

VitisGDB: The Multifunctional Database for Grapevine Breeding and Genetics

Grapevine cultivation has been gaining commercial popularity in many parts of the world due to the high yield and versatility of this horticultural crop. A recent survey from the International Organization of Vine and Wine (OIV) estimated that the global area under vine cultivation in 2018 was about 7.4 million hectares and that the world production of grapes was about 77.8 million tons in total (OIV, 2019). The majority of the global grape yield is used for producing wines, fresh fruit, and raisins, bringing in annual revenue of billions of US dollars (Alston and Sambucci, 2019). In addition to its economic value, the grapevine is also a useful model for the study of the genetic basis of clonality, fruit development, sex determination, grafting, evolution, and domestication (This et al., 2006). Furthermore, for many countries in the world traditional viniculture and viticulture are important emblems of cultural identity. All these factors have made grapevine one of the most heavily invested plants in horticultural research.

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

The framework of VitisGDB was constructed with MySQL, ThinkPHP, and FastAdmin (Figure 1 and Supplemental Note) to allow for easier data organization and a user-friendly interface. Four main modules, namely species, germplasm, phenotype, and gene (Figure 1), were created for the effective categorization and access of aggregated grapevine data. In brief, the species module provides easy retrieval of information for one European *Vitis* species (two subspecies), 19 North American *Vitis* Species, 26 East Asia *Vitis* Species, and three species from other genera (Supplemental Figure 1). The main web page for each *Vitis* species starts with a species profile information section, which includes Latin name, chromosome number, geographical distribution, and morphological description. A representative picture (if publicly available) is also provided to facilitate taxonomic identification of the species. The second section lists the statistics of all available reference genome assemblies, by which the quality of the assemblies (contig N50, scaffold N50, and BUSCO value) can

be compared. The following section details a table of sequenced germplasm with extensive ID information. The final section presents interactive graphs of the phylogenetic tree and the population genetic analyses. The phylogenetic tree shows a clear classification of major grapevine groups, and the accession label shows detailed information for each grapevine. Both the scatterplot of principal component analysis and the bar plot of ADMIXTURE analysis can be zoomed in and out for clarity. The summary statistics of agronomic trait values in the form of box-plot distributions are also presented. Finally, users have access to species-related literature that is periodically updated.

The germplasm module includes the passport data, whole-genome sequences, and published phenotypic data for 1641 *Vitis* accessions, which are reported by various resequencing projects. To resolve the issue that a single cultivar may have different names, the genetic background of each accession was determined using SNP data and cross-verified with the VIVC database. Consequently, accessions with the same genetic background are grouped under the same prime name, whereas 28 accessions that might be misidentified are highlighted with the inferred taxa in the germplasm module and the phylogenetic tree section under species module.

The phenotype module indexes numeric values or categorical values for a total of 45 grapevine phenotypic trait data from 1461 accessions. For each trait, the descriptor includes trait name, trait unit, OIV code, scale, and a brief summary of how the trait value was obtained. All phenotypic values are presented in a table with a histogram plot showing their distribution.

Gene annotation results for three chromosome-level reference genomes are integrated in the gene module. A total of 104 454 genes are curated. The web page for each gene sequentially lists summary information (gene locus ID, gene symbol, gene type, position, and transcript number). The gene structure can be viewed in an embedded JBrowse. The coding sequence (CDS) of the gene and the amino acid sequence of the protein product are provided. The identified SNPs around and within the gene are also listed to facilitate marker selection for functional verification analysis. The expression level of the gene is presented in a heatmap for easy visualization.

In addition to the main modules, VitisGDB contains a total of 25 integrated tools and external databases devoted to *Vitis* genetic research (Supplemental Figure 2). For instance, the BLAST tool is incorporated into a stand-alone web page, where 19 genome assemblies, seven CDS sequence databases, and seven protein

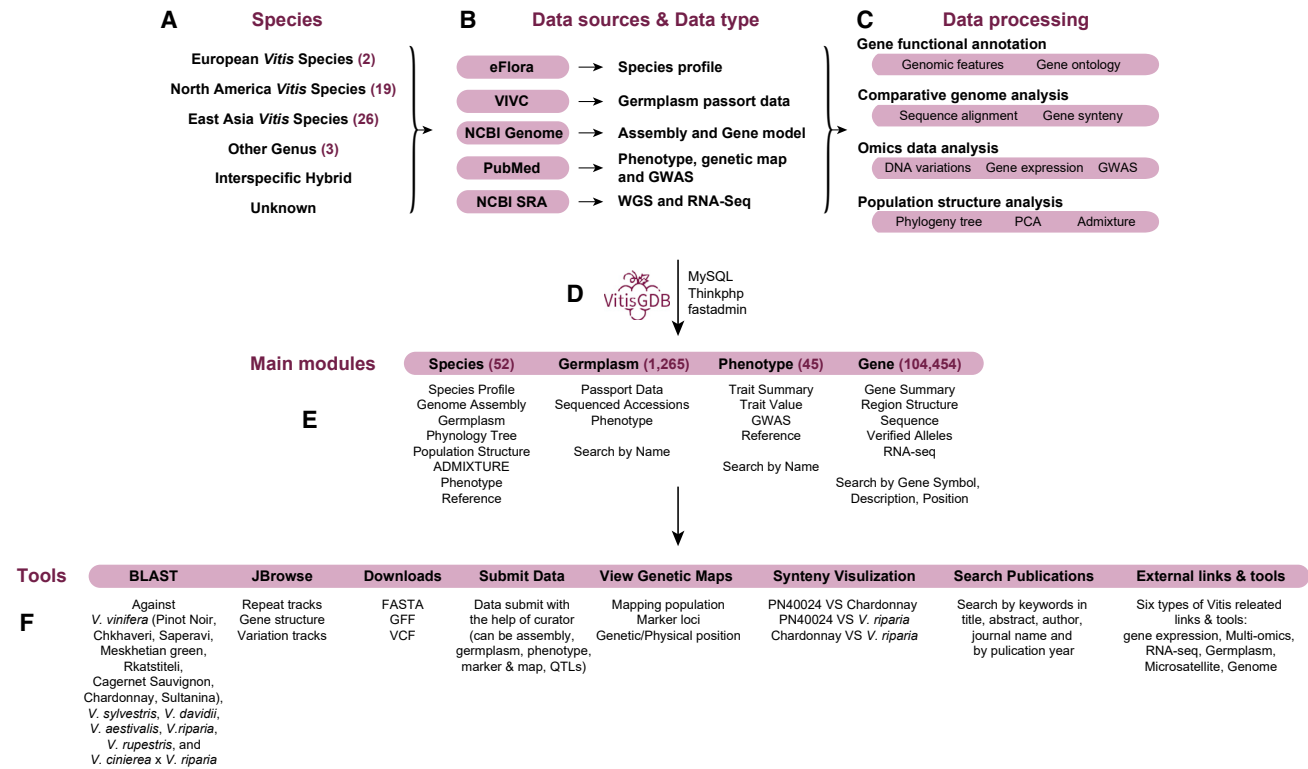


Figure 1. Schematic Illustration of VitisGDB

(A) Species information. Information for 50 species in six groups was collected. Highlighted numbers in parentheses indicate number of species under each group.
 (B) Data sources and data type. Data sources include species web sites such as eFlora, VIVC, and the public database NCBI, as well as literature. Data type includes information about species, germplasm, genome assembly, gene annotation, phenotypic trait values, genetic maps, resequencing data, and RNA-sequencing data.
 (C) Summary of data-processing methods.
 (D) Framework of VitisGDB. All data are stored in MySQL. Thinkphp and FastAdmin are used for front-end and back-end data management.
 (E) Main modules. Highlighted numbers in parentheses indicate the number of corresponding entries.
 (F) Overview of the integrated tools in VitisGDB with main functions listed below.

sequence databases are available for query of orthologous gene candidates. The input can be either plain text or a fasta sequence file. The alignment result (available in eight styles) opens up in a new page, detailing the overall alignment score, query length, and similarity between the query and subject sequences. The BLAST result allows secondary filtering, and the final subjects can be downloaded in HTML format.

JBrowse is an efficient visualization tool, which facilitates the viewing of gene models, CDS, heterozygous SNPs, and RNA sequencing data, each presented in a different color, in the context of the genomic region. At the moment, all available *Vitis* genomes and gene models are incorporated into JBrowse.

The JavaScript-based tool SynVisio is implemented to show the synteny relationships of three pairs of chromosome-level reference genomes (PN40024 versus Chardonnay, PN40024 versus *Vitis riparia*, and Chardonnay versus *V. riparia*). The visualization includes a hive plot indicating synteny between chromosomes, a dot plot indicating collinearity between two species/cultivars, and a scatterplot indicating identified signal strength. The threshold for displaying results in the hive plot and dot plot can be selected by dragging the little circle on the value bar

from the min to the max (above the dotted line at the lower left corner).

To date, three grapevine genetic maps are available, covering a total of 70 832 marker loci on the genome. For a selected genetic map, a heat plot shows the density of loci along the chromosomes and a corresponding table provides basic information about the mapping population and the genetic map. Double-clicking on the heat plot will zero in on a chromosome of interest on a new web page. The chromosome can be sized with the pointer to view regions in finer detail and show details for each locus with a hyperlink to JBrowse.

To allow personalized usage and analyses of the data, we have built a “Download” web page for all datasets available to the public. These include genome assembly sequences, annotation results, and genomic variations in FASTA, GFF, and VCF format, respectively. Considering the large size of the raw data for *de novo* assembly, resequencing projects, and RNA sequencing, we provide the NCBI BioProject ID and BioSample ID as well as the corresponding links on the web page. We have also imported the metadata for all *Vitis*-related publications from NCBI into VitisGDB for quick searching.

Correspondence

In summary, VitisGDB provides the most comprehensive view of *Vitis* genomic data to date and will be a valuable platform for studies on *Vitis* functional genomics and agronomic improvement. With the goal of becoming a community-built platform dedicated to making research results on grapevine broadly available, VitisGDB accepts the submission of all types of grapevine genetic data via the “Submit Data” page. VitisGDB will be continuously updated as genomic data from ongoing sequencing projects become available. New tools and analysis for transposable elements, non-coding RNAs, and environmental data will be added, so that VitisGDB will provide long-term support to the grapevine research community.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at *Molecular Plant Online*.

ACKNOWLEDGMENTS

Yunnan Provincial Key Programs of Yunnan Eco-friendly Food International Cooperation Research Center Project (2019ZG00908). The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received: February 19, 2020

Revised: March 20, 2020

Accepted: May 7, 2020

Published: May 14, 2020

**Xiao Dong^{1,10}, Wei Chen^{1,2,3,10,11},
Zhenchang Liang^{4,5}, Xuzhen Li^{1,6}, Peter Nick⁷,
Shanshan Chen⁸, Yang Dong^{1,3,11},
Shaohua Li^{4,*} and Jun Sheng^{6,9,*}**

¹State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan Agricultural University, Kunming 650201, China

²College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming 650201, China

³Yunnan Research Institute for Local Plateau Agriculture and Industry, Kunming 650201, China

⁴Beijing Key Laboratory of Grape Sciences and Enology, Laboratory of Plant Resources, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

⁵Sino-Africa Joint Research Center, Chinese Academy of Sciences, Wuhan 430074, China

⁶Faculty of Animal Science and Technology, Yunnan Agricultural University, Kunming 650201, China

⁷Botanical Institute, Karlsruhe Institute of Technology, Karlsruhe 76128, Germany

Molecular Plant

⁸BGI College, Zhengzhou University, Zhengzhou 450001, China

⁹Key Laboratory of Pu'er Tea Science, Ministry of Education, Yunnan Agricultural University, Kunming 650201, China

¹⁰These authors have contributed equally to this article.

¹¹Co-senior authors

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

<https://doi.org/10.1016/j.molp.2020.05.002>

REFERENCES

- Alston, J.M., and Sambucci, O. (2019). Grapes in the world economy. In *The Grape Genome*, D. Cantu and A. dWalker, eds. (Basel: Springer Nature), pp. 1–24.
- Canaguier, A., Grimplet, J., Di Gaspero, G., Scalabrin, S., Duchêne, E., Choisine, N., Mohellibi, N., Guichard, C., Rombauts, S., and Le Clainche, I. (2017). A new version of the grapevine reference genome assembly (12X. v2) and of its annotation (VCost. v3). *Genom. Data* **14**:56.
- Girolet, N., Rubio, B., Lopez-Roques, C., Valière, S., Ollat, N., and Bert, P.-F. (2019). De novo phased assembly of the *Vitis riparia* grape genome. *Sci. Data* **6**:1–8.
- Liang, Z., Duan, S., Sheng, J., Zhu, S., Ni, X., Shao, J., Liu, C., Nick, P., Du, F., and Fan, P. (2019). Whole-genome resequencing of 472 *Vitis* accessions for grapevine diversity and demographic history analyses. *Nat. Commun.* **10**:1–12.
- Minio, A., Massonnet, M., Figueroa-Balderas, R., Castro, A., and Cantu, D. (2019). Diploid genome assembly of the wine grape Carménère. *G3 (Bethesda)* **9**:1331–1337.
- OIV. (2019). OIV Statistical Report on World Vitiviniculture. <http://www.oiv.int/public/medias/6782/oiv-2019-statistical-report-on-world-vitiviniculture.pdf%20>.
- Roach, M.J., Johnson, D.L., Bohlmann, J., van Vuuren, H.J., Jones, S.J., Pretorius, I.S., Schmidt, S.A., and Borneman, A.R. (2018). Population sequencing reveals clonal diversity and ancestral inbreeding in the grapevine cultivar Chardonnay. *PLoS Genet.* **14**:e1007807.
- This, P., Lacombe, T., and Thomas, M.R. (2006). Historical origins and genetic diversity of wine grapes. *Trends. Genet.* **22**:511–519.
- Vondras, A.M., Minio, A., Blanco-Ulate, B., Figueroa-Balderas, R., Penn, M.A., Zhou, Y., Seymour, D., Ye, Z., Liang, D., Espinoza, L.K., et al. (2019). The genomic diversification of grapevine clones. *BMC Genomics* **20**:972.
- Zhou, Y., Massonnet, M., Sanjak, J.S., Cantu, D., and Gaut, B.S. (2017). Evolutionary genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *Proc. Natl. Acad. Sci.* **114**:11715–11720.

Molecular Plant, Volume 13

Supplemental Information

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, and Jun Sheng

1 SUPPLEMENTAL INFORMATION

2 **VitisGDB: the Multifunctional Database for Grapevine Breeding and Genetics**

3 Dong, X., Chen, W., et al.

4

5 **Supplementary Note:**

6 **Status quo of existing *Vitis* databases**

7 Database is a source of information where users are able to access and organize knowledge of
8 a common theme. Modern genome databases for plants (i.e. Rosacea
9 <https://www.rosaceae.org/>, maize <https://www.maizegdb.org/>, and cucurbit
10 <http://cucurbitgenomics.org/>) often incorporate bioinformatic tools and even germplasm
11 services, thereby, enabling the users to perform multiple tasks with ease (Martinez, 2016).
12 Currently, there are many *Vitis* databases where only descriptive information (i.e. species
13 name, country of origin, cultivar names, berry color, and usage etc.) and/or simple genetic
14 information (i.e. EST and SSR) for grapevine accessions are curated. These include the *Vitis*
15 International Variety Catalogue (VIVC, <http://www.vivc.de/>), the European *Vitis* database
16 (<http://www.eu-vitis.de/index.php>), the Armenian *Vitis* Database (<http://www.vitis.am/eng>),
17 the Bulgarian *Vitis* Database (<http://www.bulvitis-db.com/>) and so on. The conventional
18 genomic database such as the National Center for Biotechnology Information (NCBI,
19 <https://www.ncbi.nlm.nih.gov/>) and Ensembl (<http://ensembl.gramene.org>) does provide
20 access to all genomic data of published grapevines, but it is not a specialized platform for
21 grapevine genetics and breeding. There are also other databases and websites devoted to a
22 single aspect of grapevine genetic information. They could be grouped into several major
23 categories: (1) Gene expression browser (i.e. VTCdb; <http://vtcdb.adelaide.edu.au/Home.aspx>
24 and Grape eFP Browser; http://bar.utoronto.ca/efp_grape/cgi-bin/efpWeb.cgi). (2) Molecular
25 pathway database (i.e. *VitisNet*; [https://www.sdstate.edu/vitisnet-molecular-networks-](https://www.sdstate.edu/vitisnet-molecular-networks-grapevine)
26 [grapevine](https://www.sdstate.edu/vitisnet-molecular-networks-grapevine)). (3) Transcriptome database (<http://www.grapeworld.cn/gt/index.html>). (4)
27 Grape sRNA atlas (https://mpss.danforthcenter.org/dbs/index.php?SITE=grape_sRNA_atlas).

28 (5) Microsatellite database like SVMd (<http://www1.unine.ch/svmd/>) and GMC
29 (<http://meteo.iasma.it/genetica/gmc.html>). (6) Grapevine Genome database (i.e. UGRI;
30 <https://urgi.versailles.inra.fr/Species/Vitis> and grape genome database;
31 <http://genomes.cribi.unipd.it/grape/index.php>).

32

33 **All data sources used in the development of VitisGDB**

34 (1) Genome Assemblies. A total of 22 *Vitis* genome assemblies were included in the
35 VitisGDB. The genome assemblies of eight cultivated grapevines (Chkhaveri, Saperavi,
36 Meskhetian green, Rkatsiteli (Tabidze et al., 2017), Chardonnay clone I10V1 (Roach et al.,
37 2018), Tannat clone UY11 (Da Silva et al., 2013), Sultanina (Di Genova et al., 2014),
38 Zinfandel clone 03 (Vondras et al., 2019)), two wild species (*V. aestivalis* cv. Norton and *V.*
39 *riparia* cv. Gloire de Montpellier (Girollet et al., 2019)), and one hybrid cultivar (*V. cinerea*
40 × *V. riparia* cv. Boerner) were collected from the NCBI assembly database
41 (<https://www.ncbi.nlm.nih.gov/assembly/>), and the genome assembly of PN40024 12X.v2
42 (Jaillon et al., 2007) was downloaded from Ensembl
43 (http://ensembl.gramene.org/Vitis_vinifera/Info/Index). Six genome assemblies were
44 acquired from private websites, including Cabernet Sauvignon clone 08 (Chin et al., 2016)
45 and Carmenere clone 02 (Minio et al., 2019) from the Cantu Lab webpage
46 (<https://cantulab.github.io/data.html>), Chardonnay clone FPS 04 (Zhou et al., 2019) from
47 Zenodo (<https://zenodo.org/record/3337377#.XhQ49fkzZaR>), and three other assemblies of
48 Sultanina (Patel et al., 2018) from Open PRAIRIE of South Dakota State University
49 (https://openprairie.sdstate.edu/vitis_vinifera_sultanina/1/). Additionally, our lab sequenced
50 and assembled the genomes of one wild European grape (*V. vinifera* subsp. *sylvestris*), two
51 North America Species (*V. rupestris* and *V. riparia*) and one East Asia Species (*V. davidii*).
52 These unpublished draft assemblies will be first released in VitisGDB.

53 (2) Whole-genome Resequencing Data. Raw sequencing reads of 27 grapevine accessions
54 from Zhou et al. (2017) were downloaded from the NCBI SRA database. The sequencing
55 reads of 472 *Vitis* accessions from our group were also deposited in VitisGDB.

56 (3) RNA-Seq Data. Raw sequencing reads of 150 SRA experiments from BioProject
57 PRJNA386889 (Fasoli et al., 2018) were downloaded from the NCBI SRA database, which
58 include 17 time points of berry development from fruit set to maturity in Pinot Noir and
59 Cabernet Sauvignon.

60 (4) Phenotypic Data. Profile information of the *Vitis* species and cultivars was gathered from
61 eFlora (<http://efloras.org/>) and the passport data of all *Vitis* accessions were verified against
62 the VIVC database (<http://www.vivc.de/>). Phenotypic data and genome-wide association
63 analysis results were obtained partly from our lab (Liang et al., 2019), and partly from
64 publications (Guo et al., 2019; Laucou et al., 2018).

65

66 **Genomic Data processing**

67 Genome collinearity analysis between the chromosome-level assemblies (PN40024 (Jaillon et
68 al., 2007), Chardonnay clone FPS04 (Zhou et al., 2019), and *V. riparia* cv. Gloire de
69 Montpellier (Girollet et al., 2019)) was performed using MCScanX (Wang et al., 2012) with
70 default settings.

71

72 Whole-genome resequencing reads were mapped to the *Vitis vinifera* reference genome
73 (Jaillon et al., 2007) with BWA (Li and Durbin, 2009) using the default parameters. Software
74 SAMtools (Li et al., 2009) was used to convert mapping results into the BAM format, and
75 then filter unmapped and non-unique reads. The Picard package
76 (<http://broadinstitute.github.io/picard/>) was used to filter the duplicated reads. For variation
77 detection, the best practice workflow recommended by Genome Analysis Toolkit (GATK)
78 (McKenna et al., 2010) was applied. SNPs annotation were performed according to the
79 grapevine genome annotation result using the package ANNOVAR (Wang et al., 2010).

80

81 Based on the detected SNPs, Principal component analysis (PCA) was performed with the
82 software Genome-wide Complex Trait Analysis (GCTA) (Yang et al., 2011). The first three
83 eigenvectors were plotted in an interactive graph. Population structure was analyzed using the
84 ADMIXTURE program (Alexander et al., 2009) with K ranging from 2 to 14 by a block-
85 relaxation algorithm. The whole-genome SNPs were used to construct the ML phylogenetic
86 tree with 100 bootstrap by SNPhylo (Version: 20140701)(Lee et al., 2014).

87

88 Fastq-dump tool in the SRA Toolkit (v.2.9.6, <https://github.com/ncbi/sra-tools>) was used to
89 convert .sra files of the transcriptomic data into fastq format. Trimmomatic (v.0.32) (Bolger
90 et al., 2014) was then used to trim the sequencing reads by the following the parameters
91 “ILLUMINACLIP: Trimmomatic-0.32/adapters/TruSeq3-SE.fa:2:30:10 LEADING:3
92 TRAILING:3 SLIDINGWINDOW:4:15”. The clean RNA-seq reads were aligned to the *V.*
93 *vinifera* genome assembly using TopHat (v. 2.0.10) (Trapnell et al., 2009) with default
94 parameters and the FPKM value was calculated for each protein-coding gene by Cufflinks (v.
95 2.1.1) (Trapnell et al., 2012) using default parameters.

96

97 **Data integration**

98 Profile information of *Vitis* species, sequenced germplasm, genome assemblies, phenotypic
99 traits, and all relevant information of protein-coding genes were sorted into different
100 categories in a MySQL database (<https://www.mysql.com>). Keywords was curated for the
101 internal query and hyperlink of all information.

102

103 **Database construction**

104 Since all data were stored and managed by MySQL, ThinkPHP and FastAdmin were used for
105 frontend and backend interactive queries, respectively. All genomic features were visualized
106 in JBrowse (Buels et al., 2016) and its plugins. The BLAST server was supported by

107 Viroblast (Deng et al., 2007). Echarts (<https://www.echartsjs.com/>), JavaScript and jQuery
108 were implemented to enhance user experience. Phylogeny.IO (Nikola and S, 2019)
109 (<https://github.com/oist/phylogeny-io>) was used for interactive visualization of the
110 phylogenetic analysis result. QuiGMap (<https://github.com/MFlores2021/QuiGMap>) and
111 SynVisio (<https://github.com/kiranbandi/synvisio>) were applied to visualize genetic map and
112 genome synteny analysis, respectively. All functionalities of VitisGDB have been tested in
113 Google Chrome, Apple Safari, and 360 Browser.

114

115 **Publication query criteria**

116 By using the query in NCBI PubMed: ((((((((((trait[Title/Abstract]) OR
117 QTL[Title/Abstract]) OR gene[Title/Abstract]) OR genome[Title/Abstract]) OR
118 map[Title/Abstract]) OR microsatellite[Title/Abstract]) OR annotation[Title/Abstract]) OR
119 EST[Title/Abstract]) OR marker[Title/Abstract]) OR sequence[Title/Abstract]) OR
120 GWAS[Title/Abstract]) AND (((vitis[Title/Abstract]) OR Vitaceae[Title/Abstract]) OR
121 grape[Title/Abstract]), we obtained and imported the metadata for a total of 3,943
122 publications to VitisGDB.

123

124 **Data availability**

125 The raw sequencing data sets generated and used in the VitisGDB are available from NCBI
126 under the BioProject accessions PRJNA625615, PRJNA625617, PRJNA625619,
127 PRJNA625621 for *V. vinifera* subsp. *sylvestris*, *V. rupestris*, *V. riparia*, and *V. davidii*,
128 respectively.

129

130 **References:**

131 Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of
132 ancestry in unrelated individuals. *Genome research* 19:1655-1664.
133 Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for
134 Illumina sequence data. *Bioinformatics* 30:2114-2120.

135 Buels, R., Yao, E., Diesh, C.M., Hayes, R.D., Munoz-Torres, M., Helt, G., Goodstein, D.M.,
136 Elsik, C.G., Lewis, S.E., and Stein, L. (2016). JBrowse: a dynamic web platform for
137 genome visualization and analysis. *Genome biology* 17:66.

138 Chin, C.S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C.,
139 O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., et al. (2016). Phased diploid
140 genome assembly with single-molecule real-time sequencing. *Nat Methods* 13:1050-
141 1054.

142 Da Silva, C., Zamperin, G., Ferrarini, A., Minio, A., Dal Molin, A., Venturini, L., Buson, G.,
143 Tononi, P., Avanzato, C., and Zago, E. (2013). The high polyphenol content of
144 grapevine cultivar tannat berries is conferred primarily by genes that are not shared
145 with the reference genome. *The Plant Cell* 25:4777-4788.

146 Deng, W., Nickle, D.C., Learn, G.H., Maust, B., and Mullins, J.I. (2007). ViroBLAST: a
147 stand-alone BLAST web server for flexible queries of multiple databases and user's
148 datasets. *Bioinformatics* 23:2334-2336.

149 Di Genova, A., Almeida, A.M., Muñoz-Espinoza, C., Vizoso, P., Travisany, D., Moraga, C.,
150 Pinto, M., Hinrichsen, P., Orellana, A., and Maass, A. (2014). Whole genome
151 comparison between table and wine grapes reveals a comprehensive catalog of
152 structural variants. *BMC plant biology* 14:7.

153 Fasoli, M., Richter, C.L., Zenoni, S., Bertini, E., Vitulo, N., Dal Santo, S., Dokoozlian, N.,
154 Pezzotti, M., and Tornielli, G.B. (2018). Timing and order of the molecular events
155 marking the onset of berry ripening in grapevine. *Plant physiology* 178:1187-1206.

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 Guo, D.-L., Zhao, H.-L., Li, Q., Zhang, G.-H., Jiang, J.-F., Liu, C.-H., and Yu, Y.-H. (2019).
159 Genome-wide association study of berry-related traits in grape [*Vitis vinifera* L.]
160 based on genotyping-by-sequencing markers. *Horticulture research* 6:1-13.

161 Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisine, N.,
162 Aubourg, S., Vitulo, N., and Jubin, C. (2007). The grapevine genome sequence
163 suggests ancestral hexaploidization in major angiosperm phyla. *nature* 449:463.

164 Laucou, V., Launay, A., Bacilieri, R., Lacombe, T., Adam-Blondon, A.-F., Berard, A.,
165 Chauveau, A., de Andrés, M.T., Hausmann, L., and Ibanez, J. (2018). Extended
166 diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs.
167 *PloS one* 13.

168 Lee, T.H., Guo, H., Wang, X., Kim, C., and Paterson, A.H. (2014). SNPhylo: a pipeline to
169 construct a phylogenetic tree from huge SNP data. *BMC Genomics* 15:162.

170 Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler
171 transform. *bioinformatics* 25:1754-1760.

172 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis,
173 G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools.
174 *Bioinformatics* 25:2078-2079.

175 Liang, Z., Duan, S., Sheng, J., Zhu, S., Ni, X., Shao, J., Liu, C., Nick, P., Du, F., and Fan, P.
176 (2019). Whole-genome resequencing of 472 *Vitis* accessions for grapevine diversity
177 and demographic history analyses. *Nature communications* 10:1-12.

178 Martinez, M. (2016). Computational tools for genomic studies in plants. *Current genomics*
179 17:509-514.

180 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A.,
181 Garimella, K., Altshuler, D., Gabriel, S., and Daly, M. (2010). The Genome Analysis
182 Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing
183 data. *Genome research* 20:1297-1303.

184 Minio, A., Massonnet, M., Figueroa-Balderas, R., Castro, A., and Cantu, D. (2019). Diploid
185 genome assembly of the wine grape Carménère. *G3: Genes, Genomes, Genetics*
186 9:1331-1337.

187 Nikola, J., and S, M.A. (2019). Interactive web-based visualization and sharing of
188 phylogenetic trees using phylogeny.IO. *Nucleic Acids Research:W1*.

189 Patel, S., Lu, Z., Jin, X., Swaminathan, P., Zeng, E., and Fennell, A.Y. (2018). Comparison
190 of three assembly strategies for a heterozygous seedless grapevine genome assembly.
191 BMC genomics 19:57.

192 Roach, M.J., Johnson, D.L., Bohlmann, J., van Vuuren, H.J., Jones, S.J., Pretorius, I.S.,
193 Schmidt, S.A., and Borneman, A.R. (2018). Population sequencing reveals clonal
194 diversity and ancestral inbreeding in the grapevine cultivar Chardonnay. PLoS
195 genetics 14:e1007807.

196 Tabidze, V., Pipia, I., Gogniashvili, M., Kunelauri, N., Ujmajuridze, L., Pirtskhalava, M.,
197 Vishnepolsky, B., Hernandez, A.G., Fields, C.J., and Beridze, T. (2017). Whole
198 genome comparative analysis of four Georgian grape cultivars. Molecular Genetics
199 and Genomics 292:1377-1389.

200 Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions
201 with RNA-Seq. Bioinformatics 25:1105-1111.

202 Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg,
203 S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression
204 analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols 7:562-
205 578.

206 Vondras, A.M., Minio, A., Blanco-Ulate, B., Figueroa-Balderas, R., Penn, M.A., Zhou, Y.,
207 Seymour, D., Ye, Z., Liang, D., Espinoza, L.K., et al. (2019). The genomic
208 diversification of grapevine clones. BMC Genomics 20:972.

209 Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic
210 variants from high-throughput sequencing data. Nucleic acids research 38:e164-e164.

211 Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.-h., Jin, H., Marler, B.,
212 and Guo, H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of
213 gene synteny and collinearity. Nucleic acids research 40:e49-e49.

214 Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-
215 wide complex trait analysis. The American Journal of Human Genetics 88:76-82.


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

219 Zhou, Y., Minio, A., Massonnet, M., Solares, E., Lv, Y., Beridze, T., Cantu, D., and Gaut,
220 B.S. (2019). The population genetics of structural variants in grapevine
221 domestication. Nat Plants 5:965-979.

222

223


A Species profile



Genus	Vitis
Species	V. vinifera
Common name	Cultivated grape
Distribution	Worldwide
Chromosome number	2n = 38

Morphological Description

Plants sprawling to moderately high climbing, sparsely branched. Branches bark exfoliating in shreds or plates; nodal diaphragms 3–5 mm thick; branchlets terete to slightly angled, pubescent, sometimes glabrescent, growing tips not enveloped by unfolding leaves; tendrils along length of branchlets, persistent, tendrils (or inflorescences) at only 2 consecutive nodes; nodes not red-banded. Leaves: stipules usually more than 3.5 mm; petiole + equaling blade; blade cordate-ovate to cordate-ovate, 12–20 cm, usually 3-shouldered to 3-5-lobed, sometimes deeply so, apex acute to short acuminate, abaxial surface not glaucous, sparsely pubescent to glabrous, visible through hairs, adaxial surface usually glabrous. Inflorescences 10–20 cm. Flowers bisexual. Berries usually reddish purple to nearly black, sometimes yellow-green, + glaucous, oblong to ellipsoid, 8–25 mm diam., skin adhering to pulp; lenticels absent.



B Genome assembly

Assembly	Assembly level	Size(Mb)	Contig N50(Mb)	Scaffold N50(Mb)	Gene number	BUSCO(%)
cv. Pinot Noir clone PN42004 (B4)	Chr	493.21	0.065	2.07	30434	-
cv. Pinot Noir clone PN42004 (10X.v0)	Chr	496.26	-	22.39	22563	95.6
cv. Pinot Noir clone PN42004 (10X.v0)	Chr	496.2	0.102	22.39	41963	95.6
cv. Chihaveri	Chr	427.17	0.112	22.7	-	92
cv. Saperavi	Chr	427.04	0.112	22.69	-	92
cv. Meskhetian green	Chr	427.21	0.112	22.71	-	92
cv. Reitschel	Chr	426.6	0.112	22.67	-	92
cv. Cabernet Sauvignon clone 05	Clg	501.42 (P-contigs), 367.76 (Haplotigs)	2.173 (P-contigs), 0.779 (Haplotigs)	-	55036 (P-contigs), 40444 (Haplotigs)	95.4
cv. Chardonnay clone 10V1	Clg	400 (P-contigs), 378 (Haplotigs)	0.936 (P-contigs), 0.378 (Haplotigs)	-	-	93.54
cv. Chardonnay clone FPS-04	Chr	606	1.24	24.5	38020	93.4

Showing 1 to 10 of 17 entries

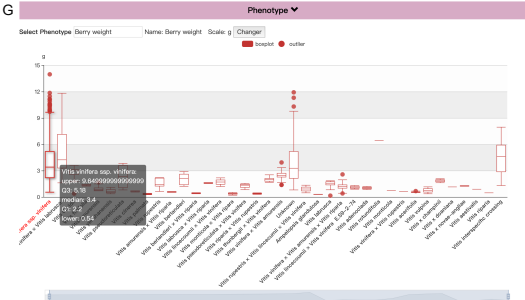
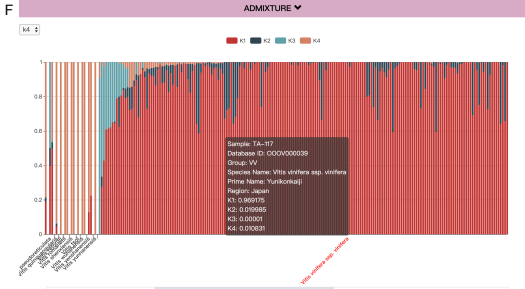
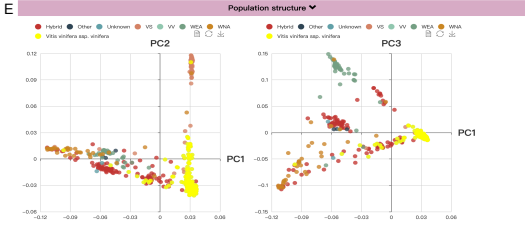
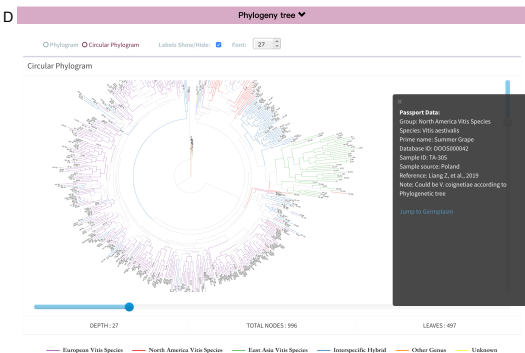
[Download](#)

C Germplasm

Database ID	Prime name	Country of Origin	VVC
ARGV000001	Fahri Kalemak	Afghanistan	4035
ARGV000002	Kandahari Shah	Afghanistan	6866
ARGV000003	Nacos	Afghanistan	8043
ARGV000004	Sekidan	Afghanistan	26057
ALBV000001	Zana	Albania	12684
ARGV000001	Criolla Grande Sanjuntina	Argentina	3261
ARGV000002	Moscakul	Argentina	8019
ARGV000003	Paizga	Argentina	6964
ARGV000004	Perlon	Argentina	9170
ARGV000005	Big Muscat Seedless	Argentina	17650

Showing 1 to 10 of 982 entries

[Download](#)



H Reference

Author	Title	Journal	Year	PMID
Amends M, Vondras et al.	The genomic diversification of clonally propagated grapevines	BioRxiv	2019	
Andrea Minio et al.	Diploid Genome Assembly of the Wine Grape Carménère	G3 (Bethesda)	2019	30923136
Guo D L, et al.	Genome-wide association study of berry-related traits in grape [Vitis vinifera L.] based on genotyping-by-sequencing markers	Horticulture Research	2019	30603096
Liang Z, et al.	Whole-genome resequencing of 472 Vitis accessions for grapevine diversity and demographic history analyses	Nature Communications	2019	30687614
Zhou Y, et al.	The population genetics of structural variants in grapevine domestication	Nature Plants	2019	31505640
Lauou V, et al.	Extended diversity analysis of cultivated grapevine Vitis vinifera with 10K genome-wide SNPs	PLoS One	2016	28420602
Poach MJ, et al.	Population sequencing reveals clonal diversity and ancestral inbreeding in the grapevine cultivar Chardonnay	PLoS Genet	2016	30458008
Sagar Patel et al.	Comparison of three assembly strategies for a heterozygous seedless grapevine genome assembly	BMC Genomics	2018	29343235
Ganaguru A, et al.	A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VvCost.v3)	Genomics Data	2017	28973018

226 **Supplementary Figure 1: An example of Species module for *V. vinifera* ssp. *vinifera*.** (A)
227 Basic species information, which include Latin name, chromosome number, distribution area
228 and description. Figures on two sides are preventive picture and geographic distribution of
229 the sequenced germplasm. (B) Available genome assembly, which listed assembled cultivar
230 and clone, as well as the assembly statistics. (C) Sequenced germplasm, which contains
231 database ID in VitisGDB, prime name, country of origin and ID in VIVC database. (D)
232 Interactive Phenology tree, which includes 497 accessions and their detailed information.
233 Branches are colored by *Vitis* group. (E) Scatter plot of principle component analysis result,
234 which indicate the species population structure with the corresponding species highlighted in
235 yellow. (F) Bar plot of ADMIXTURE result with K from 2 to 14. Zooming into the detail
236 shows the detail information of accessions of such species. (G) Box plot/bar plot which
237 present the distribution of quantitative/qualitative trait. (H) Related references.

238

239

Comp	Assembly	Downloads	Reference
Vitis vitifera esp. vitifera	cv. Pinot Noir clone FR40024_12xv0	Genome Assembly, Genes, Proteins (External Link)	Adaro-Strickland A.E. 2020
Vitis vitifera esp. vitifera	cv. Pinot Noir clone FR40024_12xv0	Raw, Genome Assembly, Genes, CDS, Proteins (External Link)	Compoint A. et al. 2020
Vitis vitifera esp. vitifera	cv. Chardonnay v1.0	Raw, Genome Assembly (External Link)	Talbot V. et al. 2017
Vitis vitifera esp. vitifera	cv. Saperavi v1.0	Raw, Genome Assembly (External Link)	Talbot V. et al. 2017
Vitis vitifera esp. vitifera	cv. Merlot Noir v1.0	Raw, Genome Assembly (External Link)	Talbot V. et al. 2017
Vitis vitifera esp. vitifera	cv. Cabernet Sauvignon clone 05_11.0	Raw, Genome Assembly, Genes, Repeat, CDS, Proteins (External Link)	Choi C.S. et al. 2016
Vitis vitifera esp. vitifera	cv. Chardonnay clone TRV1 v1.0	Raw, Genome Assembly, Genes, CDS, Proteins (External Link)	Rouch N. et al. 2018
Vitis vitifera esp. vitifera	cv. Tannin clone UY1 v1.0	Raw (External Link)	De Silva A. et al. 2019
Vitis vitifera esp. vitifera	cv. Sultanina v1.0	Raw, Genome Assembly (External Link)	Di Giampa A. et al. 2019
Vitis vitifera esp. vitifera	cv. Sultanina_PLAT_1_MERGE_and_PLAT_2_C1	(External Link)	Petit S. et al. 2018
Vitis vitifera esp. vitifera	cv. Carmenere clone 02_v1.0	Raw, Genome Assembly, Genes, Repeat, CDS, Proteins (External Link)	Wojcik A. et al. 2019
Vitis vitifera esp. vitifera	cv. Zinfandel clone 01_v1.0	Raw (External Link)	Wojcik A. et al. 2019

BLAST

Basic Search - using default BLAST parameter settings

Basic search parameters for this search:

Query sequence: [input field]

Database: [input field]

Advanced Search - verify your favorite parameters below:

Algorithm: BlastN
 Matrix: BLOSUM62
 Gap costs: 11
 Filter: None
 Maximal target seqs: 100
 Adjusted expect threshold: 0.001
 Number of hits: 1

Accession	Score	E-value	Identical	Positives	Bits
... (table continues with search results)					

Synteny Visualization

SyntViz: An Interactive Multiscale-Synteny Visualization Tool for Hi-C

Step 1: Choose Species Pair
 Pinot Noir clone FR40024_12xv0 | Choose

Step 2: Choose Plot Character
 Multi-Locus Analysis | Single Analysis
 Default Dashboard | Dot Plot | Linear Plot

Step 3: Choose Chromosomes
 Source Chromosomes: 4 items selected | Target Chromosomes: 4 items selected | GO-D

Count | Match Score | E value

JBrowse

Genome Track View

Genome: [track]

Annotations: [track]

JBrowse logo

Map Overview

View Genetic Maps

Genetic map - C0-50/Reeling_F1_Reeling

Key Data:
 Species: Vitis vinifera L.
 Software: R/GenMap
 Map Type: F1
 Population Type: F1
 Map Size: 48K
 Parental Parent: C0-50
 Parental Parent: Reeling
 Number of C0s: 10
 Number of Loci: 400
 Reference: Wang H. et al. 2018

External links & tools

Gene expression | Multi-omics | RNA-seq | Genomics | Microsatellites | Genomes

Database: **Armenian Vitis Database**
 Feature: It is a public information source, including molecular and anthropographic characterization of Armenian grape genotypes.

Database: **VITD**
 Feature: It is a web-based database, which contains the genotypes of grapevine cultivars/accessions from the grapevine collection maintained at the Institute of Viticulture and Enology, Plovdiv, Bulgaria.

Database: **The European Vitis Database**
 Feature: It contains genetic resources of European grapes.

Database: **French Plant Stage Database**
 Feature: It is a catalogue of vine growth in France.

Database: **French Network of Grapevine Sequences (FRV)**
 Feature: It includes 107 sequenced markers located in the chromosomes of several vitifera parental genotypes and selected 1000 SNPs from 107 accessions from different countries.

Search Publications

View Genetic Maps

To search for publications, please enter keywords in the text boxes below. You can limit your search by selecting the field in the dropdown box.

ANY Field | grape
 ANY Field | none
 Years | 0

Search | Reset

Found 3073 Results

Year Publication

- 1980 Identification of closely related Agrobacterium vitis isolates by chromosomal DNA probes
Ottavio L, Casavola J
Journal of Phytopathology 1980; 10: 1-10
- 1982 Physical map of the vitigen T1 plasmid pT1
Gomez EC, Chastel C, Sengul T, de la Haza E, Ottavio L
Phytopathology 1982; 72: 100-104
- 1982 Organization and functional analysis of three T-DNA clones from the vitigen T1 plasmid pT1
Casavola J, Gomez EC, Ottavio L
Journal of Phytopathology 1982; 10: 1-10

Publication overview

Title: Identification of closely related Agrobacterium vitis isolates by chromosomal DNA probes
 Authors: Ottavio L, Casavola J
 Journal: Journal of Phytopathology
 Year: 1980
 Volume: 10
 Pages: 1-10

240 **Supplementary Figure 2: User-friendly interfaces for seven integrated tools.** (A)
241 BLAST; (B) JBrowse; (C) Downloads; (D) View Genetic Maps; (E) Synteny Visualization;
242 (F) Search Publications. (G) External Links & Tools.
243