论文

中华猕猴桃基因组可变剪接事件鉴定及分析

周晏秋^①, 覃瑞^①, 刘胜毅^②, 李刚^①, 黄军艳^②, 董彩华^②, 程晓晖^②, 刘虹^{①*},
童超波^{②*}^① 中南民族大学武陵山区特色资源植物种质保护与利用湖北省重点实验室, 武汉 430074;^② 中国农业科学院油料作物研究所, 农业部油料作物生物学与遗传育种重点实验室, 武汉 430062

* 联系人, E-mail: liuhong@mail.scuec.edu.cn; tongchaobo@gmail.com

收稿日期: 2014-10-29; 接受日期: 2014-12-29; 网络版发表日期: 2015-03-10

国家自然科学基金(批准号: 31370253, 31301039 和 31471536)和武汉市青年科技晨光计划(批准号: 201271031395)资助项目
doi: 10.1360/N052014-00275

摘要 可变剪接使一个基因能产生多种 mRNA 成熟体, 极大地增加蛋白多样性. 采用中华猕猴桃基因组数据做参考数据, 利用中华猕猴桃叶片和果实 3 个不同发育时期(未成熟、半成熟和成熟期)的转录组数据, 从中华猕猴桃基因组(39040 个基因)中共鉴定出 11651 个基因(占总基因数的 29%)对应的 32180 个可变剪接事件. 在可变剪接不同类型中, 内含子保留类型的发生频率最高, 占 50%以上; 3'可变位点类型频率约为 5'端可变类型的 2 倍. GO 富集分析结果表明, 可变剪接的基因主要富集于酶调控及核苷酸结合相关功能的 GO 类别中, 而组织特有可变剪接基因功能富集热点与组织的重要功能关联, 叶片多为肌动蛋白及微管相关; 未成熟果实与双组分信号系统相关; 半成熟果实多与磷脂合成过程相关; 成熟果实多与信号传递过程相关. 另外, 55.6%的维生素合成相关基因发生可变剪接事件, 显著高于基因组水平的 29.6%, 暗示着可变剪接参与维生素合成相关基因代谢过程中的重要作用. 通过对中华猕猴桃全基因组可变剪接的分析, 为解析中华猕猴桃基因组及进一步开展相关分子育种工作提供依据.

关键词

中华猕猴桃
可变剪接
基因组
维生素

猕猴桃(*Actinidia chinensis*)原产于中国, 是营养最为全面的水果之一, 深受世界各国人民的喜爱^[1]. 猕猴桃作为一种经济作物, 不仅味美, 还拥有极高的营养及药用价值, 被誉为“水果之王”. 猕猴桃的果肉含有丰富的维生素 C、矿物质、膳食纤维和其他有益健康的营养代谢成分^[2,3]. 已有文献报道, 猕猴桃不仅可以美容养颜, 辅助治疗口腔溃疡、消化不良、坏血病等, 还可以提升人体免疫功能、预防癌症和抗衰老^[4,5]. 随着新一代测序技术的飞速发展和应用^[6], 猕猴桃全基因组序列草图于 2013 年成功绘制并公开

发表^[7], 为推动猕猴桃基因组学的研究提供了便利和基础. 通过对猕猴桃基因组层次的研究, 尤其对营养品质如维生素 C 代谢相关网络及基因的解析, 将为猕猴桃品质改良和遗传育种提供参考和依据.

可变剪接(alternative splicing, AS)是指同一基因编码产生的 mRNA 前体通过不同的剪接方式产生 2 种或 2 种以上成熟体的过程^[8]. mRNA 前体需要在一式种大分子复合体(spliceosome)的帮助下通过 RNA 聚合酶 II 催化完成剪接. 这个复合体由 170 余种蛋白和 5 种 snRNAs(small nuclear RNAs)构成^[9], 在 mRNA

引用格式: 周晏秋, 覃瑞, 刘胜毅, 等. 中华猕猴桃基因组可变剪接事件鉴定及分析. 中国科学: 生命科学, 2015, 45: 289–300

Zhou Y Q, Qin R, Liu S Y, et al. Genome-wide survey of alternative splicing in *Actinidia chinensis*. SCIENTIA SINICA Vitae, 2015, 45: 289–300, doi: 10.1360/N052014-00275

剪接过程中受 5'端 GU、3'端 AG、内含子 3'端多嘧啶序列、3'剪接位点上游 17~40 碱基的分支点(CURAY)等决定信号调控^[10]。根据可变剪接位点在基因内含子和外显子中的发生位置,可将可变剪接事件分为 4 种基本类型^[11]: (i) 内含子保留(intron retention, IR); (ii) 可变的 3'端剪接位点(alternative 3' splice site, A3SS); (iii) 可变的 5'端剪接位点(alternative 5' splice site, A5SS); (iv) 外显子跳跃(exon skipping, ES)。

可变剪接事件广泛存在于真核生物基因表达过程中^[12],发生于参与表达调控过程的基因,如信号传导通路^[13]及程序性死亡^[14]相关基因。可变剪接能影响 mRNA 翻译中顺式作用元件的效率、稳定性和定位^[15],调节细胞内基因的表达,在植物的生长发育、生理代谢、生物和非生物胁迫等过程中发挥调控作用^[16,17]。例如,拟南芥开花自主途径中 *FCA*(flowering time control protein)基因通过可变剪接产生 4 转录本,并通过对转录本表达量的调控来完成自身的负调控表达,最终影响开花时间^[18]。拟南芥(*Arabidopsis thaliana*)能通过可变剪接事件产生的异构体影响下游许多基因的表达,从而应对盐胁迫或冷胁迫刺激^[16,19]。可变剪接能产生不同的蛋白异构体,其功能可能相似,也可能相反^[20]。另外,相当一部分转录本由于可变剪接能改变其可读框,从而使终止密码子提前产生,最终进入 NMD(non-sense mediated mRNA decay)途径降解^[21]。

本研究以新完成的猕猴桃全基因组序列为参考序列,通过分析不同组织的转录组(RNA-seq)数据,鉴定猕猴桃基因组中的基因可变剪接事件,分析 4 种类型可变剪接在各组织中的频率及动态变化,考察组织共有和特异的可变剪接事件。通过 GO(gene ontology)富集分析,研究可变剪接主要发生在哪些基因类群中,并着重考察维生素代谢网络中的可变剪接事件。通过猕猴桃全基因组可变剪接的分析,为解析猕猴桃基因组及开展分子育种工作提供依据。

1 材料与方法

1.1 猕猴桃基因组及组织转录组数据下载

猕猴桃全基因组序列及相关注释文件下载自猕猴桃基因组数据库(kiwifruit genome database, <http://bioinfo.bti.cornell.edu/cgi-bin/kiwi/>)^[7]。通过 SRA (sequence read archive) 数据库(<http://www.ncbi.nlm.nih.gov/sra/>)

下载猕猴桃 4 个组织转录组成套的数据,分别对应成熟叶片(SRR653044)、未成熟果实(授粉后 20 天, SRR653045)、半成熟果实(授粉后 120 天, SRR653046)和成熟果实(授粉后 127 天, SRR653047)^[7]。这些转录组数据来自 Illumina HiSeq 2000 测序平台,测序序列片段读长(reads) 90 bp, 双末端测序,测序量分别为 4.8, 5.1, 5.3 和 5.3 G。这些数据在原文献^[7]中仅用于表达分析,尚未进行可变剪接分析,本研究将利用该数据进行猕猴桃全基因组水平可变剪接事件鉴定,研究叶片及果实成熟发育过程中可变剪接事件的特征和规律。

1.2 数据过滤, 比对及可变剪接事件鉴定

下载的转录组数据首先通过 sratoolkit. 2.3.5 软件^[22]中 fastq-dump 工具包进行格式转换。转换后的数据再通过 NGS QCToolkitv2.3.2 工具包^[23]进行质量过滤。质量过滤标准为: (i) 去掉接头序列; (ii) 截取掉 3'端质量值小于 20 的碱基,并且截取后序列长度不低于 50; (iii) 去除 N(N 表示无法确定的碱基)比例大于 10% 的序列片段; (iv) 过滤掉质量值小于 20 的碱基数目超过该序列总长度 30% 的序列片段。

采用 Tophat 软件^[24]进行转录组测序序列片段与猕猴桃参考基因组之间的比对,其中最大比对错配数设置为 2,其余采用默认参数。Tophat 能够实现序列片段劈开比对(split-mapped)基因组,从而鉴定出基因组上所有可能的序列片段支持的剪接位点(junction),剪接位点的左右两端分别为 GT-AG 碱基,定义出转录本不同外显子的边界及组合关系,每个剪接位点要求至少有 3 条序列片段支持,每条序列劈开比对后覆盖剪接位点左右边界至少 5 个碱基。参考文献中提供的方法^[11,20],对目前已经定义的 4 种可变剪接进行分类鉴定: (i) 外显子跳跃。除正常剪接位点外,要求有跨越该外显子、直接连接左右外显子的剪接位点存在; (ii) 内含子保留。除正常跨越该内含子的剪接位点存在外,在该内含子 90% 以上的区域均有唯一比对序列(unique mapping reads)覆盖,且其平均值测序深度大于两边外显子区域的平均测序深度较小一个的 15%; (iii) 5'端可变位点。除正常剪接位点外,左侧的剪接位点(donor sites)位置在左侧外显子上出现另外一种可变位置; (iv) 3'端可变位点。除正常剪接位点外,右侧的剪接位点位置在右侧外显子上出现另外一种可变位置。

1.3 GO 富集分析及猕猴桃维生素功能相关基因分析

通过 InterProScan^[25] 比对查询多个数据库, 获得包括分子功能(molecular function)、生物学过程(biological process)和细胞构成(cellular component)等类别内的 GO 注释编号. 通过 AgriGo 在线工具 (<http://bioinfo.cau.edu.cn/agriGO/index.php>), 以全基因组为背景进行功能富集分析. 利用在线工具 Venny (<http://bioinfo.gp.cnb.csic.es/tools/venny/index.html>) 分析并展示猕猴桃各组织样本中共有或特异的可变剪接事件数目, 并对组织样本共有或特有的可变剪接基因分别进行功能富集分析. 富集分析计算每个 GO 类别中所操作基因集与全基因组背景之间的差异, 采用 Fisher 统计检验并对 *P* 值进行 bonferroni 校正. 若 $P < 0.05$, 则定义所操作基因集在该 GO 类别中富集.

利用猕猴桃基因的详细 GO 注释信息、中华猕猴桃基因组的蛋白注释及与 KEGG 网站植物基因的序列比对结果, 归纳统计注释与维生素代谢相关基因. 并使用卡方检验检测可变剪接事件在这些基因类别中的频率是否与全基因组水平有显著性差异. 通过 Cufflink^[26] 软件计算基因的表达量 FPKM (fragments kilobase of exon model per million mapped reads) 值, 并使用 R 软件包绘制猕猴桃基因的表达量热图 (heatmap).

2 结果与分析

2.1 猕猴桃基因组可变剪接事件鉴定

通过叶与未成熟、半成熟及成熟果实组织的转录组数据, 共鉴定了猕猴桃基因组 32180 个可变剪接事件, 这些事件发生于 11561 个基因中, 占基因总数的

29.6% (表 1). 按可变剪接事件类型分类, 内含子保留类型含量最高, 共 15516 个, 占有可变剪接事件的 48.2%; 3'端位点可变共 8328 个, 占 25.9%; 5'端位点可变共 7249 个, 占 22.5%; 外显子跳跃共 2891 个, 占 9%. 另外, 按可变剪接基因的数量比例分析: 发生内含子保留的基因最多, 在可变剪接基因中所占比例为 61.7%; 发生 3'端位点可变、5'端位点可变、外显子跳跃的基因依次减少, 占可变剪接基因比例分别为 48.9%, 35.6% 和 19.1%. 每个可变剪接基因平均发生 2.78 个可变剪接事件, 发生内含子保留, 3'端位点可变、5'端位点可变、外显子跳跃事件的平均值分别为 1.34, 0.72, 0.63 和 0.25.

通过各组织间可变剪接事件共有性和特异性比较 (图 1), 可变剪接事件的发生在各个组织中出现动态变化, 不同组织中发生可变剪接的基因不完全相同. 11561 个发生可变剪接的基因中, 在不同组织中均发生可变剪接事件的基因只有 1811 个, 占猕猴桃全部基因的 5.6%, 占发生可变剪接基因总数的 15.7%. 另外, 仅在一个组织中发生可变剪接事件的基因, 在叶片、未成熟果实、半成熟果实和成熟果实中分别有 1913, 612, 783 和 1191 个.

可变剪接事件在叶片组织中最发生最多, 果实中随着成熟度的增加可变剪接事件总数也递增 (未成熟果实 < 半成熟果实 < 成熟果实). 不同类型可变剪接事件在不同组织中的比例发生着较大变化 (图 2). 内含子保留事件在未成熟果实、半成熟果实及成熟果实中分别占可变剪接事件的 62.7%, 63.7% 和 39.4%, 而外显子跳跃的比例在 3 个组织中分别为 3.9%, 2.8% 和 15.8%. 果实成熟后可变剪接中内含子保留事件明显减少, 而外显子跳跃事件明显增多. 不同组织中, 3'端位点和 5'端位点可变事件所占比例无显著变化.

表 1 猕猴桃基因组中鉴定的可变剪接事件^{a)}

可变剪接	共计(个)	内含子保留(个)	3'端位点可变(个)	5'端位点可变(个)	外显子跳跃(个)
叶片	7762(16160)	4451(8638)	3201(4049)	2182(2581)	769(892)
未成熟果实	4723(8463)	2713(5304)	1622(1825)	924(1003)	302(331)
半成熟果实	5555(10652)	3469(6787)	1952(2269)	1159(1301)	262(295)
成熟果实	6374(13007)	2723(5124)	2668(3460)	1972(2364)	1605(2059)
共计	11561(32180)	7137(15516)	5652(8328)	418(7249)	2212(2891)

a) 数据表示为发生可变剪接的基因个数(可变剪接事件的个数). 1 个基因可能对应多个可变剪接事件

图1 猕猴桃4个不同组织共有和特异的可变剪接基因数目

图2 猕猴桃4个不同组织中各类型可变剪接事件所占百分比

2.2 猕猴桃基因组可变剪接基因 GO 注释及富集分析

通过 InterProScan 软件, 16737(占总基因数的 43.3%)个猕猴桃基因获得了 GO 注释, 其中包括 7247 个发生可变剪接的基因. 以获得注释的基因作为分析对象, 对发生可变剪接的基因集进行 GO 富集分析 (图 3). 发现在分子功能大类别中, 可变剪接事件最活跃基因主要集中在酶调控和核苷酸结合等相关的 GO 类别, 如 GTPase 和 Ras 酶调节蛋白(GO: 0005083)、水解酶(GO: 0016818)、连接酶(GO: 0016874)和核苷酸结合域类(GO: 0000166)等. 按照 GO 分级系统, 在水解酶活性(GO: 0016817)的 GO 子类中, 含磷酸酐包括焦磷酸酶(GO: 0016462)及核苷三磷酸酶(GO: 0017111)的活性调节与可变剪接事件发生频率相对较高; 在核苷酸结合域(GO: 0000166)的 GO 子类中, 尤其是嘌呤核苷酸结合域(GO: 0032555)相关功能基因的可变剪接事件频率相对较

高. 在生物学过程中, 可变剪接基因富集途径主要有: 信号转导的调节(GO: 0023046), 如小 GTPase 介导的信号转导调控(GO: 0051056)等; 小分子代谢的调节(GO: 0044281), 如细胞内氨基酸代谢途径(GO: 0006520)等; 细胞分解代谢的调节(GO: 0009056), 如细胞大分子分解(GO: 0044265)等; 大分子定位(GO: 0033036), 如胞内转导(GO: 0046907)、蛋白定位(GO: 0034613)等. 按细胞构成分析, 发生可变剪接事件的基因主要归类在膜结合细胞器(GO: 0043231)及细胞质(GO: 0005737), 其中事件频率最高的 2 个分别为内质网系统(GO: 0012505)和核(GO: 0005654). 这些富集类型体现了可变剪接对于这些类型的基因行使其正常功能具有重要作用.

叶片、未成熟果实、半成熟果实及成熟果实中获得注释且仅在一个组织中发生可变剪接的基因数量分别为 1164, 377, 460 和 707 个. 对其分别进行富集分析, 发现不同组织中特异的可变剪接基因的热点具有明显特征(图 4): 叶片特异的热点与微管相关, 包括基于微管运动(GO: 0007018)、微管相关机制(GO: 0007017)、马达蛋白活性(GO: 0003774)、微管马达蛋白活性(GO: 0003777)、微管结合(GO: 0008017)、微管蛋白结合(GO: 0015631)和细胞骨架蛋白结合(GO: 0008092)等; 未成熟果实的富集结果仅有一个热点, 为双组分信号转导系统(GO: 0000160); 半成熟果实特异可变剪接基因的富集热点主要与磷脂生物合成过程(GO: 0008654)相关, 且主要位于内质网(GO: 0044432); 而成熟果实特有的富集热点较多, 均与信号传递相关, 包括小 GTPase 介导的信号转导(GO: 0007264)、细胞内级联信号(GO: 0007242)、信号转导(GO: 0007165)、信号传输(GO: 0023060)、信号处理(GO: 0023046)、调节信号转导(GO: 0009966)、信号流程监管(GO: 0023051)、细胞通讯调节(GO: 0010646)等, 且发生位置位于细胞质(GO: 0005737)或胞质内(GO: 0044444). 这些在组织中特异发生可变剪接对于基因在该组织中行使其功能具有重要作用.

2.3 猕猴桃维生素相关基因表达差异分析

在猕猴桃基因组中共鉴定了 83 个与维生素合成相关的基因(表 2), 其中 47 个基因发生了共 224 个可变剪接事件(表 3). 猕猴桃维生素相关基因发生可变剪接事件的频率为 56.6%, 显著高于全基因组水平(29.6%), 反映了可变剪接这一机制与猕猴桃维生素

合成相关基因表达的联系十分紧密. 在不同组织中均发生可变剪接事件的基因有 8 个, 仅在叶片、未成熟果实、半成熟果实和成熟果实中发生可变剪接事件的基因分别有 7, 2, 5 和 2 个(图 5). 排除叶片与 3 个时期的果实组织共有的 8 个可变剪接基因, 在未成熟果实、半成熟果实和成熟果实中发生可变剪接的维生素相关基因分别有 15, 17 和 14 个. 47 个发生可变剪接的维生素相关基因的表达量在叶片、未成熟果实、半成熟果实中较高, 在成熟果实中相对降低(图 6). 分析发现, 大部分维生素相关基因在果实成熟前就开始表达, 仅 2 个基因(Achn097841, Achn097841)在果实成熟后才发生可变剪接事件. 果实从半成熟过渡到成熟时期, 有 15 个在半成熟果实中发生可变剪

接事件的基因停止可变剪接现象, 且成熟果实中维生素相关基因的表达量明显降低, 这说明猕猴桃果实中维生素的合成积累主要集中在果实成熟前, 果实成熟后基因的表达有明显改变.

3 讨论与结论

可变剪接是一个基因产生的 mRNA 前体加工成多种 mRNA 成熟体, 翻译成不同蛋白的过程^[27]. 因此, 通过可变剪接一个基因可能产生多个蛋白, 极大地增加蛋白多样性和基因组的复杂性^[28-30]. 基于高通量测序数据, 在人(*Homo sapiens*)^[31]、小鼠(*Mus musculus*)^[32]、牛(*Bovini*)^[33]、黑腹果蝇(*Drosophila*

图 3 猕猴桃发生可变剪接事件基因的部分 GO 富集结果

图4 猕猴桃各组织发生特异可变剪接事件基因的富集结果

图5 猕猴桃组织共有和特异的维生素相关可变剪接基因

melanogaster)^[34]、衣藻(*Chlamydomonas*)^[35]、拟南芥^[21]、水稻(*Oryza sativa*)^[61]、线虫(*Caenorhabditis ele-*

gans)^[36]、二穗短柄草(*Brachypodium distachyon*)^[37]、杨树(*Populus*)^[38]、高粱(*Sorghum bicolor*)^[39]、小立碗藓(*Physcomitrella patens*)^[40]、白菜(*Brassica pekinensis*)^[9]和大豆(*Glycine max*)^[41]等生物中均发表了全基因组水平对可变剪接事件的检测及研究,检测到了大量新的可变剪接事件.本研究利用叶片及果实发育过程中3个不同时间点的转录组数据,共鉴定出猕猴桃全基因组中32180个可变剪接事件,涉及29.6%的基因,证实了可变剪接在猕猴桃基因中发生的普遍性.同时,这一比例接近水稻^[11]33%和玉米(*Zea mays*)^[42]37%的可变剪接基因比例,但是远低于拟南芥中61%的比例^[21,43].随着猕猴桃组织转录水平数

图 6 猕猴桃维生素相关基因在不同组织中的表达量差异

L1, F1, F2, F3 分别表示叶片、未成熟果实、半成熟果实、成熟果实;
灰色为未检测到基因表达

据的增多,这一比例还将进一步提高.因为通过不断更新的大规模数据分析,人类基因组中 95% 的多外显子基因均可发生可变剪接事件^[20,44].另外,猕猴桃中 4 种类型可变剪接事件中,内含子保留类型的事件比例最高,3'端位点可变类型比例次之,约为 5'端位点可变类型的 2 倍,这和其他植物,如苜蓿(*Medicago sativa*)、杨树(*Populus*)、拟南芥、水稻、衣藻

(*Chlamydomonas*)和二穗短柄草中的结果一致^[6,45,46].

本研究通过对猕猴桃组织特异可变剪接事件的比较分析发现,一部分基因是在特定组织发生可变剪接事件,这与目前的研究观点“可变剪接在生物不同时期不同组织中发挥基因表达的调控作用,表现出很强的组织特异性”一致^[20].通过对各个组织特有的可变剪接基因进行 GO 富集分析发现,其功能热点有明显特征:叶片中特异的功能热点与微管相关;未成熟果实特异的功能热点为双组分信号系统;半成熟果实特异的功能热点主要与磷脂合成过程相关;成熟果实特异的功能热点均与信号传递过程相关.细胞内叶绿体的运动需要肌动蛋白及微管的参与^[47],如叶绿体避光运动等^[48,49].猕猴桃叶片组织中微管相关基因可变剪接事件的富集可能与叶绿体的运动有关.在植物中双组分系统能通过感受信号(如乙烯和细胞分裂素等)来应对环境刺激及调控生长^[50].猕猴桃果实授粉后呈 S 型发育趋势,授粉 130 天后营养成分的积累达到峰值^[51,52].这符合未成熟果实中双组分系统相关基因的可变剪接事件富集及半成熟果实表现出磷脂合成相关基因的可变剪接事件富集的现象.反映了组织特异可变剪接基因的富集热点与组织的特定功能关联,组织功能的顺利实现需要这些基因可变剪接事件的参与.

对猕猴桃维生素相关基因的分析发现,维生素的合成与积累主要集中在果实成熟前,且成熟果实中可变剪接类型的比例与其他时期有较明显的差别:果实成熟后由于发生外显子跳跃事件的基因数量显著增加,内含子保留事件比例明显减少;而外显子跳跃事件比例明显增多.对成熟果实中发生可变剪接的基因进行 GO 富集分析发现,与全基因组的富集结果相比,成熟果实特有的可变剪接基因 GO 富集热点主要为信号转导相关及胞质运输途径.研究表明,果实成熟后维生素等营养物质的合成积累基本完成,果实进入成熟衰老过程,开始传递果实软化信号如加速糖代谢及分解脂类营养物质等^[53-55].成熟果实中基因表达调控网络发生重大改变,需表达相关信号转导基因并传递大量信号^[56],这与成熟果实中特有可变剪接基因富集在信号转导及胞质运输相关功能的现象相符.而外显子跳跃事件数量的显著增多,表明外显子跳跃事件可能参与调控果实的成熟衰老过程.

表 2 猕猴桃维生素相关基因的注释信息^{a)}

猕猴桃基因编号	基因注释	IPR 编号	拟南芥同源 基因编号	拟南芥同源 基因名称	KO 编号	KEGG 途径
Achn005111	维生素 C 转运体	IPR006043	AT2G27810.1		K14611	
Achn008171	—	IPR013215	AT5G17920.2			map00270
Achn014721	维生素 C 转运体	IPR006044				
Achn032341	—	—	AT1G60550.1	<i>menB</i>	K01661	map00130
Achn036641	—	IPR006968				
Achn043541	包含 CobW 结构域的蛋白	IPR011629				
Achn045051	维生素 K 环氧化还原酶(前体)	IPR012932	AT4G35760.1			
Achn053691	—	—	AT1G60550.1	<i>menB</i>	K01661	map00130
Achn053941	维生素 C 转运体	IPR006045	AT1G60030.1		K14611	
Achn055281	—	—	AT2G39770.2			
Achn058821	5-甲基四氢蝶酰三谷氨酸高半胱氨酸 甲基转移酶	IPR002629	AT5G17920.2			map00270
Achn058831	5-甲基四氢蝶酰三谷氨酸高半胱氨酸 甲基转移酶	IPR002629	AT3G03780.3			map00270
Achn060421	—	—	AT5G60540.1	<i>pdxT</i>	K08681	map00750
Achn062781	嘌呤转运体	IPR006046	AT3G10960.1	<i>pbuG</i>	K06901	
Achn068181	维生素 C 转运体	IPR006047	AT2G27810.1		K14611	
Achn071031	—	—	AT5G48840.1	<i>panC</i>	K01918	map00410
Achn072771	—	—	AT5G54770.1	<i>thi1</i>	K03146	
Achn078251	—	IPR006968				
Achn078621	—	—	AT5G48840.1	<i>panC</i>	K01918	map00410
Achn080221	—	—	AT5G54770.1	<i>thi1</i>	K03146	
Achn082061	—	—	AT5G60540.1	<i>pdxT</i>	K08681	map00750
Achn087641	—	—	AT5G49970.2			
Achn091471	—	IPR006968				
Achn095121	—	—	AT1G79600.1	<i>ADCK</i>	K08869	
Achn097841	P47K: 维生素 B12 合成相关类 CobW 蛋白	IPR011629	AT1G80480.1	<i>CNOT2</i>	K12605	
Achn107621	—	—	AT5G60540.1	<i>pdxT</i>	K08681	map00750
Achn114541	P47K: 维生素 B12 合成相关类 CobW 蛋白	IPR011629	AT1G80480.1	<i>CNOT2</i>	K12605	
Achn115661	维生素 B6 合成相关酶 PdxS	IPR001852	AT2G38230.1	<i>pdxS</i>	K06215	map00750
Achn116821	维生素 C 转运体	IPR006048	AT5G49990.1		K14611	
Achn119481	—	—	AT5G58560.1			
Achn119781	维生素 B6 合成相关酶 PdxS	IPR001852	AT5G01410.1	<i>pdxS</i>	K06215	map00750
Achn122581	—	—	AT1G23360.1	<i>COQ5</i>	K06127	map00130
Achn122891	维生素 C 转运体	IPR006049	AT1G60030.1		K14611	
Achn125081	维生素 C 转运体	IPR006050	AT2G34190.1		K14611	
Achn128531	—	IPR006968				
Achn136601	维生素 B6 激酶	IPR004625	AT5G37850.2			
Achn142361	—	IPR006968				
Achn148261	嘌呤转运体	IPR006051	AT3G10960.1	<i>pbuG</i>	K06901	
Achn149271	—	IPR002629	AT5G17920.2			map00270
Achn155031	—	—	AT4G26850.1	<i>VTC2</i>	K14190	

续表 2

猕猴桃基因编号	基因注释	IPR 编号	拟南芥同源 基因编号	拟南芥同源 基因名称	KO 编号	KEGG 途径
Achn156671	维生素 C 转运体	IPR006052	AT2G34190.1		K14611	
Achn157071	维生素 C 转运体	IPR006053	AT2G05760.1		K14611	
Achn157231	维生素 C 转运体	IPR006054				
Achn178181	—	IPR006968				
Achn191981	—	—	AT3G63410.1	<i>VTE3</i>	K12502	
Achn198991	维生素 B6 合成相关酶 PdxS	IPR001852	AT2G38230.1	<i>pdxS</i>	K06215	map00750
Achn200611	维生素 C 转运体	IPR006055	AT2G27810.3			
Achn200621	维生素 C 转运体	IPR006056	AT2G27810.2			
Achn207681	维生素 K 环氧化还原酶(前体)		AT4G35760.1			
Achn210521	—	IPR006968				
Achn212141	—	—	AT2G39770.2			
Achn224401	维生素 C 转运体	IPR006057				
Achn227781	维生素 C 转运体	IPR006058	AT2G05760.1		K14611	
Achn228621	维生素 B12 合成相关蛋白/P47K 家族蛋白	IPR011629				
Achn231571	—	—	AT1G22940.1	<i>thiDE</i>	K14153	
Achn256141	—	IPR002629	AT5G17920.2			map00270
Achn258021	—	—	AT2G39770.2			
Achn258461	嘌呤转运体	IPR006059				
Achn259241	—	IPR007373				map00730
Achn261091	—	IPR006968				
Achn275491	—	—	AT3G61530.2			
Achn276691	—	—	AT4G32770.1	<i>VTE1</i>	K09834	map00100
Achn280421	维生素 C 转运体	IPR006060				
Achn288611	—	—	AT1G22940.1	<i>thiDE</i>	K14153	
Achn288631	—	—	AT1G22940.1	<i>thiDE</i>	K14153	
Achn300301	—	—	AT3G61530.2			
Achn303541	维生素 B6 合成相关酶 PdxS	IPR001852	AT5G01410.1	<i>pdxS</i>	K06215	map00750
Achn305661	核碱基维生素 C 转运体	IPR006061	AT2G27810.3			
Achn309301	—	—	AT2G34470.1			
Achn321521	—	—	AT3G63410.1	<i>VTE3</i>	K12502	
Achn332441	核碱基维生素 C 转运体	IPR006062	AT4G38050.1		K14611	
Achn333301	—	—	AT5G54770.1	<i>thi1</i>	K03146	
Achn339231	—	—	AT4G26850.1	<i>VTC2</i>	K14190	
Achn343451	—	—	AT1G79600.1	<i>ADCK</i>	K08869	
Achn345681	维生素 C 转运体	IPR006063	AT2G34190.1		K14611	
Achn351791	—	IPR019258				
Achn357911	—	IPR007373				map00730
Achn364711	—	—	AT4G26850.1	<i>VTC2</i>	K14190	
Achn370901	维生素 C 转运体	IPR006064	AT5G49990.1		K14611	
Achn375821	嘌呤转运体	IPR006065				
Achn379291	—	IPR013215	AT5G17920.2			map00270
Achn390021	—	—	AT1G23360.3			
Achn390641	—	IPR006968				

a) 基因注释来自猕猴桃基因组草图中蛋白注释, IPR 编号为 InterProScan 结构域编号, KO 编号为 KEGG(Kyoto Encyclopedia of Genes and Genomes)数据库注释编号, 途径包含该基因归属的通路信息

表 3 猕猴桃维生素相关基因的可变剪接事件^{a)}

基因编号	叶	果实(不同时期)			基因编号	叶	果实(不同时期)		
		未成熟	半成熟	成熟			未成熟	半成熟	成熟
Achn005111	A5	RI	A3, RI	RI	Achn191981	A3, A5	—	—	—
Achn008171	A5, RI	RI	—	RI, SE	Achn200621	—	—	A5	—
Achn014721	—	—	—	RI	Achn207681	A5	A5	—	—
Achn032341	SE	SE	A3, SE	—	Achn210521	—	—	RI	—
Achn036641	A5	—	—	—	Achn212141	A3, RI	—	—	A5
Achn043541	A5, RI	A3, A5	—	RI	Achn228621	—	RI	A5	A5
Achn045051	A3	—	—	—	Achn231571	RI	—	—	RI
Achn053941	—	RI	—	—	Achn256141	A5, RI	A5	—	RI
Achn060421	A3, A5	—	—	A3, A5	Achn259241	RI	RI	RI	RI
Achn068181	A3	A3	A3	—	Achn261091	SE	A3	SE	A5, SE
Achn078251	A3, RI	—	—	A5	Achn276691	A3	—	A3, RI	—
Achn078621	RI	RI	RI	RI	Achn280421	A5	RI	A5, RI	A5
Achn087641	A3, A5, RI	—	—	A3, A5	Achn300301	A3	—	A3	—
Achn091471	A3	—	—	—	Achn305661	A5, RI, SE	—	RI	—
Achn095121	A3	—	A3	—	Achn309301	—	—	A5	—
Achn097841	—	—	—	SE	Achn321521	A3	A3	A3	—
Achn116821	A5, RI	RI	—	RI	Achn332441	—	RI	RI	—
Achn119481	—	A5	—	—	Achn339231	A3, RI	RI	RI	A5
Achn122581	A5	—	—	—	Achn345681	RI	A3	A3	—
Achn122891	—	—	A5	—	Achn357911	A5, SE	—	SE	RI, SE
Achn125081	—	—	RI	—	Achn370901	A5	—	—	—
Achn136601	A3	A3	—	A5	Achn379291	RI	RI	A5, RI	RI
Achn149271	SE	—	—	—	Achn390641	RI	RI	RI	—
Achn155031	A3, RI	A3	RI	RI, SE					

a) 可变剪接事件按类型缩写分别为: 内含子保留(IR)、3'端位点可变(A3)、5'端位点可变(A5)及外显子跳跃(ES)

参考文献

1 Arcus A C. Proteolytic enzyme of *Actinidia chinensis*. Biochim Biophys Acta, 1959, 33: 242–244

2 Cheng C H, Seal A G, Boldingh H L, et al. Inheritance of taste characters and fruit size and number in a diploid *Actinidia chinensis* (kiwifruit) population. Euphytica, 2004, 138: 185–195

3 Skinner M A, Loh J M S, Hunter D C, et al. Gold kiwifruit (*Actinidia chinensis* ‘Hort16A’) for immune support. Proc Nutr Soc, 2011, 70: 276–280

4 张菊明, 林佩芳. 中华猕猴桃多糖对巨噬细胞-T 细胞免疫介质的作用. 科技通报, 1990, 6: 284–286

5 宋文瑛, 许冠华, 张光霁. 猕猴桃根多糖对人胃癌 SGC-7901 细胞增殖、凋亡及 p-p38 表达的影响. 中国中西医结合杂志, 2014, 34: 329–333

6 Wang B B, Brendel V. Genomewide comparative analysis of alternative splicing in plants. Proc Natl Acad Sci USA, 2006, 103: 7175–7180

7 Huang S, Ding J, Deng D, et al. Draft genome of the kiwifruit *Actinidia chinensis*. Nat Commun, 2013, 4: 2640

8 Berget S M, Moore C, Sharp P A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci USA, 1977, 74: 3171–3175

9 Tong C, Wang X, Yu J, et al. Comprehensive analysis of RNA-seq data reveals the complexity of the transcriptome in *Brassica rapa*. BMC Genomics, 2013, 14: 689

10 Reddy A S. Alternative splicing of pre-messenger RNAs in plants in the genomic era. Annu Rev Plant Biol, 2007, 58: 267–294

11 Zhang G, Guo G, Hu X, et al. Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. Genome Res, 2010, 20: 646–654

12 Reddy A S, Marquez Y, Kalyna M, et al. Complexity of the alternative splicing landscape in plants. Plant Cell, 2013, 25: 3657–3683

13 Zhang X N, Mount S M. Two alternatively spliced isoforms of the *Arabidopsis* SR45 protein have distinct roles during normal plant development. Plant Physiol, 2009, 150: 1450–1458

- 14 Wang Q, Silver P A. Genome-wide RNAi screen discovers functional coupling of alternative splicing and cell cycle control to apoptosis regulation. *Cell Cycle*, 2010, 9: 4419–4421
- 15 Gu L, Guo R. Genome-wide detection and analysis of alternative splicing for nucleotide binding site-leucine-rich repeats sequences in rice. *J genet genomics*, 2007, 34: 247–257
- 16 Leviatan N, Alkan N, Leshkowitz D, et al. Genome-wide survey of cold stress regulated alternative splicing in *Arabidopsis thaliana* with tiling microarray. *PLoS One*, 2013, 8: e66511
- 17 Staiger D, Brown J W. Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell*, 2013, 25: 3640–3656
- 18 Quesada V, Macknight R, Dean C, et al. Autoregulation of FCA pre-mRNA processing controls *Arabidopsis* flowering time. *EMBO J*, 2003, 22: 3142–3152
- 19 Ding F, Cui P, Wang Z Y, et al. Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in *Arabidopsis*. *BMC Genomics*, 2014, 15: 431
- 20 Wang E T, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*, 2008, 456: 470–476
- 21 Filichkin S A, Priest H D, Givan S A, et al. Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res*, 2010, 20: 45–58
- 22 Goldberg D H, Victor J D, Gardner E P, et al. Spike train analysis toolkit: enabling wider application of information-theoretic techniques to neurophysiology. *Neuroinformatics*, 2009, 7: 165–178
- 23 Patel R K, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*, 2012, 7: e30619
- 24 Trapnell C, Pachter L, Salzberg S L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 2009, 25: 1105–1111
- 25 Quevillon E, Silventoinen V, Pillai S, et al. InterProScan: protein domains identifier. *Nucleic Acids Res*, 2005, 33: W116–W120
- 26 Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*, 2012, 7: 562–578
- 27 Rabbitts T H. Evidence for splicing of interrupted immunoglobulin variable and constant region sequences in nuclear RNA. *Nature*, 1978, 275: 291–296
- 28 Black D L. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*, 2003, 72: 291–336
- 29 Stamm S, Ben-Ari S, Rafalska I, et al. Function of alternative splicing. *Gene*, 2005, 344: 1–20
- 30 Lareau L F, Green R E, Bhatnagar R S, et al. The evolving roles of alternative splicing. *Curr Opin Struc Biol*, 2004, 14: 273–282
- 31 Xu Q, Modrek B, Lee C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res*, 2002, 30: 3754–3766
- 32 Transcriptome and genome conservation of alternative splicing events in humans and mice. In: Sugnet C W, Kent W J, Haussler D, eds. *Proc. 9th Pacific Symposium on Biocomputing*. Hawaii. 2004, Singapore: World Scientific, 2004. 66–77
- 33 Chacko E, Ranganathan S. Genome-wide analysis of alternative splicing in cow: implications in bovine as a model for human diseases. *BMC Genomics*, 2009, 10: S11
- 34 Babenko V N, Aitnazarov R B, Goncharov F A, et al. Alternative splicing landscape of the *Drosophila melanogaster* genome. *Russ J Genet*, 2010, 46: 1036–1038
- 35 Labadorf A, Link A, Rogers M F, et al. Genome-wide analysis of alternative splicing in *Chlamydomonas reinhardtii*. *BMC Genomics*, 2010, 11: 114
- 36 Ramani A K, Calarco J A, Pan Q, et al. Genome-wide analysis of alternative splicing in *Caenorhabditis elegans*. *Genome Res*, 2011, 21: 342–348
- 37 Sablok G, Gupta P K, Baek J M, et al. Genome-wide survey of alternative splicing in the grass *Brachypodium distachyon*: a emerging model biosystem for plant functional genomics. *Biotechnol Lett*, 2011, 33: 629–636
- 38 Bao H, Li E Y, Mansfield S D, et al. The developing xylem transcriptome and genome-wide analysis of alternative splicing in *Populus trichocarpa* (black cottonwood) populations. *BMC Genomics*, 2013, 14: 359
- 39 Panahi B, Abbaszadeh B, Taghizadeghan M, et al. Genome-wide survey of alternative splicing in *Sorghum Bicolor*. *Physiol Mol Biol Plants*, 2014, 20: 323–329
- 40 Wu H P, Su Y S, Chen H C, et al. Genome-wide analysis of light-regulated alternative splicing mediated by photoreceptors in *Physcomitrella patens*. *Genome Biol*, 2014, 15: R10
- 41 Shen Y, Zhou Z, Wang Z, et al. Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell*, 2014, 26: 996–1008
- 42 Li P, Ponnala L, Gandotra N, et al. The developmental dynamics of the maize leaf transcriptome. *Nat Genet*, 2010, 42: 1060–1067
- 43 Marquez Y, Brown J W, Simpson C, et al. Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Res*, 2012, 22: 1184–1195
- 44 Pan Q, Shai O, Lee L J, et al. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*, 2008, 40: 1413–1415

- 45 Walters B, Lum G, Sablok G, et al. Genome-wide landscape of alternative splicing events in *Brachypodium distachyon*. DNA Res, 2013, 20: 163–171
- 46 Baek J M, Han P, Iandolino A, et al. Characterization and comparison of intron structure and alternative splicing between *Medicago truncatula*, *Populus trichocarpa*, *Arabidopsis* and rice. Plant Mol Biol, 2008, 67: 499–510
- 47 Staiger C J. Signaling to the actin cytoskeleton in plants. Annu Rev Plant Physiol Plant Mol Biol, 2000, 51: 257–288
- 48 Kadota A, Wada M. Photoorientation of chloroplasts in protonemal cells of the fern *Adiantum* as analyzed by use of a video-tracking system. Bot Mag, 1992, 105: 265–279
- 49 Wada M, Kagawa T, Sato Y. Chloroplast movement. Annu Rev Plant Biol, 2003, 54: 455–468
- 50 Stock A M, Robinson V L, Goudreau P N. Two-component signal transduction. Annu Rev Biochem, 2000, 69: 183–215
- 51 吴家森, 管剑峰. 秦美猕猴桃果实生育及营养量变的若干特点. 浙江林学院学报, 2002, 19: 244–246
- 52 安华明, 樊卫国, 刘进平. 生育期猕猴桃果实中营养元素积累规律研究. 种子, 2003, 4: 24–25
- 53 陈昆松. ABA 和 IAA 对猕猴桃果实成熟进程的调控. 园艺学报, 1999, 26: 81–86
- 54 张玉, 陈昆松, 张上隆, 等. 猕猴桃果实采后成熟过程中糖代谢及其调节. 植物生理与分子生物学学报, 2004, 30: 317–324
- 55 涂正顺, 李华, 王华, 等. 猕猴桃果实采后香气成分的变化. 园艺学报, 2001, 28: 512–516
- 56 殷学仁, 张波, 李鲜, 等. 乙烯信号转导与果实成熟衰老的研究进展. 园艺学报, 2009, 36: 133–140

Genome-Wide Survey of Alternative Splicing in *Actinidia chinensis*

ZHOU YanQiu¹, QIN Rui¹, LIU ShengYi², LI Gang¹, HUANG JunYan², DONG CaiHua²,
CHENG XiaoHui², LIU Hong¹ & TONG ChaoBo²

*1 Hubei Provincial Key Laboratory for Protection and Application of Special Plant Germplasm in Wuling Area of China,
South-Central University for Nationalities, Wuhan 430074, China;*

*2 Oil Crops Research Institute of CAAS, Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture,
Wuhan 430062, China*

Alternative splicing can increase the diversity and complexity of proteome greatly through creating multiple mRNA transcripts from a single gene. RNA-seq data from different organs (leaf, immature fruit, half-ripe fruit and mature fruit) of *Actinidia chinensis* had been used to identify the alternative splicing events using the *A. chinensis* genome as reference, and totally 32180 alternative splicing events were discovered from the *A. chinensis* genome (39040 genes) corresponding to 11651 genes (29% of the total genome). Intron retention events showed the highest frequency (more than 50%) among all the different alternative splicing events. The frequency of events with alternative 3' splice sites was twice than the alternative 5' splice sites. GO enrichment analysis result showed that the alternative splicing gene was mainly enriched in enzyme regulation and nucleotide binding related GO categories. The alternative splicing events in different organs of *A. chinensis* seemed to preferentially occur in genes with important function. The genes of leaf were associated with the function of actin and microtubule. The genes of immature fruit decided the function of two-component signal system. The genes of half-ripe fruit were connected with phospholipid synthesis process. The genes in mature fruit were associated with signal transduction. In addition, 55.6% of vitamin related genes had been found to have alternative splicing changes, which were significantly higher than the average changes of the genome (the average splicing change level was 29.6%). Those explained that alternative splicing events played an important role in the process of vitamin synthesis. In this paper, through the genome-wide analysis of alternative splicing in *A. chinensis*, a powerful resource for understanding the complex genome of *A. chinensis* was provided. The result was also useful to molecular breeding in kiwifruit.

***Actinidia chinensis*, alternative splicing, genome, vitamin**

doi: 10.1360/N052014-00275