



HAL
open science

Bioinformatics tools in grapevine genomics

J rome Grimplet, Julie Dickerson, Anne-Francoise A.-F. Adam-Blondon,
Grant Cramer

► **To cite this version:**

J rome Grimplet, Julie Dickerson, Anne-Francoise A.-F. Adam-Blondon, Grant Cramer. Bioinformatics tools in grapevine genomics. Genetics, Genomics, and Breeding of Grapes, Science Publishers, 390 p., 2011, 978-1-57808-717-4. hal-02809785

HAL Id: hal-02809785

<https://hal.inrae.fr/hal-02809785>

Submitted on 6 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin e au d p t et   la diffusion de documents scientifiques de niveau recherche, publi s ou non,  manant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv s.

13

Bioinformatics Tools in Grapevine Genomics

Jérôme Grimplet,^{1,5,*} Julie Dickerson,²
Anne-Françoise Adam-Blondon³ and Grant Cramer⁴

ABSTRACT

With the release of the grapevine genome sequence and with the increasing affordability of high throughput analysis tools, an ever increasing wealth of grapevine bioinformatics resources have been developed over the last decade in highly diverse biological fields. The grapevine community can now access extensive databases containing information relative to molecular markers, genetic maps, the genome sequence and its annotation, gene expression, proteins and metabolites. This chapter presents how to access these resources and provides some review of their contents and use. The next challenge is the further development of systems biology for grapevine; regarding this aspect, the different tools currently available for the integration of different types of data are discussed.

Keywords: Database, maps, STS, EST, gene annotation, gene ontology, system biology

¹South Dakota State University, Horticulture, Forestry, Landscape and Parks Department, Brookings, SD, 57006, USA.

²Electrical and Computer Engineering, Iowa State University, Ames, IA, 50011-3060, USA.

³UMR INRA UEVE ERL CNRS, Génomique Végétale (URGV), 2 rue Gaston Crémieux, BP5708, 91 057 Evry cedex, France.

⁴1664 North Virginia Street, Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NY 89557, USA.

⁵Instituto de Ciencias de la Vid y del Vino (CSIC,UR, Gobierno de La Rioja), C /Madre de Dios 51, 26006 Logroño, Spain.

*Corresponding autor: jerome.grimplet@icvv.es

13.1 Introduction

The grapevine genomic data has been increasing very rapidly since the start of the International Grape Genome Program in 2001 (IGGP) (www.vitaceae.org). Starting from scratch at that time, the genome sequence has been published twice (Jaillon et al. 2007; Velasco et al. 2007), more than 360,000 expressed sequence tags (ESTs) are now available (*Vitis vinifera*: July 2010, <http://www.ncbi.nlm.nih.gov/nucest>), and large efforts in proteomics and metabolomics have been successful (Chapter 12). There is more than ever a need for common databases and bioinformatics tools to allow a wide and deep use of these important resources. The first policy of the IGGP was to advise that deposition of the resources should be placed in public databases like the National Center for Biotechnology Information (NCBI). This advice has been followed for the most part by the grapevine community. However, there is also a need for better coordination in many aspects and of an interface with NCBI. As the grapevine community is a rather small one when compared to the community working on *Arabidopsis*, it will be difficult to raise funds to end up with a highly centralized and productive system like TAIR (<http://www.arabidopsis.org>). Other more collaborative initiatives are proposed such as for the Sol project (Menda et al. 2008) and may be a good middle term system. Such a model could be followed in the future with the help of the French national repository for plant genomic data (URGI; <http://urgi.versailles.inra.fr/>).

13.2 Molecular Marker and Genetic Map Databases

The international consortium for sequencing the grape genome was strongly encouraged to deposit the marker sequences and primers in the STS (sequence tagged sites) section of the GenBank public database maintained by NCBI. Currently, there are 779 grapevine unique STS in UniSTS (http://www.ncbi.nlm.nih.gov/genome/sts/unists_stats.html). Most of these publically available markers are simple sequence repeat (SSR) markers, mainly developed from *Vitis vinifera*.

More than 500 markers, mainly simple sequence repeat (SSR) markers, have been mapped onto the 17 maps stored in UniSTS and in the 27 maps stored at URGI (<http://urgi.versailles.inra.fr/GnpMap>). These maps were first made available through the GnpMap database of the French national repository for plant genomic data maintained by URGI and then transferred to NCBI. URGI also provides access to the Cabernet Sauvignon physical map and its links to the grapevine genetic map and the grapevine reference genome sequence (Moroldo et al. 2008; <http://urgi.versailles.inra.fr/cmap/>) using the GMOD tool, CMap (www.gmod.org). A comprehensive view of the data stored at URGI is provided by the web page <http://urgi.versailles.inra.fr/index.php/urgi/Species/Vitis/>.

In the near future, large sets of single nucleotide polymorphisms (SNPs) should become available. A large resource of SNP markers was for instance developed for mapping at Istituto Agrario di San Michele All'Adige-Fundazione Edmund Mach (Italy, Troggio et al. 2007; Pindo et al. 2008; Vezzulli et al. 2008) and the maps and their links with a Pinot Noir physical map and genome sequence can be viewed at <http://genomics.research.iasma.it/iasma/>. The markers are however not yet deposited at NCBI. More recently, a project of SNP discovery using the new high-throughput sequencing technologies has been released (Myles et al. 2010). A grape Gramene database has been set up to help the visualization of the results when they are obtained (http://www.gramene.org/Vitis_vinifera/). Such projects are now possible thanks to the availability of the reference genome sequence.

13.3 Genome Databases and Gene Annotation

The published version (8×) of the grapevine reference genome annotated sequence (Jaillon et al. 2007) has been deposited at EMBL and can be queried through the GBrowse interface (www.gmod.org) provided by two members of the sequencing consortium, in France (<http://www.genoscope.cns.fr/Vitis>) and Italy (<http://www.appliedgenomics.org/vitis8x>). It has been decided by the international community to transfer the last version (12×) to URGI, which will maintain it for the future and provide regular annotation updates to NCBI. The 12x genome is now accessible at URGI (http://urgi.versailles.inra.fr/cgi-bin/gbrowse/vitis_12x_pub/) and EMBL websites. The annotation provided was generated automatically by integrating several layers of evidence (alignments with ESTs, with protein, *de novo* prediction and comparative genomics) using the software, GAZE, which was manually calibrated (Howe et al. 2002; Jaillon et al. 2007). In parallel, the genome assembly of another *Vitis vinifera* genotype (cv. Pinot Noir) and its automatic annotation is available at <http://genomics.research.iasma.it/iasma/> (Chapter 9 Section 9.4).

The next major objective is to organize a coordinated community of manual annotators. This is not simple; for example, in the *Arabidopsis* community it has been done through the parallel, integrated projects of (i) communities of bioinformaticists and expert biologists that manually annotated gene families, both at the structural and the functional level in a highly coordinated way, with in-house tools and high standards (see for example, Aubourg et al. 2005) and (ii) the TAIR repository (www.arabidopsis.org) that regularly improves and updates the automatic annotation of the *Arabidopsis* genome sequence and also manually curates the results (Swarbreck et al. 2007). The goal is to progressively include and centralize the work achieved by experts like the one stored at the Center for Bioinformatics of Peking University (Grape Transcription Factor Database: <http://plantfdb.cbi.pku.edu.cn/index.php?sp=Vvi>) on transcription factors in

a single genome browser and/or to NCBI TPA (Third Party Annotation). Several tools or databases can help such community surveys like PlantGDB, Gramene, FLAGdb++, and DFCI.

The *Vitis* section of PlantGDB (vvGDB; <http://www.plantgdb.org/VvGDB/>) encompasses primary sequence data obtained from GenBank, including the most current GSS (Genome Survey Sequences), bacterial artificial chromosome (BAC) assembly and annotation from the Genoscope, EST, cDNA sequences, and protein sequences derived from the annotated genes. The mapping of cDNA, EST and Affymetrix probesets sequences onto the genome as well as interpretations of the overall data are based on the PlantGDB spliced alignment results obtained with the GeneSeqer software (<http://deepc2.psi.iastate.edu/cgi-bin/gs.cgi>). The Grapevine (*Vitis vinifera*) Transcription Factor Database collected 867 predicted transcription factors (TFs) from wine grape transcripts using input sequences from the PlantGDB-assembled Unique Transcript-fragment derived from *Vitis vinifera* mRNAs from October 2006 (based on GenBank release 155). Extensive annotations for those TFs have been made, including similarity searches against major databases (e.g., Uniprot, RefSeq, EMBL, TRANSFAC), InterPro domain information and EST expression information extracted from UniGene. Fifty-nine transcription factor families have been identified in grape.

Manual gene family functional annotation is greatly facilitated through the use of the databases FLAGdb++ (Samson et al. 2004; <http://urgv.evry.inra.fr/projects/FLAGdb++/HTML/index.shtml>), GRAMENE (<http://www.gramene.org>) and the DFCI Grape Gene Index (<http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=grape>). FLAGdb++ allows the query of gene families using keywords of Pfam domains (Finn et al. 2008), the retrieval of all their sequences in one query, their mapping (BLAST) on the *Arabidopsis* and rice genome. GRAMENE proposes whole genome alignments with the *Arabidopsis* and the poplar genome, phylogenetic trees for each gene, identification of orthologs and paralogs, prediction of alternative splicing, knowledge on genetic variation, mapping of different kind of domains and signatures. The DFCI Grape Gene Index integrates research data from international grapevine EST sequencing and gene research projects. The latest release of the Grape Gene Index (Release 7.0; 17 April, 2010) has been built from 352,730 ESTs and 25,497 ETs (mature transcripts corresponding to the genes from the heterozygous genome). The Grape Gene Index contains 80,778 unique sequences (34,154 TCs, 31,813 singleton ESTs and 14,811 singleton ETs). Several functional annotation tools are available at the DFCI Grape Gene Index: the prediction of alternative splice variants (5,306 clusters); the comparison of EST expression between different libraries or tissues for libraries with over 50 ESTs; the classification of tentative contigs (TCs) by Gene ontology (GO) vocabularies (16,633 TC sequences with Molecular Function, 16,781 TC sequences with Biological Process,

20,570 TC sequences with Cellular Component); and the list of all 70-mer oligo predictions for the TC and the association of TCs with metabolic and signaling pathways.

13.4 Gene Expression Databases

Gene expression database resources are available on grapevine, such as repositories for microarray experiments, which are not limited to grapevine (GEO, ArrayExpress, PlexDB), MPSS (massively parallel signature sequencing), small RNA, or EST databases (VitisExpDB).

Grape MPSS (<http://mpss.udel.edu/grape/>) is a database that collected sequences based upon MPSS. It is a limited database that allows one to estimate transcript abundance in different tissues (Iandolino et al. 2008). This deep sequencing approach has the power to measure very low abundance transcripts that ordinary microarrays are not able to measure adequately. A small RNA website (<http://smallrna.udel.edu/index.php>) has also been generated from the same research group. This site allows mapping of small RNAs against the plant genome of choice including grape. One can determine if your transcript of choice might be regulated by small RNAs.

VitisExpDB (Doddapaneni et al. 2008, http://cropdisease.ars.usda.gov/vitis_at/main-page.htm) is a relational database that houses annotated EST and gene expression data. The database contains about 300,000 ESTs downloaded from the NCBI website. Several tools are available such as BLAST and the GO annotation of the ESTs. Results from a custom microarray experiment on Pierce's disease infected tissues are also available. The expression through EST experiments that are mapped to an *Arabidopsis* gene can be viewed on 25 AraCyc pathways (Zhang et al. 2005).

There are currently three public repositories where microarray experiments for grapevine have been uploaded.

Gene Expression Omnibus (GEO; www.ncbi.nlm.nih.gov/geo/) is a public gene expression/molecular abundance repository supporting MIAME compliant data submissions, and a curated, online resource for gene expression data browsing, query and retrieval. Fourteen grapevine microarray experiments data (including published works from Peng et al. 2007; Fung et al. 2008; Rotter et al. 2008; Lund et al. 2008; Mathiason et al. 2009; Ophir et al. 2009; Mica et al. 2009; Albertazzi et al. 2009; Hren et al. 2009; Bellin et al. 2009; Sreekantan et al. 2010; Polesani et al. 2010) have been submitted to GEO. GEO also includes simple tools for exploration, visualization, and analysis of the data sets.

ArrayExpress (<http://www.ebi.ac.uk/microarray-as/ae/>) is supported by the European Bioinformatics Institute (EBI). Not only is it able to store datasets from different platforms (MIAME- and MINSEQE-compliant data), but has the capability of conducting large meta-analyses across different

microarrays using different technologies. There are 13 grapevine microarray data sets that have been submitted to ArrayExpress (Pilati et al. 2007; Fung et al. 2008; D'Onofrio et al. 2009; Fernandez et al. 2010; Rotter et al. 2009; Hren et al. 2009; Camps et al. 2010; Polesani et al. 2010).

PLEXdb (Plant Expression Database, <http://www.plexdb.org/> Wise et al. 2007) is a unified public resource for gene expression for plants and plant pathogens. PLEXdb serves as a bridge to integrate new and rapidly expanding gene expression profile data sets with traditional structural genomics and phenotypic data. The integrated tools of PLEXdb allow investigators to use commonalities in plant biology for a comparative approach to functional genomics through the use of large-scale expression profiling data sets. In addition to the microarray experiments repository aspects, analysis tools are available, such as gene expression normalization and statistical comparison across experiments, graphical visualization of gene expression. PLEXdb is the largest repository for grapevine microarray experiments: it contains data for published works from Pilati et al. (2007), Deluc et al. (2007), Grimplet et al. (2007), Tattersall et al. (2007), Cramer et al. (2007), Espinoza et al. (2007), Fung et al. (2008), Lund et al. (2008), Albertazzi et al. (2009), Koyama et al. (2009) in addition to secured unpublished works and all experiments available in GEO. Experiments included studies on berry development, tissue expression profiles, abiotic stresses, viral disease and other biotic stresses. PLEXdb also provides a submission service to the GEO repository for submitted experiments.

13.5 Protein or Metabolite Databases

Proteomic and metabolite databases are less well developed than gene expression and genomic databases. We shall give a brief description of these databases. Currently there are no grape-specific databases in use, although these databases are under development in MetNet (see below).

PRIDE (<http://www.ebi.ac.uk/pride/>), which stands for Proteomics Identifications database, is an open source central repository for all proteomic information and data. This database provides a common data exchange format for deposited proteomic data and publications that can be accessed by other programs for analysis. For example, the program SkyPainter (http://www.reactome.org/cgi-bin/skypainter2?DB=gk_current) can be used to identify the statistical overabundance of a protein, its presence in specific pathways and types of reactions in a set of proteins of interest.

KNAPSAcK (<http://kanaya.naist.jp/KNAPSAcK/>) is a comprehensive species-metabolite database. The database provides information on the metabolite, species that the metabolite was found in, and the mass spectrum of the metabolite, which is very useful for the identification of the metabolites. It includes data from NMR, LC-MS and GC-MS technologies.

PRIME: (<http://prime.psc.riken.jp/>) stands for Platform for RIKEN Metabolomics. It is a web-based program for metabolomics and transcriptomics. Metabolites are measured by multi-dimensional NMR spectroscopy, GC-MS, LC-MS, and CE-MS technologies and can be linked with KNApSAcK. Additional tools are available for analysis of metabolites, transcripts, and integrated data sets.

MetNet (<http://www.metnetdb.org/>) contains software and a database that allows the global analysis of transcripts, proteins, and metabolites. One can visualize and model metabolic and regulatory networks. At the moment, the database contains functions for grapevine, *E.coli*, *Arabidopsis* and soybean data.

The UniProt (<http://www.uniprot.org>) database contains 31,254 grapevine protein sequences, mostly putative proteins from the heterozygote genome sequencing project completed by 821 other sequences (July 1, 2010). The UniProt database contains 421 reviewed grapevine proteins with highly curated annotation. UniProt displays automatic GO annotation in addition to domains and protein families automatically identified from a variety of sources (e.g., INTERPRO, Gene3D, Pfam, PIR, PRINTS, SMART, TIGRFAMs, PROSITE, and ProtoNet).

The PlantMetabolomics.org database contains metabolite data for *Arabidopsis* from a consortium of laboratories and contains both fingerprinting and targeted metabolomics data (Bais et al. 2010). Detailed tutorials on sample production are provided along with analysis tools and visualizations of results. The metabolite data can be placed in pathways using metabolite mappings from the AraCyc and MetNet pathway databases.

13.6 Genetic Resources and Phenotypes

The grapevine is a perennial and highly heterozygous species. As such, grapevine genetic resources are conserved in vineyards and distributed as cuttings most of the time, although seeds and pollen exchanges can also occur through scientists. Two databases have been established in Europe aiming first at facilitating the management of the collections and second, at their characterization by collaborative networks (Table 13-1).

The current goal is to improve these databases in order to link genotypic variation to phenotypic variation. This requires a common ontology for the description of developmental stages, phenotypes and traits. In grapevine, there is a long history of international coordination that led to a list of 128 OIV descriptors (OIV 1983) aiming at a precise identification of the accessions. Other descriptors were added later, for traits description (e.g., resistance to diseases); the European database is still under improvement through a collaborative project (<http://www1.montpellier.inra.fr/grapegen06/>). There is

now a need to go further, with the aim to contribute to the international Plant Ontology Consortium (PO; <http://www.plantontology.org/>).

Table 13-1 Public databases for grapevine genetic resources management and characterization.

Country	Network	Web Site	Main Features Publically Available
France	French Network for Grapevine Genetic Resources	http://urgi.versailles.inra.fr/siregal/	-Passport data -core-collections
Europe	European Network for Grapevine Genetic Resources Conservation and Characterization (GENRES #081)	http://www.genres.de/eccdb/vitis/	-Passport data -primary and secondary descriptors

13.7 Integration of Different Data and Systems Biology of the Grapevine

The sequencing of the grapevine genome raised numerous questions about the potential function of all the identified genes. The observed complexity of the genome led grapevine researchers to rethink their approach for studying gene function. The reductionist approach, that has been used for decades or even centuries is limited. Grapevine is a complex living organism with multiple levels of hierarchical control from the individual cell to the tissue, organ and whole plant. Every part must be fully integrated with its environment. Understanding the role of these individual parts and their higher levels of control requires the integration of information in order to understand complex processes such as growth, fruit production and reproduction. Systems biology is the study of how objects interact with each other in a system and can be used for understanding biological phenomena (Mesarovic 1968). In its current use, systems biology refers to the integrative study of the molecular parts of an organism known as transcripts, proteins and metabolites.

Tools to integrate (correlate) metabolite, protein and transcript profiles are seriously underdeveloped, but the availability will be greater in the near future.

For instance, the Tomato Functional Genomics database (<http://ted.bti.cornell.edu/>) integrates the tomato metabolite and transcript data using Pearson correlations to find links between metabolites and transcripts using a newly developed package, Plant MetGenMap (<http://bioinfo.bti.cornell.edu/cgi-bin/MetGenMAP/home.cgi>). Currently the program supports *Arabidopsis*, rice and tomato data, but will support data from other species upon request. A metabolite database is under development at EBI (European Bioinformatics Institute), the institute that has developed the ArrayExpress, UniProt and PRIDE databases (to name a few) with the long-term goal to

integrate these databases with a systems biology approach. A statistical method to analyze or make correlations with two data sets (Bylesjo et al. 2007) was recently extended to three datasets (Bylesjo et al. 2009), which is a first step toward such integrative approaches.

In parallel to these efforts towards an integration of heterogeneous data, several tools have been developed that allow transcript, protein and metabolite “omics” data to be displayed on molecular pathways. Each of these mapping programs presents specific advantages that are summarized in Table 13-2.

Table 13-2 Comparison of mapping programs available for plants. T: Transcripts, P: Proteins, M: Metabolites. N/A: numerical values about pathways composition are not available.

Program	Number of Pathways	Mapped Genes	Mapped Metabolites	Metabolic Pathways	Regulatory Pathways	“Omics Visualization”
<i>Arabidopsis data</i>						
Mapman	64	N/A	N/A	yes	yes	T/M, embedded
PMN/Aracyc	423	5335	2720	yes	no	T or P/M, embedded
REACTOME	454	2646	N/A	yes	yes	T or P/M, via Cytoscape
KaPPA-View	153	2610	1427	yes	no	T/M, embedded
MetNet	334	N/A	N/A	yes	yes	T/P/M, via Cytoscape
<i>Vitis data</i>						
DFCI GI	N/A	N/A	N/A	yes	no	No
KEGG vitis	106	2,265	N/A	yes	yes	No
VitisNet/ MetNet	219	13,145	1,593	yes	yes	T/P/M, via Cytoscape

13.7.1 Mapping Programs for Plant Data

MapMan (<http://gabi.rzpd.de/projects/MapMan/>) displays large data sets onto pictorial diagrams that symbolically depict areas of biological function. The originality of MapMan is that genes are initially organized in blocks rather than as pathways. This allows genes to be tentatively assigned, even when their function is only approximately known and to present pathways not extensively described. Fifty-seven pathways are currently available. Twenty-seven mapping files allow specific microarray platforms to be mapped on the pathways and are downloadable on the MapMan website. The mapping files exist for *Arabidopsis*, rice, potato, barley, maize, wheat, tomato, *Medicago* and grapevine. MapMan also allows the visualization of metabolite abundance.

The Plant Metabolic Network (PMN, <http://plantcyc.org/>) is a collaborative project among databases and biochemists with a common goal to build a broad network of plant metabolic pathway databases. PlantCyc provides access to manually curated or reviewed information about shared and

unique metabolic pathways present in over 340 plant species. PlantCyc-based databases include AraCyc (<http://www.arabidopsis.org/biocyc/index.jsp>) and 10 other plant “Cyc” pathway databases and have been built with the Pathway Tools software. The PlantCyc release 4.0 contains 762 pathways regrouping 11,058 enzymes and 2,966 compounds. AraCyc contains 423 pathways regrouping 5,506 enzymes and 2,719 metabolites. AraCyc includes an expression data visualization tool, Omics Viewer that paints data values from the user’s data sets onto the Cellular Overview diagram. A VitisCyc database has been created from the 12 × sequence data and will be available at the PlantCyc website.

REACTOME (<http://www.reactome.org/>) is probably the most extensive pathway database available, it contains up to 5,000 proteins and 1,000 pathways for some animal species. Pathways for *Arabidopsis* and rice are available, but they have been electronically inferred from human gene annotation. REACTOME allows the exportation of the pathways into the biological pathways standard format such as SBML and BioPAX. REACTOME does not include a molecule abundance viewing tool, but data can be exported into Cytoscape (www.cytoscape.org), where the abundance can be uploaded, visualized and analyzed with plugins provided by a very active user community. These tools include OmicsViz (Xia and Dickerson 2008) and Genoscape (Clement-Ziza et al. 2009).

KaPPA-View4 (<http://kpv.kazusa.or.jp/kpv4/>) allows users to display quantitatively data sets of transcripts and/or metabolites on comprehensive plant metabolic pathway maps. In addition, *Arabidopsis*, rice, tomato and *Lotus japonicus* genes were assigned to the pathway maps. KaPPA-View contains 153 pathways regrouping 2610 genes and 1427 metabolites for *Arabidopsis*.

The MetNet database (MetNetDB; <http://metnetonline.org> Wurtele et al. 2007) contains integrative information on networks of metabolic and regulatory and interactions in *E. coli*, *Arabidopsis*, soybean and grapevine. The grapevine pathways are manually curated and described in Grimplet et al. (2009). The pathway information is based on input from biologists in their area of expertise. In addition to the MetNet-curated interactions, AraCyc-curated pathways and AGRIS-curated regulatory networks are provided. MetNet allows export of pathways to the Cytoscape format for visualization of “omics” data.

13.7.2 Mapping Efforts for Grapevine

The grapevine nucleotide sequences have been mapped onto the KEGG pathways at the DFCI Grape Gene Index and at KEGG itself (<http://www>).

genome.jp/kegg-bin/show_organism?org=dvvi). The DFCI Grape Gene Index presents the mapping of the contigs onto the pathways. Sequences with a GO annotation linked to an E.C. number have been assigned to the pathways where that E.C. number appears. At KEGG, 2265 Genoscope gene sequences have been annotated with KAAS (KEGG Automatic Annotation Server) blast for ortholog assignment and mapped on 106 pathways. This database currently lacks the tools for the visualization of “omics” data.

To date, grapevine data integration has been limited to manual construction of maps and visualization (Cramer et al. 2007; Deluc et al. 2007; Deluc et al. 2009). However new approaches are emerging. The first approach from Zamboni et al. (2008) uses the O2PLS method (Bylesjö et al. 2007) to combine the expression of the genes obtained from microarray technology with the profiles of metabolites obtained from HPLC-MS analysis. This powerful method allows the detection of genes whose expression correlates with metabolite abundances. The second approach, called VitisNet, from Grimplet et al. (2009) is based on the construction of molecular networks that regroup genes, transcripts, proteins and metabolites within biological pathways. These “omics” data can then be uploaded onto the networks and abundance of molecules can be visualized through a user-defined color scheme with Cytoscape. The pathways are available at MetNet (<http://www.metnetonline.org/>) and at <http://vitis-dormancy.sdstate.org>. Two hundred and nineteen pathways are available that incorporate more than 13,000 genes; novel pathways involving hormone signaling, transport, and transcription factors that have not been elucidated in other species yet are included in this mapping effort.

13.8 Conclusions

There has been a rapid explosion of bioinformatics tools in the last few years. The most significant contribution to these efforts has been the sequencing of the grapevine genome. The next major task for grapevine biology is to annotate many of the unknown genes and elucidate their function. This will be done with the large toolset now available to grapevine biologists. Bioinformatics databases are essential for capturing this information and facilitating additional analyses, insights and hypotheses. Grapevine biology has never been more exciting and promises to continue that way for some time to come. It is expected that there will be a rapid acceleration in grapevine research with wide exploration of the many grapevine genetic resources that are available in the many grapevine repositories around the world. One can predict the discovery of many new genes that play important roles in development, flavor, and stress tolerance.

References

- Albertazzi G, Caffagni A, Milc JA, Francia E, Roncaglia E, Ferrari F, Tagliafico E, Stefani E, Pecchioni N (2009) Gene expression in grapevine cultivars in response to Bois Noir phytoplasma infection. *Plant Sci* 176: 792–804.
- Aubourg S, Brunaud V, Bruyère C, Cock M, Cooke R, Cottet A, Couloux A, Déhais P, Deléage G, Duclert A, Echeverria M, Eschbach A, Falconet D, Filippi G, Gaspin C, Geourjon C, Grienberger JM, Houlné G, Jamet E, Lechauve F, Leleu O, Leroy P, Mache R, Meyer C, Nedjari H, Negrutiu I, Orsini V, Peyretailade E, Pommier C, Raes J, Risler JL, Rivière S, Rombauts S, Rouzé P, Schneider M, Schwob P, Small I, Soumayet-Kampetenga G, Stankovski D, Toffano C, Tognolli M, Caboche M, Lecharny A (2005) Genefarm, structural and functional annotation of Arabidopsis gene and protein families by a network of experts. *Nucl Acids Res* 33: D641–D646.
- Bais P, Moon SM, He K, Leitao R, Dreher K, Walk T, Sucaet Y, Barkan L, Wohlgemuth G, Roth MR, Wurtele ES, Dixon P, Fiehn O, Lange BM, Shulaev V, Sumner LW, Welti R, Nikolau BJ, Rhee SY, Dickerson JA (2010) PlantMetabolomics.org: A Web Portal for Plant Metabolomics Experiments. *Plant Physiol* 152: 1807–16.
- Bellin D, Ferrarini A, Chimento A, Kaiser O, Levenkova N, Bouffard P, Delledonne M (2009) Combining next-generation pyrosequencing with microarray for large scale expression analysis in non-model species. *BMC Genom* 10: 555.
- Bylesjo M, Eriksson D, Kusano M, Moritz T, Trygg J (2007) Data integration in plant biology: the O2PLS method for combined modeling of transcript and metabolite data. *Plant J* 52: 1181–1191.
- Bylesjo M, Nilsson R, Srivastava V, Gronlund A, Johansson AI, Jansson S, Karlsson J, Moritz T, Wingsle G, Trygg J (2009) Integrated analysis of transcript, protein and metabolite data to study lignin biosynthesis in hybrid aspen. *J Proteome Res* 8: 199–210.
- Camps C, Kappel C, Lecomte P, Léon C, Gomès E, Coutos-Thévenot P, Delrot S (2010) A transcriptomic study of grapevine (*Vitis vinifera* cv. Cabernet-Sauvignon) interaction with the vascular ascomycete fungus *Eutypa lata*. *J Exp Bot* 61: 1719–37.
- Clément-Ziza M, Malabat C, Weber C, Moszer I, Aittokallio T, Letondal C, Rousseau S (2009) Genoscape: a Cytoscape plug-in to automate the retrieval and integration of gene expression data and molecular networks. *Bioinformatics* 25: 2617–2618.
- Cramer GR, Ergül A, Grimplet J, Tillett RL, Tattersall EA, Bohlman MC, Vincent D, Sonderegger J, Evans J, Osborne C, Quilici D, Schlauch KA, Schooley DA, Cushman JC (2007) Water and salinity stress in grapevines: early and late changes in transcript and metabolite profiles. *Funct Integr Genom* 7: 111–134.
- Deluc LG, Grimplet J, Wheatley MD, Tillett RL, Quilici DR, Osborne C, Schooley DA, Schlauch KA, Cushman JC, Cramer GR (2007) Transcriptomic and metabolite analyses of cabernet sauvignon grape berry development. *BMC Genom* 8: 429.
- Deluc LG, Quilici DR, Decendit A, Grimplet J, Wheatley MD, Schlauch KA, Mérillon JM, Cushman JC, Cramer GR (2009) Water deficit induces cultivar-specific effects in multiple metabolic pathways affecting important flavor and quality traits throughout grape berry ripening. *BMC Genom* 10: 212.
- Doddapaneni H, Lin H, Walker AM, Yao J, Civerolo EL (2008) Vitisexpdb: A database resource for grape functional genomics. *BMC Plant Biol* 8: 23.
- D’Onofrio C, Cox A, Davies C, Boss PK (2009) Induction of secondary metabolism in grape cell cultures by jasmonates. *Funct Plant Biol* 36: 323–338.
- Espinoza C, Vega A, Medina C, Schlauch K, Cramer G, Arce-Johnson P (2007) Gene expression associated with compatible viral diseases in grapevine cultivars. *Funct Integr Genom* 7: 95–110.
- Fernandez L, Torregrosa L, Segura V, Bouquet A, Martinez-Zapater JM (2010) Transposon-induced gene activation as a mechanism generating cluster shape somatic variation in grapevine. *Plant J* 61: 545–57.

- Finn RD, Tate J, Mistry J, Coghill PC, Sammut JS, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A (2008) The Pfam protein families database. *Nucl Acids Res* 36: D281–D288.
- Fung RW, Gonzalo M, Fekete C, Kovacs LG, He Y, Marsh E, McIntyre LM, Schachtman DP, Qiu W (2008) Powdery mildew induces defense-oriented reprogramming of the transcriptome in a susceptible but not in a resistant grapevine. *Plant physiol* 146: 236–249.
- Grimplet J, Deluc LG, Tillett RL, Wheatley MD, Schlauch KA, Cramer GR, Cushman JC (2007) Tissue-specific mRNA expression profiling in grape berry tissues. *BMC Genom* 8: 187.
- Grimplet J, Cramer GR, Dickerson JA, Mathiason K, Van Hemert J, Fennell AY (2009) *Vitis*Net: “Omics” integration through grapevine molecular networks. *PLoS One* 4: e8365.
- Howe KL, Chothia T, Durbin R (2002) GAZE: a generic framework for the integration of gene-prediction data by dynamic programming. *Genome Res* 12: 1418–1427.
- Hren M, Nikolić P, Rotter A, Blejec A, Terrier N, Ravnikar M, Dermastia M, Gruden K (2009) ‘Bois noir’ phytoplasma induces significant reprogramming of the leaf transcriptome in the field grown grapevine. *BMC Genom* 10: 460.
- Iandolino A, Nobuta K, da Silva FG, Cook DR, Meyers BC (2008) Comparative expression profiling in grape (*Vitis vinifera*) berries derived from frequency analysis of ESTs and MPSS signatures. *BMC Plant Biol* 8: 53.
- Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Huguency P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pè ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quétier F, Wincker P (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463–467.
- Koyama K, Sadamatsu K, Goto-Yamamoto N (2009) Abscisic acid stimulated ripening and gene expression in berry skins of the Cabernet Sauvignon grape. *Funct Integr Genom Epub* ahead of print.
- Lund ST, Peng FY, Nayar T, Reid KE, Schlosser J (2008) Gene expression analyses in individual grape (*Vitis vinifera* L.) berries during ripening initiation reveal that pigmentation intensity is a valid indicator of developmental staging within the cluster. *Plant Mol Biol* 68: 301–15.
- Mathiason K, He D, Grimplet J, Venkateswari J, Galbraith DW, Or E, Fennell A (2009) Transcript profiling in *Vitis riparia* during chilling requirement fulfillment reveals coordination of gene expression patterns with optimized bud break. *Funct Integr Genom* 9: 81–96.
- Menda N, Buels RM, Teclé I, Mueller LA (2008) A community-based annotation framework for linking solanaceae genomes with phenomes. *Plant Physiol* 147: 1788–1799.
- Mesarovic MD (1968) Systems theory and biology—view of a theoretician. *Syst Theor Biol* 351: 59–87.
- Mica E, Piccolo V, Delledonne M, Ferrarini A, Pezzotti M, Casati C, Del Fabbro C, Valle G, Policriti A, Morgante M, Pesole G, Pè ME, Horner DS (2009) High throughput approaches reveal splicing of primary microRNA transcripts and tissue specific expression of mature microRNAs in *Vitis vinifera*. *BMC Genom* 10: 558.
- Moroldo M, Paillard S, Marconi R, Fabrice L, Canaguier A, Cruaud C, De Berardinis V, Guichard C, Brunaud V, Le Clainche I, Scalabrin S, Testolin R, Di Gaspero G, Morgante M, Adam-Blondon AF (2008) A physical map of the heterozygous grapevine ‘cabernet sauvignon’ allows mapping candidate genes for disease resistance. *BMC Plant Biol* 8: 66.
- Myles S, Chia J-M, Hurwitz B, Simon C, Zhong GY, Buckler E, Ware D (2010) Rapid Genome Characterization of the Genus *Vitis*. *PLoS One* 5: e8219.
- [OIV] Organisation Internationale de la Vigne et du Vin (1983) Code des caractères descriptifs des variétés et espèces de *Vitis*. OIV, 18 rue d’Aguesseau, Paris, France.

- Ophir R, Pang X, Halaly T, Venkateswari J, Lavee S, Galbraith D, Or E (2009) Gene-expression profiling of grape bud response to two alternative dormancy-release stimuli expose possible links between impaired mitochondrial activity, hypoxia, ethylene-ABA interplay and cell enlargement. *Plant Mol Biol* 71: 403–23.
- Peng FY, Reid KE, Liao N, Schlosser J, Lijavetzky D, Holt R, Martínez Zapater JM, Jones S, Marra M, Bohlmann J, Lund ST (2007) Generation of ESTs in *Vitis vinifera* wine grape (Cabernet Sauvignon) and table grape (Muscat Hamburg) and discovery of new candidate genes with potential roles in berry development. *Gene* 402: 40–50.
- Pilati S, Perazzolli M, Malossini A, Cestaro A, Demattè L, Fontana P, Dal Ri A, Viola R, Velasco R, Moser C (2007) Genome-wide transcriptional analysis of grapevine berry ripening reveals a set of genes similarly modulated during three seasons and the occurrence of an oxidative burst at veraison. *BMC Genom* 8: 428.
- Pindo M, Vezzulli S, Coppola G, Cartwright DA, Zharkikh A, Velasco R, Troggio M (2008) SNP high-throughput screening in grapevine using the SNPLEX™ genotyping system. *BMC Plant Biol* 8: 12.
- Polesani M, Bortesi L, Ferrarini A, Zamboni A, Fasoli M, Zadra C, Lovato A, Pezzotti M, Delledonne M, Polverari A (2010) General and species-specific transcriptional responses to downy mildew infection in a susceptible (*Vitis vinifera*) and a resistant (*V. riparia*) grapevine species. *BMC Genom* 11: 117.
- Rotter A, Hren M, Baebler S, Blejec A, Gruden K (2008) Finding differentially expressed genes in two-channel DNA microarray datasets: how to increase reliability of data preprocessing. *Omics* 12: 171–182.
- Rotter A, Camps C, Lohse M, Kappel C, Pilati S, Hren M, Stitt M, Coutos-Thévenot P, Moser C, Usadel B, Delrot S, Gruden K (2009) Gene expression profiling in susceptible interaction of grapevine with its fungal pathogen *Eutypa lata*: extending MapMan ontology for grapevine. *BMC Plant Biol* 9: 104.
- Samson F, Brunaud V, Duchêne S, De Oliveira Y, Caboche M, Lecharny A, Aubourg S (2004) Flagdb++: a database for the functional analysis of the *Arabidopsis* genome. *Nucl Acids Res* 32: D347–D350.
- Sreekantan L, Mathiason K, Grimplet J, Schlauch K, Dickerson JA, Fennell AY (2010) Differential floral development and gene expression in grapevines during long and short photoperiods suggests a role for floral genes in dormancy transitioning. *Plant Mol Biol* 73: 191–205.
- Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E (2007) The *Arabidopsis* information resource (TAIR): gene structure and function annotation. *Nucl. Acids Res* 36: D1009–D1014.
- Tattersall EA, Grimplet J, DeLuc L, Wheatley MD, Vincent D, Osborne C, Ergül A, Lomen E, Blank RR, Schlauch KA, Cushman JC, Cramer GR (2007) Transcript abundance profiles reveal larger and more complex responses of grapevine to chilling compared to osmotic and salinity stress. *Funct Integr Genom* 7: 317–333.
- Troggio M, Malacarne G, Coppola G, Segala C, Cartwright DA, Pindo M, Stefanini M, Mank R, Moroldo M, Morgante M, Grando MS, Velasco R (2007) A dense single-nucleotide polymorphism-based genetic linkage map of grapevine (*Vitis vinifera* L.) anchoring pinot noir bacterial artificial chromosome contigs. *Genetics* 176: 2637–2650.
- Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J, Malacarne G, Iliev D, Coppola G, Wardell B, Micheletti D, Macalma T, Facci M, Mitchell JT, Perazzolli M, Eldredge G, Gatto P, Oyzerski R, Moretto M, Gutin N, Stefanini M, Chen Y, Segala C, Davenport C, Demattè L, Mraz A, Battilana J, Stormo K, Costa F, Tao Q, Si-Ammour A, Harkins T, Lackey A, Perbost C, Taillon B, Stella A, Solovyyev V, Fawcett JA, Sterck L, Vandepoele K, Grando SM, Toppo S, Moser C, Lanchbury J, Bogden R, Skolnick M, Sgaramella V, Bhatnagar SK, Fontana P, Gutin A, Van de Peer Y, Salamini F, Viola R (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE* 2: e1326.

- Vezzulli S, Micheletti D, Riaz S, Pindo M, Viola R, This P, Walker MA, Troglio M, Velasco R (2008) A SNP transferability survey within the genus *Vitis*. *BMC Plant Biol* 8: 128.
- Wise RP, Caldo RA, Hong L, Shen L, Cannon EK, Dickerson JA (2007) BarleyBase/PLEXdb: A Unified Expression Profiling Database for Plants and Plant Pathogens. In: D Edwards (ed) *Plant Bioinformatics—Methods and Protocols*. *Methods in Molecular Biology*, vol 406 Humana Press, Totowa, NJ, USA, pp 347–363.
- Wurtele ES, Li L, Berleant D, Cook D, Dickerson JA, Ding J, Hofmann H, Lawrence M, Lee EK, Li J, Mentzen W, Miller L, Nikolau BJ, Ransom N, Wang Y (2007) MetNet: Systems Biology Software for *Arabidopsis*. In: BJ Nikolau, ES Wurtele (eds) *Concepts in Plant Metabolomics* Springer Dordrecht, The Netherlands.
- Xia T, Dickerson JA (2008) OmicsViz: Cytoscape plug-in for visualizing omics data across species. *Bioinformatics* 24: 2557–2558.
- Zhang P, Foerster H, Tissier C, Mueller L, Paley S, Karp P, Rhee S (2005) MetaCyc and AraCyc Metabolic Pathway Databases for Plant Research. *Plant Physiol* 138: 27–37.
- Zamboni AL, Minoia L, Ferrarini A, Chimento A, Di Carli M, Desiderio A, Benvenuto E, Guzzo F, Toffali K, Delledonne M, Pezzotti M (2008) Systems biology of berry ripening and postharvest withering processes. 8th Int Symp on Grapevine Physiology and Biotechnology. Adelaide, Australia, 23–28 Nov 2008.