A crop of maize variants

Xuehui Huang & Bin Han

Three new studies report large-scale resequencing and comparative genomic analysis of diverse maize varieties. The authors conducted a comprehensive characterization of sequence variation in maize genomes and identified signals of selection in maize domestication and breeding.

Maize (*Zea mays* ssp. *mays*) is one of the most important cereals in the world. As an outcrossing species, maize has a high level of sequence diversity and shows rapid decline of linkage disequilibrium. Genetic diversity and agronomically favorable alleles in maize germplasm have long been exploited by breeders for grain yield improvements. Comprehensive characterization of the extensive genomic variation in diverse maize varieties would be widely useful for maize genetics research and breeding. However, this endeavor has been technically challenging due to the complexity of the large maize genome, which is 2.3 Gb in total size and is ~85% repetitive sequences[1]. In this issue, Doreen Ware and colleagues report large-scale whole-genome resequencing of *Z. mays*, including the genomes of 60 modern cultivars, 23 landraces and 19 wild relatives, with each sequenced at ~4.2× coverage[2]. Using a novel population genetics scoring model, the authors constructed a second-generation haplotype map of the maize genome (Maize HapMap2), which is more comprehensive than the previous map[3]. Jeffrey Ross-Ibarra and colleagues carried out further population genetics analyses on a subset of these data to evaluate the evidence for selection during the initial domestication and subsequent improvement of maize landraces[4]. Another study by Jinsheng Lai and colleagues reports the resequencing of 278 modern temperate maize cultivars, each to twofold coverage, which was used to assess artificial selection during modern maize breeding[5].

Xuehui Huang and Bin Han are at the National Center for Gene Research, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China.
e-mail: xhhuang@ncgr.ac.cn or bhan@ncgr.ac.cn

Characterization of complex variation

Chia *et al.*[2] generated 13 billion reads across the 103 inbred maize lines and initially used two complementary algorithms for variant calling. Although these two algorithms are widely used for human data, Chia *et al.*[2] found that they did not yield consistent results in maize, which the authors attributed to paralogy in the maize genome. A population genetics–based quality control pipeline, which takes into account patterns of linkage disequilibrium and genotype segregation, was then applied to filter the uncertain calls, although many of the discarded calls might represent real variants. In addition to SNPs and small insertion and deletions (indels), there are a large number of structural variants in the maize genomes, including causal polymorphisms for phenotypic variations[6]. The authors assessed the usefulness of the newly generated data by combining it with data from maize HapMap1 SNPs in genome-wide association studies (GWAS) in the nested association mapping (NAM) population[7] for five leaf-development and disease-resistance traits. In many instances, they found stronger association at HapMap2 SNPs relative to SNPs from the previous HapMap version, with the HapMap2 SNPs accounting for most of the significantly associated loci[8,9].

Despite these successes, characterization of variation across the genome in maize from short-read sequencing is still a big challenge, as maize is an ancient tetraploid with extensive paralogy. Multiple reference sequences from a whole-genome *de novo* assembly[10] would be desirable in maize, as many fragments (including those containing protein-coding genes) in diverse varieties are not shared with the reference B73 genome. For ongoing efforts to sequence more complex crop species (for example, barley and wheat), the pioneering work in maize will undoubtedly provide guiding examples and benchmarks.

Screening of selective sweeps

Selection by humans on crops (including rice, wheat and maize) resulted in selective sweeps around the genes that were the targets of selection during domestication and improvement[11,12]. The polymorphism data generated by Chia *et al.*[2] and Jiao *et al.*[5] have enabled genome-wide detection of these sweeps in maize (**Fig. 1**). Hufford *et al.*[4] identified 484 domestication and 695 improvement loci by using a likelihood method and suggest that there was stronger selection during domestication than in improvement. They also conclude that improvement loci have a smaller average size and contain fewer genes than domestication loci[4].

The sampling of Jiao *et al.*[5] of 278 temperate maize inbred lines that were bred during the past few decades is quite diverse. The authors assessed the overall quality of their genotype data by performing GWAS of three traits (cob color, silk color and date to anthesis) after SNP imputation. The GWAS results showed significant *P* values near loci known to be associated with these traits, although a better-powered GWAS remains to be performed. In addition to their detection of selective sweeps during improvement, Jiao *et al.*[5] also report the results from deep sequencing (27× on average, including the data previously reported[13]) of four inbreeding maize lines with known pedigree information, which they used to assess the rate of genetic change per generation that has occurred during breeding.

Together, the sequencing data and accompanying analyses offer important resources

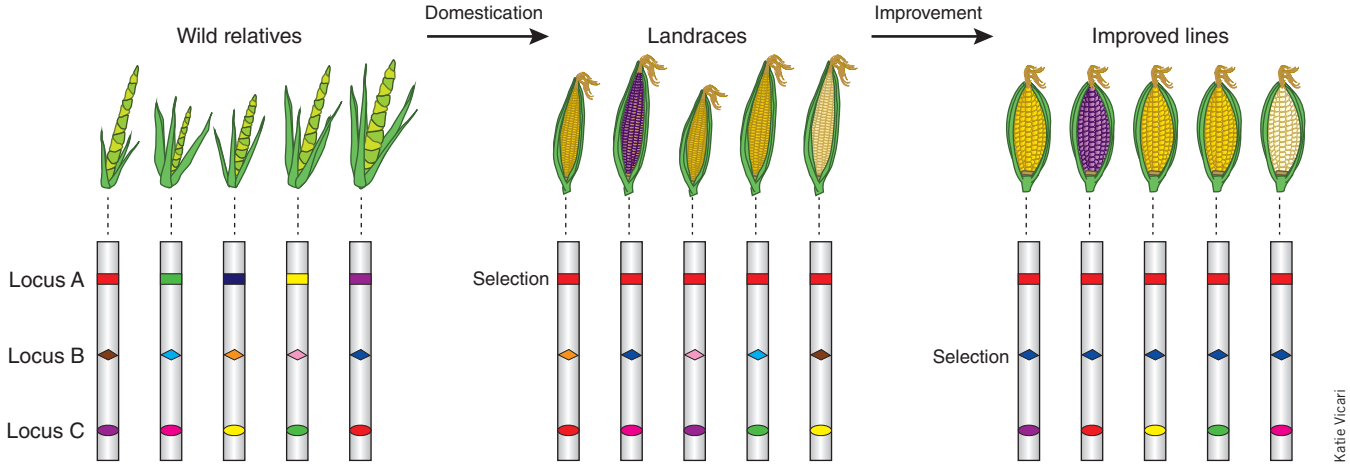

Figure 1 Whole-genome screening of selection signatures during maize domestication and subsequent improvement. Evidence for selection across the genome during both domestication and improvement was evaluated[4,5] in landraces versus wild relatives for domestication and in improved lines versus landraces for improvement, using the whole-genome polymorphism data generated by the large-scale resequencing of maize varieties[2,5]. The selected genomic loci are expected to have a much higher allele frequency in landraces than in wild relatives (for example, a selective sweep in maize domestication; locus A) or in modern cultivars than in landraces (for example, a selective sweep in maize improvement; locus B), whereas other loci are expected to be nearly neutral in different panels (for example, a random genomic region without selection; locus C).

for association mapping, genomic selection and evolutionary studies in maize. It is anticipated that these three studies will facilitate the genotyping of larger populations of maize germplasm for GWAS and will accelerate the identification of the key causal genes that were modified during domestication and improvement in maize.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Schnable, P.S. *et al. Science* **326**, 1112–1115 (2009).
2. Chia, J.-M. *et al. Nat. Genet.* **44**, 803–807 (2012).
3. Gore, M.A. *et al. Science* **326**, 1115–1117 (2009).
4. Hufford, M.B. *et al. Nat. Genet.* **44**, 808–811 (2012).
5. Jiao, Y. *et al. Nat. Genet.* **44**, 812–815 (2012).
6. Studer, A., Zhao, Q., Ross-Ibarra, J. & Doebley, J. *Nat. Genet.* **43**, 1160–1163 (2011).
7. McMullen, M.D. *et al. Science* **325**, 737–740 (2009).
8. Tian, F. *et al. Nat. Genet.* **43**, 159–162 (2011).
9. Kump, K.L. *et al. Nat. Genet.* **43**, 163–168 (2011).
10. Gan, X. *et al. Nature* **477**, 419–423 (2011).
11. Doebley, J.F., Gaut, B.S. & Smith, B.D. *Cell* **127**, 1309–1321 (2006).
12. Takeda, S. & Matsuoda, M. *Nat. Rev. Genet.* **9**, 444–457 (2008).
13. Lai, J. *et al. Nat. Genet.* **42**, 1027–1030 (2010).

Tuning gene expression with nucleosome-disfavoring sequences

Timothy Palpant & Jason Lieb

A new study shows that alteration of poly(dA:dT) tracts in promoters offers a broadly applicable genetic mechanism for predictably tuning gene expression with high resolution. By systematically manipulating these tracts in a controlled yeast system, the authors demonstrate quantitative mechanistic relationships linking regulatory DNA sequences, nucleosome occupancy, transcription factor binding and gene expression.

In eukaryotic cells, transcription factor binding is modulated by nucleosomes that restrict access to potential binding sites. Altering local nucleosome organization in regulatory regions is therefore one way to vary transcription factor binding and gene expression. On page 743 of this issue, Eran Segal and colleagues[1] show that varying the characteristics of nucleosome-disfavoring poly(dA:dT) tracts in a controlled yeast system can be employed to specify nucleosome occupancy and tune gene expression at a specific promoter. Their results advance the quantitative understanding of the regulatory code and suggest a technique for engineering promoters with precisely specified gene expression.

Timothy Palpant and Jason Lieb are at the Department of Biology at The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.
e-mail: jlieb@bio.unc.edu

Tunable components

Transcription factors regulate gene expression by binding to short consensus sequences in the regulatory regions of their target genes. However, transcription factors bind to only a small subset of potential sites *in vivo*, and nucleosome-mediated DNA accessibility is a strong determinant of binding site selection[2,3]. Although many factors contribute to nucleosome organization, intrinsic sequence preferences are a major component that influences nucleosome occupancy *in vivo*[4]. In particular, poly(dA:dT) tracts are nucleosome disfavoring and constitute a discrete element for preventing nucleosome formation near transcription factor–binding sites[5,6]. Poly(dA:dT) tracts are found ubiquitously in eukaryotic promoters[7], yet a quantitative understanding of their effects on nucleosome