ZEAMAP, a comprehensive database adapted to the maize multi-omics era

Songtao Gui[1], Linfeng Yang[1], Jianbo Li[2], Jingyun Luo[1], Xiaokai Xu[2], Jianyu Yuan[2], Lu Chen[1], Wenqiang Li[1], Xin Yang[2], Shenshen Wu[1], Shuyan Li[1], Yuebin Wang[1], Yabing Zhu[2], Qiang Gao[2], Ning Yang[1*], Jianbing Yan[1*]

[1]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China.
[2]Beijing Genomics Institute-Shenzhen, Shenzhen 518083, China.

*To whom correspondence should be addressed: yjianbing@mail.hzau.edu.cn and yangningyingji@126.com

Abstract

As one of the most extensively cultivated crops, maize (*Zea mays* L.) has been extensively studied by researchers and breeders for over a century. With advances in high-throughput detection of various omics data, a wealth of multi-dimensional and multi-omics information has been accumulated for maize and its wild relative, teosinte. Integration of this information has the potential to accelerate genetic research and generate improvements in maize agronomic traits. To this end, we constructed ZEAMAP (http://www.zeamap.com), a comprehensive database incorporating multiple reference genomes, annotations, comparative genomics, transcriptomes, open chromatin regions, chromatin interactions, high-quality genetic variants, phenotypes, metabolomics, genetic maps, genetic mapping loci, population structures and domestication selection signals between teosinte and maize. ZEAMAP is user-friendly, with the ability to interactively integrate, visualize and cross-reference multiple different omics datasets.

Introduction

Maize (*Zea mays* L.) is one of the most important crops for food, feed and fuel, and is also a model species for genetic and genomic research. As the cost of sequencing has been decreased and new omics technologies have arisen, there has been explosive growth in the amount of biological information available for maize. The maize B73 reference genome has recently been updated (1), and four more maize genome assemblies have been released during the last two years (2-5). The previous two-dimensional genome has recently been resolved in three dimensions with the mapping of open chromatin and the identification of chromatin interactions based on ChiA-PET and Hi-C technologies (6,7). Omics data, including deep DNA resequencing, transcriptome and metabolome, have been accumulated at the population scale (8-15). There are many different applications for these new data sets, including gene cloning and the study of regulatory networks. These new and comprehensive data sets provide valuable resources for maize research, and have the potential to completely revolutionize breeding (16).

Comprehensive databases are needed to store, maintain and analyze the multi-omics data which is now available for maize. Several maize genomics and functional genomics databases have been developed, including the Maize Genetics and Genomics Database (MaizeGDB) (https://www.maizegdb.org/), which collects maize reference sequences, stocks, phenotypic and genotypic data and also provides useful tools for maize data mining (17,18). Panzea (https://www.panzea.org/) collects genotypic and phenotypic information for several maize populations (19). while MaizeNet (http://www.inetbio.org/maizenet/) provides a genome-scale co-functional network of maize genes (20). Other generic databases such as Genbank (https://www.ncbi.nlm.nih.gov/genbank/), Gramene (http://www.gramene.org/) and ePlant (http://bar.utoronto.ca/eplant_maize/) also collect maize omics data. Despite being very useful, these databases are designed either to collect general maize genomic and genetic information or to focus on one specific omics area. To make the best use of the multi-omics information for maize research and breeding, researchers currently need to either systematically integrate omics data generated from different sources (21), or use multi-omics data which were all generated from the same panel.

MODEM (http://modem.hzau.edu.cn/) is the first attempt to integrate multi-omics datasets, including various types of genetic variants, expression data and metabolomic data (22). Despite the importance of wild relatives in understanding the domestication of modern crops, no existing maize multi-omics databases incorporate teosinte (23). To fill this gap, we have developed ZEAMAP (http://www.zeamap.com/), a multi-omics database for maize research and breeding, which integrates omics data generated from 527 elite inbred lines (in an association mapping panel, AMP) and 183 teosinte accessions (24). ZEAMAP includes genome assemblies and annotations of four inbred lines, B73 (1), Mo17 (3), SK (4) and HuangZaoSi (HZS) (5), and a teosinte accession (*Zea mays* ssp. *mexicana*) (25), expression patterns of tissues from different development stage of same inbred line (4,9) and same tissue of different samples within the AMP (10), three dimensional chromatin interactions and open chromatins of B73 (6), genetic variations including single-nucleotide polymorphisms (SNPs), small insertions and deletions (InDels) and large structure variations (SVs) generated from the deep sequencing of the AMP and the comparison among reference genome assemblies, the phenotypes and metabolome of the AMP and the related loci mapped by genome-wide association studies (GWAS), expression quantitative trait locus (eQTL)

and linkage analysis, the population structure and pedigrees of each germplasm and the selective signals between different teosinte subspecies and maize. ZEAMAP generated comprehensive functional annotations for the annotated gene models in each assembly, and provided useful tools for users to search, analyze and visualize all these different omics data.

Materials and methods

Data collection

The pedigree information of elite inbred lines in the maize association mapping panel was collected from (24). The B73 reference genome assembly (AGPv4) and annotation files (Version 4.43) were downloaded from Gramene (https://www.maizegdb.org/genome/genome_assembly/Zm-B73-REFERENCE-GRAMENE-4.0), while the Mo17 reference genome (version 1.0) and annotations (version 1.0) were downloaded from MaizeGDB (https://ftp.maizegdb.org/MaizeGDB/FTP/Zm-Mo17-REFERENCE-CAU-1.0/). The SK reference genome (version 1.0), annotations (version 1.0) and RNA-seq data of nine SK tissues were collected from (4). The HZS genome assembly, annotations and RNA-sequencing data were retrieved from Genome Sequence Archive in Beijing Institute of Genomics (BIG) Data Center (http://bigd.big.ac.cn/gsa) with project ID PRJCA001247. The genome assembly and annotations of *Zea mays* ssp. *mexicana* were from (25). RNA-sequencing of different B73 tissues were from (9), and RNA-seq data of developing maize kernels from 368 AMP inbred lines were from (10). The chromatin interaction and histone modification data were from (6), the chromatin accessibility data were from (7), and the DNA methylation data of the AMP were from (26). Phenotypes of the AMP, including agronomic traits, kernel amino acid contents, kernel lipid contents and metabolomic data were collected from previously reported studies (11,27,28). Linkage maps and QTL mapping results were collected from (29).

Functional annotation

For each genome annotation, the protein sequences of the predicted genes were compared against the InterPro database using InterProScan 5 (30) to identify functional protein domains. The proteins were further compared against the GenBank non-redundant protein (nr) database using Basic Local Alignment Search Tool (BLAST) with the options "-p blastp –e 1e-05 –b 5 –v 5 –a 4 –m 7 –F F". The BLAST results against the nr database and the Interpro results were further analyzed by Blast2GO (31) to assign gene ontology (GO) terms. Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations were performed by running BLAST against the KEGG database (version 84.0) with options "-p blastp -e 1e-05 -a 4 -m 8 -F F". The proteins were also searched against PFAM version 32.0 (32) using HMMer 3.1b2 (33) with default parameters. To identity gene orthologs and clusters of orthologous group (COG) annotations, the proteins were mapped to eggNOG orthology database (version 4.5.1) (34) using emapper-1.0.3 (35). To add gene-product annotations, the proteins were searched against UniProt database (version 2019_04) (36) using Diamond (v0.8.22.84) (37) with the options "--evalue 1e-05 --max-target-seqs 1", the UniProt and EggNog search results were combined to get the gene and product names using Gene2Product v1.32 (https://github.com/nextgenusfs/gene2product). Possible proteolytic enzymes were annotated by searching the proteins against the MEROPS database (version 12.0) (38) using Diamond with the options "--evalue 1e-05 --max-target-seqs 1". The proteins were also searched against the BUSCO (version 2.0) (39) dikarya models using HMMer with default options.

Comparative genomics

To identify synteny blocks, we first compared proteins from one genome to those from another using BLASTP with an E-value cutoff of 1e-10 and a maximum number of alignments of 5. The significant hits were then analyzed by MCScanX (40) with parameters "-k 50 -g -1 -s 5 -e 1e-10 -m 25 -w 5" to obtain synteny blocks. The whole genome comparison between two genome assemblies was performed using MUMmer 3 (41) with default parameters. The whole genome alignments between two genomes in ZEAMAP were performed using minimap2 (version 2.17-r941) (42) , with parameters "-c -x asm5 -B5 -O4,16 --no-long-join -r 85 -N 50 -s 65 -z 200 --mask-level 0.9 --min-occ 200 -g 2500 --score-N 2", and the raw alignment results were filtered to get the best alignment for each contig with QUAST-LG (43).

Genotyping and annotating of genetic variations

SNPs and InDels were identified and genotyped using DNA deep re-sequencing data (~20x) of the AMP (NCBI Bioproject: PRJNA531553) (4) in combination with GATK v3.8 (44) and samtools v1.9 (45). For each sample, raw reads were trimmed with Trimmomatic v0.36 (46) and the clean reads were used to call SNPs through the following protocol: 1) Reads were first aligned to B73 reference genome (AGPv4) using Bowtie2. 2) Uniquely mapped reads were used to generate local realignments with GATK RealignerTargetCreator and IndelRealigner. 3) The raw variants were called using GATK UnifiedGenotyper and Samtools mpileup. 4) The concordances of the two variant datasets were further filtered by screening out records with criteria "QualByDepth (QD) < 20.0, ReadPosRankSum < -8.0, FisherStrand (FS) > 10.0, QUAL < Mean quality of all SNPs". 5) The remaining variants were used to recalibrate base quality scores using GATK BaseRecalibrator to generate the ready-to-call-variants bam alignments. 6) The recalibrated bam files were subjected to variant calling using Samtools mpileup and GATK UnifiedGenotyper. 7) The concordances were filtered by keeping records with depth in a range of 5 to 200 and a quality value no less than 20 to get the variants for each sample. 8) Variants from different samples were merged using GATK CombineVariants and then genotyped using GATK UnifiedGenotyper. 9) Variants with a missing rate greater than 75% in the population were filtered out to get the final SNPs and InDels genotype matrix. The final SNPs and InDels were annotated using the Ensembl variant effect predictor (VEP) (47) according to B73 gene annotation v4.43. The polymorphic SVs between B73 and SK, as well as their genotypes in the AMP, were retrieved from (4), and annotated according to B73 gene annotation v4.43 using SURVIVOR v1.0.6 (48). Finally, haplotype blocks and tag SNPs were identified using Haploview (49).

Mapping, filtering and visualizing genetic loci

To perform genome-wide association studies for the collected phenotypic traits, the SNPs were then further filtered to keep only records with a minor allele frequency (MAF) of at least 5%. A mixed linear model accounting for the population structure (Q) and familial relationship (K) was used to examine the association between the SNPs and each trait using Tassel3 (50). The *P* value of each SNP was calculated, and significance was defined with Bonferroni corrected *P* value cutoff of 1/N,

where N is the total number of markers used. To prevent the interactive GWAS viewers and the tabular loci browser from operating too slowly, the volume of GWAS results was reduced by filtering out SNPs which had very low significance values (*P* value > 1e-4). The pairwise LD r^2 values of the remaining SNPs for each trait within 500 Kb windows were calculated using PopLDdecay (51).

To perform eQTL mapping, raw RNA-seq reads were trimmed with Trimmomatic v0.36 (46) then mapped to the B73 AGPv4 reference genome using STAR v2.7.2 (52) Read counts of each gene were calculated using HTSeq (53) and normalized by library sequencing depth using the R package DESeq2 (54) and Box–Cox transformation. eQTL mapping was performed for each gene using EMMAX (55) with hidden factors calculated using PEER (56) and multidimensional scaling coordinates calculated form the SNP dataset as covariates. We kept *cis*-eQTLs alone by retaining only the SNPs within 1 Mb of each gene (57). High quality cis-eQTL SNPs were selected by only retaining those with a *P* value smaller than the Bonferroni corrected *P* value cutoff of 1/N. The pairwise LD r^2 values of the remaining SNPs for each gene were calculated using PopLDdecay (51).

The visualization tools for GWAS results were developed using LocusZoom.js (https://github.com/statgen/locuszoom), a JavaScript embeddable plugin for interactively visualizing statistical genetic data, and ECharts (https://www.echartsjs.com), an open-sourced JavaScript visualization tool. The eQTL visualizer was modified from GTEx visualizations (https://github.com/broadinstitute/gtex-viz) (57).

Selective signals

We have previously deeply re-sequenced 183 teosinte samples, including 81 *Zea mays ssp. mexicana* lines, 70 *Zea mays ssp. parviglumis* lines, 12 *Zea nicaraguensis* lines, 2 *Zea diploperennis* lines, 2 *Zea perennis* lines, one *Zea luxurians*, one *Zea mays ssp. huehuetenangensis* and 14 mixed teosinte lines, through Illumina NovaSeq 6000 with a mean sequencing depth of ~20x (unpublished data). The aforementioned variant calling pipeline was also applied to this teosinte population to obtain a merged genotype matrix of both the AMP and the teosinte populations. The two major teosinte sub-species (*mexicana* and *parviglumis*) were compared to the maize AMP to calculate F_{ST} values using vcftools v0.1.16 (58) and selective sweep signals were detected using XP-CLR values within a 1kb siding window calculated using a previously described method (59). The visualizer for selection signals was also developed using LocusZoom.js and ECharts.

Population structures

The population structures were infered based on the populational SNP genotyping data. The principal component analyses (PCA) for a combination of the AMP and teosinte populations, and for each sub-population, were performed using smartpca embedded in EIGENSOFT version 6.1.4 (60). The ancestries for the whole Zea population and each sub-population were inferred from an unsupervised clustering analysis using ADMIXTURE (61). The interactive PCA dot plots and admixture bar plots were generated by ECharts.

CRISPR/Cas9 single-guide RNA designing

CRISPR/Cas9 sgRNAs for each maize reference genome were designed using CRISPR-Local (62) with default options. Results were converted into gff format with in-house perl scripts to format them for JBrowse.

Database contents and features

Overview structures

ZEAMAP is comprised of a user account management system, a main database, a full-site search engine, and a set of analysis and visualization tools (Figure S1). The multi-omics data in ZEAMAP are categorized into six main content modules involving genomic, genetic, variation, population, evolution and epigenetic information (Figure 1). ZEAMAP construction utilized the biological community database construction toolkit Tripal (63), which combines the content management system Drupal (https://www.drupal.org) with the standard biological relational database storage backend, Chado (64). Each feature in ZEAMAP has its own page and features are linked to each other by sequence ontology relationships.

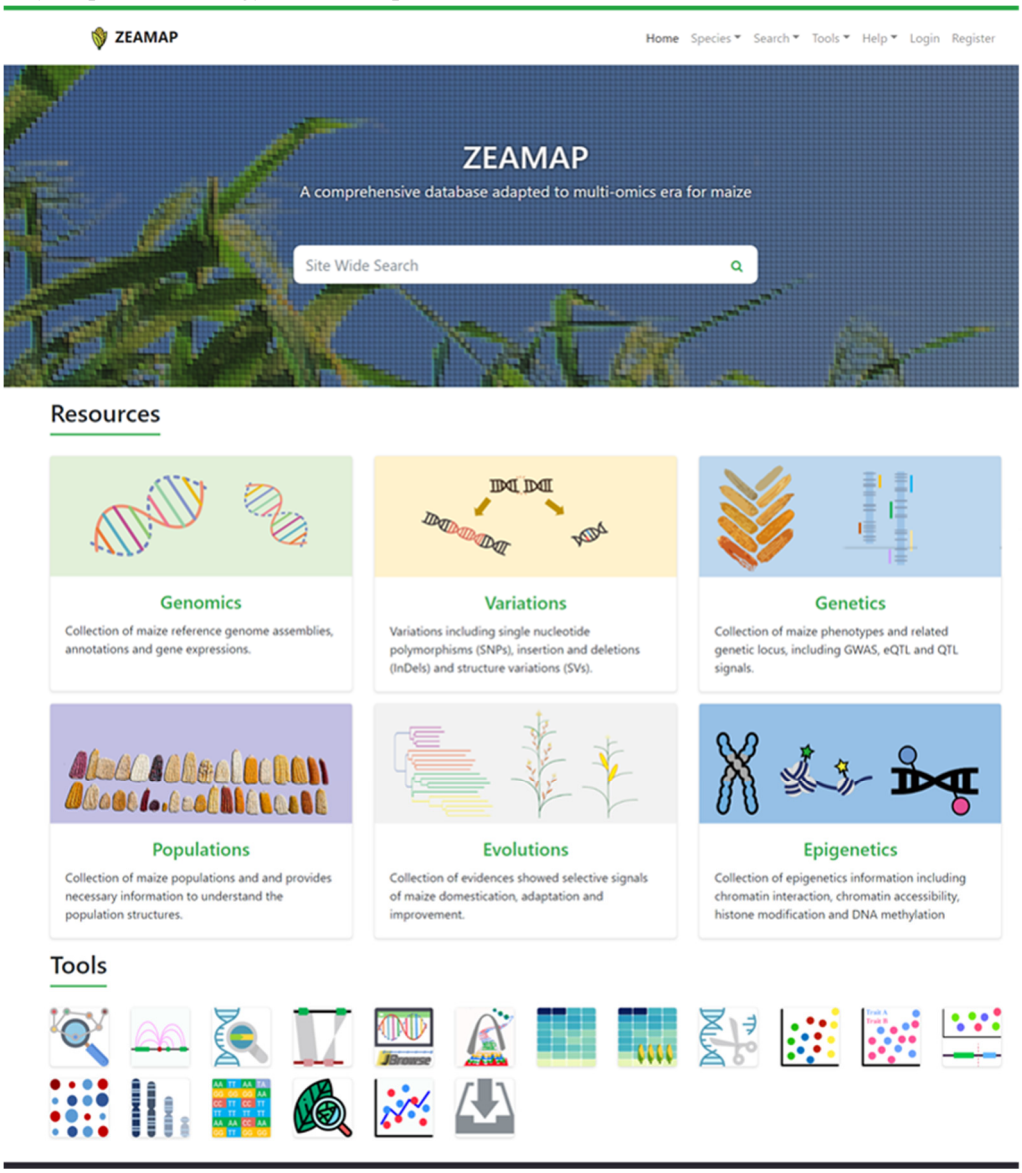


Figure 1. A screenshot of the ZEAMAP home page. The home page of ZEAMAP consists of a top menu bar, a site-wide search engine, access to the six biological modules and miscellaneous tools.

The Genomics module

The Genomics module collects reference genome assemblies, gene expression profiles and comparative genomics information related to the available genomes and populations in ZEAMAP. Currently, ZEAMAP contains reference genome assemblies of four maize inbred lines (B73, SK, Mo17 and HZS) and one teosinte species (*Zea mays ssp.* mexicana). Each genome assembly has its own page which contains general information for each genome assembly and sub-menus with links to access various related information and bioinformatic analysis tools (Figure S2). The mRNA and predicted protein for gene models in each genome assembly were assigned functional annotations including gene ontologies (GO), Kyoto Encyclopedia of Genes and Genomes pathways (KEGG), clusters of orthologous groups (COG), orthology groups, known gene-product annotations, proteolytic enzymes and homologs found in InterPro, PFAM, BUSCO and NCBI nr databases. The genome features for each assembly, including genes, mRNAs, proteins and transposable elements, can be searched by their IDs or locations through Chado feature search (Figure S3A). Genes (as well as mRNAs and proteins) can also be searched by their functional annotation descriptions (Figure S3B). Each annotated genome feature has its own page with multiple sub-menus displaying summary information (resource type, accession, organism, name, identifier and others), sequences, annotations, cross references linked to the same feature in MaizeGDB or NCBI, as well as related parent and child features, orthologs and synteny blocks.

Two genome browsers, JBrowse (65) and WashU Epigenome Browser (66), were embedded to display the genome sequences, annotated genomic features and other genomic information for all the available reference genome assemblies in ZEAMAP (Figure 2A). Both genome browsers are designed to easily add tracks, search for certain information in specific regions and export data as well as figures. These two genome browsers share some common features, including genome sequences and genomic annotations. However, each one has unique information tracks (see sections below) because JBrowse performs better when dealing with large piecemeal features such as variations and the WashU Epigenome Browser is specially designed to display epigenomic tracks.

ZEAMAP provides comparative genomic information for each pair of the available reference genome assemblies, including both synteny blocks identified from gene collinearities and whole genome alignment details. The synteny blocks are managed and displayed through the Tripal synteny viewer module (https://github.com/tripal/tripal_synview). Each synteny block has its own unique block ID and can be searched by it in both the Synteny block browser page and in the full-site search engine (Figure 2B). The detailed whole genome sequence alignments can also be accessed through the genome browser (Figure 2B).

We have collected gene expression patterns in different tissues for each maize genome assemblies, as well as expression profiles of kernels for 368 inbred lines of the AMP (10) based on B73 reference annotations. Expression patterns in different tissues for each gene can be visually displayed through heat maps after being queried in the “Tissue Overview” page (Figure 2C). ZEAMAP also enables users to browse the expression patterns of several genes among different tissues or samples and cluster the genes and tissues/samples based on the gene expression patterns (Figure 2C). Both functions provide download links to a raw expression matrix of the queries.

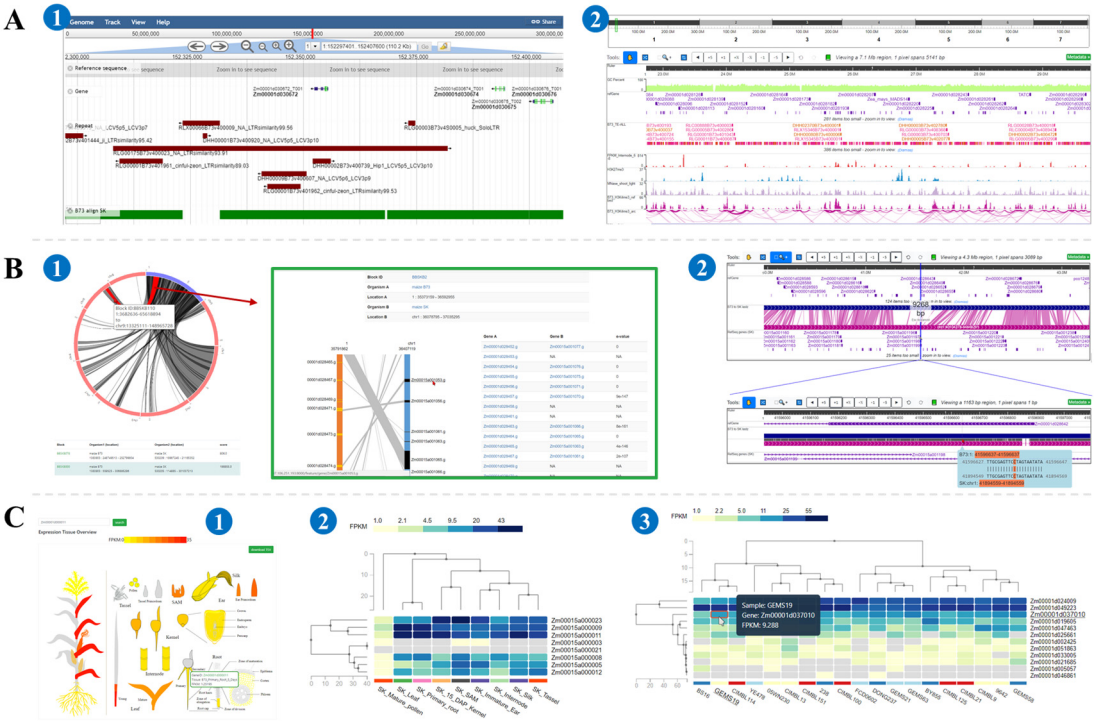


Figure 2. Features of the ZEAMAP genomics module. (A) Schematic of the two genome browsers embedded in ZEAMAP, JBrowse① and the WashU Epigenome Browser②. (B) Comparative genomic information in ZEAMAP, including Gene synteny blocks displayed by the synteny viewer with interactive circos plots and links to the detailed collinearity of the included genes①. Whole genome sequence alignments between two genomes are also accessible through the WashU Epigenome Browser, with Zoom-In and -Out functions and mouse-over display of the detailed alignments②. (C) Gene expression functions in ZEAMAP. The Tissue Overview function shows the expression of a gene in different tissues, with more detailed information available upon click①. ZEAMAP also has functions to cluster and display the expression patterns of several genes by tissue type② or sample③, with the gene IDs linked to pages with more detailed information.

The Variations module

The Variations module collects the genotypes and annotations of polymorphic variations including SNPs, InDels and SVs among the AMP in reference to the B73 reference genome, as well as a haplotype map generated from the SNP genotype matrix (See Materials and methods section for the source and the analysis used to generate the related data). The general variation information of a gene, including variation positions, allele types and annotations can be queried by their IDs or locations and displayed through tabular view (Figure S4A). The variations can also be browsed through JBrowse. Upon clicking each variant block in JBrowse tracks, the detailed information about that variant, including the annotations and the genotype of each germplasm will be shown. There is also a genotype overview for the variations in the current JBrowse display panel when the related “variant matrix” track is selected (Figure S4B). ZEAMAP also provides a function to query for the detailed genotype matrix for specified germplasms within certain regions (Figure S4C).

The Genetics module

ZEAMAP has collected phenotypic data from the AMP, including 21 agronomic traits, 31 kernel lipid content-related traits, 19 kernel amino acid content-related traits and 184 known metabolites of maize kernels. All these phenotypes can be searched and filtered by their threshold values using the "Search Trait Evaluation" tool (Figure S5). We have identified loci significantly associated with these phenotypes using GWAS and provided a tabular data search function to find specific loci by trait names, variant IDs, chromosome regions and significant *P* values (Figure S6). Three GWAS visualization tools ("GWAS-Single-Trait", "GWAS-Multi-Trait" and "GWAS-Locus") were developed to better browse the GWAS results and compare the significant signals among different traits. Querying a trait and navigating to specific regions can be easily accomplished by inputting boxes or through the interactive navigational Manhattan plot (in "GWAS-Single-Trait" and "GWAS-Multi-Trait" tools). The GWAS-Single-Trait tool displays all signals associated with the selected trait as a scatter plot, with colors indicating the linkage disequilibrium (LD) r^2 values between the user-selected reference variant and all the other variants (Figure 3A). The GWAS-Multi-Trait tool was designed to compare GWAS signals among two or more traits, with the colors indicating different traits (Figure S7), while the GWAS-Locus tool displays GWAS signals of all traits that show significant association with the query variant (Figure S8). These three tools are provided with a lightweight genome browser which indicates the gene models within the current region. Each element in the plot is also interactive and links to other related information.

Genetic variations can impact gene expression through many factors, including alterations in splicing, noncoding RNA expression and RNA stability (67). Expression quantitative trait locus (eQTL) mapping is a powerful approach to detect the possible variants which alter gene expression. In ZEAMAP, we have collected *cis*-eQTL signals with a total of 18,039 gene expression patterns in maize kernels, and provided a tabular tool to search and filter eQTL signals by gene IDs, gene locations, distances from transcription start site (TSS), effect sizes, and significance values (Figure S9). A visualization tool was also developed to browse all *cis*-eQTLs affecting the selected gene, with significance values, effect size and pairwise LD information displayed interactively (Figure 3B).

ZEAMAP has currently collected 12 published genetic maps constructed from different artificial maize segregating populations using genotypes generated from the Illumina MaizeSNP50 BeadChip (Illumina Inc., San Diego, CA, USA), as well as 813 quantitative trait loci (QTLs) identified from 15 plant architecture-related traits (29). The genetic markers can be searched and filtered by their IDs, genomic locations and genetic linkage group (Figure S10). QTLs can be searched by traits and QTL labels, resulting in detailed records of the genetic markers located in or adjacent to that QTL. By employing the TripalMap extension module (https://github.com/ksbuble/TripalMap), the linkage maps, including all related markers and QTLs, can be visualized and compared with another map interactively (Figure 3C).

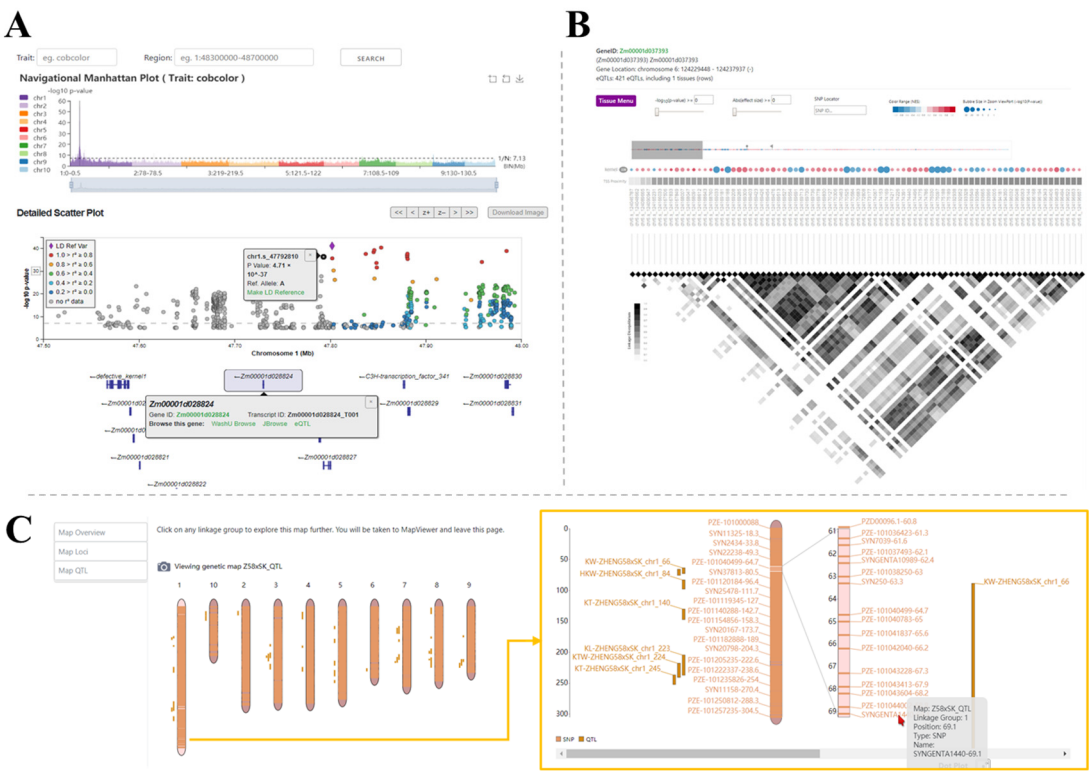


Figure 3. Features of the ZEAMAP genetics module. (A) Schematic of the "GWAS-Single-Trait" tool. The trait and region of interest can be queried through the top input boxes. Regions can be easily browsed through by clicking on the histogram of the interactive "Navigational Manhattan Plot" track. The "Detailed Scatter Plot" track plots the variants according to their chromosome locations and by the significance of their *P* values. The colors of each dot indicate the LD r^2 values between that variant and the reference variant (the purple diamond dot, can be reset by selecting the "Make LD reference" link on the popup page for each variant). The bottom track shows the gene annotations in the selected region, with a popup for each gene element which links to a detailed information page, genome browsers and the eQTL visualizer for that gene. (B) Schematic of the eQTL visualization tool. The significant *cis*-eQTL site for each gene is sized by the significance of its *P* value and colored by the effect size (beta value). The heatmap indicates pairwise LD r^2 values of the variants. (C) Schematic of the TripalMap tool in ZEAMAP. This tool displays the detailed genetic markers and mapped QTLs for each linkage group. Both the markers and the QTLs link to their own detailed information page.

The Populations module

It is often useful to dissect the genetic diversity, population structure and pedigrees of maize lines for both evolutionary studies and molecular breeding. ZEAMAP provides interactive information about the population structures assessed by principal component analysis (PCA) and ancestries inferred from an unsupervised clustering analysis using ADMIXTURE (61) for the whole *Zea* population and each sub-population in the database (Figure 4A). We have also added a table that lists the origins or pedigree information for each sub-population (Figure 4B).

Figure 4. Features of the ZEAMAP Populations module. (A) Interactive PCA diagram (top two dot plots) and structure diagram (stacked bar plot). Each diagram is zoomable and shows detailed information, including germplasm names and PCA/structure values when an element is moused over. (B) A table browser is provided to search for germplasm by pedigree, origin and subpopulation information.

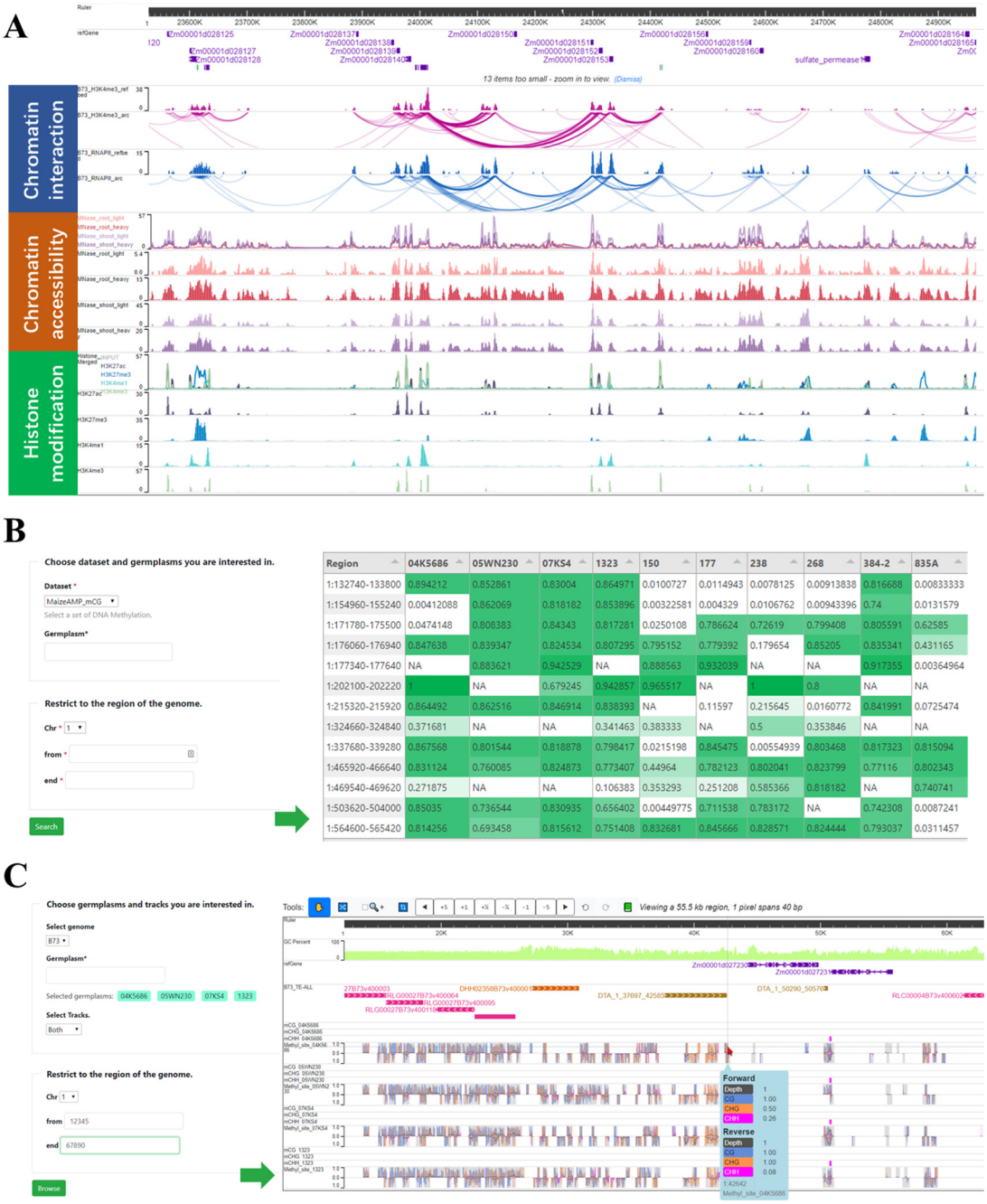
The Evolutionary module

Although maize is a well-domesticated and cultivated crop, it still have much to "learn" from their wild relatives, such as the ability to withstand biotic and abiotic stresses (68). In order to provide a general guide for adding new alleles from teosinte into maize breeding programs, ZEAMAP provides selection signals and genetic affinities between maize and its two main teosinte relatives, *Zea mays* ssp. *mexicana* and *Zea mays* spp. *parviglumis*. The evolutionary selection signals can be browsed graphically through an interactive "Selective signals browser", which is similar to the aforementioned GWAS viewer but with an additional Y-axis indicating the genetic variance by F_{ST} values (Figure S11A). Signals can also be analyzed and downloaded in a tabular format (Figure S11B), or viewed in the WashU Epigenome Browser (Figure S11C).

Epigenetics module

Eukaryotic gene expression has been shown to be altered by three-dimensional DNA interactions, which are affected by chromatin accessibility. Additionally, the modifications of epigenetic states

on histones and nucleotides add another layer of control to gene expression regulation (69). These regulatory factors are crucial for the ability of sessile plants to respond to diverse environmental challenges (70). In ZEAMAP, we have collected the chromatin interaction maps associated with RNA polymerase II occupancy and the histone mark H3K4me3 according to the B73 reference genome (6) . Open chromatin regions are based on micrococcal nuclease (MNase) digestion (7), histone acetylation and methylation regions, and populational DNA methylation information generated from the third leaves at V3 of the 263 AMP inbred lines (26). This information can be accessed through a tabular data browser or visualized through the WashU Epigenome Browser (Figure 5A). For DNA methylation information from the AMP, customized interfaces were developed to easily select multiple samples with differentially methylated regions (DMRs) in the table browser (Figure 5B) and visualize both DMR and DNA methylation sites in the WashU Epigenome Browser (Figure 5C).



Region	04K5686	05WN230	07KS4	1323	150	177	238	268	384-2	835A
1:132740-133800	0.894212	0.852861	0.83004	0.864971	0.0100727	0.0114943	0.0078125	0.00913838	0.816688	0.00833333
1:154960-155240	0.00412088	0.862069	0.818182	0.853896	0.00322581	0.004329	0.0106762	0.00943396	0.74	0.0131579
1:171780-175500	0.0474148	0.808383	0.84343	0.817281	0.0250108	0.786624	0.72619	0.799408	0.805591	0.62585
1:176060-176940	0.847638	0.839347	0.824534	0.807295	0.795152	0.779392	0.179654	0.85205	0.835341	0.431165
1:177340-177640	NA	0.883621	0.942529	NA	0.888563	0.932039	NA	NA	0.917355	0.00364964
1:202100-202220	1	NA	0.679245	0.942857	0.965517	NA	1	0.8	NA	NA
1:215320-215920	0.864492	0.862516	0.846914	0.838393	NA	0.11597	0.215645	0.0160772	0.841991	0.0725474
1:324660-324840	0.371681	NA	NA	0.341463	0.383333	NA	0.5	0.353846	NA	NA
1:337680-339280	0.867568	0.801544	0.818878	0.798417	0.0215198	0.845475	0.00554939	0.803468	0.817323	0.815094
1:465920-466640	0.831124	0.760085	0.824873	0.773407	0.44964	0.782123	0.802041	0.823799	0.77116	0.802343
1:469540-469620	0.271875	NA	NA	0.106383	0.353293	0.251208	0.585366	0.818182	NA	0.740741
1:503620-504000	0.85035	0.736544	0.830935	0.656402	0.00449775	0.711538	0.783172	NA	0.742308	0.0087241
1:564600-565420	0.814256	0.693458	0.815612	0.751408	0.832681	0.845666	0.828571	0.824444	0.793037	0.0311457

Figure 5. Features of the ZEAMAP Epigenetics module. (A) Schematic of chromatin interaction, chromatin accessibility and histone modification tracks displayed in the WashU Epigenome Browser. (B) Populational DNA methylation table browser. This tool filters population DNA methylation information by the DNA methylation type, germplasm and genomic region of interest, with the resulting matrix displaying DMRs for each selected germplasm within the query region. (C) Interface of the population DNA methylation genome browser. This interface provides options to display DNA methylation information by DMRs or DNA methylation sites of the selected germplasms within specified regions.

Additional tools

In addition to the aforementioned major biological modules, ZEAMAP also offers several additional tools. The currently available additional tools include a site-wide search engine, BLAST server, a CRISPR browser and an FTP data downloader.

Although there are already independent search tools for several of these analyses, a site-wide search engine powered by Chado is still useful since it enables users to quickly search for all items related to their queries. The ZEAMAP site-wide search engine was built using the Tripal Elasticsearch module (https://github.com/tripal/tripal_elasticsearch), with a search box which is accessible in both the home page and the status bar of each page. The search engine supports advanced search behaviors including wildcards, fuzzy searches, regular expressions and Boolean operators. The search results are also categorized by their entity type in the database (Figure 6A).

We have also implemented an instance of NCBI's BLAST tool in ZEAMAP using SequenceServer (71) , which provides a user-friendly interface with text-based and interactive visual outputs (Figure 6B). ZEAMAP currently has BLAST databases for whole genome sequences, mRNAs, CDSs and predicted proteins for each reference genome assembly.

ZEAMAP also includes an imbedded tool to search for reliable single-guide RNAs (sgRNAs) targeting the genes in each maize genome assembly in the database in order to support genome editing experiments using the CRISPR/Cas9 system. The sgRNA information can be browsed tabularly when querying a gene ID or a genomic region, and graphically through JBrowse. Both the tabular and the graphical results provide information about the editing positions and possible off-target genes (Figure S12).

Additionally, we have provided an FTP server to store a backup of all the publicly released datasets used in ZEAMAP through h5ai (https://larsjung.de/h5ai/), an open source file indexer with an enhanced user interface, text preview and directory download.

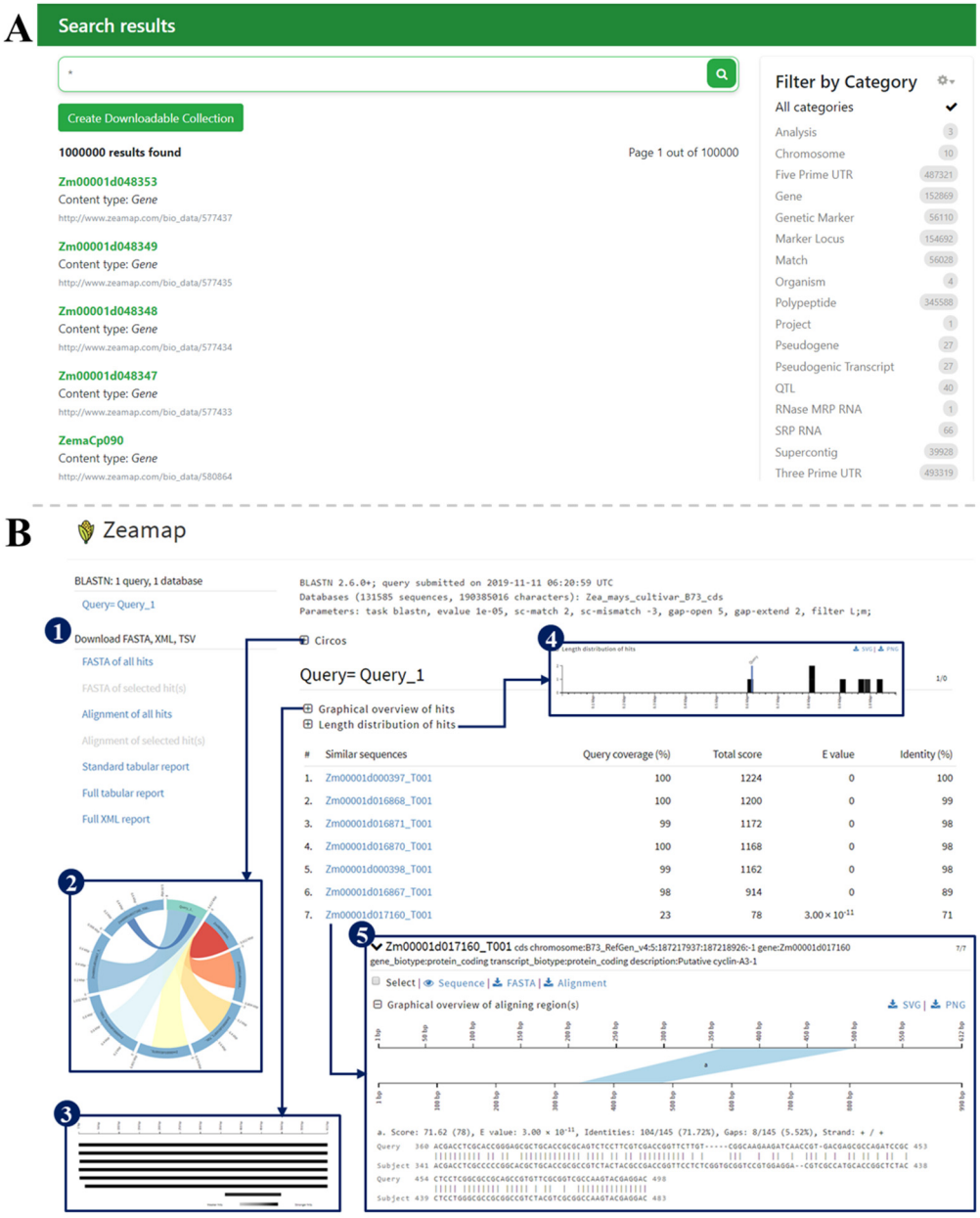


Figure 6. Features of the additional functional tools in ZEAMAP. (A) An example search result by the site-wide search engine queried with an asterisk wildcard. The resulting items are categorized by their feature types and have links to their detailed information pages. (B) An example result of the BLAST server in ZEAMAP. The result page provides download links of sequences① with reports available in different formats. Also included are interactive plot views of each alignment including Circos plots②, NCBI BLAST-like alignment hits visualization③ and length distribution of hits④. Each alignment hit has detailed alignment information, including a graphic view of the aligned regions and detailed alignments⑤.

Conclusion and future directions

We have created ZEAMAP, a database for maize research and breeding that collects multi-dimensional omics information, including genome assemblies, comparative genomics, transcriptomes, open chromatin, chromatin interactions, genetic variants, phenotypes, metabolomics, genetic maps, genetic mapping loci, population structures and pedigrees, and evolutionary selective signals between teosinte and maize. Most of the datasets were generated from the same maize population, which makes it possible to cross-reference these multi-omics data to support maize research in a more uniform and comprehensive manner. To make the acquisition and analysis of information more effective and flexible, ZEAMAP provides several convenient modules, including a site-wide search function, dataset-specific search tools, a BLAST server, a gene expression pattern analyzer, tabular browsers, genome browsers and specialized visualizers for different datasets.

ZEAMAP will be carefully maintained and continuously updated with the latest genomic and genetic advances. Currently, ZEAMAP has mainly focused on collection, query and visualization of pre-analyzed datasets, with several lightweight online analysis tools embedded. More online analysis tools (software for LD and PCA analyses, for example) will be embedded in ZEAMAP in the near future. Ultimately, we plan to systematically integrate all available omics data and make ZEAMAP a platform to analyze relationships between genotypes and phenotypes in order to predict complex traits for maize researchers and breeders.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Qing Li and Ms Jing Xu (National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University) for providing the DNA methylation data of AMP.

FUNDING

This research was supported by the National Key Research and Development Program of China (2016YFD0101001, 2016YFD0100303) and the National Natural Science Foundation of China (31525017, 31900494).

Reference

1. Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M.C., Wang, B., Campbell, M.S., Stein, J.C., Wei, X., Chin, C.-S. *et al.* (2017) Improved maize reference genome with single-molecule technologies. *Nature*, **546**, 524.
2. Springer, N.M., Anderson, S.N., Andorf, C.M., Ahern, K.R., Bai, F., Barad, O., Barbazuk, W.B., Bass, H.W., Baruch, K., Ben-Zvi, G. *et al.* (2018) The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nature genetics*, **50**, 1282.
3. Sun, S., Zhou, Y., Chen, J., Shi, J., Zhao, H., Zhao, H., Song, W., Zhang, M., Cui, Y., Dong, X. *et al.* (2018) Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nature genetics*, **50**, 1289.
4. Yang, N., Liu, J., Gao, Q., Gui, S., Chen, L., Yang, L., Huang, J., Deng, T., Luo, J., He, L. *et al.* (2019) Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nature genetics*, **51**, 1052.
5. Li, C., Song, W., Luo, Y., Gao, S., Zhang, R., Shi, Z., Wang, X., Wang, R., Wang, F., Wang, J. *et al.* (2019) The HuangZaoSi Maize Genome Provides Insights into Genomic Variation and Improvement History of Maize. *Molecular Plant*, **12**, 402-409.
6. Peng, Y., Xiong, D., Zhao, L., Ouyang, W., Wang, S., Sun, J., Zhang, Q., Guan, P., Xie, L., Li, W. *et al.* (2019) Chromatin interaction maps reveal genetic regulation for quantitative traits in maize. *Nature communications*, **10**, 2632.
7. Rodgers-Melnick, E., Vera, D.L., Bass, H.W. and Buckler, E.S. (2016) Open chromatin reveals the functional maize genome. *Proceedings of the National Academy of Sciences*, **113**, E3177-E3184.
8. Hirsch, C.N., Foerster, J.M., Johnson, J.M., Sekhon, R.S., Muttoni, G., Vaillancourt, B., Peñagaricano, F., Lindquist, E., Pedraza, M.A., Barry, K. *et al.* (2014) Insights into the maize pan-genome and pan-transcriptome. *The Plant Cell*, **26**, 121-135.
9. Walley, J.W., Sartor, R.C., Shen, Z., Schmitz, R.J., Wu, K.J., Urich, M.A., Nery, J.R., Smith, L.G., Schnable, J.C. and Ecker, J.R. (2016) Integration of omic networks in a developmental atlas of maize. *Science*, **353**, 814-818.
10. Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., Han, Y., Chai, Y., Guo, T., Yang, N. *et al.* (2013) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nature genetics*, **45**, 43.
11. Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., Liu, J., Liu, H., Chen, W., Luo, J. *et al.* (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nature communications*, **5**, 3438.
12. Wang, B., Regulski, M., Tseng, E., Olson, A., Goodwin, S., McCombie, W.R. and Ware, D. (2018) A comparative transcriptional landscape of maize and sorghum obtained by single-molecule sequencing. *Genome research*, **28**, 921-932.
13. Hu, L., Robert, C.A., Cadot, S., Zhang, X., Ye, M., Li, B., Manzo, D., Chervet, N., Steinger, T., Van Der Heijden, M.G. *et al.* (2018) Root exudate metabolites drive plant-soil feedbacks on growth and defense by shaping the rhizosphere microbiota. *Nature communications*, **9**, 2738.
14. Xu, G., Cao, J., Wang, X., Chen, Q., Jin, W., Li, Z. and Tian, F. (2019) Evolutionary Metabolomics Identifies Substantial Metabolic Divergence between Maize and its Wild Ancestor, Teosinte. *The Plant Cell*, tpc. 00111.02019.

15. Zhou, S., Zhang, Y.K., Kremling, K.A., Ding, Y., Bennett, J.S., Bae, J.S., Kim, D.K., Ackerman, H.H., Kolomiets, M.V., Schmelz, E.A. *et al.* (2019) Ethylene signaling regulates natural variation in the abundance of antifungal acetylated diferuloylsucroses and Fusarium graminearum resistance in maize seedling roots. *New Phytologist*, **221**, 2096-2111.
16. Wallace, J.G., Rodgers-Melnick, E. and Buckler, E.S. (2018) On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. *Annual review of genetics*, **52**, 421-444.
17. Lawrence, C.J., Dong, Q., Polacco, M.L., Seigfried, T.E. and Brendel, V. (2004) MaizeGDB, the community database for maize genetics and genomics. *Nucleic acids research*, **32**, D393-D397.
18. Portwood, J.L., Woodhouse, M.R., Cannon, E.K., Gardiner, J.M., Harper, L.C., Schaeffer, M.L., Walsh, J.R., Sen, T.Z., Cho, K.T., Schott, D.A. *et al.* (2018) MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucleic acids research*, **47**, D1146-D1154.
19. Zhao, W., Canaran, P., Jurkuta, R., Fulton, T., Glaubitz, J., Buckler, E., Doebley, J., Gaut, B., Goodman, M., Holland, J. *et al.* (2006) Panzea: a database and resource for molecular and functional diversity in the maize genome. *Nucleic acids research*, **34**, D752-D757.
20. Lee, T., Lee, S., Yang, S. and Lee, I. (2019) MaizeNet: a co-functional network for network-assisted systems genetics in Zea mays. *The Plant Journal*, **99**, 571-582.
21. Rajasundaram, D. and Selbig, J. (2016) More effort—more results: recent advances in integrative 'omics' data analysis. *Current opinion in plant biology*, **30**, 57-61.
22. Liu, H., Wang, F., Xiao, Y., Tian, Z., Wen, W., Zhang, X., Chen, X., Liu, N., Li, W., Liu, L. *et al.* (2016) MODEM: multi-omics data envelopment and mining in maize. *Database*, **2016**, baw117.
23. Tian, J., Wang, C., Xia, J., Wu, L., Xu, G., Wu, W., Li, D., Qin, W., Han, X., Chen, Q. *et al.* (2019) Teosinte ligule allele narrows plant architecture and enhances high-density maize yields. *Science*, **365**, 658-664.
24. Yang, X., Gao, S., Xu, S., Zhang, Z., Prasanna, B.M., Li, L., Li, J. and Yan, J. (2011) Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Molecular Breeding*, **28**, 511-526.
25. Yang, N., Xu, X.-W., Wang, R.-R., Peng, W.-L., Cai, L., Song, J.-M., Li, W., Luo, X., Niu, L., Wang, Y. *et al.* (2017) Contributions of Zea mays subspecies mexicana haplotypes to modern maize. *Nature communications*, **8**, 1874.
26. Xu, J., Chen, G., Hermanson, P.J., Xu, Q., Sun, C., Chen, W., Kan, Q., Li, M., Crisp, P.A., Yan, J. *et al.* (2019) Population-level analysis reveals the widespread occurrence and phenotypic consequence of DNA methylation variation not tagged by genetic variation in maize. *Genome biology*, **20**, 243.
27. Yang, N., Lu, Y., Yang, X., Huang, J., Zhou, Y., Ali, F., Wen, W., Liu, J., Li, J. and Yan, J. (2014) Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLoS Genetics*, **10**, e1004573.
28. Liu, H., Wang, X., Warburton, M.L., Wen, W., Jin, M., Deng, M., Liu, J., Tong, H., Pan, Q., Yang, X. *et al.* (2015) Genomic, transcriptomic, and phenomic variation reveals the complex adaptation of modern maize breeding. *Molecular plant*, **8**, 871-884.
29. Pan, Q., Li, L., Yang, X., Tong, H., Xu, S., Li, Z., Li, W., Muehlbauer, G.J., Li, J. and Yan, J.

(2016) Genome-wide recombination dynamics are associated with phenotypic variation in maize. *New Phytologist*, **210**, 1083-1094.

30. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G. *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**, 1236-1240.
31. Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674-3676.
32. Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J. *et al.* (2014) Pfam: the protein families database. *Nucleic acids research*, **42**, D222-D230.
33. Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R. and Finn, R.D. (2018) HMMER web server: 2018 update. *Nucleic Acids Research*, **46**, W200-W204.
34. Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M. *et al.* (2015) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research*, **44**, D286-D293.
35. Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C. and Bork, P. (2017) Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, **34**, 2115-2122.
36. Consortium, T.U. (2018) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, **47**, D506-D515.
37. Buchfink, B., Xie, C. and Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, **12**, 59-60.
38. Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A. and Finn, R.D. (2017) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*, **46**, D624-D632.
39. Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210-3212.
40. Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.-h., Jin, H., Marler, B., Guo, H. *et al.* (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic acids research*, **40**, e49-e49.
41. Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C. and Salzberg, S.L. (2004) Versatile and open software for comparing large genomes. *Genome biology*, **5**, R12.
42. Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094-3100.
43. Mikheenko, A., Prjibelski, A.D., Saveliev, V., Antipov, D. and Gurevich, A. (2018) Versatile genome assembly evaluation with QUAST-LG. *intelligent systems in molecular biology*, **34**, i142–i150.
44. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*,

20, 1297-1303.

45. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Subgroup, G.P.D.P. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.
46. Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120.
47. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P. and Cunningham, F. (2016) The ensembl variant effect predictor. *Genome biology*, **17**, 122.
48. Jeffares, D.C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., Balloux, F., Dessimoz, C., Bähler, J. and Sedlazeck, F.J. (2017) Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nature communications*, **8**, 14061.
49. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2004) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263-265.
50. Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633-2635.
51. Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M. and Yang, T.-L. (2018) PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, **35**, 1786-1788.
52. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15-21.
53. Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166-169.
54. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, **15**, 550.
55. Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.-y., Freimer, N.B., Sabatti, C. and Eskin, E. (2010) Variance component model to account for sample structure in genome-wide association studies. *Nature genetics*, **42**, 348.
56. Stegle, O., Parts, L., Piipari, M., Winn, J. and Durbin, R. (2012) Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nature protocols*, **7**, 500.
57. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N. *et al.* (2013) The genotype-tissue expression (GTEx) project. *Nature genetics*, **45**, 580.
58. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.
59. Chen, H., Patterson, N. and Reich, D. (2010) Population differentiation as a test for selective sweeps. *Genome research*, **20**, 393-402.
60. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, **38**, 904.
61. Alexander, D.H., Novembre, J. and Lange, K. (2009) Fast model-based estimation of ancestry

in unrelated individuals. *Genome research*, **19**, 1655-1664.

62. Sun, J., Liu, H., Liu, J., Cheng, S., Peng, Y., Zhang, Q., Yan, J., Liu, H.-J. and Chen, L.-L. (2018) CRISPR-Local: a local single-guide RNA (sgRNA) design tool for non-reference plant genomes. *Bioinformatics*, **35**, 2501–2503.
63. Ficklin, S.P., Sanderson, L.-A., Cheng, C.-H., Staton, M.E., Lee, T., Cho, I.-H., Jung, S., Bett, K.E. and Main, D. (2011) Tripal: a construction toolkit for online genome databases. *Database*, **2011**, bar044.
64. Jung, S., Lee, T., Ficklin, S., Yu, J., Cheng, C.-H. and Main, D. (2016) Chado use case: storing genomic, genetic and breeding data of Rosaceae and Gossypium crops in Chado. *Database*, **2016**, baw010.
65. Buels, R., Yao, E., Diesh, C.M., Hayes, R.D., Munoz-Torres, M., Helt, G., Goodstein, D.M., Elsik, C.G., Lewis, S.E., Stein, L. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome biology*, **17**, 66.
66. Zhou, X., Lowdon, R.F., Li, D., Lawson, H.A., Madden, P.A., Costello, J.F. and Wang, T. (2013) Exploring long-range genome interactions using the WashU Epigenome Browser. *Nature methods*, **10**, 375.
67. Gilad, Y., Rifkin, S.A. and Pritchard, J.K. (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends in genetics*, **24**, 408-415.
68. Mammadov, J., Buyyarapu, R., Guttikonda, S.K., Parliament, K., Abdurakhmonov, I.Y. and Kumpatla, S.P. (2018) Wild relatives of maize, rice, cotton, and soybean: treasure troves for tolerance to biotic and abiotic stresses. *Frontiers in plant science*, **9**, 886.
69. Dekker, J. (2008) Gene regulation in the third dimension. *Science*, **319**, 1793-1794.
70. He, Y. and Li, Z. (2018) Epigenetic environmental memories in plants: establishment, maintenance, and reprogramming. *Trends in Genetics*, **34**, 856-866.
71. Priyam, A., Woodcroft, B.J., Rai, V., Moghul, I., Mungala, A., Ter, F., Chowdhary, H., Pieniak, I.L., Gibbins, M.A., Moon, H. *et al.* (2019) Sequenceserver: a modern graphical user interface for custom BLAST databases. *Molecular Biology and Evolution*, **msz185**, 1-6.