

植物功能基因网络及其应用

苏晓华, 刘琦, 宁坤, 刘成功

(林木遗传育种国家重点实验室, 国家林业局林木培育重点实验室, 中国林业科学研究院林业研究所, 北京 100091)

摘要:功能基因网络既能够衡量基因之间的功能关联关系,也可以预测基因间的直接相互作用,可为未知功能基因的功能注释提供重要信息,本文简要介绍功能基因网络的概念、功能基因关联挖掘的计算方法和实验方法、功能基因网络的分析方法以及在植物和林木中的应用研究进展。随着林木生物信息学大数据的不断增长,功能基因网络将得到更深入的应用。

关键词:功能基因网络;共表达;系统发生谱;林木;杨树

中图分类号:S792

文献标识码:A

文章编号:1001-1498(2018)01-0094-11

Functional Gene Network and Its Application in Forestry

SU Xiao-hua, LIU Qi, NING Kun, LIU Cheng-gong

(State Key Laboratory of Tree Genetics and Breeding; Key Laboratory of Tree Breeding and Cultivation, State Forestry Administration; Research Institute of Forestry, Chinese Academy of Forestry; Beijing 100091, China)

Abstract: Functional gene network measures the functional association among genes, also predicts the direct interaction among genes, which can provide important information for the functional annotation of unknown functional genes. This review briefly introduces the concept and function of functional gene network, computational method and experimental method for exploring gene association relation, the method for functional gene network analysis, as well as the research and application progress in plant science and forest science. With the increasing of forest bioinformatics data, it is expected that functional gene network will be in-depth applied in research.

Keywords: Function gene network; Co-expression; Phylogenetic profile; Forestry tree; Poplar

拟南芥 (*Arabidopsis thaliana* (L.) Heynh) 全基因组的公布开创了植物基因组学研究的新时代^[1-2],其基因组的成功测序使植物学家开始关注其他植物基因组。由于多数植物具有较大的基因组或较为复杂的倍性,使其基因组的解析面临巨大的挑战。下一代测序技术(NGS)的兴起改变了基因组学的研究规则^[3],越来越多的植物基因组已成功测序,例如近年来已经测序完成的多倍体小麦^[4]、藜麦^[5]、巨桉^[6]和胡杨^[7]等。虽然,越来越多物种的基因组已成功测序公布,但对植物大多数基因的功能仍然知之甚少,例如在拟南芥和水稻中分别只有40%和1%左右的基因是基于实验研究的证据而进

行的注释^[8-9],其他植物实验注释的基因数量相对更少,这阻碍了进一步了解生物学过程中植物是如何发生及演化的问题。

NGS也可用于研究植物不同品种和家系的重测序,可通过全基因组关联分析来鉴定性状关联的基因组位点,例如水稻^[10]、玉米^[11-12]和大豆^[13]。虽然遗传学和基因组学得到了很大发展,但是鉴定复杂性状相关基因和鉴定这些性状潜在的通路仍然较困难,已有的实验方法也只是提供少量的功能和表型注释。系统生物学方法已被用来进行基因-表型的关联研究,关联推定原则^[14-15]已在很多物种中被用来基于基因网络以系统鉴定与特定功能或表型相

关的基因,例如拟南芥和水稻等^[16-17],而这种方式有效性的前提是该物种具备准确完善的功能基因网络,因此,构建一个高质量的功能基因网络是进行关联推定的重要前提。

功能基因网络是指网络中的 2 个基因相互关联以行使相同的功能,即如果 2 个基因处于同一生物学过程或者通路,代表这 2 个基因网络中的 2 个节点在网络中是相互连接的^[18]。功能基因网络中的连接并不一定表示基因产物的直接物理相互作用,但基因产物具有直接物理作用的 2 个基因一定在网络中形成连接,因此,功能基因网络比直接物理相互作用(如蛋白-蛋白相互作用网络和蛋白-DNA 相互作用网络)能更抽象也更广泛地描述生物系统。功能基因网络可以整合不同类型的数据形成一个单独网络,而不是将代表不同分子作用关系的网络简单地叠合在一起;此外,其准确度和覆盖度是通过统一标准去衡量不同类型的数据获得的功能关联是否参与同一过程或者通路来表示的,可对不同类型的数据直接进行比较。功能基因网络的一个重要用途是进行基因功能预测^[19-20],如基因 A 在功能基因网络中与基因 B 相互连接,已知基因 B 参与功能 X,则通过关联推导,基因 A 也参与功能 X。如果基因 A 在网络中与参与功能 X 的基因 C、D 和 E 相互连接,则更能说明基因 A 参与功能 X。

林木大多为多年生木本植物,为人们提供了大量的生物质材料,并由此带来巨大的商业价值;然而,这些植物通常具有很长的生长周期且携带的基因组相对较大,对其直接进行分子生物学实验相对困难,这就更需要通过一些模式树种的深入研究来解析其共有机制,如木材形成与发育机制等^[21]。杨树、云杉等不仅是研究林木的模式植物^[22-24],而且还具有重要的经济价值,如杨树在我国 3/4 的国土面积上均能种植,是现有人工林中适生范围最大,用途最广的林木,已成为我国人造板工业材和纸浆材的主要原料^[25],研究这些木本植物的生长发育和木材的形成等生理过程也有助于了解其它林木植物,对其表型的机理研究也能更加有效地促进优良品种的培育与发掘^[26-28]。虽然杨树、云杉等木本植物的基因组已成功测序^[29],一些与杨树特征相关的基因也被鉴定,但仍然有大量的基因以及与重要特性相关的基因有待进一步研究和注释。随着高通量测序的不断发展,基于 RNA-seq 的转录组研究以及基于重测序的标记关联分析,产生了大量的功能基因组

数据^[30-32],这些数据对了解林木基因型及表型提供了重要信息,在功能基因网络水平上极大地促进了对林木特征进行的系统研究。

1 功能基因网络的构建

1.1 功能基因关联推断的计算方法

除了研究较多的拟南芥、水稻和玉米等植物^[33-35]外,很多植物物种都缺乏其大规模的功能基因网络,挖掘功能基因关联是构建功能基因网络的基础,用来推断基因间的功能关联包含多种计算方法,通过功能关联的整合能够获得更高覆盖度和准确度的功能基因网络。主要计算方法有共表达法、基因组上下文法、同源映射法、文献挖掘法、基因融合法和结构域共发生法 6 种。

1.1.1 共表达法 共表达方法是基于基因的共表达模式,同一生物过程和通路的基因在不同的实验条件下倾向于表现出相类似的表达动态性^[36]。基于高维表达实验数据中表达模式的相似性,可以衡量 2 个基因间功能关联的概率大小。通过共表达方法推断植物的功能基因关联有 2 个明显的优势,首先,共表达模式不仅存在于保守生物学过程的基因中,在物种特异的通路中也能被检测到,因此,共表达方法可作为同源映射法的一个很好补充;其次,公共数据库中已有大量的植物共表达数据可供使用,如 GEO 数据库^[37]、ArrayExpress^[38] 和 PLEdb^[39]。随着高通量转录组测序技术的应用,会有越来越多的共表达数据添加到公共数据库中。

衡量共表达的方式可分为直接相关度量和交互关联秩度量(MR),直接相关度量有皮尔逊相关系数(PCC)和斯皮尔曼相关系数(SCC),它们是衡量 2 个变量相关性的常用方法,但是直接相关度量需要考虑包括表达数据标准化^[40]在内的多个统计学问题。目前,也有一些研究采用交互关联秩度量的方式来构建共表达网络^[41]。Microarray 是通过检测芯片上探针杂交的 cDNA 的量来衡量基因表达,由于存在交叉杂交和非特异性杂交的影响,会带来一定的误差。近些年来,随着高通量 RNA-seq 测序技术的发展^[42],很多物种产生了大量的表达数据,RNA-Seq 与 Microarray 相比,对基因表达的估计通常具有更高的准确性。RNA-Seq 实验条件的维度相对 Microarray 来说,仍十分有限,这也限制了其在共表达网络构建中的大规模应用,但随着测序技术的迅猛发展和价格成本的不断降低,RNA-Seq 将会逐渐替

代 Microarray, 以用于共表达网络推断。

1.1.2 基因组上下文法 在同一通路的基因倾向于具有相同的基因组上下文, 系统发生谱法和基因临近法是 2 种广泛用于网络功能关联推断的基因组上下文的方法, 其在微生物的功能基因网络推断中也受到广泛运用^[43-45]。系统发生谱是指一个基因在其他物种中的直系同源基因发生模式的向量^[46], 在物种分化中, 一个通路中的 2 个相互关联的基因, 由于功能约束常同时出现或者同时保留, 如果在其他物种中观察到 2 个基因的直系同源具有类似的现象, 表明 2 个基因一起行使了相关功能, 并通过系统发生谱的相似性, 去衡量基因间的功能约束; 基因临近法通过衡量 2 个基因在基因组上的临近关系进行功能基因关联推断^[47]。在细菌基因组上, 功能相关的基因通过排布在很接近的基因组位置上进行协同调控, 其中操纵子便是一个典型案例^[48]。原核生物基因的这种临近关系与功能关联的相关性也可以扩展到植物和动物的直系同源基因, 如果一种植物在原核生物中的 2 个直系同源基因在原核生物基因组上趋向于定位在临近位置, 表明这 2 个植物基因也具有功能关联^[49]。目前, 有 2 种方式用来衡量这种基因临近关系, 一种是基因组物理距离, 另外一种是临近概率, 2 种衡量方式已被证明可以相互补充进而用来推断功能关联^[50]。

1.1.3 同源映射法 很多网络是基于拟南芥的功能基因网络进行同源映射而获得, 在进化保守的基因间已越来越多的采用映射关联构建网络。在物种形成时, 一个祖先序列经过复制进而在 2 个物种中分开, 这 2 个直系同源基因通常具有相似的功能。基于该假设, 在 2 个物种中直系同源基因对的功能关联也都被保留。一般来说, 通过 BLAST^[51] 搜索的相似序列可考虑为直系同源基因, 如果需要更严格, 可进行双向最佳匹配 (BBH), 即来源 2 个不同物种的基因互为对方基因组的最佳匹配基因。BBH 虽然更准确, 但是也会遗漏很多近期复制的基因^[52]。InParanoid 作为另外一种检测直系同源的算法^[53], 可以鉴定协同直系同源, 同时获得比 BBH 更多的直系同源; 但 InParanoid 获得的大部分协同直系同源也具有功能分化的旁系同源, 因此, 通过这些协同直系同源映射的功能关联会在网络中带入假阳性; 通过 BBH 和 InParanoid 构建直系同源各有利弊。植物基因网络也可通过直系同源从其他非植物物种进行映射, 酵母和人类的基因网络映射在拟南芥功能基

因网络中, 这对预测植物特异性表型相关基因具有重要贡献^[16]。动植物间大量歧化的表型具有保守的基因网络, 这些不明显的直系同源表型也使得从非植物的同源映射变得可行。考虑到植物中的功能基因网络数据仍然有限, 非植物的功能基因网络可作为植物基因网络推断的重要参考。

1.1.4 文献挖掘法 文献中有大量的蛋白相互作用等基因功能关联信息, 目前文献信息数量巨大且呈指数迅速增长, 传统的人工阅读方式已无法在如此巨大的文献信息中及时、高效地整理并提取基因关联信息, 文献挖掘法已成为文献资源大数据中自动挖掘文献信息的有效途径^[54]。美国国立生物技术信息中心 (NCBI) 在 Pubmed 数据库中收集并维护了数量巨大的生物学相关文献摘要信息, 截止 2016 年底, 已收录超过 2 700 万篇文献, 为文献挖掘法奠定了良好基础。同时也提出了一些基于文献挖掘和围绕该概念的信息提取方法, 发展至今已从生物文献中开发多种挖掘方法用于自动挖掘蛋白相互作用或基因关联的研究中, 归纳起来主要分为 3 类^[54-55]: (1) 基于共出现和规则的方法; (2) 基于计算语言学方法; (3) 基于机器学习方法。共出现和规则方法是最早用来从生物文献句子中识别蛋白质相互作用关系^[56-57], 该方法主要是使用自然语言处理技术 (NLP) 获取相关的信息片段, 然后按照预先定义的模型或算法将其收集和提取^[58], 文献中蛋白相互作用关系的描述具有一定的规则模式, 如基因/蛋白名称在句子中共同出现, 词序列呈现特定的排列且包含特定的描述相互作用的词汇, 如: 互作、结合和激活, 通过定义或抽取规则, 在句子中使用模式匹配提取蛋白质相互作用关系。计算语言学方法是利用词性标注和句法分析等, 基于语法规则创建句法结构以判断蛋白之间是否具有相互作用关系。机器学习方法由于能结合基于规则的方法和自然语言处理方法进而提取文本中累积的信息, 在句子中用来作为有效学习特征进而判断蛋白是否具有相互作用关系^[59]。近年来, 该方法已在基于文献挖掘的蛋白相互作用提取中获得广泛应用, 并取得了良好效果^[60]。

1.1.5 基因融合方法 基因融合方法也叫罗塞塔石碑 (Rosetta stone) 方法, 一个物种中功能关联的基因在另外一个物种中融合成一个基因^[61-63], 这种现象的一个自然解释是如果两个基因具有功能关联, 通过融合成一个基因更有助于它们共同行使作用。

Marcotte 等通过该方法在大肠杆菌和酵母中分别鉴定出 6 809 个和 45 502 个功能关联对,功能关联对的两个基因与其他某些基因组的一个单独基因具有明显的序列相似性。通过与 DIP 数据库中实验获得的蛋白相互作用数据比较,有 6.4% 的蛋白相互作用可以通过基因融合方法获得^[61]。

1.1.6 结构域共发生法 蛋白结构域是一个蛋白的重要功能单元,由于其蛋白的特殊功能和相互作用的角色,具有相似结构域发生谱的两个蛋白很可能也具有相关联的功能^[16]。通过在基因组蛋白序列中注释其结构域组成,如 PFAM 结构域^[64]、Inter-Pro 结构域^[65]等,构建结构域发生谱,通过与系统发生谱相类似的分析方法推断功能基因关联。

1.2 功能基因关联推断的实验方法

1989 年首次应用酵母双杂交技术,经过多次改进后,已开始应用于研究基因编码蛋白的直接相互作用^[66]。蛋白的直接相互作用意味着其编码基因间的功能关联。真核生物的转录因子通常具有至少两个不同的结构域,一个是直接与启动子 DNA 序列结合(BD),另一个是激活转录的结构域(AD),只有两个结构域结合在一起才能启动转录因子进行转录。酵母双杂交使用最多的是酵母 GAL4 转录因子,该转录因子结合 UAS 启动子以启动下游基因的表达。酵母双杂交实验中,诱饵蛋白和捕获蛋白分别与两个结构域融合在一起,两个蛋白的结合使两个结构域一起产生功能性的 GAL4 转录因子,进而激活 UAS 驱动的报告基因的表达^[67-68]。在筛选基因组范围的直接相互作用时,酵母双杂交方法通过两种途径进行改进^[69-70],一种是矩阵方法,即在一个平板上创建一个捕获蛋白矩阵,每个点表示一个特定的捕获蛋白,然后每个诱饵蛋白菌株与一系列捕获蛋白菌株进行配对,基于报告基因的表达和在平板的位置选择蛋白相互作用;另外一种文库方法,即每个诱饵蛋白对含有未定义的随机片段或开放阅读框(ORF)的捕获蛋白文库中进行筛选,根据在特定的平板上是否生长进而筛选出阳性相互作用,相互作用的蛋白由 DNA 测序确定。酵母双杂交技术已在植物蛋白相互作用鉴定中得到广泛应用,利用酵母双杂交技术研究拟南芥所有 MADS 基因蛋白的相互作用,发现相似发育过程中的蛋白聚集在一起^[71]。利用酵母双杂交技术研究水稻非生物胁迫以及种子发芽相关基因的蛋白相互作用,构建由 116 个具有代表性的水稻激酶以及 254 个相互作用

的蛋白构成的网络,并在水稻蛋白激酶的作用中得到应用^[72]。

AP-MS 方法是亲和纯化与质谱技术的结合,在该方法中,一个诱饵蛋白与一个亲和标签,如 His 标签、Flag 标签和 TAP 标签融合以在体内进行表达,然后在细胞裂解物中通过亲和纯析分离出多组分的蛋白复合体,并通过下游的 MS 或 MS/MS 进行分析,将质谱与蛋白质谱数据库进行搜索获得相互作用的候选目标蛋白^[73]。AP-MS 方法具有明显的优势^[73-75],首先,AP-MS 可以真实反映体内生理条件下的相互作用;其次,AP-MS 可以获得不同条件下动态蛋白相互作用的关系;最后,AP-MS 可被用来捕获蛋白复合体。目前,AP-MS 已被成功应用于植物蛋白相互作用的研究中^[76-78]。

双分子荧光互补(BiFC)是基于对无荧光的 GFP 突变体进行重建,以生成具有活性的能发出荧光信号的蛋白复合体的一种功能基因关联推断的实验方法^[79]。在该方法中,诱饵蛋白和目标蛋白分别与互补的 GFP 蛋白部分融合,诱饵蛋白和目标蛋白的相互结合使 GFP 的两个部分进行融合,可通过荧光电镜对其发出的荧光进行观察^[80]。由于植物自身也可以发出荧光,在植物 BiFC 研究中,大多使用黄色荧光蛋白(YFP)以及红色荧光蛋白(RFP)代替绿色荧光蛋白(GFP)^[81-82]。BiFC 可在各种细胞组件中灵敏地检测出微弱的和瞬时的相互作用,并且已被应用于拟南芥和烟草等物种的蛋白复合体的研究^[83-84]。

1.3 整合基因关联构建功能基因网络

通过不同方法和数据来源获得的功能关联通常是互补的,因此,对网络进行整合能提高基因网络的准确度和覆盖度。虽然也有其他的机器学习算法,如决策树、支持向量机^[85],但由于贝叶斯框架可整合多种类型数据来源的功能基因关联,其已成功在拟南芥、水稻、大豆的功能基因关联网络的整合和评估中得以应用^[16-17,86]。如果不同类型的数据是相互独立的,可以将不同来源的功能关联进行简单的汇总(朴素贝叶斯);然而,从不同数据集挖掘功能关联时,都或多或少的利用了相互关联的实验或计算方法,因此,就需要对朴素贝叶斯进行修正以补偿数据集之间的关联。基因关联打分的关键过程是构建一个高质量的功能关联标准集,通过具有相同生物学过程和通路对其功能关联质量进行评估。Gene Ontology 是一种分层组织的注释系统^[87],广泛用于

基因网络的测试。用于基因网络测试的其他系统还包括 KEGG 通路^[88]、MapMan 代谢通路^[89]和 Plant-Cyc 代谢通路^[90],这些注释系统之间的重叠较少,因此,对于增加关联评估效能具有很好的互补性。通过贝叶斯框架整合的网络中,每一条边都可用一个概率得分来表示该功能关联的可行性,即该网络是概率功能基因网络,通过改变概率阈值,还可获得概率更高但覆盖度低或概率更低但覆盖度高的网络。

2 功能基因网络的分析

2.1 功能基因网络的评估

在构建功能基因网络后,需要对其进行评估,最好的评估方式是使用多种独立的实验去验证功能关联;然而,用传统方式验证这些功能关联费时费力。一种替代方式是使用已有的功能关联实验数据,这些实验数据集要求尽可能大且尽可能包括该物种多方面的生物学过程或通路,同时也可用于网络构建或者功能关联打分的过程。网络测试的方式有两种,其中,一种是使用该物种表型已知基因,评估这些基因在网络中相对随机选取的基因是否具有更多的相互连接;另外一种测试方式是使用表型相关已知基因的一个集合(种子基因),评估功能网络能否预测得到其他表型相关的基因。在后一种测试方式中,通常采用受试者特征曲线(ROC)去检测网络的预测能力^[91],首先,对网络中所有的节点(包括种子基因)用与其连接的种子基因的关联概率得分总和对其进行打分,与种子基因具有高关联概率的基因将获得高分,反之,与种子基因只有一个低关联概率的基因获得低分,与种子基因没有任何关联的基因得零分;然后根据得分对所有基因进行排序,使用被预测得到的种子基因的比例(敏感度)对预测得到的非种子基因的比例(1-特异度)构建 ROC 曲线,ROC 曲线中真阳性率作为 Y 轴,假阳性率作为 X 轴,使用 ROC 下曲线面积(AUC)来衡量预测的好坏。如果是一个完美的预测,网络应该返回所有的阳性关联且无假阳性,此时 AUC 等于 1,如果网站质量非常差,预测出真阳性和假阳性具有同等概率,此时 AUC 等于 0.5^[92]。

2.2 功能基因网络的基本性质分析

功能基因网络分析主要包括网络的基本性质分析和网络聚类分析。网络的基本性质分析,包括网络节点的度分布、网络中心性、聚类系数和最短路径长度分析等,基因网络通常具有无标尺属性^[93],即

网络少量节点具有很大的连接度(关联基因),大量节点只有较小的连接度,功能基因网络作为基因网络的一种,也应具有类似的无标尺网络属性。聚类系数是网络中节点聚集程度的衡量,聚类系数越大,表明节点间比随机设立相互连接的概率越大,越趋向于建立一组严密的聚集关系^[94]。功能基因网络的聚类技术主要分为:分层聚类法、划分聚类法和基于密度的方法。常用的聚类方法包括聚类效率高的马尔科夫聚类算法(MCL)以及能发现重叠聚类的 CFinder^[95]。聚类的质量可从紧密度、连接度、空间隔离、预测能力及聚类的稳定性等方面进行衡量。此外,聚类也可以使用其他标准数据集和策略来进行验证,如完整度和聚类之间的相似度。

2.3 功能基因网络的可视化

对已有的基因功能网络进行可视化展示将有助于更直观的认识功能网络,并从中发掘出一些不易从数据中获得的信息,目前已经有大量的软件和在线工具对基因网络进行可视化展示,常用的有 Cytoscape^[96]、Gephi (<https://gephi.org/>)和 Pajek (<http://mrvar.fdv.uni-lj.si/pajek/>)。

Cytoscape 是一个开放源代码的可视化分子网络和生物学路径的软件平台,其最初设计用于生物学研究,现已成为进行复杂网络分析和可视化的通用平台,该软件可把基因网络、基因注释信息、基因表达谱信息和其他类型的数据整合起来。Cytoscape 包含的核心软件以提供数据整合和可视化基本的功能,并可下载专门的插件扩展,如显示布局、编程扩展以及与数据库的连接等其他功能。

Gephi 是 JAVA 编写的一个开源的大型网络可视化平台,支持 Windows、Linux、Mac OS 操作系统,Gephi 的 UI 以可视化为中心,类似图像处理工具 Photoshop,因此有“数据可视化领域的 Photoshop”之称。Gephi 通过内置的 OpenGL 引擎,可适应大型网络需求,甚至能够达到上百万个元素的可视化网络,可实时完成对网络节点的所有操作(布局、过滤、拖拽等),基于 Netbeans 平台构建图形架构,可通过 API 扩展和重用,也可通过插件扩展其他功能。

Pajek 是一个专门为处理大数据集而设计的网络分析和可视化程序,仅支持 Windows 操作系统。Pajek 可同时处理数百万个节点的超大型网络,也可将大型网络分解成几个较小的网络,除普通网络外,Pajek 还支持多关系网络、二分图和动态图。Pajek 提供了多种数据输入方式,可从网络文件(扩展名

NET)中引入 ASCII 格式的网络数据,也可将多列的 NCOL 文件转成 NET 文件进行处理。网络文件中包含节点列表和弧/边列表,只需指定存在的联系即可,以便高效地输入大型网络数据。Pajek 的强项是图形功能,可以方便地调整图形以及指定图形所代表的含义。由于大型网络难以在一个视图中显示,因此,Pajek 会区分不同的网络亚结构分别予以可视化。每种数据类型在 Pajek 中都有自己的描述方法。此外,通过 Pajek 还可在网络中搜索模块、重点节点的邻居,解析和显示属于同一模块的节点,反映节点的连接关系,在模块内收缩节点并显示模块间的关系等。

3 植物功能基因网络的应用

3.1 基因网络数据库

目前,已通过不同方法或者方法组合构建出多个植物物种的功能基因网络,同时基于构建的功能

基因网络进行构建数据库平台。功能基因网络的数据库平台主要分为两种类型,一种是物种特异性的数据库,包括通过同源映射、共表达、基因组上下文等方法的拟南芥功能基因网络数据库 AraNet^[33]、水稻功能基因网络及其数据库 RiceNet^[34]以及大豆功能基因网络 SoyNet^[86]、通过共表达方法构建的大豆功能网络数据库 SFGD^[100]、葡萄功能基因网络数据库 VTCdb^[102]和棉花共表达网络数据库 ccNET^[97],其中,在 SFGD、VTCdb 和 ccNET 中,不仅包含通过共表达构建的功能基因网络,同时也对基因进行了深入注释;另一种是植物的功能基因网络数据库,包括多个物种,如通过共表达方法构建的 PlaNet 数据库^[113]包含拟南芥、大麦、二尾短柄草和苜蓿等多种物种的共表达网络,ATTED-II^[103]含有拟南芥、玉米和葡萄等物种的共表达网络数据,STRING 数据库^[111]中含有多种方法获得的拟南芥和水稻等几种植物的基因功能关联信息。

表 1 植物功能基因网络数据库

Table 1 Databases for plant functional gene networks

网络数据库名称 Name of network database	包含物种 ^① Organisms	功能关联推断方法 ^② Link inference method	参考文献 References
单物种功能基因网络数据库 Databases containing the gene network of single species	AraNet	At	AA, CE, GC [33]
	CCNet	Cc	CE [97]
	PPIM	Zm	CE, GC, TM [35]
	PRIN	Os	AA [98]
	RED	Os	CE [99]
	RiceNet	Os	AA, CE, GC [34]
	SFGD	Gm	CE [100]
	TomatoNet	Le	AA, CE, GC [101]
	PoplarNet	Pt	AA, CE, GC [20]
	VTCdb	Vv	CE [102]
多物种功能基因网络数据库 Databases containing the gene networks of multiple species	ATTED-II	At, Vv, Zm, Me, Pt, Os, Gm	CE [103]
	BMRF	At, Me, Pt, Os, Gm, Le	AA, CE [104]
	CoP	At, Ba, Vv, Zm, Pt, Os, Gm, Ta	CE [105]
	CORNET	At, Zm	AA, CE [106]
	PlaNet	At, Ba, Br, Me, Pt, Os, Gm, Tc, Ta	CE [107]
	PlantExpress	At, Os	CE [108]
	PlantGenIE	At, Pa, Pt	CE [109]
	PLANEX	At, Br, Vv, Zm, Os, Gm, Le, Ta	CE [110]
	STRING	At, Br, Vv, Pt, Os, Sg	AA, CE, GC [111]
	VirtualPlant	At, Mc, Zm, Os, Gm	AA, CE [112]

注:①At: 拟南芥;Ba: 大麦;Br: 二尾短柄草;Mc: 木薯;Cc: 棉花;Vv: 葡萄;Zm: 玉米;Me: 苜蓿;Pa: 云杉;Pt: 杨树;Os: 水稻;Gm: 大豆;Sg: 高粱;Le: 番茄;Tc: 烟草;Ta: 小麦;②AA: 同源映射;CE: 共表达;GC: 基因组上下文;TM: 文献挖掘

Note: ①At: Arabidopsis, Ba: Barley, Br: Brachypodium, Mc: Cassava, Cc: Gossypium, Vv: Grapevine, Zm: Maize, Me: Medicago, Pa: Conifer, Pt: Poplar, Os: Rice, Gm: Soybean, Sg: Sorghum, Le: Tomato, Tc: Tobacco, Ta: Wheat.

3.2 林木功能基因网络和木材形成功能基因网络

Lorenz W 等^[114]使用 cDNA 芯片研究云杉 12 个基因型与环境组合的基因表达模式,使用共表达方式构建功能基因网络,通过网络的拓扑分析获得了

根组织耐旱相关的候选基因。Grönlund 等^[115]基于杨树表达谱芯片数据通过多层网络模型构建共表达网络,分析发现,该网络表现出明显的模块化构成;Cai 等^[116]利用公共数据库的杨树表达芯片数据构

建杨树的共表达基因网络,通过网络聚类分析,发现 6 个模块与植物细胞壁生物合成相关,进一步通过顺式作用元件分析发现,在其中两个模块中鉴定到 10 个潜在的重要顺式作用元件。Kavka 等^[117]研究杨树在磷缺乏条件下营养相关的共表达功能基因网络,通过聚类分析在差异表达基因的子网络中获得 11 个模块,发现其中 1 个模块与磷缺乏响应具有较大的相关性。Dash 等^[118]研究杨树根部在响应低氮条件下,根部发育的共表达基因网络并鉴定与促进根部发育以响应低氮密切相关的网络模块。

Lamara 等^[119]结合关联分析和共表达网络分析,从 1 694 株白杉群体所携带的 2 652 个基因中,鉴定到其含有的 SNPs 与木材密度、硬度、微纤维角度和木材胸径等性状显著关联,在木质部的共表达功能网络中发现了 200 多个显著关联的基因。拟南芥虽然是草本植物,但被证明它可作为研究木材形成发育的优良模式植物^[120]。众多研究表明,在草本植物和林木植物中微管及木材形成的分子机制是非常保守的,大量的转录因子、植物激素以及其他因子在拟南芥和木本植物中都参与了调控木质部的发育^[120]。Taylor-Teeples 等^[121]在拟南芥中采用酵母单杂交技术构建转录因子与次生细胞壁代谢基因之间的蛋白-DNA 调控网络,发现了大量新的基因功能关联。Davin 等^[122]使用拟南芥 *soc1*ful 突变体研究木质部形成中差异表达基因的功能关联网络,发现功能基因网络中的 hub 节点在草本组织和木本组织中具有明显地差异表达。Jokipii-Lukkari 等^[123]以云杉形成层和木质部的多个部位的组织为样本通过 RNA-Seq 测序构建基因表达谱,进而构建共表达功能基因网络,通过与拟南芥和杨树木材发育共表达网络的比较,发现了云杉的某些特有次生细胞壁发育的功能基因关系,同时还构建了 NorWood 网络服务平台可供研究者查询云杉木材形成基因共表达网络信息。Raherison 等^[124]通过 *Piceaglauca* 的芯片表达数据构建共表达网络,以研究 *Piceaglauca* 功能网络的模块化构成以及模块功能和进化特征,并重点鉴定了木材形成发育基因子网络,进而研究了微管组织发育相关基因的表达和功能分化。

4 展 望

生物学过程的调控是多基因通过形成网络来协同参与相关功能,目前对林木的研究大多集中在特定的基因或者有限的基因上,从基因组整体水平构

建的功能基因网络,可为林木未知功能基因的功能注释提供重要信息,同时也可利用已知林木表型特征相关的基因信息挖掘新的候选基因。此外,基于功能基因网络,利用计算机技术构建的生物信息学平台资源,可提供林木基因的功能关联信息和基于功能基因网络的分析功能,并整合相关的生物信息学工具。在平台中,可基于构建的功能基因网络,通过网络的基因关联和已知功能的基因对常规同源搜索方法不能注释的未知功能基因进行功能注释,同时,基于网络指导的搜索策略,可通过与某些功能特征相关的已知基因鉴定新的表型特征关联的基因。此外,还可在整体的功能网络中,分离获得部分与某些生物学过程相关的子网络,然后对网络进行重要功能模块分析,在子网络的层面对该生物学过程进行研究。通过对整体功能基因网络的构建以及基因生物信息学平台的构建,将有助于林木分子水平的相关研究,并为林木遗传育种提供信息和辅助手段。

参 考 文 献:

- [1] Initiative T A G. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana* [J]. *Nature*, 2000, 408 (6814): 796-815.
- [2] Ossowski S, Schneeberger K, Clark R M, *et al.* Sequencing of natural strains of *Arabidopsis thaliana* with short reads [J]. *Genome Res*, 2008, 18(12):2024-2033.
- [3] van Dijk E L, Auger H, Jaszczyszyn Y, *et al.* Ten years of next-generation sequencing technology [J]. *Trends in Genetics*, 2014, 30 (9):418-426.
- [4] Mayer K F, Waugh R, Brown J W, *et al.* A physical, genetic and functional sequence assembly of the barley genome [J]. *Nature*, 2012, 491(7426):711-716.
- [5] Jarvis D E, Ho Y S, Lightfoot D J, *et al.* The genome of *Chenopodium quinoa* [J]. *Nature*, 2017, 542(7641):307-312.
- [6] Myburg A A, Grattapaglia D, Tuskan G A, *et al.* The genome of *Eucalyptus grandis* [J]. *Nature*, 2014, 510(7505):356-362.
- [7] Ma T, Wang J, Zhou G, *et al.* Genomic insights into salt adaptation in a desert poplar [J]. *Nature Communications*, 2013, 4:2797.
- [8] Lamesch P, Berardini T Z, Li D, *et al.* The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools [J]. *Nucleic Acids Res*, 2012, 40(Database issue):D1202-1210.
- [9] Youens-Clark K, Buckler E, Casstevens T, *et al.* Gramene database in 2010: updates and extensions [J]. *Nucleic Acids Res*, 2011, 39 (Database issue):D1085-1094.
- [10] Huang X, Lu T, Han B. Resequencing rice genomes: an emerging new era of rice genomics [J]. *Trends in Genetics*, 2013, 29(4): 225-232.
- [11] Tian F, Bradbury P J, Brown P J, *et al.* Genome-wide association study of leaf architecture in the maize nested association mapping

- population[J]. *Nat Genet*,2011, 43(2):159–162.
- [12] Xiao Y, Liu H, Wu L, *et al.* Genome-wide association studies in maize: praise and stargaze[J]. *Molecular Plant*,2017, 10(3):359–374.
- [13] Zhang J, Singh A, Mueller D S, *et al.* Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean[J]. *Plant J*,2015, 84(6):1124–1136.
- [14] Aravind L. Guilt by association: contextual information in genome analysis[J]. *Genome Res*,2000, 10(8):1074–1077.
- [15] 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*,2017, 1860(1):53–63.
- [16] Lee I, Ambaru B, Thakkar P, *et al.* Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana*[J]. *Nat Biotechnol*,2010, 28(2):149–156.
- [17] Lee I, Seo Y S, Coltrane D, *et al.* Genetic dissection of the biotic stress response using a genome-scale gene network for rice[J]. *Proc Natl Acad Sci USA*,2011, 108(45):18548–18553.
- [18] Lee I, Date S V, Adai A T, *et al.* A probabilistic functional network of yeast genes[J]. *Science*,2004, 306(5701):1555–1558.
- [19] Costello J C, Dalkilic M M, Beason S M, *et al.* Gene networks in *Drosophila melanogaster*: integrating experimental data to predict gene function[J]. *Genome Biol*,2009, 10(9):R97.
- [20] Liu Q, Ding C, Chu Y, *et al.* PoplarGene: poplar gene network and resource for mining functional information for genes from woody plants[J]. *Scientific Reports*,2016, 6:31356.
- [21] Neale D B, Kremer A. Forest tree genomics: growing resources and applications[J]. *Nat Rev Genet*,2011, 12(2):111–122.
- [22] Taylor G. *Populus: Arabidopsis* for forestry. Do we need a model tree? [J]. *Ann Bot*,2002, 90(6):681–689.
- [23] Wullschlegel S D, Tuskan G A, DiFazio S P. Genomics and the tree physiologist[J]. *Tree Physiol*,2002, 22(18):1273–1276.
- [24] 张勇, 张守攻, 齐力旺, 等. 杨树——林木基因组学研究的模式物种[J]. *植物学通报*,2006(3):286–293.
- [25] 苏晓华, 丁昌俊, 马常耕. 我国杨树育种的研究进展及对策[J]. *林业科学研究*,2010, 23(1):31–37.
- [26] 林善枝. 植物学:中国杨树分子遗传改良研究进展[J]. *中国学术期刊文摘*,2007(1):5–5.
- [27] 胥猛, 潘惠新, 张博, 等. 林木遗传改良中的分子生物学研究进展[J]. *林业科学*,2009, 45(1):136–143.
- [28] 苏晓华. 我国杨树育种发展策略[C]. 全国林木遗传育种大会:2008.
- [29] Tuskan G A, Difazio S, Jansson S, *et al.* The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray)[J]. *Science*,2006, 313(5793):1596–1604.
- [30] Fahrenkrog A M, Neves L G, Resende M F, *et al.* Genome-wide association study reveals putative regulators of bioenergy traits in *Populus deltoides*[J]. *New Phytol*,2017, 213(2):799–811.
- [31] Du Q, Gong C, Wang Q, *et al.* Genetic architecture of growth traits in *Populus* revealed by integrated quantitative trait locus (QTL) analysis and association studies[J]. *New Phytol*,2016, 209(3):1067–1082.
- [32] Xie J, Tian J, Du Q, *et al.* Association genetics and transcriptome analysis reveal a gibberellin-responsive pathway involved in regulating photosynthesis[J]. *Journal of Experimental Botany*,2016, 67(11):3325–3338.
- [33] Lee T, Yang S, Kim E, *et al.* AraNet v2: an improved database of co-functional gene networks for the study of *Arabidopsis thaliana* and 27 other nonmodel plant species[J]. *Nucleic Acids Res*,2015, 43(Database issue):D996–1002.
- [34] Lee T, Oh T, Yang S, *et al.* RiceNet v2: an improved network prioritization server for rice genes[J]. *Nucleic Acids Res*,2015, 43(W1):W122–127.
- [35] Zhu G, Wu A, Xu X J, *et al.* PPIM: A protein-protein interaction database for maize[J]. *Plant Physiol*,2016, 170(2):618–626.
- [36] Rhee S Y, Mutwil M. Towards revealing the functions of all genes in plants[J]. *Trends Plant Sci*,2014, 19(4):212–221.
- [37] Barrett T, Wilhite S E, Ledoux P, *et al.* NCBI GEO: archive for functional genomics data sets-update[J]. *Nucleic Acids Res*,2013, 41(Database issue):D991–995.
- [38] Kolesnikov N, Hastings E, Keays M, *et al.* ArrayExpress update-simplifying data submissions[J]. *Nucleic Acids Res*,2015, 43(Database issue):D1113–1116.
- [39] Dash S, Van Hemert J, Hong L, *et al.* PLEXdb: gene expression resources for plants and plant pathogens[J]. *Nucleic Acids Res*,2012, 40(Database issue):D1194–1201.
- [40] Usadel B, Obayashi T, Mutwil M, *et al.* Co-expression tools for plant biology: opportunities for hypothesis generation and caveats[J]. *Plant, Cell & Environment*,2009, 32(12):1633–1651.
- [41] Mutwil M, Usadel B, Schutte M, *et al.* Assembly of an interactive correlation network for the *Arabidopsis* genome using a novel heuristic clustering algorithm[J]. *Plant Physiol*,2010, 152(1):29–43.
- [42] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics[J]. *Nat Rev Genet*,2009, 10(1):57–63.
- [43] 姜春雷, 赵锐, 吕林峰, 等. 茄科雷尔氏菌蛋白质相互作用网络预测及分析[J]. *应用与环境生物学报*,2012, 18(1):139–146.
- [44] Liu Q, Jiang C, Xu Z, *et al.* Analysis of protein interaction network and function of *Staphylococcus aureus*[J]. *Acta Microbiologica Sinica*,2009, 49(1):56–63.
- [45] 孙景春, 徐晋麟, 曹建平, 等. 钩端螺旋体蛋白质相互作用网络预测与系统分析[J]. *科学通报*,2006, 51(9):1049–1057.
- [46] Jiang Z. Protein function predictions based on the phylogenetic profile method[J]. *Critical Reviews in Biotechnology*,2008, 28(4):233–238.
- [47] Korbel J O, Jensen L J, von Mering C, *et al.* Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs[J]. *Nat Biotechnol*,2004, 22(7):911–917.
- [48] Price M N, Huang K H, Alm E J, *et al.* A novel method for accurate operon predictions in all sequenced prokaryotes[J]. *Nucleic Acids Res*,2005, 33(3):880–892.

- [49] Bowers P M, Pellegrini M, Thompson M J, *et al.* Prolinks: a database of protein functional linkages derived from coevolution [J]. *Genome Biol*,2004, 5(5):R35.
- [50] Shin J, Lee T, Kim H, *et al.* Complementarity between distance- and probability-based methods of gene neighbourhood identification for pathway reconstruction [J]. *Molecular BioSystems*,2014, 10(1):24–29.
- [51] Altschul S F, Gish W, Miller W, *et al.* Basic local alignment search tool[J]. *Journal of Molecular Biology*,1990, 215(3):403–410.
- [52] Remm M, Storm C E, Sonnhammer E L. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons [J]. *Journal of Molecular Biology*,2001, 314(5):1041–1052.
- [53] Sonnhammer E L, Ostlund G. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic [J]. *Nucleic Acids Res*,2015, 43(Database issue):D234–239.
- [54] Zhou D, He Y. Extracting interactions between proteins from the literature [J]. *Journal of Biomedical Informatics*,2008, 41(2):393–407.
- [55] Li C, Liakata M, Rebholz-Schuhmann D. Biological network extraction from scientific literature: state of the art and challenges [J]. *Briefings in Bioinformatics*,2014, 15(5):856–877.
- [56] Huang M, Zhu X, Hao Y, *et al.* Discovering patterns to extract protein-protein interactions from full texts [J]. *Bioinformatics*,2004, 20(18):3604–3612.
- [57] Hao Y, Zhu X, Huang M, *et al.* Discovering patterns to extract protein-protein interactions from the literature: Part II [J]. *Bioinformatics*,2005, 21(15):3294–3300.
- [58] Temkin J M, Gilder M R. Extraction of protein interaction information from unstructured text using a context-free grammar [J]. *Bioinformatics*,2003, 19(16):2046–2053.
- [59] Lazzarini N, Widera P, Williamson S, *et al.* Functional networks inference from rule-based machine learning models [J]. *BioData Mining*,2016, 9(1):28.
- [60] Papanikolaou N, Pavlopoulos G A, Theodosiou T, *et al.* Protein-protein interaction predictions using text mining methods [J]. *Methods*,2015, 74:47–53.
- [61] Marcotte E M, Pellegrini M, Ng H L, *et al.* Detecting protein function and protein-protein interactions from genome sequences [J]. *Science*,1999, 285(5428):751–753.
- [62] Enright A J, Iliopoulos I, Kyripides N C, *et al.* Protein interaction maps for complete genomes based on gene fusion events [J]. *Nature*,1999, 402(6757):86–90.
- [63] Marcotte C J, Marcotte E M. Predicting functional linkages from gene fusions with confidence [J]. *Applied Bioinformatics*,2002, 1(2):93–100.
- [64] Finn R D, Coggill P, Eberhardt R Y, *et al.* The Pfam protein families database: towards a more sustainable future [J]. *Nucleic Acids Res*,2016, 44(D1):D279–285.
- [65] Finn R D, Attwood T K, Babbitt P C, *et al.* InterPro in 2017-beyond protein family and domain annotations [J]. *Nucleic Acids Res*,2017, 45(D1):D190–D199.
- [66] Fields S, Song O. A novel genetic system to detect protein-protein interactions [J]. *Nature*,1989, 340(6230):245–246.
- [67] Causer B. Studying the interactome with the yeast two-hybrid system and mass spectrometry [J]. *Mass Spectrometry Reviews*,2004, 23(5):350–367.
- [68] Phizicky E, Bastiaens P I, Zhu H, *et al.* Protein analysis on a proteomic scale [J]. *Nature*,2003, 422(6928):208–215.
- [69] Bartel P L, Roecklein J A, SenGupta D, *et al.* A protein linkage map of *Escherichia coli* bacteriophage T7 [J]. *Nat Genet*,1996, 12(1):72–77.
- [70] Walhout A J, Sordella R, Lu X, *et al.* Protein interaction mapping in *C. elegans* using proteins involved in vulval development [J]. *Science*,2000, 287(5450):116–122.
- [71] de Folter S, Immink R G, Kieffer M, *et al.* Comprehensive interaction map of the *Arabidopsis* MADS Box transcription factors [J]. *Plant Cell*,2005, 17(5):1424–1433.
- [72] Ding X, Richter T, Chen M, *et al.* A rice kinase-protein interaction map [J]. *Plant Physiol*,2009, 149(3):1478–1492.
- [73] Forler D, Kocher T, Rode M, *et al.* An efficient protein complex purification method for functional proteomics in higher eukaryotes [J]. *Nat Biotechnol*,2003, 21(1):89–92.
- [74] Van Leene J, Witters E, Inze D, *et al.* Boosting tandem affinity purification of plant protein complexes [J]. *Trends Plant Sci*,2008, 13(10):517–520.
- [75] Gingras A C, Gstaiger M, Raught B, *et al.* Analysis of protein complexes using mass spectrometry [J]. *Nature Reviews Molecular Cell Biology*,2007, 8(8):645–654.
- [76] Rohila J S, Chen M, Chen S, *et al.* Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice [J]. *Plant J*,2006, 46(1):1–13.
- [77] Rohila J S, Chen M, Cerny R, *et al.* Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants [J]. *Plant J*,2004, 38(1):172–181.
- [78] Schoonheim P J, Veiga H, Pereira Dda C, *et al.* A comprehensive analysis of the 14–3–3 interactome in barley leaves using a complementary proteomics and two-hybrid approach [J]. *Plant Physiol*,2007, 143(2):670–683.
- [79] Magliery T J, Wilson C G, Pan W, *et al.* Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism [J]. *Journal of the American Chemical Society*,2005, 127(1):146–157.
- [80] Miller K E, Kim Y, Huh W K, *et al.* Bimolecular Fluorescence Complementation (BiFC) Analysis: advances and recent applications for genome-wide interaction studies [J]. *Journal of Molecular Biology*,2015, 427(11):2039–2055.
- [81] Jach G, Pesch M, Richter K, *et al.* An improved mRFP1 adds red to bimolecular fluorescence complementation [J]. *Nature Methods*,2006, 3(8):597–600.
- [82] Bracha-Drori K, Shichrur K, Katz A, *et al.* Detection of protein-protein interactions in plants using bimolecular fluorescence complementation [J]. *Plant J*,2004, 40(3):419–427.
- [83] Puts G S, Spadafora N. Detection of protein-protein interactions in

- tobacco BY-2 cells using bimolecular fluorescence complementation [J]. *Methods Mol Biol*,2016, 1342:269 – 277.
- [84] Boruc J, Inze D, Russinova E. A high-throughput bimolecular fluorescence complementation protein-protein interaction screen identifies functional *Arabidopsis* CDKA/B-CYCD4/5 complexes [J]. *Plant Signaling & Behavior*,2010, 5(10):1276 – 1281.
- [85] Ni Y, Aghamirzaie D, Elmarakeby H, *et al.* A machine learning approach to predict gene regulatory networks in seed development in *Arabidopsis*[J]. *Frontiers in Plant Science*,2016, 7:1936.
- [86] Kim E, Hwang S, Lee I. SoyNet: a database of co-functional networks for soybean *Glycine max*[J]. *Nucleic Acids Res*,2017, 45 (D1):D1082 – D1089.
- [87] Consortium T G O. Expansion of the gene ontology knowledgebase and resources [J]. *Nucleic Acids Res*, 2017, 45 (D1): D331 – D338.
- [88] Kanehisa M, Goto S, Sato Y, *et al.* Data, information, knowledge and principle: back to metabolism in KEGG[J]. *Nucleic Acids Res*,2014, 42(Database issue):D199 – 205.
- [89] Usadel B, Poree F, Nagel A, *et al.* A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize [J]. *Plant, Cell & Environment*,2009, 32 (9):1211 – 1229.
- [90] Dreher K. Putting the plant metabolic network pathway databases to work : going offline to gain new capabilities[J]. *Methods Mol Biol*, 2014, 1083:151 – 171.
- [91] Davis J, Goadrich M. The relationship between precision-recall and ROC curves [C]// *Proceedings of the 23rd international conference on Machine learning*; Pittsburgh, Pennsylvania, USA. 1143874; 2006: 233 – 240.
- [92] Pepe M S, Cai T, Longton G. Combining predictors for classification using the area under the receiver operating characteristic curve [J]. *Biometrics*,2006, 62(1):221 – 229.
- [93] Arita M. Scale-freeness and biological networks[J]. *J Biochem*, 2005, 138(1):1 – 4.
- [94] Girvan M, Newman M E. Community structure in social and biological networks[J]. *Proc Natl Acad Sci U S A*,2002, 99(12):7821 – 7826.
- [95] Adamcsek B, Palla G, Farkas I J, *et al.* CFinder: locating cliques and overlapping modules in biological networks[J]. *Bioinformatics*,2006, 22(8):1021 – 1023.
- [96] Shannon P, Markiel A, Ozier O, *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks [J]. *Genome Res*,2003, 13(11):2498 – 2504.
- [97] You Q, Xu W, Zhang K, *et al.* ccNET: Database of co-expression networks with functional modules for diploid and polyploid *Gossypium*[J]. *Nucleic Acids Res*,2017, 45(D1):D1090 – D1099.
- [98] Gu H, Zhu P, Jiao Y, *et al.* PRIN: a predicted rice interactome network[J]. *BMC Bioinformatics*,2011, 12:161.
- [99] Xia L, Zou D, Sang J, *et al.* Rice Expression Database (RED): An integrated RNA-Seq-derived gene expression database for rice [J]. *Journal of Genetics and Genomics*,2017, 44(5):235 – 241.
- [100] Yu J, Zhang Z, Wei J, *et al.* SFGD: a comprehensive platform for mining functional information from soybean transcriptome data and its use in identifying acyl-lipid metabolism pathways [J]. *BMC Genomics*,2014, 15:271.
- [101] Kim H, Kim B S, Shim J E, *et al.* TomatoNet: A Genome-wide co-functional network for unveiling complex traits of tomato, a model crop for fleshy fruits [J]. *Molecular Plant*,2017, 10(4): 652 – 655.
- [102] 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.
- [103] Aoki Y, Okamura Y, Tadaka S, *et al.* ATTED-II: A plant co-expression database towards lineage-specific coexpression[J]. *Plant & Cell Physiology*,2016, 57(1):e5.
- [104] Kourmpetis Y A, van Dijk A D, van Ham R C, *et al.* Genome-wide computational function prediction of *Arabidopsis* proteins by integration of multiple data sources[J]. *Plant Physiol*,2011, 155 (1):271 – 281.
- [105] Ogata Y, Suzuki H, Sakurai N, *et al.* CoP: a database for characterizing co-expressed gene modules with biological information in plants[J]. *Bioinformatics*,2010, 26(9):1267 – 1268.
- [106] De Bodt S, Hollunder J, Nelissen H, *et al.* CORNET 2.0: integrating plant coexpression, protein-protein interactions, regulatory interactions, gene associations and functional annotations[J]. *New Phytol*,2012, 195(3):707 – 720.
- [107] Proost S, Mutwil M. PlaNet: Comparative Co-Expression Network Analyses for Plants [J]. *Methods Mol Biol*, 2017, 1533: 213 – 227.
- [108] Kudo T, Terashima S, Takaki Y, *et al.* PlantExpress: A database integrating oryzaExpress and arthaExpress for single-species and cross-species gene expression network analyses with microarray-based transcriptome data[J]. *Plant & Cell Physiology*,2017, 58 (1):e1.
- [109] Sundell D, Mannapperuma C, Netotea S, *et al.* The plant genome Integrative explorer resource: PlantGenIE.org [J]. *New Phytol*, 2015, 208(4):1149 – 1156.
- [110] Yim W C, Yu Y, Song K, *et al.* PLANEX: the plant co-expression database[J]. *BMC Plant Biology*,2013, 13:83.
- [111] Franceschini A, Szklarczyk D, Frankild S, *et al.* STRING v9.1: protein-protein interaction networks, with increased coverage and integration[J]. *Nucleic Acids Res*,2013, 41(Database issue): D808 – 815.
- [112] Katari M S, Nowicki S D, Aceituno F F, *et al.* VirtualPlant: a software platform to support systems biology research [J]. *Plant Physiol*,2010, 152(2):500 – 515.
- [113] Mutwil M, Klie S, Tohge T, *et al.* PlaNet: combined sequence and expression comparisons across plant networks derived from seven species[J]. *Plant Cell*,2011, 23(3):895 – 910.
- [114] Lorenz W W, Alba R, Yu Y S, *et al.* Microarray analysis and scale-free gene networks identify candidate regulators in drought-stressed roots of loblolly pine (*P. taeda* L.) [J]. *BMC Genomics*,2011, 12:264.
- [115] Gronlund A, Bhalerao R P, Karlsson J. Modular gene expression

- in Poplar: a multilayer network approach[J]. *New Phytol*,2009, 181(2):315–322.
- [116] Cai B, Li CH, Huang J. Systematic identification of cell-wall related genes in *Populus* based on analysis of functional modules in co-expression network[J]. *PLoS One*,2014, 9(4):e95176.
- [117] Kavka M, Polle A. Dissecting nutrient-related co-expression networks in phosphate starved poplars[J]. *PLoS One*,2017, 12(2):e0171958.
- [118] Dash M, Yordanov Y S, Georgieva T, *et al.* A network of genes associated with poplar root development in response to low nitrogen [J]. *Plant Signaling & Behavior*,2016, 11(8):e1214792.
- [119] Lamara M, Raheison E, Lenz P, *et al.* Genetic architecture of wood properties based on association analysis and co-expression networks in white spruce[J]. *New Phytol*,2016, 210(1):240–255.
- [120] Zhang J, Elo A, Helariutta Y. *Arabidopsis* as a model for wood formation[J]. *Current Opinion in Biotechnology*,2011, 22(2): 293–299.
- [121] Taylor-Teeple M, Lin L, de Lucas M, *et al.* An *Arabidopsis* gene regulatory network for secondary cell wall synthesis[J]. *Nature*, 2015, 517(7536):571–575.
- [122] Davin N, Edger P P, Hefer C A, *et al.* Functional network analysis of genes differentially expressed during xylogenesis in social woody *Arabidopsis* plants[J]. *Plant J*,2016, 86(5):376–390.
- [123] Jokipii-Lukkari S, Sundell D, Nilsson O, *et al.* NorWood: a gene expression resource for evo-devo studies of conifer wood development[J]. *New Phytol*,2017, 216(2):482–494.
- [124] Raheison E S, Giguere I, Caron S, *et al.* Modular organization of the white spruce (*Picea glauca*) transcriptome reveals functional organization and evolutionary signatures[J]. *New Phytol*,2015, 207(1):172–187.

(责任编辑:詹春梅)