# iScience



# Article

ZEAMAP, a Comprehensive Database Adapted to the Maize Multi-Omics Era



Gui et al., iScience 23, 101241 June 26, 2020 © 2020 The Author(s). https://doi.org/10.1016/ j.isci.2020.101241



# iScience

# Article ZEAMAP, a Comprehensive Database Adapted to the Maize Multi-Omics Era

Songtao Gui,<sup>1</sup> Linfeng Yang,<sup>1</sup> Jianbo Li,<sup>2</sup> Jingyun Luo,<sup>1</sup> Xiaokai Xu,<sup>2</sup> Jianyu Yuan,<sup>2</sup> Lu Chen,<sup>1</sup> Wenqiang Li,<sup>1</sup> Xin Yang,<sup>2</sup> Shenshen Wu,<sup>1</sup> Shuyan Li,<sup>1</sup> Yuebin Wang,<sup>1</sup> Yabing Zhu,<sup>2</sup> Qiang Gao,<sup>2</sup> Ning Yang,<sup>1,3,\*</sup> and Jianbing Yan<sup>1,\*</sup>

#### SUMMARY

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 multidimensional 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, 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 populational DNA methylation signals within maize inbred lines. 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 researches. As the cost of sequencing has been decreased and new omics technologies have arisen, there has been an explosive growth in the amount of biological information available for maize. The maize B73 reference genome has recently been updated (Jiao et al., 2017), and four high-quality maize genome assemblies have been released during the last 2 years (Li et al., 2019; Springer et al., 2018; Sun et al., 2018; Yang et al., 2019). 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 (Peng et al., 2019; Rodgers-Melnick et al., 2016). Omics data, including deep DNA resequencing, transcriptome, and metabolome, have been accumulated at the population scale (Hirsch et al., 2014; Hu et al., 2018; Li et al., 2013; Walley et al., 2016; Wang et al., 2018; Wen et al., 2014; Xu et al., 2019a; Zhou et al., 2019). There are many different applications for these new datasets, including gene cloning and the study of regulatory networks. These new and comprehensive datasets provide valuable resources for maize research and have the potential to completely revolutionize breeding (Wallace et al., 2018).

Comprehensive databases are needed to store, maintain, and analyze the multi-omics data that are 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, and phenotypic and genotypic data and also provides useful tools for maize data mining (Lawrence et al., 2004; Portwood et al., 2018). Panzea (https://www.panzea.org/) collects genotypic and phenotypic information for several maize populations (Zhao et al., 2006), whereas MaizeNet (http://www.inetbio.org/maizenet/) provides a genome-scale co-functional network of maize genes (Lee et al., 2019). 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 (Rajasundaram and Selbig, 2016) or use multi-omics data that were all generated from the same panel.

<sup>1</sup>National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China <sup>2</sup>BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China <sup>3</sup>Lead Contact \*Correspondence: yangningyingji@126.com

(N.Y.), yjianbing@mail.hzau.edu.cn (J.Y.)

https://doi.org/10.1016/j.isci. 2020.101241









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 (Liu et al., 2016). Despite the importance of wild relatives in understanding the domestication of modern crops, no existing maize multi-omics databases incorporate teosinte (Tian et al., 2019). 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 507 elite inbred lines (in an association mapping panel, AMP) (Yang et al., 2011) and 183 teosinte accessions (unpublished data). ZEAMAP includes genome assemblies and annotations of four inbred lines, B73 (Jiao et al., 2017), Mo17 (Sun et al., 2018), SK (Yang et al., 2019), and Huang-ZaoSi (HZS) (Li et al., 2019), and a teosinte accession (Zea mays ssp. mexicana) (Yang et al., 2017), expression patterns of tissues from different developmental stages of the same inbred line (Walley et al., 2016; Yang et al., 2019) and same tissue of different samples within the AMP (Li et al., 2013), three dimensional chromatin interactions and open chromatins of B73 (Peng et al., 2019), 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 (GWASs), expression quantitative trait locus (eQTL) and linkage analysis, the population structure and pedigrees of each germplasm, and the populational DNA methylation signals within maize inbred lines. 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.

#### RESULTS

#### **Overview Structures of ZEAMAP Database**

ZEAMAP comprises 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 five main content modules involving genomic, genetic, variation, population, and epigenetic information (Figures 1 and S1). ZEAMAP construction utilized the biological community database construction toolkit Tripal (Ficklin et al., 2011), which combines the content management system Drupal (https://www.drupal.org) with the standard biological relational database storage backend, Chado (Jung et al., 2016). Each feature in ZEAMAP has its own page, and features are linked to each other by sequence ontology relationships.

#### **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 bioinformatics 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 relationships, Single Copy Orthologs (SCO), known gene-product annotations, proteolytic enzymes, and related items in InterPro, PFAM, and NCBI nr databases (Figure S3). 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 S4A). Genes (as well as mRNAs and proteins) can also be searched by their functional annotation descriptions (Figure S4B). 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 (Buels et al., 2016) and WashU Epigenome Browser (Zhou et al., 2013), 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.







#### 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.

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).



# iScience Article



#### Figure 2. Features of the ZEAMAP Genomics Module

(A) Schematic diagram of the two genome browsers embedded in ZEAMAP, Jbrowse(1) and the WashU Epigenome Browser(2).

(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 genes (1). 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 alignments (2).

(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 click (1). ZEAMAP also has functions to cluster and display the expression patterns of several genes by tissue type (2) or sample (3), with the gene IDs linked to pages with more detailed information.

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 (Li et al., 2013) based on B73 reference annotations. Expression patterns in different tissues for each gene can be visually displayed through heatmaps 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.

#### **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 Transparent Methods 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 S5A). 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 S5B). ZEAMAP also provides a function to query for the detailed genotype matrix for specified germplasms within certain regions (Figure S5C).

# iScience Article





#### Figure 3. Features of the ZEAMAP Genetics Module

(A) Schematic diagram 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 by clicking on the histogram of the interactive "Navigational Manhattan Plot" track. The "Detailed ScatterPlot" 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<sup>2</sup> 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 that links to a detailed information page, genome browsers, and the eQTL visualizer for that gene.

(B) Schematic diagram 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 diagram 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 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 S6). 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 S7). 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 scatterplot, with colors indicating the linkage disequilibrium (LD) r<sup>2</sup> 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 S8), whereas the GWAS-Locus tool displays GWAS signals of all traits that show significant association with the query variant (Figure S9). These three tools are provided with a lightweight genome browser that 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 (Gilad et al., 2008). Expression quantitative trait locus (eQTL) mapping is a powerful approach to detect the possible variants that alter gene expression. In ZEAMAP, we have collected *cis*-eQTL signals with gene expression patterns in maize kernels based on B73 annotation 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 S10). 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, San Diego, CA, USA), as well as 813 quantitative trait loci (QTLs) identified from 15 plant architecture-related traits (Pan et al., 2016). The genetic markers can be searched and filtered by their IDs, genomic locations, and genetic linkage group (Figure S11). 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).

#### **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 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 inbred line of maize AMP (Figure 4B).

#### **The 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 (Dekker, 2008). These regulatory factors are crucial for the ability of sessile plants to respond to diverse environmental challenges (He and Li, 2018). 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 (Peng et al., 2019). Open chromatin regions are based on micrococcal nuclease (MNase) digestion (Rodgers-Melnick et al., 2016), histone acetylation and methylation regions, and populational DNA methylation information generated from the third leaves at V3 of the 263 AMP inbred lines (Xu et al., 2019b). 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).

#### **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 that 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).

# iScience Article





в

ID 🔷	Lines ^	Predigree	Origin 🔶	Sub Populations 🔶
filter column	filter column	filter column	filter column	filter column
1	150	Unkown	China	Mixed
2	177	Unkown	China	Mixed
3	238	Unkown	China	NSS
4	268	(shen5003xU8112) F3x32	China	NSS
5	501	Unknown	China	NSS

#### 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.

We have also implemented an instance of NCBI's BLAST tool in ZEAMAP using SequenceServer (Priyam et al., 2019), 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





Α		Ruler								
			1 23600K 2370			24100K 24200K			24600K 24700K	24800K 24900K
		refGene	2m00001d028125	Zm00001d028137	Z	m00001d028150	29151	Zm00001d028156		Zm00001d028164 Zm00001d028165
			Em00001d02812	Zm00001d02 Zm00001d02	28139	Zm00001	028152	Zm00001d028159 Zm00001d0281	50 <b>m</b>	Zm0(
			Zm00001d0281	128 Zm00001	d028140	Zm000	01d028153		sulfate_perme	ase 1
						13 items too small - zoo	n in to view. (Dismiss)			
		B73_H3K4me3_ref	38 7		1					
					- 14 6 A . 64	64.6	to h			
		B73_H3K4me3_arc			19		IT	$\sim$		
	ti a				1		// \			
5	ĔΫ						1 1			
		B73_RNAPII_refbe	15		. A	1	114			
	<u>ר</u> ש ב	873 RNAPII arc	0	į .		a half a second	14 4	- <b>A</b>	اب . اسه	14 44 7 1 1 <b>1</b> 1
7	Unromatin	B73_RNAPII_arc								
									X	
		MNase_root_light	57 7							
		MNase_root_heavy	10 Au		111		1.	M die	A di	1. 2
	- 2	MNase_shoot_light MNase_shoot_heavy	A A A A A A A A A A A A A A A A A A A	in a star	Aller	man All A a muse stars	AA A	a Manual M	A . Man	all a de se have , he
		MNase_root_light 5	147	a set and a set of the set of						
Č	<b>p</b> i a		a matter del de la des	مقلال بالله	A he had a h	de a child it structure beau	All the set	a life a server bld	all daring have	when my whenda the
	E is	MNase_root_heavy		a and makes an and a second	di at at i i	in their miniates in the	1.	1 101	id a	16 I I.
9	O N		ALL AND MALLA	Add when a different		A WELL & Milara A Mile	M. A. M.	And A. L. Maria and Maria	M. Level Leve	will & Mary Marken & I
	unromatin accessibility	MNase_shoot_light	45		A 11	- 1		h in	a sul	
C C	ັ່ຼ	MNase_shoot_heav ;	as a shi titu the	and the sheet		m	Ah A	a da	M. a. M. A.	the day Amar Ad
		y y	20		de al contra de als	and the second	1.	li ia l	1.6	a. 1 1 .
		Histone_INPUT	0	a shall be a shall be a	AL AL MIL A	add a seadly a fible a dis	All A	a_dh A_i , mai a i _dhi	M. A. will See &	h h A and Ab Ab
	100	Merged H3K27ac	h hu		1.1		T.			
	5	H3K27me3 H3K4me1	· In MA		A la ba	6.1	IA I	1	A	
	nistone modification	H3K27acH3K4me3 :	30	and the second s		(A.A	6.49.19			
	a a		o le bad		Ante		111	1		A
	<u>e.</u>		35]	-						
	s =		•				1.	A she		
	EZ	H3K4me1	15							12
92	Ĕ		0	1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1				and the surgers of		A
	- Q <b>-</b> S	H3K4me3	57		1.1		E a			
			•		11.11		1.1	1	11	2

#### в

Choose dataset and germplasms you are interested in.	Region 🔶	04K5686 🔺	05WN230 🔶	07KS4 🔶	1323 🔶	150 🔶	177 🔶	238 🔶	268 🔶	384-2 🔶	835A 🔶
Dataset *	1:132740-133800	0.894212	0.852861	0.83004	0.864971	0.0100727	0.0114943	0.0078125	0.00913838	0.816688	0.00833333
MaizeAMP_mCG	1:154960-155240	0.00412088	0.862069	0.818182	0.853896	0.00322581	0.004329	0.0106762	0.00943396	0.74	0.0131579
elect a set of DNA Methylation.	1:171780-175500	0.0474148	0.808383	0.84343	0.817281	0.0250108	0.786624	0.72619	0.799408	0.805591	0.62585
Sermplasm <sup>-</sup>	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	0.0036496
	1:202100-202220	1	NA	0.679245	0.942857	0.965517	NA	1	0.8	NA	NA
Restrict to the region of the genome.	1:215320-215920	0.864492	0.862516	0.846914	0.838393	NA	0.11597	0.215645	0.0160772	0.841991	0.0725474
Chr • 1 •	1:324660-324840	0.371681	NA	NA	0.341463	0.383333	NA	0.5	0.353846	NA	NA
from *	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
nd *	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
arch	1:564600-565420	0.814256	0.693458	0.815612	0.751408	0.832681	0.845666	0.828571	0.824444	0.793037	0.0311457

с







#### Figure 5. Features of the ZEAMAP Epigenetics Module

(A) Schematic diagram 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.

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.

#### DISCUSSION

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 populational DNA methylation signals within maize inbred lines. 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. 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.

#### **Limitations of the Study**

Currently, ZEAMAP has mainly focused on collection, query, and visualization of pre-analyzed datasets, with only several lightweight online analysis tools embedded. The lack of comprehensive online analysis tools that make the best of the multi-omics data in ZEAMAP to help users better understand their custom data is the main limitation for the current version of ZEAMAP database.

#### **Resource Availability**

#### Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Ning Yang (yangningying)i@126.com).

#### Materials Availability

This study did not generate new unique reagents.

#### Data and Code Availability

All the strategies and data included in this paper are available from ZEAMAP (http://www.zeamap.com). The related source codes are available at https://github.com/ZEAMAP.

#### **METHODS**

All methods can be found in the accompanying Transparent Methods supplemental file.

#### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101241.





Content type: Gene Introductions of Mark 277437     Analysis Content type: Gene Introductions of Mark 277437     Analysis Content type: Gene Introductions of Mark 277437       Zm00001d04343 Content type: Gene Introductions complex data 277438     Benefit Marker Introductions of Mark 277437       Zm00001d04343 Content type: Gene Introductions complex data 277438     Preve dogenic Transcript Introductions of Mark 277438       Zm0001d04343 Content type: Gene Introductions complex data 277438     Preve dogenic Transcript Introductions of Mark 277438       Zmacp090 Content type: Gene Introductions of Mark 277438     Supercontig Introductions of Mark 277438       Zmacp091 Content type: Gene Introductions of Mark 277438     Market 120030       Zmacp092 Content type: Gene Introductions of Mark 277438     Supercontig Introductions of Mark 277438       Zmacp091 Content type: Gene Introductions of Mark 277438     Market 120138       Zmacp092 Content type: Gene Introductions of Mark 277438     Supercontig Introductions of Mark 277438       Zmacp091 Content type: Gene Introductions of Mark 277438     Market 120138       Zmacp092 Content type: Gene Introductions of Mark 277438     Supercontig Introductions of Mark 277438       Zmacp093 Content type: Gene Introductions of Mark 277438     Supercontig Introductions of Mark 277438       Zmacp094 Content type: Gene Introductions of Mark 277438     Market Introduction of Mark 277438       Zmootodocome Note Status and Content type Introductions of Mark 277438     Supercontig Introduction of Mark 277438       Zmootodocome Introduct	*	٩	Filter by Catego	orv 🌣
Career Uswaliostable Collection       Analysis         100000 results found       Page 1 out of 100000       File Winnessense         Content type: Gree       Gene       Gene         Integration assamp combine, data/STRES       Match       Gene         Content type: Gree       Match       Gene       Gene         Integration assamp combine, data/STRES       Match       Gene       Gene         Content type: Gree       Match       Gene				.,
100000 results found     Page 1 out of 100000     Chromosome       2media     Genetit Marker     Genetit Marker       2media     Genetit Sper. Gene     Project       10me//www.ammaccom/bio.data/57633     Superconfig     Genetit Marker       2media     Genetit Sper. Gene     Superconfig     Genetit Marker       10me//www.ammaccom/bio.data/57636     Genetit Sper. Gene     Superconfig     Genetit Marker       10me//www.ammaccom/bio.data/57636     Genetit Sper. Gene     Superconfig     Genetit Marker       10med//www.ammaccom/bio.data/57636     Genetit Sper. Gene     Superconfig     Genetit Marker       10med//wwww.ammaccom/bio.data/57641     Superconfig <th>Create Downloadable Collection</th> <th></th> <th></th> <th></th>	Create Downloadable Collection			
Zumoon Lokessas     Five Prime UTR     Gene       Content type: Gene     Marker Locus     Marker Locus       Marker Locus     Marker Locus     Marker Locus       Thydrows assumption high, 184, 1977,83     Organism       Zeneoor I dokasaa     Polypetide       Thydrows assumption high, 184, 1977,83     Predogene       Zeneoor I dokasaa     Project       Thydrows assumption high, 184, 1977,83     Predogene       Zeneoor I dokasaa     Project       Thydrows assumption high, 184, 1977,83     Predogene       Zeneoor I dokasaa     Predogene       Thydrows assumption high, 184, 1977,83     Predogene       Zeneoor I dokasaa     Star RAA       Sperconting     Three Prime UTR       Witz/Iwre assumption high, 184, 1977,83     Star RAA       Sperconting     Three Prime UTR       Witz/Iwre assume Conducts, 484, 1978, 1978, 1989, 1977, 198     Star RAA       Sperconting     Three Prime UTR       Witz/Iwre assumption high, 184, 1977, 198     Star RAA       Sperconting     Three Prime UTR       Witz/Iwre assumption high, 184, 1977, 198     Star RAA       Sperconting     Three Prime UTR       Witz/Iwre assumption high, 184, 1977, 198     Star RAA       Sperconting     Star RAA       Sperconting     Star RAA       Sperconting	1000000 results found	Page 1 out of 100000		1
Content type: Grave     Grave     Grave       Purport/instrumentstamm.prom/biol.state/STR45     Grave     Grave       Purport/instrumentstamm.prom/biol.state/STR45     Project       Purporting     Strans.prog0       Content type: Grave     BLSTN1 12.6.6.9: govery subsitted on 2015-11-11 06:20:59 UTC       Deversoond of NSTA, NML, TSOV     Purporting       Purporting     BLSTN1 12.6.6.9: govery subsitted on 2015-11-11 06:20:59 UTC       Deversoond of NSTA, NML, TSOV     Purporting       Postad and tables     BLSTN1 12.6.6.9: govery subsitted on 2015-11-10 06:20:59 UTC       Deversoond of NSTA, NML, TSOV     Purporting       Postad and tables     Purporting       Standard tables     Corpstical overview of hits       Standard tables     Purporting       Purporting     Standard tables       Purporting     Standard tables       Standard tables     Open purportexit	Zm00001d048353			48732
thtp://www.zamap.com/biol.dtat/57732 Zmotoolidd/3339 Content type: Grine http://www.zamap.com/biol.dtat/57733 Zmotoolidd/3343 Zmotoolidd/3347 Content type: Grine http://www.zamap.com/biol.dtat/57733 Zmotoolidd/3347 Content type: Grine http://www.zamap.com/biol.dtat/57733 Zmotoolidd/3347 Content type: Grine http://www.zamap.com/biol.dtat/57733 Zencoolidd/3347 Content type: Grine http://www.zamap.com/biol.dtat/57733 Zencoolidd/3408 Zencoolidd/3467 Content type: Grine http://www.zamap.com/biol.dtat/57733 Zencoolidd/3408 Zencoolidd/3467 Content type: Grine http://www.zamap.com/biol.dtat/57733 Zencoolidd/3408 Zencoolidd/347733 Zencoolidd/3408 Zencoolidd/347733 Zencoolidd/3408 Zencoolidd/347733 Zencoolidd/3408 Zencoolidd/347733 Zencoolidd/3408 Zencoolidd/347733 Zencoolid			Gene	15286
Lindou 10042499       Match         Content type: Gene       Match         Time: Why://www.samp.com/time_data/57735       Polycet         Content type: Gene       Project         Time: Why://www.samp.com/time_data/57735       Project         Content type: Gene       Project         Mitc//www.samp.com/time_data/57735       Project         Zenoool 10045347       Project         Content type: Gene       Project         Mitc//www.samp.com/time_data/57735       Project         Zenoool 10045347       RNase MRP RNA         Seperconig       Three Prime UTR         Mitc//www.samp.com/time_data/57735       Project         Zenoool 10045347       RNase MRP RNA         Superconig       Three Prime UTR         Mitc//www.samp.com/time_data/57763       Project         Zenoood Addistits 2.6.0m; guery subsitted on 3059-11-11 66:20:59 UTC       Detablaser (21355 sequences, 10355056 characters); Tax_says_coll/bar_P32_cd5         Query Oury_1       Extitution of hits       Cuery Oury_1         Mitch datablar report       Fill Stall Adata Batan, evalue 1=69, sc=atal 2, sc=at			Genetic Marker	5611
Content type: Gene http://www.asampc.com/biol.plail.plaif.p	Zm00001d048349		Marker Locus	15469
http://www.samap.com/bio_data/57743. Content type: Content type			Match	5602
Zmodouł dołasia       Project         Content type: Gene       Project         Impl/mon-zemap.com/bio.dsta/57733       Pseudogenic         Zmodouł dołasia       Gut         Zmodouł dołasia       Gut         Zmodouł dołasia       Gut         Zmodouł dołasia       Skip RNA         Szenac poła       Skip RNA         Zmodouł dołasia       Skip RNA         Zmodouł dołasia       Skip RNA         Szenac poła       Skip RNA         Skip RNA       Supercontig         Intra//www.semap.com/bio.dsta/5006c1       Three Prime UTR         Intra//www.semap.com/bio.dsta/5006c1       Databases (13155 sequences, 19036506 charactera): Tag_mays_colitivar_B73_cds         Paraeters: task blacto, evalue 18-05, sc-math 2, sc-math 2, sc-mismatch -3, gap-open 5, gap-extend 2, filter Lym;         Intra/www.semap.com/bio.dsta/5000037_T001       Citos         Standard tabular report       Full XML report         Full XML report       Sconoouddousers_T001       100         Interport       Sconoouddousers_T001       100       1224         Interport       Sconoouddousers_T001       100       1224       0         Interport       Sconoouddousers_T001       100       1224       0       2       2       2       0 <td></td> <td></td> <td>Organism</td> <td></td>			Organism	
Content type: Gene http://www.asamag.com/bio.dstat/577434 Zan0001104344577434 Zan000110434547 Content type: Gene http://www.asamag.com/bio.dstat/577434 Zencopool Content type: Gene http://www.asamag.com/bio.dstat/577434 Zencopool Model Amage and	7:::00001d049349		Polypeptide	34558
http://www.rzemsp.com/biog.data/S77244 ZmoColol 1048347 Content type: Gene http://www.rzemsp.com/biog.data/S77243 ZmaCp090 Content type: Gene http://www.rzemsp.com/biog.data/S70604 With://www.rzemsp.com/biog.data/S70604 With://www.rzemsp.com/biog.data/S70604 BLASTN: 1 quety, 1 database Quety, 1 database Quety Quety, 1 Databases (13565 sequences, 19038506 characters): 2.8e_mays_cultiver_B72_cds Parameters: task blastn, evalue 1e-65, sc-atch 2, sc-mismatch -3, gap-open 5, gap-extend 2, filter t;m; Poundod ASTA, X0ML, TSV FASTA of all hils Alignment of align ergin(h) Alignment of align ergin(h) Align			Project	
Zm00001d0434347       OTL         Content type: Gene       RNase MRP RNA         SRP RNA       Supercontig         Tht://www.zeamap.com/bio.dsta/377433       Supercontig         Zm0001d0434347       Supercontig         Zm0001d0434347       Supercontig         RNase MRP RNA       Supercontig         Supercontig       Supercontig         Tht://www.zeamap.com/bio.dsta/3700064       Three Prime UTR         BLASTR: 1 guery, 1 database       BLASTR 2.6.0+; guery submitted on 2019-11-11 06:20:59 UTC         Download FASTA, XML_TSV       Parameters: task blasts, evalue 1e-05; sc-match 2; sc-match 3; gap-open 5; gap-extend 2; filter Lip;         Download FASTA, XML_TSV       Circos         Very Query_1       Graphical overview of hits         BLastre report       Standard tabular report         Full Maluer report       1. Zm00001d016867_T001       100       1224       0         Standard tabular report       2. Zm00001d016867_T001       100       126       0       2. Zm00001d016867_T001       100       126       0         Standard tabular report       Scropoot1d016867_T001       100       126       0       2. Zm00001d016867_T001       100       126       0       2. Zm00001d016867_T001       123       78       3.000 t10 <sup>21</sup>			Pseudogene	2
Zm00001r04/9347       OTL         Content type: Gene       RNase MRP RNA         Strandpoint       Strandpoint         Content type: Gene       Three Prime UTR         Thtt://www.treamap.com/biol.data/\$77433         Zemacp090         Content type: Gene         Thtt://www.treamap.com/biol.data/\$20064         Staffer 1 guery, 1 database         BLASTN: 1 guery, 1 database         Query: Query_1         Btaffer 2 guery submitted on 2019-11-11 66:20:59 UTC         Download FASTA, XML, TSV         Parameters: task blastn, evalue 1e-05, sc-match 2, sc-mismatch -3, gap-open 5, gap-extend 2, filter Lip;         Download FASTA, XML, TSV         Picts of all hits         Alignment of all hits         Alignment of all hits         Alignment of all hits         Alignment of all hits         Standard tabular report         Full tabular repo				2
Content type: C				4
SRP RNA Superconfig Three Prime UTR SRP RNA Superconfig Three Prime UTR Superconfig Three Prima UTR Superconfig Three				
Zemacp090 Entert type: Gene Inter/Inversemana.com/Bio_data/500664       Supercontig Three Prime UTR         BLASTN: 1 query, 1 database Query=Query_1       BLASTN: 2.6.0+; query submitted on 2019-11-11 06:20:59 UTC Databases (131555 sequences, 199355016 characters): Zea.mays.cultivar_B72_cds Parameters: task blastn, evalue 1e-05, sc-match -3, gap-open 5, gap-extend 2, filter L;m;         Ownhoad FASTA, XML, TSV       Circos         FASTA of all hits       Query=Query_1         Alignment of all hits       Circos         Alignment of all hits       Circos         Standard tabular report       Similar sequences         Full tabular report       100         Full XML report       9         Strandord UseR_TODI       100         Tormooold01687_T001       100         Tormooold01687_T001       23         Tormooold016887_T001       23				6
Contenting type: General type:			Supercontig	3992
BLSTN: 1 query, 1 database Query Query_1          BLSTN: 1 query, 1 database       BLASTN 2.6.0+; query submitted on 2019-11-11 06:20:59 UTC Databases (131555 sequences, 190385016 characters): Zeq.mays_cultivar_B73_cds Parameters: task blastn, evalue 1e-05, sc-entich 2, sc-entismatch -3, gap-open 5, gap-extend 2, filter L;s;         Ownload FASTA, XML, TSV       Circos         FASTA of all his       Query=Query_1         Alignment of selected hit(s)       Graphical overview of hits         Standard tabular report       Graphical overview of hits         Vancouldules*101       100       1024         Vancouldules*101       100       100         1.27mosoold006885, 1001       100       1224         2.7mosoold016871, 100       100       1224       0         3.7mosoold016870, 100       100       1224       0         3.7mosoold016877, 1001       39       1172       0         3.7mosoold016870, 100       100       1168       0         3.7mosoold016870, 1001       38       94       0         3.7mosoold016870, 1001       38       94       0         3.7mosoold01687160, 1001       38       94       0         3.7mosoold016877, 1001       38       3.00 × 10 <sup>41</sup> 10         3.7mosoold016877, 1001       38       3.00 × 10 <sup>41</sup> 10				49331
Alignment of selected hild(s)         Standard tabular report         Full soular report         Full soular report         Full soular report         Full soular report         Composition for the sequences         Query coverage (%)         Total score         E Length distribution of hits         *         Similar sequences         Query coverage (%)         Total score         E Value         Identity         Zmocooldoids87_T001         100       11224         2         Zmocooldoids87_T001         99       1162         0         Zmocooldoids87_T001         99       1162         0       Zmocooldoids87_T001         10       123         78       3.00×10*1         Comoooldoids7160_T001       23         78       3.00×10*1         Comoooldoil7160_T001       23         78       3.00×10*1         Comoooldoil7160_T001       23         78       3.00×10*1         Comoooldoil7160_T001       24         8elect () Sequence () ASATA () Alignment         B caphical overview of aligning	BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV	Databases (11585 sequences, 199385016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-mismatch -3, gap-open 5, gap Circos		G
Standard tabular report         Similar sequences         Query coverage (%)         Total score         E value         Identity           Full tabular report         1         2.mo0001d00397_T001         100         1224         0         12         <	BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits	Databases (13155 sequences, 19935516 characters): Zea_mayi_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-match -3, gap-open 5, gap Circos Query=Query_1	± svc] ± mo	5
Full tabular report       1. Zmo0001d000097_1001       100       1224       0         Full XML report       2. Zmo0001d01688_1001       100       1200       0         3. Zmo0001d01687_1001       99       1172       0         4. Zmo0001d01687_1001       99       1162       0         5. Zmo0001d01687_1001       99       1162       0         6. Zmo0001d016867_1001       23       78       3.00 × 10 <sup>11</sup> Totopoold017160_T001       23       78       3.00 × 10 <sup>11</sup> Select (* Sequence) & FASTA (& Alignment)         Select (* Sequence) & FASTA (& Alignment)         Graphical overview of aligning region(s)	BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of selected hit(s) Alignment of all hits	Databases (131585 sequences, 199385016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-mismatch -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits	± svc] ± mo	6 1/0
Full XML report       2. </td <td>BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits Alignment of all hits Alignment of selected hit(s)</td> <td>Databases (13155 sequences, 199355016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-mismatch -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits D Length distribution of hits</td> <td></td> <td>1/0 Identity (%</td>	BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits Alignment of all hits Alignment of selected hit(s)	Databases (13155 sequences, 199355016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-mismatch -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits D Length distribution of hits		1/0 Identity (%
2     1     0       1     2     100       1     100   <	BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits Alignment of all hits Alignment of selected hit(s) Standard tabular report	Databases (13155 sequences, 199355016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue 1e-05, sc-match 2, sc-mismatch -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits B Length distribution of hits # Similar sequences Query coverage (%) Total so	A stol A res	
4.         Zmooold016870_T001         100         1168         0           5.         Zmooold00398_T001         99         1162         0           6.         Zmooold016667_T001         98         914         0           7.         Zmooold017160_T001         23         78         3.00×10 <sup>41</sup> Sector         Zmooold017160_T001 cds chromosome873_Ref6en_v45:187218726-1 geneZm00001d017160         gene_biotypeprotein_coding transcript_biotypeprotein_coding description=Putative cyclin A3-1         SetCt   @ Sequence   & FASTA   & Alignment           B         Graphical overview of aligning region(s)         & \$XVG   & \$XVG	BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits Alignment of all hits Alignment of selected hit(s) Standard tabular report	Databases (13155 sequences, 199355016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue 1e-05, sc-match 2, sc-mismatch -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits B Length distribution of hits # Similar sequences 1. Zm00001d000397_T001 100 12	Livel Live re Evalue 24 0	Identity (%
3.       Zm0001d00338, T001       99       1162       0         6.       Zm0001d016867_T001       98       914       0         7.       Zm0001d017160_T001       23       78       3.00 × 10 <sup>31</sup> Setect   % Sequence   Å FASTA   Å Alignment         Setect   % Sequence   Å FASTA   Å Alignment         Graphical overview of aligning region(s)	BLASTN: 1 query, 1 database Query:: Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits Alignment of all hits Alignment of selected hit(s) Standard tabular report Full tabular report	Databases (13155 sequences, 19935516 characters): Zea_mayi_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-match -3, gap-open 5, gap Circos Query= Query_1 Graphical overview of hits E Length distribution of hits # Similar sequences 1. Zm00001d000357_T001 2. Zm00001d016868_T001 100 12	L soci L rec 3 2 2 2 0 00 0	Identity (%
6.         Zmo0001d018667_T001         98         914         0           7.         Zmo001d017160_T001         23         78         3.00 × 10 <sup>-11</sup> 7.         Zm0001d017160_T001         23         78         3.00 × 10 <sup>-11</sup> 8.         Zm0001d017160_T001         cds dromosome873_RefGen_v45:1872118926-1 geneZm00001d017160 gene_biotypeprotein_coding transcript_biotypeprotein_coding description=Putative cyclin A3-1         Image: Sequence   d= FASTA   d= Allgimment           Image: Graphical overview of aligning region(s)         d=         2 % G = d         d=         3 d         d=	BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits Alignment of all hits Alignment of selected hit(s) Standard tabular report Full tabular report	Databases (13155 sequences, 19935516 characters): Zea_mayi_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-match -3, gap-open 5, gap Circos Query= Query_1 Graphical overview of hits Length distribution of hits # Similar sequences 1. Zrm00001d00357_T001 2. Zm00001d016587_T001 3. Zm00001d016871_T001 99 111	A total A free           1         1           2         0           72         0	Identity (% 10/ 9!
7.         Zm00001d017160_T001         23         78         3.00 × 10 <sup>-11</sup> ***         Zm00001d017160_T001 eds chromosome.873_BetGen_v45:1872178970:187218926-1 gene:Zm00001d017160 pre_biotype:grotein_coding transcript_biotype:grotein_coding description=Putative cyclin A3.1         Select   * Sequence   & FASTA   & Alignment           ©         Graphical overview of aligning region(s)         & 2         2	BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits Alignment of all hits Alignment of all hits Standard tabular report Full tabular report	Databases (13155 sequences, 19935016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue 1e-05, sc-match 2, sc-match -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits E Length distribution of hits W Similar sequences Similar sequences Query coverage (%) Total so 1. Zmo0001d010587_T001 2. Zm00001d016871_T001 3. Zm00001d016871_T001 100 110 110 110 110 110 110	A two   A two           an         an	Identity (% 10) 9: 94
Zm00001d017160_T001_cts_thromosome#73_BrtGren_v45:187217973187218926-1 gene2m0001d017160 gene_biotype_protein_coding transcript_biotype_protein_coding description=Putative cyclin A3-1     Select   © Sequence   & FASTA   & Alignment     Graphical overview of aligning region(s)	BLASTN: 1 query, 1 database Query:: Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits Alignment of all hits Alignment of selected hit(s) Standard tabular report Full tabular report	Databases (13155 sequences, 19935016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-match -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits B Length distribution of hits W Similar sequences 1. Zmoooold000397_T001 1. Zmoooold01686_T001 1. Zmoooold01686_T001 1. Zmoooold016870_T001 1. Zmoooold016870_T001 1. Zmoooold016870_T001 1. Zmoooold016870_T001 1. Zmoooold016870_T001 1. Zmoooold016870_T001 1. Zmoooold016870_T001 1. Zmoooold00039_T001 2. Zmoooold016870_T001 2. Zmoooold016870_T001 2. Zmoooold016870_T001 3. Zmoooold016870_T001 3. Zmoooold00039_T001 3. Zmooold00039_T001 3. Zmoooold00039_T001 3. Zmoooold00039_T001 3. Zmooold00039_T001 3. Zmooold0	A tool A root ore E value 24 0 00 0 72 0 68 0 62 0	Identity (% 10) 9) 94 94 94
Cm00001d01/160_r100_r1cs_etable_rule_star21193718718721892et-1genez/2n00001d01/160     gene_biotypeprotein_coding transcript_biotypeprotein_coding description=Putative cyclin A3-1     Select   © Sequence   & FASTA   & Alignment     Graphical overview of aligning region(s)     \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$	BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits Alignment of all hits Alignment of selected hit(s) Standard tabular report Full tabular report	Databases (13155 sequences, 19935016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-mismatch -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits B Length distribution of hits W Similar sequences 2 Zmo0001d000397_T001 1 Zm00001d0087_T001 2 Zm00001d01687_T001 3 Zm00001d00388_T001 9 9 11 5 Zm00001d00388_T001 9 9 11 6 Zm00001d00388_T001 9 9 11 100 100 100 100 100 100 100	List) List re Evalue 24 0 00 0 72 0 68 0 62 0 14 0	Identity (% 100 99 91 91 91 91 91 91 85
Graphical overview of aligning region(s)     S	BLASTN: 1 query, 1 database Query:: Query1 Download FASTA, XML, TSV FASTA of all hits FASTA of selected hit(s) Alignment of all hits Alignment of selected hit(s) Standard tabular report Full tabular report	Databases (13155 sequences, 19935016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-match -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits B Length distribution of hits W Similar sequences 1. Zmoooold00037_T001 1. Zmoooold01037_T001 1. Zmoooold016870_T001 1. Zmoooold016870_T001 1. Zmoooold016870_T001 1. Zmoooold016870_T001 1. Zmoooold016870_T001 1. Zmoooold016870_T001 1. Zmoooold016870_T001 1. Zmoooold016870_T001 1. Zmoooold016870_T001 2. Zmoooold01687_T001 2. Zmooold01687_T001 2. Zmoooold01687_T001 2. Zmooool	A total A rec           ore         E value           24         0           00         0           72         0           68         0           62         0           14         0           76         3.00×10 <sup>+11</sup>	Identity (% 10) 9) 94 94 94
	BLASTN: 1 query, 1 database Query:: Query1 Download FASTA, XML, TSV FASTA of all hits FASTA of selected hit(s) Alignment of all hits Alignment of selected hit(s) Standard tabular report Full tabular report	Databases (13155 sequences, 19935516 characters): Zea_mayi_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-match -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits E Length distribution of hits # Similar sequences 2, Zmo0001d016870_T001 2, Zm00001d016870_T001 4, Zm00001d016870_T001 5, Zm00001d016870_T001 6, Zm00001d016870_T001 6, Zm00001d016870_T001 6, Zm00001d016870_T001 6, Zm00001d016870_T001 6, Zm00001d016870_T001 6, Zm00001d017160_T001 6, Zm00001d017160_T001 6, Zm00001d017160_T001 6, Zm00001d017160_T001 6, Zm00001d017150_T001 6, Zm00001d017150_T001 7, Zm00001d017150_T001 7, Zm00001d017150_T001 7, Zm0000	A total A free           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           b         b           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           a         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a	Identity (% 100 99 91 91 91 91 91 91 85
- 10 km	BLASTN: 1 query, 1 database Query:: Query1 Download FASTA, XML, TSV FASTA of all hits FASTA of selected hit(s) Alignment of all hits Alignment of selected hit(s) Standard tabular report Full tabular report	Databases (13155 sequences, 199355016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-match -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits Length distribution of hits Similar sequences y Similar sequences Query coverage (%) Total so 1. Zrm0001d015871_T001 1. Zrm0001d015871_T001 1. Zrm0001d015871_T001 1. Zrm0001d015871_T001 1. Zrm0001d015871_T001 1. Zrm0001d015871_T001 2. Zrm0001d015871_T001 2. Zrm0001d015871_T001 2. Zrm00001d015871_T001 2. Zrm0001d015871_T001 2. Zrm0001d01587_Zrm001 2. Zrm0001d01587_Zrm001 2. Zrm0001d01587_Zrm001 2. Zrm0001d01587_Zrm001 2. Zrm0001d01587_Zrm001 2. Zrm0001d01587_Zrm001 2. Zrm0001d01587_Zrm001 2. Zrm0001d01587_Zrm001 2. Zrm0001d01587_Zrm001 2. Zrm001587_Zrm001 2. Zrm001587_Zrm001587_Zrm001 2. Zrm001587_Zrm001587_Zrm001587_Zrm001587_Z	A total A free           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           b         b           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           a         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a	Identity (% 100 99 91 91 91 91 91 91 85
	BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits Alignment of all hits Alignment of all hits Standard tabular report Full tabular report	Databases (13155 sequences, 199355016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue 1e-05, sc-match 2, sc-match -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits E Length distribution of hits Similar sequences Very coverage (%) Total so 1. Zmocooldou397_T001 1. Zmocooldou397_T001 1. Zmocooldou388_T001 1. Zmocooldou388_T001 2. Zmocooldou388_T001 2. Zmocooldou388_T001 2. Zmocooldou388_T001 2. Zmocooldou388_T001 2. Zmocooldou387_T001 2. Zmocooldou388_T001 2. Zmocooldou387_T001 2. Zmocooldou3757_T001 2. Zmocooldou387_T001 2. Zmocooldou387_T001 2. Zm	L tots   A rec           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           b         a <td>Identity (% 100 99 91 91 91 91 91 91 85</td>	Identity (% 100 99 91 91 91 91 91 91 85
	BLASTN: 1 query, 1 database Query:: Query1 Download FASTA, XML, TSV FASTA of all hits FASTA of selected hit(s) Alignment of all hits Alignment of selected hit(s) Standard tabular report Full tabular report	Databases (13155 sequences, 190355016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-match -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits B Length distribution of hits D Length distribution of hits B Length distribution of hits B Length distribution of hits C zmocooldous88_T001 100 12. Zmocooldous88_T001 100 11. Zmocooldous88_T001 100 100 100 100 100 100 100	L tots   A rec           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           b         a <td>Identity (% 100 99 94 94 83 7.</td>	Identity (% 100 99 94 94 83 7.
	BLASTN: 1 query, 1 database Query:: Query1 Download FASTA, XML, TSV FASTA of all hits FASTA of selected hit(s) Alignment of all hits Alignment of selected hit(s) Standard tabular report Full tabular report	Databases (13155 sequences, 190355016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-match -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits B Length distribution of hits D Length distribution of hits B Length distribution of hits B Length distribution of hits C zmocooldous88_T001 100 12. Zmocooldous88_T001 100 11. Zmocooldous88_T001 100 100 100 100 100 100 100	L tots   A rec           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           b         a <td>Identity (% 100 99 94 94 83 7.</td>	Identity (% 100 99 94 94 83 7.
	BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits Alignment of all hits Alignment of selected hit(s) Standard tabular report Full tabular report	Databases (13155 sequences, 190355016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-match -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits B Length distribution of hits D Length distribution of hits B Length distribution of hits B Length distribution of hits C zmocooldous88_T001 100 12. Zmocooldous88_T001 100 11. Zmocooldous88_T001 100 100 100 100 100 100 100	L tots   A rec           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           b         a <td>Identity (% 100 99 94 94 83 7.</td>	Identity (% 100 99 94 94 83 7.
a. Score: 71.62 (78), E value: 3.00 x 10 <sup>-11</sup> , Identities: 104/145 (71.728), Gaps: 8/145 (5.528), Strand: + / +	BLASTN: 1 query, 1 database Query:: Query1 Download FASTA, XML, TSV FASTA of all hits FASTA of selected hit(s) Alignment of all hits Alignment of selected hit(s) Standard tabular report Full tabular report	Databases (13155 sequences, 190355016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue 1e-05, sc-match 2, sc-match -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits B Length distribution of hits B Similar sequences Query coverage (%) Total so 1. Zrm00001400357_T001 2. Zrm000014016570_T001 3. Zrm000014016570_T001 4. Zrm000014016570_T001 5. Zrm00001401571_T001 6. Zrm000014017160_T001 7. Zrm000014017160_T001 99 11 6. Zrm000014017160_T001 98 5 7. Zrm00014017160_T001 98 5 8 2 2 2 2 2 2 2 2 2 2 2 2 2	L tots   A rec           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           b         a <td>Identity (% 100 99 94 94 83 7.</td>	Identity (% 100 99 94 94 83 7.
	BLASTN: 1 query, 1 database Query:: Query1 Download FASTA, XML, TSV FASTA of all hits FASTA of selected hit(s) Alignment of all hits Alignment of selected hit(s) Standard tabular report Full tabular report	Databases (13155 sequences, 199355016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-match -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits Length distribution of hits Similar sequences y Similar sequences Query coverage (%) Total so 1. Zrm0001d015871_T001 1. Zrm0001d015871_T001 1. Zrm0001d015871_T001 1. Zrm0001d015871_T001 1. Zrm0001d015871_T001 1. Zrm0001d015871_T001 2. Zrm0001d015871_T001 2. Zrm0001d015871_T001 2. Zrm00001d015871_T001 2. Zrm0001d015871_T001 2. Zrm0001d01587_Zrm001 2. Zrm0001d01587_Zrm001 2. Zrm0001d01587_Zrm001 2. Zrm0001d01587_Zrm001 2. Zrm0001d01587_Zrm001 2. Zrm0001d01587_Zrm001 2. Zrm0001d01587_Zrm001 2. Zrm0001d01587_Zrm001 2. Zrm0001d01587_Zrm001 2. Zrm001587_Zrm001 2. Zrm001587_Zrm001587_Zrm001 2. Zrm001587_Zrm001587_Zrm001587_Zrm001587_Z	A total A free           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           b         b           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           a         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a           b         a	Identity (% 100 99 91 91 91 91 91 91 85
	BLASTN: 1 query, 1 database Query:: Query1 Download FASTA, XML, TSV FASTA of all hits FASTA of selected hit(s) Alignment of all hits Alignment of selected hit(s) Standard tabular report Full tabular report	Databases (13155 sequences, 190355016 characters): Zea_mays_cultivar_B73_cds Parameters: task blastn, evalue le-05, sc-match 2, sc-match -3, gap-open 5, gap Circos Query=Query_1 Graphical overview of hits B Length distribution of hits D Length distribution of hits B Length distribution of hits B Length distribution of hits C zmocooldous88_T001 100 12. Zmocooldous88_T001 100 11. Zmocooldous88_T001 100 100 100 100 100 100 100	L tots   A rec           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           a         a           b         a <td>Identity (% 100 99 94 94 83 7.</td>	Identity (% 100 99 94 94 83 7.

#### 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 sequences (1) with reports available in different formats. Also included are interactive plot views of each alignment including Circos plots (2), NCBI BLAST-like alignment hits visualization (3), and length distribution of hits (4). Each alignment hit has detailed alignment information, including a graphic view of the aligned regions and detailed alignments (5).

#### ACKNOWLEDGMENTS

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). 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.

# **iScience**

Article

#### **AUTHOR CONTRIBUTIONS**

Conceptualization, N.Y., L.Y., and S.G.; Methodology, S.G., L.Y., and J. Li; Software, S.G., X.X., and J. Yuan; Resources, J. Luo, L.C., W.L., and Y.W.; Formal Analysis, S.G., X.Y., X.X., J. Yuan, S.W., S.L., and Y.Z.; Writing - Original Draft, S.G. and N.Y.; Writing - Review & Editing, S.G., N.Y., and J. Yan; Funding Acquisition, J. Yan and N.Y.; Supervision, J. Yan and Q.G.

#### **DECLARATION OF INTERESTS**

The authors declare no competing interests.

Received: April 3, 2020 Revised: May 21, 2020 Accepted: June 1, 2020 Published: June 26, 2020

#### REFERENCES

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 Biol. 17, 66.

Dekker, J. (2008). Gene regulation in the third dimension. Science *319*, 1793–1794.

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.

Gilad, Y., Rifkin, S.A., and Pritchard, J.K. (2008). Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet. *24*, 408–415.

He, Y., and Li, Z. (2018). Epigenetic environmental memories in plants: establishment, maintenance, and reprogramming. Trends Genet. 34, 856–866.

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 pangenome and pan-transcriptome. Plant Cell *26*, 121–135.

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. Nat. Commun. *9*, 2738.

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.

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.

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 Res. *32*, D393–D397. Lee, T., Lee, S., Yang, S., and Lee, I. (2019). MaizeNet: a co-functional network for network-assisted systems genetics in *Zea mays*. Plant J. *99*, 571–582.

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. Mol. Plant 12, 402–409.

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. Nat. Genet. 45, 43.

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.

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 Phytol. 210, 1083–1094.

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. Nat. Commun. 10, 2632.

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 multigenome genetics and genomics database. Nucleic Acids Res. 47, D1146–D1154.

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. Mol. Biol. Evol. *msz185*, 1–6.

Rajasundaram, D., and Selbig, J. (2016). More effort—more results: recent advances in integrative 'omics' data analysis. Curr. Opin. Plant Biol. 30, 57–61.

Rodgers-Melnick, E., Vera, D.L., Bass, H.W., and Buckler, E.S. (2016). Open chromatin reveals the

functional maize genome. Proc. Natl. Acad. Sci. U S A 113, E3177–E3184.

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. Nat. Genet. *50*, 1282.

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. Nat. Genet. 50, 1289.

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.

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. Annu. Rev. Genet. 52, 421–444.

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.

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 singlemolecule sequencing. Genome Res. 28, 921–932.

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. Nat. Commun. *5*, 3438.

Xu, G., Cao, J., Wang, X., Chen, Q., Jin, W., Li, Z., and Tian, F. (2019a). Evolutionary metabolomics identifies substantial metabolic divergence between maize and its wild ancestor, teosinte. Plant Cell 31, 1990–2009.

Xu, J., Chen, G., Hermanson, P.J., Xu, Q., Sun, C., Chen, W., Kan, Q., Li, M., Crisp, P.A., Yan, J., et al. (2019b). Population-level analysis reveals the widespread occurrence and phenotypic





consequence of DNA methylation variation not tagged by genetic variation in maize. Genome Biol. *20*, 243.

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. Nat. Genet. *51*, 1052.

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. Nat. Commun. *8*, 1874. 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. Mol. Breed. *28*, 511–526.

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 Res. *34*, D752– D757. 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 Phytol. 221, 2096–2111.

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. Nat. Methods 10, 375.

## iScience Article

iScience, Volume 23

# **Supplemental Information**

# ZEAMAP, a Comprehensive Database

# Adapted to the Maize Multi-Omics Era

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

# **Supplemental Files**

### **Supplemental Figures**

Figure S1. Sketch of ZEAMAP database structure.
Figure S2. Species detailed page of *Zea mays* cultivar B73.
Figure S3. Functional annotation information in each gene's detailed page.
Figure S4. Feature search and gene search functions in ZEAMAP.
Figure S5. Features of ZEAMAP variations module.
Figure S6. Search trait function in ZEAMAP.
Figure S7. GWAS table browser in ZEAMAP.
Figure S8. GWAS-Multi-Trait visualization tool in ZEAMAP.
Figure S9. GWAS-Locus visualization tool in ZEAMAP.
Figure S10. eQTL table browser in ZEAMAP.
Figure S11. Genetic marker search functions in ZEAMAP.
Figure S12. Crispr sgRNA function in ZEAMAP.

### **Transparent Methods**

**Supplementary References** 

# **Supplemental Figures**



**Figure S1. Sketch of ZEAMAP database structure, Related to Figure 1.** The multi-omics data were stored into database and managed by chado sequence ontology and tripal modules. Two genome browsers, custom table browsers and visualization tools were developed for data querying and visualization. The frontend of ZEAMAP were divided into five main biological sections (Genomics, Variations, Genetics Populations and Epigenetics), and miscellaneous tools including site-wide search engine, BLAST server, Crispr sgRNA searcher and FTP Download server.

Zea mays cult	tivar:	B73	Site Wide Search Q
Summary	>	SUMMARY	
Annotations		Resource Type	Organism
Cross Reference		Abbreviation	873
Publication		Genus	Zea
Relationship		Species	mays cultivar:B73
		Common Name	maize 873
		Description	The maize inbred line B73 is a represent germplasm of the Reid yello dent maize variety group. Its reference genome has been updated several times since its initial release in 2009. The genome assembly used here is <b>B73_RefGen_v4.42</b> .
			View more about the germplasm information at GRIN Download the genome assembly used in this databse from Gramene
			Download the genome annotations used in this databse from Gramene
		Organism Image	

**Figure S2. Species detailed page of Zea mays cultivar B73, Related to Figure 2.** The species page shows the general information about the current species/germplasm, including brief introductions, external links to the germplasm information and the accessions of related omics data.



**Figure S3. Functional annotation information in each gene's detailed page, Related to Figure 2.** (A) The "Functional annotation" tag in the detailed page of each gene/mRNA/ protein (taking the detailed page of protein Zm00001d044266\_P013 for example here), the annotated items within each database were linked to their detailed information page (B-F). A specific page was developed to get the detailed information of the related ortholog group including the phylogenetic relationships, the gene ontology items, and the sequence alignments within the ortholog group from EggNOG database (G).

			_	45 records were returned						Download Table
		and the sequence source. To select multiple can be downloaded in FASTA or CSV tabular		Name	Uniquename	Туре	Organism	Source		Location
format.			1	Zm00015a011942_P001	Zm00015a011942_P001	polypeptide	Zea mays cultivar:SK	Zea mays cultivar:SK Whole Gen Annotation	nome Assembly and	chr3: 17424839 17440222
Species	Any Zea mays cultivar:873		2	ZMex08g026281	ZMex08g026281	gene	Zea mays ssp. mexicana	Zea mays ssp. mexicana Genom Annotation	ne Assembly and	8: 10014738 1001
	Zea mays cultivar:Mo17 Zea mays cultivar:SK 🔻		3	Zm00014a009861_P002	Zm00014a009861_P002	polypeptide	Zea mays cultivar:Mo17	Zea mays cultivar:Mo17 RefSeq	Genome	chr5: 211530022 211533543
Туре	Any ^		4	Zm00001d021846_T015	Zm00001d021846_T015	mRNA	Zea mays cultivar:873	Zea mays cultivar:B73 RefSeq G	Senome	7: 164385354 164417958
	mRNA polypeptide ~		5	Zm00015a023356	Zm00015a023356	gene	Zea mays cultivar:SK	Zea mays cultivar:SK Whole Gen Annotation	nome Assembly and	chr5: 143053835 143058464
Source	Any		6	ZMex06t021864_P01	ZMex06t021864_P01	polypeptide	Zea mays ssp. mexicana	Zea mays ssp. mexicana Genom Annotation	ne Assembly and	6: 29006050 2901
	Zea mays cultivar:B73 RefSeq Genome Zea mays cultivar:Mo17 RefSeq Genom Zea mays cultivar:SK Whole Genome A		7	Zm00014a010455	Zm00014a010455	gene	Zea mays cultivar:Mo17	Zea mays cultivar:Mo17 RefSeq	Genome	chr4: 243182415 243183690
Location	Any * between	and	8	Zm00015a004640	Zm00015a004640	gene	Zea mays cultivar:SK	Zea mays cultivar:SK Whole Gen Annotation	nome Assembly and	chr1: 239094036 239095412
Name	contains •		9	ZMex01t001328_T01	ZMex01t001328_T01	mRNA	Zea mays ssp. mexicana	Zea mays ssp. mexicana Genom Annotation	ne Assembly and	1: 29525762 2953
			10	Zm00014a029257_T001	Zm00014a029257_T001	mRNA	Zea mays cultivar:Mo17	Zea mays cultivar:Mo17 RefSeq	I Genome	chr8: 22958439 22962152
File Upload	Choose File No file chosen		11	Zm00001d015722_P001	Zm00001d015722_P001	polypeptide	Zea mays cultivar:B73	Zea mays cultivar:873 RefSeq G	Senome	5: 112752607 112755567
Provide seque	nce names in a file. Separate each name by a n	ew line.		Zm00001d035004_T032	Zm00001d035004_T032	mRNA	Zea mays cultivar:B73	Zea mays cultivar:B73 RefSeq G	Senome	6: 1572540 15784
			13	Zm00014a044091	Zm00014a044091	gene	Zea mays cultivar:Mo17	Zea mays cultivar:Mo17 RefSeq	Genome	chr4: 203693418 203694704
Search by							Term			
File Upload	Name contains   Choose File No file chosen enames in a file. Separate each name by a new line.	(e.g. act)	1	Zm00001d027402_P004	Zea mays 213 cultivar:873	polypeptide	ONL93052 protein	N 3_Zm00001d027402 cl 9	"Helical backbone" me NAMED. pep chromosome:873_Ref0 gene:Zm00001d02740	5en_v4:1:4417837:44 2
File Upload	Choose File No file chosen names in a file. Separate each name by a new line.	(eg.10)	1		cultivar:873		ONL93052 protein ZEAMMB7 [Zea mays	N 3_Zm00001d027402 cl 9 tr 9 tr 9 tr	NAMED. pep chromosome:B73_Ref0 gene:Zm00001d02740 transcript:Zm00001d0 gene_biotype:protein_ transcript_biotype:pro	Gen_v4:1:4417837:44 2 27402_T004 coding tein_coding
File Upload Provide sequence Search by Source	Choose Tell No file chosen entres in a file. Separate each name by a new line. Assembly Dear mays cultivart \$73 Brefseq Genome bister Dear mays cultivart \$75 Br	a re-	1	Zm00001d027402_P004 Zm00001d027402_P004	cultivar:873	polypeptide	ONL93052 protein ZEAMMB7 [Zea mays ONL93052 protein	N 3_Zm00001d027402 ft 1 hypothetical 3_Zm00001d027402 g 9 9	VAMED. pep chromosome: 873_Ref( gene:Zm0000102; ranscript:Zm0000100; gene_biotype:protein_ ranscript_biotype:prot 'Helical backbone' me VAMED. pep chromosome: 873_Ref( gene:Zm00001002740	Sen_v4:1:4417837:44 2 27402_1004 coding tein_coding tal receptor. FAMILY Sen_v4:1:4417837:44 2
File Upload Provide sequence Search by	Choose File No file chosen annes in 3 file Sponte each rame by a rev Ins. Assembly Les mays cuttion 737 British Genome Au Bater Zea mays cuttion 737 British Genome Au Exe mays cuttion 737 British British Genome Au Exe mays cuttion 737 British British Genome Au Exe mays cuttion 737 British Brit	a re-	2	Zm00001d027402_P004	cultiva:873 Zea mays 213 cultiva:873	polypeptide	ONL93052 protein ZEAMM87 [Zea mays ONL93052 protein ZEAMM [Zea mays]	N 3,2m00001d027402 9 1 1 hypothetical 3,2m00001d027402 9 1 1 9 1 1 9 1 1 9 1 1 1 9 1 1 1 9 1 1 1 9 1 1 1 9 1 1 1 9 1	vAMED. pep chromosome:873_Ref( gene:Zm00001d02740 transcript:Zm00001d02 gene_biotypeprotein_ transcript_biotypeprot 'Helical backbone' me vAMED. pep chromosome:873_Ref( gene:Zm00001d02740 transcript:Zm00001d02740 transcript_biotype:protein_ transcript_biotype:protein_	5en_v4:1:4417837:44 27402_T004 coding tal receptor. FAMILY 6en_v4:1:4417837:44 2 27402_T004 coding tein_coding
File Upload Provide sequence Search by Source Type	CONNETIEN No Te de charant reneral est de la serie est a rener les y a ren ten. Automotion est a fait de la serie de la serie de la Reneral de la serie de la s	a II or metry and Annotation yand Annotation	2		cultiva:873 Zea mays 213 cultiva:873		ONL93052 protein ZEAMM2 Zea mays ONL93052 protein ZEAMM2 Zea mays ONL93052 protein ZEAMM2	X, Zm000014027402     d	VAMED, pep chromosome: 873, 8/efG pere-Zm00001 d02740 transcript_Zm0001 d02740 ranscript_biotype protein ranscript_biotype protein chromosome:873, 8/efG pene-Zm00001 d02740 gene_biotype protein, amscript_Zm0001 d02740 gene_biotype protein, thoronsome:873, 8/efG	Sen_v4:1:4417837:44 2 27402_T004 coding ttal receptor. FAMILY Sen_v4:1:4417837:44 2 27402_T004 coding tein_coding tal receptor. FAMILY Sen_v4:1:4417837:44
File Upload Provide sequence Search by Source Type Search by GO Term	CONNETIEN To The Character transmise Ada Sector and the same transmission A second by the Character 27 Medices Connerses there are an explorational Connerses there are a the explorational Connerses the explorational Conner	If m * * * * * * * * * * * * * * * * * *	2	Zm00001d027402_P004	cultiva::873 Zea mays 213 Cultiva::873	polypeptide	ONL93052 protein ZEAMM8 [Zea mays] protein ZEAM87 [Zea mays] ONL93052 protein	1 hypothetical 7 1 hypothetical 7 1 hypothetical 7 1 hypothetical 7 1 hypothetical 7 1 hypothetical 7 1 hypothetical 7 9 9 9 9 9 9 9 9 9 9 9 9 9	VAMED. pep thomosome 373, Beff gene:Zm00001 d02740 transcript:Zm0001 d02740 theical backbone" m VAMED. pep thomosome 373, Beff pere:Zm00001 d02740 thical backbone" m VAMED. pep thomosome 373, Beff pere:Zm0001 d02740 thical backbone" m VAMED. pep thomosome:873, Beff pere:Zm0001 d02740 thical backbone" m VAMED. pep	Een_v4:1:4417837:44 2 27402_T004 coding tein_coding tal receptor. FAMILY 2 27402_T004 coding tein_coding tein_coding tein_coding telreceptor. FAMILY Sen_v4:1:4417837:44 2 27402_T004 coding
File Upload Provide sequence Search by Source Type Search by	COUNCE THE No THE character Interview is Allia. State of a character by a real total Interview is Allia. State of a character by a real total Interview is Allia. State of a character by a real total Interview is a character by a real total I	a II or metry and Annotation yand Annotation	2	Zm00001d027402_P004	cultivar.873 Zea mays 213 cultivar.873 Zea mays 213 cultivar.873	polypeptide	ONL93052 protein ZEAAMAE Zea mays protein ZEAMMB7 [Zea mays ONL93052 protein ZEAMMB7 [Zea mays]	Thypothetical     Thypothetical     Thypothetical     Thypothetical     Thypothetical     Thypothetical     Thypothetical     Thypothetical     Thypothetical	VAMED, pep chromosome:873, RefG pere:Zm00001d02740 transcript:Zm0001d02 Helical backbone" me VAMED, pep pere:Zm00001d02740 ranscript:Zm0001d02 pere:Zm00001d02 Helical backbone" me VAMED, pep chromosome:873, RefG pere:Zm00001d02740 ranscript:Zm00001d02740 ranscript:Zm00001d02740 ranscript:Zm000000 ranscript:Zm00000 ranscript:Zm00000 ranscript:Zm00	Een_v4:1:4417837:44 2 27402_1004 coding teln_coding teln_coding teln_coding teln_coding teln_coding teln_coding teln_coding teln_coding teln_coding teln_coding teln_coding teln_coding teln_coding

**Figure S4. Feature search and gene search functions in ZEAMAP, Related to Figure 2.** The feature search function (A) in ZEAMAP enables searching for all annotated features on certain genome assemblies by their locations and/or names. The locations of the search results were displayed and linked to genome browser for visualization. The gene search function (B) is dedicated to search for annotated genes, mRNAs and polypeptides by their names and/or their functional annotations, with the detailed functional annotations displayed in the search results for each record.

Population	maiz	e association mapping panel						•	0 20,000,000			000 000 120 000 000 ○ ○ ○ ○ 2 • 2	173715490 173716126 (637 b)			_
Variation Name					0				715.500 O Reference sequence	173.716.625		15750	173735875		. 716.000	
Variation Class	SNP			Variation Type	Am				© Gene							
Genome	Any								© 107 genomes she feat	energy lange mar	15620 Jan 2 3	0124_173715712	01/2 1 173715861	1712 4 173715937	01/2 4 173716001	
Chr/Scaffold	Ann	• between		1.8	nd	bp				SNY1-C	173715836 UN2.1, 1737					
Maf	betwee		and							01/2 1 17 17 15576 01/2 SNV C-+ T		2 8_173715736 VA-+ 0		chr2.s_173715915 SNV G -+ A		
Variation effect										01/2 6 173715504 SNV T A	042 6 173715478	042 4, 173715745 SNV 0 - A		In mail of the mail		
Variation effect	Anv			W. Martin				<u> </u>		0H2.8_173735591 SNV T-+ 0		0h2 s 173715765 5NV C = 0		NS IN A STATUS	4	
<i>a</i>	Search		Down	sload CSV		Cloar						5W G T				
Population	Ref_Genome	VariantID * VariantType	* Chr	the local division of	^ Alt ~ MAF			_	© 507 genomes ano mate				-	ME_H I Devolgion (RT)		
Maize_AMP Maize_AMP	873_V4	chr2.s_173715868 SNP chr2.s_173715916 SNP	2	173715868 A	G REER	< 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1				11/1/	~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~			Antiparti ater Antiparti Antiparti Antiparti	1000	
Maize AMP	873 V4	chr2.s 173715937 SNP	2	173715916 G	3 prime_UTR van	1978		-						- 44	M	
Maize AMP	873 V4	chr2.s 173715945 SNP	2	173715945 G	T									14100 14173 11101	- (A) - (A) - (A) - (A) - (A) - (A)	n
					A	• Ibros									41.01.42 41.147 41.147	n n
Maize_AMP	873_V4	chr2.s_173715946 SNP	2	173715946 C												
Maize_AMP Maize_AMP	B73_V4 B73_V4	chr2.s_173715946 SNP chr2.s_173716001 SNP	2	173715946 C 173716001 T	C .					-			- 23			
Maize_AMP			2				se internet	ALT -	04K5672	04K5686	04K5702	05W002	05WN230	- 07KS4	0.00002	*
Msize_AMP	B73_V4 hotype Search			173716001 T	C	• Jbrov	REF	ALT –	04K5672 4	04K5686	04K5702 ~	05W002	05WN230			
Maize_AMP ZEAMAP Ger Browse the genotype internation.	B73_V4 notype Search of cetain gemplasma.3	chr2.s_173716001 SNP		173716001 T	c Position	<ul> <li>Jbrow</li> <li>diamondal</li> </ul>	REF C	Sec. 8						- 07KS4	- 1323	
Maize_AMP ZEAMAP Gen Internation Choose datar Dataset * MaloALIP Jugi	B73_V4 notype Search of certain gemplasms set and geemplasms	chr2.s_173716001 SNP	me of geringi	173716001 T	C Position 1:2356	ID chr1.s_2356	REF C 5 C	T	0 0	0 1	1 1	1[1	1 1	<ul> <li>07KS4</li> <li>0 1</li> </ul>	* <b>1323</b> 0 1	*
Maize_AMP ZEAMAP Gen Internation Choose datar Dataset * MaloALIP Jugi	B73_V4 notype Search of certain gemplasms set and geemplasms	chr2.4_173716001 SNP	me of geringi	173716001 T	C Position 1:2356 1:28115	ID // 1D // 10 //	REF C 5 C 1 A	T T	0 0	0 1 0 0	1 1 0 0	1 1 0 1	1 1 0 1	<ul> <li>07KS4</li> <li>0 1</li> <li>0 1</li> </ul>	* 1323 0 1 0 1	*
Maize_AMP ZEAMAP Ger Internation. Choose datas Dataset * MaloxAT Japa Select a set of ge Germplasm*	873_V4 notype Search of certain germplasms set and germplasms toppe_tag.variast. •] metic variants. Learn noo	chr2a_1713716001 SNP ingly select a distance and then selece the na- roy our are interested in:	me of geringi sis soci	173716001 T	C Position 1:2356 1:28115 1:28701	ID chr1.s_2356 chr1.s_2811 chr1.s_2870	REF           C           5           1           A           9	T T G	0 0 0 0 1 1	0 1 0 0 1 1	1 1 0 0 1 1	1 1 0 1 1 1	1 1 0 1 1 1	<ul> <li>07KS4</li> <li>0 1</li> <li>0 1</li> <li>1 1</li> </ul>	<ul> <li>1323</li> <li>0 1</li> <li>0 1</li> <li>1 1</li> </ul>	*
Maize_AMP ZEAMAP Ger Internation. Choose datas Dataset * MaloxAT Japa Select a set of ge Germplasm*	873_V4 notype Search of certain germplasms set and germplasms toppe_tag.variasts.*] metic variants. Learn noo	chr2.s_173716001 SNP	me of geringi sis soci	173716001 T	C Position 1:2356 1:28115 1:28701 1:33879	<ul> <li>ID</li> <li>chr1.s_2356</li> <li>chr1.s_2811</li> <li>chr1.s_2870</li> <li>chr1.s_3877</li> <li>chr1.s_3495</li> </ul>	REF           C           5           C           1           A           9           C           1           G	T T G T	0 0 0 0 1 1 0 0	0 1 0 0 1 1 0 0	1 1 0 0 1 1 0 0	1 1 0 1 1 1 1 1	1 1 0 1 1 1 1 1	<ul> <li>о7кs4</li> <li>о[1</li> <li>0[1</li> <li>1[1</li> <li>1]1</li> </ul>	<ul> <li>1323</li> <li>0 1</li> <li>0 1</li> <li>1 1</li> <li>1 1</li> </ul>	*
Matter, AMP ZEAMAP Gent These the grantinge distancion. Choose data Dataset * MaterNAP Just Gentra at of op Ge	873_V4 notype Search of certain germplasms set and germplasms toppe_tag.variasts.*] metic variants. Learn noo	chr24,173716001 540 Imply when a disease, and then error the re- you are interested in.	me of geringi sis soci	173716001 T	C Position 1:2356 1:28715 1:28701 1:33879 1:34951 1:35636	<ul> <li>ID</li> <li>chr1.s_2356</li> <li>chr1.s_2810</li> <li>chr1.s_2870</li> <li>chr1.s_387</li> <li>chr1.s_3495</li> <li>chr1.s_563</li> </ul>	REF           C           5           C           1           A           9           C           1           G           C	T T G T A T	0 0 0 0 <b>1 1</b> 0 0 0 0 0 0	0 1 0 0 1 1 0 0 0 0 0 0	1 1 0 0 1 1 0 0 1 1 0 0	1 1 0 1 1 1 1 1 0 0 0 0	1 1 0 1 1 1 1 1 0 0 0 0	<ul> <li>07KS4</li> <li>0[1</li> <li>0[1</li> <li>1[1</li> <li>1[1</li> <li>0[0</li> <li>0[0</li> </ul>	* 1323 0 1 0 1 1 1 1 1 0 0 0 0	*
ZEAMAP Ger Trave the penatype information Choose data Distant <sup>1</sup> MetaAtty Jug Selected germpli	873,344 Intrype Search of testals geregations. 7 let and geregations. 7 integer signatures. 7 Integer signatu	chr24,173716001 540 Imply when a disease, and then error the re- you are interested in.	me of geringi sis soci	173716001 T	C Position 1:2356 1:28115 1:28701 1:33879 1:34951 1:35636 1:35864	ID chr1.s_2356 chr1.s_2870 chr1.s_3877 chr1.s_3495 chr1.s_3586 chr1.s_3586	REF           C           5         C           1         A           9         C           1         G           6         C           4         G	T T G T A T A	0 0 0 0 1 1 0 0 0 0 0 0 1 1	0 1 0 0 1 1 0 0 0 0 0 0 0 0	1 1 0 0 1 1 0 0 1 1 0 0 1 1	1 1 0 1 1 1 1 1 0 0 0 0 0 0	1 1 0 1 1 1 1 1 0 0 0 0 0 0	<ul> <li>O7KS4</li> <li>O[1</li> <li>O[1</li> <li>I]1</li> <li>I]1</li> <li>I]1</li> <li>O[0</li> <li>O[0</li> <li>O[0</li> </ul>	1323     0 1     0 1     1 1     1 1     0 0     0 0     0 0     0 0	*
Maze_AMP ZEAMAP Gen These the genetype internation Choose data Dataset * MaterNAT2 Jung Genetypeary Genetypeary Genetypeary Genetypeary Restrict to th	873,344 Intrype Search of testals geregations. 7 let and geregations. 7 integer signatures. 7 Integer signatu	dr22_172716021 500 any union global g	me of geringi sis soci	173716001 T	C Position 1:2356 1:28115 1:28701 1:33879 1:34951 1:35864 1:35864 1:35819	ID         Accession           chr1.s_2356         chr1.s_2870           chr1.s_3877         chr1.s_3875           chr1.s_3563         chr1.s_3563           chr1.s_3586         chr1.s_3586           chr1.s_32871         chr1.s_3586	REF           C           5         C           1         A           9         C           1         G           6         C           4         G           9         T	T T G T A T A C	0 0 0 0 1 1 0 0 0 0 0 0 1 1 0 0	0 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0	1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0	1 1 0 1 1 1 1 1 0 0 0 0 0 0 1 1	1 1 0 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1	<ul> <li>07KS4</li> <li>0[1</li> <li>0[1</li> <li>1[1</li> <li>1[1</li> <li>0]0</li> <li>0]0</li> <li>0]0</li> <li>0]0</li> <li>1[1</li> </ul>	1323 0 1 0 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1	
Maze, AMP ZEAMAP Ger Trease the genetype volumentar. Choose data Dataset MetaAMP Age Genetian	873,344 Intrype Search of testals geregations. 7 let and geregations. 7 integer signatures. 7 Integer signatu	dr22_172716021 500 any union global g	me of geringi sis soci	173716001 T	C Position 1:2356 1:28115 1:28701 1:33879 1:34951 1:35636 1:35864	ID chr1.s_2356 chr1.s_2870 chr1.s_3877 chr1.s_3495 chr1.s_3586 chr1.s_3586	<ul> <li>REF</li> <li>C</li> <li>C</li> <li>C</li> <li>C</li> <li>1</li> <li>A</li> <li>9</li> <li>C</li> <li>1</li> <li>G</li> <li>6</li> <li>C</li> <li>4</li> <li>G</li> <li>9</li> <li>T</li> <li>3</li> <li>A</li> </ul>	T T G T A T A	0 0 0 0 1 1 0 0 0 0 0 0 1 1	0 1 0 0 1 1 0 0 0 0 0 0 0 0	1 1 0 0 1 1 0 0 1 1 0 0 1 1	1 1 0 1 1 1 1 1 0 0 0 0 0 0	1 1 0 1 1 1 1 1 0 0 0 0 0 0	<ul> <li>O7KS4</li> <li>O[1</li> <li>O[1</li> <li>I]1</li> <li>I]1</li> <li>I]1</li> <li>O[0</li> <li>O[0</li> <li>O[0</li> </ul>	1323     0 1     0 1     1 1     1 1     0 0     0 0     0 0     0 0	*

Figure S5. Features of ZEAMAP variations module, Related to Figure 1. In ZEAMAP, variants could be displayed tabularly and filtered by IDs, locations and annotated effects (A), with each resulted record has links to its location on Jbrowse (B). The popup for each variant in Jbrowse shows the detailed information including annotations and genotype in each sample (inset in B). Jbrowse has also provided an additional track shows an overview of genotype matrix (red: reference genotype; light blue: homozygous genotype; light pink: heterozygous; grey: no call). (C) ZEAMAP has also provided a "Genotype Search" function to get the genotype information by genomics regions and germplasms of interest, with a resulted matrix shows the variant positions, their IDs, reference and alternative alleles and genotypes in each samples ( "0" for reference allele and "1" for alternative allele; red: homozygous genotype ).

		Qualitative	e Trait Quant	itative Trait							
	/	Trait 1 Trait 2 Trait 3	cobcolor • Any • Any •	Value 1 Value 2 Value 3	red  Any red white Any	AND •					
Qualita	ative Trait Quant	titative Trait									
Trait 1	100grainwei	ght •	between	11.7899		and	30.5579		AND •		
Trait 2	cobweight	٠	between	4.49766		and	31.4859		AND *		
Trait 3	kernellength	ı <b>T</b>	between	6.88922		and	10.9851				
Iralt 5											
Search	Reset rds were returned Germplasm		Species			100grain	weight	cobweight		Downloa kernellength	ıd Tab
Search 7 recor	rds were returned		<b>Species</b> Zea mays culti	ivar:B73		100grain 18.6644	weight	cobweight 18.7322	_	partie control of the	id Tab
Search 7 recor	rds were returned Germplasm						weight			kernellength	ıd Tak
earch 7 recor	rds were returned Germplasm 150	3	Zea mays culti	ivar:B73	_	18.6644	weight	18.7322		kernellength 8.71982	ıd Tab
Search 7 recor	rds were returned Germplasm 150 177	3	Zea mays culti Zea mays culti	ivar:B73 ivar:B73		18.6644 22.1752	weight	18.7322 16.6977		kernellength 8.71982 8.4085	id Tal
Search 7 recor	rds were returned Germplasm 150 177 238		Zea mays culti Zea mays culti Zea mays culti	var:B73 var:B73 var:B73	_	18.6644 22.1752 21.9434	weight	18.7322 16.6977 17.4338		kernellength 8.71982 8.4085 8.83602	ıd Tab

**Figure S6. Search trait function in ZEAMAP, Related to Figure 3.** Both qualitative and quantitative trait could be searched by their trait values with multiple filter conditions supported. The search result shows all the germplasms that passed the filter conditions and their triat values.

Population					٣
trait				0	divided by semicolon
Variant ID					divided by semicolon
Chromosome	Any	between		and	Ьр
- Log10(P value)	between		and		
S	earch		Download CSV		Clear
Sitrait	earch	^ chr	Download CSV	<ul> <li>-log(P Value)</li> </ul>	Clear
		^ chr		<ul> <li>-log(P Value)</li> <li>5.61376671730847</li> </ul>	
trait	<sup>▲</sup> variantID	chr	▲ posi		action ^
<b>trait</b> C180_C200	variantID chr1.s_1079408	1	<b>posi</b> 1079408	5.61376671730847	action visualize by trait; visualize by variant;

**Figure S7. GWAS table browser in ZEAMAP, Related to Figure 3.** The GWAS signals could be searched by traits, variant IDs and variant locations, and filtered by significant P values. Each record in the search result has links to the GWAS visualization tools.



**Figure S8. GWAS-Multi-Trait visualization tool in ZEAMAP, Related to Figure 3.** This tools displays GWAS signals of multiple traits, with logic similar to GWAS-Single-Trait tool (as indicated in Figure 2A). The only differences are that the traits here are multi-selectable, and the colors in the detailed scatter plot indicate variants for different traits rather than LDs. A "data layers" button has been added in the control panel of the detailed scatter plot to fade, hide, order or remove certain trait layers.



**Figure S9. GWAS-Locus visualization tool in ZEAMAP, Related to Figure 3.** This tool displays all significantly associated signals between the query variant and all available traits, with a yellow highlight line and a red dashed line respectively indicated the general and detailed position of the query variant.

ne								±.	divi	ided by sem	icolon
iromosome	Any	,	between	n			and				bp
stance tss	between				and						
eta value	between				and						
og10(pvalue)	between				and						
2	Search				Downlo	ad CSV					Clear
gene	position	-	genome 4	·	focus	^ abs_c	distance_tss	^ beta	value 1	• -I •	action 🔶
+ ENSRNA0494649	1:7044629-7	044731	B73_RefSeq	-	chr1.s_7	044 475		-0.67	'04	11.9992	visualize by gene ID;
- ENSRNA0494638	46 5:3203379-3	203539	B73_RefSeq	-	chr5.s_3	204 1261		-0.51	85	7.582	visualize by gene ID;
locus_id		≜ abs_d	istance_tss			betavalu	e		<u> </u>	log10(pva	lue)
filter column		Min		Max		Min	Ma	х	I	Min	Max
chr5.s_3203074		465				-0.4165			1	0.1494	
chr5.s_3204800		1261				-0.5185			7	.582	
		27177				0.4642			8	.433	
chr5.s_3230716											
chr5.s_3230716 chr5.s_3231267		27728				0.4684			9	.1799	

**Figure S10. eQTL table browser in ZEAMAP, Related to Figure 3.** Using this tool, eQTL signals could be filtered by gene IDs and locations, as well as the distance from transcription start site, the effect size (beta value) and the significance (p value) of the most significant variant within each gene. The search result shows one gene per record, with links to the visualization of each gene. Each record has a sub-table which lists all the *cis*-eQTL signals significantly associated with this gene.

dvanced Marker Search			# Name	Type Sp		Мар	Linkage Group	Genetic Location	Genomic Location
			1 abph1.15	SNP Ze	a mays cultivar:873	MO17xTeo_BC2F5	MO17xTeo_BC2F5.2	91.6	2.28867080.2886708
	· ·		2 abph1.15		na mays cultivar:873	ZHENG58xSK_F6	ZHENG58x5K_F6.2	81.6	2:28867080.2886708
Advanced Marker Search QTL Nearby Markers	Wild card (*) can be used to match any QTL la	label	3 abph1.15	SNP Ze	sa mays cultivar:873	MO17xTeo_BC2F5	MO17xTeo_8C2F5.2	91.6	2-28867080.2886708
Advanced Marker Search QIL Nearby Markers	→ Wild Calld ( ) call be died to match any Gre in	later.	4 abph1.15	SNP Ze	na mays cultivar:873	ZHENG58x5K_F6	ZHENG58x5K_F6.2	81.6	2.288670802886708
	QTL		5 abph1.22		na mays cultivar:873				
Search by Name			6 ae1.4		sa mays cultivar:873				
a contractor - contractor			7 ae1.8		a mays cultivar:873	B73x8Y804_RIL	873x8Y804_RIL5	104.3	5:17242449917242
Marker Name contains *	Distance	сM	8 an1.3	SNP Ze	sa mays cultivar:873				
File Upload Choose File No file chosen	Search Reset	(	2	SEQUENCE	5				
Provide marker names in a file. Separate each name by a new line.				Sequence			SATCCCCGT66T6ATCAT6TCCT		1
Provide markes markes in a me, separate each name by a new line.							TACTGCATATTTT		
						9			
				1					
Marker Type Any  Marker Mapped in Genome Marker Developed from S	pecies		abph1.22	1	CUMUANY				
"	pecies		abph1.22 Summary						
Marker Mapped in Genome Marker Developed from S	pecies			1	SUMMARY Resource	Туре		Genetic Ma	arker
Marker Mapped in Genome Marker Developed from S	pecles		Summary	,				Genetic M.	arker
Marker Mapped in Genome Marker Developed from S		ьр	Summary Sequence	,	Resource				arker cultivar:873
Marker Mapped in Genome Marker Developed from S Any    Restrict by Location		bp	Summary Sequence Annotations	,	Resource Accession				
Marker Mapped in Genome Marker Developed from S Any  Any  Restrict by Location Genome Location Any between		ьр	Summary Sequence Annotations Contact	) ince	Resource Accession Organism			Zea mays o	
Marker Marker Developed from S Any • Any • Any • Any • • • • • • • • • • • • • • • • • • •		bp	Summary Sequence Annotations Contact Cross Refer	) ince	Resource Accession Organism Name			Zea mays o abph1.22	

**Figure S11. Genetic marker search functions in ZEAMAP, Related to Figure 3.** The genetic markers in ZEAMAP could be searched by their names, features and locations through "Advanced Marker Search" function (A) or by their distances with certain QTL through "QTL Nearby Marker" function (inset in A). The search result lists general information of genetic and physical locations for each record, with physical locations linked to Jbrowse visualization (B). Click on each marker name would lead to the detailed page for this marker including the flanking sequence and the location on maps (C).



**Figure S12.** Crispr sgRNA function in ZEAMAP, Related to Figure 1. The sgRNA information could be searched by IDs and genome locations of target genes through table browser (A), the resulted records have links to their locations on Jbrowse (B) and the detail page of each Cas9\_sgRNA element on Jbrowse shows the information of target and off-target genes (C).

# **Transparent Methods**

### Data collection

The pedigree information of elite inbred lines in the maize association mapping panel was collected from (Yang et al., 2011). 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 (Yang et al., 2019). 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 (Yang et al., 2017). RNA-sequencing of different B73 tissues were from (Walley et al., 2016), and RNA-seq data of developing maize kernels from 368 AMP inbred lines were from (Li et al., 2013). The chromatin interaction and histone modification data were from (Peng et al., 2019), the chromatin accessibility data were from (Rodgers-Melnick et al., 2016), and the DNA methylation data of the AMP were from (Xu et al., 2019). Variants, GWAS and eQTL signals, and phenotypes of the AMP, including agronomic traits, kernel amino acid contents, kernel lipid contents and metabolomic data were collected from previously reported studies (Liu et al., 2015; Wen et al., 2014; Yang et al., 2014). Linkage maps and QTL mapping results were collected from (Pan et al., 2016).

### **Functional annotation**

For each genome annotation, the protein sequences of the predicted genes were compared against the InterPro database using InterProScan 5 (Jones et al., 2014) 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 (Conesa et al., 2005) 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 (Finn et al., 2014) using HMMer 3.1b2 (Potter et al., 2018) with default parameters. To identify gene orthologs and clusters of orthologous group (COG) annotations, the proteins were mapped to eggNOG orthology database (version 4.5.1) (Huerta-Cepas et al., 2015) using emapper-1.0.3 (Huerta-Cepas et al., 2017). To add gene-product annotations, the proteins were searched against UniProt database (version 2019\_04) (Consortium, 2018) using Diamond (v0.8.22.84) (Buchfink et

al., 2015) 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 (<u>https://github.com/nextgenusfs/gene2product</u>). Possible proteolytic enzymes were annotated by searching the proteins against the MEROPS database (version 12.0) (Rawlings et al., 2017) using Diamond with the options "--evalue 1e-05 --max-target-seqs 1". The proteins were also searched against the embryophyta single copy ortholog models from BUSCO Datasets (embryophyta\_odb9, update date: 2017-02-13) (Simao et al., 2015) 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 (Wang et al., 2012) with parameters "-k 50 -g -1 -s 5 -e 1e-10 -m 25 -w 5" to obtain synteny blocks. The whole genome alignments between two genomes in ZEAMAP were performed using minimap2 (version 2.17-r941) (Li, 2018), 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 (Mikheenko et al., 2018).

### Annotating of genetic variations

The SNPs and InDels were annotated using the Ensembl variant effect predictor (VEP) (McLaren et al., 2016) 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 (Yang et al., 2019), and annotated according to B73 gene annotation v4.43 using SURVIVOR v1.0.6 (Jeffares et al., 2017). Haplotype blocks and tag SNPs were identified using Haploview (Barrett et al., 2004).

### Mapping and filtering of 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 (Bradbury et al., 2007). 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 (Zhang et al., 2018).

We kept *cis*-eQTLs alone by retaining only the SNPs within 1 Mb of each gene (Lonsdale et al., 2013). 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 (Zhang et al., 2018).

### Interactive visualization tools

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 gene expression pattern viewer and the eQTL visualizer were modified from GTEx visualizations (https://github.com/broadinstitute/gtex-viz) (Lonsdale et al., 2013). The principal component analyses (PCA) dot plot and the ancestries stacked histogram were also developed using ECharts.

### CRISPR/Cas9 single-guide RNA designing

CRISPR/Cas9 sgRNAs for each maize reference genome were designed using CRISPR-Local (Sun et al., 2018) with default options. Results were converted into gff format with in-house perl scripts to format them for JBrowse.

# **Supplemental References**

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.

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.

Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nature Methods *12*, 59-60.

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.

Consortium, T.U. (2018). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research *47*, D506-D515.

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.

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.

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.

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. 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.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094-3100.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.