

甘薯‘徐薯18’转录组分析

颜朗, 魏昌赫, 张义正*

四川大学生命科学学院, 四川省分子生物学及生物技术实验室, 成都610064

摘要: ‘徐薯18’是我国种植时间长和区域广的甘薯品种, 该品种的转录组数据库构建及其重要基因的挖掘, 对于甘薯遗传多样性分析和分子育种都具有重要意义。本文综述了我们近年来对‘徐薯18’转录组的研究进展, 包括利用第2代和第3代测序技术构建‘徐薯18’全长转录组数据库, 从转录组数据库中挖掘重要的功能基因以及分析它们的表达等, 并展望了未来甘薯转录组研究的发展方向。

关键词: 甘薯; ‘徐薯18’; 转录组; 基因挖掘; 基因表达

虽然甘薯(*Ipomoea batatas*)是重要的粮食作物之一, 但是与其他主要作物和模式生物相比, 在过去几十年中, 甘薯的基础研究很少受到重视, 分子生物学资源十分匮乏。其主要原因是甘薯是六倍体植物($2n=6x=90$), 基因组较大($2.2 \times 10^9 \sim 3.0 \times 10^9$ bp), 加之甘薯为无性繁殖植物, 一般具有自交不亲和性, 而且也存在杂交不亲和性, 故其遗传结构十分复杂, 遗传分析很困难。其次, 甘薯作为发展中国家居民的主食, 几乎集中在发展中国家就地生产和就地消费, 相应的科学研究在发达国家未引起足够重视(Varshney等2010)。

近年来, 测序技术和生物信息学的快速发展为甘薯的组学研究带来了希望, 通过高通量的第2代测序技术(next-generation sequencing, NGS), 研究者们可以获得大量遗传信息, 特别是通过转录组测序获得丰富的基因表达信息(Martin和Wang 2011; McGettigan 2013; Mutz等2013)。这些由NGS得到的基因表达谱对于研究甘薯不同品种, 不同组织、器官及各个发育时期, 不同环境条件下的基因表达差异及其调控途径具有重要的意义。此外, 通过高通量的转录组测序所富集的大量基因表达信息, 对于挖掘新基因、分析基因的进化及非编码RNA基因的功能等研究也提供了有力的支撑(陈昊和谭晓风2014; Jiang等2015)。NGS的发展和转录组测序的普及, 将甘薯这种无参考基因组的非模式生物的遗传学研究推向了数字化时代。借助计算机软件和统计学方法的应用, 我们可以同时评价甘薯中成千上万个基因的表达活性, 即每个转录本的表达量, 比通过基因克隆方法研究单个基因功能有了实质性的突破(McGettigan 2013)。

1 甘薯转录组研究概况

目前, 我国甘薯转录组的研究已经取得了较

大进展。2010年, 广东农业科学研究院率先利用Illumina双末端测序(paired-end sequencing)技术对甘薯品种‘广薯87’的块根组织进行了转录组测序, 以挖掘甘薯块根形成与发育相关基因及其表达(Wang等2010)。2012年, 四川大学对我国甘薯代表性栽培品种‘徐薯18’进行了根、茎、叶组织混合转录组测序, 并对该品种的7个组织及生长发育期样品进行表达谱测序, 获得了甘薯不同组织及生长发育时期的基因差异表达和特异表达信息(Tao等2012); 次年发表了该品种在干旱胁迫下开花的转录组研究结果(Tao等2013)。同年, 中国林科院与美国南卡罗来纳州立大学合作, 对紫薯品种‘京薯6号’的块根组织进行了转录组测序, 挖掘到3 553个参与淀粉、生物碱、花色甙色素和维生素生物合成相关基因(Xie等2012)。通过与黄色薯品种的转录组相比较, 该研究发现UDP-葡萄糖-类黄酮-3-O-葡萄糖基转移酶(UDP-glucose-flavonoid-3-O-glucosyltransferase, UFGT)基因是紫薯中花青素生物合成途径中的关键酶基因, 其产物花色苷-3-葡萄糖苷是紫薯中花色甙色素的主要成分。近两年来, 像上述通过两个性状对应的甘薯品种的比较转录组方法, 研究相关代谢通路基因表达调控成为甘薯转录组研究的热点。例如, 2015年, 江苏农业科学研究院利用紫薯品种‘宁薯1号’与其白皮白心的野生种进行比较转录组分析, 深入挖掘了其花青素生物合成途径相关基因及转录因子(Ma等2016)。同年, 中国农业大学对一个具有橙色薯肉的甘薯品种‘维多利’进行了转录组测序, 并

收稿 2017-02-15 修定 2017-05-08

资助 国家科技支撑计划项目(2007BAD78B03)和四川省“十一五”重点攻关项目(07SG111-003-1)。

* 通讯作者(E-mail: yizhang@scu.edu.cn)。

与其具有高类胡萝卜素含量的突变体‘HVB-3’的转录组进行了比较分析,挖掘到22个功能基因和31个转录因子参与了胡萝卜素的生物学合成代谢及富集相关功能(Li等2015)。

国外有关甘薯转录组的第1篇研究论文发表于2010年,国际马铃薯中心利用454焦磷酸测序法对干旱胁迫下的叶和茎进行了转录组测序,成功构建了含有31 685条重叠DNA片段的转录组数据库(Schafleitner等2010)。2013年,以色列植物所通过对甘薯块根早期发育的转录组研究发现,木质素生物合成相关基因表达下调,而淀粉合成基因的表达则上调(Firon等2013)。

除了对不同甘薯品种进行转录组测序和功能基因挖掘外,国内外研究者对于甘薯近缘物种和野生种的转录组分析也表现出浓厚兴趣,希望揭示甘薯近缘种或野生种的优良性状相关基因表达模式和调控途径,以助于甘薯的分子育种工作。例如假厚藤(又名海滩牵牛花, *I. imperati*)是甘薯的近缘物种之一,因其生长在海滩附近而具有超强的盐适应性。Solis等(2016)通过比较转录组方法挖掘该植物中抗盐相关基因及其代谢途径,发现除了WRKY转录因子、阳离子:阳离子逆向转运体(cation:cation antiporter)以及乙烯应答因子(ethylene responsive factor)等在甘薯中同样存在的抗盐相关基因外,在假厚藤中还发现了一些甘薯中没有的新基因表达,如DREB基因家族等。Leslie和Baucom (2014)对另一种甘薯近缘种圆叶牵牛(*I. purpurea*)的转录组进行测序,揭示了具有除草剂甘膦抗性和除草剂敏感型圆叶牵牛间与除草剂抗性相关基因的表达差异。2016年,被认为是甘薯直接起源的野生种三浅裂野牵牛(*I. trifida*)的转录组完成了测序,并通过与甘薯栽培种间的比较转录组分析,挖掘了2 848个包含bZIP、MYB、bHLH、Orphans、ERF、C2H2、NAC、WRKY等在内的51个转录因子家族,这些转录因子对于植物应对干旱、盐碱等逆境胁迫具有关键的作用(Cao等2016)。该物种的遗传信息对于甘薯的分子育种非常重要,不仅有利于构建甘薯的转基因体系,而且基于该野生种中功能基因特别是抗性基因的挖掘,有利于阐述甘薯在长期驯化和染色体加倍过程中基因的丢失与变异。

在甘薯中,‘徐薯18’是在我国种植时间长、推广面积大、性状稳定的甘薯品种,对于今后开展甘薯中重要农艺性状相关功能基因的研究具有重要意义。本文综述了我们对‘徐薯18’转录组的研究工作进展,包括转录组数据库的建立和重要基因的挖掘。

2 ‘徐薯18’转录组分析

2.1 甘薯转录组组装策略

在获得了大量的第2代测序的转录组数据后,如何将短序列的reads片段进行组装,将直接关系到转录组数据库的质量及其后续分析的可靠性。我们在对‘徐薯18’根、茎、叶及其不同发育阶段的混合RNA样品所测得的reads片段进行组装时,通过对不同组装策略的组装结果比较发现,利用组合式从头组装新策略的效果较好,即利用不同的组装软件和不同的参数进行组装,然后根据contig(重叠群)的平均长度、N50长度及最长contig长度等评价指标选择最优contig库进行再次组装(Tao等2012)。为了证明组合式从头组装策略(combined *de novo* transcriptome assembly, CDTA)比单软件单参数组装策略(single-assembler single-parameter, SASP)和单软件多参数组装策略(single-assembler multiple-parameter, SAMP)更适合于六倍体甘薯转录组的组装,我们将六倍体甘薯、二倍体水稻以及二倍体真菌(粗毛栓菌)的转录组测序数据,用3种不同的组装软件和组装策略分别进行组装,然后对组装质量进行比较。结果显示,用contig长度分布、准确性和完整度、长开放阅读框(open reading frame, ORF)数量等3个指标评价,组合式从头组装策略在甘薯的转录组组装中表现最优,适合多倍体物种的转录组组装,而基于Oases软件的单软件多参数组装策略比较适合水稻和粗毛栓菌等二倍体物种。将这3种策略应用到二倍体玉米和四倍体小麦转录组组装进行验证,所得结果与上述结果一致,即转录组组装策略与生物的倍性相关(He等2015)。

2.2 ‘徐薯18’转录组数据库的构建

利用组合式从头组装策略,我们成功地构建了‘徐薯18’的首个根、茎、叶混合转录组数据库。该数据库共有70 412条长度大于200 bp的转录本,总长度为44.2 Mb,其中长度大于1 000和2 000

bp的转录本分别有12 434和2 407条;具有ORF且长度大于600、900、1 200和1 500 bp的转录本分别为14 398、8 115、4 572和2 494条;以参考序列法对组装结果进行质量评估显示,该转录组的完整度和准确度分别为0.93和0.73;将所有reads进行映射分析发现,约有84%的reads能成功映射到转录本上。此外,我们还分别构建了7个cDNA标签文库,利用Illumina公司的数字基因表达谱(digital gene expression tag profiling, DGE)测序技术进行了高通量测序,通过分析获得了‘徐薯18’不同组织和发育阶段的基因表达(Tao等2012)。

为了进一步提高‘徐薯18’转录组数据库的质量,建立能够覆盖甘薯所有器官的‘徐薯18’全长转录本数据库,我们还采用了第3代测序技术,对‘徐薯18’的根、茎、叶、花以及嫁接的甘薯茎、叶、花共7个样品进行混合全长转录本测序(Iso-seq)。在获得的544 277条长reads中,共发现192 517条全长非嵌合型reads (FLNC reads),全长转录本序列占总转录本的35.4%,高于甜菜和丹参全长转录组中全长转录本的比例(27.8%和28.1%)(Minoche等2015; Xu等2015)。从Iso-seq的测序转录本长度看,全长转录本的平均长度(2 396 bp)、最大长度(19 457 bp)以及N50长度(2 766 bp)已经达到甚至优于第2代测序的组装结果。也就是说,基于Iso-seq的全长非嵌合型转录本的长度更长,更接近于物种本身转录后的转录本全长,且其连续性更好(结果未发表)。

在甘薯全长转录组数据库中首次发现3 255个新转录本,这些新转录本有的可能是甘薯中基因转录加工过程中产生的可变剪接体。从率先开展全长转录组测序且具有参考基因组的玉米、高粱中发现的新转录本结果可以推测,甘薯中这些新发现的转录本也有可能来自尚未发现的新基因转录本,如高粱中通过第3代全长转录组测序新发现11 000多个新转录本以及2 100多个新基因(Abdel-Ghany等2016; Wang等2016)

为了验证‘徐薯18’转录组中组装序列的可靠性,我们对多个基因编码序列进行了克隆和测序,结果显示,实测序列和组装序列基本上是一致的,而且每个基因都有3个同源性很高的转录本(Cao等2012; Chang等2012; Lai等2014; Shao等2011; Wang等2015; Zhou等2016)。

3 甘薯‘徐薯18’转录组中重要功能基因挖掘

基于‘徐薯18’根、茎、叶混合转录组和花转录组数据库,我们获得了相对比较丰富的基因表达信息,如淀粉合成、抗逆、病毒、转座子以及开花等相关基因的表达丰度。

3.1 淀粉和蔗糖代谢途径相关基因

淀粉和糖是甘薯各种用途的主要物质来源。在‘徐薯18’混合转录组中,共挖掘到922条转录本参与甘薯淀粉和蔗糖代谢途径,编码淀粉合成酶(starch synthase, SS)、蔗糖合成酶(sucrose synthase, SuSy)、 α -淀粉酶和 β -淀粉酶等39种酶,且这些蛋白编码基因的表达水平普遍较高,如催化淀粉生物合成的限速酶1-磷酸葡萄糖腺苷转移酶(glucose-1-phosphate adenylyltransferase),也称为ADP-葡萄糖焦磷酸化酶(ADP-glucose pyrophosphorylase, AGPase)以及具有淀粉合成与降解双重功能的淀粉磷酸酶(starch phosphorylase, SP)等的基因都在甘薯中高表达,它们的高表达对甘薯淀粉的积累有重要作用(Tao等2012)。

通过甘薯块根不同生长发育阶段的基因差异表达分析后发现,UDP-葡萄糖焦磷酸化酶(UDP-glucose pyrophosphorylase, UGPase)编码基因在根组织中的表达量明显上调,尤其是从须根到块根再到块根膨大的过程中,该基因的表达量呈大幅度上升,而在块根收获时,其表达量则与须根中相当。参与淀粉合成过程中的AGPase和淀粉合成酶基因,在根中具有绝对的表达优势,尤其是在初始块根和收获期块根中表达量最高。在块根膨大时期,AGPase和SS基因的表达水平却要低得多,但UGPase基因的表达量却达到了最高点。也就是说,该研究发现甘薯中淀粉的合成主要是在初期和收获期块根中,而在块根膨大时期淀粉的合成相对减缓,而大量的Glc-1-P则可能用于块根膨大所需的其他碳水化合物的合成(Tao等2012)。

3.2 抗逆性相关基因

‘徐薯18’在我国长时间、大面积的种植过程中,表现出很强的环境适应性和非生物/生物胁迫抗性,通过转录组研究也发现下述的大多数抗逆相关基因的表达丰度都很高:例如编码液泡膜质子转运无机焦磷酸酶(vacuolar-type H^+ -translocating inorganic pyrophosphatase, V-PPase)、热激蛋白

(heat shock protein, HSP)、胚胎发育晚期丰富蛋白(late embryogenesis abundant protein, LEA)、类金属硫蛋白(metallothionein-like protein, MTLP)、超氧化物歧化酶(superoxide dismutase, SOD)、过氧化氢酶(catalase, CAT)以及贮藏蛋白(sporamin)等基因。因此, 抗逆相关基因的高表达可能是‘徐薯18’具有很强适应性的原因之一。

LEA是植物在逆境胁迫环境下, 特别是干旱和盐胁迫条件下诱发表达的一种应激蛋白, 是重要的植物抗逆功能蛋白。通过‘徐薯18’数字基因表达谱和半定量RT-PCT分析, 均确定了*Ib-LEA2*基因在甘薯不同组织和发育阶段的表达图谱; 当把该基因克隆到大肠杆菌中进行盐胁迫耐受性分析时发现, 较之对照菌株, 表达*Ib-LEA2*的菌株能够更好地耐受高浓度的NaCl (陈娇等2014)。在‘徐薯18’转录组中还挖掘到另一种涉及多种胁迫抗性的非特异脂质转运蛋白(nonspecific lipid transportor, nsLTP)。通过qRT-PCR分析发现, *nsLTP1*和*nsLTP2*基因在天然生长和盐胁迫条件下具有明显的表达差异, 同时通过重组大肠杆菌进行耐盐性分析, 证实了nsLTP在甘薯盐胁迫反应中的作用(Li等2016)。在盐胁迫条件下, 甘薯甜菜碱醛脱氢酶(betaine aldehyde dehydrogenase, BADH)基因*IbBADH*具有相似的表达规律(Chen等2014)。

甘薯贮藏蛋白是甘薯块根中主要的贮藏蛋白, 占甘薯块根可溶性蛋白的60%~80%。它除了为萌发的甘薯幼苗提供氮源和碳源外, 还是一种蛋白酶抑制剂(Senthilkumar和Yeh 2012)。甘薯中的贮藏蛋白编码基因以家族形式存在, 不同亚家族中具有多种不同的编码基因, 使甘薯贮藏蛋白基因的分离和鉴定工作比较困难, 因此至今缺乏全面的研究。然而, 通过‘徐薯18’转录组数据库的建立和贮藏蛋白编码基因的挖掘, 获得了所有的贮藏蛋白基因contig序列, 据此设计相应的引物, 成功克隆到6个*sporamin A*亚家族基因和10个*sporamin B*亚家族基因(赖先军等2013)。

3.3 ‘徐薯18’的病毒相关基因

病毒感染会直接影响甘薯产量和品质。甘薯转录组测序为病毒检测提供了新的手段, 它能发现所有的甘薯病毒种类在不同组织中的表达分布, 包括未知的新病毒, 并且能获得甘薯在自然栽培条件下的病毒基因在不同组织中的表达水平。

我们分别从‘徐薯18’营养器官和花器官转录组中挖掘病毒基因的转录本, 以调查甘薯在自然栽培条件下的病毒种类。从‘徐薯18’的营养器官转录组数据库中, 共挖掘到12种不同的病毒, 其中2种为杆状DNA病毒属(*Badnavirus*)的sweet potato badnavirus A (SPBV-A)和sweet potato badnavirus B (SPBV-B), 其余10种均为RNA病毒, 分别为马铃薯Y病毒属(*Potyvirus*)的sweet potato feathery mottle virus (SPFMV)、sweet potato virus G (SPVG)、sweet potato latent virus (SwPLV)、*Ipomoea vein mosaic virus* (IVMV)、sweet potato virus B2 (SPVB2)、yam mosaic virus (YMV)、turnip mosaic virus (TuMV)和sunflower mosaic virus (SuMV), 胞质弹状病毒属(*Cytorhabdovirus*)的northern cereal mosaic virus (NCMV)以及蚕豆病毒属(*Fabavirus*)的*Mikania micrantha* mosaic virus (MMMV)。在这些病毒中, SPFMV和SPVG是reads数量最多、组装的核苷酸序列最长且表达水平最高的两种病毒, 表明‘徐薯18’营养器官中受到这2种病毒的侵染最为严重。在‘徐薯18’花器官转录组数据库中, 仅挖掘到SPFMV、SPVG、sweet potato leaf curl virus (SPLCV)和*Cymbidium* mosaic virus (CymMV)等4种病毒相关基因, 且表达量均较低。值得注意的是, 从‘徐薯18’转录组中挖掘到的病毒基因种类, 除了SPFMV、SPVG、SwPLV和SPVB2外, 其余所有病毒种类在国内甘薯中都是首次发现。同时, 基于‘徐薯18’营养器官数字基因表达谱的结果, 对不同甘薯病毒在不同器官和不同发育时期的差异表达进行了分析, 并通过RT-PCR进行验证。结果发现, 不同的病毒基因间以及同一基因在不同器官或发育阶段的表达水平具有很大差异, 即须根中的表达量最高, 其次是块根和茎, 而在嫩叶和成熟叶中病毒基因的表达量几乎检测不到。也就是说, 甘薯病毒基因主要活跃在甘薯须根和块根中, 通过影响其他基因的表达, 从而影响甘薯块根的形成和发育, 进而造成甘薯产量和品质的降低(Gu等2014)。该研究结果对于甘薯的病毒检测时取样具有重要的指导意义。

3.4 ‘徐薯18’的转座子基因家族

转座元件可以通过自身的复制和移动, 对基因组结构和功能产生重要影响(Biémont 2010; Lisch 2013)。由于甘薯主要是通过无性繁殖方式将遗传

物质传递给后代, 缺少有性繁殖时配子间基因重组和遗传物质交流的机会, 因此, 转座子可能在甘薯为了适应自然环境进行基因组重组中起到更加重要的作用。

在‘徐薯18’转录组数据库中挖掘到1 405个转座酶转录本, 包括883个反转录转座子和552个DNA转座子, 并将其归类到Ty1/Copia、Ty3/Gypsy、Tc1/Mariner、hAT、Mutator、PIF-harbinger、CACTA和Helitron等超家族中(Yan等2014)。另外, 将‘徐薯18’和其他甘薯品种中挖掘的转座子进行差异表达分析后发现了411个在‘徐薯18’中特异性表达的转座子; 通过‘徐薯18’营养器官和花器官的转录组比对后发现, 各有374个和292个转座子分别在上述两个器官中特异性表达; 通过‘徐薯18’中建立的7个组织的数字基因表达谱, 发现101个转座子只在某一个器官或发育阶段特异性表达。在这些‘徐薯18’特异性或表达活性较高的转座元件中, Helitron超家族转座子在数量上占据较大优势, 其次是hAT超家族, 而反转录转座元件中的Ty1/Copia超家族元件也在‘徐薯18’中大量表达, 这是区别于其他甘薯品种和其他物种的主要转座子表达特征。在甘薯转录组数据库中挖掘到的转座子基因信息, 填补了一些公认的重复元件数据库(如PlantTribes和Rebase)中没有任何甘薯转座元件信息的空白。这种搜寻和鉴定转座元件的新方法, 适用于如甘薯这样的未知基因组的非模式生物, 相较于通过基因克隆逐个获得转座元件序列, 然后进行转录和翻译研究的传统方法具有明显的优势。

3.5 ‘徐薯18’开花相关基因的表达

我们利用干旱胁迫和嫁接两种方式促使‘徐薯18’开花, 对花器官进行转录组测序, 然后挖掘甘薯花器官特异性表达基因, 分析甘薯在不同胁迫条件下开花的分子调控机理。

比较转录组分析结果显示, 有2 959个花器官及2 928个营养器官特异性转录本。在花器官中, 特异性表达的基因主要富集在角质、软木质和蜡质生物合成, 戊糖和葡萄糖醛酸酯互变, 氧化磷酸化及氮代谢, 而营养器官中的基因主要富集在石黄春生物合成、类固醇激素生物合成等代谢途径。在许多模式植物的研究中, 已阐明了自主途径、春化途径、赤霉素途径及光周期途径等多种

开花控制途径(Corbersier等2007; 宋杨等2014)。在对‘徐薯18’开花的调控网络进行解析后发现, 首先, 植物中开花途径调控基因FLOWERING LOCUS C (FLC)在甘薯花器官中表达量显著上调, FLOWERING LOCUS T (FT)的表达量无明显变化, 而SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)在花中的表达量比营养器官中的表达量低。从上述3个基因相互调控的关系可知, 由于FLC可整合植物开花的自主途径、春化途径以及PAF1复合物途径等3条开花调控途径的信号, 且FLC阻遏SOC1和FT的表达, 故FLC的上调理应降低SOC1和FT的表达量, 而SOC1和FT的表达量却无明显变化或低表达。因此, 甘薯在干旱条件下开花与植物开花的自主途径、春化途径以及PAF1复合物途径无直接联系。其次, 由于开花控制因子CONSTANS (CO)可促进SOC1和FT的表达, 虽然发现CO在甘薯花器官中的表达量有上调现象, 但因为SOC1和FT的表达量低, 因此甘薯在干旱条件下的开花调控也不属于光周期途径。最后, 由于甘薯花分生组织诱导基因LEAFY (LFY)和APETALAI (API)的表达量在花器官中显著高于营养器官, 因此推测甘薯在干旱胁迫条件下的开花生理主要受赤霉素途径的调节。通过进一步分析赤霉素途径中的几个关键基因的表达, 发现赤霉素途径中的主要调控蛋白基因DELLA在甘薯花器官中的表达上调, 且赤霉素受体基因GID1也表达上调。DELLA表达丰度的提高在一定程度上可提高植物的抗旱能力, 但其在赤霉素信号传导通路中起到阻遏作用, 而GID1的表达上调使得更多的活性赤霉素分子与GID1蛋白结合, 进一步通过级联放大效应将此信号放大而降解DELLA并促进开花(Tao等2013)。

为探索嫁接甘薯开花的分子机理, 我们还将‘徐薯18’与甘薯近缘物种牵牛(*I. nil*)嫁接, 收集接穗甘薯的茎、叶、青蕾、开放花, 砧木的茎、根, 对照甘薯的茎和叶, 对照牵牛的茎和根等样品, 分别进行转录组测序(Wei等2015)。通过比较转录组分析后发现, 多数基因在接穗茎和叶中的表达量明显低于对照甘薯茎和叶中的表达量; 与干旱胁迫开花相比, 接穗甘薯花中大量基因的表达也被下调, 暗示嫁接在一定程度上对甘薯造成了逆境

胁迫, 导致接穗甘薯茎和叶中大量的基因表达被抑制(结果未发表)。通过接穗茎叶与对照甘薯间的比较转录组发现, 接穗茎叶中蔗糖代谢和转运的关键酶基因, 如*SuSy*基因、蔗糖转运蛋白基因(*sucrose transporter, SUT*)等, 其表达量均发生上调, 因此, 嫁接胁迫并未抑制甘薯与牵牛的嫁接体植株的蔗糖代谢和转运途径, 反而在嫁接体植株的开花期接穗叶中蔗糖的合成和代谢过程很活跃, 与已有的研究结果相似(Chincinska等2008)。由于嫁接体植株缺少甘薯块根库器官, 其地上部可能会积累较高水平的蔗糖, 而植物地上部尤其是顶端分生组织蔗糖水平的提高可以促进开花(Bolouri Moghaddam和Van den Ende 2013)。也就是说, 甘薯的嫁接过程会影响到营养物质向地下部的转运, 而库源关系的改变可以使营养物质更多地积累在地上部的库器官中, 从而有利于茎、叶积累足够的营养物质以形成花芽和促进甘薯开花。

利用权重基因共表达网络分析(weighted gene co-expression network analysis, WGCNA)方法(Langfelder和Horvath 2008)构建甘薯基因共表达网络, 可以分析甘薯基因在嫁接后青蕾期花芽、成熟花, 干旱胁迫全花器官, 嫁接茎叶组织和对照茎叶组织等7个组织间的表达差异, 挖掘甘薯开花相关控制基因共表达模块。在获得的22个基因共表达模块(M1~M22)中, 5个模块与组织特异性相关, 即M2(干旱胁迫全花)、M6(嫁接成熟花)、M7(嫁接花芽)、M15(嫁接茎)、M21(嫁接叶)。在经过基因富集分析后发现, M2主要涉及孢子花粉素生物合成及花药发育相关基因、儿茶酚氧化酶活性、果胶分解代谢以及乙醛酸脱氢酶; M6中富集的基因都参与器官衰老、脱落(脱落酸信号通路途径调控)以及自我吞噬; M7中的基因主要参与ATP生物合成(包括能量耦合质子转运及呼吸作用电子转运链等); M15和M21中富集的基因主要参与叶绿素合成及代谢过程、乙醛酸循环及光合磷酸化作用, 并涉及丝裂原活化蛋白激酶(mitogen-activated protein kinases, MAPKs)级联反应过程。也就是说, 甘薯经嫁接后, 叶组织光合及呼吸作用增强, 开花过程中有多种调节细胞生长、分化及适应环境应激基因的高度表达。相对于嫁接的成熟花中所富集的大量脱落、衰老及自噬基因表达,

干旱条件下的高表达基因更有利于花器官形成及稳定发育(结果未发表)。

3.6 其他重要功能基因的挖掘

在‘徐薯18’的转录组数据库中, 存在大量的长片段叶绿体基因转录本, 有的转录本的长度超过10 kb。通过‘徐薯18’转录组中叶绿体基因转录本和‘徐薯18’叶DNA测序及组装, 获得了甘薯叶绿体的全长序列以及它的结构(Yan等2015)。由于长非编码RNA是非常重要的调控因子(Zhang等2013), 我们从‘徐薯18’转录组序列中挖掘到17 228条长链非编码RNA(long noncoding RNA, LncRNA)序列(结果未发表)。此外, 还从‘徐薯18’转录组中挖掘到4 249条cDNA-derived simple sequence repeats(cSSRs)序列(Tao等2012)。

4 小结与展望

虽然我们近几年来对甘薯‘徐薯18’转录组数据库的构建和重要功能基因的挖掘开展了一些研究工作, 但是对于甘薯这种遗传结构复杂的物种来说是远远不够的, 因此, 今后需要重点开展以下研究工作。

4.1 建立更加完整的甘薯转录组数据库

一个完整的转录组数据库, 可以为甘薯基因组的基因注释提供强有力的支撑, 也可以获得具有全长编码序列的基因, 为基因克隆和功能研究奠定基础。由于基因的表达具有时空特异性, 而且不同基因的表达量差异很大, 因此, 要构建完整的甘薯转录组数据库需要通过长期的努力。有的基因转录本可以通过可变剪接, 产生不同序列的mRNA, 利用第2代测序和组装无法获得这种可变剪接转录本, 所以需要利用第3代测序技术进行转录组测序, 因为第3代测序技术可以获得更多的转录本信息(Bleidorn 2016; Lee等2016)。

4.2 基于甘薯转录组的共表达网络及E-TWAS研究

转录组研究的重要应用价值之一在于生物体表现出的重要性状所对应的基因表达调控等分子机理, 因此, 很多关联物种性状和基因的重要研究方法和算法应运而生。虽然这些研究方法更多是用于有参考基因组背景的大规模重测序群体, 如全基因组关联分析(genome wide association study, GWAS)及群体结构和进化分析等(Bush和Moore 2012; Korte和Farlow 2013), 但是基于转录组测序

和基因挖掘更能直观地对基因表达和特定性状进行关联。更重要的是, 由于转录组中挖掘的基因表达信息直接受到外界环境影响, 因此能够直观地反映不同外界环境或胁迫对于生物体基因表达和调控的影响。WGCNA和全转录组关联分析(transcriptome wide association study, TWAS), 都是借助多个平行转录组数据对某些重要性状进行基因调控通路上的挖掘和关联, 也将成为多时间点多梯度研究多个转录组基因表达信息的新手段(Langfelder和Horvath 2008; Gusev等2016)。

4.3 甘薯转录组学与其他组学研究

如上所述, 甘薯转录组的研究已取得较大进展, 而甘薯基因组(Hirakawa等2015; Si等2016)和蛋白质组学(Jiang等2012; Shekhar等2016)的研究则还刚刚起步, 与其他作物相比, 仍有很大差距。随着基因组、蛋白组、代谢组、表观遗传组等多个技术手段的突破和发展, 基因及其分子生物学的研究成为多个组学的交叉、结合和应用。目前, 转录组和代谢组的结合研究成为了新热点, 代谢产物是基因表达和生物体表型间的一道桥梁, 基因的表达信息与代谢产物的富集信息能够更精确、更可靠地判断生物体在特定环境下的表型及其调控机制(Bielecka等2015; Chen等2016)。随着多组学交叉结合研究思路的影响及多种测序手段和技术的突破, 未来的分子生物学研究将会依托于大数据背景, 向着多组学和多维度的系统生物学研究方向发展。

参考文献

- Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Ben-Hur A, Reddy ASN (2016). A survey of the sorghum transcriptome using single-molecule long reads. *Nat Commun*, doi: 10.1038/ncomms11706
- Bielecka M, Watanabe M, Morcuende R, Scheible WR, Hawkesford MJ, Hesse H, Hoefgen R (2015). Transcriptome and metabolome analysis of plant sulfate starvation and resupply provides novel information on transcriptional regulation of metabolism associated with sulfur, nitrogen and phosphorus nutritional responses in *Arabidopsis*. *Front Plant Sci*, 5: 805
- Biémont C (2010). A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics*, 186: 1085–1093
- Bleidorn C (2016). Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. *Syst Biodivers*, 14 (1): 1–8
- Bolouri Moghaddam MR, Van den Ende W (2013). Sugars, the clock and transition to flowering. *Front Plant Sci*, 4: 22
- Bush WS, Moore JH (2012). Chapter 11: Genome-wide association studies. *PLoS Comput Biol*, 8: e1002822
- Cao QH, Li A, Chen JY, Sun Y, Tang J, Zhang A, Zhou ZL, Zhao DL, Ma DF, Gao S (2016). Transcriptome sequencing of the sweet potato progenitor (*Ipomoea trifida* (HBK) G. Don.) and discovery of drought tolerance genes. *Trop Plant Biol*, 9 (2): 63–72
- Cao QH, Shao HH, Gu YH, Tao X, Huang C, Zhang YZ (2012). Cloning, characterization and expression analysis of *HSP70* gene from sweet potato. *Afr J Microbiol Res*, 6: 6325–6332
- Chang MJ, Tao X, Gu YH, Lin G, Shao HH, Cao QH, Zhang YZ (2012). Cloning and characterization of the 14-3-3 protein gene from *Ipomoea batatas* (L.) Lam. *Afr J Microbiol Res*, 6: 1990–1999
- Chen H, Tan XF (2014). Excavation of genic resources based on next generation sequencing technologies. *Plant Physiol J*, 50 (8): 1089–1095 (in Chinese with English abstract) [陈昊, 谭晓风 (2014). 基于第二代测序技术的基因资源挖掘. *植物生理学报*, 50 (8): 1089–1095]
- Chen J, Jiang YS, Tao X, Tan XM, Zhang YZ (2014). Cloning and expression profile of betaine aldehyde dehydrogenase gene of *Ipomoea batatas* in response to salt stress. *Russ J Plant Physiol*, 61 (4): 476–483
- Chen J, Jiang YS, Zhang YZ, Tan XM (2014). Cloning and expression analyses of LEA2 gene from *Ipomoea batatas*. *Chin J Appl Environ Biol*, 20 (2): 204–210 (in Chinese with English abstract) [陈娇, 姜玉松, 张义正, 谭雪梅(2014). 甘薯LEA2基因的克隆与表达分析. *应用与环境生物学报*, 20 (2): 204–210]
- Chen W, Wang WS, Peng M, Gong L, Gao YQ, Wan J, Wang SC, Shi L, Zhou B, Li ZM, et al (2016). Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nat Commun*, 7: 12767
- Chincinska IA, Liesche J, Krügel U, Michalska J, Geigenberger P, Grimm B, Kühn C (2008). Sucrose transporter StSUT4 from potato affects flowering, tuberization, and shade avoidance response. *Plant Physiol*, 146: 515–528
- Corbesier L, Vincent C, Jang SH, Fornara F, Fan QZ, Searle I, Giakountis A, Farrona S, Gissot L, Turnbull C, et al (2007). FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science*, 316 (5827): 1030–1033
- Firon N, LaBonte D, Villordon A, Kfir Y, Solis J, Lapis E, Perlman TS, Doron-Faigenboim A, Hetzroni A, Altham L, et al (2013). Transcriptional profiling of sweetpotato (*Ipomoea batatas*) roots indicates down-regulation of lignin biosynthesis and up-regulation of starch biosynthesis at an early stage of storage root formation. *BMC Genomics*, 14: 460
- Gu YH, Tao X, Lai XJ, Wang HY, Zhang YZ (2014). Exploring the polyadenylated RNA virome of sweet potato through high-throughput sequencing. *PLoS ONE*, 9: e98884
- Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, Jansen R, De Geus EJC, Boomsma DI, Wright FA, et al (2016). Integrative approaches for large-scale transcriptome-wide association

- studies. *Nat Genet*, 48: 245–252
- He B, Zhao SR, Chen YH, Cao QH, Wei CH, Cheng XJ, Zhang YZ (2015). Optimal assembly strategies of transcriptome related to ploidies of eukaryotic organisms. *BMC Genomics*, 16 (1): 65
- Hirakawa H, Okada Y, Tabuchi H, Shirasawa K, Watanabe A, Tsuruoka H, Minami C, Nakayama S, Sasamoto S, Kohara M, et al (2015). Survey of genome sequences in a wild sweet potato, *Ipomoea trifida* (H. B. K.) G. Don. *DNA Res*, 22 (2): 171–179
- Jiang YS, Chen C, Tao X, Wang JX, Zhang YZ (2012). A proteomic analysis of storage stress responses in *Ipomoea batatas* (L.) Lam. tuberous root. *Mol Biol Rep*, 39 (8): 8015–8025
- Jiang ZH, Zhou X, Li R, Michal JJ, Zhang SW, Dodson MV, Zhang ZW, Harland RM (2015). Whole transcriptome analysis with sequencing: methods, challenges and potential solutions. *Cell Mol Life Sci*, 72: 3425–3439
- Korte A, Farlow A (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, 9: 29
- Lai XJ, Gu YH, Wang HY, Zhang YZ (2013). Cloning and expression analyses of sporamin family genes in sweet potato. *Chin J Appl Environ Biol*, 19 (2): 215–223 (in Chinese with English abstract) [赖先军, 古英洪, 王海燕, 张义正(2013). 甘薯Sporamin家族基因克隆与表达分析. 应用与环境生物学报, 19 (2): 215–223]
- Lai XJ, Gu YH, Tao X, Zhang YZ, Wang HY (2014). Cloning and characterization of uridine diphosphate glucose dehydrogenase gene from *Ipomoea batatas*. *Russ J Plant Physiol*, 61: 298–308
- Langfelder P, Horvath S (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9: 559
- Lee H, Gurtowski J, Yoo S, Nattestad M, Marcus S, Goodwin S, McCombie WR, Schatz MC (2016). Third-generation sequencing and the future of genomics. *BioRxiv*, 486603
- Leslie T, Baucom RS (2014). *De Novo* assembly and annotation of the transcriptome of the agricultural weed *Ipomoea purpurea* uncovers gene expression changes associated with herbicide resistance. *G3 (Bethesda)*, 4 (10): 2035–2047
- Li RJ, Zhai H, Kang C, Liu DG, He SZ, Liu QC (2015). *De novo* transcriptome sequencing of the orange-fleshed sweet potato and analysis of differentially expressed genes related to carotenoid biosynthesis. *Int J Genomics*, 2015: 843802
- Li XD, Wei CH, Shao HH, Wu Y, Zhang YZ, Wang HY (2016). Cloning and expression of nonspecific lipid transfer protein genes from sweetpotato [*Ipomoea batatas* (L.) Lam.]. *J Plant Sci*, 34: 583–592
- Lisch D (2013). How important are transposons for plant evolution? *Nat Rev Genet*, 14: 49–61
- Ma PY, Bian XF, Jia ZD, Guo XD, Xie YZ (2016). *De novo* sequencing and comprehensive analysis of the mutant transcriptome from purple sweet potato (*Ipomoea batatas* L.). *Gene*, 575: 641–649
- Martin JA, Wang Z (2011). Next-generation transcriptome assembly. *Nat Rev Genet*, 12: 671–682
- McGettigan PA (2013). Transcriptomics in the RNA-seq era. *Curr Opin Chem Biol*, 17: 4–11
- McGregor CE, Miano DW, LaBonte DR, Hoy M, Clark CA, Rosa GJM (2009). Differential gene expression of resistant and susceptible sweetpotato plants after infection with the causal agents of sweet potato virus disease. *J Am Soc Hortic Sci*, 134: 658–666
- Minoche AE, Dohm JC, Schneider J, Holtgräwe D, Viehöver P, Montfort M, Sörensen TR, Weisshaar B, Himmelbauer H (2015). Exploiting single-molecule transcript sequencing for eukaryotic gene prediction. *Genome Biol*, 16: 184
- Mutz K-O, Heilkenbrinker A, Lönne M, Walter J-G, Stahl F (2013). Transcriptome analysis using next-generation sequencing. *Curr Opin Biotechnol*, 24: 22–30
- Schafleitner R, Tincopa LR, Palomino O, Rossel G, Robles RF, Alagon R, Rivera C, Quispe C, Rojas L, Pacheco JA, et al (2010). A sweet potato gene index established by *de novo* assembly of pyrosequencing and Sanger sequences and mining for gene-based microsatellite markers. *BMC Genomics*, 11: 604
- Senthilkumar R, Yeh KW (2012). Multiple biological functions of sporamin related to stress tolerance in sweet potato (*Ipomoea batatas* Lam). *Biotechnol Adv*, 30 (6): 1309–1317
- Shao HH, Cao QH, Tao X, Gu YH, Chang MJ, Huang CL, Zhang YZ, Feng H (2011). Cloning and characterization of ATP synthase CF1 α gene from sweet potato. *Afr J Biotechnol*, 10: 19035–19042
- Shekhar S, Mishra D, Gayali S, Buragohain AK, Chakraborty S, Chakraborty N (2016). Comparison of proteomic and metabolomic profiles of two contrasting ecotypes of sweetpotato (*Ipomoea batata* L.). *J Proteomics*, 143: 306–317
- Si ZZ, Du B, Huo JX, He SZ, Liu QC, Zhai H (2016). A genome-wide BAC-end sequence survey provides first insights into sweetpotato (*Ipomoea batatas* (L.) Lam.) genome composition. *BMC Genomics*, 17: 945
- Solis J, Baisakh N, Brandt SR, Villordon A, La Bonte D (2016). Transcriptome profiling of beach morning glory (*Ipomoea imperati*) under salinity and its comparative analysis with sweetpotato. *PLoS ONE*, 11: e0147398
- Song Y, Dou LD, Zhang HJ (2014). Molecular and genetic mechanisms of control of floral induction in higher plants. *Plant Physiol J*, 50 (10): 1459–1468 (in Chinese with English abstract) [宋杨, 窦连登, 张红军(2014). 高等植物成花诱导调控的分子和遗传机制. 植物生理学报, 50 (10): 1459–1468]
- Tao X, Gu YH, Jiang YS, Zhang YZ, Wang HY (2013). Transcriptome analysis to identify putative floral-specific genes and flowering regulatory-related genes of sweet potato. *Biosci Biotechnol Biochem*, 77: 2169–2174
- Tao X, Gu YH, Wang HY, Zheng W, Li X, Zhao CW, Zhang YZ (2012). Digital gene expression analysis based on integrated *de novo* transcriptome assembly of sweet potato [*Ipomoea batatas* (L.) Lam.]. *PLoS ONE*, 7: e36234
- Varshney RK, Glaszmann J-C, Leung H, Ribaut J-M (2010). More genomic resources for less-studied crops. *Trends Biotechnol*, 28: 452–460
- Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, Lu Z, Olson A, Stein JC, Ware D (2016). Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat Commun*, 7: 11708

- Wang JX, Jiang YS, Yan Y, Tao X, Zhang YZ (2015). Cloning, expression, and characterization of *phyA* gene from *Ipomoea batatas*. *Russ J Plant Physiol*, 62 (1): 109–115
- Wang ZY, Fang BP, Chen JY, Zhang XJ, Luo ZX, Huang LF, Chen XL, Li YJ (2010). *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). *BMC Genomics*, 11: 726
- Wei CH, Tao X, Li M, He B, Yan L, Tan XM, Zhang YZ (2015). *De novo* transcriptome assembly of *Ipomoea nil* using Illumina sequencing for gene discovery and SSR marker identification. *Mol Genet Genomics*, 290: 1873–1884
- Xie F, Burklew CE, Yang Y, Liu M, Xiao P, Zhang B, Qiu D (2012). *De novo* sequencing and a comprehensive analysis of purple sweet potato (*Ipomoea batatas* L.) transcriptome. *Planta*, 236: 101–113
- Xu ZC, Peters RJ, Weirather J, Luo HM, Liao BS, Zhang X, Zhu YJ, Ji AJ, Zhang B, Hu SN, et al (2015). Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. *Plant J*, 82: 951–961
- Yan L, Gu YH, Tao X, Lai XJ, Zhang YZ, Tan XM, Wang HY (2014). Scanning of transposable elements and analyzing expression of transposase genes of sweet potato [*Ipomoea batatas*]. *PLoS ONE*, 9: e90895
- Yan L, Lai XJ, Li XD, Wei CH, Tan XM, Zhang YZ (2015). Analyses of the complete genome and gene expression of chloroplast of sweet potato [*Ipomoea batata*]. *PLoS ONE*, 10 (4): e0124083
- Zhang J, Mujahid H, Hou YX, Nallamilli BR, Peng ZH (2013). Plant long ncRNAs: a new frontier for gene regulatory control. *Am J Plant Sci*, 4 (5): 1038–1045
- Zhou YX, Chen YX, Tao X, Cheng XJ, Wang HY (2016). Isolation and characterization of cDNAs and genomic DNAs encoding ADP-glucose pyrophosphorylase large and small subunits from sweet potato. *Mol Genet Genomics*, 291: 609–620

Transcriptomic analyses in sweetpotato [*Ipomoea batatas* (L.) Lam. cv. ‘Xushu 18’]

YAN Lang, WEI Chang-He, ZHANG Yi-Zheng*

College of Life Sciences, Sichuan Key Laboratory of Molecular Biology and Biotechnology, Sichuan University, Chengdu 610064, China

Abstract: The sweetpotato (*Ipomoea batatas* cv. ‘Xushu 18’) has a long period and wide area of cultivation in China. The construction of transcriptomic databases and the mining of functional genes of this cultivar will play significant roles in the genetic diversity analysis and molecular breeding of sweetpotato. This paper reviews the recent research progresses in ‘Xushu 18’ transcriptome, including the construction of full-length transcriptome databases using RNA-seq and Iso-seq, the mining of important functional genes and their expression. The future direction is discussed in the study of sweetpotato transcriptome.

Key words: sweetpotato; ‘Xushu 18’; transcriptome; gene mining; gene expression

Received 2017-02-15 Accepted 2017-05-08

This work was supported by the National Science & Technology Pillar Program of China (Grant No. 2007BAD78B03) and the “Eleven-Five” Key Project of Sichuan Province (Grant No. 07SG111-003-1).

*Corresponding author (E-mail: yizzhang@scu.edu.cn).