HeveaDB: a genetic resource database for rubber tree genomic study

Han Cheng^{1 *}, Yanhua Xia², Jiannan Zhou¹, Chaorong Tang³, Huasun Huang¹

1, Rubber Research Institute, Chinese Academy of Tropical Agricultural Science, Xueyuan Road, Haikou, Hainan 571101, P.R. China.

2, Beijing Biomarker Technologies Co. Ltd, Fuqian Street 12, Nanfaxin Town, Shunyi District, Beijing, China 101300

3, Hainan University, Haikou, Hainan, P.R. China

***Correspondence author:** Dr. Han Cheng

E-mail: forcheng@gmail.com

Tel: 0086-898-66961226

Fax: 0086-898-66961226

Address: Rubber Research Institute, Chinese Academy of Tropical Agricultural Science, Haikou, Hainan, 571737, P.R. China

Keywords: *Hevea brasiliensis*; genome; database; transcriptome; germplasm

Total counts of words: 2197

Total counts of figure: 8

Total counts of supplementary tables: 2

Abstract

With the proceedings in the genome study in rubber tree, more and more data were produced. However, these data are still deposited in public databases as raw format, and the utilization of these NGS data is hindered for the researchers in rubber tree community. HeveaDB was therefore constructed and maintained under the organization of International Rubber Research and Development Board. The HeveaDB 1.0 (<http://hevea.catas.cn>) mainly stores 4 versions of Hevea draft genome, 99 NGS transcriptomes, 18451 annotated EST sequences, 12 curated gene families, 30200 gene annotations, 5049 wild germplasm phenotype data, and 18328 Wickham rubber clonal information. Several bioinformatic tools are integrated in the database to facilitate the utilization of the data for the users who are not skilled in bioinformatics. With the periodical update and functional improvement, the HeveaDB will serve as a potential platform for genomic studies and breeding technology in rubber tree.

Introduction

The Para rubber tree (*Hevea brasiliensis*) is an economically important tropical tree species that produces natural rubber, an essential industrial raw material. Though more than 2000 plant species can produce nature rubber, the rubber trees produce more than 98% of the nature rubber in the world [1]. Now rubber trees are widely planted in the tropical regions of South Asia and Africa, and the produced rubber is one of the major incomes of the small holding farmers in these countries. The *Hevea* genus has about 10 members, all of them are exclusively originated in tropical South America. In 1886, Henry Alexander Wickham collected over 70 000 rubber tree seeds from the central Amazon basin, and send back to the Kew garden, London. About 2300 rubber seedlings was obtained from these seeds. The bulk of seedlings were dispatched to Henarathgoda, Ceylon. Then 22 plants were transported to Singapore in 1877. Thirteen of them went to Kuala Kangsar and nine retained at Singapore Botany Garden. The modern rubber clones are almost all descended from these trees, and they therefore are denoted as Wickham germplasm.

With the proceedings in rubber breeding programs, the yield of rubber tree clones increased substantially. However, due to the long-life cycle, the rubber tree breeding programs takes about thirty years [2], which makes the rubber clone improvement rather difficult. Recent progresses in genomics may accelerate the crop breeding program through the promising genome selection technique [3], in which genome-wide markers are used to predict the breeding value of individuals in a breeding population. However, this technique requires the advances on genetics and genomics. Recent progress in next-generation sequencing technique greatly promoted the genomic study in rubber tree [4–13]. At current stage, 4 versions of draft genome in rubber trees were compiled [14–17] and more than 100 NGS transcriptomic data were produced [18–25]. However, these data are still deposited in NCBI and other public database in raw format, which hindered the NGS data utilization for the scientists in rubber tree community. Besides, huge amount of phenotypic data were

produced from rubber tree clones and wild germplasms [26,27]. These data need to be stored and mined in a comprehensive data hub. Considering the present needs in the rubber community, the International Rubber Research and Development Board (IRRDB) decided to construct an integrative genome hub in 2017 International Rubber Conference (Jakarta), which would be used to store the *Hevea* draft genomes, germplasm genome re-sequencing data, clone genome variation data, transcriptomic data, proteomic data, and phenotype data *etc.*, and to provide a platform for data archiving, distribution and utilization in rubber tree community. The HeveaDB 1.0 database (<http://hevea.catas.cn>) was therefore constructed and provided online data access.

Methods

Genome data collection and annotation

Basically, two types of rubber tree data were deposited in the HeveaDB: nucleic acid and clonal information data. So far, 4 versions of genome drafts were published in rubber tree, among which the CATAS7-33-97 genome has less scaffold number and longer N50 [28]. Totally 30 200 functional genes were annotated from 43 791 predicted genes. The HeveaDB uses the CATAS7-33-97 genome as the reference. Totally 43 791 protein coding genes were predicted from the reference genome. The expression data were obtained from the published transcriptomes. Totally there are 142 transcriptomes in the GenBank SRA database, among which 99 high quality samples were selected (> 1 Gb) for gene expression calculation. Totally 18451 EST sequences were collected, in which 15 870 were sequenced from a transformation-ready full-length cDNA library [29], 2851 were collected from NCBI [30]. The EST were annotated against Nr, KEGG, Trembl, Swissprot InterPro and GO databases. The successfully annotated EST sequences were deposited under each database classification.

Germplasm data collection

Besides the nucleotide data, the HeveaDB database also stored rubber tree germplasm data. Totally 5,049 IRRDB 1981' wild germplasm accession phenotypic data and 18,328 Wickham rubber clone's information were deposited. IRRDB 1981' wild germplasm data was collected from the National Germplasm Garden of Rubber Tree (Danzhou, Hainan) [27], while the Wickham rubber clone's information was collected from the internal circulated publications and publicly published literatures elsewhere.

Gene expression analysis

Ninety-nine transcriptomes were mapped to reference genome (Reyan7-33-97) with tophat2 [31], and the FPKM values of each gene were extracted. The 99 transcriptomes belong to 74 samples, some of which have 2 or 3 biological replicates (Supplemental Table 1). The expression value for the samples with biological replicates were calculated from the means, while the value for the samples without replicate was taken from the transcriptome directly. A gene expression matrix was then obtained and used for expression display and co-expression network construction.

Co-expression network construction

A WGCNA (weighted gene co-expression network analysis) co-expression network [32] was constructed from gene expression data of 74 samples used in expression visualization module. The Pearson's coefficient threshold value was set as >0.3 to build the network. Some of the genes have too many near nodes, which make the network too complicated to be recognized. We therefore only display the top 20 nodes in the co-expression network. The information of target node genes is also displayed below the network.

Database construction

The HeveaDB database are currently running on a Linux (CentOS release 6.6) virtual cluster platform. Eight-core processors with 32 GB memory resources were allocated for HeveaDB database. The software environments are set up as the following: Apache Tomcat 7.0.64, Java openjdk-1.8.0.161, Perl 5.10.1, Apache HTTP server 2.2.15, MySQL 5.6.38. Genome were browsed in GBrowse 2.54 and JBrowse 1.12.3. The gene expression is visualized by ECharts 3.2.2, and the co-expression network is drawn with Cytoscape.js 2.7.2. BLAST 2.2.26 and BLAT 3.5 are used to align with the genome. Other pages are generated as static web pages.

Results and Discussion

An intergrated genomic database for rubber tree study

The HeveaDB database is highlighted in four features: 1, The HeveaDB collected substantial data of the rubber trees, including 4 versions of draft genome, 99 NGS transcriptomes, 18 451annotated EST sequences, 12 gene families, 30 200 gene annotations, 5 049 wild germplasm phenotype data, and 18 328 rubber clone's information etc. (**Table 1**); 2, The HeveaDB provide user-friendly interface to facilitate scientists to utilize the data. Several tools are integrated in the database, including genome browser, blast, blat, text search, sequence retrieve and download etc.; 3, The HeveaDB mined the raw data and provide easy-to-use info to the user. The transcriptomes were analyzed and the gene expression are presented as heatmap and digital FPKM formats. A co-expression network was constructed to show possible correlation between genes; 4, The HeveaDB has a mechanism to receive and share data from rubber tree community under the framework of IRRDB.

Database structure and organization

The database has 5 major functional modules: Data, Search, Tools, Download and Submit (**Figure 1**). To enable the researcher to explore the rubber tree genome data in

a friendly interface, a drop-down menu was provided to access the functions at the upper part of the homepage. A quick search box is presented in the top-right of the page for searching the genes and germplasms.

Gene page

Detailed info of each gene is displayed in a compiled gene page. The info includes: ID, loci, structure, expression heatmap, co-expression network, sequences (genomic sequence, transcripts, CDS and deduced peptide), gene annotations (Figure 2). Besides the sequence of CDS and deduced peptide, the up- and down-stream sequences (promoter, terminator etc.) can also be retrieved by entering the numbers to be shown in the genomic sequence panel. The gene pages can be accessed through hyperlinks elsewhere in the HeveaDB database.

Genome browser

The reference genome from CATAS7-33-97 clone is visualized in the Gbrowse and Jbrowse respectively (Figure 3). Each gene is linked to its detail information page by clicking the gene names in the browser (Figure 2). The Gbrowse detail page includes the following basic information of this gene: name, position in the scaffold, length, CDS parts, and sequence. The Jbrowse detail page is showed in a popup window. The primary data and attributes of a gene are displayed, including name, position, length, and region sequences and their sub-features. The default scaffold in the genome browsers is scaffold0001, the user can shift to view particular scaffold by entering the name in the searching form.

Data search systems

Six types of searching system were provided in HeveaDB (Table 2): gene search, marker search, germplasm search, gene expression search, and sequence homology search (blast and blat). Gene search utilized ID or annotation as keywords to search

the functional annotated genes in rubber tree. The genes can be viewed in a compiled gene page, which includes the general gene information, genes structure, gene expression, co-expression network, gene sequences (including CDS and deduced peptide), and annotation (Figure 4). The gene structure was extracted from genome annotation and viewed in the Gbrowse, while the expression of this gene in 74 samples is displayed as heatmap. The rubber genetic map can also be searched by linkage group or marker name (Figure 5). Since there are limited markers in HeveaDB, the markers can also be browsed by clicking the linkage group names at the left side. In gene expression searching system, up-to 20 gene ID can be investigated once. Users can also narrow down the sample numbers through the check boxes which are arranged by clones, tissues and treatments. The expression results are displayed as heatmaps or FPKM values according to the user's needs (Figure 6). The blast and blat programs were incorporated to provide homology search against genome, CDS and protein sequences (Figure 7). Finally, the information of the Wickham and IRRDB1981 wild germplasm can also be retrieved by search the clone names or germplasm accession (Figure 8). For the Wickham clones, the breeding organism, country, parentages and obtention year are displayed, while for the wild germplasms, both the quantitative (two types of latex yields) and qualitative traits (Girth increment, pulvinus groove, petiolule length, nectary shape, leaf shape, leaf margin, leaf base, leaflets separation, latex color, sprouting stage, defoliation stage and latex yield evaluation) data are displayed (Figure 8).

Other pages

Other data and information, including data submission instructions, are also incorporated in the HeveaDB database. IRRDB information pages, which provide basic descriptions about this database and IRRDB community.

Perspectives and concluding remarks

The fast proceeding of sequencing technology greatly benefits the genome study in non-model plant species, such as *Hevea brasiliensis*. The appearance of huge amount of dataset requires a comprehensive database for the researchers in rubber tree community. The HeveaDB is a datahub to store, share, distribute and re-utilize rubber tree NGS data. Using CATAS7-33-97 genome as the reference, HeveaDB provide the genome visualization, gene annotation, co-expression network, and expression visualization. Besides, the data is also searchable through the searching module, and downloadable through the web engine. Though HeveaDB is still a preliminary genomic database system, it has achieved much attention and usage in rubber community. Until now, HeveaDB achieved 8,499 pageviews from 1491 unique viewers of 14 countries and districts in the past one year (Supplemental Table 2). To make HeveaDB easier to access and more powerful to employ, further improvement will be conducted. With the progress on rubber tree genome study, the data will be updated periodically. More functions will be developed to facilitate post-genomic study analysis online. Now the rubber community has began to implement the resequencing project of Hevea germplasm. The HeveaDB will focus on the data collection of both genomic and phenomic data, and to establish a platform for the potential new discoveries for genomic studies and breeding technology in rubber tree.

Author contributions

CH, XY and HH designed the database, CH collected the data and calculated the gene expression from transcriptomes. ZJ coordinated the cooperation between IRRDB members, CH wrote the manuscript and HH revised.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

The HeveaDB datahub project was led by the IRRDB Molecular Biology and Physiology (MBP) and was financially supported by the Central Public-interest Scientific Institution Basal Research Fund for CATAS (No. 1630022018007) and the Fund for Innovative Research Team Program of CATAS (NO. 17CXTD-28). I thank Dr. Chow Keng See from Malaysian Rubber Board, Dr. Fang Yongjun and Dr. An Zewei from CATAS for their insightful suggestions for the construction of HeveaDB database.

References

- [1] Cheng H, Chen X, Zhu J, Huang H. Overexpression of a *Hevea brasiliensis* ErbB-3 Binding protein 1 Gene Increases Drought Tolerance and Organ Size in *Arabidopsis*. *Front Plant Sci* 2016;7. doi:10.3389/fpls.2016.01703.
- [2] Huasun Huang. Fifty years of rubber breeding in China. China Agricultural Press; 2005.
- [3] Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de los Campos G, et al. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science* 2017;22:961–75. doi:10.1016/j.tplants.2017.08.011.
- [4] Cheng H, Wang Y, Sun M. Comparison of Gene Expression Profiles in Nonmodel Eukaryotic Organisms with RNA-Seq. In: Wang Y, Sun M, editors. *Transcriptome Data Analysis*, vol. 1751, New York, NY: Springer New York; 2018, p. 3–16. doi:10.1007/978-1-4939-7710-9_1.
- [5] Zou Z, Gong J, An F, Xie G, Wang J, Mo Y, et al. Genome-wide identification of rubber tree (*Hevea brasiliensis* Muell. Arg.) aquaporin genes and their response to ethephon stimulation in the laticifer, a rubber-producing tissue. *BMC Genomics* 2015;16:1001. doi:10.1186/s12864-015-2152-6.
- [6] Liu H, Deng Z, Chen J, Wang S, Hao L, Li D. Genome-wide identification and expression analysis of the metacaspase gene family in *Hevea brasiliensis*. *Plant Physiology and Biochemistry* 2016;105:90–101. doi:10.1016/j.plaphy.2016.04.011.

- [7] Xiao X-H, Yang M, Sui J-L, Qi J-Y, Fang Y-J, Hu S-N, et al. The calcium-dependent protein kinase (CDPK) and CDPK-related kinase gene families in *Hevea brasiliensis* -comparison with five other plant species in structure, evolution, and expression. *FEBS Open Bio* 2017;7:4–24. doi:10.1002/2211-5463.12163.
- [8] Guo D, Zhou Y, Li H-L, Zhu J-H, Wang Y, Chen X-T, et al. Identification and characterization of the abscisic acid (ABA) receptor gene family and its expression in response to hormones in the rubber tree. *Sci Rep* 2017;7:45157. doi:10.1038/srep45157.
- [9] Sui J-L, Xiao X-H, Qi J-Y, Fang Y-J, Tang C-R. The SWEET gene family in *Hevea brasiliensis* - its evolution and expression compared with four other plant species. *FEBS Open Bio* 2017;7:1943–59. doi:10.1002/2211-5463.12332.
- [10] Wang Y, Zhan D-F, Li H-L, Guo D, Zhu J-H, Peng S-Q. Transcriptome-Wide Identification and Characterization of MYB Transcription Factor Genes in the Laticifer Cells of *Hevea brasiliensis*. *Front Plant Sci* 2017;8. doi:10.3389/fpls.2017.01974.
- [11] Zou Z, Liu J, Yang L, Xie G. Survey of the rubber tree genome reveals a high number of cysteine protease-encoding genes homologous to *Arabidopsis* SAG12. *PLoS ONE* 2017;12:e0171725. doi:10.1371/journal.pone.0171725.
- [12] Guo D, Li H-L, Zhu J-H, Wang Y, An F, Xie G-S, et al. Genome-wide identification, characterization, and expression analysis of SnRK2 family in *Hevea brasiliensis*. *Tree Genetics & Genomes* 2017;13:86. doi:10.1007/s11295-017-1168-2.
- [13] Zou Z, Xie G, Yang L. Papain-like cysteine protease encoding genes in rubber (*Hevea brasiliensis*): comparative genomics, phylogenetic, and transcriptional profiling analysis. *Planta* 2017;246:999–1018. doi:10.1007/s00425-017-2739-z.
- [14] Tang C, Yang M, Fang Y, Luo Y, Gao S, Xiao X, et al. The rubber tree genome reveals new insights into rubber production and species adaptation. *Nature Plants* 2016;2:16073. doi:10.1038/nplants.2016.73.
- [15] Rahman AYA, Usharraj AO, Misra BB, Thottathil GP, Jayasekaran K, Feng Y, et al. Draft genome sequence of the rubber tree *Hevea brasiliensis*. *BMC Genomics* 2013;14:75. doi:10.1186/1471-2164-14-75.
- [16] Pootakham W, Sonthirod C, Naktang C, Ruang-Areerate P, Yoocha T, Sangrakru D, et al. De novo hybrid assembly of the rubber tree genome reveals evidence of paleotetraploidy in *Hevea* species. *Scientific Reports* 2017;7:41457. doi:10.1038/srep41457.
- [17] Lau N-S, Makita Y, Kawashima M, Taylor TD, Kondo S, Othman AS, et al. The rubber tree genome shows expansion of gene family associated with rubber biosynthesis. *Scientific Reports* 2016;6. doi:10.1038/srep28594.
- [18] Gébelin V, Argout X, Engchuan W, Pitollat B, Duan C, Montoro P, et al. Identification of novel microRNAs in *Hevea brasiliensis* and computational prediction of their targets. *BMC Plant Biology* 2012;12:18. doi:10.1186/1471-2229-12-18.
- [19] Li D, Deng Z, Qin B, Liu X, Men Z. De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). *BMC Genomics* 2012;13:192. doi:10.1186/1471-2164-13-192.
- [20] Duan C, Argout X, Gébelin V, Summo M, Dufayard J-F, Leclercq J, et al. Identification of the *Hevea brasiliensis* AP2/ERF superfamily by RNA sequencing. *BMC Genomics* 2013;14:30. doi:10.1186/1471-2164-14-30.

- [21] Salgado LR, Koop DM, Pinheiro DG, Rivallan R, Le Guen V, Nicolás MF, et al. De novo transcriptome analysis of *Hevea brasiliensis* tissues by RNA-seq and screening for molecular markers. *BMC Genomics* 2014;15:236. doi:10.1186/1471-2164-15-236.
- [22] Chao J, Chen Y, Wu S, Tian W-M. Comparative transcriptome analysis of latex from rubber tree clone CATAS8-79 and PR107 reveals new cues for the regulation of latex regeneration and duration of latex flow. *BMC Plant Biol* 2015;15. doi:10.1186/s12870-015-0488-3.
- [23] Liu J-P, Xia Z-Q, Tian X-Y, Li Y-J. Transcriptome sequencing and analysis of rubber tree (*Hevea brasiliensis* Muell.) to discover putative genes associated with tapping panel dryness (TPD). *BMC Genomics* 2015;16:398. doi:10.1186/s12864-015-1562-9.
- [24] Fang Y, Mei H, Zhou B, Xiao X, Yang M, Huang Y, et al. De novo Transcriptome Analysis Reveals Distinct Defense Mechanisms by Young and Mature Leaves of *Hevea brasiliensis* (Para Rubber Tree). *Scientific Reports* 2016;6:33151. doi:10.1038/srep33151.
- [25] Cheng H, Chen X, Fang J, An Z, Hu Y, Huang H. Comparative transcriptome analysis reveals an early gene expression profile that contributes to cold resistance in *Hevea brasiliensis* (the Para rubber tree). *Tree Physiol* 2018;38:1409–23. doi:10.1093/treephys/tpy014.
- [26] Adifaiz AF, Maiden NA, Shamin NA, Zarawi AG, Rafii MY. Potential Genotypes of the 1995 RRIM *Hevea* Germplasm Collection for Future Rubber Breeding and Selection Programme 1995:20.
- [27] Zhen X, Hu D. Catalogue of Rubber Germplasm Resources in China. Ed 1st. China Agricultural Press; 1994.
- [28] Tang C, Yang M, Fang Y, Luo Y, Gao S, Xiao X, et al. The rubber tree genome reveals new insights into rubber production and species adaptation. *Nature Plants* 2016;2:16073. doi:10.1038/nplants.2016.73.
- [29] Cheng H, Gao J, Cai H, Zhu J, Huang H. Gain-of-function in *Arabidopsis* (GAINA) for identifying functional genes in *Hevea brasiliensis*. *SpringerPlus* 2016;5:1853. doi:10.1186/s40064-016-3523-4.
- [30] Chow K-S, Wan K-L, Isa MNM, Bahari A, Tan S-H, Harikrishna K, et al. Insights into rubber biosynthesis from transcriptome analysis of *Hevea brasiliensis* latex. *J Exp Bot* 2007;58:2429–40. doi:10.1093/jxb/erm093.
- [31] Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013;14:R36. doi:10.1186/gb-2013-14-4-r36.
- [32] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559. doi:10.1186/1471-2105-9-559.
- [33] Xia Z, Xu H, Zhai J, Li D, Luo H, He C, et al. RNA-Seq analysis and de novo transcriptome assembly of *Hevea brasiliensis*. *Plant Mol Biol* 2011;77:299–308. doi:10.1007/s11103-011-9811-z.
- [34] Triwitayakorn K, Chatkulkawin P, Kanjanawattanawong S, Sraphet S, Yoocha T, Sangrakru D, et al. Transcriptome Sequencing of *Hevea brasiliensis* for Development of Microsatellite Markers and Construction of a Genetic Linkage Map. *DNA Res* 2011;18:471–82. doi:10.1093/dnares/dsr034.

Table 1. Data content and sources of each database in HeveaDB version 1.0.

Category	Contents	Number	Data source	References
Genome	Draft genomes	4	NCBI	[15–17,28]
Transcriptome	NGS transcriptomes	99	NCBI	[19,22–25,33,34]
Gene annotation	NR	30200	Tang et al. 2016.	[28]
	GO	43792	Tang et al. 2016.	[28]
	Pfam	43877	Tang et al. 2016.	[28]
Gene family	Curated gene family	12	This study	
EST	EST sequences	41903	This study	[29,30]
	EST annotations	21619	This study	
Germplasm	Wickham germplasms	18328	Collected from internally circulated publications	
	Wild IRRDB 1981' germplasm	5049	Zhen et al. 1994. [27]	
Gene expression dataset	Gene expression	99	From analysis of 99 transcriptomes	
Co-expression network	WCGNA co-expression network constructed basing on expression data	1	This study	

Table 2. Search engines embedded in HeveaDB database.

Search Engine	Database	Keyword type	Address
Gene Search	Gene annotation	Gene ID; Protein ID; Annotation	http://hevea.catas.cn/search/v1/toGeneSearch
Marker Search	Genetic map	Marker name:	http://hevea.catas.cn/search/v1/toMarker
Network Search	Co-expression network	Gene ID	http://hevea.catas.cn/coexpression/v1/toCoexpression
Gene Expression Search	Gene expression database	Gene ID; Sample names	http://hevea.catas.cn/tool/v1/toExpression
Sequence homology search	Genome database (blast, blat)	Query sequence	http://hevea.catas.cn/tool/v1/toBlat ; http://hevea.catas.cn/tool/v1/toBlast
Germplasm Search	Wickham germplasm; Wild germplasm	Clone accession	http://hevea.catas.cn/search/v1/toGermplasms ; http://hevea.catas.cn/search/v1/toPhenotype .

Legends

Figure 1. The structure of HeveaDB database. There are 5 major functional modules in the database: Data, Search, Tools, Download and Submit. Each module is connected with gene ID and integrated into gene info pages. The data content and functions are annotated under the block in each module.

Figure 2. The organization of the gene page. The gene page includes: ID, loci, structure, expression heatmap, co-expression network, sequences (genomic sequence, transcripts, CDS and deduced peptide), gene annotations. The user can retrieve the up- and down-stream sequences of the gene by entering the numbers to be shown in red box.

Figure 3. Genome scaffolds are displayed in genome browsers Gbrowse (A) and Jbrowse (B).

A

B

Figure 4. Gene searching engine (A) and multiple genes returned through annotation keywords search (B). Gene ID, Protein ID and annotation text can be used as keywords to search the functional annotated genes in rubber tree. When annotation text is used as keywords, multiple hits will be returned as a list.

A

Search for genes, sequences

Search genes and transcripts by species, dataset, genome location, name and/or keyword. For keyword, enter any protein name or homologs, KEGG term/CO number, GO term, or InterPro term.

1. Select a Target

Target: Please select species

2. Build your query

Search type: Gene ID Protein ID Annotation

The following is the input box, which can put a gene ID or a keyword that can be annotated (pay attention to one class at a time). You can also put more than one gene ID

Search

浏览... 未选择文件

B

Genome Annotation

The rubber tree genome annotation was based on the version of genome published by Tang et al. 2016. The gene functional annotation is mainly based on the alignment of protein sequences. The sequences of 10 genes is compared with the NCBI and EMBL databases, and the corresponding functional annotation information is obtained. You may retrieve the annotation of each individual gene by searching the gene ID, or search gene ID list by annotation keywords.

Hevea brasiliensis

Nr WRKY

Q

Please select

Gene ID	Database	Annotation_ID	Annotation
scaffold002_2_505480 (gene/v1.0) GeneDatal/scaffold002_2_505480 (annotation)	Nr	—	Probable WRKY transcription factor 11
scaffold007_4_3676915 (gene/v1.0) cGeneDatal/scaffold007_4_3676915 (annotation)	Nr	—	Probable WRKY transcription factor 4
scaffold0029_3200731 (gene/v1.0) cGeneDatal/scaffold0029_3200731 (annotation)	Nr	—	WRKY transcription factor 44
scaffold004_2577483 (gene/v1.0) cGeneDatal/scaffold004_2577483 (annotation)	Nr	—	Probable WRKY transcription factor 42
scaffold0043_3058540 (gene/v1.0) cGeneDatal/scaffold0043_3058540 (annotation)	Nr	—	Probable WRKY transcription factor 3
scaffold00046_2243620 (gene/v1.0) cGeneDatal/scaffold00046_2243620 (annotation)	Nr	—	Probable WRKY transcription factor 19
scaffold004_2007822 (gene/v1.0) cGeneDatal/scaffold004_2007822 (annotation)	Nr	—	Probable WRKY transcription factor 53
scaffold0060_2329873 (gene/v1.0) cGeneDatal/scaffold0060_2329873 (annotation)	Nr	—	Probable WRKY transcription factor 65
scaffold0053_948487 (gene/v1.0) cGeneDatal/scaffold0053_948487 (annotation)	Nr	—	Probable WRKY transcription factor 19
scaffold0059_2368229 (gene/v1.0) cGeneDatal/scaffold0059_2368229 (annotation)	Nr	—	Probable WRKY transcription factor 72

< < 1 2 3 4 5 > >

Figure 5. Molecular marker searching system (A) and a genetic map with 18 linkage groups to be displayed by clicking (B).

A

The molecular marker library contains all kinds of molecular marker information used in the mapping, and different molecular markers have different attributes.

Marker type:

Clone accession:

Chromosome/Linkage group:

Genome Location: and (bp)

Marker name: I.E. : Marker275929

B

Figure 7. The integrated Blast (A) and Blat (B) programs to search against genome, CDS and protein sequences.

A

Blast

Choose program to use and database to search:

Program Database

Enter sequence below in [FASTA](#) format

Or load it from disk 未选择文件。

Set subsequence: From To

B

Blat

biocloud Web BLAT

Database: Query Type:

Sort By: Output:

Figure 8. The germplasm searching systems. A, B, the Wickham clone search system and the information provided. C, D, the IRRDB1981 wild germplasm searching system and the information provided.

A Wickham clones

Wickham germplasm represent the major clone resources for cultivation in rubber tree. The Wickham clone database collected the information for major part of clones developed in rubber plantation countries from Asia, Africa and South America. If you are interested in individual clone, please search clone name in the following form. We encourage the user to submit information about the clones not in the database, or the updating information for the existing clones. If you are submitting clone data, please go to Submit page, or send email to ichang@rubber.cn

Clone accession: ng : 018

Country:

C Wild Germplasm

Wild germplasm represent the species richness in rubber tree. The wild clone database collected the information for the IRRDB germplasm and other wild germplasm. If you are interested in individual germplasm, please search accession name in the list encourage the user to submit information about the wild germplasm not in the database, or the updating information for the existing clones. If you are submitting clone data, please go to Submit page, or send email to: ichang@rubber.cn

Clone accession: ng : AC 100

B Select Result

Wickham clone report for CATAS/1-23-97

Clone name:	CATAS/1-23-97
Species:	NR 50884588
Breeding organism:	South China Academy of Tropical Crops, China
Country:	China
Parentage:	NR 50884588 X NR 127
Genetic background:	Wicham
Accession year:	1997
Development stage:	

D Select Result

Wild Germplasm report for IRRDB1981

Clonal ID:	IRRDB1981
Germplasm:	AC 100

Quantitative trait

Parental clone (year)	2010
Latent clone (year)	2010

Qualitative trait

Birth year:	1997
Polynomial genome:	40,000
Polynomial length:	100,000
Secondary phase:	100
Leaf shape:	100
Leaf length:	100
Leaf width:	100
Leaflet separation:	100
Leaflet color:	100
Spreading stage:	100
Germination stage:	100
Polynomial value (year)	100
Latent clone (year)	100

Download

Gene info

Organism: *Hevea brasiliensis*

Gene Name: scaffold0001_1573087.mRNA1

Location: scaffold0001:1570532..1575627 (-)

Gene structure

Gene expression

Gene network

Gene sequence

Genomic sequence [5096]

Key:CDS

Show flanking sequence :

upstream:

downstream:

Submit

>scaffold0001_1573087.mRNA1

AAAAATAGAGCCGCGAGTGTCTATTTTCCTTTTCCTACTTTTGTCCAATAAAACAGCACAAAGATGCTTTTGGTACAAATGTGAACGTAAAAAACACCAGGAACACAAACAGGCCCCGAGGCA
AATCAAGAAGATGCAAAATGCCTCAGATATCTATCCATTGCGTAAATTTCAATATAAGTTTAATAAGAAGAAATATGAATATACCATTAAACAAGATCATCCATTGGACAAGATGTGAACCAGAT
AAGGAGTCTGTATGTGAGGTACGAAGTTCAATAGAAAATCTACAGTTTCCCCTGATGGATCAACAGTTGTGATCAAACCCCAACCCACCAGAAATCACAGCTATCTGCTCTTTGATCATGCT
GCTGATAATAGCTATTAAATGCATATGATATATTTCCAGGCCAGCTGATATTGAAACAGGATCCACCAGGTGACAGTGCAGTGCATCAGCAACTGAACATGCATCCATGGCATTGGTACTT
GCATTAGACAAGTCTTTATTATTGTTTACAACACACCCTTTGAAGAAAGATATACTACATTCTTCGCATTCTCTGGTTTCTTCAAACCTTGGCCAAAATCAATACGATATTTGGCCTGACCAT

> Transcript sequence[1784]

bioRxiv preprint doi: <https://doi.org/10.1101/708305>; this version posted July 30, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

> CDS sequence [1459]

> Peptide sequence [487]

Gene Annotation

ID	Type	Description
PF00332	Pfam	Glycosyl hydrolases family 17
GO:0005975	GO	carbohydrate metabolic process
PF07983	Pfam	X8 domain
GO:0004553	GO	hydrolase activity, hydrolyzing O-glycosyl compounds
--	Nr	Glucan endo-1,3-beta-glucosidase 9

A

HeveaDB database: 500 kbp from scaffold0753:71,399..571,398

Browser [Select Tracks](#) [Snapshots](#) [Community Tracks](#) [Custom Tracks](#) [Preferences](#)

Search

Landmark or Region:

scaffold0753:71,399..571,395 Search

Examples: scaffold0001:402942..446537.

Data Source

HeveaDB database

Download Track Data Configure... Go

Save Snapshot | Load Snapshot

Scroll/Zoom: Show 500 kbp Flip

Overview

Region

Details

: 571.4 kbp

100 kbp

★ **Protein-coding genes**

B

Search genes and transcripts by species, dataset, genome location, name and/or keyword. For keyword, enter any protein name of homologs, KEGG term/EC number, GO term, or InterPro term.

1. Select a Target

Target

Please select species

bioRxiv preprint doi: <https://doi.org/10.1101/708305>; this version posted July 30, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Viridiplantae

Euphorbiaceae

Japan rubber tree

Malaysia rubber trees

Thailand rubber trees

CN rubber trees

2. Build your query

Search type

Gene ID

Protein ID

Annotation

The following is the input box, which can put a gene ID or a keyword that can be annotated (pay attention to one class at a time). You can also put more than one gene ID

Search

浏览... 未选择文件。

Genome Annotation

The rubber tree genome annotation was based on the version of genome published by Tang et al. 2016. The gene functional annotation is mainly based on the alignment of protein sequences. The sequence of the gene is compared with the Nr, GO and Pfam databases, and the corresponding functional annotation information is obtained. You may retrieve the annotation of each individual gene by searching the gene ID, or search gene ID list by annotation keywords.

Hevea brasiliensis

Nr

WRKY

Q

Download

Please select

Gene ID	DataBase	Annotation_ID	Annotation
scaffold0012_505480 (/gene/v1/toGeneDetail?scaffold0012_505480&annotation)	Nr	--	Probable WRKY transcription factor 11
scaffold0014_3616915 (/gene/v1/toGeneDetail?scaffold0014_3616915&annotation)	Nr	--	Probable WRKY transcription factor 4
scaffold0029_3200731 (/gene/v1/toGeneDetail?scaffold0029_3200731&annotation)	Nr	--	WRKY transcription factor 44
scaffold0041_2577483 (/gene/v1/toGeneDetail?scaffold0041_2577483&annotation)	Nr	--	Probable WRKY transcription factor 42
scaffold0043_3058540 (/gene/v1/toGeneDetail?scaffold0043_3058540&annotation)	Nr	--	Probable WRKY transcription factor 3
scaffold0046_2243620 (/gene/v1/toGeneDetail?scaffold0046_2243620&annotation)	Nr	--	Probable WRKY transcription factor 19
scaffold0047_2007822 (/gene/v1/toGeneDetail?scaffold0047_2007822&annotation)	Nr	--	Probable WRKY transcription factor 53
scaffold0050_2329873 (/gene/v1/toGeneDetail?scaffold0050_2329873&annotation)	Nr	--	Probable WRKY transcription factor 65
scaffold0053_1948487 (/gene/v1/toGeneDetail?scaffold0053_1948487&annotation)	Nr	--	Probable WRKY transcription factor 19
scaffold0059_2368229 (/gene/v1/toGeneDetail?scaffold0059_2368229&annotation)	Nr	--	Probable WRKY transcription factor 72

A

The molecular marker library contains all kinds of molecular marker information used in the mapping, and different molecular markers have different attributes.

Marker type:

SNP

Clone accession:

Thainland

Chromosome/Linkage group:

LG1

Genome Location:

1000

and

2000

(bp)

Marker name:

I.E. : Marker275929

Reset

Search

B

Expression Visualization

Enter feature
unique
names:

scaffold0001 1573087,scaffold0001 2292053,scaffold0001 2304224,scaffold0001 2698931,scaffold0001 3060027

Example feature unique names:scaffold0001 1573087.scaffold0001 2292053.scaffold0001 2304224.scaffold0001 2698931.scaffold0001 3060027

Experiments:

BT3410

☒ Check All

leaf

☒ 1years

CATAS7-20-59

☒ Check All

latex

☒ NA☒ 7years

CATAS7-33-97

☒ Check All

bark

☒ 10years☒ earlyCK☒ lateCK☒ earlyCOR☒ lateCOR

flower

☒ female☒ male

Display Expression Heatmap

Clear

Expression Result

Display FPKM

A

Blast

Choose program to use and database to search:

[Program](#)

blastn ▾

[Database](#)

CN_genome ▾

Enter sequence below in [FASTA](#) format

Or load it from disk

浏览...

 未选择文件。

Set subsequence: From To

Clear sequence

Search

B

Blat

bioRxiv preprint doi: <https://doi.org/10.1101/708305>; this version posted July 30, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

biocloud Web BLAT

Database:

CN_genome ▾

 Query Type:

BLAT's Guess ▾

 Sort By:

query,score ▾

 Output:

hyperlink ▾

Submit

A

Wickham clones

Wickham germplasms represent the major clone resources for cultivation in rubber tree. The Wickham clone database collercted the information for major part of clones developed in rubber plantation countries from Asia, Africa and South America. If you are interested in individual clone, please search clone name in the following form. We encourage the user to submit information about the clones not in the database, or the updating information for the existing clones. If you are submitting clone data, please go to Submit page, or sent email to: hcheng@catas.cn

Clone accession: CATAS7-33-97 eg : BD5

Country:

Reset Search

B

Select Result

Wickham clones report for CATAS7-33-97

Clone name: CATAS7-33-97

Species: H. brasiliensis

Breeding organism: South China Academy of Tropical Crops, China

Country: China

Parentage: RRIM600 X PR107

Genetic background: Wickham

Obtention year: 1966

Development stage:

C

Wild Germplasms

Wild germplasms represent the species richness in rubber tree. The wild clone database collercted the information for the IRRD germplasms and other wild germplasms. If you are interested in individual germplasm, please search accession name in the fol encourage the user to submit information about the wild germplasm not in the database, or the updating information for the exis you are submitting clone data, please go to Submit page, or sent email to: admin@heveadb.org

Clone accession: eg : AC 100

Reset Search

D

Select Result

SNP Report for XJA00001

Unify No: XJA00001

Germplasm: AC 100

Quantitative trait

Petiolule latex yield: 39.6

Lateral vein latex yield: 2

Qualitative trait

Girth inrement: Fairly slow

Pulvinus groove: Absence

Petiolule length: Medium

Nectary shape: Flat

Leaf shape: Elliptic

Leaf margin: Wavy

Leaf base: Acute

Leaflets separation: White

Latex color: White

Sprouting stage: Late

Defoliation stage: Early

Petiolule latex yield evaluation: Low

Lateral vein latex yield evaluation: Fairly low