

DOI:10.12113/202106010

网络首发地址: <http://kns.cnki.net/kcms/detail/23.1513.Q.20220613.2021.002.html>

# VitisMod: 一个葡萄基因共表达数据库

刘伟, 魏晓玲, 何华勤\*

(福建农林大学 生命科学学院, 福州 350002)

**摘要:** 葡萄转录组已在各种组织、发育阶段、生物胁迫、非生物胁迫和其他条件下被测定。目前, 仍没有简单实用的网络工具来探索这些宝贵的数据。本文从美国国立生物技术信息中心的基因表达数据库(NCBI GEO)下载 1019 个基因表达芯片数据, 进行权重基因共表达网络分析(WGCNA)。鉴定到 41 个基因共表达模块。功能富集分析发现这些模块具有不同的功能, 并与实验/表型相关。通过模块内连接度筛选枢纽基因, 这些基因可能具有重要功能。通过关联推定(Guilt-by-association)原理对模块内功能未知的基因进行功能预测。最后, 构建了免费的网络工具 VitisMod, 为葡萄的基因功能研究提供新资源, 网址为: <http://bioinformatics.fafu.edu.cn/grape>。

**关键词:** 基因共表达; 枢纽基因; 连接度; 葡萄; 网络工具

**中图分类号:** 文献标志码: A 文章编号: 1672-5565(2023)01-060-08

## VitisMod: A database for co-expressed gene modules of grape

LIU Wei, WEI Xiaoling, HE Huaqin\*

(College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China)

**Abstract:** Grape (*Vitis vinifera* L.) transcriptomes have been extensively profiled in different tissues, at various developmental stages, under biotic or abiotic stresses, and other conditions. However, there is still no user-friendly tool to explore these precious data. A total of 1 019 microarray sample data were downloaded from the National Center for Biotechnology Information Gene Expression Omnibus database (NCBI GEO). Weighted gene correlation network analysis (WGCNA) was conducted. Forty-one gene co-expression modules were identified. Module functional annotation suggested that these modules performed different functions and were associated with experimental conditions/phenotypes. Hub genes, which bear important functions, were screened based on intra-module connectivity. Module genes with unknown functions were annotated according to the “guilt-by-association” rationale. Finally, a web tool VitisMod was developed as a new resource for grape gene function investigation. VitisMod is freely available at <http://bioinformatics.fafu.edu.cn/grape/>.

**Keywords:** Gene co-expression; Hub gene; Connectivity; *Vitis vinifera* L.; Web tool

葡萄(*Vitis vinifera* L.)是一种营养丰富的水果, 也常用于酿酒。基因表达芯片技术十分成熟, 价格低廉, 可以在基因组水平对基因表达进行测量<sup>[1]</sup>。大量葡萄基因表达芯片数据保存于公共数据库, 如美国国立生物技术信息中心(NCBI)的基因表达数据库(GEO)<sup>[2]</sup>。通过生物信息学工具对这些高通量数据重新分析以挖掘基因功能, 并获得基因表达与实验条件之间的关系, 如温度和干旱等非生物胁迫, 白粉病菌等生物胁迫和组织特异性, 可以为葡萄

抗病和育种提供有用的信息。传统的差异基因表达分析只适合小样本数据分析, 而不能有效分析来自不同实验或条件下的大量异质性样本<sup>[3]</sup>。在高通量数据分析中, 降维是一种常用的数据简化方法。显然, 应根据研究目的选择合适的分析技术。基因共表达网络(GCN)分析方法被广泛应用于转录组的全局和时空表达分析。权重基因共表达网络分析(WGCNA)是一种常用的GCN方法, 其已被应用到拟南芥、水稻、玉米和棉花等<sup>[4-7]</sup>。目前关于葡萄基

收稿日期: 2021-06-11; 修回日期: 2022-03-27; 网络首发日期: 2022-06-14.

基金项目: 福建农林大学科技创新专项基金项目(No.CXZX2019050G); 福建省自然科学基金项目(No.2019J01381).

作者简介: 刘伟, 男, 讲师, 硕士生导师, 研究方向: 生物信息学. E-mail: weilau@fafu.edu.cn.

\* 通信作者: 何华勤, 男, 教授, 博士生导师, 研究方向: 生物信息学. E-mail: hehq3@fafu.edu.cn.

因共表达的研究不多。5-氮杂胞苷可促进葡萄浆果发育,研究者对 5 个发育时期 5-氮杂胞苷处理葡萄浆果进行基因共表达分析,鉴定到 8 个和发育时期相关的模块及一些枢纽基因<sup>[8]</sup>。葡萄白腐病是一种重要病害,研究者使用 WGCNA 对葡萄白腐病菌侵染 12 h 和 36 h 后叶片的基因表达进行分析,鉴定得到 4 个模块,其中 1 个模块可分为 50 个基因簇,其中一个基因簇可能和葡萄抗性相关<sup>[9]</sup>。对 7 个不同品系来源的葡萄皮基因共表达分析,共鉴定 64 个模块,和葡萄糖累积、节律等相关,提示这些功能可能在葡萄浆果成熟过程中发挥作用<sup>[10]</sup>。所有这些研究都包括几十个特定条件下的样本,只鉴定和特定条件相关的模块,这可能限制共表达分析的

能力。因此,为鉴定更多的模块,需要更多来自不同研究的样本。组织来源或实验处理可被认为是一种扰动,这是我们整合不同研究数据集的原因。它有助于识别跨实验的共有模块,这不同于样本仅限于特定条件的模块分析<sup>[11]</sup>。我们收集了来自 35 个实验的 1 019 个葡萄转录组样本数据,应用 WGCNA 将数据维度从 16 437 个探针集减少到 41 个模块,并对模块基因进行功能注释。通过关联推定 (Guilty-by-association) 来注释功能未知的基因。模块与性状的关系进行关联分析。最后,建立一个简单实用的网页工具,为用户提供分析得到的模块信息、连接度信息,也提供样本比较等功能。分析框架如图 1 所示。



图 1 VitisMod 数据库构建流程图

Fig. 1 Workflow for VitisMod database construction

## 1 材料与方法

### 1.1 基因表达数据采集与处理

从 NCBI GEO 数据库下载葡萄基因表达芯片数据,平台为 GPL1320,该平台目前具有最多的葡萄转录组数据。该平台使用 Affymetrix *Vitis vinifera* (Grape) Genome 芯片 (<http://www.afymetrix.com>)。根据制造商提供的注释文件 (Release 36),阵列包含 16 437 个探针集,代表 9 192 个 *Entrez* 基因。使用 MAS5 方法在 Expression Console (v1.4.1.46) 中分析

了原始基因芯片数据,得到探针水平的基因表达数据。去除对照探针后,在 R (v3.3.1) 中使用 Normalization.Quantiles 函数进行分位数归一化,并保留所有探针以保证最大的信息。为了方便,文中将探针集称为其对应的基因。

### 1.2 加权基因共表达网络分析 (WGCNA)

使用 Bioconductor WGCNA 软件包 (v1.63) 进行共表达网络分析,参数如下:networkType = 'signed', softPower = 16, minModuleSize = 30, deepSplit = 4, cutHeight = 0.99, MEDissThres = 0.2。根据基因表达矩阵构建符号共表达网络 (Signed Network)。首先计

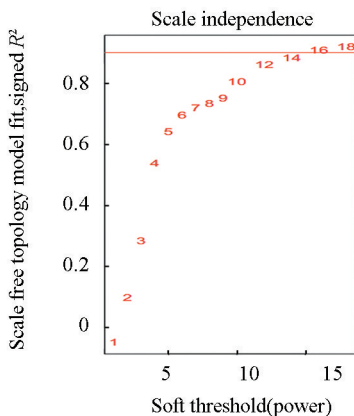
算每对基因间的 Pearson 相关系数, 获得相关系数矩阵; 相关矩阵进行幂指数加权将其转换为邻接矩阵 (Adjacency Matrix)<sup>[12]</sup>; 其次, 幂指数设置为 16, 以满足无标度网络拓扑标准。两个基因之间的关系不仅取决于它们的相关性, 还取决于它们共享的邻域。WGCNA 使用拓扑重叠网络 (Topological Overlap, TO) 来度量基因间关系。以 1-TO 为基因间距离进行分层聚类, 鉴定高度共表达的基因<sup>[12]</sup>; 最后, 使用动态树切割算法鉴定基因共表达模块<sup>[13]</sup>。每个模块通过奇异值分解获得模块特征基因表达 (Module Eigengene, ME), 因此每个模块的表达由其第一个主成分表示。ME 解释了模块表达水平的最大变化量, 被认为是模块中最具代表性的基因表达。设置模块中包含的基因个数最小为 30。考虑到样本量大, 为了鉴定尽可能多的模块, 将划分模块的敏感度 deepSplit 设为最高值 4。为减少冗余模块, 对相似性高于 0.8 的模块进行合并, 因此 MEDissThres 设为 0.2。测试模块稳定性的方法为: 原始连接度与随机采样 1 000 次的一半样本的连接度之间的平均相关性<sup>[14]</sup>。对于每个样本, 每个模块都有 1 个 ME 值。最终原始基因表达数据降维到样本数×模块数。

### 1.3 模块的功能注释

使用 g:Profiler 对共表达模块进行功能富集分析<sup>[15]</sup>。它将基因映射到已知功能信息源, 如基因本体论 (GO) 数据库, 并检测与模块基因集相关联的特定 GO 术语和背景分布是否存在差异, 计算出统计显著性。本研究中, 显著性阈值设定为校正后的  $P$  值小于 0.05。

### 1.4 基于 R Shiny 的数据库构建

VitisMod 数据库构建主要使用 R shiny 包和 shiny server 搭建<sup>[16]</sup>。和该流程相关的代码和文件可在以下链接免费获取: <http://bioinformatics.fafu.edu.cn/shiny-grape.zip>。



(a) 符号网络无尺度网络模型拟合  $R^2$  和软阈值 power 取值的关系

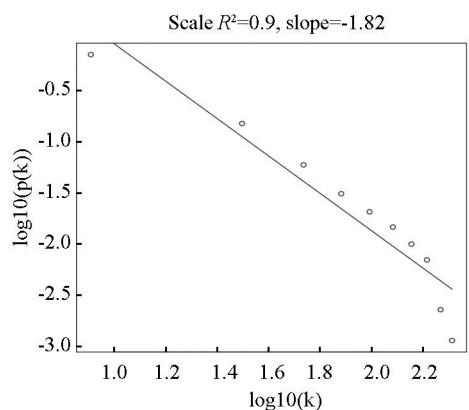
## 2 结果与分析

### 2.1 葡萄基因共表达网络构建

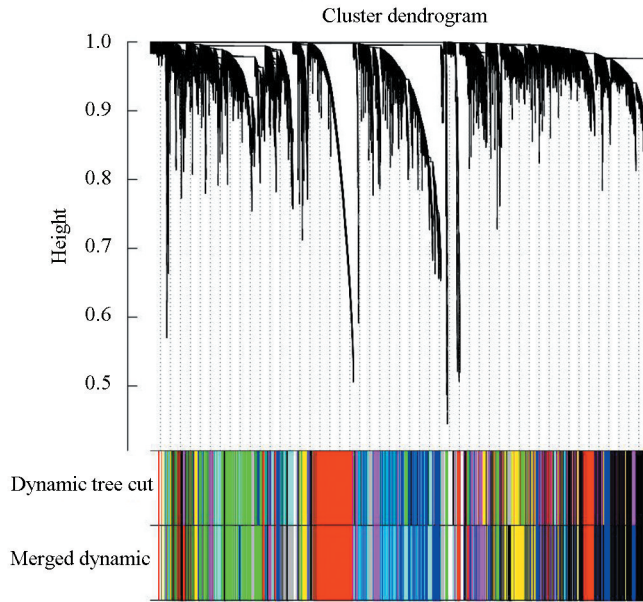
使用 WGCNA 分析 1 019 个葡萄样本数据, power 设置为 16, 以构建无标度 (Scale-free) 的基因共表达网络 (见图 2a, 2b)。通过动态树切割法对网络基因进行聚类和切割 (见图 2c)。为了方便, 每个模块都分配一个数字编号。相似簇合并后形成 41 个基因共表达模块。通过计算原始连接度与从每个模块的 1 000 个半采样连接度值计算出连接度相关性, 来测试模块的稳定性 (见图 2d)。结果表明除了 M25, 所有模块的平均连接度相关性都大于 0.95 (模块命名为 M 加上模块编号, 例如 M4。在 WGCNA 期间会合并一些相似的模块, 因此编号不是连续的)。M4 的模块稳定性最高, 而 M25 的模块稳定性最低。利用 Shiny server<sup>[16]</sup> 建立了一个网页工具来浏览模块基因和模块功能注释信息 (<http://bioinformatics.fafu.edu.cn/grape>)。在“Genelist”选项卡中, 用户可以浏览模块的连接度信息 (见图 3)。

### 2.2 基于连接度的分析

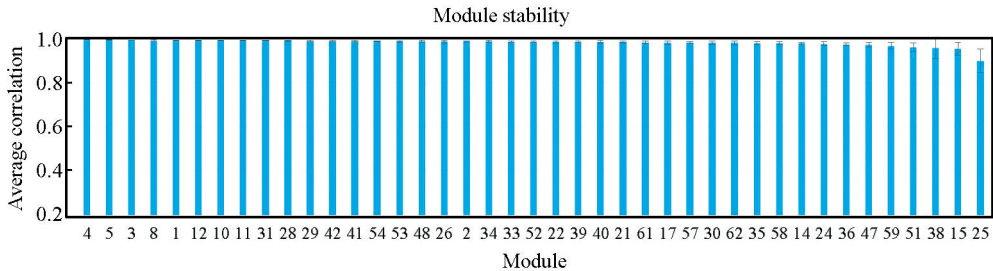
在网络中, 基因的连接度 (Connectivity) 定义为其与所有其他基因的连接之和。枢纽基因 (Hub Gene) 通常在模块中具有高连接度, 因此可能更重要。在 WGCNA 中, 连接度分 2 种类型。全局连接度 (Global Connectivity) 是整个网络内的连接度, 而模块内连接度 (Intramodule Connectivity) 是模块内的连接度。全局连接度中排名前 100 的基因显著富集于叶绿体 (Chloroplast,  $P = 5 \times 10^{-28}$ )。第一名高连接度的基因是 VIT\_00031272001 (Chloroplast stem-loop binding protein of 41 kDa b), 第二名高连接度的基因是 VIT\_00010012001 (Thioredoxin M-type, chloroplastic-like)。



(b) Power 设为 16 时, 葡萄基因共表达网络遵守幂律定律, 线性拟合度满足无尺度网络的要求 ( $R^2 = 0.9$ )



(c) 以平均连接距离对所有基因聚类的树状图, 下方水平条颜色代表基因分配到不同模块



(d) 每个模块进行1 000次随机半采样, 其模块内连接度和原连接度的相关性以平均值±标准差表示并降序排序, 做出柱状图

图 2 从葡萄基因共表达网络中鉴定到 41 个模块

Fig. 2 Forty-one modules were identified from grape gene co-expression network

The screenshot shows the VitisMod web interface. At the top, there is a navigation menu with tabs: "Network summary", "Genelist", "Hub genes", "Annotations", "Network", "Gene function investigation", and "Module signature". Below the menu, there are buttons for "Finding similar experiments", "Compare experiments", and "Help". A search bar is present on the right. The main content is a table with columns: "Probid", "Module", "kTotal", "kWithin", "kOut", "kDiff", "EntrezID", "Symbol", and "Ensembl". The table lists 10 entries for module M42. Below the table, there are "Download Module" and "Download Genelist" buttons. At the bottom, it says "Showing 1 to 10 of 69 entries" and has pagination controls.

Probid	Module	kTotal	kWithin	kOut	kDiff	EntrezID	Symbol	Ensembl
179	1616434_s_at	M42	16.28	3.38	12.89	-9.51	100232862	LOC100232862 VIT_16s0098g00850
218	1607945_at	M42	8.09	2.67	5.42	-2.75	100233090	PEL VIT_14s0219g00230
235	1619703_at	M42	4.54	1.18	3.36	-2.18	100233000	PIP1;1 VIT_13s0067g00220
283	1617739_at	M42	23.04	3	20.04	-17.04	100232906	LOC100232906 VIT_01s0150g00460
312	1614643_at	M42	16.22	5.66	10.56	-4.9	100233087	LOC100233087 VIT_03s0063g00140
614	1619223_s_at	M42	32.66	3.79	28.87	-25.08	100249279	LOC100249279 VIT_11s0016g00470
2616	1619450_s_at	M42	26.71	5.58	21.12	-15.54	100854172	LOC100854172 VIT_16s0098g00850
2748	1610311_at	M42	16	3.1	12.9	-9.79	100243744	LOC100243744 VIT_12s0028g03120
2773	1617667_at	M42	20.59	7.15	13.44	-6.29	---	---
2863	1614836_s_at	M42	29.64	6.45	23.19	-16.74	100243560	METK5 VIT_07s0005g02230

图 3 VitisMod 网络工具截图

Fig. 3 Snapshots of the web tool VitisMod

注:在“Genelist”选项卡中,用户可以浏览选定模块的信息,包括探针集 ID、基因模块间和模块内连接度、Entrez 数据库 ID、基因符号及 Ensembl 数据库 ID。

模块内连接度表示特定模块内基因间的关系。模块间连接度(Intermodule Connectivity)等于全局连接度减去模块内连接度,表明基因在模块间协调功能中的贡献。基于模块的基因连接度分析显示,M2

模块间连接度比例最低,而 M53 模块间连接度比例最高。M2 参与脂质应答。M53 无明确注释,但其枢纽基因的共表达基因富集二甲基烯丙基二磷酸酯代谢。用户可在“Hub gene”选项卡中浏览每个模块的

前3个枢纽基因。例如, M4 主要和质体相关, 其枢纽基因是磷酸甘油酸激酶。

### 2.3 将模块与实验条件/特征相关联

鉴定的模块通过其第一主成分 ME 来概括模块基因的表达。计算每个样品中的每个模块表达, 获得的 ME 向量可用来表征实验条件/特征。可以在“Module signature”选项卡中查看模块高表达的30个样本。例如, 节律性光照处理的葡萄浆果具有最高的 M58 表达, M58 和 DNA 拓扑异构酶有关, 这和已有报道一致<sup>[17]</sup>。缺水12天的葡萄茎尖具有最高的 M42 表达, 功能富集分析提示 M42 和细胞壁合成有关。这些模块可作为特定条件的特征。M35 富集热响应基因, 根据“Module signature”选项卡, 该模块在转色期绿果阶段表达最高, 而在更成熟阶段如红果阶段表达相对下调, 提示热响应基因在葡萄浆果成熟过程发挥重要作用。

在“Find similar experiments”选项卡中提供了一个功能, 通过 ME 向量的 Pearson 相关性来计算不同实验/条件的相似性, 即用户可选择一个感兴趣的实验, VitisMod 返回30个最相似的实验(见图4a)。关于实验的详细信息, 可点击实验的超链接来访问查看。例如, 如果选择 GSM1525182(紫外辐射 C 处理 10 min 后 6 h 采样的叶片), 该工具将返回 GSM1525184 和 GSM1525183, 这是除其自身外的最相似实验。此外, 用户可通过“Compare experiment”选项卡, 选择感兴趣的数据来查看任意多个样本间的相似性(见图4b)。例如, 如果选择样本 GSM207647、GSM207649、GSM207651、GSM207653、GSM207655、GSM207657 和 GSM207659, 这些样本对应转色前浆果热处理 0 h、1 h、4 h、7 h、9 h、14 h 和 24 h, 结果发现相邻时间点间, 4 h 和 7 h 的处理、14 h 和 24 h 的处理更为相似, 均为 0.94, 这表明 7 h 和 24 h 提

供冗余的信息。因此, 4 h 和 14 h 可能是更关键的采样时间点。如果选择三个生物学重复样本(GSM1016867, GSM1016868 和 GSM1016869), 发现这些重复非常相似, 相关性均高于 0.99, 表明实验重复性非常好。

### 2.4 基因功能预测

基因共表达分析可用于基因功能预测<sup>[18]</sup>, 它基于“关联推定”的原理, 即共表达的基因在功能上相关, 将功能分配给同一模块中注释较差的基因。一个例子是 M59, 这是一个只有 39 个基因的小模块, 注释为冰冻应答。M59 枢纽基因包括 VIT\_08s0007g07520、VIT\_07s0205g00010 和一个 21 kDa 蛋白。VIT\_08s0007g07520 是 PPR(Pentatricopeptide repeat) 蛋白家族成员, 在冷胁迫等非生物胁迫下发挥重要作用<sup>[1]</sup>。基因组中, VIT\_07s0205g00010 和 VIT\_08s0007g07520 基因表达相关性最高(Pearson 相关系数  $r=0.9$ ,  $P < 0.01$ ), 但注释为功能未知蛋白质。其在拟南芥的同源基因为 AT5G65660, 是一个富含羟脯氨酸糖蛋白(Hydroxyproline-rich glycoproteins), 具有感应温度变化的功能<sup>[19]</sup>。21 kDa 蛋白没有功能注释, 但其序列含有一个 PME1-like\_3 保守结构域, PEMI 最近被证实和冰冻耐受有关<sup>[20]</sup>。

### 2.5 网络模块可视化

模块通常以网络视图显示, 基因间关系简明。在“Network”选项卡中, 用户可选择感兴趣的模块并可视化高度连接的基因(见图5)。由于可视化过程消耗大量计算机资源, VitisMod 仅绘制了最相关的基因。用户可通过左侧的“Download Module”和“Download Genelist”两个按钮分别下载模块信息和模块的基因列表。所有的模块基因及其连接度信息均可以下载。

Sample_info	Cor_similarity	NCBI_link
leaves treatment: 6 h after the initiation of treatment, leaves age: 30 days old, detached	1.0000	<a href="#">GSM1525182</a>
leaves treatment: 6 h after the initiation of treatment, leaves age: 30 days old, detached	0.9965	<a href="#">GSM1525184</a>
leaves treatment: 6 h after the initiation of treatment, leaves age: 30 days old, detached	0.9944	<a href="#">GSM1525183</a>
leaves treatment: 12 h after the initiation of treatment, leaves age: 30 days old, detached	0.9788	<a href="#">GSM1525185</a>
leaves treatment: 12 h after the initiation of treatment, leaves age: 30 days old, detached	0.9740	<a href="#">GSM1525186</a>
leaves treatment: 12 h after the initiation of treatment, leaves age: 30 days old, detached	0.9188	<a href="#">GSM1525187</a>

(a) “Finding similar experiments”选项卡截图, 返回排名前30最相似的样本



图 4 VitisMod 工具中基于模块表达的 2 项功能截图

Fig.4 Snapshots of two functions of VitisMod tool based on module eigengene (ME) expression

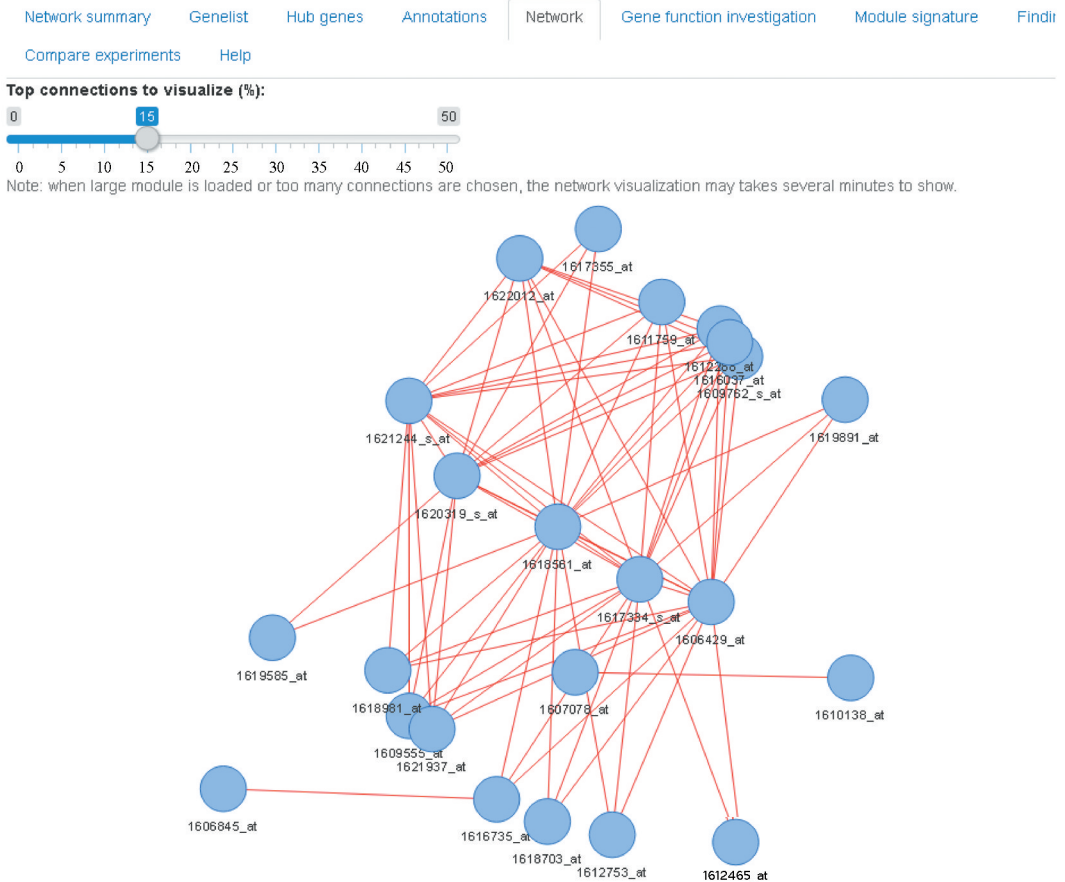


图 5 选项卡“Network”截图，返回不同比例的高连接度基因的共表达网络

Fig.5 Snapshots of tab “Network”. Gene co-expression network visualized by choosing different percentages of high-connectivity genes

### 3 讨论

基因表达芯片是一种成熟的技术,检测成本较低。公共数据库中有大量基于基因表达芯片的转录组数据,这些宝贵的数据可用于知识挖掘。由于多重比较负担和样本量的变化,传统的成对差异基因表达分析不可行<sup>[21]</sup>。尽管 RNA-Seq 比基因表达芯片更具优势,但二代测序的文库类型多样,测序深度不同和仪器类型差异,导致难以合并来自不同研究的数据<sup>[3]</sup>。有研究表明基因表达芯片数据更适合于基因共表达网络分析,数据越多,鉴定模块越多<sup>[22-23]</sup>。另外,特定条件下转录组通常是静态的,如果分析局限于某一条件,鉴定出的共表达模块也仅限于该条件<sup>[4]</sup>。因此,不能期望从一种条件下的样本中识别出许多模块。如果将条件/扰动视为输入,将模块视为输出,则重复或相似的输入可能会导致相同或相似的输出。因此,数据集整合有助于解决此问题,当更多的扰动或条件下的数据添加到分析中时,将鉴定出更多显著性强的模块<sup>[24]</sup>。因此,大样本数有助于提高 WGCNA 的统计效能。有研究整合了 374 个葡萄基因芯片数据,构建基因共表达网络,鉴定得到 557 个模块,但没有提供一个可使用的数据库<sup>[25]</sup>。VTCdb 是一个葡萄基因共表达数据库,它整合了约 800 个样本的基因表达芯片数据,为基因模块鉴定、基因功能注释及候选基因排序等定性分析提供工具<sup>[26]</sup>。本文整合了更多数据集并使用 WGCNA 对不同实验条件下的葡萄转录组进行分析。其次,进行了深入分析以挖掘更为精确的定量信息,这些信息包括基因连接度,模块枢纽基因,条件的特征模块和基因功能预测。最后,提供了一个简单的网络工具,以帮助葡萄研究人员探索这些信息或产生可检验的假设。

通过模块功能富集分析和模块稳定性测试验证了所鉴定模块的有效性。确定了 41 个具有生物学意义的模块。使用者可通过模块功能或高度相关基因的功能来推断基因功能。模块中的基因由于其连接度差异而对模块贡献不同,可从其模块内连接度高低推断基因的重要性。枢纽基因通常是重要的高连接度基因,这些信息有助于选择候选基因进行下一步的实验验证。使用者可在“Network”选项卡中可视化模块网络。

WGCNA 是一种降维方法,有助于减少数据冗余<sup>[12]</sup>。每个样本都可由 41 个模块表示。通过 ME 值对样本进行排名可为特定实验提供潜在的特征模块。特征模块中的枢纽基因是该样本的潜在生物标

志物,值得验证。此外,研究者需要了解样本之间的相似性,相似的样本可能表明相似的表达模式或机制。样本的 ME 向量使样本间比较或发现相似的样本成为可能。

本研究构建了一个简单易用的数据库,可用于葡萄研究中候选基因的筛选或者功能未知基因的功能预测。例如实验获得了差异基因列表,但哪个基因更重要并值得验证是未知的,此时可以在本数据库获取基因的连接度信息,模块内连接度高的基因往往更重要。在 Gene function investigation 选项卡中输入要查询的基因,会获得和该基因表达相关系数高的基因列表及其功能富集信息,同时可以查看该基因位于哪个模块。Module signature 选项卡可查看该模块和哪个实验条件最相关,为实验验证提供重要信息。

### 4 结论

本分析提供了在不同条件下葡萄转录组的系统视图,鉴定到 41 个稳定的又具有不同功能的基因共表达模块,发现高全局连接度基因显著富集于叶绿体、模块内高连接度基因可以用于基因功能预测,挖掘到和实验相关的特征模块,如 M58 和节律性光照有关,M42 和缺水处理有关,M35 和热响应有关。开发了一个简单的葡萄基因共表达网络数据库来帮助研究者探索与模块基因相关的信息,如模块基因连接度、模块基因网络可视化、模块功能富集情况、模块在不同实验基因表达情况和基于模块的样本相似性等,为实验验证产生新的假设。

### 参考文献(References)

- [1] CHEN G, ZOU Y, HU J, et al. Genome-wide analysis of the rice PPR gene family and their expression profiles under different stress treatments [J]. *BMC Genomics*, 2018, 19 (1): 720. DOI: 10.1186/s12864-018-5088-9.
- [2] BARRETT T, WILHITE S E, LEDOUX P, et al. NCBI GEO: archive for functional genomics data sets-update [J]. *Nucleic Acids Research*, 2013, 41: D991-995. DOI: 10.1093/nar/gks1193.
- [3] 刘伟,李立,叶桦,等.权重基因共表达网络分析在生物医学中的应用 [J]. *生物工程学报*, 2017, 33 (11): 1791-1801. DOI: 10.13345/j.cjb.170006.
- LIU Wei, LI Li, YE Hua, et al. Weighted gene co-expression network analysis in biomedicine research [J]. *Chinese Journal of Biotechnology*, 2017, 33 (11): 1791-1801. DOI: 10.13345/j.cjb.170006.
- [4] LIU W, LIN L, ZHANG Z, et al. Gene co-expression net-

- work analysis identifies trait-related modules in *Arabidopsis thaliana* [J]. *Planta*, 2019, 249(5): 1487–1501. DOI: 10.1007/s00425-019-03102-9.
- [5] 杨宇昕, 桑志勤, 许诚, 等. 利用 WGCNA 进行玉米花期基因共表达模块鉴定 [J]. *作物学报*, 2019, 45(2): 161–174. DOI: 10.3724/SP.J.1006.2019.83053.
- YANG Yuxin, SANG Zhiqin, XU Cheng, et al. Identification of maize flowering gene co-expression modules by WGCNA [J]. *Acta Agronomica Sinica*, 2019, 45(2): 161–174. DOI: 10.3724/SP.J.1006.2019.83053.
- [6] 李旭凯, 李任建, 张宝俊. 利用 WGCNA 鉴定非生物胁迫相关基因共表达网络 [J]. *作物学报*, 2019, 045(009): 1349–1364. DOI: 10.3724/SP.J.1006.2019.82061.
- LI Xukai, LI Renjian, ZHANG Baojun. Identification of rice stress-related gene co-expression modules by WGCNA [J]. *Acta Agronomica Sinica*, 2019, 45(9): 1349–1364. DOI: 10.3724/SP.J.1006.2019.82061.
- [7] 巨飞燕, 张思平, 刘绍东, 等. 利用 WGCNA 进行棉花果枝节间伸长相关基因共表达模块鉴定 [J]. *棉花学报*, 2019, 31(5): 403–413. DOI: 10.11963/1002-7807.jfypcy.20190904.
- JU Feiyan, ZHANG Siping, LIU Shaodong, et al. Identification of co-expression modules of genes related to internode elongation of cotton fruiting branches by WGCNA [J]. *Cotton Science*, 2019, 31(5): 403–413. DOI: 10.11963/1002-7807.jfypcy.20190904.
- [8] GUO D L, LI Q, JI X R, et al. Transcriptome profiling of 'Kyoho' grape at different stages of berry development following 5-azaC treatment [J]. *BMC genomics*, 2019, 20(1): 825. DOI: 10.1186/s12864-019-6204-1.
- [9] ZHANG Y, YAO J L, FENG H, et al. Identification of the defense-related gene VdWRKY53 from the wild grapevine *Vitis davidii* using RNA sequencing and ectopic expression analysis in *Arabidopsis* [J]. *Hereditas*, 2019, 156: 14. DOI: 10.1186/s41065-019-0089-5.
- [10] GHAN R, PETEREIT J, TILLET R L, et al. The common transcriptional subnetworks of the grape berry skin in the late stages of ripening [J]. *BMC plant biology*, 2017, 17(1): 94. DOI: 10.1186/s12870-017-1043-1.
- [11] HE F, MASLOV S. Pan- and core-network analysis of co-expression genes in a model plant [J]. *Scientific Reports*, 2016, 6: 38956. DOI: 10.1038/srep38956.
- [12] LANGFELDER P, HORVATH S. WGCNA: An R package for weighted correlation network analysis [J]. *BMC Bioinformatics*, 2008, 9: 559. DOI: 10.1186/1471-2105-9-559.
- [13] LANGFELDER P, ZHANG B, HORVATH S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R [J]. *Bioinformatics*, 2008, 24(5): 719–720. DOI: 10.1093/bioinformatics/btm563.
- [14] YE H, LIU W. Transcriptional networks implicated in human nonalcoholic fatty liver disease [J]. *Molecular Genetics and Genomics*: MGG, 2015, 290(5): 1793–1804. DOI: 10.1007/s00438-015-1037-3.
- [15] REIMAND J, ARAK T, ADLER P, et al. G:Profiler—a web server for functional interpretation of gene lists (2016 update) [J]. *Nucleic Acids Research*, 2016, 44(W1): W83–W89. DOI: 10.1093/nar/gkw199.
- [16] CHANG W, CHENG J, ALLAIRE J J, et al. Shiny: Web application framework for R [Z]. *R Package Version 0.11*, 2015, 1(4): 106.
- [17] KUPRIYANOVA E V, ALBERT E V, BLIZNINA A I, et al. *Arabidopsis* DNA topoisomerase I alpha is required for adaptive response to light and flower development [J]. *Biology Open*, 2017, 6(6): 832–843. DOI: 10.1242/bio.024422.
- [18] VAN DAM S, VOSA U, VAN DER GRAAF A, et al. Gene co-expression analysis for functional classification and gene-disease predictions [J]. *Briefings in Bioinformatics*, 2018, 19(4): 575–592. DOI: 10.1093/bib/bbw139.
- [19] PINSKI A, BETEKHTIN A, SALA K, et al. Hydroxyproline-rich glycoproteins as markers of temperature stress in the leaves of *brachypodium distachyon* [J]. *International Journal of Molecular Sciences*, 2019, 20(10). DOI: 10.3390/ijms20102571.
- [20] CHEN J, CHEN X, ZHANG Q, et al. A cold-induced pectin methyl-esterase inhibitor gene contributes negatively to freezing tolerance but positively to salt tolerance in *Arabidopsis* [J]. *Journal of Plant Physiology*, 2018, 222: 67–78. DOI: 10.1016/j.jplph.2018.01.003.
- [21] DALMASSO C, BROET P, MOREAU T. Multiple comparison procedures: principles, limits. Applications to microarray phenotype-genotype analysis [J]. *Revue D'épidemiologie et de Sante Publique*, 2004, 52(6): 523–537. DOI: 10.1016/s0398-7620(04)99092-x.
- [22] GIORGI F M, DEL FABBRO C, LICAUSI F. Comparative study of RNA-seq- and microarray-derived coexpression networks in *Arabidopsis thaliana* [J]. *Bioinformatics*, 2013, 29(6): 717–724. DOI: 10.1093/bioinformatics/btt053.
- [23] SONG L, LANGFELDER P, HORVATH S. Comparison of co-expression measures: mutual information, correlation, and model based indices [J]. *BMC Bioinformatics*, 2012, 13(1): 328. DOI: 10.1186/1471-2105-13-328.
- [24] SCHAEFER R J, MICHNO J M, MYERS C L. Unraveling gene function in agricultural species using gene co-expression networks [J]. *Biochimica et Biophysica Acta. Gene Regulatory Mechanisms*, 2017, 1860(1): 53–63. DOI: 10.1016/j.bbagr.2016.07.016.
- [25] LIANG Y H, CAI B, CHEN F, et al. Construction and validation of a gene co-expression network in grapevine (*Vitis vinifera*. L.) [J]. *Horticulture Research*, 2014, 1: 14040. DOI: 10.1038/hortres.2014.40.
- [26] WONG D C, SWEETMAN C, DREW D P, et al. VTCdb: a gene co-expression database for the crop species *Vitis vinifera* (grapevine) [J]. *BMC Genomics*, 2013, 14: 882. DOI: 10.1186/1471-2164-14-882.