

# Homotools: a suite of genomic tools for homologous retrieval and comparison

Hui Liu<sup>1</sup>, Olamide Adesina<sup>1</sup>, Ravi Bika<sup>1</sup>, Rishabh Singh<sup>2</sup>, Mithila Jugulam<sup>2</sup> and Sanzhen Liu<sup>1\*</sup>

<sup>1</sup> Department of Plant Pathology, Kansas State University, Manhattan, KS 66506-5502, USA

<sup>2</sup> Department of Agronomy, Kansas State University, Manhattan, KS 66506-5502, USA

\* Corresponding author, E-mail: [liu3zhen@ksu.edu](mailto:liu3zhen@ksu.edu)

## Abstract

Genome sequencing and assemblies offer fundamental data for comprehensively exploring genomic variation among individuals. Genomic variation in genic regions is of particular interest. However, identifying homologous genes or sequences and DNA variation, including structural variation, from multiple genomes is a tedious process. Here we present the software package Homotools which includes multiple modules for retrieval of best-hit homologs, variant discovery, variant annotation, and visualization of structural comparison. These modules facilitate all these processes and leverage genomic resources for single-gene studies. The tools can be used for any species as long as assembled genomes and their genome annotations are available. In a case study using Homotools, it is shown that tolerance to the herbicide nicosulfuron is associated with multiple independent genomic variants found in various maize inbred lines, including structural variants due to insertions of transposable elements. The results from Homotools generate valuable testable hypotheses for further examination. Scripts of all modules are publicly available in GitHub ([liu3zhenlab/homotools](https://github.com/liu3zhenlab/homotools)).

**Citation:** Liu H, Adesina O, Bika R, Singh R, Jugulam M, et al. 2024. Homotools: a suite of genomic tools for homologous retrieval and comparison. *Genomics Communications* 1: e002 <https://doi.org/10.48130/gcomm-0024-0002>

## Introduction

With the advance of high-throughput long-read sequencing and the development of effective genome assembly algorithms, the community of genomics has constructed a wealth of high-continuity genome assemblies, frequently in the form of chromosome-level pseudomolecules, of individuals from a wide variety of species<sup>[1,2]</sup>. Such genome assemblies provide fundamental resources for the identification of genomic variation, including single nucleotide polymorphisms, small insertion and deletion, and structural variation<sup>[3,4]</sup>. Genomic computational tools have been developed to facilitate genome alignments and variant discovery. For example, NUCMER and minimap2 are two effective sequence aligners that could handle alignments between genome sequences<sup>[5,6]</sup>. For alignments between complex genomes with large structure variation, AnchorWave was designed for improved alignment accuracy<sup>[7]</sup>. Genomic alignment results can be further processed to identify genomic polymorphisms<sup>[8]</sup>. Such analyses are frequently conducted for global genomic comparisons although the process is computationally intensive.

Genomic variation in and around genes is of particular interest. Without knowing homologous genes across genomes, extracting data related to single genes and their homologs from the genome-wide comparative results among multiple genomes is not straightforward. In a scenario where genes of all genomes are well annotated, genic comparison could be directly conducted by aligning sequences of gene models at the DNA and protein levels. However, complete and precise gene annotation is a daunting task, and alternative splicing of a gene complicates such analysis. In addition, with substantial resources of genome assemblies, multiple rounds of sequence retrieval, homolog search, variant identification, and annotation are needed for examining a gene of interest and its homologs. The process also typically involves manual examination and judgment. Therefore, a handy computational pipeline is needed for an efficient process. An online tool, BridgeCereal, represents one such effort to discover structural variation among multiple genomes<sup>[9]</sup>. Homotools, including a suite of scripts was developed to ease sequence retrieval of homologs and comparisons among

them from multiple genomes. Comprehensive genomic variation can be uncovered, annotated, and visualized by Homotools.

## Materials and methods

### Genomic data

Maize genomes of 28 inbred lines, including A188, B73, Mo17, and other 25 parents of the Nested Association Mapping (NAM) population, were used for analysis<sup>[10–12]</sup>. All data can be downloaded from <https://download.maizgdb.org>.

### Development of Homotools

The majority of scripts were developed using Shell scripting and Perl. Plotting was implemented through R. Bioinformatics tools frequently used include Bedtools<sup>[13]</sup>, BLAST+<sup>[14,15]</sup>, and NUCMER<sup>[5]</sup>.

### Plant materials and herbicide treatment

Maize inbred lines were provided by the North Central Regional Plant Introduction Unit, USDA, USA. Ten seeds per inbred line were sown in 10.8 cm × 10.8 cm × 9.5 cm pots. The plants were cultivated in greenhouse conditions under 28/21 °C day/night (d/n) temperature and a 16/8 h d/n photoperiod, supplemented with 600 μmol m<sup>-2</sup> s<sup>-1</sup> illumination provided by sodium vapor lamps. Seedlings at the 3–5 leaf stage (2- to 3-weeks old) were used for herbicide treatment. At least two seedlings from each inbred line were sprayed with nicosulfuron (Accent Q®, DuPont, Wilmington, USA) at a dose of 137 g active ingredient per hectare. Herbicide application was performed using a bench-track sprayer (DeVries Manufacturing, Hollandale, MN, USA) equipped with a flat-fan nozzle (8002 TeeJet® tip, Spraying Systems, Wheaton, IL, USA), calibrated to deliver 187 liters per hectare at a speed of 4.85 km h<sup>-1</sup> at 207 kPa pressure<sup>[16]</sup>. Two weeks after treatment, the plants were phenotyped and photographed.

## Results

### Overview of the Homotools package

Homotools includes five major modules, geneseq, homocomp, homograph, homostack, and homomine (Fig. 1). Module geneseq can

extract gene sequences, including sequences of genes, transcripts, and coding regions based on a query gene name and related databases. Module homocomp implements homologous searching and visualization of the comparison between the query and the target sequences. A homologous gene could be found if a related gene annotation is supplied. The alignment between the two homologous sequences is plotted. Module homostack can stack input sequences and visualize alignments between two neighboring sequences. Module homograph visualizes alignments of multiple homologous sequences, determines the number of haplotypes, and views the haplotype alignment graph. All scripts are publicly available in GitHub (liu3zhenlab/homotools).

These modules can be combined. For example, module geneseq is typically run first to extract genomic and related sequences for an input gene from a reference genome, with which module homocomp can be run multiple times to extract homologs from other genomes. For an easier implementation, module homomine was developed to combine geneseq and homocomp for sequence retrieval, homologous search, and annotation of variants, including nucleotide substitutions, small insertion, small deletion, and large structural variation. In this manuscript, example codes to illustrate the usages of these modules are provided (Supplemental File 1).

### An application example of geneseq, homocomp, and homostack

Here, homologous genes of Zm00001eb001720 (*knox1* encoding a homeobox protein) from the inbred line B73 in the other four inbred lines A188, CML333, Ki3, and P39 were examined<sup>[17]</sup>. First, the genomic sequence of the gene and the annotation were extracted from the B73 reference genome through module geneseq. The lengths of sequences flanking the gene can be set based on the purpose. Through implementing homocomp, the *knox1* sequence was then used to find a homologous sequence in other genomes. The alignments between the *knox1* sequence and a homologous sequence from the A188 genome are visualized in an alignment connection format (alnplot) (Fig. 2a) and in a dotplot view (Fig. 2b). From the alnplot output figure, both structural variation and polymorphic levels of alignments can be informed.

In module homocomp, BLAST+ is used for an initial alignment search for the query sequence in a targeted genome<sup>[14]</sup>. The genomic region with the highest alignment score (bit score) with the query is anchored and then the region is expanded from both sides to identify the alignment start and end sites. Alignments in repetitive regions could be excluded for the identification of the highest-score alignment if the corresponding parameter is specified. The sequence in the alignment range identified by BLAST+ is then

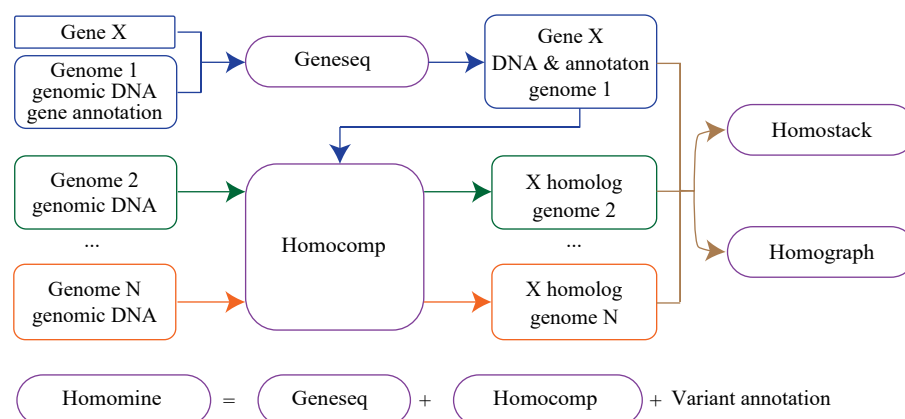
compared with the query sequence using NUCMER, which identifies pairs of sequence blocks and reports their similarities<sup>[5]</sup>.

Multiple sequences retrieved using homocomp from the other genomes of the four inbred lines, along with the query sequence, can be used for sequential sequence comparisons *via* module homostack (Fig. 2c). The order of these sequences is based on the input order of these sequences. Sequence regions of any sequences can be highlighted with additional annotation inputs. The gene annotation, including exon and coding regions, of *knox1* is color-highlighted (Fig. 2c). From the homostack result, P39 and A188 alleles are similar; Ki3 and CML333 alleles are structurally similar but the identity is relatively low; and three presence and absence variations (PAV) were observed among these five alleles from the five maize inbred lines.

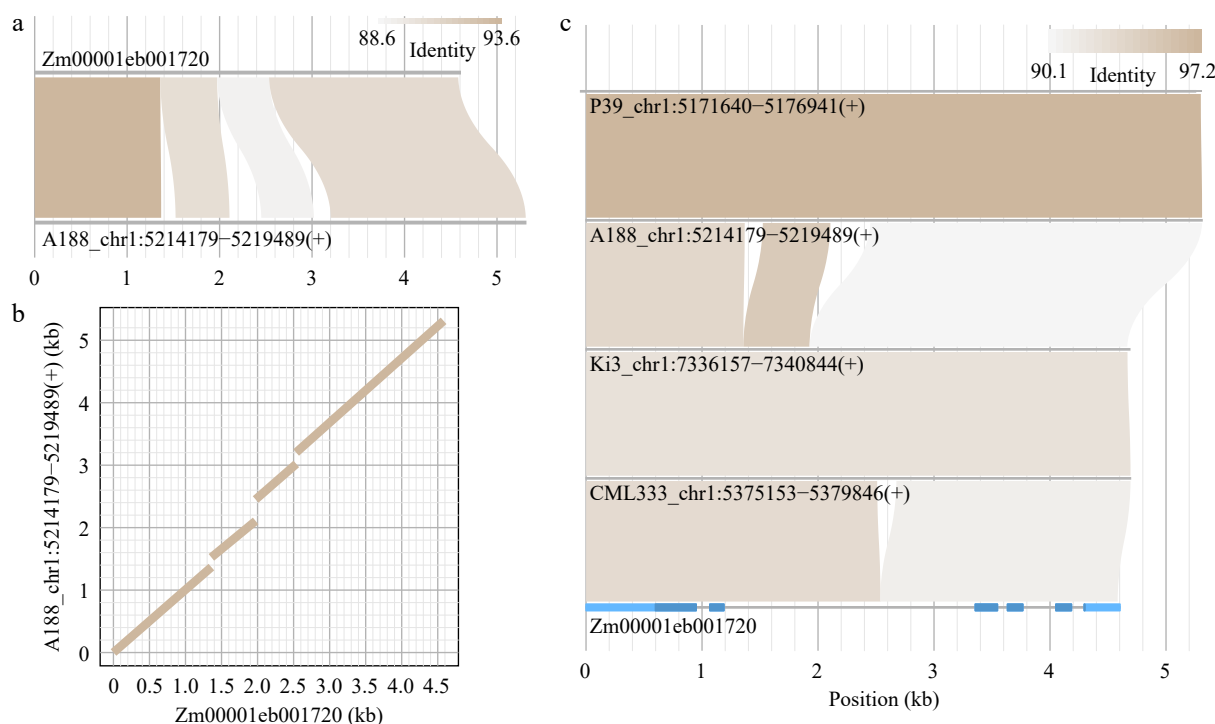
### Homograph for clustering and sequence block analysis

To provide detailed polymorphisms among multiple sequences, the module homograph was developed. The homograph algorithm starts with multiple sequence alignment (MSA) with input DNA sequences (or taxa), including a reference sequence. Three MSA software packages are available: Clustal Omega<sup>[18]</sup>, MUSCLE<sup>[19]</sup>, or MAFFT<sup>[20]</sup>. All sequences are clustered with cd-hit based on the user-defined clustering criteria<sup>[21]</sup>. The non-gap consensus sequences of MSA are then split into blocks at all transitions between a gap and a nucleotide base at the end of an alignment. In each block, the number of polymorphisms and the number of haplotypes are determined. An informative block graph is plotted to represent MSA (Fig. 3a). Blocks are color-coded to indicate the sequence conservation and the polymorphic levels as compared to the consensus sequence. The haplotype sequence of each block is named and stored in a FASTA output (Fig. 3b) and the genotyping result per block in each taxon is output (Fig. 3c).

Using sequences from the previous *knox1* case, the five sequences were grouped into three clusters using the default criteria: c1, c2, and c3. The alleles of A188 and P39 were grouped to c1, the CML333 and Ki3 alleles were grouped to c2, and the B73 allele was in c3. Consistently with the homostack alignment output, three PAV were observed among these five alleles. As compared with the homostack output, more details related to DNA substitution polymorphisms and small insertions or deletions (INDELs) are provided in the homograph output (Fig. 3d). Briefly, DNA substitution polymorphisms spread in most regions. Only a few small regions are conserved across all five sequences, and interestingly, most of the



**Fig. 1** Overview of Homotools. Module geneseq extracts sequences and related information from databases for a gene (e.g., gene X); homocomp implements homolog searching for gene X and visualizes the comparison of two homologs; homograph determines haplotypes of multiple homologs, and displays an alignment graph of a set of homologs; and homostack stacks input sequences and shows alignments; Module homomine is a pipeline implementing modules geneseq and homocomp for homologous searching and annotating all variants between the query sequence and the resulting homolog.



**Fig. 2** Alignment visualization via homocomp and homostack. (a), (b) A target homologous sequence of the query sequence (Zm00001eb001720) was identified in the A188 genome using module homocomp. (a) An alignment plot, and (b) a dotplot were output. (c) Multiple sequences retrieved by homocomp, along with the query sequence, were supplied to homostack to show the sequential alignments. The gene structure of Zm00001eb001720 was plotted, in which light and dark blue represent untranslated exon regions and coding regions, respectively.

conserved sequences are not in coding regions. No INDELs were observed in the protein-coding regions.

### The integrated pipeline homomine discovers and annotates genomic variants

An integrated module, homomine, was developed to simplify the procedure for sequence retrieval, homologous searching, variant calling, and annotation of variants, including structural variants. For the example of Zm00001eb001720, the input gene name was used to retrieve genomic DNA sequence data and the gene annotation in the query genome (e.g., B73) first. The homologous region and the homologous gene will then be identified in a target genome (e.g., A188). In the result, alignments between two homologous regions will be visualized (Fig. 4a), and the coordinates of detailed alignments are provided (Fig. 4b). To discover variants, including large insertions and deletions, a script to convert alignments using an MSA algorithm to genomic variants was developed. All genomic variants were annotated using SNPEff<sup>[22]</sup>. In particular, genomic variants annotated to cause moderate or high impacts on the gene function were extracted and reported in the outputs (Fig. 4c). These genomic variants, particularly high-impact variants, likely contribute to form a functionally divergent allele. The homomine outputs are extracted to produce a HTML report for overviewing the major results (Supplemental Dataset 1).

To facilitate the application of homomine, a web-based Shiny application for homologous retrieval and variant annotation for maize genes from the B73 reference genome was developed. The webpage (129.130.89.83:3838/maizeHM) is publicly available. Each B73 gene can be input to find homologs in 27 other maize genomes<sup>[10–12]</sup>. A HTML report is generated for each of the 27 genomes.

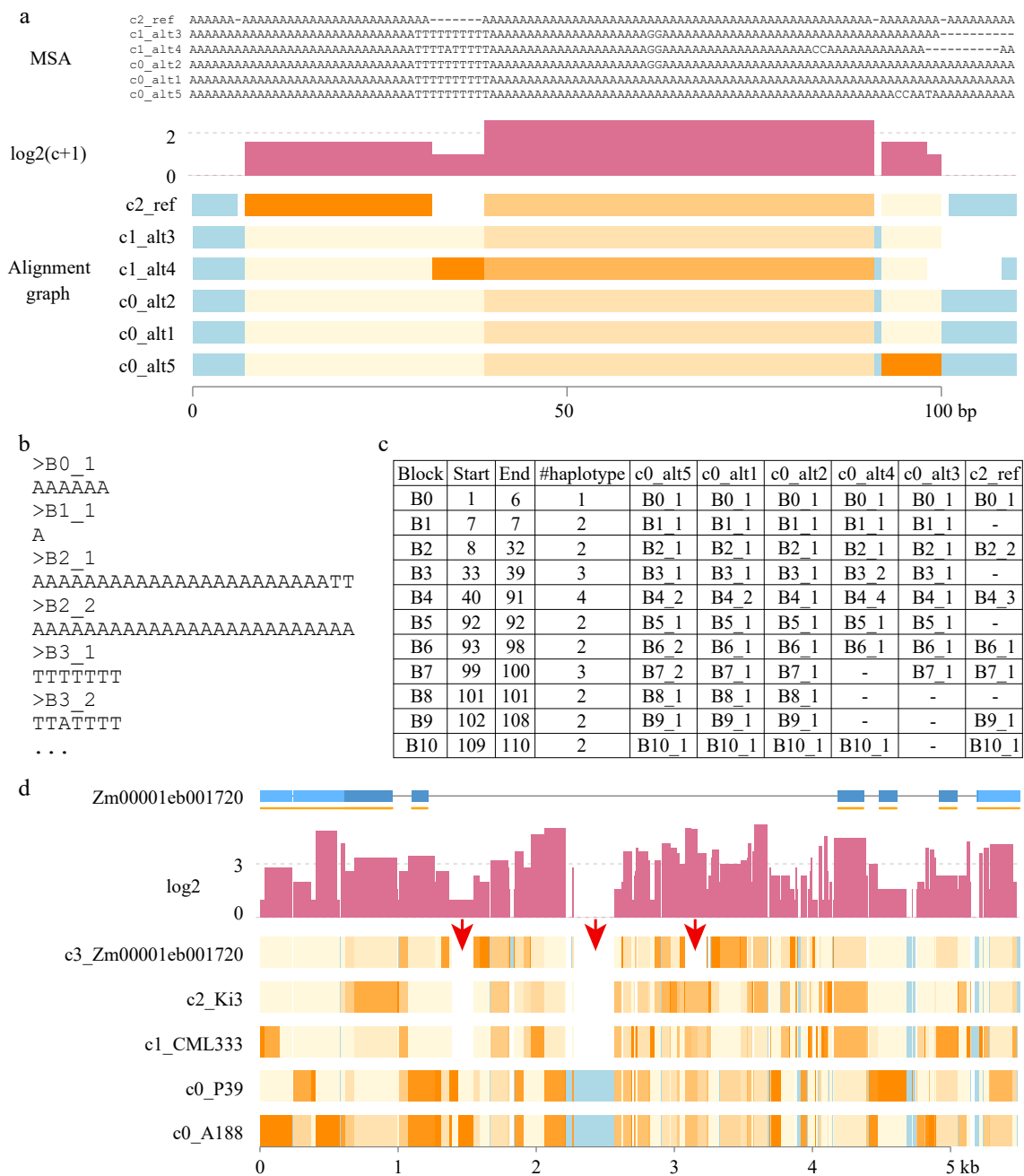
### Genomic variation of nicosulfuron tolerance gene in multiple genomes

The maize gene, *nsf1* or *ben1*, encoding a cytochrome P450 monooxygenase (CYP81A9) confers tolerance to a herbicide

nicosulfuron<sup>[23–26]</sup>. The B73 allele (Zm00001eb214410) is a functional allele that contributes to the tolerance (Fig. 5a). Nineteen maize inbred lines were tested, including B73, for their sensitivity to nicosulfuron (Supplemental Table S1). Three sensitive lines were identified, namely, A188, CML228, and Il14H (Fig. 5b–d). Homomine analysis using the B73 allele as a query found that the A188 allele contains an insertion of 392 bp at the end of the first exon and that the CML228 allele contains an insertion of 5512 bp in the middle of the same exon (Fig. 5e, 5f). Both insertions were associated with the *Gypsy* retrotransposon and predicted to cause shifts of the open reading frames, presumably disrupting the gene function of herbicide tolerance. As compared with the B73 allele, multiple structural variation events were found in the Il14H allele, including a 120-bp insertion, a 297-bp insertion in the upstream of the transcription start site, and the absence of the partial 3' untranslated regions (3'UTR) (Fig. 5g, Supplemental Table S2). These structural variants can also be found between at least a nicosulfuron tolerance inbred line and B73, indicating they may not be the causal variants causing the loss of herbicide tolerance (Supplemental Table S2). Two variants of Il14H exhibiting moderate impacts on the gene function are not in any of 16 herbicide tolerant lines: a SNP causing a missense alteration and a 6-bp insertion introducing a 2-amino-acid inframe insertion (Supplemental Table S2). One of the two variants is likely the causal variant for herbicide sensitivity. In this case, through using homomine, pangenome data were leveraged for exploring the genomic variation of genic regions, facilitating the association of all types of genomic variants with phenotypic data.

### Discussion

The development of Homotools was motivated by the tedious procedure for retrieving sequence data and related information of a gene from a reference genome and homologous genes from other genomes. Multiple modules have been developed to simplify identification and comparison of homologs. The tools are particularly

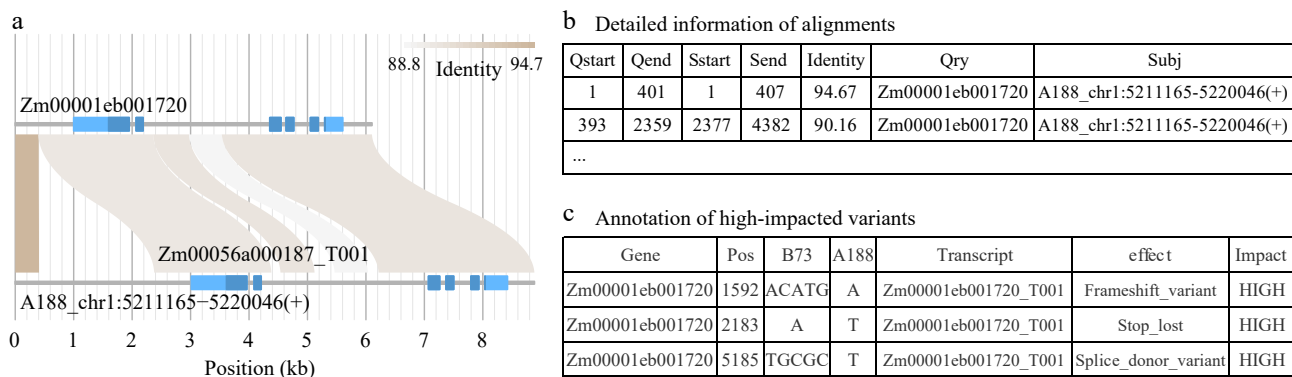


**Fig. 3** Homograph outputs. (a) Three parts are included: MSA, log2(c+1), and Alignment graph. MSA displays a graphic view of multiple sequence alignment (MSA) for helping understand the alignment graph. The MSA was divided into blocks (the transition between a gap ('-') and a nucleotide base). For each block, the single nucleotide substitution sites, excluding gaps, were counted. The Y-axis unit is at the scale of log2 of (c) the substitution count, which is adjusted by adding 1. In the alignment graph, each rectangle represents an alignment block without a gap. Light blue represents conserved regions with no substitution polymorphisms. The gradient colors from light yellow to orange indicate the divergent levels from the consensus sequences, from low divergence to high divergence, respectively. Names of taxa are prefixed with cluster groups (e.g., c0). (b) The allele sequence of each block is stored in a FASTA file. (c) A table output lists the allele type of each taxon at each block. The Start and End coordinates are the MSA alignment positions. The number of haplotypes (#haplotype) specifies the number of alleles of a block, which includes the allele of a gap if it exists. (d) A graphic output of homograph analysis of *knox1* homologs. Red arrows point at regions with presence and absence variation. The gene structure of Zm00001eb001720 was plotted, in which light and dark blue represent untranslated exon regions and coding regions, respectively. Each orange line under the gene structure indicates an exon, which could be split by alignment gaps.

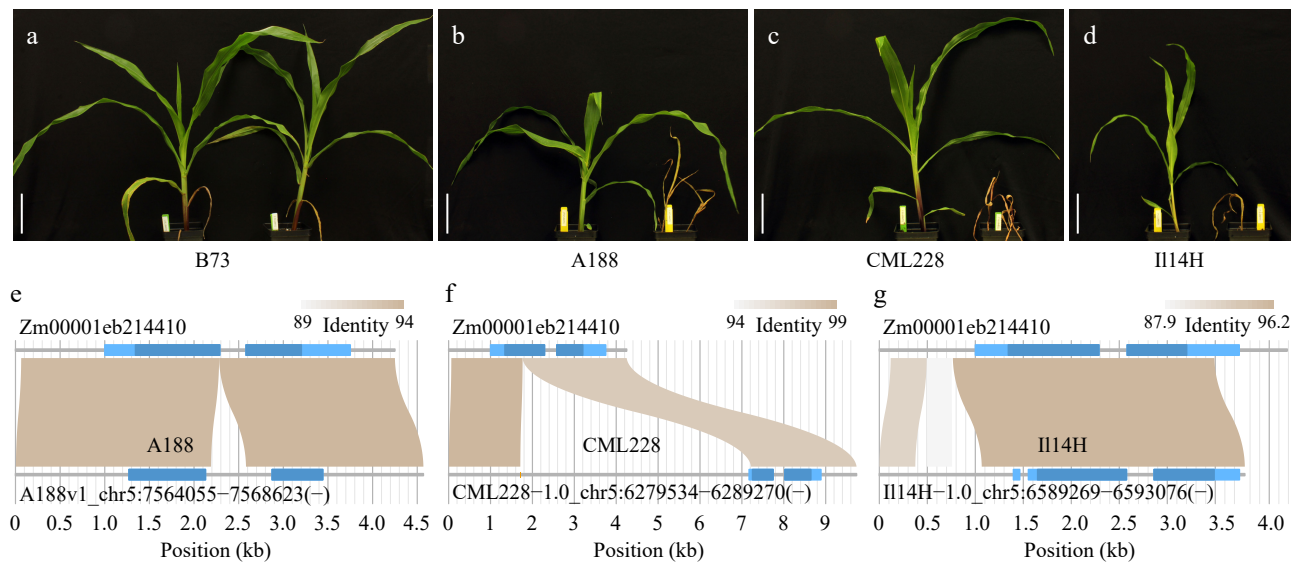
useful for single-gene genomic studies of a species with multiple genome sequences. Genomic databases with standard formats (e.g., FASTA and General Transfer Format, GTF) are required. Therefore, the tools can be used for genomic analysis for any species. With Homotools, pangenomic data related to an input gene can be readily collected and related genomic variation, including structural variation, can be accurately identified and annotated. In addition, publishable high-resolution figures are output from Homotools. Specifically,

module geneseq can extract sequences of genomic DNA, coding regions, proteins, and related transposable elements as long as databases are supplied. Modules homocomp and homomine facilitate homologous retrieval, structural comparison, and variant identification. Modules homostack and homograph graphically visualize sequential alignments and multiple sequence alignments, respectively. Because many software dependencies are required, guidance for creating a Conda environment is provided.





**Fig. 4** Partial outputs from module homomine. (a) Homomine searches homologous sequences of a query gene Zm00001eb001720 in a targeted genome (A188). The flanking 1 kb on the 5' end and 500 bp on the 3' end are included. Homomine identifies the matching region that harbors Zm00056a000187 in the A188 genome. Light and dark blue in gene structures represent untranslated exon regions and coding regions, respectively. (b) In the homomine output, detailed alignments between the two homologs are provided. (c) The homomine output lists variants identified between the two homologs, including structural variation. Variants are annotated and moderate- and high-impacted variants are highlighted in the output. The table shows the annotation of variants with a high impact.



**Fig. 5** Phenotypic results of nicosulfuron treatments and allelic comparisons. (a)–(d) Maize plants with nicosulfuron treatment at a dose of 137 gram active ingredient per hectare (right) and a mock treatment (left). Bar = 10 cm. (e)–(g) The structural comparison between each allele of A188, CML228, and II14H with the B73 allele of gene Zm00001eb214410. Light and dark blue in gene structures represent untranslated exon regions and coding regions, respectively.

Modules of Homotools could be further improved to enhance computational effectiveness and efficiency. First, the mapping of a query gene to a genome and identification of homologous genes in the homocomp and homomine modules could fail due to repetitive sequences. Approximately 3% of genes are subject to this issue in highly repetitive maize genomes. The issue may be mitigated by specifying the target chromosome and/or a rough target region. Alternatively, a sophisticated algorithm such as the algorithm of CHOICE (Clustering HSPs for Ortholog Identification via Coordinates and Equivalence), for identifying homologous regions in a genome could be adapted in the future<sup>[9]</sup>. Second, the variant annotation in module homomine relies on the software SNPEff, which groups variants into high, moderate, low, and modifier impacts<sup>[22]</sup>. Most moderate-impacted variants are missense polymorphisms for which the functional impacts can be further quantified using other annotation tools including artificial intelligence (AI) based approaches developed to assess the functional impacts of missense polymorphisms<sup>[27,28]</sup>. Third, homostack performs alignments and plotting based on the order of DNA sequences that users input. Users may be uncertain about the order to input. An algorithm

could be added to enable automatic ordering by determining pairwise similarities of input sequences. Fourth, homologous sequences of a gene collected using Homotools from pangenomes could be used to build a pangenome graph of genic sequences<sup>[29]</sup>. Module homograph visualizing alignments of MSA is capable of tackling simple structural variations such as insertion and deletion. Inversion and translocation would be better represented in more complex graphical structures<sup>[30]</sup>. In addition, the development of algorithms aiding in hypothesizing evolutionary trajectories, events of mutation and recombination among homologous genes would be highly valuable.

In summary, it has been shown that Homotools can be useful for readily identifying best-hit homologous genes and genomic polymorphisms at the gene level. An online application has been developed that is tailored for the maize community, with the potential to extend similar platforms to a wide range of species. As the Homotools package continues to be refined and enhanced to ensure greater reliability and efficiency, community feedback will play a vital role for future improvement. We look forward to collaborative efforts to further advance the utility of this tool.

## Author contributions

The authors confirm contribution to the paper as follows: herbicide tolerance experiment design and conduction: Liu H, Adesina O, Singh R, Jugulam M; Homotools use and test: Liu H, Adesina O, Bika R; Homotools package development: Liu S; draft manuscript preparation: Liu H, Liu S. All authors reviewed and revised the manuscript.

## Data availability

All related scripts have been deposited at GitHub (liu3zhenlab/homotools).

## Acknowledgments

We thank Jinliang Yang at the University of Nebraska-Lincoln and Won Yim at the University of Nevada for valuable suggestions. We thank funding provided by the USDA NIFA award (2018-67013-28511 and 2021-67013-35724), the NSF awards (1741090, 2011500, and 2311738) and the Department of Energy award (DE-SC0023138). This is a contribution 25-016-J from the Kansas Agricultural Experiment Station, Manhattan, Kansas (USA).

## Conflict of interest

The authors declare that they have no conflict of interest. Dr. Sanzhen Liu is the Editorial Board member of *Genomics Communications* who was blinded from reviewing or making decisions on the manuscript. The article was subject to the journal's standard procedures, with peer-review handled independently of this Editorial Board member and the research groups.

**Supplementary Information** accompanies this paper at (<https://www.maxapress.com/article/doi/10.48130/gcomm-0024-0002>)

## Dates

Received 7 August 2024; Revised 4 September 2024; Accepted 4 September 2024; Published online 19 September 2024

## References

- Li H, Durbin R. 2024. Genome assembly in the telomere-to-telomere era. *Nature Reviews Genetics* 25:658–70
- Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, et al. 2021. Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 592:737–46
- Gong Y, Li Y, Liu X, Ma Y, Jiang L. 2023. A review of the pangenome: how it affects our understanding of genomic variation, selection and breeding in domestic animals? *Journal of Animal Science and Biotechnology* 14:73
- Groza C, Schwendinger-Schreck C, Cheung WA, Farrow EG, Thiffault I, et al. 2024. Pangenome graphs improve the analysis of structural variants in rare genetic diseases. *Nature Communications* 15:657
- Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, et al. 2018. MUMmer4: a fast and versatile genome alignment system. *PLoS Computational Biology* 14:e1005944
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–100
- Song B, Marco-Sola S, Moreto M, Johnson L, Buckler ES, et al. 2022. AnchorWave: sensitive alignment of genomes with high sequence diversity, extensive structural polymorphism, and whole-genome duplication. *Proceedings of the National Academy of Sciences of the United States of America* 119:e2113075119
- Goel M, Sun H, Jiao WB, Schneeberger K. 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology* 20:277
- Zhang B, Huang H, Tibbs-Cortes LE, Vanous A, Zhang Z, et al. 2023. Streamline unsupervised machine learning to survey and graph indel-based haplotypes from pan-genomes. *Molecular Plant* 16:975–78
- Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou S, et al. 2021. *De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes*. *Science* 373:655–62
- Chen J, Wang Z, Tan K, Huang W, Shi J, et al. 2023. A complete telomere-to-telomere assembly of the maize genome. *Nature Genetics* 55:1221–31
- Lin G, He C, Zheng J, Koo DH, Le H, et al. 2021. Chromosome-level genome assembly of a regenerable maize inbred line A188. *Genome Biology* 22:175
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–42
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403–10
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421
- Shyam C, Borgato EA, Peterson DE, Dille JA, Jugulam M. 2021. Predominance of metabolic resistance in a six-way-resistant palmer amaranth (*Amaranthus palmeri*) population. *Frontiers in Plant Science* 11:614618
- Hake S, Smith HMS, Holtan H, Magnani E, Mele G, et al. 2004. The role of Knox genes in plant development. *Annual Review of Cell and Developmental Biology* 20:125–51
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7:539
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792–97
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30:3059–66
- Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–59
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6:80–92
- Liu XM, Xu X, Li BH, Yao XX, Zhang HH, et al. 2018. Genomic and transcriptomic insights into cytochrome P450 monooxygenase genes involved in nicosulfuron tolerance in maize (*Zea mays* L.). *Journal of Integrative Agriculture* 17:1790–99
- Pataky JK, Williams MM, Riechers DE, Meyer MD. 2009. A common genetic basis for cross-sensitivity to mesotrione and nicosulfuron in sweet corn hybrid cultivars and inbreds grown throughout North America. *Journal of the American Society for Horticultural Science* 134:252–60
- Liu X, Bi B, Xu X, Li B, Tian S, et al. 2019. Rapid identification of a candidate nicosulfuron sensitivity gene (Nss) in maize (*Zea mays* L.) via combining bulked segregant analysis and RNA-seq. *Theoretical and Applied Genetics* 132:1351–61
- Zhang Y, Zhang Q, Liu Q, Zhao Y, Xu W, et al. 2024. Fine mapping and functional validation of the maize nicosulfuron-resistance gene CYP81A9. *Frontiers in Plant Science* 15:1443413
- Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, et al. 2023. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* 381:eadg7492
- Brandes N, Goldman G, Wang CH, Ye CJ, Ntranos V. 2023. Genome-wide prediction of disease variant effects with a deep protein language model. *Nature Genetics* 55:1512–22
- Eizenga JM, Novak AM, Sibbesen JA, Heumos S, Ghaffari A, et al. 2020. Pangenome graphs. *Annual Review of Genomics and Human Genetics* 21:139–62
- Guarracino A, Heumos S, Nahnsen S, Prins P, Garrison E. 2022. ODGI: understanding pangenome graphs. *Bioinformatics* 38:3319–26



Copyright: © 2024 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.