual Precipitation (mm)
- 0 - 200
- 200 - 400
- 400 - 600
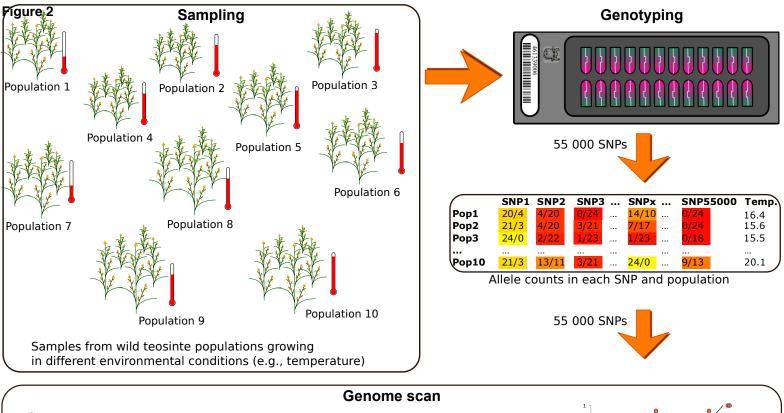- 600 - 800
- 800 - 1,000
- 1,000 - 1,200
- 1,200 - 1,400
- >1,400

N   0 125 250   500 Kilometers

(c)

Days to Pollen
- mexicana
- parviglumis
- landraces

Days to Silk
- mexicana
- parviglumis
- landraces

(d)

- *Tripsacum dactyloides*
- *Zea nicaraguensis*
- *Zea luxurians*
- *Zea diploperennis*
- *Zea perennis*
- *Zea mays* ssp. *huehuetenangensis*
- *Zea mays* ssp. *mexicana*
- *Zea mays* ssp. *parviglumis*
- *Zea mays* ssp. *mays*

Figure 2

Sampling

Population 1
Population 2
Population 3
Population 4
Population 5
Population 6
Population 7
Population 8
Population 9
Population 10

Samples from wild teosinte populations growing
in different environmental conditions (e.g., temperature)

Genotyping

55 000 SNPs

	SNP1	SNP2	SNP3	...	SNPx	...	SNP55000	Temp.
Pop1	20/4	4/20	0/24	...	14/10	...	0/24	16.4
Pop2	21/3	4/20	3/21	...	7/17	...	0/24	15.6
Pop3	24/0	2/22	1/23	...	1/23	...	0/18	15.5
...	...	...	...		...		...	...
Pop10	21/3	13/11	3/21	...	24/0	...	9/13	20.1

Allele counts in each SNP and population

55 000 SNPs

Genome scan

SNPs correlating with temperature along teosinte genome (black circles)

60 candidate SNPs that associate with
an environmental variable

Information from databases and previous studies

Gene X   Gene X

Knock-out mutation has a phenotypic
effect in maize

Interesting tissue specific expression under
heat stress in maize (e.g., in leaves)

**Verifying the phenotypic
effect in maize**

Transformation or designed
zinc-finger nucleases

Maize with maize variant of gene X

Common garden experiments

AG   AG   AA   AA   AA   GG   GG   AA

Natural genetic variation in gene X has a phenotypic
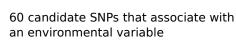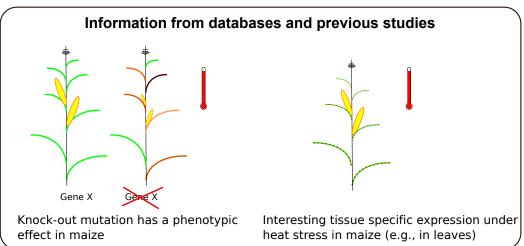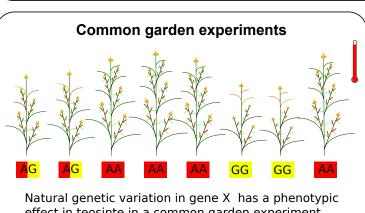effect in teosinte in a common garden experiment

10 SNPs that have strong
potential to have desired
phenotypic effect in maize

More heat tolerant maize with
teosinte variant of gene X

1 2 3 4 5 6 7 8 9 10

Z. mays ssp. *parviglumis*

ssp. *mexicana*