SCIENTIFIC REPORTS

Received: 13 December 2017 Accepted: 14 February 2018 Published online: 26 February 2018

OPEN GourdBase: a genome-centered multi-omics database for the bottle gourd (Lagenaria siceraria), an economically important cucurbit crop

Ying Wang¹, Pei Xu¹, Xiaohua Wu¹, Xinyi Wu¹, Baogen Wang¹, Yunping Huang², Yaowen Hu¹, Jiandong Lin¹, Zhongfu Lu¹ & Guojing Li¹

GourdBase is an integrative data platform for the bottle gourd to examine its multifarious intuitive morphology and annotated genome. GourdBase consists of six main modules that store and interlink multi-omic data: the genome (with transcriptomic data integrated) module, the phenome module, the markers/QTLs module, the maps (genetic, physical and comparative) module, the cultivars module, and the publications module. These modules provide access to various type of data including the annotated reference genome sequence, gene models, transcriptomic data from various tissues, physical and comparative genome maps, molecular markers in different types, phenotypic data for featuring traits including fruit shape and umami taste, and guantitative trait loci (QTLs) that underlie these traits. GourdBase is intuitive, user-friendly and interlinked and is designed to allow researchers, breeders and trained farmers to browse, search and fetch information on interests and assist in genomics-driven studies and breeding. The knowledge base and web interface can be accessed at http://www.gourdbase.cn/.

The bottle gourd [Lagenaria siceraria (Mol.) Standl.] (2n = 2x = 22), native to Africa and believed to be independently domesticated in Africa and Asia¹, is an important horticultural and medicinal crop that belongs to the Cucurbitaceae family¹⁻³. The fruit is the primary economic organ of the bottle gourd, which can be used as a vegetable, a container, a decoration item or an instrument^{4,5}.

Due to the rapid progress of genomic technologies^{6,7}, obtaining a high-quality genome assembly is no longer a bottleneck for most of the plant species. We previously reported on the first dense genetic map and framework genome of the bottle gourd⁸ and have recently assembled the entire genome for the Chinese bottle gourd landrace "Hangzhou gourd" for food use. We have also generated other types of omics data including spatial transcriptomic profiles and gene co-expression networks. However, precise and high-throughput phenotyping still remains a challenge. For example, fruit shape is a composite and complex trait controlled by multiple QTLs⁹⁻¹³. Although tremendous genetic diversity in fruit shapes exists in the bottle gourd germplasm⁸, only a few genes responsible for fruit shape have been identified thus far. SUN, OVATE, WOX and YABBY are the genes that have been cloned and characterized for determining fruit shape in the fleshy crop tomato model¹⁴⁻¹⁷. An OVATE-like gene, CaOVATE, was shown to control fruit shape in pepper¹⁸. Although many fruit shape QTLs were mapped in major cucurbit crops, including in melons, cucumbers and watermelons⁹⁻¹³, little is known about the genes underlying these QTLs.

Fruit taste quality is another important agricultural trait of the bottle gourd that is difficult to dissect. Cultivars with a peculiar umami taste in fruit are more favored in the market. In many vegetables, free amino acids are recognized to be the umami ingredients. For example, tomato fruit ripening was reported to be accompanied by an increase of the concentration of amino acids. As a result, fully ripened tomato fruits usually contain higher

¹Institute of Vegetables, Zhejiang Academy of Agricultural Sciences, Hangzhou, 310021, PR, China. ²Institute of Vegetables, Ningbo Academy of Agricultural Sciences, Ningbo, 315040, China. Correspondence and requests for materials should be addressed to P.X. (email: peixu@mail.zaas.ac.cn) or G.L. (email: ligj@mail.zaas.ac.cn)



Bottle gourd [Lagenaria siceraria (Mol.) Standl.], also known as calabash, is an edible, medical, container and grafting stock plant cultivated all over the tropics. Amazingly, high genetic variability exists in bottle gourd, especially in fruit size and shape, which can be round, oblate, pyriform, elongated curvilinear, dipper, slender straight, tubby, snake-like and more.



Figure 1. The GourdBase homepage. Six main modules are displayed in this interface and include the following: the genome (with transcriptomic data integrated), phenome, markers/QTLs, maps (genetic, physical and comparative), released cultivars, and publications modules.





levels of free glutamic acid^{19,20}. Free glutamate is the main ingredient responsible for the umami taste in edible mushrooms²¹. However, which components contribute to the umami taste of bottle gourd fruits is unclear.

In addition, with the rapid expansion of bioinformatics data, how these data should be stored and interlinked and displayed in an interface requires special efforts. From a breeder's perspective, easy-to-use tools such as a marker browser, genetic maps with information on trait QTLs and designing tools for marker-assisted selection are needed. At present, there are two web servers for the cucurbits that are publically available, including the Cucurbit Genomics Database (CGD, http://cucurbitgenomics.org/), and the Cucumber Genome DataBase (http://cucumber.genomics.org.cn/page/cucumber/index.jsp). However, only CGD holds a section for the bottle gourd, in which the genome sequence of the *Lagenaria siceraria* cv. USVL1VR-Ls is retrievable. Other information that is essential for molecular breeding, such as phenotypic information and gene regulatory markers is missing. Therefore, there is an urgent need to construct a dedicated and more comprehensive bottle gourd multi-omics database to allows researchers, breeders and trained farmers to search and access various type of data.

In this study, a dedicated database named GourdBase was developed to assist in the study of biological traits and molecular breeding in the bottle gourd. The main purpose of GourdBase is to provide a reference genome A

С

Genome P	henome -	B Sea	arch By Re	aion	
Browse		000	aron by ric	gion	
Search By Gene	e ld				
Search By Region			Chromosomou	(on ohrdd
Sequence Fetch	1		chromosome:		e.g., chi 11
Multiple Sequer	nces Fetch		Start:		e.g., 15
Blastn Against (Gene		End:		e.g., 600000
Blastn Against (Genome				
Genome Brows	er			Search	
Search By Reg	gion				
Chromosome:	chr11	✓ e.g., chr11			
Start:	15	🖌 e.g., 15			
End:	600000	✓ e.g., 60000	0		

Select the columns to display:

Chr Start End Dir KEGG Ipr Inter Pro Go Planet KEGG Planet Nr ID

Gene ID	Chr	Start	End	Dir
BG_GLEAN_10002540	Chr11	41478	47365	-
BG_GLEAN_10002541	Chr11	68631	71394	+
BG_GLEAN_10002544	Chr11	86164	87124	
BG_GLEAN_10002545	Chr11	113903	116077	-
BG_GLEAN_10002546	Chr11	117745	119951	-
BG_GLEAN_10002547	Chr11	125024	133048	+
BG_GLEAN_10002548	Chr11	134649	136643	-
BG_GLEAN_10002549	Chr11	140808	149126	+
BG_GLEAN_10002550	Chr11	151560	152327	+
BG_GLEAN_10002551	Chr11	152763	153614	+

Figure 3. A detailed view of the 'Genome' function. (**A**) 'Genome' consists of 'Browse', 'Search By Gene Id', 'Search By Region', 'Sequence Fetch', 'Multiple Sequences Fetch', 'Blastn Against Gene', 'Blastn Against Genome' and 'Genome Browser'. (**B**) Demonstration of the 'Search By Region' dialog box. (**C**) An example of the search result.

resource (annotation genes, transcriptomics and proteins), a phenome (fruit shape and umami taste phenotypes), molecular markers (InDels, SSRs, SNPs), and maps (genetic, physical and comparative) to enable marker-assisted breeding. GourdBase also features hundreds of fruit shape images, secondary geometric fruit shape parameters and QTLs, which make the database valuable for studying fruit shapes.

Results

The database contents of GourdBase. GourdBase consists of six main modules: the genome (with transcriptomic data integrated) module, the phenome module, the markers/QTLs module, the maps (genetic, physical and comparative) module, the released cultivars module, and the publications module (Fig. 1). Currently, GourdBase contains the following information: 27,500 genes (25,502 protein-coding genes) with tissue expression patterns; over 6000 various types of molecular markers (SNPs, InDels, SSR, AFLP, etc.); fruit shape phenotypes for 137 diverse accessions, umami taste phenotypes that are expressed as free amino acid contents for 135 accessions²²; 56 QTLs for various traits including 17 umami taste-related QTLs, 37 fruit shape-related QTLs and 2 bitterness taste-related QTLs; 4 genetic, physical and comparative maps; 4 released cultivars with detailed information. GourdBase also harbors a publications module in which currently 18 publications supporting the aforementioned information can be freely read and downloaded with the authors' permission. All of the modules and datasets in the GourdBase combine and interrelate with one another (Fig. 2).



Figure 4. A regional view of the genome using the GourdBase genome browser. (**A**) A graphic view of the region 2759194 bp to 2800211 bp on Chr.1. (**B**) The interface after clicking on 'BG_GLEAN_10011834'.

The application of the web interface. *Genome Search.* The GourdBase genome database provides publicly available information for the recently assembled genome of the Chinese bottle gourd landrace 'Hangzhou Gourd' in the authors' lab and of the Indian accession 'cv. USVL1VR-Ls' that was initially released in the CGD. We predicted 25,502 protein-coding genes in the 'Hangzhou Gourd' genome, exceeding what was predicted in the watermelon (23,440, east Asia watermelon cultivar 97103) and cucumber (23,248, *Cucumis sativus var. sativus*) genomes^{23,24}. When one drags the mouse over the 'Genome', the column header label appears, which reads 'Browse', 'Search By Gene Id', 'Search By Region', 'Sequence Fetch', 'Multiple Sequences Fetch', 'Blastn Against Gene,' Blastn Against Genome' and 'Genome Browser' (Fig. 3A). The user can click on any one of these labels to the sublinks and search for the information needed. For example, if one clicks on 'Search By Region', then an extra layer appears that reads 'Chromosome', 'Start', and 'End' (Fig. 3B). The user can fill the query information in the box as demonstrated, and when he or she clicks on 'Search', large amounts of data appear quickly as requested (Fig. 3C). In the GourdBase genome database, the Basic Local Alignment Search Tool (BLAST) is also available to align sequences to determine similarities. 'Blast Sequence Alignment Query', 'Blast Against Gene' and 'Blast Against Genome' allow one to search the diverse genome resources.

Genome Browser. The genome browser of GourdBase includes tracks describing genes, mRNAs, peptides, and other features of interest and provides a graphical display of annotations on the bottle gourd genome (Fig. 4). Users can browse gene models on chromosomes and unanchored contigs and scaffolds. For example, if one sets up the genomic region of 2759194 bp to 2800211 bp on Chr.1 for browsing, all of the genes in this region will appear in order (Fig. 4A). When clicking on 'BG_GLEAN_10011834', an extra layer will appear with the detailed information such as mRNAs, CDS, peptides, and other features (Fig. 4B).

Transcriptomic data. The transcriptomic database of GourdBase stores and displays different RNA-seq datasets from different tissues including roots, leaves, ovaries, and the tender fruits of 'Hangzhou Gourd'. Currently, the transcriptomics data has been integrated into the GourdBase genome database, which provides a simple platform for visualizing the expression data of one particular gene and as part of the detailed information of this gene in the genome module. The RNA-Seq information is shown at the bottom of each gene information page in the form of box-illustrations.



Fruit shape phenotypes in image form

Select the columns to display:

Accession name	Collection Site	Туре	Fruit Shape	Fruit Shape Image		
J104	Hunan province(Central China)		Round			
YD-11	Guangdong province(Southern China)		Straight	NA		
Mini gourd	Liaoning province(Northeast China)	Landrace	Calabash	\$		
Songtao gourd	Guizhou province(Southwest China)	Landrace	Corbel			

Figure 5. Shape diversity of bottle gourd fruits. (**A**) Highly variable fruit shapes in the bottle gourd germplasm pool. (**B**) A snapshot of the GourdBase phenome database.

Phenotypic data. The GourdBase phenome database currently contains the following data: fruit shape phenotypes in image form, fruit shape phenotypes in digital form, and umami taste phenotypes in numeric form. Highly variable fruit shapes are found in the bottle gourd germplasm, and the majority of the fruit shape types are shown in the phenome database of GourdBase (Fig. 5A). The image data covers the fruit shapes of round, pyriform, hulu (double-gourd), slender straight, corbel, long handled round, and tubby, to name a few. The image also provides information on the name, site of origin, and types for 137 bottle gourd accessions (Fig. 5B). In addition, to acquire more precise data on fruit shapes, we digitalized the fruit images to extract the shape outline and used the Elliptic Fourier analysis to fit the outline of each shape. The simulated shape outlines were overlaid and displayed in the fruit shape phenotypes in the digital form section.

Umami taste is an important target trait for bottle gourd breeding. We have determined that free glutamate content serves as a key factor that confers umami taste in the bottle gourd by analyzing correlation between free amino acid content and the umami taste scoring²². In the GourdBase phenome database, the umami taste trait represented by the free glutamate content, as well as the content of 18 more different types of amino acids, were deposited in numeric form. The umami taste phenotypes in numeric form of the GourdBase phenome database



SSR Markers

Keyword: Search

Select the columns to display:

🗷 F primer 🚿 R primer 🗷 Product size 🐨 SSR sequences 🚿 Source contigs/singletons 🐨 Forward-Chr 🗷 Forward-Position

Marker name	F primer 0	R primer 0	Product size	SSR sequen 0	Source cont	Forward-Chr	Forward-Pos
LSR001	CGGAGGGGTTC TTTATGTTGT	GAGTGACGTGG AGTTGCTTCTA	162	CTTTTCCTTTTC CTTTTCCTTTTC CTTTTC	contig02871	Chr03	9908189- 9908209
LSR002	ACTCAATCCAAC CCACGAATAG	TGGCATTTGAAC GCTAAGTTG	189	TTTCTTTTCTTT TTCTTTTCTTTT TCTTT	contig01664	Chr08	11871785- 11871764
LSR003	TACTACCAGTGT CCGCTTTTCC	GAAGGTTGAGG AGAGTTTGCAG	128	CCCCTCCCCTC CCCTCCCCTCC CCTC	contig01795	#N/A	#N/A
LSR004	AAGCCAAGAATA CAAGTCCCCT	CCACGTTCATGT AAGAAAATGAGA G	392	TTTATTTTATTTTA TTTTATTTTATTTT	contig02748	Chr04	24971956- 24971977
LSR005	TAAGAGGTTTAG GTAGCCTGCG	CATGGCTGAGA AGTTGATTCTG	376	ATATATATATATAT ATATATATATAT	contig03386	Chr01	2661245- 2661266

Figure 6. A snapshot of the SSR marker information window in the Markers/QTLs database.

holds 135 accessions, which were assayed in 2014 and 2015. The free amino acids content was affected by the environment, which led to the differences in the umami taste phenotypes between 2014 and 2015²².

Markers/QTLs Search. The GourdBase markers/QTLs database collects over 6,000 molecular markers in various types including SNPs, Indels, SSR and AFLP markers, with key information such as source sequences, primer sequences, and product sizes that are readily retrievable (Fig. 6). This module allows users to search markers by type when dragging the mouse over the 'Marker' label. At present, all primers are displayed in tabular form. The markers/QTLs database also holds information on 17 published umami taste-related QTLs, 37 fruit shape-related QTLs and 2 bitterness taste-related QTLs. The QTL positions, left and right flanking markers, LOD value and PVE (%) are shown at present.

Genetic, physical and comparative genome maps. The GourdBase Map database contains two genetic maps, a genetic-physical correspondence map, and a genome comparative map between the bottle gourd and its relatives. One genetic map (the 'JJ' map) was constructed based on the 139-individual F_2 population derived from the cross of the 'Hangzhou gourd' and 'J129' using SNP markers from RAD-Seq data⁸, and the other genetic map was constructed based on the F₂ population derived from the 'Hangzhou gourd' and YD-4 cross using ddRAD-Seq data. The genetic-physical map correspondence information was previously published in Xu *et al.*⁸. Comparative genomic information (synteny maps) between the bottle gourd and its relatives (cucumber, watermelon and melon) have already been published⁸. Circos illustrations of the genome synteny are also available in this database.

Cultivars and publications. In addition to dedicated genomics-driven research on the bottle gourd, the ZAAS bottle gourd research program looks at the breeding of new cultivars for market use. During the past decade, a series of bottle gourd cultivars with elite fruit shapes and umami taste traits have been released, including 'Zhepu' No. 2, 'Zhepu' No. 6, 'Zhepu' No. 8 and 'Zhepu' No. 9. 'Zhepu' No. 2 bears long, straight fruit and is suitable for protected cultivation in the winter and early spring due to its strong growth potential, good fruit setting and early maturity. 'Zhepu' No. 6 bears beautiful, slender and straight fruit and is a highly resistant line to low temperatures and weak light. 'Zhepu' No. 8 bears medium-long straight fruit and is an elite, heat-resistant variety. 'Zhepu' No. 9, with its short straight fruit shape and strong umami taste, is suitable for cultivation during the summer and autumn. Detailed information and photos of the aforementioned bottle gourd cultivars are shown in the 'Cultivars' module of GourdBase.

The publications module of GourdBase currently harbors 18 publications supporting the aforementioned information can be freely read and downloaded with the authors' permission. Among them, Xu *et al.* developed a Bayesian method for calling genotypes from an F₂ population in the bottle gourd and constructed a high-density genetic map that provides a bright future of applying RAD-Seq to population genomic research for nonmodel species⁸. Xu *et al.* reported on the partial sequencing of the bottle gourd genome and revealed markers that are

useful for phylogenetic analysis and breeding²⁵. Wu *et al.* have determined free glutamate content to be a key factor conferring umami taste in the bottle gourd by genome-wide association analysis²².

Discussion

The current version of GourdBase is v1.0. As more and novel omics data, such as re-sequencing, structural variation (SV), Chip-Seq and epigenomic data from the bottle gourd are continually available, GourdBase will always evolve over time. Future development will emphasize on the update, integration and centralization of the information to provide an easier and more comprehensive access to the large amount of data.

In the short term, more transcriptomics and expression data will be added continuously, and advanced functionality will be available gradually. For markers/QTLs database, a click-and-activation function will be incorporated to allow for the display of marker information along the genome in a graphic form. Currently, none of the maps allow one to click on a feature (*e.g.*, a marker, gene, QTL) to obtain detailed information or external resources, and future development will add a greater number of feature types, more map sets, and a new search, display, and download options. In the long term, GourdBase is committed to building a complete and comprehensive information system that integrates spatial-temporal data of genomics, transcriptomics, proteomics, metabolomics, phenomics and more. With more cucurbits becoming available in the omics data, GourdBase will serve not only the bottle gourd research community but also researchers and breeders of the broader cucurbit crops.

References

- 1. Erickson, D. L. et al. An Asian origin for a 10,000-year-old domesticated plant in the Americas. Proc. Natl. Acad. Sci. USA 102, 18315–18320 (2005).
- 2. Whitaker, T. W. Endemism and pre-Columbian migration of bottle gourd, *Lagenaria siceraria* (Mol.) Standl. *Man Across The Sea*. 78–218 (1971).
- Beevy, S. S. & Kuriachan, P. Chromosome numbers of south Indian Cucurbitaceae and a note on the cytological evolution in the family. J Cytol Genet. 31, 65–71 (1996).
- 4. Lee, J. M. Cultivation of grafted vegetables I. Current status, grafting methods, and benefits. Hort Science. 29, 235–239 (1994).
- 5. Yetisir, H. & Sari, N. Effect of different rootstock on plant growth, yield and quality of watermelon. *Aust J Exp Agr.* **43**, 1269–1274 (2003).
- 6. Harris, T. D. et al. Single-molecule DNA sequencing of a viral genome. Science. 320, 106-109 (2008).
- 7. Travers, K. J. *et al.* A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res.* **38**, e159–e159 (2010).
- 8. Xu, P. *et al.* Population genomic analyses from low-coverage RAD-Seq data: a case study on the non-model cucurbit bottle gourd. *Plant J.* 77, 430–442 (2014).
- 9. Diaz, A. et al. A consensus linkage map for molecular markers and quantitative trait loci associated with economically important traits in melon (Cucumis melo L.). BMC Plant Boil. 11, 111 (2011).
- Tomason, Y. et al. Map-based molecular diversity, linkage disequilibrium and association mapping of fruit traits in melon. Mol Breeding. 31, 829–841 (2013).
- 11. Ren, Y. *et al.* An integrated genetic map based on four mapping populations and quantitative trait loci associated with economically important traits in watermelon (Citrullus lanatus). *BMC Plant Boil.* **14**, 33 (2014).
- Bo, K. et al. Molecular mapping reveals structural rearrangements and quantitative trait loci underlying traits with local adaptation in semi-wild Xishuangbanna cucumber (Cucumis sativus L. var. xishuangbannanesis Qi et Yuan). Theor Appl Genet. 128, 25–39 (2015).
- 13. Wei, Q. et al. Rapid identification of fruit length loci in cucumber (Cucumis sativus L.) using next-generation sequencing (NGS)based QTL analysis. Sci Rep. 6, 27496 (2016).
- 14. Liu, J. et al. A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. Proc. Natl. Acad. Sci. USA 99, 13302–13306 (2002).
- Xiao, H. et al. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science. 319, 1527–1530 (2008).
- Rodríguez, G. R. et al. Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity. Plant Physiol. 156, 275–285 (2011).
- van der Knaap, E. et al. What lies beyond the eye: the molecular mechanisms regulating tomato fruit weight and shape. Front. Plant Sci. 5, https://doi.org/10.3389/fpls.2014.00227 (2014).
- Tsaballa, A. *et al.* Multiple evidence for the role of an Ovate-like gene in determining fruit shape in pepper. *BMC Plant Boil.* 11, 46 (2011).
- 19. Petro-Turza, M. Flavor of tomato and tomato products. Food Rev Int. 2, 309-351 (1986).
- 20. Oruna-Concha, M. J. *et al.* Differences in glutamic acid and 5'-ribonucleotide contents between flesh and pulp of tomatoes and the relationship with umami taste. *J Agr Food Chem.* **55**, 5776–5780 (2007).
- 21. Zhang, Y. et al. Recent developments on umami ingredients of edible mushrooms-A review. Trends Food Sci Tech. 33, 78-92 (2013).
- 22. Wu, X. et al. Genome-wide association analysis of free glutamate content, a key factor conferring umami taste in the bottle gourd [Lagenaria siceraria (Mol.) Standl.]. Sci Hortic-Amsterdam. 225, 795–801 (2017).
- 23. Li, Z. et al. RNA-Seq improves annotation of protein-coding genes in the cucumber genome. BMC Genomics. 12, 540 (2011).
- 24. Guo, S. *et al.* The draft genome of watermelon (Citrullus lanatus) and resequencing of 20 diverse accessions. *Nat Genet.* **45**, 51–58 (2013).
- 25. Xu, P. et al. Partial sequencing of the bottle gourd genome reveals markers useful for phylogenetic analysis and breeding. BMC Genomics. 12, 467 (2011).

Acknowledgements

This study was partially supported by the National Natural Science Foundation of China (31401880), the Research Project for Public Interest in Zhejiang Province (2017C32027), the Major Science and Technology Project of Plant Breeding in Zhejiang Province (2016C02051) and the International Exchange Project (to X.W.) from Zhejiang Academy of Agricultural Sciences. Partial financial support was also from the National Program for Support of Top-notch Young Professionals (to P.X.). We thank Zequn Zheng (VGenomics, Shanghai) for technical assistances in website construction.

Author Contributions

P.X., G.L. and Y.W. conceived the research. Y.H. and J.L. conducted the field experiments and taken the photographs. X.W., X.W., Y. H. and Z.L. analyzed the data. B.W. created the website. Y.W. and P.X. wrote the paper.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018