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,¹ Linfeng Yang,¹ Jianbo Li,² Jingyun Luo,¹ Xiaokai Xu,² Jianyu Yuan,² Lu Chen,¹ Wenqiang Li,¹ Xin Yang,² Shenshen Wu,¹ Shuyan Li,¹ Yuebin Wang,¹ Yabing Zhu,² Qiang Gao,² Ning Yang,^{1,3,*} and Jianbing Yan^{1,*}

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., 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.

¹National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

²BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China ³I ead 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.



🛞 ZEAMAP

Home ▼ Species ▼ Search ▼ Tools ▼ Help ▼ My account Log out



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).







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







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 r² 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) r² 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).





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	1					E.		
			1	23600K	23700K	23800K	23900K	24000K	24100K	24200K	24300K	24400K	24500K	24600K	24700K	24800	K 24900K	
		refGene	100	Em00001d028	8125	Zm00001	d028137		Zm00001d028150	7-00000	40204548	Zm00001	d028156				Zm00001d02816	4
			120	Em0000	01d028127		Zm00001d02813	39		Zm000	01d028152		Zm00001d02	028160			211000010020	Zm0(
				2 Zm000	001d028128		Zm00001d02	81403		Zm	00001d028153				sulfate_per	mease 1		
_		_							13 items t	too small - z	oom in to view. (Disc	niss)						
		B73_H3K4me3_ref : bed	38															
	2 2	B73 H3K4me3 arc	0		.,			at the set	···· 44 4		hal	* .					\$ \$ 7	20
	<u>9</u> . 9			-11				2			11						Y	
	ct g																	
	ra n	B73_RNAPII_refbe	15]					-			111							
	z ē		0	the stille	1.10		4.1	A 11	and the .		14 1			. Ankan	. d	14		1.
t	52	B73_RNAPII_arc					~							>	1		X	\sim
								14				X		~ >				
		MNase_root_light	57 7							~		-		-	~ ~	/		
	_	MNase_root_heavy		11			1	NII I			4	14		MININ	A		1	۸.,
	ੂ ਦੇ ਹ	MNase_shoot_heavy	0	M. MAN	A AMA MA	- An	Akun	MAL.	som Maan	mark	And.	marks	umalah	MARA	. Allan	mallia	Andran	N
		MNase_root_light 5	5.4	in white	1000		L at	Allera	to will		1.	4		LLL.	4	- La	- 10 m	11
	ie n	MNase root heavy	13 7	all dellade	All Marks an		her all life and some	AL MA JAK.	to the same different as	landa.ni hia	ar shile the s	با المله هدف	in also added the second		And Milling	A have been a little has a	life as shield dates	44
	ss or			A MAN.	d cills dit	1		1.1.1	i ha hit i A wi	HALL & MA	a 14 a .	10 44	L. MALLE	Aller 1			the s back as	11
	E S	MNase_shoot_light	45]				A									All shall a	HIR of a HIR AL BAN	
Ī	ב כ		0	Adding the	a liter at	- he	h	Al I Int	. in What	una a la	. Ann	- Ala	A . Minut & A.	ALLA .	. Maria		the section and	11
	- 10	MNase_shoot_heav ;	20]	· he				hal .				h.		to la a	. 41			
		Histone_INDIT	57 7	de interne	A Ada ad	-ake	a al har as a	AL LI MI	h a fible as All a	mark & Ma	a Ah A	AMAA	. A . Alank A.A.		and the	- Annally A.	A all	AL
		Merged H3K27ac		her				1.1.1			L	1						
	5	H3K4me1	0	IN NA	1.			A LA N	fal				Δ.	M	1	4	ο Λ	M
	ē :⊒	H3K27ac ^{H3K4me3}	30	1				11.										
	ca D		0	Is had				Able	.6.1		111	1		h a				1.
	ĭ, iž	H3K2/me3	35	14											4		4	1
	≌i;	H3K4me1	15 7					- 1	and Andrews					Man			A A	11
12	± 0										14							
	E	H3K4me3	57	1 1				1.1.1			T	1						
			0	h. r.al				11.11				I		11				k .

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
Germolasm*	1:171780-175500	0.0474148	0.808383	0.84343	0.817281	0.0250108	0.786624	0.72619	0.799408	0.805591	0.62585
	1:176060-176940	0.847638	0.839347	0.824534	0.807295	0.795152	0.779392	0.179654	0.85205	0.835341	0.431165
	1:177340-177640	NA	0.883621	0.942529	NA	0.888563	0.932039	NA	NA	0.917355	0.00364964
	1:202100-202220	1	NA	0.679245	0.942857	0.965517	NA	1	0.8	NA	NA
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
end *	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
Search	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 information by DMRs or 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 (yangningyi@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.





*			۹ Filter by C	ategory
			All categories	
Create Downloadable Collection			Analysis	
1000000 results found		Page 1 out of 1	00000 Chromosome	
7m00001d048353			Five Prime UTR	487
Content type: Gene			Gene	152
http://www.zeamap.com/bio_data/577437			Genetic Marker	56
Zm00001d048349			Marker Locus	154
Content type: Gene			Match	56
http://www.zeamap.com/bio_data/577435			Organism	
Zm00001d048348			Polypeptide	345
Content type: Gene			Project	
http://www.zeamap.com/bio_data/577434			Pseudogene Pseudogenic Tr	ancerint
Zm00001d048347			OTI	unactipe
Content type: Gene http://www.zeamap.com/bio.data/577432			RNase MRP RN	A
https://www.camapiconiyoro_datajor/465			SRP RNA	
ZemaCp090			Supercontig	39
http://www.zeamap.com/bio_data/580864			Three Prime UT	R 493
BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits	BLASTN 2.6.0+; query submitted on 2019-11-11 06:20 Databases (131585 sequences, 139385016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, ef Circos	9:59 UTC : Zea_mays_cultivar_873_ ; sc-mismatch -3, gap-oper phdistribution of bits	cds n 5, gap-extend 2, filter L	, m; , m;
BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of selected hit(s)	BLASTN 2.6.0+; query submitted on 2019-11-11 06:20 Databases (13155 sequences, 139355016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Circos Query=Query_1	9:59 UTC : Zea_mays_cultivar_873_ ; sc-mismatch -3, gap-ope phdivibulin utbis	cds n 5, gap-extend 2, filter L	;m; 4 310 4 790
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	BLASTN 2.6.0+; query submitted on 2019-11-11 06:20 Databases (13155 sequences, 139355016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Circos Query=Query_1	0:59 UTC 1: Zea_mays_cultivar_073_ 1: ce-mismatch -3, gap-ope gh-duobudum d'hbu	cds n 5, gap-extend 2, filter L	;m; _ mc _ Pec
BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of selected hit(s) Alignment of selected hit(s)	BLASTN 2.6.0+; query submitted on 2019-11-11 06:20 Databases (131565 sequences, 130365016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Circos Query=Query_1 Graphical overview of hits Length distribution of hits similar sequences	2:59 UTC : Zea_mays_cultivar_B73 sc-mismatch -3, gap-ope ph doubles of bas <u>1</u> <u>1</u> <u>1</u> <u>1</u> <u>2</u> Ouery coverage (%)	cds n 5, gap-extend 2, filter L	;m; <u>A</u> two j A rec <u>A</u> two j
BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of selected hit(s) Alignment of selected hit(s) Standard tabular report	BLASTN 2.6.0+; query submitted on 2019-11-11 06:20 Databases (131565 sequences, 190365016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Circos Query=Query_1 Graphical overview of hits E. Length distribution of hits wimilar sequences 1. Zmooonid000397 TOOI	0:59 UTC : Zea_mays_cultivar_B73 sc-mismatch -3, gap-oper ph dechades of bas Query coverage (%) 100	cds n 5, gap-extend 2, filter L Total score E value	;m; A too j A mo A too j A too
BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits Alignment of selected hit(s) Standard tabular report Full tabular report	BLASTN 2.6.0+; query submitted on 2019-11-11 06:20 Databases (131595 sequences, 19035016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Circos Query=Query_1 Graphical overview of hits Length distribution of hits Similar sequences 1. Zmooooldoocsay, Tool 2. Zmoooldoocsay, Tool	0:59 UTC 1: Zea_mays_cultivar_B73_ pt decoductor at the Query coverage (%) 100 100	cds n 5, gap-extend 2, filter L Total score 1224 0 1200 0	im) A troj A troi Martin Identity (
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 Standard tabular report Full tabular report Full XML report	BLASTN 2.6.0+; query submitted on 2019-11-11 06:20 Databases (131585 sequences, 190385016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Circos Query=Query_1 Graphical overview of hits Length distribution of hits % Similar sequences 1. Zmooooid00397_T001 2. Zmooooid016888_T001 3. Zmooooid016887_T001	0:59 UTC 1: Zea_mays_cultivar_B73_ sc-mismatch -3, gap-open ph dividues of this Query coverage (%) 100 100 99	cds n 5, gap-extend 2, filter L Total score Evalue 1224 0 1200 0 1172 0	im)
BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits FASTA of selected hit(s) Alignment of selected hit(s) Standard tabular report Full tabular report Full tabular report Full XML report	BLASTN 2.6.0+; query submitted on 2019-11-11 06:20 Databases (131585 sequences, 190385016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Circos Query=Query_1 Graphical overview of hits E Length distribution of hits # Similar sequences 1. Zmoooidd016887_T001 2. Zmooooidd01687_T001 4. Zmooooidd01687_T001	2:59 UTC : Zee_mays_cultivar_B73_ ; sc-mismatch -3, gap-open en denbudes othen Query coverage (%) 100 100 99 100	cds n 5, gap-extend 2, filter L Total score Evalue 1224 0 1200 0 1172 0 1168 0	, a so (2 min)
BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits Alignment of selected hit(s) Alignment of all hits Alignment of all hits Standard tabular report Full tabular report Full tabular report Full XML report	BLASTN 2.6.0+; query submitted on 2019-11-11 06:20 Databases (13155 sequences, 190305016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Circos Query= Query_1 Graphical overview of hits Length distribution of hits Similar sequences 1. Zmoooold000397_T001 2. Zmoooold016870_T001 3. Zmoooold016870_T001 5. Zmoooold00588_T001	2:59 UTC : Zee_mays_cultivar_B73_ ; sc-mismatch -3, gap-open en dentedeu otex Query coverage (%) 100 99 100 99	cds n 5, gap-extend 2, filter L Totalscore Evalue 1224 0 1200 0 1172 0 1168 0 1162 0	, a troi à ren dentity (
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 Alignment of all hits Standard tabular report Full tabular report Full XML report	BLASTN 2.6.0+; query submitted on 2019-11-11 06:20 Databases (13155 sequences, 19938906 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Circos Query=Query_1 Graphical overview of hits E Length distribution of hits E Length distribution of hits Similar sequences I. Zmoooold000397_T001 2. Zmoooold016868_T001 3. Zmoooold016870_T001 6. Zmoooold006857_T001 6. Zmoooold006857_T001	2:59 UTC : Zea_mays_cultivar_B73 sc-mismatch -3, gap-ope ph debloem of bas Query coverage (%) 100 100 99 100 99 98	cds n 5, gap-extend 2, filter L Total score E value 1224 0 1200 0 1172 0 1168 0 1162 0 914 0	;m; A too; A the Identity (
BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits Alignment of selected hit(s) Standard tabular report Full tabular report Full XML report 2	BLASTN 2.6.0+; query submitted on 2019-11-11 06:20 Databases (13155 sequences, 19083016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Circos Query=Query_1	2:59 UTC 1: Zea_mays_cultivar_073 sc-mismatch -3, gap-open ph debbes of bas Query coverage (%) 100 99 100 99 98 23	cds n 5, gap-extend 2, filter L Total score E value 1224 0 1200 0 1172 0 1168 0 1162 0 914 0 78 3.00 × 10 ²¹	;es; A to() A to() Identity (
BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XNL, TSV FASTA of all hits FASTA of all hits Alignment of selected hit(s) Standard tabular report Full tabular report Full XNL report Q	BLASTN 2.6.0+; query submitted on 2019-11-11 06:20 Databases (13155 sequences, 199839016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Circos Query=Query_1 Graphical overview of hits E Length distribution of hits E Length distribution of hits Similar sequences Zmoooo1d010397_T001 Zmoooo1d016868_T001 Zmoooo1d01687_T001 Zmoooo1d016867_T001 Zmoooo1d016867_T001 Zmoooo1d016667_T001 Zmoooo1d017160_T001 (es chromosor	2:59 UTC 1: Zea_mays_cultivar_873 sc-mismatch -3, gap-oper ph debbers of bus Query coverage (%) 100 99 100 99 98 23 meB73. RefGen. v45.(872119373)	cds n 5, gap-extend 2, filter L Total score E value 1224 0 1200 0 1172 0 1168 0 1162 0 914 0 78 3.00×10 ⁴¹	;es; ivo; j & rec i i i i i i i i i i i i i
BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits Alignment of selected hit(s) Standard tabular report Full tabular report Full XML report Q	BLASTN 2.6.0+; query submitted on 2019-11-11 06:20 Databases (131565 sequences, 190365016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Circos Query=Query_1 Graphical overview of hits E Length distribution of hits Similar sequences 1. Zmoooo1d000397,T001 2. Zmoooo1d003687,T001 3. Zmoooo1d016872,T001 4. Zmoooo1d016872,T001 5. Zmoooo1d016872,T001 6. Zmoooo1d016872,T001 7. Zmoooo1d016872,T001 7. Zmoooo1d016875,T001 7. Zmoooo1d016875,T001 7. Zmoooo1d016875,T001 7. Zmoooo1d016875,T001 7. Zmoooo1d017160,T001 7. Zmoooo1d017160,T001 7. Zmoooo1d017160,T001	2:59 UTC 1: Zea_mays_cultivar_073 sc-mismatch -3, gap-open m dechades of his Query coverage (%) 100 100 99 100 99 98 23 meB/71_Ref6m_v46.18721195171.	cds n 5, gap-extend 2, filter L Total score E value 1224 0 1200 0 1172 0 1168 0 1162 0 914 0 78 3.00 × 10 ⁴¹ 87218926-1 game.2mm0001d017160 e cyclin.43-1	;e; i voj a rec i i i i i i i i i i i i i
BLASTN: 1 query, 1 database Query=Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of selected hit(s) Alignment of selected hit(s) Standard tabular report Full tabular report Full XML report	BLASTN 2.6.0+; query submitted on 2019-11-11 06:20 Databases (131565 sequences, 19035016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Circos Query=Query_1 Graphical overview of hits Length distribution of hits Similar sequences 1. Zmoooold000397,1001 2. Zmoooold000387,1001 3. Zmoooold016870,1001 3. Zmoooold016870,1001 3. Zmoooold016870,1001 3. Zmoooold016870,1001 4. Zmoooold016870,1001 5. Zmoooold016870,1001 5. Zmoooold016870,1001 5. Zmoooold017160,1001 5. Select (Sequence) ASTA & Alig	2:59 UTC : Zea_mays_cultivar_873 sc-mismatch - 3, gap-open get debbetes of bas Query coverage (%) 100 100 99 100 99 98 23 me:873_RefGen_v45:187217937:1 rotoin, coding deciripion.Putativo gnment	cds n 5, gap-extend 2, filter L Total score E value 1224 0 1200 0 1172 0 1168 0 1162 0 914 0 78 2.00×10 ⁻¹¹ 87218926-1 geneZm00001d017160 e cyclin A3-1	;m; A too; A two identity (
BLASTN: 1 query, 1 database Query: Query_1 Download FASTA, XML, TSV FASTA of all hits FASTA of all hits Alignment of selected hit(s) Standard tabular report Full tabular report Full XML report	BLASTN 2.6.0+; query submitted on 2019-11-11 06:20 Databases (131565 sequences, 19035016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Clicos Query=Query_1 Graphical overview of hits Length distribution of hits Similar sequences 1. Zmocool:do:00397_1001 2. Zmocool:do:0887_1001 3. Zmocool:do:0887_1001 4. Zmocool:do:0887_1001 5. Zmocool:do:0867_1001 5. Zmocool:do:0867_1001 5. Zmocool:do:0867_1001 5. Zmocool:do:1687_1001 5. Zmocool:do:1687_1001 6. Zmocool:do:1687_1001 7. Zmocool:do:1687_1001	21:59 UTC 1: Zea_mays_cultivar_873_ 2: comismatch - 3, gap-oper ph distribution of hose Query coverage (%) 100 100 100 99 98 23 meB73_RefGen_v45.187217937:1 rotein_coding description.Plustive gament 9)	cds n 5, gap-extend 2, filter L Total score 1224 0 1220 0 1172 0 1168 0 1162 0 914 0 78 3.00×10 ⁻¹¹ 87218926-1 gene2m00001d017160 e cyclin-A3-1	;e; ▲ too (▲ rec
BLASTN: 1 query, 1 database Query=Query_1 Download FASTA, XML, TSV FASTA of selected hit(s) Alignment of selected hit(s) Standard tabular report Full tabular report Full XML report	BLASTN 2.6.0+; query submitted on 2019-11-11 06:20 Databases (131565 sequences, 19035016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Circos Query=Query_1 Graphical overview of hits Lamooolidousographical similar sequences 1. Zmoooolidousographical 2. Zmoooolidousographical 3. Zmoooolidousographical 3. Zmoooolidousographical 3. Zmoooolidousographical 4. Zmoooolidousographical 5. Zmoooolidousographical 5. Zmoooolidousographical 6. Zmoooolidousographical 7. Zmoooolidousogra	2:59 UTC : Zea_mays_cultivar_873 sc-mismatch -3, gap-open ph debhdes rhss Query coverage (%) 100 100 99 98 23 me873_RefGen_v451871793711 rotein_coding deciripion R-btable comment s) 2 2 2 3 2 3 3 3 4 5 5 5 5 5 5 5 5 5 5 5 5 5	cds n 5, gap-extend 2, filter L Total score E value 1224 0 1200 0 1172 0 1168 0 1162 0 914 0 778 3.00×10 ⁻¹¹ 87218926-1 gene2m00001d017160 ecyclin-X3-1	;;;;; ▲ 100 ▲ 190 ▲ 190 ▲ • Identity (• SVG ▲ • SVG ▲
BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XNH, TSV FASTA of all hits FASTA of all hits Alignment of selected hit(s) Standard tabular report Full XML report Full XML report	BLASTN 2.6.0+; quary submitted on 2019-11-11 06:20 Databases (131585 sequences, 190385016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Circos Query= Query_1 Graphical overview of hits Length distribution of hits Similar sequences 1. Zmo001d016897_T001 2. Zmo001d016897_T001 3. Zmo001d016897_T001 5. Zmo001d016897_T001 6. Zmo001d016897_T001 7. Zmo001d016897_T001 6. Zmo001d016897_T001 7. Zmo001d016897_T001 6. Select[& Sequence] & FASTA & Mij Graphical overview of aligning region(s # # # # # # # # #	2:59 UTC 2: Zea_mays_cultivar_B73_ 5: cr-mismatch - 3, gap-open mt diothesis at this Query coverage (%) 100 100 100 99 98 23 meB73_Ref(em_v45.187217927.1 99 98 23 meB73_Ref(em_v45.187217927.1 100 99 98 23 100 100 99 98 23 100 100 99 98 23 100 100 99 98 23 100 100 99 98 23 100 100 99 98 23 100 100 99 98 23 100 100 99 98 23 100 100 99 98 23 100 100 99 98 23 100 100 100 100 100 100 100 10	cds n 5, gap-extend 2, filter L Total score Evalue 1224 0 1200 0 1172 0 1168 0 1162 0 914 0 78 3.00 × 10 ⁴¹ 78 3.00 × 10 ⁴¹	jej inci ≜ rec identity i statistical s
BLASTN: 1 query, 1 database Query= Query_1 Download FASTA, XMI, TSV FASTA of all hits Alignment of all hits Alignment of selected hit(s) Standard tabular report Full tabular report Full XML report	BLASTN 2.6.0+; quary submitted on 2019-11-11 06:20 Databases (131585 sequences, 190385016 characters) Parameters: task blastn, evalue 1e-05, sc-match 2, Circos Query=Query_1 Graphical overview of hits Length distribution of hits Similar sequences 1. Zmo001d016897_T001 2. Zm0001d016897_T001 3. Zm00001d016897_T001 4. Zm0001d016897_T001 5. Zm00001d016897_T001 6. Zm00001d017160_T001 cds chromoson rgr.e.je.loypeprotein_colong transcrite_biotypepr Select [Sequence] & Sequence [Archard L, biotypepr Graphical overview of aligning region(f # Select [Sequence] & Sequence [Archard L, aligning region(f # Select [Sequence] & Sequence [Archard L, aligning region(f # Select [Sequence] & Sequence [Archard L, aligning region(f # Select [Sequence] & Sequence [Archard L, aligning region(f # Select [Sequence] & Sequence [Archard L, aligning region(f # Select [Sequence] & Sequence [Archard L, aligning region(f # Select [Sequence] & Sequence [Archard L, aligning region(f # Select [Sequence] & Sequence [Archard L, aligning region(f # Select [Sequence] & Select [Select [Sequence] & Select [Select [Select] & Select [Sequence] & Select [Select] & Select [Select] & Select] & Select [Select] & Select] & Select [Select] & Select] & Select] & Select [Select] & Select] & Select] & Select] & Select [Select] & Select	2:59 UTC 1: Zea_mays_cultivar_073_ 1: Zea_mays_cultivar_073 1: Zea_	cds n 5, gap-extend 2, filter L Total score E value 1224 0 1200 0 1172 0 1168 0 1162 0 1164 0	;m; icit icit i me identity (

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.



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, 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 cul	tivar:l	B73	Site Wide Search Q
Summary	>	SUMMARY	
Annotations		Resource Type	Organism
Cross Reference		Abbreviation	B73
Publication		Genus	Zea
Relationship		Species	mays cultivar:B73
		Common Name	maize B73
		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 B73_RefGen_v4.42 . 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).

marines, rou ma	ay also filter results by sequence type	and the sequence source. To select multiple	-	Marrie	11- fair and a second	Terre	Quantan			A
options click wi format.	while holding the "ctrl" key. The result	s can be downloaded in FASTA or CSV tabular	1	Zm00015a011942_P001	Zm00015a011942_P001	polypeptide	Zea mays	Zea mays cultivar:SK Whole Ge	enome Assembly and	chr3: 17424839
Species A	Any		2	ZMex08g026281	ZMex08g026281	gene	cultivar:SK Zea mays ssp.	Annotation Zea mays ssp. mexicana Genor	me Assembly and	17440222 8: 10014738 10
Ze	Zea mays cultivar:873 Zea mays cultivar:Mo17		3	Zm00014a009861_P002	Zm00014a009861_P002	polypeptide	mexicana Zea mays	Annotation Zea mays cultivar:Mo17 RefSec	q Genome	chr5: 211530022
-	tea mays culovar.sk		4	Zm00001d021846_T015	Zm00001d021846_T015	mRNA	cultivar:Mo17 Zea mays	Zea mays cultivar:873 RefSeq 0	Genome	211533543 7: 164385354
Type 9	any o gene mRNA		5	Zm00015a023356	Zm00015a023356	gene	cultivar:873 Zea mays	Zea mays cultivar:SK Whole Ge	enome Assembly and	164417958 chr5: 143053835
p	polypeptide -		6	ZMex06t021864 P01	ZMex06t021864 P01	polypeptide	cultivar:SK Zea mays ssp.	Annotation Zea mays ssp. mexicana Genor	me Assembly and	143058464
Source Ar	Any Zea mays cultivar:873 RefSeq Genome		7	70000140010455	7000014+010455		mexicana Zeo maur	Annotation	a Conomo	chr4: 342103415
Ze	Zea mays cultivar:Mo17 RefSeq Genon Zea mays cultivar:SK Whole Genome A	ne Assembly and Annotation 👻	1	2.11000143010433	2.11000148010433	gene	cultivar:Mo17	Zea mays convariant in Reiser	d denome	243183690
Location A	Any • between	and	8	2m00015a004640	2m00015a004640	gene	cultivar:SK	Zea mays cultivar:SK Whole Ge Annotation	enome Assembly and	239095412
Name	contains •		9	ZMex01t001328_T01	ZMex01t001328_T01	mRNA	Zea mays ssp. mexicana	Zea mays ssp. mexicana Genor Annotation	me Assembly and	1: 29525762 29
			10	Zm00014a029257_T001	Zm00014a029257_T001	mRNA	Zea mays cultivar:Mo17	Zea mays cultivar:Mo17 RefSec	q Genome	chr8: 22958439 22962152
File Ch Upload	hoose File No file chosen		11	Zm00001d015722_P001	Zm00001d015722_P001	polypeptide	Zea mays cultivar:873	Zea mays cultivar:873 RefSeq 0	Genome	5: 112752607 112755567
Provide sequence	e names in a file. Separate each name by a n	new line.	12	Zm00001d035004_T032	Zm00001d035004_T032	mRNA	Zea mays cultivar:B73	Zea mays cultivar:B73 RefSeq 0	Genome	6: 1572540 157
Search	Neset	— ,	13	Zm00014a044091	Zm00014a044091	gene	Zea mays cultivar:Mo17	Zea mays cultivar:Mo17 RefSee	q Genome	chr4: 203693418 203694704
				Mama	Organize Langth	Turne	GO PLAST	VIGG	INTERPROA	
Search by Na	lame			Name	Organism Length	Туре	GO BLAST Term	KEGG	INTERPROA	
Search by Na Gene/Feature Nau File Upload Provide sequence nem Search by As	ame contant • .	(0.9.40)	# 1	Name Zm00001d027402_P004	Organism Length Zea mays 213 cultivar:873	Type polypeptide	GO BLAST Term ONL93052 protein ZEAMM87 [Zea mays	KEGG 2.1 hypothetical 73_Zm00001d027402]	INTERPRO▲ "Helical backbone" me NAMED, pep chromosome:873,Reft gene:Zm00001d02740 transcriptZm00001d0 gene_biotype:protein_ transcript_biotype:protein_	ital receptor. FAMII Sen_v4:1:4417837:4 2 27402_T004 coding tein_coding
Search by Na Gene/Feature Nas File Upload Incide sequence nam Search by As Source	Cancer III III IIII IIIIIIIIIIIIIIIIIIIIIII	(P § +R) CII or analy and Angeleration by your Accession	# 1 2	Name Zm00001d027402_P004 Zm00001d027402_P004	Organiam Length Zea mays 213 cultivar:873 Zea mays 213 cultivar:873	Type polypeptide polypeptide	GO BLAST Term ON1:93052 protein ZEAMM87 [Zea mays ON1:93052 protein ZEAMM87 [Zea mays	KEGG 2.1 hypothetical 73 _Zm000010027402 0 2.1 hypothetical 73 _Zm000010027402 0	INTERPROA "Helical backbone" nm NAMED, pep chromosome 873, 8ef gene_Zm00001 402 740 gene_Diotype protein, transcript_Diotype protein, thanscript_Zm0001 402 740 chromosome 873, 8ef gene_Zm00001 402 740	tal receptor. FAMI Sen_v4:1:4417837: 2 27402_1004 coding tein_coding tein_coding tal receptor. FAMI Sen_v4:1:4417837: 2 2
Search by Na Gent/Feature Nas File Upload hexite separce nam Search by As Source	Anne	(B § +20) CB or a mantip year Anomation * 1	# 1 2	Name Zm00001d027402_P004 Zm00001d027402_P004	Organism Length Zea mays 213 cultivar:873 Zea mays 213 cultivar:873	Type polypeptide polypeptide	GO BLAST Term ONL93052 protein ZEAMM87 [Zea mays [Zea mays [Zea mays ONL93052	KEGG 2.1 hypothetical 3.2.mo00010027402 3 2.1 hypothetical 3.3.2mo00010027402 1 3.1.buoduloinol	INTERPROA "Helical backbone" nm NAMED, pep chromosome 873, Beffe gene Zm00001 d02740 transcript Zm0001 d0 gene, biotype protein, transcript Jioitype pro Helical backbone" m Bene Zm00001 d02740 transcript Zm0001 d0 gene, biotype protein, transcript, biotype protein, transcript, biotype protein, transcript, biotype protein,	tal receptor, FAMI Sen_v4:1:4417837: 2 27402_T004 coding tal receptor, FAMI sen_v4:1:4417837: 2 27402_T004 coding tein_coding tein_coding
Search by Na Gene/Feature Na File Upload Provide sequence nam Search by As Source Type Search by Pu	table contain a	(ing. ant) Citi or they prove the research in the bity and Annexation. *	# 1 2 3	Name Zm00001d027402_P004 Zm00001d027402_P004 Zm00001d027402_P004	Organium Length Zea mays cultivar:873 213 Zea mays cultivar:873 213 Zea mays cultivar:873 213	Type polypeptide polypeptide	GO BLAST Term ONL93052 ZEAMM87 [Zea mays ONL93052 IZea mays ZEAMM87 [Zea mays Protein ZEAMM87 Zeamas	KEGG 2.1 hypothetical 73_2m00001d022402 1 2.1 hypothetical 73_2m00001d022402 1 2.1 hypothetical 73_2m00001d022402 2.1 hypothetical	INTERPROA "Helical backbone" mr NAMED, pp chromosome.873, Bef gene.2m00001d02740 transcript.2m0001d0 Helical backbone" mr NAMED, pp chromosome.873, Bef gene.2m0001d02740 gene.biotypeprotein, transcript.2m0001d0 gene.biotypeprotein, transcript.biotypeprotein, homosome.873, Bef gene.2m0001d02740 Helical backbone" mr	tal receptor. FAMI See, y4:14417837. 2 27402_1004 coding tal receptor. FAMI tal receptor. FAMI see, coding tal receptor. FAMI see, coding tal receptor. FAMI tal receptor. FAMI tal receptor. FAMI
Search by Na Gens/Feature Nas File Upbad Nextle sequere non Search by As Source Type Search by Pu Go Term BLAST Description	Anne	(in g. add) CE in we have a second of the s	# 1 2	Name Zm00001d027402_P004 Zm00001d027402_P004 Zm00001d027402_P004	Organism Length Zea mays cultivar/873 213 Zea mays cultivar/873 213	Type polypeptide polypeptide	GO BLAST Term CALL ST PORTONIC ST IZEAMAGE ZEAMAGE ZEAMAGE ZEAMAGE ZEAMAGE ZEAMAGE ZEAMAGE	KEGG 2.1 hypothetical 73_2m00001d027402 1 2.1 hypothetical 2.2 hypothetical 1 2.1 hypothetical 73_2m00001d027402 1	INTERPROA "Helical backbone" me NAMED, pep chromosome 273, Bef gene_Zhodoo1 402,740 gene_Zhodoo1402,740 Telecial chrospep otelin, NAMED, pep chromosome 873, Bef gene_Zmodoo1402,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_Zhodoo140,740 transcript_	tal receptor. FAMI January 1, 41: 144 17837- 2 27402, TOO4 coding tal receptor. FAMI tal receptor. FAMI 27702, TOO4 coding ten v.v1: 144 17837- 2 27402, TOO4 coding 2 27402, TO7
Search by Na Genu/Feature Nar File Upload Newle sepance nen Search by As Source Type Search by Pu Go Term BLAST Description KEGG Description	Anne	(4.9, add) (5.9 m entropy and Averagians) by and Averagians) (4.9, G27 banding, bary add) (4.8 work of Planetary bary add) (4.8 work of Planetary bary add) (4.8 work of Planetary bary add) (4.9 m) (4.1 m) (# 1 2 3	Name Zm00001d027402_P004 Zm00001d027402_P004 Zm00001d027402_P004 Zm00001d027402_P004	Organism Length Zea mays 213 cultivar873 213 Zea mays 213 Cultivar873 213 Zea mays 213 Zea mays 213 Zea mays 213	Type polypeptide polypeptide polypeptide	GO BLAST Tem ON193052 protein [Zea mays protein zCAMME] [Zea mays [Zea mays [Zea mays [Zea mays [Zea mays] [Zea mays] [Zea mays] [Zea mays]	KEGG 2:1 hypothetical 73_2m00001d027402 0 2:1 hypothetical 3_2m00001d027402 0 2:1 hypothetical 73_2m00001d027402 0 2:1 hypothetical 73_2m00001d027402 0 2:1 hypothetical 2:1 hypothetical	INTERPROA Telefal backbon* nm NAMID, pp (consonne 87), 240, 400, 402, 404 gene Zondon 402, 404 gene Zondon 402, 404 gene Zondon 402, 404 genez Zondon 402, 404 transcript Zindongenez Talkica backbon* m Harica Jackbon* m	tal receptor, FAMI Sen, v4:14417837, 2 27402_T004 tal receptor, FAMI Sen, v4:14417837, 2 7402_T004 coding emic, coding emic, coding emi

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.

roputation	ma	ize association mapping panel									2000 1.1	2 173715490 173716126 (637 1	54 Go 🐊		
Variation Name					0			C715.500	1717/6/65		73,715,750	173,735,875		13,716,000	
Variation Class	SNI	P/		Variation Type	Am	,		© Gene			an one of the				
Senome	Aim	i						() 527 genomes she fe	100000	115620 Jun 2 11 173715	495 UN2 4_173715702	Lang 1, 1737 150	61 art2 1_173715907	1 chr2.s_173716001	н
hr/Scaffold	Am	, • between		1.9	and	bp		* 1 5NV N/2.5, 173715487	0-24 SNY T-0 C	173715636 (N2.1.)	173718713 042 4 1737181	1 1 1 1 1 1 1 1 1 1 1 1 1 1	5068 0x2 s 1737150	145 012 173718	5005
	hate	-							01/2 x 1737 15576 1040	2 173715642	012 8_173715736		CHI2 & 1737159915		
	Den		204						5002 + 173715804	012 6 173715678	50V2 6 173715745		In mile starsme		
ariation effect	Am	·							08/2 s 173715591 5NV T-+ 0		6h2.s.173716766 5NV C -> 0		NO NO	Andre III.	
	Search		Down	load CSV		Cloar					SWY G -> T		Security Cont	NO NO N	
Population *	Ref_Genome	* VariantID * VariantTy	pe Chr	Posi ^ Ref	^ Alt ~ MAF	* ANN_Code * Action	is *	© 507 genomes snp m	utrix				NELH 1 Devolution (NT)		
Maize_AMP	873_V4	ctv2.s_173715868 SNP	2	173715868 A	G REAL	i 👐 🖌 ibrows				~~~~				-	
Maize_AMP	B73_V4	chr2.s_173715916 SNP	2	173715916 G	3.prime_UTR_var	Ibrows Ibrows								10 100 10 10	
Maize AMP	873 V4	chr2s_173715945_SNP	2	173715945 G	T	Jbrows							1400A	10 (K) (10) 10 (K) (10)	*
Maize_AMP	873_V4	chr2.s_173715946 SNP	2	173715946 C	A	• Jbrows								4.5.47 4.5.47	-
														#A #2	-
Maize_AMP	B73_V4	chr2s_173716001 SNP	2	173716001 T	С	Drows									-
Maize_AMP	B73_V4	chr2.s_173716001 SNP	2	173716001 T	Position	A ID A	REF 🔶 ALT	▲ 04K5672	- 04K5686	04K5702	• 05W002	- 05WN230	- 07KS4	÷ 1323	à.
Maize_AMP ZEAMAP Gen rowse the genotype ricenation.	B73_V4	chr2.s_173716001 SNP	2 e name of germpla	173716001 T	Position 1:2356	 ID chr1.s_2356 	REF ALT	▲ 04K5672 0]0	• 04K5686 •	04K5702	 05W002 1 1 	05WN230	• 07KS4	* 1323 0 1	
Maize_AMP	B73_V4	chr2s_173716001 SNP	2 e name of germpia	173716001 T	Position 1:2356 1:28115	 ID chr1.s_2356 chr1.s_28115 	REF ALT C T C T	▲ 04K5672 0[0 0]0	04K5686	04K5702	 o5w002 1 1 0 1 	 05WN230 1 1 0 1 	откя 01 01	* 1323 0 1 0 1	*
Maize_AMP CEAMAP Gen rouse the genatype formation. Choose datass Dataset * MalanAMP_Jugh	B73_V4 notype Search of certain germplane and germplane botppa_tag_variates	chr2s_173716001 SNP 	2 e name of germpla	173716001 T	Position 1:2356 1:28115 1:28701	 ID chr1.s_2356 chr1.s_28115 chr1.s_28701 	REF ALT C T C T A G	 04K5672 0(0 0(0 1(1 	04K5686 011 010 111	04K5702 1 1 0 0 1 1	O5W002 1[1 0]1 1]1	05WN230 1/1 0/1 1/1	07KS4 0 1 0 1 1 1	 1323 0 1 0 1 1 1 	*
Maize_AMP CEAMAP Gen revue the genatype formation. Choose datas Dataset * MalosAttP-Japh Select a set of ge Germplasm*	B73_V4 notype Search of certain germplasm set and germplasm topps_tag_variants *	dw2.4,173716031 SNP . Strepy refert a drawer, and then even th ns you are interested in.	2 e name of germp% malysis tool	173716001 T	Position 1:2356 1:28115 1:28701 1:33879	 ID chr1.s_2356 chr1.s_28115 chr1.s_28701 chr1.s_33879 	REF ALT C T C T A G C T	 04K5672 0 0 0 0 1 1 0 0 	 04K5686 0 1 0 0 1 1 0 0 	04K5702 1 1 1 0 0 1 1 0 0	 o5w002 1 1 0 1 1 1 1 1 	 o5WN230 1 1 0 1 1 1 1 1 1 1 	07KS4 0 1 0 1 1 1 1 1	 1323 0 1 0 1 1 1 1 1 	*
Malze AMP EEAMAP Gerr rease the genetype formation: Choose datas Dasset * MataxMP Jupit Select a sit of ge Gemplean*	B73_V4 nokype Search of cetsils gremplasm set and germplasm totype_sag.variats •) metic variants. Learn m	dw2.4,173716001 SNP . Strepy select a dataset, and then enter th in you are interested in. . our elect each dataset using not learn hit	2 In name of gerergia majoris took	173716001 T	Position 1:2356 1:28115 1:28701 1:33879 1:34951	 ID chr1.s_2356 chr1.s_28115 chr1.s_28701 chr1.s_33879 chr1.s_34951 	REF ALT C T C T A G C T G A		04K5686 4 0[1 0[0 1]1 0[0 0]0 0[0	04K5702 1 1 0 0 1 1 0 0 1 1	 05W002 1 1 0 1 1 1 1 1 0 0 	 05WN230 1 1 0 1 1 1 1 1 0 0 	07KS4 0 1 0 1 1 1 1 1 1 1 0 0	* 1323 0 1 0 1 1 1 1 1 0 0	*
Malze_AMP EAMAP Gerr tornation Choose datas Dataset * MalandM1-bags Gerriphsen* Selected gerriphs	B73_V4 hotype Search of cetais gemplane iet and geemplane intege sig.variate • hotpe sig.variate •	chi-2.5,173716031 SNP Steppy where a disease, and then every th in you are interested in. ever always exchange on legenth A	2 r name of geength nalysis tool.	173716001 T	Position 1:2356 1:28115 1:28701 1:33879 1:34951 1:35636	 ID chr1.s_2356 chr1.s_28701 chr1.s_28701 chr1.s_3879 chr1.s_34951 chr1.s_35636 	REF ALT C T C T A G C T G A C T	O4K5672 Olo Olo	O4K5686 Ol1 Ol0 Ol0	O4K5702 1]1 0]0 1]1 0]0 1]1 0]0 1]1 0]0 1]1 0]0	o5woo2 1[1 0]1 1[1 1]1 0]0 0]0 0]0	 o5WN230 1 1 0 1 1 1 1 1 0 0 0 0 0 0 	07KS4 0 1 0 1 1 1 1 1 0 0 0 0	 1323 0 1 0 1 1 1 1 1 0 0 0 0 	*
Malze, AMP EAMAP Gerr ows the genetype formation. Choose datase Dataset * MalandM: Jugg Select a set of ge Gerrgliser Selected genetyle Restrict to the	B73,14 notype Search of cetais gemptions and gemptions toppe (sg.venate.) toppe (sg.venate.) area (sg.venate.) area (sg.venate.)	chicky statistics (see the statistics) (see the statistic of the statistic	2 e name of gerengia majosis tool 200 (07K54)	173716001 T	Position 1:2356 1:28115 1:28701 1:33879 1:34951 1:35636 1:35864	 ID chr1.s_2356 chr1.s_28701 chr1.s_3879 chr1.s_34951 chr1.s_35636 chr1.s_35864 	REF ALT C T C T A G C T G A C T G A C T G A	O4K5672 O(0 O(0 I(1 O(0 O(0 O(0 O(0 O(0 O(0 O(0 O(0 O(0 I(1 I(1	O4K5686 Ol1 Ol0 Ol0	очк5702 о 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1	oswoo2 1[1 0[1 1[1 1[1 0]0 0]0 0]0 0]0	O5WN230 I 1 O 1 I 1 I 1 O 0 O 0 O 0 O 0	07KS4 0 1 0 1 1 1 1 1 0 0 0 0 0 0 0 0	 1323 0 1 0 1 1 1 1 1 0 0 0 0 0 0 0 0 	*
Make, AMP EAMAP Gerr asset for generitype formation. Choose datas Dataset* MasaatAT: Juga Select as of a dia Genergiaan* Selected genergia Restrict to the To opticipite the is	B73,14 hotype Search of cetain germplane let and	childra, 17771001 State Strapy where a down and the same the strapy where a down and hold methods are short such information and your family the strategy information and you family the strategy information and you family the strategy information and the same strate in the strategy information and the same strate in the strategy in the strategy information and the same strategy in the strategy in	2 e name of gerengin natypis took	173716001 T	Position 1:2356 1:28115 1:28701 1:33879 1:34951 1:35836 1:35864 1:35864 1:39119	 ID chr1.s_2356 chr1.s_28701 chr1.s_3879 chr1.s_34951 chr1.s_35636 chr1.s_35636 chr1.s_35864 chr1.s_31919 	REF ALT C T C T A G C T G A C T G A C T G A C T G A T C		O4K5686 O1 O1 O O O O O O O O O O O O O O O O O O	 04K5702 1 1 0 0 	oswoo2 1[1 0[1 1[1 1[1 1[1 1[1 0[0 0[0 0[0 0[0 1[1	O5WN230 I 1 O 1 I 1 I 1 O 0 O 0 O 0 I 1	07KS4 0 1 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0		*
Make, AMP EAMAP Gen town the genetype formation Choose datas Selected genetybe Restrict to the To applefice the so Choose (1 - 1)	B73_V4 httpp://www.second.org/actions/ https://www.second.org/actions/ https://www.second.org/actions/ https://www.second.org/actions/ https://www.second.org/ https://wwwwwwwwwwwwwwwwwwwwwww	0-022,17771001 90 Chepy white a dataset, and then enter the set you are instructed in . we should also be able of the enter the instruct instruct instruct instruct instruct instruct instruct instruct instruct instruct instruct instruct instruct instruct instruct	2 e name of greenpli mayole tool 200 07/604	173716001 T	Position 1:2356 1:28115 1:28701 1:34951 1:35636 1:35864 1:35919 1:51053	 ID * chr1.s_2356 chr1.s_28701 chr1.s_38791 chr1.s_3864 chr1.s_3864 chr1.s_3864 chr1.s_31919 chr1.s_51053 	REF ALT C T C T G A C T G A T C A T		O4K5686 O1 O1 O	O4K5702 Ill Ol Ol		O5WN230 1/1 0/1 1/1 1/1 1/1 0/0 0/0 0/0 1/1 0/0	O7KS4 0 1 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	* 1323 0 1 0 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1	*
Make, AMP	B73_V4 httpp://search of cetails generalism int and germplane int and germplane int and germplane int and germplane interveloped to a search generalism interveloped to a search generalism interveloped to a search interveloped to a search i	chic2_117711001 SP thirdy which 3 charact and then enter the thirdy which 2 charact and then enter the train of the thirdy which 2 characteristic the train of the thirdy which 2 characteristic the third	2 e name of geength natypic tool.	173746001 T	Position 1:2356 1:28115 1:28701 1:33879 1:34951 1:35636 1:35864 1:39119 1:51053 1:56788	 ID chr1.s_2356 chr1.s_24701 chr1.s_3879 chr1.s_34951 chr1.s_35664 chr1.s_35864 chr1.s_51053 chr1.s_56788 	REF ALT C T C T A G C T G A T C A T A T A G	O4K5672 O(0	O4K5686 O[1 O[0 O[0	04K5702 111 010 111 010 111 010 111 010 111 010 111 010 111 010 111 010 010 010	oswooz 1/1 0/1 1/1 1/1 0/0 0/0 0/0 0/0 1/1 1/1	OSWN230 IJ1 IJ1 UI1 IJ1 UI UI	 O7KS4 O[1 O[1 O[1 O[1 O[0 O[0	* 1323 0 1 0 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1	

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 Trait Quanti	tative Trait				
	,	Trait 1 cobcolor • Trait 2 Any • Trait 3 Any •	Value 1 red • Any Value 2 red white Value 3 Any •	AND •			
Qualitat	ive Trait Quantita	tive Trait					
Trait 1	100grainweight	• between	11.7899	and	30.5579	AND	•
Trait 2	cobweight	* between	4.49766	and	31.4859	AND	•
Trait 3	kernellength	• between	6.88922	and	10.9851		
Search 467 record	Reset Is were returned						Download Table
#	Germplasm	Species		100grain	weight	cobweight	kernellength
1	150	Zea mays culti	var:B73	18.6644		18.7322	8.71982
2	177	Zea mays culti	var:B73	22.1752		16.6977	8.4085
3	238	Zea mays culti	var:B73	21.9434		17.4338	8.83602
4	268	Zea mays culti	var:B73	18.8663		14.0561	7.96233
5	501	Zea mays culti	var:B73	25.8311		17.2623	8.97662
6	812	Zea mays culti	var:B73	22.9926		11.902	9.0297

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					T
trait					divided by semicolon
Variant ID					divided by semicolon
Chromosome	Any	* between		and	bp
- Log10(P value)	between		and		
Sea	arch		Download CSV		Clear
trait	variantID	↑ chr	🔷 posi	-log(P Value)	▲ action
C180_C200	chr1.s_1079408	1	1079408	5.61376671730847	visualize by trait; visualize by variant;
C160_C180	chr1.s_1921683	1	1921683	5.88796891847108	visualize by trait; visualize by variant;
C160_C180	chr1.s 1922018	1	1922018	5.10452516470998	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								1	divided by sem	nicolon
romosome	Any	¥	betweer	n			and			bp
tance tss	between			а	nd					
a value	between			а	nd					
g10(pvalue)	between			а	nd					
Sea	rch				Downloa	id CSV				Clear
jene 🍝	position	^	genome 🗳	· 4	locus	âbs_c	listance_tss	betavalu	e 🔶 -I 🔶	action
+ ENSRNA049464994	1:7044629-7044	731	B73_RefSeq	-	chr1.s_704	44 475		-0.6704	11.9992	visualize by gene ID;
- ENSRNA049463846	5:3203379-3203	539	B73_RefSeq	-	chr5.s_320	04 1261		-0.5185	7.582	visualize by gene ID;
locus_id	^	abs_dis	stance_tss		^	betavalu	e	^	-log10(pva	lue)
locus_id filter column	•	abs_dis Min	stance_tss	Max	^	betavalu Min	e Max	·	- log10(pva Min	l ue) Max
locus_id filter column chr5.s_3203074		abs_dis Min 465	stance_tss	Max		betavalu Min -0.4165	e Max		-log10(pva Min 10.1494	Max
locus_id filter column chr5.s_3203074 chr5.s_3204800		abs_dis Min 465 1261	stance_tss	Max		betavalu Min -0.4165 -0.5185	e Max		-log10(pva Min 10.1494 7.582	Max
Iocus_id filter column chr5.s_3203074 chr5.s_3204800 chr5.s_3230716		abs_dis Min 465 1261 27177	stance_tss	Max		betavalu Min -0.4165 -0.5185 0.4642	e Max		-log10(pva Min 10.1494 7.582 8.433	Max
Iocus_id filter column chr5.s_3203074 chr5.s_3204800 chr5.s_3230716 chr5.s_3231267		abs_dis Min 465 1261 27177 27728	stance_tss	Max		betavalu Min -0.4165 -0.5185 0.4642 0.4684	e Max		-log10(pva Min 10.1494 7.582 8.433 9.1799	Max

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	pecies	Map	Linkage Group	Genetic Location	Genomic Location
avancea marker scaren			1 abph1.15	SNP Ze	ea mays cultivar.873	MO17xTeo_BC2F5	MO17xTeo_BC2F5.2	91.6	2:28867080.2886708
			2 abph1.15	SNP Ze	ea mays cultivar.873	ZHENG58xSK_F6	ZHENG58x5K_F6.2	81.6	2:28867080.2886708
and the second second second second	Wild card (*) can be used to match any OTU	label	3 abph1.15	SNP Ze	ea mays cultivar.873	MO17xTeo_BC2F5	MO17xTeo_BC2F5.2	91.6	2:28867080.2886708
Advanced Marker Search QIL Nearby Markers		aben.	4 abph1.15	SNP Ze	ea mays cultivar.873	ZHENG58xSK_F6	ZHENGS8xSK_F6.2	81.6	2.28867080.2886708
	OTL		5 abph1.22	SNP Ze	ea mays cultivar.873				
Search by Name			6 ae1.4	SNP Ze	ea mays cultivar.873				
			7 ae1.8	SNP Ze	ea mays cultivar.873	873x8Y804_RIL	873x8Y804_RIL5	104.3	5:17242449917242
Marker Name contains *	Distance	cM	8 an1.3	SNP Ze	ea mays cultivar.873				
File Upload Choose File No file chosen	Search Reset	C	2	SEQUENCE	s				
Provide marker names in a file. Separate each name by a new line.				Sequence		AGGETGAADGA AAGGATCAAC[SATCECCGTGGTGATCATGTECT	CAGAGAACGTGCCGAC AACGCATATACCTMTT	÷
Marker Type Any			abph1.22						
Marker Type Any Marker Mapped in Genome Any Any	pecies		abph1.22 _{Summary}	/ ,	SUMMARY				
Marker Type Any Marker Mapped in Genome Marker Developed from S Any	pacies		abph1.22 Summary Sequence	,	SUMMARY Resource	Туре		Genetic Ma	arker
Marker Type Any Marker Mapped in Genome Marker Developed from S Any Restrict by Location	pecies		abph1.22 Summary Sequence Annotations	,	SUMMARY Resource Accession	Туре		Genetic M.	arker
Marker Type Any Marker Mapped in Genome Any Restrict by Location Genome Location Any between	pedes	Бр	abph1.22 Summary Sequence Annotations Contact	,	SUMMARY Resource Accession Organism	Туре		Genetic M. Zea mays (arker Lultivar:873
Marker Type Any Marker Magned In Genome Marker Developed from S Any Restrict by Location Genome Location Map Any	and	bp	abph1.22 Summary Sequence Annotations Contact Cross Refere		SUMMARY Resource Accession Organism Name	Туре		Genetic M. Zea mays o abph1.22	arker cultivar:873
Marker Type Any Marker Magned In Genome Any Restrict by Location Genome Location Map Any Chr Number Any	pecies and	bp	abph1.22 Summary Sequence Annotations Contact Cross Refere Location on	ence Map	SUMMARY Resource Accession Organism Name Identifier	Туре		Genetic M Zea mays o abph1.22 abph1.22	arker Lultivar:873
Marker Type Any Marker Type Any Marker Napped in Genome Any Restrict by Location Genome Location Map Any Chr Number Any	and	bp	abph1.22 Summary Sequence Annotations Contact Cross Refere Location on Publication	ence Map	SUMMARY Resource Accession Organism Name Identifier Is Analysi	Type		Genetic Mi Zea mays of abph1.22 abph1.22 No	arker tultivar:873

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).

A												C				
	Genome	B73				Strand		All			*	U	Cas9_sgRNA			
													Primary Data			
	Gene ID	Zm00001d0	27230								۵		Туре	Cas9_sgRNA		
	Region	chr	1		between			and					Score	0.406026		
			-	_				_			_		Position	1:4430144301 (+ 1	trand)	
	Sea	rch			Downlo	ad CSV				Clear			Attributes			
	GenelD	Chr.	Car .	Deci	Crient	C	047	047.	0457.	Minm	Action A		Crispr seq	CGTGCGGGGCCTC	ICTCCTC+TGG	
	Genero	Chir	Su	POSI	crispr	Score	On la	Onna	On la	wiisin	Action		Mismatch	4		
	Zm00001d027230	1	2	44310	GAGGGAA	0.609387	Zm00001d	10:+88570	GGGGGAA.	. 4	Jbrowse		Off target cr	sprseq CGGGGGGGGG	CCTCCCTCCTC+C#	VG
	Zm00001d027230	1	-	44317	AGGGCTG	0.589956	Zm00001d	10:+12906	AGGGCTC	4	Jbrowse		Off target in	0 10+104273688		
	Zm00001d027230	1		44328	GACCACC	0.581558	Zm00001d	10:+11436	GAGCACC	4	Jbrowse		Sgrna id	Cas9sgR_1:+44301		
	Zm00001d027230	1	-	44336	GCGGCCAT	0.586449	Zm00001d	9:-20629500	GTGTCCTT	4	Jbrowse		Target gene	d Zm00001d027230		
	7m00001d027230	1		44337	GGCGGCC	0.620183	Zm00001d	10:+14094	66066	4	Ibrowse		Region sequ	ence		
				11007		0.020100	Linocoordan	1011110510	/		10.01120		Not displaying	inderlying reference seque	nce, feature is longer	than maximum of 0 bp
D									/				-			
D	0 20.000.000	40.000.000	60 000 000	80.000.00	0 100.000	120.0	00.000 140	000.000 1	30.000.000	180 000 00	200.000.000	29000	240.00	260 000 000	280.000.000	300.000.000
	20,000,000			00,000,00				A	4-44470 4	4984 (804 b)		~ Ø	.,		200,000,000	000,000,000
	4	1 250			44 375		444	44 500	1.441724	4804 (09419)	44 625	00 00		44 750		
	C Reference sequence															
									m							
	G Gene Zm0000	14027230_T001		_												
	2m000	010027230		-												
	Crispr		0.1	++ +++	4 + 4 +	+ +	+ ++++	++ ++	++ ++		++++	+ + +	+ + + +	• • • • • • •	H H	+++ ++
				+++++++++++++++++++++++++++++++++++++++	4 4 4	-	1		4 4	11 1	<u> </u>	44.	4 4 4 4	4 4 4 4 4 1	1	1
					1	-	-	H.	-i -	-1-1			1	ને ને	•	-1
				+			-	-		-				-		
							÷									

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 files annotation (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-b5-v5-a4-m7-FF". 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 - FF". 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 (<u>https://github.com/statgen/locuszoom</u>), a JavaScript embeddable plugin for interactively visualizing statistical genetic data, and ECharts (<u>https://www.echartsjs.com</u>), an open-sourced JavaScript visualization tool. The gene expression pattern viewer and the eQTL visualizer were modified from GTEx visualizations (<u>https://github.com/broadinstitute/gtex-viz</u>) (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.