

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/341414228>

# VitisGDB: The Multifunctional Database for Grapevine Breeding and Genetics

Article in *Molecular Plant* · May 2020

DOI: 10.1016/j.molp.2020.05.002

CITATIONS

0

READS

122

9 authors, including:



**Xiao Dong**

Beijing Genomics Institute

8 PUBLICATIONS 21 CITATIONS

SEE PROFILE



**Wei Chen**

Yunnan Agricultural University

106 PUBLICATIONS 429 CITATIONS

SEE PROFILE



**Zhenchang Liang**

Chinese Academy of Sciences

88 PUBLICATIONS 1,085 CITATIONS

SEE PROFILE



**Peter Nick**

Karlsruhe Institute of Technology

349 PUBLICATIONS 5,453 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Abiots stress [View project](#)

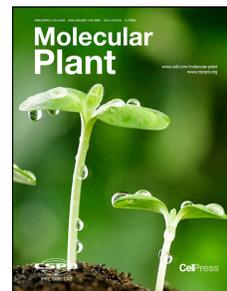


Use of CRISPR/cas9 for various studies [View project](#)

# Journal Pre-proof

VitisGDB: The Multifunctional Database for Grapevine Breeding and Genetics

Xiao Dong, Wei Chen, Zhenchang Liang, Xuzhen Li, Peter Nick, Shanshan Chen,  
Yang Dong, Shaohua Li, Jun Sheng



PII: S1674-2052(20)30137-4  
DOI: <https://doi.org/10.1016/j.molp.2020.05.002>  
Reference: MOLP 929

To appear in: *MOLECULAR PLANT*  
Accepted Date: 7 May 2020

Please cite this article as: **Dong X., Chen W., Liang Z., Li X., Nick P., Chen S., Dong Y., Li S., and Sheng J.** (2020). VitisGDB: The Multifunctional Database for Grapevine Breeding and Genetics. *Mol. Plant*. doi: <https://doi.org/10.1016/j.molp.2020.05.002>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

All studies published in *MOLECULAR PLANT* are embargoed until 3PM ET of the day they are published as corrected proofs on-line. Studies cannot be publicized as accepted manuscripts or uncorrected proofs.

© 2020 The Author

# 1 **VitisGDB: The Multifunctional Database for Grapevine Breeding and Genetics**

2

3 Xiao Dong<sup>1,10</sup>, Wei Chen<sup>1,2,3,10,11</sup>, Zhenchang Liang<sup>4,5</sup>, Xuzhen Li<sup>1,6</sup>, Peter Nick<sup>7</sup>, Shanshan  
4 Chen<sup>8</sup>, Yang Dong<sup>1,3,11</sup>, Shaohua Li<sup>4,\*</sup>, Jun Sheng<sup>6,9,\*</sup>

5

6 <sup>1</sup>State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan  
7 Agricultural University, Kunming 650201, China.

8 <sup>2</sup>College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming  
9 650201, China.

10 <sup>3</sup>Yunnan Research Institute for Local Plateau Agriculture and Industry, Kunming 650201,  
11 China.

12 <sup>4</sup>Beijing Key Laboratory of Grape Sciences and Enology, Laboratory of Plant Resources,  
13 Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China.

14 <sup>5</sup>Sino-Africa Joint Research Center, Chinese Academy of Sciences, Wuhan 430074, China.

15 <sup>6</sup>Faculty of Animal Science and Technology, Yunnan Agricultural University, Kunming  
16 650201, China.

17 <sup>7</sup>Botanical Institute, Karlsruhe Institute of Technology, Karlsruhe 76128, Germany.

18 <sup>8</sup>BGI College, Zhengzhou University, Zhengzhou 450001, China

19 <sup>9</sup>Key Laboratory of Pu'er Tea Science, Ministry of Education, Yunnan Agricultural  
20 University, Kunming 650201, China

21 <sup>10</sup>These authors have contributed equally to this work.

22 <sup>11</sup>Co-senior authors.

23

24 \*Correspondence: Shaohua Li (email: shhli@ibcas.ac.cn), Jun Sheng (email:  
25 shengjun@dongyang-lab.org).

26

27

28

29 Grapevine cultivation has been gaining commercial popularity in many parts of the world due  
30 to the high yield and versatility of the horticultural crop. A recent survey from the  
31 International Organization of Vine and Wine (OIV) estimated that the global area under vines  
32 in 2018 was about 7.4 million hectares and the world production of grapes was about 77.8  
33 million tons in total (OIV, 2019). The majority of the global grape yield is used for producing  
34 wines, fresh fruit, and raisins, bringing in multibillion revenue annually in US dollar (Alston  
35 and Sambucci, 2019). In addition to its economic value, grapevine is also a great model to  
36 study the genetic basis of clonality, fruit development, sex determination, grafting, evolution,  
37 and domestication (This et al., 2006). Furthermore, for many countries in the world, the  
38 sustenance of traditional viniculture and viticulture are important emblems of cultural identity.  
39 All these factors have determined that grapevine is one of the heavily-invested plants in  
40 horticultural research.

41  
42 The rise of the genome sequencing technologies has facilitated the release of reference-grade  
43 genetic codes and individual-level genetic variations for many grapevine species and  
44 cultivars (Canaguier et al., 2017; Girollet et al., 2019; Liang et al., 2019; Minio et al., 2019;  
45 Roach et al., 2018; Vondras et al., 2019; Zhou et al., 2017). Despite the increasing genomic  
46 data, a reliable platform for comparing and mining *Vitis* genomic information is not available.  
47 To fill this gap, we have developed VitisGDB, an online genus-level multifunctional  
48 genomics database for grapevine (Figure 1; Supplementary Note; URL:  
49 <http://vitisgdb.yнау.edu.cn/>). VitisGDB aggregates genetic information for 50 out of 60  
50 extant *Vitis* species, provides the results and visualization of a series common genetic  
51 analyses, and implements easy-to-use bioinformatic tools to enable the investigation of  
52 economically important traits for breeding new grapevine cultivars.

53  
54 The framework of VitisGDB is constructed with MySQL, ThinkPHP, and FastAdmin (Figure  
55 1; Supplementary Note) for easier data organization and user-friendly interface. Four main

56 modules, including species, germplasm, phenotype and gene (Figure 1), are created for the  
57 effective categorization and access of aggregated grapevine data. In brief, the species module  
58 provides easy retrieval of information for one European *Vitis* species (two subspecies), 19  
59 North American *Vitis* Species, 26 East Asia *Vitis* Species, 3 species from other genera  
60 (Supplementary Figure 1). The main webpage for each *Vitis* species starts with a species  
61 profile information section, which includes Latin name, chromosome number, geographical  
62 distribution, and morphological description. A representative picture (if publicly available) is  
63 also present to facilitate the taxonomic identification of the species. The second section lists  
64 the statistic of all available reference genome assemblies, by which the quality of the  
65 assemblies (contig N50, scaffold N50, and BUSCO value) can be compared. The following  
66 section details a table of sequenced germplasm with various ID information. The final section  
67 presents interactive graphs of the phylogenetic tree and the population genetic analyses. The  
68 phylogenetic tree shows a clear classification of major grapevine groups, and accession label  
69 shows detailed information of each grapevine. Both the scatter plot of principle component  
70 analysis and the bar plot of ADMIXTURE analysis can be zoomed in and out for clarity. The  
71 summary statistics of agronomic trait values in the form of box plot distributions is also  
72 presented. Finally, users have access to species-related literatures that are periodically  
73 updated.

74

75 The germplasm module includes the passport data, whole-genome sequences and published  
76 phenotypic data for 1641 *Vitis* accessions, which are reported in various re-sequencing  
77 projects. To resolve the issue that a single cultivar may have different names, the genetic  
78 background of each accession is determined with SNP data and cross-verified with the VIVC  
79 database. Consequently, accessions with the same genetic background are grouped under the  
80 same prime name, whereas 28 accessions that might be misidentified were highlighted with  
81 the inferred taxa in the phylogenetic tree.

82

83 The phenotype module indexed numeric values or categorical values for a total of 45  
84 grapevine phenotypic traits data from 1461 accessions. For each trait, the descriptor includes  
85 trait name, trait unit, OIV code, scale, and a brief summary of how this trait value was  
86 obtained. All phenotypic values are present in a table with a histogram plot showing its  
87 distribution.

88

89 Gene annotation results of three chromosome-level reference genome are integrated in the  
90 gene module. A total of 104,454 genes are curated. The webpage for each gene sequentially  
91 lists summary information (gene locus ID, gene symbol, gene type, position, and transcript  
92 number). The gene structure could be viewed in an imbedded JBrowse. The coding sequences  
93 of the gene and protein product are provided. The identified SNPs around and within the gene  
94 are also listed for the marker selection in the function verification analysis. The expression  
95 level of the gene is presented in a heatmap for easy visualization.

96

97 In addition to the main modules, VitisGDB contains a total of 25 integrated tools and external  
98 databases devoted to *Vitis* genetic research (Supplementary Figure 2). For instance, the  
99 BLAST tool is incorporated into a stand-alone webpage, where 19 genome assemblies, 7  
100 CDS sequence databases, and 7 protein sequence databases are available for query of  
101 orthologous gene candidates. The input can be either plain text or a fasta sequence file. The  
102 alignment result (available in eight styles) opens up in a new page, detailing the overall  
103 alignment score, query lengths, and similarity between the query and subject sequences. The  
104 BLAST result allows secondary filtering and the final subjects can be downloaded in HTML  
105 format.

106

107 JBrowse is an efficient visualization tool, which facilitates the viewing of gene models, CDS,  
108 heterozygous SNPs, and RNA-seq in the context of genomic region in different colors . At  
109 the moment, all available *Vitis* genomes and gene models are incorporated into JBrowse.

110

111 A JavaScript based tools SynVisio is implemented to show the synteny relationships of three  
112 pairs of chromosome-level reference genomes (PN40024 vs Chardonnay, PN40024 vs *V.*  
113 *riparia*, and Chardonnay vs *V. riparia*). The visualization includes a hive plot indicating  
114 synteny between chromosomes, a dot plot indicating collinearity between two  
115 species/cultivars, and a scatter plot indicating identified signals strength. By dragging on the  
116 little circle on the value bar from the min to max, the hive plot and the dot plot will show the  
117 corresponding results at the selected threshold (above the dotted line at the lower left corner).

118

119 Up to this point, there are three grapevine genetic maps available, covering a total of 70,832  
120 marker loci on the genome. For a selected genetic map, a heat plot shows the loci density  
121 along the chromosomes and a corresponding table describes the basic information of the  
122 mapping population and the genetic map. Double click in the heat plot zeroes in on a  
123 chromosome of interest in a new webpage. Finer details of a region on the chromosome can  
124 be sized with the pointer to show details of each locus with hyperlink to JBrowse.

125

126 In order to assist the usage and analyses of data to the user's personal preference, we build a  
127 'Download' webpage for all datasets available to the public. These include genome assembly  
128 sequences, annotation results, and genomic variations in FASTA, GFF, and VCF format,  
129 respectively. Considering the large size of raw data for *de novo* assembly, re-sequencing  
130 projects, and RNA-seq, we provide the NCBI BioProject ID, BioSample ID as well as the  
131 corresponding links on the webpage. We also import the metadata of all *Vitis* related  
132 publications from NCBI into VitisGDB for quick search.

133

134 In summary, VitisGDB has provided the most comprehensive view of the *Vitis* genomic data  
135 to date and will be a valuable platform for studies on *Vitis* functional genomics and  
136 agronomic improvement. With the goal of becoming a community-built platform dedicated to  
137 making research result on grapevine broadly available, VitisGDB accepts the submission of  
138 all types of grapevine genetic data via the ‘Submit Data’ page. VitisGDB will be  
139 continuously updated when genomic data from ongoing sequencing projects become  
140 available. New tools and analysis for transposable elements, non-coding RNAs,  
141 environmental data will be added, so that VitisGDB provides long-term support to the  
142 grapevine research community.

143

#### 144 **Acknowledgement**

145 Yunnan Provincial Key Programs of Yunnan Eco-friendly Food International Cooperation  
146 Research Center Project (2019ZG00908).

147

148

149 **Reference**

- 150 Alston, J.M., and Sambucci, O. (2019). Grapes in the World Economy. In: The Grape  
151 Genome: Springer. 1-24.
- 152 Canaguier, A., Grimplet, J., Di Gaspero, G., Scalabrin, S., Duchêne, E., Choisine, N.,  
153 Mohellibi, N., Guichard, C., Rombauts, S., and Le Clainche, I. (2017). A new version  
154 of the grapevine reference genome assembly (12X. v2) and of its annotation (VCost.  
155 v3). *Genomics data* 14:56.
- 156 Girollet, N., Rubio, B., Lopez-Roques, C., Valière, S., Ollat, N., and Bert, P.-F. (2019). De  
157 novo phased assembly of the *Vitis riparia* grape genome. *Scientific data* 6:1-8.
- 158 Liang, Z., Duan, S., Sheng, J., Zhu, S., Ni, X., Shao, J., Liu, C., Nick, P., Du, F., and Fan, P.  
159 (2019). Whole-genome resequencing of 472 *Vitis* accessions for grapevine diversity  
160 and demographic history analyses. *Nature communications* 10:1-12.
- 161 Minio, A., Massonnet, M., Figueroa-Balderas, R., Castro, A., and Cantu, D. (2019). Diploid  
162 genome assembly of the wine grape Carménère. *G3: Genes, Genomes, Genetics*  
163 9:1331-1337.
- 164 OIV. (2019). OIV Statistical Report on World Vitiviniculture.
- 165 Roach, M.J., Johnson, D.L., Bohlmann, J., van Vuuren, H.J., Jones, S.J., Pretorius, I.S.,  
166 Schmidt, S.A., and Borneman, A.R. (2018). Population sequencing reveals clonal  
167 diversity and ancestral inbreeding in the grapevine cultivar Chardonnay. *PLoS*  
168 *genetics* 14:e1007807.
- 169 This, P., Lacombe, T., and Thomas, M.R. (2006). Historical origins and genetic diversity of  
170 wine grapes. *TRENDS in Genetics* 22:511-519.
- 171 Vondras, A.M., Minio, A., Blanco-Ulate, B., Figueroa-Balderas, R., Penn, M.A., Zhou, Y.,  
172 Seymour, D., Ye, Z., Liang, D., Espinoza, L.K., et al. (2019). The genomic  
173 diversification of grapevine clones. *BMC Genomics* 20:972.

174 Zhou, Y., Massonnet, M., Sanjak, J.S., Cantu, D., and Gaut, B.S. (2017). Evolutionary  
175 genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. Proceedings of the  
176 national academy of sciences 114:11715-11720.

177

178

179

180

### 181 **Figure Legend**

182 **Figure 1 Schematic of VitisGDB.** (A) Species information. Information on 50 species in six  
183 groups were collected. Highlighted numbers in the brackets indicate number of species under  
184 each group. (B) Data sources and data type. Data sources include websites such as eFlora,  
185 VIVC, public database NCBI, and literatures. Data type includes information for species,  
186 germplasm, genome assembly, gene annotation, phenotypic trait values, genetic map,  
187 resequencing data, and RNA-seq data. (C) Summary of data processing method. (D)  
188 Framework of VitisGDB. All data are stored in MySQL. Thinkphp and FastAdmin are used  
189 for frontend and backend data management. (E) Main modules. Highlighted numbers in the  
190 brackets indicate the corresponding entries. (F) Overview of the integrated tools in VitisGDB  
191 with main functions listed below.

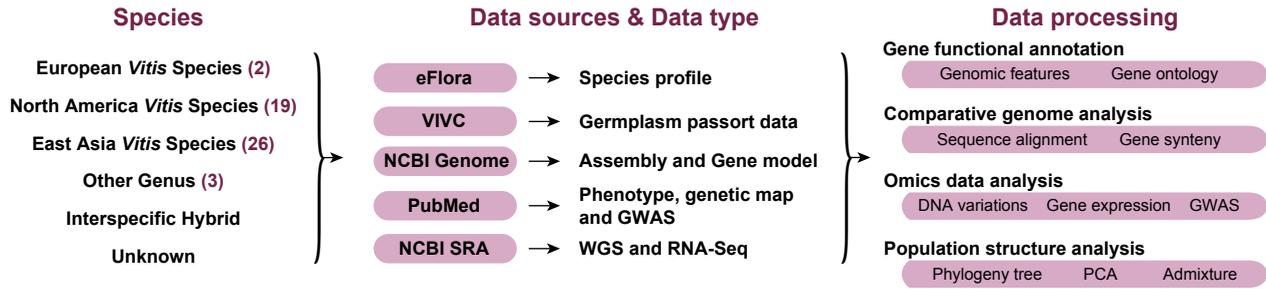
### 192 **Supplementary Figure 1: An example of Species module for *V. vinifera* ssp. *vinifera*.** (A)

193 Basic species information, which include Latin name, chromosome number, distribution area  
194 and description. Figures on two sides are preventive picture and geographic distribution of  
195 the sequenced germplasm. (B) Available genome assembly, which listed assembled cultivar  
196 and clone, as well as the assembly statistics. (C) Sequenced germplasm, which contains  
197 database ID in VitisGDB, prime name, country of origin and ID in VIVC database. (D)  
198 Interactive Phenology tree, which includes 497 accessions and their detailed information.  
199 Branches are colored by *Vitis* group. (E) Scatter plot of principle component analysis result,  
200 which indicate the species population structure with the corresponding species highlighted in  
201 yellow. (F) Bar plot of ADMIXTURE result with K from 2 to 14. Zooming into the detail

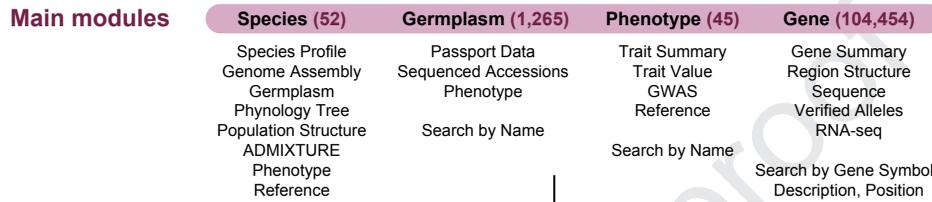
202 shows the detail information of accessions of such species. (G) Box plot/bar plot which  
203 present the distribution of quantitative/qualitative trait. (H) Related references.

204 **Supplementary Figure 2: User-friendly interfaces for seven integrated tools.** (A) BLAST;  
205 (B) JBrowse; (C) Downloads; (D) View Genetic Maps; (E) Synteny Visualization; (F) Search  
206 Publications. (G) External Links & Tools.

Journal Pre-proof



VitisGDB  
MySQL  
Thinkphp  
fastadmin



Tools	BLAST	JBrowse	Downloads	Submit Data	View Genetic Maps	Synteny Visualization	Search Publications	External links & tools
	Against <i>V. vinifera</i> (Pinot Noir, Chkhaveri, Saperavi, Meskhetian green, Rkatsiteli, Cagernet Sauvignon, Chardonnay, Sultanina), <i>V. sylvestris</i> , <i>V. davidii</i> , <i>V. aestivalis</i> , <i>V. riparia</i> , <i>V. rupestris</i> , and <i>V. cinifera</i> x <i>V. riparia</i>	Repeat tracks Gene structure Variation tracks	FASTA GFF VCF	Data submit with the help of curator (can be assembly, germplasm, phenotype, marker & map, QTLs)	Mapping population Marker loci Genetic/Physical position	PN40024 VS Chardonnay PN40024 VS <i>V. riparia</i> Chardonnay VS <i>V. riparia</i>	Search by keywords in title, abstract, author, journal name and by publication year	Six types of <i>Vitis</i> related links & tools: gene expression, Multi-omics, RNA-seq, Germplasm, Microsatellite, Genome