# RESEARCH

### **Open Access**

# PPGV: a comprehensive database of peach population genome variation



Yanlin An<sup>1†</sup>, Qiuyan Ban<sup>4†</sup>, Li Liu<sup>1</sup>, Feng Zhang<sup>1</sup>, Shirui Yu<sup>1</sup>, Tingting Jing<sup>3\*</sup> and Shiqi Zhao<sup>2\*</sup>

#### Abstract

Peach tree is one of the most important fruit trees in the world, and it has been cultivated for more than 7,500 years. In recent years, the genome and population resequencing of peach trees have been published continuously, which has effectively promoted the research of peach tree genetics and breeding. In order to promote the further mining and utilization of these data, we integrated and constructed a comprehensive peach genome and variation database (PPGV, http://peachtree.work/home). The PPGV contains 10 sets of published peach tree genome data, as well as genomic variation information for 1,378 peach tree samples (the resequencing data of 1,378 samples were aligned with the high-quality genomes of Lovell, CN14 and Chinesecling, respectively, for mutation detection). A variety of useful and flexible tools, such as BLAST, Gene ID Convert, KEGG/GO Enrichment, Primer Design and Gene function, were also specially designed for searching data and assisting in breeding.

Keywords Peach tree, Genome, Variation, Database, Tools module

#### Introduction

Peach tree (*Prunus persica*) is one of the most representative and widely cultivated crops of the rose family that originated in China [1]. Its fruit has a wonderful flavor and rich nutrients, thus giving the peach fruit high economic and health value. Peach trees were already planted in China at least 7,500 years ago and spread to most of the world over the millennia since. Today, peach is the

<sup>†</sup>Yanlin An, Qiuyan Ban these authors have contributed equally to this work.

\*Correspondence: Tingting Jing jtt0127@163.com Shiqi Zhao zhaosq@zjou.edu.cn <sup>1</sup>Department of Food Science and Engineering, Moutai Institute, Renhuai, China <sup>2</sup>School of Fishery, Zhejiang Ocean University, Zhoushan, Zhejiang 316022, China <sup>3</sup>State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei, China

<sup>4</sup>College of Horticulture, Henan Agricultural University, Zhengzhou, China

fourth largest fruit after apples, pears and grapes, with 1,527,052 hectares harvested worldwide [2, 3].

The advancement of sequencing technology and the rapid reduction of sequencing costs have greatly promoted the research process of plant genome and population evolution. Up to now, more than 1,300 plant genomes have been published and are still growing rapidly (https://www.plabipd.de/index.ep). For peach trees, at least five more peach tree genomes have been released since the first genome was published in 2013 [1, 2, 4-6]. Studies have shown that the genomes of different peach tree varieties have significant differences, such as "LHSM" and "124 Pan" with genome sizes of 257.2 Mb and 206 Mb, respectively [4, 6]. To reveal the genetic diversity and evolutionary history of peach trees, many population resequencing studies have been widely carried out based on the above reference genomes. For example, Cao et al. [7] identified 409 selected genes based on resequencing analysis of 10 wild and 74 cultivated peach, and proved that human breeding activities led to a sharp decline in the heterozygosity proportion of SNPs;



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Guan et al. [8] identified a 1.67 Mb structural variation controlling the shape of peach fruit through genome and resequencing analysis of 149 peach trees; Li et al. [9] and Yu et al. [6] analyzed the domestication patterns and the biochemical basis of fruit flavor in the breeding process of peach trees through resequencing data analysis of 480 and 564 peach accessions, respectively. Extensive large-scale resequencing not only greatly promoted the research progress of peach trees, but also released a large amount of raw sequencing data. Effectively integrating and utilizing these data is undoubtedly quite challenging [10, 11].

For many food and cash crops, various functional databases have been successfully constructed [12–14]. For example, Gui et al. [15] constructed a comprehensive maize multiomics database; based on thousands of soybean germplasm resources, Zheng et al. [16] built a large database containing 65,374,688 SNPs and 10,952,749 InDels. In addition, databases of genetic variation of species such as *Brassica napus*, sorghum and tea plants have been successfully published [17–19]. In the previous research, although some databases of *Rosaceae* plants have been reported (https://www.rosaceae.org/), the database dedicated to the genetic variation of peach tree population has not been constructed.

#### **Materials and methods**

#### Retrieval of genome and resequencing data resources

Before constructing the database, we first obtain peach tree genome sequences, gff files, protein sequences, and a large amount of resequencing data from public databases such as NCBI (https://www.ncbi.nlm.nih.gov/), CNCB (https://www.cncb.ac.cn/) and GDR (https://www.rosaceae.org/). In addition to the ten sets of genomic data (including 'Lovell', 'CN14', 'Chinesecling', 'Pfe', 'Pka', 'Pmi', 'Pda', 'LHSM', 'RYP1', '124Pan'), another 1,378 samples of resequencing data were downloaded for analysis. For the resequencing data, detailed information can be found in the Overview section of this database (http://peachtree. work/overview).

# Identification of the peach tree genome and population genetic variation

In this study, we chose the 'Lovell' (about 220 Mb), 'CN14' (about 222 Mb), and 'Chinesecling' (about 236 Mb) peach tree genomes as reference genomes and used the BWA software to build indexes for each of them [19]. The downloaded sequences were filtered using the following parameters of the fastp software: --length\_required=25 --cut\_front --cut\_window\_size 4 --cut\_mean\_quality 15 [20]. High quality reads of all samples were aligned to the reference genome with BWA software under the default reference. Following that, use the samtools software to convert the SAM file to the BAM format, then sort and

build an index. PCR duplicates were removed using the Picard software [21]. Subsequently, the HaplotypeCaller mode of GATK software is used to identify SNP/INDEL mutation sites with default parameters, and finally the original mutation data set is obtained. Due to differences in the depth of resequencing among populations, we used the --minQ parameter to filter SNPs (--minQ 180) and INDELs (--minQ 50) rseparately, and annotated all the mutation sites by snpEFF software [22]. At the same time, structural variations were identified using the default parameters of the LUMPY software [23].

#### Genomic and protein sequence analysis

We first used EDTA software to perform a preliminary identification of transposons in each genome, and TEsorter software was used to classify unknown types of transposons based on the above identification results, and then further identification was performed using EDTA software to ensure the reliability of the results [24, 25]. The MISA software was used to identify SSR (Simple Sequence Repeat) loci on different genomes, with the specific definition (unit\_size, min\_repeats) parameters as follows: 2–6, 3–6, 4–5, 5–4, 6–4. At the same time, we annotate the protein sequence with eggNOG online website (http://eggnog5.embl.de/#/app/home) and pfamscan software respectively. In addition, transcription factors were further identified and annotated by iTAK website (http://itak.feilab.net/cgi-bin/itak/index.cgi).

#### Integration of other biological analysis tools

On this website, we offer several practical bioinformatics tools such as get sequence, blast, primer design, and enrichment analysis. These tools are primarily implemented by utilizing bioinformatics software and R packages such as seqtk, blast, primer3, and clusterProfiler. In addition, some softwares such as ORFfinder, SSRMMD and TBtools are also used for data analysis or to provide support for this site [26, 27].

#### **Database construction**

We used Flask (a web package encoded by Python) and Dash (a fast Python web framework package) to build the basic framework of this website, and used Hypertext Markup Language (HTML) and Cascading Style Sheets (CSS) to refine the web layout. All interactive functions were implemented through the third-party plugin fefferyantd-components, which provides rich UI components and interactive functions to help developers quickly build interactive web applications. At the same time, we also used the cacheout caching technology to enhance the response speed of the website. All the data were imported into the MySQL database to improve the query speed. Finally, the website was deployed on Ubuntu 22.04 LTS system provided by Tencent Cloud.

#### Results

#### Summary of the PPGV database

In this study, we constructed a comprehensive peach genome and population variation database, which includes genetic variation information for 1,378 peach samples aligned to the "Lovell", "CN14", and "Chinesecling" genomes, as well as genome information from 10 publicly available peach genomes. Apart from the Home and Overview pages, PPGV mainly consists of five functional modules: Variation, Genome, JBrowse, Tools, and Download. In the following, we will provide a detailed introduction to these modules (Fig. 1). However, it is particularly important to note that due to some studies only publishing genome sequences without accompanying protein sequences, gff files, or accompanying files that cannot be used, it is difficult to perform transcript factor, collinearity, and other analyses.

#### Variation module

In the current module, users can choose any one of the three reference genomes, "CN14", "Lovell", and "Chinesecling", and then search for SNP, INDEL, or SV variations in different samples (Fig. 2A). According to the different types of mutations, we provide diverse search modes. For SNP and INDEL, after selecting the reference genome, further selection of the sample name and chromosome information is required, and the range interval of the starting and ending positions of the variant site needs to be entered (Fig. 2B). After that, clicking the "Submit" button will submit the query conditions and view the corresponding results (Fig. 2C). At the same time, users can enter the gene ID to further filter the above search results to view specific gene-related mutation sites. They can also query specific types of mutations based on SnpEff annotation results, such as "splice acceptor variant" and "MODERATE" (Fig. 2D). For INDEL and SV, users can also filter the query results based on the mutation type and length (Fig. 2E).

#### Genome module

In the Genome module, PPGV provides three search functions including TE/SSR/Synteny, Function search, and Transcription factor (Fig. 3A). Taking the TE/SSR/ Synteny sub-module as an example, when users enter the interactive interface, they need to select the genome first, and then click the submit button to get the corresponding results (Fig. 3B). Based on the result list, users can further filter the results based on chromosome or the type of transposon/SSR (Fig. 3C). Meanwhile, through the Genomic Synteny function, users can query the tandem duplicated genes within different peach tree genomes and the collinear gene pairs between genomes (Fig. 3D). In the Function search sub-module, users can easily obtain the GO, KEGG and Pfam annotation information of different peach tree genomes, and visualize the gene structure of specific genes (Fig. 3E). In addition, we provide a transcription factor search function. Users can not only quickly download all the sequences of specific transcriptome factors in each genome, but also observe the quantitative characteristics of transcription factors in different peach genomes through column charts (Fig. 3F).

#### **Tools module**

To enhance the functionality of the PPGV site, we have configured several commonly used biology tools in the Tools module, including GetSequence, BLAST, IDConvert, orfFinder, Primerdesign, PolySSR, and EnrichmentAnalysis. In this module, users can not only obtain the corresponding sequences according to the input location information or gene ID (Fig. 4A), but also perform blastn, blastp and TEblast analysis through the alignment function (Fig. 4B). At the same time, through the gene ID converter, homologous genes of different genes can be obtained (Fig. 4C). In addition, primer design and identification of polymorphic SSR sites are also built into the Tools module, and users can enter sequences and parameters and click the submit button to obtain the corresponding results (Fig. 4D, E). It should be noted that for PPGV, all input sequences need to be in Fasta format. Moreover, PPGV can also perform orffinder and KEGG/ GO enrichment analysis. As shown in Fig. 5A, after selecting the GO/KEGG enrichment analysis, the user can upload the target gene set or paste it directly into the input box, and then select the more detailed analysis parameters on the right side of the page and click submit, and wait for a short time to download the corresponding analysis results (Fig. 5B).

#### JBrowse, download and overview module

Due to the volume of population variation data obtained based on three different reference genomes exceeding 1 Tb, therefore, in the JBrowse module, we only provide the browsing function for genome and GFF files to ensure a good user interaction experience. The Overview module displays a list of all samples and density and quantity statistics graphs related to variations (Fig. 5C). Especially through the download module, users can not only download SNP/SV and INDEL mutation files of all samples, but also obtain genome sequences, protein sequences, gff files, and SSR locus file (Fig. 5D). All these genomic data sets and resources promote the rapid and simple genetic analysis of peach crops.

#### Discussion

In recent years, the explosive growth of genome and other sequencing data has promoted the progress of plant database research. More and more databases are being released that integrate massive amounts of data [28–30].

# 😡 PPGV 1.0

Home



#### Welcome to Peach tree population genome variation database.

Peach tree is one of the most important fruit trees in the world, and it has been cultivated for more than 7500 years. In recent years, the genome and population resequencing of peach trees have been published continuously, which has effectively promoted the research of peach tree genetics and breeding. In order to promote the further mining and utilization of these data, we integrated and constructed a comprehensive peach genome and variation database (PPGV). The PPGV contains 10 sets of published peach tree genome data, as well as genomic variation information for 1378 peach tree samples (The resequencing data of 1378 samples were aligned with the high-quality genomes of Lovell, CN14 and Chinesecling, respectively, for mutation detection). A variety of useful and flexible tools, such as BLAST, Gene ID Convert, KEGG/GO Enrichment, TE blast, Primer Design and Gene function identification, were also specially designed for searching data and assisting in breeding. We hope you enjoy using PPGV and cite it if you find the data benifit for your research. CHEERS! Due to the lack of protein sequences or Gff/Gft files in some genomes, transcript factors, TEs, and enrichment analyses cannot be performed.





zhaosq@zjou.edu.cn jtt0127@163.com

Fig. 1 Screenshot of the homepage of PPGV database, showing a subset of available genome sequences, search and analysis tools



Fig. 2 Flow chart of mutation search module. (A) Select reference genome. (B) Selection of samples. (C) Display of search results. (D) Search according to the annotation results. (E) Filter the results according to mutation type or length

At present, ten peach tree genomes and thousands of resequencing data have been published, however, these data are not yet effectively integrated for deep utilization.

Although some published *Rosaceae* databases contain some peach tree data, these databases are still not comprehensive enough. Such as the famous GDR (https:// www.rosaceae.org/) database, which only contains a small amount of peach tree population variation information. The newly released ROFT database contains the transcription and expression information of four Rosaceae plants, including peach trees [31]. In this study, we collected 10 sets of peach genome and 1, 378 samples of resequencing data, and through multiple analysis strategies, tens of millions of variation information including SNP/INDEL/SV were integrated into the PPGV database. At the same time, users can not only easily obtain tandem repeat genes, collinear genes and transposon sequences on the genome, but also query SSR loci and transcription factor sequences through the interactive interface. At the same time, PPGV is equipped with a variety of bioinformatics tools, such as Blast, primer design, and enrichment analysis, which enhance the practicality of the database and will contribute to the genetic breeding research of peach trees.

However, it needs to be acknowledged that some of the genomic protein data and GFF files in this database still need to be supplemented. At present, the genome quality of peach trees still needs to be improved, and a large number of resequencing data are still dominated by next-generation sequencing. In the future, we will continue to update this site with the release of peach tree T2T genome and long-read population sequencing data



Fig. 3 Functional schematic of Genome module. (A) Select the submodule. (B) Select the genome. (C) Example of transposon search results. (D) Example plot of tandem repeat and genomic collinearity results (E) Examples of gene function search results. (F) Transcription factor submodule

(this database will integrate the newly released genome and resequencing data every 12 months until a more complete database is released). In conclusion, PPGV will promote the study of peach tree genome and population genome, and contribute to peach tree breeding and fruit improvement.

Chinese	eclina	✓ Enter locati	on informatio	n 💿 Chr	02 V F	rom(bp)	12000 <b>To(</b>	<b>bp)</b> 12800	011a			
Chr02	2.12001-12	800										
GTAA												
GGCA	AAAGGTAA.	AAT Enter genei	b	GACA	AACATT	TGCCGGCA	TAGACCTGTGC	CGACAAGA	ATTCGTCG			
AGTG	AGTGTTACTATTTACAAGGTATAGCCGGGCGGCTATACTTAAAAATATTCTTTTTTGGGGTATAGCCGTCGGCTATACT											
TTAA	GATATTATA	TTTTACAGTATA	GCCGGTCG	GCTGTACT	TAAAAA/		ATATTTTTGGAG	TATAGCCGI	GCGGCT			
GGTT	TAAATATTA	TATTGACGGAG	TATAGCCGC	CCGGCTA	TACTATT	TAAAGGTAT	AGAATAGTAA	TAGAATATT	TCAAAGO			
Progra	am blasti	n 🗸 Databa	se(s) Ch	ninesecling	J			$\sim$	search			
	29 Le 30	Length=35631242										
	31 5	1 Score = 3967 bits (2000), Expect = 0.0 2 Identities = 2000/2000 (100%), Gaps = 0/2000 (0%)										
	33 Strand=Plus/Plus											
	35 Qu	Jery 4 Co	CACTCGAGGT	тессевесе	GTGCAGCA	GGGTTTGTTC	TATTGGTGAGGCC	GAGGATAG	63			
	36 37 St	jct 34001 C	CACTCGAGGTC	TGCCGGGCG	GTGCAGCA	GGGTTTGTTC	TATTGGTGAGGCO	GAGGATAG	34060			
	38 39 Duony 64 GACCATCES		ACCATCOSTC	TEGEAAGGATEGAAGAEGGGTTEATAGTTAGETGAAAAAGGAGATATGGGT								
	40											
	41 St 42	jct 34061 G	ACCATCOGTCO	CAAGGATCG	AAGACGGG	STTCATAGTT4	AGGTGAAAAAGGAG	SATATGGGT	54120			
	43 Qu 44	ery 124 Te	SATTGTCAATC	GAAATTCCT	ACAGACGG	GACCAATTTO	TTCGCGCTAAAAG	TAAATAGA	183			
	45 St	jct 34121 T	SATTGTCAATC	GAAATTCCT	ACAGACGG	GCACCAATTTO	TTCGCGCTAAAAA	TAAATAGA	34180			
						D						
Geneid Pf	fe08g2906		Convert.		Download	General Primer Picking (	Condition					
						Primer GC%	Min. 40	<b>Opt.</b> 50	Max. 60			
Gen	nome to Genome	Gene ID1		Gene ID2		Primer Size	Min. 18	<b>Opt.</b> 20	Max. 22			
	Dineseding_Pie	evm.model.contig	126.13	Pfe08g2908		Primer Tm	Min. 40	<b>Opt.</b> 50	Max. 60			
	Pfe_124Pan	Pfe08g290	3	P124PAN00313		PRIMER NUM RETURN:	3	PRIMER PAIR MAX DIFF TH	5			
	Pie Chineseding	Pfe08g290	3	evm.model.contig12	6.12							
	Ple CN14	Pfe06g290	3	Pp02G004290.4		Product size length Fron	From 100	to 400	Query			
	Pre-Lovel	Pfe06g290	9	Prupe.2G034000.	.6	1 PRIMER PIC	KING RESULTS FOR Scaffold121	29112283				
Pic.pla Pic.pla		Pfe08a290	Df#08x 2008		Pria07n27d7		<ol> <li>No misprining library specified</li> <li>Using 0-based sequence positions</li> </ol>					
		01-00-200		evm.TU.ptg000344L5		5 0.150 start len tn gc7 any,th 3'th hairpin ang 6 LEFT PRIMER 165 20 59.68 50.60 0.00 0.00 ACGANTCAAAAGA 7 REGORT PRIMER 378 20 59.57 55.60 0.00 0.00 0.00 GGCCAAACTITIT 8 SEQUENCE SIZE: 639						
		Fleoog250										
	Pla.Pmi	Pte08g290		Pm102g0332		9 INCLUDED R	EGION SIZE: 639					
Pila Pile evm.TU.ptg000344L5		evm.TU.ptg000	4415	Pfe06g2908		11 PRODUCT SI	ZE: 214, PAIR ANY_TH COMPL:	2.03, PAIR 3'_TH COMPL:	0.00			
	No. No.	Pka03g000	1	Pfe08g2908	÷	13 0 ATCC 14	GGAGATTGAAAATAGTTACCCTT1666A	ICTAGCAGTAATGATTACATTTCA	IGACAA			
	100-10			共10 叙	记录 < 1 >	15 16 60 GATC	TCTCTGGTATGGACAATATTCTATCTGT	GTOGGATTGTGAATTAGGTTCTT	ACAGT			
put sequenci	e 1(Fosta) J	Upload			Input seque	nce 2(Fasta)	Upload		Que			
put sequence	e 1(Fasta) L	Upload	ATGATTACATTICA	SACAAGATC	Input seque	nce 2(Fasta) 1	Upload		Que			
put sequence > lest1 ATCCGGAGA ICTCTGGTAT	<ul> <li>P 1(Fasta)</li> <li>L</li> <li>TIGAAAATAGITAI</li> <li>GGACAATATICTA</li> <li>TITTIAGATAATAG</li> </ul>	Upload ICCTTTGGGACTAGCAGT/ ICTGTGGCGATTGTGAAT ITGTGGCCAAAACGAATG	ATGATTACATTTCCA FAGGTTCTTACAGT	SACAAGATC CATTGTATTA	Input seque >lest2 ATCCGGA0 TCTCTGGT	nce 2(Fasta) 1 GATTGAAAATAGTI TATGGACAATATTCI ATTTTAGGACTTAATT	Upload	GTAATGATTACATTT TTAGGTTCTTACAGT	Que CAGACAAGATC ACATTGTATTAA			
put sequence > test1 ATCCGGAGA ATCCGGGTAT AGGAGCCTA CGAGGTGAAA	e 1(Festa) L VITGAAAATAGTTAI VITTGAAAATAGTTAI VITTTAGATTACTATA SGCAGATTICCA	Upload CCTTTIGGGACTAGCAGT CTGTGTGGGGATTGTGAAT TTGTGGCCAAAACGAATC TATCAGCCAAAGTCTTTA CAAGAATCTGACTCT	VATGATTACATTTCA TAGGTTCTTACAGT AAAAGAGGGCGGG ITGCATCTTTTAACT TATCCAACCACAAC	GACAAGATC ACATTGTATTA VCAAATGTAT TTTGGAAAAG IGTTAGATGG	Input seque > lest2 ATCCGGA0 TCTCTGGT GGAGCCT AGAGGGC GTGAAAG	nce 2(Fasta) L GATTGGAAAATAGTT TATGGACAATATTCT ATTTTAGATTAATTAC ATTACCGATTGAATTAA ATTACCGATTGATCC	Upload ACCCTTTGGGACTAGCA TATCTGTGGGATTGTGAA GTGGCCAAAACGAATCA TCAGCCAAAGTCTTTATT AAGAATTCTGCATCTT	GTAATGATTACATTT TTAGGTTCTTACAGT AAAGAGGCGGAAC GCATCTTTTACAACCACAAGG	CAGACAAGATC ACATTGTATTAA AAATGTATCCAA TGGAAAGAGG TTAGATGGCTG			
put sequence >test1 AGGAGGCTA CCCTGGTAT AGGAGCCTA CCCAAGAGGCTGAAA GGGTGAAAAGA CTGATGAGAG	е 1(Fosta) Ц иттералагастика иттералагастика сословатистах итералагаса одотассостора и истора истора истора истора истора истора истора истора истора истора истора истора истора истора истора истора и истора и истора и истора и и истора и и и и и и и и и и и и и	Upload CCCTTTGGGACTAGCAGTA ICTGTGTGGCAAAGGAAT TGTGGSCCAAAGGAAT CCAAGAATGCTTGCAACT TCAACCACCGGCCASCTG TCGAACACTGGCATTCTGAA	VATGATTACATTICA TAGGTTCTTACAGT AAAAGAGGGGGGG TITGCATCTTTAACT TATCCAACCAAGAGT GGACAAAAAGCTT VAGGCTTTTGTTC	GACAAGATC ACATTGTATTA VCAAATGTAT TTTGGAAAG IGTTAGATGG SGCCCCACC JAGAGTGACC	Input sequent >test2 ATCCGGAA TCTCTGGT AGAGGGC GTGAAAG GTGAAAG GTCTGTGA	nce 2(Fasta) L GATTGAAAATAGTT TATGGACAATATTCT ATTTTAGATTAATTT AGATTGCAATTAA ATTACCGTTGATCG SGAAGCTTCCTGTG AAGGAATTGAGATA	Upload	GTAATGATTACATTT TTAGGTTCTTACAGT GCATCTTTAACATT TATCCAACCACAACG GGACAAAAACGCTTG GGACAAAAAAGCTTG GGACCAAAAAAGCTTG	CAGACAAGATC ACATTGTATTAA AAATGTATCCAA TGGAAAGAGG TTTAGATGGCTG GCCCCACCCTT AGAGTGACCCC			
put sequence Nest1 ATCCGGAGA CCCTGGATA AGGGTGAAA TGATGAGA TTGATGAGA CCATTCGCA CCATTCGCA	е 1(Fasta) Ц чтобалаластоти иттобалаластоти иттобалаластоти басластисата засабастисата засабассалатт табаластисата сластосто исто каластосто исто наслассоссалатт табаластисата засабасто исто насласто исто насласто н	Upload	IATGATTACATTTCA TAGGTTCTTACAGTA AAAAGAGGGGGG TIGCATCTTTAACT TATCCAACCACAAC GGACAAAAAGCTT NAGGTCGAALTTCTC NGCACCACTGAACC	GACAAGATC ACATTGTATTA CCAAAGTATA TTTTGGAAAG GGTTAGAAGG SGCCCCACC IAGAGTGACC CTGTGTTGG ATGCCAAGTA	Input seque >:test2 ATCCGGAX TCTCTGGT GGAGCCT AGAGGGG GTGAAAG ATGAGAAG ATGAGAAG GTCTGTGA ATCTCCAA	nce 2(Fasta)	Upload	GTAATGATTACATTT TRAGGTCTTACAGTT AAAGAGGGCGGAC GCATCTTTAACTTT TATCCAACCAACAACAA GGACTAAAACAACTTG GAGGTCGAATTTCTC GAGGCGCATGAACA	CAGACAAGATC ACATTGTATTAA AAATGTATCAA TGGAAAAAGAG GCCCCACCCTT AGAGTGACCCC CTGTGTTGGCC AGTGCAAGTAG			
put sequence lest1 TCCG6GAGA C(C)CGGAGA C(C)CGGAGA C)CGAGAGA C)CGAGAGA C)CGACCACCA C)CCACCACCACCACCACCACCACCACCACCACCACCACC	• 1(Feste) L ITTGGAAAATAGTTAI ITTGGAAAATAGTTAI ITTTGGAACAATAGTTAI GGAAGTTCATG JAAGGAACTTGCTG IAAGGAACTGACAA ITGGAAGTTCCTG IAAGTCACCAGCT JAATCTGAAGTGTCC F8 IAAAC	Upload	NTGATTACATTICA TAGGTTCTTACAGT TTGCATCCTTTAAGT TATCCAACCAGAAG GGGCGAAAAAGCTT NGGCTTTTTGTTT NGGCCCACTGAACG NGGTCGAATTTCTG NGCACACTGAACG NATAGTAGTGAGTTAGTGG	GACAAGATC ACATTGTATTA ACAAATGTAT TTTGGAAAG GGCCCCACC SIGGAGGACC CTGTGTTGG ATGCAAGTA GTTTTAGTGA 9	Input sequent > test2 ATCCGGAT GGAGGCCT AGAGGGC GTGAAAG ATGAGAT GTCTGTGA ATCTCCAA CGTTTGCT AGAAACA	nce 2(Fasta) J GAITIGAAAATAGTI ATIGAACAATATICI AGAITIGAATTAATI AGAATIGCAATTAA ATTACGATGATCC SGAAGCTICCTGTG CAGCGCCAAATTG GCAATIGGAAT GCAACTGCAAGTG SAATCTGAGTGTTC	Upload	GTAATGATTACATTT TAGGTTGTTACAGT GCATCTTTAACTT ATCCAACCACAAGC GGACAAAAAGCTG GGACAAAAAGCTTG AGGACGAAAATTCTC AGGACCACCTGAACC GGACATGCACCTGAACC	CAGACAAGATC ACATTGTATTAA AAATGTATTAA TGGAAAGAGG TITAGATGGCTG GCCCCACCTT TAGATGGCACCCC CTGTGTTGGCC CAAAGTGTGTG CAAAGTGGGTC GTTTTAGTCAG			
put sequence lest1 TICCGGAGA TICAGGAGA TICATGAGA CAAACAAC CCAATCAACA TICATCAGA CCAATCAACAACA IVMORPHILE SI TICACTATA IVMORPHILE SI	• 1(Feste) L UTIGAAAATAGTTA' UTIGAAAATAGTTA' UTITGAAAATAGTTA' GGACAATATCTA' UTITGGAACTICCIG GGACAATIGAAG GGACATICCIG GGACAATIGAAG CACAGCGCCAAATI TGCAATGGACTICCIG SAAICTGAAGTGIC SAAICTGAAGTGIC SR locus R locus	Upload	NATGATTACATTICA TAGGTTCTTACAGTI AAAASGGGGGGG TIGCATCTITAACT TAGCAACCTTTAACT NGGACAAAAGCTI NGGACAAAAGCTI NGGACCAACTIGAACG NGGACAAAAGCTI NGCACCACTIGAACG NGCATIGATTAGTGG NGCATIGATTAGTGG NGCATIGATTAGTGG NGCATIGATTAGTGG NGCATIGATTAGTGG NGCATIGATTAGTGG NGCATIGATTAGTGG NGCATIGATTAGTGG NGCATIGATTAGTGG NGCATIGATTAGTGG NGCATIGATTAGTGG NGCATIGATTAGTGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGA NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATGG NGCATIGATTACAGTIGATTAGTGG NGCATIGATTACAGTIGATTAGTGG NGCATIGATTAGTG NGCATIGT NGCATIGT NGCATIGT NGCATIGATTAGTG NGCATIGT	GACAAGATC ACATTGTATTA XCAAATGTAT TITIGGAAAG GGCCCCACC SIGTAGATGG GGCCCCACC SIGTAGATGG GGCTGTTGG AAGCCAGT GTTTAGTCA GTTTAGTCA	Input seque > lest2 ATCC5GAT GGGACCT. AGAGGGC GTGAAAG ATCTCCAA ACTGTGGA AGAAACA AACTATAG	nce 2(Fasta) J GAITGAAAATAGTI ATGGACAATATCI ATTTIAGAITAATI AGAITGATTAAT AGGAITGCATTAA AGGAATGCAATAG GCAATGGACAGCG AATCGGACGCCAAGTG AATCGGACGTC fasta2_moti	Uplead ACCCTITIGGACTAGCA ATTCTGTGGACTAGCA ATTCTGTGGACTAGCA ATTCTGTGACTATGA AGACATGTATTCTGACATCT AACACAGGATAGCATCTGTA AACACTGTATTATTTG AACACTGTATTATTTG AGACTTGATAGGCCTTT AGACTTGATAGGCCTTT		CAGACAAGATC ACATIGTATTAA AAATGTATCAA TIGGAAAGAGG TITAGATGACTG GCCCCACCCTT AGAGTGACCCC CITGTGTGGCC CAAAGCTGGTG CAAAGCTGGTC CGTUTTAGTCAG			
put sequence lesi1 TCCGGAGA CTCTGGTAT UGGAGAGA TGATGAGAGA TGATGAGAGA TGATGAGAGA TCGATGAGAGA CCATTGCTGGA TGATGAGAAA TGATGAGAGAA TGATGAGAA TGATGAGAA TGATGAGAGAA TGATGAGAA TGATGAGAGAA TGATGAGAGAA TGA		Upload CCTTTIGGGACTAGCAGTA ICTGTGGCGAAACGAAT TIGTGGCGAAACGAAT TIGTGGCGAAACGAAT TIGTGGCGAAACGAT CAAGAATGGCATTGCAACGAT CAAGAATGGACACGGCCAGGT TGAAGATGGACACGGCCATGT SACCATGGATAGGCCTTTT fasta1_repeat_numb or	VATGATTACATTICA TAGGTTCTTACAGT AAAAGAGGGCGGA TTGCATCTTTAACT TATCCAACCACAA GGACAAAAAGCTT VAGGCTATTTGTGT VGGACCAACGACACTGAACA VAGGTCGAATTAGTG IATATGCAGTTAGTG Fastal_start	GACAAGATC NCATTGTATTA VCAAAGTATT TTTGGAAAG GGCCCCACC JAGAGTAACC CITGTGTTGG AGGCCAACTA CAAGTACAAGTA GTTTTAGTCA GTTTTAGTCA fasta1_end	Input seque > lest2 ATCCGGAT GGAGCCT GGAGCCT AGAGGG GGTGTGT AGAGAG GTCTGTG AGAACA ATGAAACA AGTTTGCT AGAAACA AGTTTGCT AGAAACA AGTTTGCT AGAAACA AGTTTGCT AGAAACA AGTTTGCT AGAAACA AGTTTGCT AGAAACA AGTTTGCT AGAAACA AGTTTGCT AGAAACA AGTTTGCT AGAAACA AGTTGCT AGAACA AGTTGCT AGAACA AGTTGCT AGAACA AGTTGCT AGAACA AGTTGCT AGAACA AGTTGCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGACCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGACCT AGAACA AGACCT AGAACA AGACCT AGAACA AGACCT AGAACA AGACCT AGAACA AGACCT AGAACA AGTTCCT AGAACA AGCTCT AGAACA AGCTCT AGAACA AGTTCCT AGTTCCT AGAACA AGTTCCT AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGTTCCT AGAACA AGAA	псе 2(Festa) Ц GATTGAAAATAGTT ATTGAACAATATCT ATTTAGACAATATCT ATTACGATGACAATAA ATTACCGTTGATCC GAAATTGACAATTA GAAATTGACAATTGACAATTG GAAATTGACAATTG GAAATTGACAATTG GAATTGACAAGTG GAATCTGACGAGTG f f attCTGACGAGTG f f attCTGACGAGTG f	Uplead TACCCTITIGGGACTAGCA TATCTGTGGGACTAGCA TGGCCCAAAACGAATGA TGGCCCAAAACGAATGA TAACAACGAATGACAACTGA TAACAACGAATTACTACAACGAA TAACAACGAATTACTTATTTIG TGGCCTGAAAACCCATGA TGGCCTGAAAAGCCATGA TGGCCTGAAAACGCATGA TGGCCTGAAAAGCCATGA TGGCCTGAAAAGCCATGA TGGCCTGAAAACGCATGA TGGCCTGAAAAGCCATGAAAGCCATGA TGGCCTGAAAAGCATGAAAGCCATGA TGGCCTGAAAAGCATGAAAGCATGAAAGCATGAAAGCACTGAAGCACTGAAGCACTGAAGCACTGAAAGCACTGAAAGCACTGAAGCA	GTAATGATTACATTT TRAGGTTCTTACAGT AAAGAGGGGGAG GGATCTTTAACTTT ATCCAAGCACAGAG SGACAAAAAGCTTG AGGATCGAAATTTCT AGCACCACTGAACC GGATATGCAGTTAGTG TAGCAGTTAGTGCAGTTATTG TAGCAGTTAGTGCAGTTATTG Tagta2 start	CAGACAAGATC CAGACAAGATG CAATIGTATTAA AAATGTATCCAA TGGAAAGAGG GGCCCCACCTT TTAGATGGCTG GGCCCCACCTT TTAGATGGCTG GGCCCACCTT TAGATGGCCG CAAAGCTGGTC GGTTTTAGTCAG fasta2 end			
put sequence -test1 TCCGGAGA CTCTGGTAT SGGAGCTA CCAAGAGA CCATTCCAAGA CCACTATCA Immorphic S3 fasta1 id seq1 seq1	e 1(Fasta) L ITGAAAATAGTTAI IGGACAATATGTTAI IGGACAATATGTATA SGCASATTGCAAT ITGGAAGCTTGAGA ACAGGCCAAATT IGGAAGCTGGACT SAAGGAATTGAGA ACAGGCCAAGT SAAGGACTGAAT SAAGGACTGAAT ACATCACACGAGCA SA ICGGAGTGTGACT A T	Upload CCTTTIGGGACTAGCAGTA ICTGTGTGGCAATGTGGAAT TGTGTGGCCAAAGCGAATG TGTGTGGCCAAAGCGATG CAAGGAATGTGCAACGATT TCAACCAACGACGCCAGCGC CAAGGAATTCAATTAGTTGCAACGATT TCAACCAACGACGACCCAGGG JAAACATGTATTAGTTGCA AGCCTTGAAAGGCCTTTGC fasta1_repeat_numb or 12 10	VATGATTACATTICA TAGGTTCTTACAGT AAAASAGGGCGGA TAGCAACTITAACT TAGCAACCTTTAACT TAGCAACACCAACAA GGACAAAAAGCTT VAGGCTCTAATTGTG GGACAAAAAGCTT VAGGCCAACTAGTAGT SACAACAACTAGTAGT SACAACAACTAGTAGT TAGTAGTAGTAGT TAGTAGTAGT TAGTAGTAGT TAGTAGT	GACAAGATC XCATGTATA XCATGTATA XCAATGTAT TITIGGAAGA GBCCCAAC XGAAGTTACC CIGGTITIGG XGAAGTTACC CIGGTITIGG TITIAGTCA GTITIAGTCA GTITIAGTCA 3491 10429	Input seque less2 artccgaArt gGAACCT artcsfaArt gGAACCT artcsfaArt	Ince 2(Fests) L GATTGAAAATAGTT ATTGGACAATATCC ATTTAGGACAATATC ATTACCATGACAATAA ATTACCCATGATCCAATAA ATTACCCATGACAATTG GAAATTGACAAT GACACCAAGTG GACACCAAGTG GACACCAAGTG GACACCAAGTG GACACCAAGTG GACACCAAGTG GACACCAAGTG A T	Uplend TACCCTITIGGGACTAGCA TATCTGTGGGACTAGCA TGGCCAAAACGAATGA TGGCCAAAACGAATGA TAACAACGAATGACAACTGA TAACAACGAATTATTTG TGGCTTGACAGCCCATGG TGGCTTGACAGCCCATGG TGGCTTGACAGCCCATG TGGCTTGACAGCCCATG TGGCTTGACAGCCCATG TGGCTTGACAGCCCATG TGGCTTGACAGCCCATG TGGCTTGACAGCCCATG TGGCTTGACAGCCCATG TGGCTTGACAGCCCATG TGGCTTGACAGCCCATG TGGCTTGACAGCCCATG TGGCTTGACAGCCCATG TGGCTTGACAGCCCATG TGGCTTGACAGCCCATG TGGCTTGACAGCCCATG TGGCTTGACGCCATG TGGCTTGACGCCATG TGGCTTGACGCCATG TGGCTTGACGCCATG TGGCTTGACGCCCATG TGGCTTGACGCATG TGGCTTGACGCCATG TGGCTTGACGCATG TGGCTTGACGCATG TGGCTTGACGCATG TGGCTTGACGCATG TGGCTTGACGCATG TGGCTTGACGCATG TGGCTTGACGCATG TGGCATG TGGCTTGACGCA	GTAATGATTACATTI TTAGGTTCTTACAGT AAAGAGGGGGAG GGATCTITAACTTI YATCCAAGCACAGAG GAGATGGAAATTICTC AGGATGGAAATTICTC AGGATGGAATTAGTGG GBATGCAGTTAGTGG Batag start 11906 25247	Curcador Control Contr			
put sequence -test1 NTCCGGAGA NTCCGAGAGA CCAAAGAGC NTGTCTGGTGAAA TTGTCTGGTGAAA CCAATTGATGAAA CCAATTGATGAAA SAGAAA2 (CAACTATAG Mymorphile S SagaA	e 1(Festa) L VITGAAAATAGTTAI IGGACAATAIGTTAI IGGACAATAIGTTA IGGACAATAIGCAT SacAastTiccAat Acadagacatticaat Acadagacatticaat Acadagacatticaat Acadagacatticaat Salogaatticaat Salogaatticaat Salogaatticaat Salogaatticaat Salogaatticaat Salogaatticaat Salogaattica	Upload CCTTTGGGACTAGCAGTA ICTGTGTGGCAATGGAAT ITGTGGCCAAAGGAATG ITGTGGCCAAAGGAATG ICTGTGGCCAAGATTGGAA CAGTGATAGTGACACTGTG CAGTGATAGTGACACTGTG CAGTGATAGTGACACTGTG ICTGAAGATGGCATTGGA ICTGATAGTGACACTGTG ICTGATAGTGACACTGTG ICTGATAGTGACACTGTG ICTGATAGTGACACTGTG ICTGATAGTGACACTGTG ICTGATAGTGACACTGTG ICTGATGTGATAGTGACACTGTG ICTGATGTGATGTGACACTGTG ICTGATGTGATGTGACACTGTG ICTGATGTGATGTGACGTGTG ICTGATGTGATGTGACGTGTG ICTGATGTGATGTGACGTGTG ICTGATGTGATGTGACGTGTG ICTGATGTGATGTGACGTGTG ICTGATGTGATGTGACGTGTG ICTGATGTGATGTGACGTGTG ICTGATGTGACGTGTG ICTGATGTGACGTGTG ICTGATGTGACGTGTG ICTGATGTGACGTGTG ICTGATGTGACGTGTG ICTGATGTGGACGTGG ICTGATGTGGACGTGG ICTGATGTGACGTGTG ICTGATGTGGACGTGTG ICTGATGTGGACGTG ICTGATGTGGACGTG ICTGATGTGACGTGTG ICTGATGTGGACGTG ICTGATGTG ICTGA	VATGATTACATTICA TAGGTTCTTACAGTI AAAASAGGGCGGA TATCCAACCACAA GGACAAAAAGCTI VAGGCTTTTGTTT VAGGCTTTGTTT VAGGCACACCACGAACA SATATGCACTGAACA SATATGCACTTAGTGC SATATGCACGTTATGG 10420 29165	GACAAGATC IXCATGTATA XCAATGTAT TTTGGAAAG GGCCCAAC XTGGTTGG AGAAGTGAC CAAAGTGAC CAAAGTGAC GGTTTAGCAAGTA GGTTTAGCAAGTA festal end 3491 10429 29186	Input seque less2 ATCCGAAC GGAACCT GGAACCT GGAACCT GGAACCT AGAAACA ATCTCCAA ATCTCCAA ACTTATAG Fasta2 id seq1 seq1	Ince 2(feets) L GATTGAAAATAGTT ATTGGACAATATCC ATTTAGACAATATC ATTACCATGACCATTAA ATTACCCATGATCC GAAATCCCATGATC GAAATCCAAGTG GAAATCCAAGTG GAAATCCAAGTG GAAACCAAGTG GAAATCCAAGTG A T AG	AUPlend TACCCTITIGGGACTAGCA TATCTGTGGGACTAGCA TGGCCAAAACGAATGA TGGCCAAAACGAATGA TAACAATGATACAACGATGA TAACAATGATATTTTTTG AAACATGATTATTTTG AAACATGATTATTTTG Tasta2_repeat_num er 11 11 15	GTAATGATTACATTI TRAGGTTCTTACAGT AAAGAGGGGGGAG GGATCTITTAGTTA AGCACCACAGG GGATCATTAGTG GGACAAAAAGCTG AGGGTGGAAATTICT AGCACCACTGAACC GGATGGACAGTAGTG Fasta2 start 11906 26247 71963	CAGACAMGATC ACATGTATTAA ANATGTATTAA ANATGTATTAA ANATGTATTAA ANATGTATTAA AGATGACTAG GCCCACCCT LIGGTATGCCC CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC			
aut sequence testi TCCGGAGA TCCGGAGA TCGCGAGA GGGGGGAGA CCAACCTATAC CAACCTATAC CAACCTATAC CAACCTATAC TAGTAGAS Tastal id seq1 seq1 seq1	* 1(Fasta) L NITGGAAAAIAGTTAI IGGACAAIATGTAI SGCAGATIGCAA NITTAGAATAGATTGAAA ACAGCGCCAAITT IGGAAGCTICCTG SR locus fasta1_moti f A T AG A T AG A	Upload  CCCTTTIGGGACTAGCAGTA  CCGGTGGGACTAGCAAGTA  CCGGTGGGACAAGCGAATC  TAGTCAGCAAGCAATCGCAATC  TAGTCAGCAACGAATCGCAATCGCAACCGA  CCAGGAATTGGCACACTGGCAACCGA  CAGGAATAGGCACACGGAGCAGCGA  CAGGAATAGGCACACGGAGCAGCGA  CAGGAATAGGCACACGGAGCAGCGA  TAGAACATGGATAGGCACTTGGA  AGCCTTGAAAGGGCCATTGG  AGCCTGGAAGGGCCATGG  TAGAACATGGATAGGCACTTGGA  AGCCTGGAAGGGCCATGG  TAGAACATGGATAGGCACTTGGA  AGCCTGGAAGGCCATGG  TAGAACATGGATAGGCACTTGGA  CAGGAGCAGCGAGCGAGCGA  CGGAGCAGCGAGCG	VATGATTACATTICA TAGGTTCTTACAGT AAAASAGGGCGGA TAGCATCTTTAACT TATCCAACCACAA GGACAAAAAGCTT VAGGCTTTTGTTT SGGACAAAAAGCTT VAGGACTAACACCACGAACA SIGATAGTACACCACGAACA SIGATGACTTAGTGC SIGATGACTTAGTGC SIGATGACTTACTGC SIGATTACCACGACGACGACGACGACGACGACGACGACGACGACGA	GACAAGATC LCATTGTATA VCAATGTAT TTTGGAAAG GGCCCCACC JKGGTTTGG JKGGAGTA CCAAGGTG CCAAGGTG GTTTTAGTCA GTTTTAGTCA GTTTTAGTCA J491 10429 29186 31169	Input seque less2 rectant r	Ince 2(fasta) L GATTGAAAATAGTT ATTGGACAATATIC: ATTTAGACAATATIC: ATTAGACAATATA ATTACCGTTGATCC SGAAGTCCGTGATCC GACGCCAAATTG GAATCGAGGGTC GATCGGAGGGAGGGAGGGAGGGAGGGAGGGAGGGAGGGGAGGGAGGGAGGGG	A Upload TACCCTTTIGGGACTAGCA TATCTGTGGGACTAGCA TGGCCAAAACGAATCA TAGGAATTICTGCATCTTTA TAGGAATTICTGCATCTTTT TAGGAATGGCATTCTTGA TAGGACTGCATTCTTGA TAGGACTGCATTCTTTG TAGGACTGCATTCTTTTG TGGCCTGCATGGCCTTT TGGCCTGCATGGCCTTT TGGCCTGCATGGCCTTT TGGCCTGCATGGCCTTT TGGCCTGCATGGCCTTT TGGCCTGCATGGCCTTT TGGCCTGCATGGCCTTT TGGCCTGCATGGCCTTT TGGCCTGCATGGCCTTT TGGCCTGCATGGCCTTT TGGCCTGCATGGCCTTT TGGCCTGCATGGCCTTT TGGCCTGCTGCATGGCCTTT TGGCCTGCTGCCTGCTTT TGGCCTGCTGCCTTTT TGGCCTGCTGCCTGC	GTAATGATTACATTI TRAGGTCTTACAGT AAAGAGGGGGAG GACITTTAACTTI ATCCAACCACAGG GACTATTACTI AGCACCACCAGAAC SACATAGATAGTG GACATAGATAGTG GACATAGATAGTG GACATAGATAGTG Fasta2 start 11906 26247 71963 73967	CAGAC ANGINE CAGAC ANGINE CACATGTATIAA ANATGTATICAA ANATGTATICAA ANATGTATICAA AGATGACA CACAAGETGGT CACAAGAGETGGT CACAAGAGT CACAA			
put sequence -test1 TCCGGAGA TCCTCTGTAT SGAGCCTA SGAGCCTA SGAGCCTA TTGTCGAGA SGAGCCTA SGAGCCTA SGAGCCAACTATAG Seq1 Seq1 Seq1 Seq1 Seq1 Seq1 Seq1	LIFERENT L L LIFERENT L L LIFERENTE LIFERENT LIFER	Upload CCCTTTIGGGACTAGCAGTA CCTGTGTGGGATTGTGGAT TGTGGCCAAAACGAATC TTATGTGSCCAAAACGAATC TATCAGCAACGAATC TATCAGCAACGAATC TCAACCAACGAATCATTATTTTGCAACCATTGT CCAACGAATCATTATTTTTGCAACACTGTC TGAACATGGATAGGACACTGTC TGAACATGGATTCTAA GCCTTGAAAGGACCATGGA TGAACATGGATTCTAA GCCTTGAACGACGCATTGG 12 10 11 10 10 11 11	VATGATTACATTICA TAGGTTCTTACAGTA AAAASGGGGGGG TTGCATCTTTAACT TATCCAACGAAA AGGTCGAATTTCK GGCACAAAAAGCTT VAGGCTCTACTGAACGA MAGTCGAATTAGTG AGACAAAAAGCTT AGGACGAACAGTATTGC GGACAAAAAGCT AGGACGAACGACGACGA MAGTCGAACGACGACGACGACGACGACGACGACGACGACGACGA	GACAAGATC LCATTGTATA TTTGGAAAG STTAGATGG GGCCCACC JAGAGTGAC CAAAGTGAC CAAAGTGAC CAAAGTGAC GGTTTAGTCA GGTTTAGTCA GTTTAGTCA 3491 10429 29186 31169 33998	Input seque Ples2 ATCCGAN TCTCTGAT GGA3CCT AGAAGGC GGA3CCT AGAAACA ATCTCCAA ATCTCCAA ATCTCCAA ATCTCCAA AGTTATAC Fasta2 id seq1 seq1 seq1 seq1 seq1	Ince 2(fasta)	A Upload ACCCTITIGGGACTAGCA ATTCTGTGGGACTAGCA ATTCTGTGGACTGCA ATTCGTGGACTTGAT AGGAATTCTGCATCTTTA AGGAATTCTGCATCTTTA AGGAATTCTGCATCGATCGATCTTA AGGAATTCTGCATCGATCTTA AGGAATTCTGCATCGATCTTA AGGAATTCTGCATCGATCTTA AGGAATTCTGCATCGATCTTA AGGAATTCTGCATCGATCTTA AGGAATTCTGCATCGATCTTA AGGAATTCTGCATCGATCTTA AGGAATTCTGCATCGATCTTA AGGAATTCTGCATCGATCTTA AGGAATTCTGCATCGATCTTA AGGAATTCTGCATCTTA AGGAATTCTGCATCGATCTTA AGGAATTCTGCATCGATCTTA AGGAATTCTGCATCGATCTTA AGGAATTCTGCATCGATCTTA AGGAATTCTGCATCGATCTTA AGGAATTGCATCGATCTTA AGGAATTGCATCGATCTTA AGGAATTGCATCGATCTTA AGGAATTGCATCGATCTTA AGGAACTTGCATCGATCTTA AGGAATTGCATCGATCTTA AGGAATTGCATCGATCGATCTTA AGGAATTGCATCGATCGATCTTA AGGAATTGCATCGATCGATCTTA AGGAATTGCATCGATCGATCGATCTTA AGGAATTGCATCGATCGATCGATCTTA AGGAATTGCATCGATCGATCGATCTTA AGGAATTGCATCGATCGATCGATCGATCGATCGATCGATC	GTAATGATTACATTI TRAGGTGCTACAGT AAAGAGGGGGAG GACGTTTAACTTI ATCCAACGCAGAG GACGTGAAATTICT AGCACCACGAAC AGGGTGGAAATTICT AGCACCACGGAAC GGLUTGCAGTATICG b fasta2_start 11906 26247 71963 73967 75504	CAGACAMSNG ACATGTATTAA ACATGTATTAA ACATGTATTAA ACATGTATTAA ACATGTATTAA ACATGTATTAA ACATGTATTAA ACATGACATG			
put sequence stesti accaded accord accaded accaded accord accaded accord accaded accaded accord accaded accaded accord accaded accaded accaded accord accaded accaded accaded accord accaded accaded	ee 1(Feate) L NTIGAAAATAGTTAI IGGACAATATCTA SGCAGATATCCTA SGCAGATGCAAT TITTAGATTAATAA SGCAGATIGCAAT TGGAAGCCCACATGAT ACATCGACCAGCA SAICTGAAGTGAACT SR Iocus fasta1_moti f A T AG A T CT T	Upload CCCTTTIGGGACTAGCAGTA CCTGTGTGGCAATGGAAAT CCTGTGTGGCAAAGCGAATG CCAAGAATTGTGAACT TGTGGCCAAAGCGAATG CCAAGAATTGCACTCTT CCAACCACGGCCAGCCG SCACAGCATTCTAAL CAGTGATAGTGACACTGTG SAAACATGGATTCTAAL GCCTTGATAGGCCCTTGG SCACTGATAGGCCCTTGG SCACGAGCATTGTAAL GC SCACGAGCAGCAGCGCAGGCG SAAACATGATTATTTTG SCACGAGCAGCAGCGAGGCG SAAACATGATTATTTTG SCACGAGCAGCAGGCGTTGGAA GCCTGATAGGGCCTTGGA SAACATGATTATTTTG SCACGAGCAGCGAGGCG SAAACATGATTATTTTG SCACGAGCGAGGCG SAAACATGATTATTTTG SCACGAGGCGAGGCG SAAACATGATTGAGGCGTTGGA SAACATGAGGCGAGGGCG SAAACATGAGGCGTTGGAA GCCTGGAGGGCGTGGG SAAACATGAGGCGTTGGAGGCGTTGGAGGCGTTGGAGGCGTTGGGAGGCGGTGG SAAACATGAGTGGAGGCGTTGGAGGCGGGGGGGGGGGGGG	VATGATTACATTICA TAGGTTCTTACAGTA AAAASGGGGGGG TTGCATCTTTAACT TATCCAACCACAA GGCCAAAAAGCTT VAGGCTTTTGTT VAGGCACCACTGAAC2 NATATGCACGTTAGTG NATATGCACGTTAGTG NATATGCACGTATGCACGTATGTG NATATGCACGTATGCACGTATGTG NATATGCACGTATGTG NATATGCACGTAGTG NATATGCACGTAGTG NATATGCACGTATGTG NATATGCACGTATGGACGTAGTG NATATGCACGTATGTG NATATGCACGTAGTG NATATGCACGTAGTG NATATGCACGTAGTG NATATGCACGTATGTG NATATGCACGTATGTG NATATGCACGTATGTG NATATGCACGTAGTGCACGTAGTG NATATGCACGTAGTG NATATGCACGTAGTG NATATGCAC	GACAAGATC LCATTGTATA TTTGGAAAG GGCCCCACC JAGAATGAZ CCAAAGGTA GGCCCCACC JAGAATGACC CCAAGGTA GGCCCCACC JAGAATGACC CCAAGGTA GGCTTTAG CAAAGGTA GGTTTAGTCA JAGA GGTTTAGTCA JAGA GGTTTAGTCA JAGA JAGA JAGA JAGA JAGA JAGA JAGA JA	Input seque > less2 ATCCGAN TCTCTGAT GGASCCT AGAAGGC GGASCCT AGAAGGC GGASCAT AGTATGC AGAAGAC Fasta2_id seq1 seq1 seq1 seq1 seq1 seq1 seq1	nce 2(festa) L GATTGAAAATAGTT ATTGAACAATATCC ATTTAGACAATATCC ATTTAGACAATAA ATTACCGATGATCC GAAGCCCCAAATTG GAAATTCCAAGCTG GAACCCCAAGCTG GAACCCCAAGCTG A T A G A T A G A T T C T	A Upload ACCCTTTGGGACTAGCA ATCCTGTGGGATTGGAA GGGCCAAAACGAATCA GGGCCAAAACGAATCA ACCACCGACGCAGCTG GGAGATGGCATTCTGAA AGACAGTGATTATTTG GGCCTGAAAGCCCATGG AGACAGTGATTATTTG GGCCTGAAAGCCCATGG GGCCTGAAAGCCATGG GGCCTGAAAGCCATGG GGCCTGAAAGCCATGG GGCCCATGGCCATGG GGCCCATGGCCATGGC GGCCCATGGCCATGGCCATGG GGCCCATGGCCATGGCCATGGC GGCCCATGGCCATGGCCATGGC GGCCCATGGCCATGGCCATGGC GGCCCATGGCCATGGCCATGGC GGCCCATGGCCATGGCCATGGC GGCCCATGGCCATGGCCATGGC GGCCCATGGCCATGGCCATGGC GGCCCATGGCCATGGCCATGGC GGCCGGCGCAGGCCATGGCCATGGC GGCCGGCGCGCAGGCCATGGCCATGGC GGCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	GTAATGATTACATTI TRAGGTGCTKACAGT AAACAGGGCGACA GACIGTTAACTIT ATCCAACGACAGA GACIGGAATTACTG GACAACAAGATTAGTG GACAACAAGATTAGTG GACAACAAGATTAGTG GACAACAAGATTAGTG GACAACAAGATTAGTG GACAACAAGATTAGTG GACAACAAGATTAGTG GACAACAAGATTAGTG GACAACAAGATTAGTG GACAACAAGATTAGTG Festa2 stert 11906 26247 71963 73967 76504 82578 105025	dui AcArtantaria AcArtantaria AcArtantaria AcArtantaria AcArtantaria AcArtance Constructure AcArtance AcArt			
put sequence > testi attCGGGAGA AGGGIGAAA AGGAGCCTA CCAACGAGG AGGGIGAAA CCAACGAGA iymorphic S fastal id seq1 seq1 seq1 seq1 seq1 seq1 seq1 seq1	ee 1(Feste) L VITGGAAAATAGTTAI IGGACAATATCTA IGGACAATATCTA SGCAGATIGCAAT VITTAGAATTAATAA SGCAGATIGCAAT TGGAAGCTCGTGA ACAGCGCCAATT FI AGG A T AG AG T CT T A	Upload CCCTTTIGGGACTAGCAGT. CCTGTGTGGCAAAAACGAATC TGTGGCCAAAACGAATC TAATCASCCAAAGGCATTCTGAACCACTGC CCAGGAATTCTGAACCACTGGCCAGCCG SCACAAGCATTCTAAACACTGGAAAGGCATTCTGA CCCTTGAACAGTGGAACACTGGT SAAACATGGAATAGGCATTCGAA Gasta1_repeat_numb or I2 I0 I1 I0 I1 I0 I1 I1 I0 I4 I1	AATGATTACATTICA TAGGTTCTTACAGTA AAAASGGGGGGG TTGCATCTTTAACT TATCCAAGCAACAA GGCCAAAAAGCTT VAGGCTTTTGTT AGGACCAACAACACT AGGACGAACACTGAACC AGGACGAACACTGAACC AGGACGAACACTGAACC AGGACGAACACTGAACC AGGACGAACACTGAACC AGGACGACACTGAACC AGGACGACACTGAACC AGGACGACACTGAACC AGGACGACGACACTGAACCACTGAACCAC AGGACGACGACGACGACGACGACGACGACGACGACGACGA	GACAAGATC ACATGTATA ACATGTATA ACAATGTAT TITIGGAAAG GGCCCCACC JAGAGTAGCC CCAAGGTA CCAAAGCTG GGTTTAGTCA GGTTGAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGT	Input seque > less2 ATCCGAN TCTCTGAT GGASCCT AGAAGGC GGASCCT AGAAGGC GGASCAT AGCTATAC Fasta2.id seq1 s	nce 2(festa) L GATTGAAAATAGTT ATTGAACAATATCC ATTTAGACAATATCC ATTTAGACAATAA ATTACCGATGATCC GAAATTCCAATTAA ATTACCGATGATCC GAAATTCCAAGTG GAAATTCGAGTGTC fasta2_moti f a T A T AG A T CT A A	A Upload ACCCTTTGGGACTAGCA ATCCTGTGGGATTGGAA GGGCCAAAACGAATCA AGGAATTGGAATGGCATCTTAAT AGGAATGGCATCTGAA AGACAGTGATTATTTGG AGACAGTGATTATTTG AGACAGTGATTATTTG AGACAGTGATTATTTG AGACAGTGATTATTTG AGACAGTGATTATTTG AGACAGTGATTATTTG AGACAGTGATTATTTG AGACAGTGATTATTTG AGACAGTGATTATTG AGACAGTGATTG AGACAGTGATTG AGACAGTGATTG AGACAGTGATTG AGACAGTG AGACAGTGATTG AGACAGTGATTG AGACAGTG AGACGTG AGACGTG AGACGT AGACGTG AGACGTG AGACGTG AGACGTG AGACGTG AGAC	GTAATGATTACATTI TRAGGTGCTKACATT AAACAGGGCGACA GACITTTAACTIT ATCCAACGACAAG AGGTGGAATTACTG AGGATGCAACTAGTG AGGATGCAACTAGTG AGGATGCAACTAGTG AGGATGCAACTAGTG AGGATGCAACTAGTG AGGATGCAACTAGTG AGGATGCAACTAGTG AGGATGCAACTAGTG Festa2 stert 11906 26247 71963 73967 76504 82578 105025 115074	CAGACAMGAIC ACATGATATATA ACATGATICAA ACATGATICAA ACATGATCAA TIGGAMAGAG GCCCAACGAT AAGTGATCCAA TIGGAMAGAG GCCCAACGAT ACATGATCCAC CAAGAGATGATCCAC TIGGAGATAGAC GCCCAACGAT ACATGATCAA TIGGAGATAGAT TIGGAGAT T			
put sequence > text11 ATCCGGAGAA TCCTGTGTAT AGGAGCCTA CGGATGAAA CTGATGAGAA CTGATGAGAA CTGATGAGAA CCCATTGCTGTG GGGAAAACC CCCATTGCTGTG Seq1 Seq1 Seq1 Seq1 Seq1 Seq1 Seq1 Seq1	ee 1(Feste) L VITGAAAATAGTTAI IGGACAATATCTA' IGGACAATATCTA' AGGAATACCATGAAT NGGAAGTTGCAGTGAAT NGGAAGCTCCTG SaAGGAATTGAAA ACAGCGCAAAT FGCAATGGACT SR lecus festa 1_moti f A T AG A T CT T A A	Upload CCCTTTI3GGACTAGCAGT. CCTGTGTGGCAAACGAATC TGTGTGGCCAAAGCGAATC TAATCASCCAAAGGCATTCTGAACCACTGCCCAGCGACCACCGGCCAGCCA	AATGATTACATTICA TAGGTTCTTACAGT, AAAAGAGGGCGGA TTGCATCTTTAACT TATCCAACCACAA GGACAAAAGCTT VAGGTCTAATTAGTG SGACAAAAAGCTT VGGACTACAACACTGAAC2 SATAIGCAGTAATTGC SGACAAAAGCTT AGGTCGAATTAGTG SGACAAAAGCTT AGGTCGAATTAGTG SGACAAAAGCTT AGGTCGAATTAGTG SGACAAAAGCTT AGGTCGAATTAGTG SGACAAAAGCT SGACAACAC SGACAAAAGCT SGACAACACACACACACACACACACACACACACACACACA	SACAAGATC         SACATGATA           ACATGATA         STACAAGATA           TITIGGAAAG         STACAAGAAGA           SGCCCCACC         SAGAAGAACC           SAGAAGAACC         STACAAGAAGA           TAGAAGAAGA         STACAAGAAGA           TAGAAGAAGA         STACAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA	Input seque       >test2       ATCCGAN       TCTTTGGT       GGACCT       AGGAGCA       AGGAGCA       AGGAGCA       Fasta2 id       seq1	nce 2(festa) L GATTGAAAATATCT ATTTAGACAATATCT ATTTAGACAATATCAATTAGATAATTACCGTGATCC CAACTGCAAATTAGAATTA ATTACCGTGATCC SGAAACTCCAAGTG GAACCGCAACTAGC GAACCGCAACTAGC F Gasta2_motf F A T A G A T C T A A G A A T A G A A A A A A A A	A Upload ACCCTTTGGGACTAGCA ATCCTGTGGGATTGGAA STGGCCAAAACGAATCA ACCACCGACCAACTGCAACTCTTAAT AAGAAATGCATTCTGACACCCATGC ACCACGACGACTCTTAAT AAGAAGTGCATTCTTAATTTGG ACCACGACGACCACTGC ACCACGACGACCACTGC ACCACGACGACCACTGC ACCACGACGACTCTCTAAT ACCACGACGACTACTCTCACACCCATGC ACCACGACTACTCTCACACCCATGC ACCACGACTACTCTCACACCCATGC ACCACGACTACTCCACACCCATGC ACCACGACTACTCTCACACCCATGC ACCACGACTACTCTCACACCCATGC ACCACGACTACTCTCACACCCATGC ACCACGACTACTCTCACACCCATGC ACCACGACTACTCTCACACCCATGC ACCACGACTACTCTCACACCCATGC ACCACGACTACTCTCACACCCATGC ACCACGACTACTCTCACACCCATGC ACCACGACTACTCTCACACCCATGC ACCACGACTACTCTCACACCCATGC ACCACGACTACTCTCACACCCATGC ACCACGACTACTACTCTCACACCCATGC ACCACGACTACTCTCACACCCATGC ACCACGACTACTACTCTCACACCCATGC ACCACGACTACTACTCTCACACCCATGC ACCACGACTACTACTACTCTCACACCCATGC ACCACGACTACTACTACTCTCACACCCATGC ACCACGACTACTACTACTACTCTCACACCCATGC ACCACGACTACTACTACTACTACTCTCACACCCATGC ACCACGACTACTACTACTACTACTACTACTCCACACCCATGC ACCACGACTACTACTACTACTACTACTACTACTACTACTACTACTA	GTAATGATTACATTI TTAGGTTCTTACAGT AAACAGGGCGAC, GACTTTTAACTIT ATCCAACGCCAAGC GACATATAGTG GACAAAAAGCTTG AAGGTCGAAATTAGTG GACAACAAGTTAGTG GACAACAAGTTAGTG GACAACAAGTTAGTG GACAACAAGTTAGTG GACAACAAGTTAGTG GACAACAAGTTAGTG GACAACAAGTTAGTG GACAACAAGTTAGTG Festa2 stert 11906 26247 71963 72967 76504 82578 105025 115074 115846	011 CAGAC MAGNE ACATEGRA MAGNE ACATEGRA MAGNE TEGRA MAGNE GEOC CACCETT ACATEGRA MAGNE GEOC CACCETT ACATEGRA CACCET TAGATEGRA CACCET TAGATEGRA CALANGE TAGATEGRA CALANGE TAGATEGRA TAGAT			

Fig. 4 Tools built-in biology tool example. (A) Sequence fetch function. (B) Sequence alignment function. (C) Gene ID conversion. (D) Primer design. (E) Polymorphism SSR site identification

## A Enrichment analysis

GO KEGG			
Input gene list	CN14 V <b> upload</b>	ile Analysis	
evm.model.conti evm.model.conti evm.model.conti evm.model.conti evm.model.conti evm.model.conti evm.model.conti evm.model.conti evm.model.conti evm.model.conti evm.model.conti	ig95.7 ig95.58 ig95.21 ig95.57 ig95.65 ig95.23 ig95.45 ig98.3 ig98.1 ig99.3 ig99.3		P.value: 0.05 (value: 0.05) (value: 0.05 (value: 0.05) (value:
evm.model.cont	1999.4	С	Table 1. Details of 1,378 samples
obsolete intr transition r negative regu- eredopeptidasi negative regu- peptidasi negative regu- negative regu- negative regu- peptidasi negative regu- negative regu- nega	racellular metalion restasts ulation of cheotysis a chivity ulation of a chivity uncession a chivity a chivity		
*	Genera	Eile Name (2)	File Size
0		- File (Valie U	1/14
	<	04-1-139	1414
C	CN14/indel	04-1-152	1111
C	CN14/snp	04-1-31	MET
C	Chinesecling/snp	04-1-36	14M
	Lovell/indel	04-1-6	13M
C	Lovell/snp	04-1-76	14M
C	SSR	04-2-103	12M
C	r genome/124Pan	▼ 04-2-112	11M
C	r 重置	<b>新定</b> 04-2-139	11M

Fig. 5 Examples of enrichment analysis and other functional modules. (A) and (B) represent the enrichment analysis input page and the result display, respectively. (C) Overview page. (D) Download page

#### Abbreviations

PPGV	Peach Population Genome Variation
SNP	Single nucleotide polymorphism
INDEL	Insertion-deletion
SV	Structure Variantion
-	- II I I

TE Transposable element

#### Author contributions

S.Z. and T.J. conceived the project and was responsible for project initiation. Y.A. and Q.B. supervised and managed the project and research. The data collection and bioinformatics analyses were led by Y.A., Q.B., L.L., F.Z., and S.Y. The construction of the database and the writing of the manuscript were mainly done by Y.A. All authors read and approved the manuscript.

#### Funding

This research was jointly supported by grants from the Natural Science Foundation of Guizhou Province (ZK [2023]453); the Zunyi Technology and Big data Bureau, Moutai Institute Joint Science and Technology Research and Development Project (ZSKHHZ [2022] No.166 and ZSKHHZ [2021] No.327); the National Natural Science Foundation of China (32160441); research Foundation for Scientifc Scholars of Moutai Institute (mygccrc [2022]074, mygccrc [2022]088, mygccrc [2022]089), Engineering Research Center supported by Guizhou Provincial Education Department (No: Qian Jiao he KY zi [2020]022) and the Special Fund Project of Central Government Guiding Local Science and Technology Development ([2019]4006).

#### Data availability

All materials and related data in this study are provided in the PPGV database.

#### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### **Consent for publication**

Not applicable.

#### Conflict of interest

The authors declare that they have no conflict of interest.

# Received: 28 January 2024 / Accepted: 18 July 2024

Published online: 24 July 2024

#### References

- Cao K, Yang X, Li Y, Zhu G, Fang W, Chen C, Wang X, Wu J, Wang L. New highquality peach (Prunus persica L. Batsch) genome assembly to analyze the molecular evolutionary mechanism of volatile compounds in peach fruits. Plant J. 2021;108(1):281–95.
- Lian X, Zhang H, Jiang C, Gao F, Yan L, Zheng X, Cheng J, Wang W, Wang X, Ye X, et al. De novo chromosome-level genome of a semi-dwarf cultivar of Prunus persica identifies the aquaporin PpTIP2 as responsible for temperature-sensitive semi-dwarf trait and PpB3-1 for flower type and size. Plant Biotechnol J. 2022;20(5):886–902.
- Cao K, Wang B, Fang W, Zhu G, Chen C, Wang X, Li Y, Wu J, Tang T, Fei Z, et al. Combined nature and human selections reshaped peach fruit metabolome. Genome Biol. 2022;23(1):146.
- Zhang A, Zhou H, Jiang X, Han Y, Zhang X. The Draft Genome of a Flat Peach (Prunus persica L. cv.'124 Pan') Provides Insights into Its Good Fruit Flavor Traits. Plants (Basel) 2021, 10(3).
- Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F, et al. The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet. 2013;45(5):487–94.
- Yu Y, Guan J, Xu Y, Ren F, Zhang Z, Yan J, Fu J, Guo J, Shen Z, Zhao J, et al. Population-scale peach genome analyses unravel selection patterns and biochemical basis underlying fruit flavor. Nat Commun. 2021;12(1):3604.
- Cao K, Zheng Z, Wang L, Liu X, Zhu G, Fang W, Cheng S, Zeng P, Chen C, Wang X, et al. Comparative population genomics reveals the domestication history

of the peach, Prunus persica, and human influences on perennial fruit crops. Genome Biol. 2014;15(7):415.

- Guan J, Xu Y, Yu Y, Fu J, Ren F, Guo J, Zhao J, Jiang Q, Wei J, Xie H. Genome structure variation analyses of peach reveal population dynamics and a 1.67 mb causal inversion for fruit shape. Genome Biol. 2021;22(1):13.
- Li Y, Cao K, Zhu G, Fang W, Chen C, Wang X, Zhao P, Guo J, Ding T, Guan L, et al. Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. Genome Biol. 2019;20(1):36.
- Yu Y, Fu J, Xu Y, Zhang J, Ren F, Zhao H, Tian S, Guo W, Tu X, Zhao J, et al. Genome re-sequencing reveals the evolutionary history of peach fruit edibility. Nat Commun. 2018;9(1):5404.
- Zhao YL, Li Y, Cao K, Yao JL, Bie HL, Khan IA, Fang WC, Chen CW, Wang XW, Wu JL et al. MADS-box protein PpDAM6 regulates chilling requirement-mediated dormancy and bud break in peach. Plant Physiol 2023.
- Xie L, Liu M, Zhao L, Cao K, Wang P, Xu W, Sung WK, Li X, Li G. RiceENCODE: a comprehensive epigenomic database as a rice encyclopedia of DNA elements. Mol Plant. 2021;14(10):1604–6.
- Ma S, Wang M, Wu J, Guo W, Chen Y, Li G, Wang Y, Shi W, Xia G, Fu D, et al. WheatOmics: a platform combining multiple omics data to accelerate functional genomics studies in wheat. Mol Plant. 2021;14(12):1965–8.
- 14. Wang X-j, Wei Y-f, Liu Z, Yu T, Fu Y-h. Song X-m: TEGR: a comprehensive Ericaceae Genome Resource database1. J Integr Agric 2023.
- Gui S, Yang L, Li J, Luo J, Xu X, Yuan J, Chen L, Li W, Yang X, Wu S et al. ZEAMAP, a Comprehensive Database Adapted to the Maize Multi-Omics Era. *iScience* 2020, 23(6):101241.
- Zheng T, Li Y, Li Y, Zhang S, Ge T, Wang C, Zhang F, Faruquee M, Zhang L, Wu X, et al. A general model for germplasm-omics data sharing and mining: a case study of SoyFGB v2.0. Sci Bull (Beijing). 2022;67(17):1716–9.
- Cui X, Hu M, Yao S, Zhang Y, Tang M, Liu L, Cheng X, Tong C, Liu S. BnaOmics: a comprehensive platform combining pan-genome and multi-omics data from Brassica napus. Plant Commun 2023:100609.
- Luo H, Zhao W, Wang Y, Xia Y, Wu X, Zhang L, Tang B, Zhu J, Fang L, Du Z, et al. SorGSD: a sorghum genome SNP database. Biotechnol Biofuels. 2016;9:6.
- An Y, Zhang X, Jiang S, Zhao J, Zhang F. TeaPVs: a comprehensive genomic variation database for tea plant (Camellia sinensis). BMC Plant Biol. 2022;22(1):513.
- Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90.
- Ercolano MR, Di Donato A, Sanseverino W, Barbella M, De Natale A, Frusciante L. Complex migration history is revealed by genetic diversity of tomato samples collected in Italy during the eighteenth and nineteenth centuries. Hortic Res. 2020;7:100.
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly. 2014;6(2):80–92.
- 23. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 2014;15(6):R84.
- 24. Zhang RG, Li GY, Wang XL, Dainat J, Wang ZX, Ou S, Ma Y. TEsorter: an accurate and fast method to classify LTR-retrotransposons in plant genomes. Hortic Res 2022, 9.
- Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol 2019, 20(1).
- Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools an integrative toolkit developed for interactive analyses of big biological data. Mol Plant 2020.
- Gou X, Shi H, Yu S, Wang Z, Li C, Liu S, Ma J, Chen G, Liu T, Liu Y. SSRMMD: a Rapid and Accurate Algorithm for Mining SSR feature loci and candidate polymorphic SSRs based on assembled sequences. Front Genet 2020, 11.
- Droc G, Martin G, Guignon V, Summo M, Sempéré G, Durant E, Soriano A, Baurens F-C, Cenci A, Breton C et al. The banana genome hub: a community database for genomics in the Musaceae. Hortic Res 2022, 9.
- Su X, Yang L, Wang D, Shu Z, Yang Y, Chen S, Song C. 1 K Medicinal Plant Genome Database: an integrated database combining genomes and metabolites of medicinal plants. Hortic Res 2022, 9.
- Yu T, Ma X, Liu Z, Feng X, Wang Z, Ren J, Cao R, Zhang Y, Nie F, Song X. TVIR: a comprehensive vegetable information resource database for comparative and functional genomic studies. Hortic Res 2022, 9.

 Li M, Mount SM, Liu Z. Rosaceae Fruit Transcriptome Database (ROFT) – a useful genomic resource for comparing fruits of apple, peach, strawberry, and raspberry. Hortic Res 2023.

#### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.