

三倍体香蕉优良品种‘Grand Nain’全基因组变异挖掘

邵秀红^{1,2,#}, 胡春华^{2,#}, 盛鸥², 毕方铨², 邓贵明², 杨乔松^{2,*}, 易干军^{2,*}

¹湖南农业大学园艺园林学院, 长沙410128

²广东省农业科学院果树研究所, 农业部南亚热带果树生物学与遗传资源利用重点实验室, 广东省热带亚热带果树研究重点实验室, 广州510640

摘要: 为全面揭示香蕉(*Musa* spp.)优良品种的全基因组变异类型, 本研究采用高通量重测序技术, 对当前香蕉主栽品种‘Grand Nain’(AAA group, Cavendish subgroup)开展全基因组重测序, 测序深度53.79 X。与野生香蕉‘DH-Pahang’参考基因组比对, 共检测到4 598 633个单核苷酸多态位点(SNP), 484 752个小片段插入缺失位点(Indel), 57 047个结构变异(SV), 共导致36 277个基因变异; 代谢通路分析(KEGG)发现, 植物激素信号转导途径相关基因的变异最多, 存在442个基因变异, 其中乙烯合成和信号转导途径中1-氨基环丙烷-1-羧酸氧化酶基因(*ACO*)、1-氨基环丙烷-1-羧酸合成酶基因(*ACS*)、*EIN3/EILs*、*ERS*、*RAN*、*EBF*、*EIN2*等基因都存在变异。特别是序列分析发现‘Grand Nain’中的*MaACO1*基因与参考基因组相比存在2个变异位点并导致氨基酸的突变, 且在A和B基因组中*MaACO1*基因存在2个相邻的变异位点。本研究为香蕉贮藏保鲜等相关分子标记开发、基因功能研究, 以及基于基因组编辑技术的香蕉遗传改良提供依据。

关键词: 香蕉; 全基因组重测序; 单核苷酸多态; 小片段插入缺失; 结构变异; 基因变异

香蕉(*Musa* spp.)为芭蕉科(Musaceae)芭蕉属植物, 起源于东南亚, 在世界范围内130多个热带、亚热带国家广泛种植(Dash和Rai 2016)。香蕉既是重要经济作物, 也是仅次于水稻(*Oryza sativa*)、小麦(*Triticum aestivum*)、玉米(*Zea mays*)之后的第四大粮食作物; 中国是世界香蕉原产地之一, 有3 000多年种植历史, 2014年我国香蕉种植面积为39.55万hm², 产量1 179.19万t, 是世界上第二大香蕉生产国(艾建安等2016)。

香蕉是典型的呼吸跃变型果实, 香蕉的成熟和衰老受一系列发育信号、环境信号、激素代谢等调控, 其中乙烯的合成代谢在香蕉果实成熟过程中扮演了重要角色(Iqbal等2017)。1-氨基环丙烷-1-羧酸合成酶(1-aminocyclopropane-1-carboxylate synthase, *ACS*)和1-氨基环丙烷-1-羧酸氧化酶(1-aminocyclopropane-1-carboxylate oxidase, *ACO*)是乙烯合成关键酶, 国内外学者先后克隆了香蕉*ACO*和*ACS*基因(王新力等2000; 金志强等2002; 黄永红等2006; Huang等2006; Jourda等2014; 董晨等2017); 胡春华等(2015)获得了*ACO*反义基因转化的‘夫人指蕉’(AAA group), 采后常温条件下比对照推迟10 d成熟; Jourda等(2014)发现香蕉基因组中有11个*MaACS*基因和12个*MaACO*基因, 并利用转录组结合实时荧光定量PCR (quantitative real-time PCR, qPCR)分析表明*MaACO1*和*MaACS1*在香蕉成熟过

程中发挥重要作用; *MaERFs*、*MaMADS5*、*MaWRKY31*参与*MaACO*和*MaACS*基因的调控进而参与香蕉的果实成熟过程(Choudhury等2012; Xiao等2013; 范中奇等2017)。乙烯的生物学效应是在乙烯合成后通过乙烯信号转导途径过程中的乙烯受体、CTR (Constitutive Triple Response 1)、EIN2 (Ethylene Insensitive 2)、EIN3/EILs、AP2/ERF (Ethylene Response Factor)转录因子等元件来实现的(徐倩等2014; 范中奇等2015), 其中乙烯受体和CTR1均为负调控元件(徐倩等2014), 而下游的EIN2、EIN3/EILs和AP2/ERF等都是乙烯信号的正调控元件(范中奇等2015), 其中EIN3/EILs和AP2/ERF由两类植物特异的转录因子基因家族编码, 转录调控乙烯响应基因, 是目前乙烯生物学研究领

收稿 2017-09-11 修定 2018-03-14

资助 国家自然科学基金(31372018、31772267和31601720)、农业部“948”项目(2011-G16和2016-X22)、国家香蕉产业技术体系项目(CARS-32-01)、公益性行业(农业)科研专项经费(201403036)、广东省省级科技计划项目(2012-A020100007、2014B050502007、2016B020233002和2017A030303044)、“十二五”农村领域国家科技计划项目(2013BAD02B02)和广州市科技计划项目(2017-07010154)。

并列第一作者。

* 共同通讯作者: 易干军(yiganjun@vip.163.com)、杨乔松(soyang@hotmail.com)。

域的重点(Shan等2012; Lakhwani等2016; Iqbal等2017)。EIN3/EILs可激活下游乙烯响应基因, Jourda等(2014)发现香蕉中有19个EIN3/EILs成员, 在果实发育过程和外源乙烯处理后也呈现不同的表达模式; Mbéguié-A-Mbéguié等(2008)从香蕉果实中分离了5个EIN3-like基因, qPCR结果表明*MaEIL2*在果实成熟进程中上调表达, 并且能够被外源乙烯诱导表达, *MaEIL1/4*呈下调表达; *MaEIL5*则在果实成熟中显著下调, 并且可以与F-box蛋白MaEBF2互作, 形成蛋白复合体(Kuang等2013)。这些结果表明不同EIN3/EIL成员调控果实成熟的机制可能有差异。Lakhwani等(2016)在小果野蕉(*M. acuminata*, AA, $2n=2x=22$)和野蕉(*M. balbisiana*, BB, $2n=2x=22$)中分别鉴定出265和318个AP2/ERF转录因子, 主要归类为ERF、DREB、AP2、RAV和Soloist亚家族。*MaERF9*和*MaERF11*与香蕉果实成熟密切相关, 其中*MaERF9*作为转录激活子激活*MaACO1*的启动子, 而*MaERF11*作为转录抑制子抑制*MaACS1*和*MaACO1*启动子的活性, 同时利用酵母双杂交和双分子荧光互补(bimolecular fluorescence complementation, BiFC)证明了*MaERF9*和*MaERF11*在蛋白水平上与*MaACO1*存在相互作用, 共同调控乙烯相关基因的表达, 影响乙烯的合成, 从而共同调控香蕉果实的成熟过程(Xiao等2013)。

随着香蕉基因组信息的发布和完善, AA基因组(小果野蕉‘DH-Pahang’; D’Hont等2012)、BB基因组(野蕉‘Pisang Klutuk Wulung’; Davey等2013)和AB基因组(阿宽蕉, *M. itinerans*; Wu等2016)提供了大量的基因资源, 结合果实成熟过程中转录组(Asif等2014)、蛋白组(Toledo等2012; Esteve等2013)、代谢组(Dong等2016; Pandey等2016)等技术的应用, 为香蕉采后保鲜分子机理研究提供了理论支撑。然而复杂三倍体(AAA)特别是优良品种基因组的研究尚未见报道, ‘Grand Nain’ (‘大矮蕉’)具有植株中秆偏短、果梳整齐、商品率高等优点, 是当前市场主栽品种, 占全世界香蕉栽培面积的70% (冯慧敏等2016; Dash和Rai 2016)。因此, 本研究对三倍体香蕉优良品种‘Grand Nain’进行了全基因组重测序及变异分析, 并解析了乙烯合成代谢途径基因的变异, 以期对香蕉贮藏保鲜等相关性状的遗传改良提供理论依据。

1 材料与方法

1.1 植物材料

三倍体($2n=3x=33$)香蕉(*Musa* sp.)品种‘Grand Nain’ (AAA group, Cavendish subgroup; 简称GN)胚性细胞悬浮系由本实验室建立并保存。

1.2 胚性悬浮细胞DNA的提取

吸取适量胚性悬浮细胞, 吸除多余的悬浮培养液后, 立即用液氮速冻, 参照Michiels等(2003)的方法利用十六烷基三甲基溴化铵(CTAB)法提取基因组DNA, 并利用NanoDrop ND-2000c超微量分光光度计和琼脂糖凝胶电泳鉴定DNA浓度和质量, 送北京奥维森基因科技有限公司进行全基因组重测序。

1.3 ‘Grand Nain’全基因组重测序

质检合格的DNA利用超声波将其片段化, 然后对片段化的DNA进行片段纯化、3’端加A、连接测序接头等末端修饰, 再通过琼脂糖凝胶电泳选择片段大小, 进行PCR扩增形成测序文库, 质检合格的文库用Illumina HiSeq™ 2500测序。对测序得到的原始reads进行质量评估并过滤得到clean reads, 用于后续生物信息学分析。

1.4 基因组变异检测与注释

生物信息学分析工作由北京诺禾致源科技股份有限公司开展, 将测序原始序列(raw reads)进行质控得到的干净数据(clean data)通过Burrows-Wheeler Alignment软件(BWA; 参数: mem-t4 -k32 -M)比对到参考基因组上(小果野蕉‘DH-Pahang’), 比对结果经Sequence Alignment/Map tools (参数: rmdup)去除重复; 根据比对结果, 采用SAMtools (mpileup -m2 -F0.002 -d1000)进行单核苷酸多态性(single nucleotide polymorphism, SNP)检测(Li等2009), 为降低SNP检测错误率, 选用如下标准进行过滤: (1) SNP的reads支持数不低于4; (2) SNP的质量值(MQ)不低于20, 利用ANNOVAR软件对SNP检测结果进行注释(Wang等2010)。利用SAMtools (mpileup -m2 -F0.002 -d1000)检测长度小于50 bp的小片段插入与缺失(insertion-deletion, Indel), 然后用ANNOVAR软件对检测出的Indel进行注释。利用BreakDancer软件, 基于pair-end reads比对到参考基因组上面的关系及实际insert size大小检测样品与参考基因组间的插入(insertion, INS)、缺失(deletion, DEL)、倒置(inversion, INV)、染色体内部迁

移(intra-chromosomal translocation, ITX)和染色体间的迁移(inter-chromosomal translocation, CTX) (Chen等2009)。过滤去掉pair-end reads支持数小于2的结构变异(structural variation, SV)结果, 并且对其中的DEL、INS、INV通过ANNOVAR进行变异注释。通过CNVnator (参数: -call 100)进行检测 (Abyzov等2011), 即通过基因组上不同的reads覆盖深度, 判断潜在的删除(deletion)和重复(duplication), 并通过ANNOVAR进行变异注释。

1.5 DNA水平变异基因分析

通过检索参考基因组与‘Grand Nain’间发生非同义突变SNP、编码序列(coding sequence, CDS)区发生的Indel及SV的基因, 寻找‘Grand Nain’与参考基因组之间可能存在功能差异的基因, 并利用Blast将变异基因与基因本体分析(gene ontology, GO; Conesa和Götz 2008)和代谢通路分析(Kyoto Encyclopedia of Genes and Genomes, KEGG; Mao

等2005)数据库进行比对, 得到这些基因的注释用于后续基因功能的分析。

2 实验结果

2.1 ‘Grand Nain’全基因组重测序结果

为揭示GN基因组信息, 本研究利用Illumina HiSeq™ 2500对GN总DNA开展全基因组重测序, 总共获得175 619 306个有效测序数据, 除去带接头的或低质量的reads后共有160 565 648个reads比对到香蕉参考基因组上, 其中参考基因组大小为472 960 417 bp, GN的比率为91.43%, 对参考基因组(排除N区)的平均覆盖深度为53.79 X, 1 X覆盖度(至少有一个碱基的覆盖)在97.41%以上(表1)。根据染色体各位点的覆盖度情况作图(图1-A), 可见覆盖深度在染色体上的分布比较均匀, 说明测序随机性较好, 其中深度不均一的位置可能由PCR的偏好性或重复序列导致。根据变异检测分析, 共

表1 ‘Grand Nain’测序数据统计

Table 1 Summary of sequencing data of ‘Grand Nain’

比对到基因组reads条数	有效测序数据reads总条数	比率为/%	平均测序深度/X	覆盖度/%
160 565 648	175 619 306	91.43	53.79	97.41

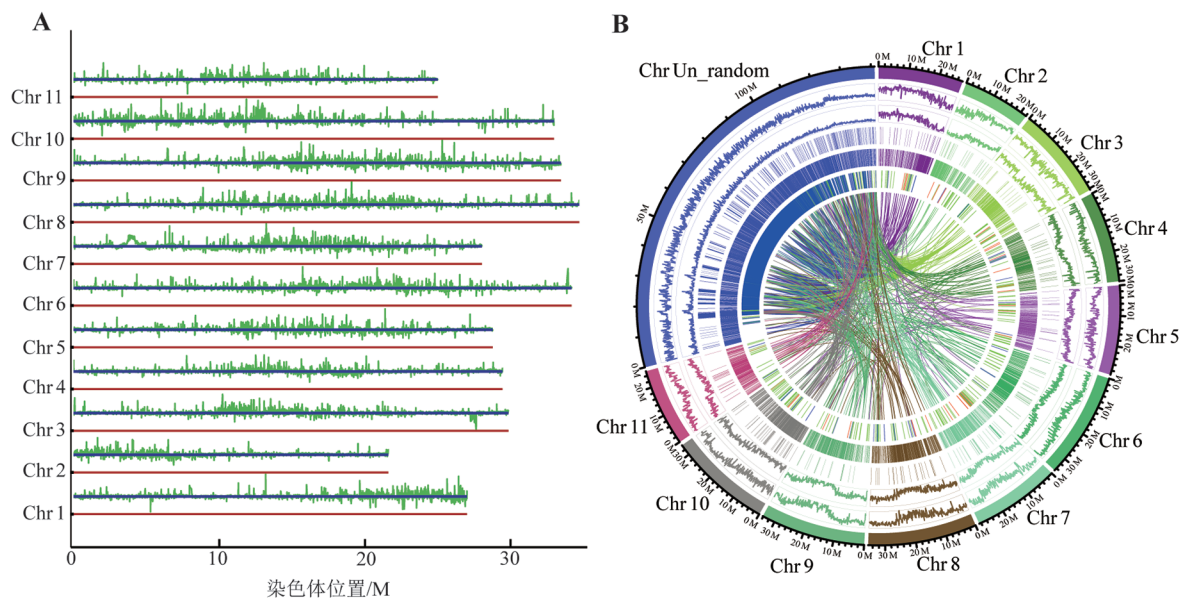


图1 ‘Grand Nain’测序数据分析

Fig.1 Analysis of sequencing data of ‘Grand Nain’

A: ‘Grand Nain’染色体覆盖图, 其中Chr 1~11为染色体编号。B: ‘Grand Nain’各类变异在染色体上分布图(Circos图), 其中ChrUn_random为0号染色体; Circos图从外到内依次是染色体, SNP, Indel, CNV重复(duplication)与CNV缺失(deletion), SV插入(insertion)、SV缺失(deletion)与SV倒置(inversion), SV ITX与SV CTX, 连线部分为SV ITX。

获得4 598 633个SNP数、484 752个Indel、57 047个SV和14 131个拷贝数变异(copy-number variation, CNV), 变异检测分布具体见图1-B; 由于参考基因组为A基因组二倍体(AA), 而GN为A基因组三倍体(AAA), 通过图1-B可以看出, 有部分数据未能匹配到染色体上。

2.2 ‘Grand Nain’基因组SNP检测与注释

为揭示GN的单核苷酸变异信息, 本研究采用SAMtools软件对GN基因组DNA进行SNP检测, 共获得4 598 633个SNP, 其中以转换(transitions, Ts)类型的SNP (2 841 060个)为主, 占SNP总数的61.78%, 颠换(transversions, Tv)类型的SNP (1 757 573个)仅占38.22%, 转换与颠换的比率(Ts/Tv)为1.616; 另一方面, 杂合型(heterozygosity) SNP (3 139 642个)为主, 占总SNP数的68.27%, 而纯合型(homozygosity) SNP (1 458 991个)仅占31.73%, 全基因组SNP杂合率(杂合SNP个数/基因组大小)为8.007% (表2)。以上结果表明GN的SNP变异主要以转换为主, 且杂

合型SNP数目是纯合型的2.15倍。后续通过ANNOVAR软件对SNP检测结果进行注释(表3), 发现基因间区(1 983 595个)和内含子区(1 410 005个)存在主要的SNP变异, 约占SNP总变异数目的73.80%; 发生在CDS区域的SNP (421 171个)仅占SNP变异总数的9.16%, 其中同义突变211 452个, 非同义突变206 269个, 这些SNP共导致32 760个基因发生了突变(表4)。

2.3 ‘Grand Nain’基因组small Indel检测与注释

为探索GN品种的小片段的插入和缺失信息, 本研究利用SAMtools软件对GN基因组DNA进行了Indel检测(表5), 在GN基因组中共发现484 752个Indel, 其中基因间区(181 553个)和内含子区(158 323个)约占基因组Indel总变异的70.11%, 而CDS区Indel数(13 678个)仅占基因组Indel总数的2.82%; 在CDS区域存在插入突变(6 219个)和缺失突变(7 459个), 其中CDS杂合突变为9 606个, 纯合突变为4 072个; 利用ANNOVAR软件对全基因组Indel进

表2 ‘Grand Nain’的SNP统计

Table 2 Summary of SNPs detected in ‘Grand Nain’

SNP数	Ts类型SNP数	Tv类型SNP数	Ts/Tv	杂合型SNP数	纯合型SNP数	全基因组SNP杂合率/%
4 598 633	2 841 060	1 757 573	1.616	3 139 642	1 458 991	8.007

表3 SNP注释结果

Table 3 SNP annotations

类型	SNP数	区域
基因间区	1 983 595	
内含子	1 410 005	
基因上游区域(1 kb以内)	364 210	
基因下游区域(1 kb以内)	351 572	
基因的5'非翻译区(UTR)内	18 152	
基因的3'UTR内	82 383	
基因上游1 kb区域同时在另一基因的下游1 kb区域	65 941	
剪切位点区域(内含子中靠近外显子/内含子边界的2 bp)	2 339	
同义突变	211 452	CDS
非同义突变	206 269	CDS
终止密码子获得	2 533	CDS
终止密码子丢失	717	CDS

表4 ‘Grand Nain’变异基因统计

Table 4 Summary of gene variations in ‘Grand Nain’

发生非同义SNP基因数	发生small Indel的基因数	发生SV变异基因数	变异基因总数
32 760	35 513	13 614	36 277

表5 ‘Grand Nain’的small Indel统计

Table 5 Summary of small Indels detected in ‘Grand Nain’

CDS插入数	CDS缺失数	CDS杂合Indel数	CDS纯合Indel数	CDS Indel总数	基因组Indel总数
6 219	7 459	9 606	4 072	13 678	484 752

行注释(表6), 发现CDS区域的Indel主要有密码子插入(3 360个)、密码子删除(3 664个)、移码突变(6 233个)、终止子的获得(214个)和丢失(62个)等类型, 这些Indel共导致35 513个基因突变(表4)。根据GN在CDS区和全基因组范围的Indel长度进行统计(图2), 发现插入类型(≥ 1)和缺失类型(≤ -1)在CDS区分别有6 401和5 829个, 在CDS区域存在较多的插入1个碱基(1 528个)和3个碱基(1 806个)类型的突变, 而在基因组范围中插入1个碱基(100 060个)和缺失1个碱基(97 886个)的突变类型较多, 分别占基因组Indel总变异数的20.64%和20.19%, 插入3个碱基(24 016个)和缺失3个碱基(25 475个)类型的突变相对较少。

2.4 ‘Grand Nain’基因组SV检测与注释

为明确GN品种大片段的插入、缺失、倒置、易位等序列信息, 本研究利用BreakDancer软件开展GN基因组SV的检测, 共检测到57 047个SV。由表7可见, SV的变异主要以缺失(22 661个)为主, 占SV总数的39.72%; 插入(9 695个)和反转(10 419个)分别占16.99%和18.26%, 染色体内易位(7 737个)及染色体间易位(6 535个)则相对较少, 仅

表6 Small Indel注释结果

Table 6 Small Indel annotations

类型	Indel数	区域
基因间区	181 553	
内含子	158 323	
基因上游区域(1 kb以内)	58 448	
基因下游区域(1 kb以内)	60 983	
基因的5'UTR内	2 259	
基因的3'UTR内	14 459	
剪切位点区域(内含子中靠近外显子/内含子边界的2 bp)	391	
基因上游1 kb区域同时在另一基因的下流1 kb区域	11 198	
密码子插入(3的整倍数)	3 360	CDS
密码子删除(3的整倍数)	3 664	CDS
移码突变(非3的整倍数插入)	2 859	CDS
移码突变(非3的整倍数删除)	3 374	CDS
终止密码子获得	214	CDS
终止密码子丢失	62	CDS
其他无法得到准确判断的突变	170	
总的插入	235 335	
总的删除	249 417	

占SV总数的13.56%和11.46%; 对插入、缺失、反转这三种类型进一步分析表明(表8), 外显子区域存在16 927个SV, 其中反转(9 773个)最多, 占

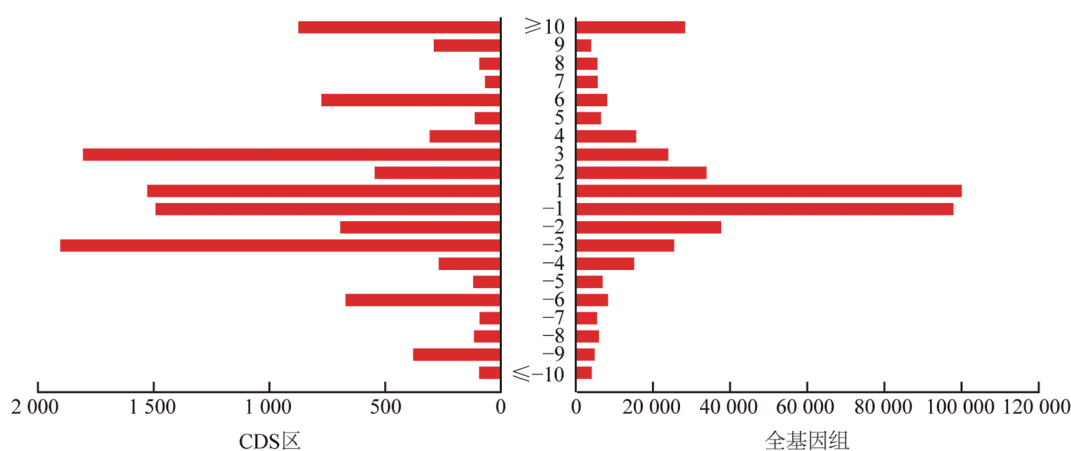


图2 全基因组和CDS区Indel长度分布图

Fig.2 Length distribution of Indels in whole genome and CDS

横坐标表示Indel变异数; 纵坐标上的正整数(x)分别表示插入碱基数(x bp), 负整数(-y)分别表示缺失碱基数(y bp)。

表7 ‘Grand Nain’的SV统计

Table 7 Summary of SVs detected in ‘Grand Nain’

插入数	缺失数	反转数	染色体内 易位数	染色体间 易位数	SV总数
9 695	22 661	10 419	7 737	6 535	57 047

表8 部分SV类型

Table 8 Classification of partial SVs

类型	外显子区SV数	内含子区SV数	基因间区SV数
插入	1 011	1 012	6 610
缺失	6 143	4 407	8 118
反转	9 773	112	424

57.74%，缺失(6 143个)次之，插入(1 011个)最少；内含子区域含有5 531个SV，主要为缺失(4 407个，79.68%)类型，插入(1 012个，18.30%)次之，反转(112个)类型最少，仅占2.02%；基因间区存在15 152个SV，与内含子中的SV类型相似，以缺失(8 118个，53.58%)和插入(6 610个，43.62%)为主，反转(424个，2.80%)最少；综上所述，在基因组、内含子及基因间区缺失是GN的主要变异类型，这些SV共导致13 614个基因发生变异(表4)。

2.5 ‘Grand Nain’基因组CNV检测与注释

CNV指基因组片段的拷贝数增加或者减少。为挖掘GN品种的CNV信息，本研究通过CNVnator软件检测GN全基因组(Abyzov等2011)，发现GN共有14 131个CNV(表9)，这些变异主要存在于基因间区(9 170个)和外显子区(3 291个)，分别占总数的64.89%和23.29%；基因上游1 kb区域(688个)和基因下游1 kb区域(545个)变异数目较少；而同时存在

表9 ‘Grand Nain’的CNV统计

Table 9 Summary of CNVs detected in ‘Grand Nain’

类型	CNV数
基因上游1 kb区域	688
外显子	3 291
内含子	262
基因下游1 kb区域	545
基因上游1 kb区域同时在另一个基因的下游1 kb区域	175
基因间区	9 170
拷贝数增加	4 109
拷贝数减少	10 022
总计	14 131

于基因上游1 kb和另一个基因下游1 kb范围内的变异最少，为175个，仅占总数的1.24%。这些变异共导致4 109个拷贝数的增加及10 022个拷贝数的减少。

2.6 ‘Grand Nain’在DNA水平变异基因分析

通过寻找参考基因组与GN基因组间发生非同义突变SNP、CDS区发生的Indel与SV的基因，发现与参考基因组相比，GN基因组共存在36 277个基因变异，其中发生非同义SNP突变的基因32 760个，发生Indel的基因35 513个，发生SV的基因13 614个(表4)，DNA水平变异基因分析的结果表明，GN基因组中存在多个基因同时存在多种类型的突变。为进一步挖掘这些变异基因的功能，本研究对发生突变的基因开展GO分析(图3)，结果表明参与生物学过程(biological process)中细胞过程(cellular process)的有13 798个基因，占总变异基因数的38.04%；代谢过程(metabolic process)中有13 403个基因，占总变异基因数的36.95%；而细胞组分(cellular component)中变异基因数目较少，其中该过程中数目最多的为细胞(cell)及膜(membrane)，均有6 681个基因，占总变异基因数的18.42%；在分子功能(molecular function)过程中两个数量较多的为催化活性(catalytic activity, 11 671个)和结合(binding, 14 005个)，分别占总变异基因数的32.17%和38.61%(图3)，GO分析结果表明相同的变异基因可能参与不同的生物学过程。进一步对发生变异的基因开展KEGG富集分析(图4)，发现植物激素信号转导途径(plant hormone signal transduction)和植物病原菌互作(plant-pathogen interaction)分别有442和216个基因存在变异，这些基因的变异为更好地适应环境条件和风味提供基因基础。

2.7 乙烯合成与代谢信号通路变异基因分析

香蕉果实的成熟主要受乙烯调控，进一步分析发现(表10)，在植物激素信号转导途径中，控制乙烯合成的酶包括ACO和ACS两个关键酶，以及乙烯信号转导途径中的乙烯受体、EILs、RTH、RAN、ERS(ethylene response sensor)、EBF、EIN、CTR等元件，且均由多基因家族编码共同调控乙烯的合成与代谢(徐倩等2014)。本研究对乙烯调控途径基因的变异分析发现，12个MaACO及14个MaACS基因中，分别有8个基因发生了1~4个非同义突变，其

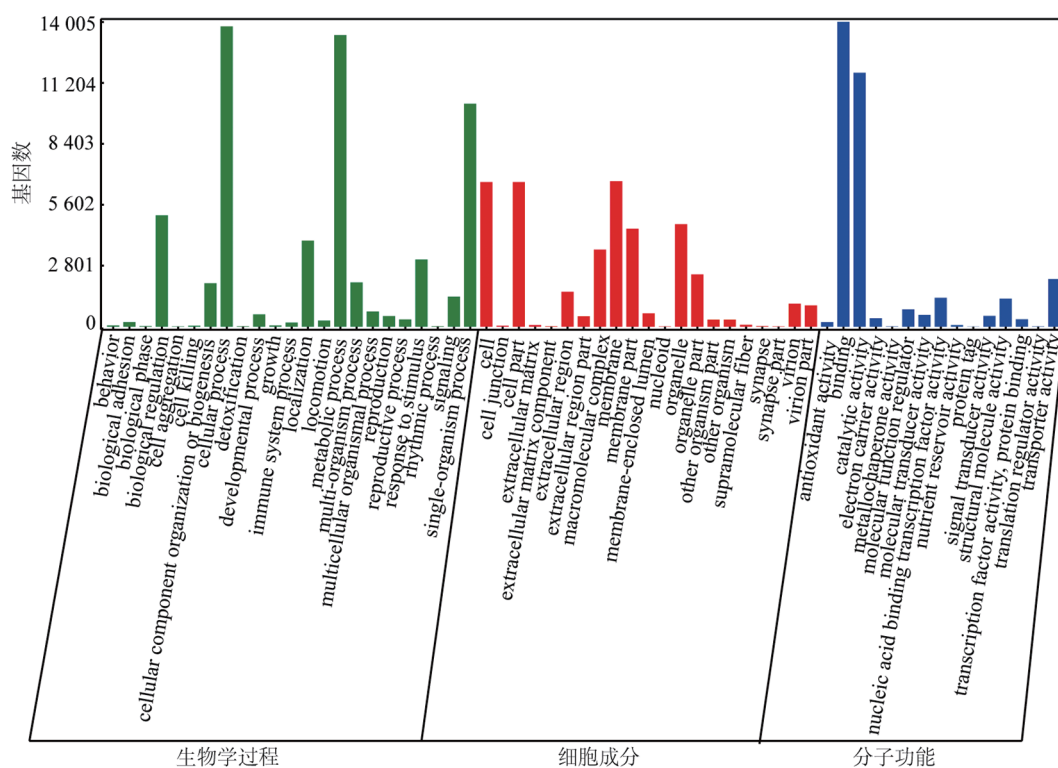


图3 变异基因GO注释分类

Fig.3 Classification of gene variations compared with GO database

横坐标为GO三个大类的下一层级的GO term, 纵坐标为注释到该term下(包括该term的子term)的差异基因个数。横坐标变异基因GO注释如下: behavior (反应)、biological adhesion (生物附着)、biological phase (生物相)、biological regulation (生物调节)、cell aggregation (细胞凝集)、cell killing (细胞杀伤)、cellular component organization or biogenesis (细胞成分组成或生物合成)、cellular process (细胞过程)、detoxification (解毒作用)、developmental process (发育过程)、growth (生长)、immune system process (免疫系统过程)、localization (定位)、locomotion (移动)、metabolic process (代谢过程)、multi-organism process (多肌体过程)、multicellular organismal process (多细胞生物过程)、reproduction (繁殖)、reproductive process (繁殖过程)、response to stimulus (刺激反应)、rhythmic process (节律调节过程)、signaling (信号)、single-organism process (单组织过程); cell (细胞)、cell junction (细胞连接)、cell part (细胞部分)、extracellular matrix (细胞外基质)、extracellular matrix component (细胞外基质成分)、extracellular region part (细胞外区域部分)、macromolecular complex (高分子复合物)、membrane (膜)、membrane part (膜部分)、membrane-enclosed lumen (膜包围内腔)、nucleoid (细胞核)、organelle (细胞器)、organelle part (细胞器部分)、other organism part (其他有机部分)、other organism (其他有机体)、supramolecular fiber (超分子纤维)、synapse (突触)、synapse part (突触部分)、virion (病毒体)、virion part (病毒体部分); antioxidant activity (抗氧化活性)、binding (结合)、catalytic activity (催化活性)、electron carrier activity (电子载体活性)、metallochaperone activity (金属伴侣活性)、molecular function regulator (分子功能调节)、molecular transducer activity (分子传感器活性)、nucleic acid binding transcription factor activity (核酸结合转录因子活性)、nutrient reservoir activity (营养受体活性)、protein tag (蛋白标签)、signal transducer activity (信号转导活性)、structural molecule activity (结构分子活性)、transcription factor activity、protein binding (蛋白结合转录因子活性)、translation regulator activity (转录调节因子活性)、transporter activity (运输活性)。

中*MaACSI*发生了密码子删除; 在17个*MaEIL*成员中有10个基因发生了非同义突变和密码子删除, 其中*MaEIL8*基因中存在11个非同义突变; *MaEBF5*和*MaEBF1*基因中存在移码突变(表11)。通过对调控乙烯合成的关键酶*MaACO1*基因序列分析发现, 在来源于不同的材料中均具有A/C和C/A的序列突变, 从而导致氨基酸的序列由谷氨酸(E)变异为丙氨酸(A)及天冬氨酸(D)变异为谷氨酸(E)(图5)。

3 讨论

香蕉基因组测序的完成以及高通量测序技术的发展, 使得在全基因组水平开展香蕉种质资源的变异分析成为可能(D'Hont等2012; Davey等2013; Wu等2016)。因栽培香蕉多为三倍体作物, 常规育种手段很难有效运用, 因此通过基因组的变异分析, 从全基因组层面鉴定和研究基因家族的分

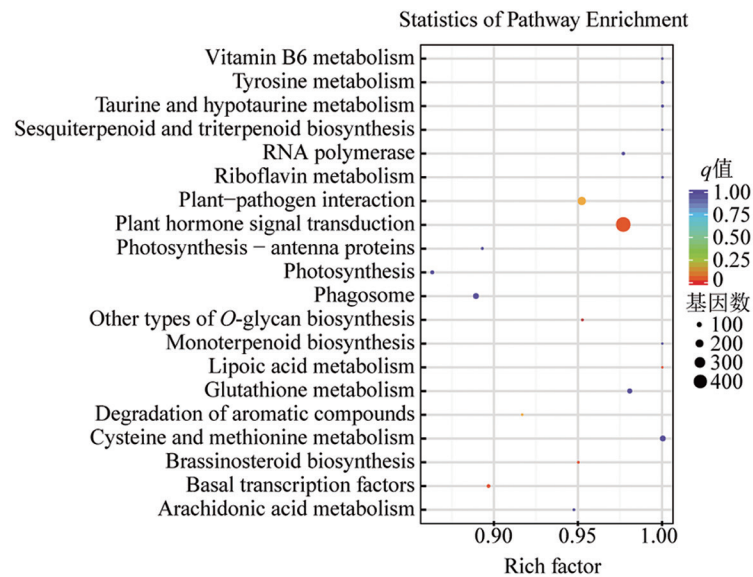


图4 变异基因KEGG注释分类

Fig.4 Classification of gene variations compared with KEGG database

横坐标Rich factor指差异表达的基因中位于该pathway条目的基因数与注释基因中位于该pathway条目的基因总数的比值。 q 值是作过多重假设检验校正之后的 P 值, q 值的取值范围为[0, 1], 越接近于0, 表示富集越显著; q 值的大小用点的颜色来表示, q 值越小则颜色越接近红色, 每个pathway下包含的差异基因的多少用点的大小来表示。纵坐标代谢途径分别为: Vitamin B6 metabolism (维生素B6代谢)、Tyrosine metabolism (酪氨酸代谢)、Taurine and hypotaurine metabolism (牛磺酸和亚牛磺酸代谢)、Sesquiterpenoid and triterpenoid biosynthesis (类倍半萜和三萜生物合成)、RNA polymerase (RNA聚合酶)、Riboflavin metabolism (核黄素合成)、Plant-pathogen interaction (植物病原菌互作)、Plant hormone signal transduction (植物激素信号转导)、Photosynthesis-antenna proteins (光合作用天线蛋白)、Photosynthesis (光合作用)、Phagosome (吞噬体)、Other types of *O*-glycan biosynthesis (其他类型 O 聚糖合成)、Monoterpenoid biosynthesis (类单萜生物合成)、Lipoic acid metabolism (硫辛酸代谢)、Glutathione metabolism (谷胱甘肽代谢)、Degradation of aromatic compounds (芳香族化合物降解)、Cysteine and methionine metabolism (半胱氨酸和蛋氨酸代谢)、Brassinosteroid biosynthesis (油菜素内脂生物合成)、Basal transcription factors (基础转录因子)、Arachidonic acid metabolism (花生四烯酸代谢)。

表10 乙烯合成与代谢信号通路相关变异基因

Table 10 Ethylene synthesis and metabolism signaling pathway-related gene variations

基因名称	注释	基因名称	注释
GSMUA_Achr7G18420_001	<i>MaACO1</i>	GSMUA_Achr11G25120_001	<i>MaEIL8</i>
GSMUA_Achr1G18250_001	<i>MaACO4</i>	GSMUA_Achr8G21550_001	<i>MaEIL9</i>
GSMUA_Achr5G09690_001	<i>MaACO5</i>	GSMUA_Achr3G09440_001	<i>MaEIL10</i>
GSMUA_Achr6G00870_001	<i>MaACO6</i>	GSMUA_Achr3G11160_001	<i>MaEIL11</i>
GSMUA_Achr6G12910_001	<i>MaACO7</i>	GSMUA_Achr8G25580_001	<i>MaEIL12</i>
GSMUA_Achr10G16070_001	<i>MaACO8</i>	GSMUA_Achr7G21630_001	<i>MaEIL13</i>
GSMUA_AchrUn_randomG20420_001	<i>MaACO9</i>	GSMUA_Achr6G23310_001	<i>MaEIL15</i>
GSMUA_AchrUn_randomG20440_001	<i>MaACO11</i>	GSMUA_Achr5T26800	<i>MaRAN1</i>
GSMUA_Achr4G29150_001	<i>MaACS1</i>	GSMUA_Achr8T06970	<i>MaRAN2</i>
GSMUA_Achr2G08600_001	<i>MaACS2</i>	GSMUA_Achr11G20580_001	<i>MaERS3</i>
GSMUA_Achr5G08880_001	<i>MaACS5</i>	GSMUA_Achr11G02140_001	<i>MaERS5</i>
GSMUA_AchrUn_randomG12290_001	<i>MaACS8</i>	GSMUA_Achr8G14350_001	<i>MaERS4</i>
GSMUA_Achr4G24930_001	<i>MaACS9</i>	GSMUA_Achr10G26950_001	<i>MaEBF5</i>
GSMUA_Achr4G30870_001	<i>MaACS10</i>	GSMUA_Achr2G15920_001	<i>MaEBF1</i>
GSMUA_Achr10G27570_001	<i>MaACS11</i>	GSMUA_Achr4G30680_001	<i>MaEBF2</i>
GSMUA_Achr1T18880	<i>MaACS14</i>	GSMUA_Achr6G05220_001	<i>MaEBF3</i>
GSMUA_Achr9G17080_001	<i>MaEIL5</i>	GSMUA_Achr9G28510_001	<i>MaEBF4</i>
GSMUA_Achr2G11540_001	<i>MaEIL6</i>	GSMUA_Achr6G31160_001	<i>EIN2-1</i>
GSMUA_Achr5G27670_001	<i>MaEIL7</i>	GSMUA_Achr7G01280_001	<i>EIN2-2</i>

表11 乙烯合成与代谢信号通路基因变异汇总

Table 11 Summary of ethylene synthesis and metabolism signaling pathway-related gene variations

基因名称	变异类型	变异信息
GSMUA_Achr7G18420_001	非同义突变	21 215 336 bp, A/C; 21 215 533 bp, C/A; 21 216 362 bp, A/G
GSMUA_Achr1G18250_001	非同义突变	13 635 748 bp, A/G; 13 635 989 bp, C/G; 13 636 237 bp, C/T
GSMUA_Achr5G09690_001	非同义突变	6 911 752 bp, A/G; 6 913 855 bp, T/C; 6 914 086 bp, A/G
GSMUA_Achr6G00870_001	非同义突变	598 985 bp, A/G; 599 068 bp, C/T
GSMUA_Achr6G12910_001	非同义突变	8 623 660 bp, G/T
GSMUA_Achr10G16070_001	非同义突变	23 745 857 bp, C/G; 23 747 128 bp, C/A; 23 747 140 bp, C/G
GSMUA_AchrUn_randomG20420_001	非同义突变	99 012 968 bp, A/G; 99 013 414 bp, A/G; 99 014 563 bp, C/G
GSMUA_AchrUn_randomG20440_001	非同义突变	99 034 405 bp, A/G
GSMUA_Achr4G29150_001	密码子删除	27 244 210 bp, AGAAGAGTTACGCA/-
GSMUA_Achr2G08600_001	非同义突变	12 545 410 bp, T/C
GSMUA_Achr5G08880_001	非同义突变	6 453 974 bp, A/C; 6 454 063 bp, C/A; 6 454 377 bp, A/C
GSMUA_AchrUn_randomG12290_001	非同义突变	58 535 331 bp, A/C
GSMUA_Achr4G24930_001	非同义突变	24 594 882 bp, T/C; 24 595 037 bp, A/T
GSMUA_Achr4G30870_001	非同义突变	28 275 857 bp, G/C
GSMUA_Achr10G27570_001	非同义突变	30 682 367 bp, A/G; 30 682 419 bp, A/G; 30 682 884 bp, T/C; 30 682 910 bp, A/G
GSMUA_Achr1T18880	非同义突变	14 122 818 bp, T/C
GSMUA_Achr9G17080_001	非同义突变	11 614 316 bp, C/T; 11 615 932 bp, T/A
	密码子删除	11 615 869 bp, CTGCTC/-
GSMUA_Achr2G11540_001	非同义突变	14 352 663 bp, A/C; 14 353 114 bp, T/A
GSMUA_Achr5G27670_001	非同义突变	28 142 601 bp, T/G; 28 142 980 bp, G/A
GSMUA_Achr11G25120_001	非同义突变	24 349 704 bp, C/A; 24 349 806 bp, C/A; 24 349 963 bp, C/T; 24 350 212 bp, G/A; 24 350 416 bp, A/T; 24 350 485 bp, C/T; 24 350 525 bp, A/G; 24 350 617 bp, T/A; 24 350 789 bp, G/C; 24 350 825 bp, T/C; 24 351 182 bp, T/C
	密码子删除	24 350 011 bp, AGA/-
GSMUA_Achr8G21550_001	非同义突变	26 410 780 bp, G/C
GSMUA_Achr3G09440_001	非同义突变	6 824 088 bp, A/C
GSMUA_Achr3G11160_001	非同义突变	8 298 691 bp, T/G
GSMUA_Achr8G25580_001	非同义突变	29 405 165 bp, T/G
GSMUA_Achr7G21630_001	非同义突变	24 372 005 bp, T/A; 24 373 151 bp, C/G
GSMUA_Achr6G23310_001	非同义突变	23 410 162 bp, C/G; 23 410 273 bp, C/A
GSMUA_Achr5T26800	非同义突变	27 498 596 bp, T/A; 27 501 328 bp, C/T; 27 502 223 bp, A/G; 27 504 943 bp, A/G; 27 505 030 bp, C/G
GSMUA_Achr8T06970	非同义突变	4 526 056 bp, T/G
GSMUA_Achr11G20580_001	非同义突变	21 500 771 bp, G/T; 21 506 186 bp, A/C
GSMUA_Achr11G02140_001	非同义突变	1 467 325 bp, C/T
GSMUA_Achr8G14350_001	非同义突变	11 050 969 bp, C/G; 11 051 207 bp, G/C; 11 051 260 bp, C/A; 11 054 212 bp, T/C
GSMUA_Achr10G26950_001	非同义突变	30 265 820 bp, A/G; 30 265 968 bp, T/C; 30 266 084 bp, G/A; 30 266 457 bp, G/T; 30 267 446 bp, T/C
	移码突变	30 267 266 bp, -/TGAAGC
GSMUA_Achr2G15920_001	移码突变	17 039 567 bp, -/C
GSMUA_Achr4G30680_001	非同义突变	28 169 075 bp, G/A; 28 169 751 bp, A/T; 28 170 056 bp, T/C
GSMUA_Achr6G05220_001	非同义突变	3 527 472 bp, G/A
GSMUA_Achr9G28510_001	非同义突变	32 471 069 bp, A/G; 32 471 166 bp, T/C; 32 471 244 bp, C/T; 32 471 515 bp, A/C; 32 471 541 bp, C/T; 32 471 854 bp, C/A; 32 472 191 bp, C/T; 32 472 428 bp, G/A
GSMUA_Achr6G31160_001	非同义突变	31 038 502 bp, A/C
GSMUA_Achr7G01280_001	非同义突变	1 000 115 bp, T/C; 1 000 598 bp, C/T; 1 002 014 bp, G/A; 1 003 456 bp, T/G; 1 008 287 bp, C/A

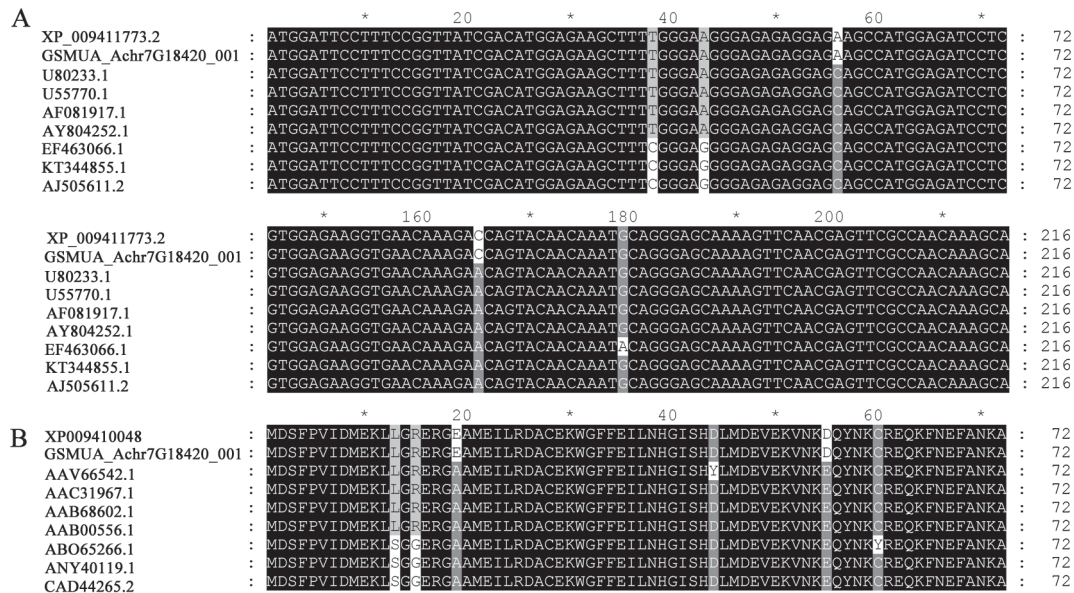


图5 不同香蕉品种MaACO1基因DNA和蛋白质序列比对

Fig.5 DNA and protein sequence alignment of *MaACO1* gene in different banana varieties

A: *MaACO1*基因的DNA序列比对; B: *MaACO1*蛋白质序列比对。DNA和蛋白质序列编号信息: GSMUA_Achr7G18420_001 (*M. acuminata* cv. DH-Pahang, AA, 马来西亚); XP_009411773.2 (*M. acuminata* subsp. *malaccensis*, AA, wild Malaysian banana), 对应蛋白序列XP009410048; U80233.1 (*M. acuminata* cv. Hsien Jin Chiao, AAA, dwarf banana), 对应蛋白序列AAB68602.1; U55770.1 (*M. acuminata* Cavendish cv. Grand Nain, AAA, dessert banana), 对应蛋白序列AAB00556.1; AF081917.1 (*M. acuminata* Cavendish cv. Williams, AAA, dessert banana), 对应蛋白序列AAC31967.1; AY804252.1 (*M. acuminata* cv. Baxi, AAA, dwarf banana), 对应蛋白序列AAV66542; EF463066.1 (*M. acuminata* × *balbisiana* cv. Fenjiao, ABB), 对应蛋白序列ABO65266.1; KT344855.1 (*M. sp.* cv. Nanjanagudurasabale, ABB), 对应蛋白序列ANY40119.1; AJ505611.2 (*M. sp.* cv. Ney Poovan, ABB), 对应蛋白序列CAD44265.2; AA、AAA分别代表A基因组二倍体和三倍体, ABB代表含有A和B基因组的三倍体。

类、进化特征和功能预测是当前香蕉功能基因研究的热点(Lakhwani等2016; Hu等2015, 2016; Goel等2016), 为进一步挖掘和有效利用优异的育种材料, 开展和优化香蕉育种过程提供了很好的技术手段(Dash和Rai 2016)。可食用的栽培香蕉(鲜食蕉和煮食蕉)是由AAA、AAB或ABB基因组组成的三倍体, 其中A基因组香蕉与收获前期发育和收获后期成熟密切相关(Davey等2013), 预示着拥有更多A基因组的栽培品种具有高产量、长果指和长期储存的优良特性(Hu等2015)。本研究通过对三倍体香蕉优良品种‘Grand Nain’ (AAA)进行重测序, 与‘DH-Pahang’参考基因组相比, 共获得4 598 633个SNP、484 752个Indel、57 047个SV, 以及36 277个基因变异, 其中多个基因同时存在多种类型的突变; 针对这些突变位点进行相关标记的开发、优异基因的挖掘将为分子标记辅助育种提供标记资源。

香蕉为呼吸跃变型果实, 乙烯在采后果实成

熟过程中扮演重要角色。国内外学者结合基因组和转录组数据分析在香蕉果实成熟和发育过程及功能基因研究方面取得了较好的结果(Asif等2014; Jourda等2014; 苗红霞等2013; Jourda等2016; Lakhwani等2016; Hu等2016; Goel等2016; 张远森等2013), 为解析香蕉果实成熟过程分子机理提供了宝贵的信息。本研究对来源于NCBI数据库中不同基因型(AAA、AAB、ABB)的*MaACO1*基因序列分析发现, 与香蕉‘DH-Pahang’参考基因组相比存在2个变异位点(A/C和C/A), 从而导致了氨基酸突变[谷氨酸(E)/丙氨酸(A)、天冬氨酸(D)/谷氨酸(E)]; 进一步分析发现, A基因组(AA、AAA)和B基因组(ABB)的*ACO*基因序列也存在2个相邻的变异(T/C和A/G)及氨基酸的突变[亮氨酸(L)/丝氨酸(S)、精氨酸(R)/甘氨酸(G)], 这为研究含A和/或B基因组香蕉品种成熟过程中*ACO1*基因功能提供了依据。在乙烯信号转导方面, *MaEIL5*在果实成熟中呈显著下调表达, 并且可以与F-box蛋白MaEBF2

互作, 形成蛋白复合体参与香蕉成熟过程(Kuang等2013), 通过分析发现GN与参考基因组相比, *MaEIL5* 存在2个非同义突变和1个密码子删除, *MaEBF2* 存在3个非同义突变, 这些变异为开展基因功能研究提供了理论参考。高度同源基因或多拷贝基因的存在是由功能分化及不等量表达等因素导致的(张彦威等2016), 如12个*ACO*和14个*ACS*中均有8个基因存在变异, 且*MaACO1*和*MaACSI*在香蕉成熟过程中表达较高, 其他基因表达较弱(Jourda等2014), 为从基因变异角度开展基因不等量表达提供了理论依据。综上所述, 本研究首次通过对三倍体香蕉优良品种‘Grand Nain’的全基因组变异分析以及KEGG富集获得了乙烯合成代谢途径如*ACO1*、*ACSI*、*EIN3/EILs*、*RAN*、*ERS*和*EBF*等家族基因存在序列变异的结果, 这些基因的变异为香蕉贮藏保鲜等相关分子标记开发、基因功能研究, 以及基于基因组编辑技术的香蕉遗传改良等工作提供了理论支撑。

参考文献(References)

- Abyzov A, Urban AE, Snyder M, et al (2011). CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res*, 21 (6): 974–984
- Ai JA, Xiong RQ, Dong LP, et al (2016). Development situation and counter measures of Guangdong banana industry in 2015. *Guangdong Agr Sci*, 43 (4): 11–15 (in Chinese with English abstract) [艾建安, 熊瑞权, 董丽萍等(2016). 2015年广东香蕉产业发展形势与对策建议. *广东农业科学*, 43 (4): 11–15]
- Asif MH, Lakhwani D, Pathak S, et al (2014). Transcriptome analysis of ripe and unripe fruit tissue of banana identifies major metabolic networks involved in fruit ripening process. *BMC Plant Biol*, 14: 316
- Chen K, Wallis JW, McLellan MD, et al (2009). BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods*, 6 (9): 677–681
- Choudhury SR, Roy S, Nag A, et al (2012). Characterization of an AGAMOUS-like MADS box protein, a probable constituent of flowering and fruit ripening regulatory system in banana. *PLoS ONE*, 7 (9): e44361
- Conesa A, Götz S (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int J Plant Genom*, 2008: 619832
- D’Hont A, Denoeud F, Aury JM, et al (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*, 488 (7410): 213–217
- Dash PK, Rai R (2016). Translating the “banana genome” to delineate stress resistance, dwarfing, parthenocarpy and mechanisms of fruit ripening. *Front Plant Sci*, 7: 1543
- Davey MW, Gudimella R, Harikrishna JA, et al (2013). “A draft *Musa balbisiana* genome sequence for molecular genetics in polyploid, inter- and intra-specific *Musa* hybrids”. *BMC Genomics*, 14: 683
- Dong C, Hu H, Hu Y, et al (2016). Metabolism of flavonoids in novel banana germplasm during fruit development. *Front Plant Sci*, 7: 1291
- Dong C, Li W, Zhang C, et al (2017). Phylogenetic analysis of *ACS* gene family in *Musa acuminata*. *Mol Plant Breed*, 15 (2): 425–432 (in Chinese with English abstract) [董晨, 李伟才, 张纯等(2017). 香蕉*ACS*基因家族的系统进化分析. *分子植物育种*, 15 (2): 425–432]
- Esteve C, D’Amato A, Marina ML, et al (2013). In-depth proteomic analysis of banana (*Musa* spp.) fruit with combinatorial peptide ligand libraries. *Electrophoresis*, 34 (2): 207–214
- Fan Z, Zhuo Y, Fu Y, et al (2017). Molecular characterization of a transcription factor MaWRKY31 and its regulation on *MaACSI*, *MaACO1* and *MaSAG1*. *Food Sci*, 38 (6): 13–19 (in Chinese with English abstract) [范中奇, 卓远芳, 符亚莉等(2017). 香蕉MaWRKY31转录因子的特性及对*MaACSI*、*MaACO1*和*MaSAG1*的调控. *食品科学*, 38 (6): 13–19]
- Fan ZQ, Kuang JF, Lu WJ, et al (2015). Advances in research of the mechanism of transcription factors involving in regulating fruit ripening and senescence. *Acta Horticult Sin*, 42 (9): 1649–1663 (in Chinese with English abstract) [范中奇, 邝健飞, 陆旺金等(2015). 转录因子调控果实成熟和衰老机制研究进展. *园艺学报*, 42 (9): 1649–1663]
- Feng HM, Chen Y, Wu YT (2016). Research progress of genetic diversity and evolution of bananas based on genomics. *J Hainan Trop Ocean Univ*, 23 (5): 87–91 (in Chinese with English abstract) [冯慧敏, 陈友, 武耀廷(2016). 基于基因组学的香蕉种质资源遗传多样性与进化研究进展. *海南热带海洋学院学报*, 23 (5): 87–91]
- Goel R, Pandey A, Trivedi PK, et al (2016). Genome-wide analysis of the *Musa WRKY* gene family: evolution and differential expression during development and stress. *Front Plant Sci*, 7: 299
- Hu C, Wei Y, Sheng O, et al (2015). Study on transformation of Furenzhi banana (*Musa* spp. AA group) with an antisense *ACO* gene. *Mol Plant Breed*, 13 (7): 1553–1558 (in Chinese with English abstract) [胡春华, 魏岳荣, 盛鸥等(2015). 香蕉*ACO*反义基因转化夫人指蕉的研究. *分子植物育种*, 13 (7): 1553–1558]
- Hu W, Wang L, Tie W, et al (2016). Genome-wide analyses of the bZIP family reveal their involvement in the development, ripening and abiotic stress response in banana. *Sci Rep*, 6: 30203

- Hu W, Zuo J, Hou X, et al (2015). The auxin response factor gene family in banana: genome-wide identification and expression analyses during development, ripening, and abiotic stress. *Front Plant Sci*, 6: 742
- Huang FC, Do YY, Huang PL (2006). Genomic organization of a diverse ACC synthase gene family in banana and expression characteristics of the gene member involved in ripening of banana fruits. *J Agr Food Chem*, 54 (11): 3859–3868
- Huang Y, Yi G, Zhou B, et al (2006). Cloning and sequence analysis of ACO gene of *Musa acuminata* × *balbisiana* cv. 'Fenjiao' (ABB Group). *Mol Plant Breed*, 4 (5): 633–639 (in Chinese with English abstract) [黄永红, 易干军, 周碧容等(2006). 粉蕉ACC氧化酶cDNA的克隆及其序列分析. *分子植物育种*, 4 (5): 633–639]
- Iqbal N, Khan NA, Ferrante A, et al (2017). Ethylene role in plant growth, development and senescence: interaction with other phytohormones. *Front Plant Sci*, 8: 475
- Jin Z, Li R, Xu B, et al (2002). Cloning of a banana pulp specific expressed cDNA encoding ACC synthase and construction of a vector with the antisense cDNA. *J Agr Biotechnol*, 10 (3): 305–306 (in Chinese with English abstract) [金志强, 李瑞珍, 徐碧玉等(2002). 香蕉果实特异性ACC合成酶基因的克隆及反义载体的构建. *农业生物技术学报*, 10 (3): 305–306]
- Jourda C, Cardi C, Gibert O, et al (2016). Lineage-specific evolutionary histories and regulation of major starch metabolism genes during banana ripening. *Front Plant Sci*, 7: 1778
- Jourda C, Cardi C, Mbéguié-A-Mbéguié D, et al (2014). Expansion of banana (*Musa acuminata*) gene families involved in ethylene biosynthesis and signalling after lineage-specific whole-genome duplications. *New Phytol*, 202 (3): 986–1000
- Kuang JF, Chen L, Shan W, et al (2013). Molecular characterization of two banana ethylene signaling component MaEBFs during fruit ripening. *Postharvest Biol Tec*, 85: 94–101
- Lakhwani D, Pandey A, Dhar YV, et al (2016). Genome-wide analysis of the AP2/ERF family in *Musa* species reveals divergence and neofunctionalisation during evolution. *Sci Rep*, 6: 18878
- Li H, Handsaker B, Wysoker A, et al (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25 (16): 2078–2079
- Mao X, Cai T, Olyarchuk JG, et al (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics*, 21 (19): 3787–3793
- Mbéguié-A-Mbéguié D, Hubert O, Fils-Lycaon B, et al (2008). EIN3-like gene expression during fruit ripening of Cavendish banana (*Musa acuminata* cv. Grande naine). *Physiol Plant*, 133 (2): 435–448
- Miao HX, Jin ZQ, Liu WX, et al (2013). Correlation between resistant starch metabolism and soluble sugar content of banana fruit. *Plant Physiol J*, 49 (8): 743–748 (in Chinese with English abstract) [苗红霞, 金志强, 刘伟鑫等(2013). 香蕉果实中抗性淀粉代谢与可溶性糖含量变化的相关性. *植物生理学报*, 49 (8): 743–748]
- Michiels A, Van den Ende W, Tucker M, et al (2003). Extraction of high-quality genomic DNA from latex-containing plants. *Anal Biochem*, 315 (1): 85–89
- Pandey A, Alok A, Lakhwani D, et al (2016). Genome-wide expression analysis and metabolite profiling elucidate transcriptional regulation of flavonoid biosynthesis and modulation under abiotic stresses in banana. *Sci Rep*, 6: 31361
- Shan W, Kuang JF, Chen L, et al (2012). Molecular characterization of banana NAC transcription factors and their interactions with ethylene signalling component EIL during fruit ripening. *J Exp Bot*, 63 (14): 5171–5187
- Toledo TT, Nogueira SB, Cordenunsi BR, et al (2012). Proteomic analysis of banana fruit reveals proteins that are differentially accumulated during ripening. *Postharvest Biol Tec*, 70: 51–58
- Wang K, Li M, Hakonarson H (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*, 38 (16): e164
- Wang XL, Peng XX, Li H (2000). Cloning and sequence of cDNA encoding ACC synthase specifically expressed in banana fruit. *Chin J Biotechnol*, 16 (2): 18–20 (in Chinese with English abstract) [王新力, 彭学贤, 李宏(2000). 香蕉果实特异性ACC合酶的cDNA克隆及序列分析. *生物工程学报*, 16 (2): 18–20]
- Wu W, Yang YL, He WM, et al (2016). Whole genome sequencing of a banana wild relative *Musa itinerans* provides insights into lineage-specific diversification of the *Musa* genus. *Sci Rep*, 6: 31586
- Xiao YY, Chen JY, Kuang JF, et al (2013). Banana ethylene response factors are involved in fruit ripening through their interactions with ethylene biosynthesis genes. *J Exp Bot*, 64 (8): 2499–2510
- Xu Q, Yin XR, Chen KS (2014). EIN3/EIL and AP2/ERF are involved in transcriptional regulation on fruit quality. *Acta Horti Sin*, 41 (9): 1913–1923 (in Chinese with English abstract) [徐倩, 殷学仁, 陈昆松(2014). 基于乙烯受体下游转录因子的果实品质调控机制研究进展. *园艺学报*, 41 (9): 1913–1923]
- Zhang YS, Qin XM, Yan JP, et al (2013). Bioinformatics analysis of the MADS-Box gene family in banana. *Plant Physiol J*, 49 (10): 1030–1040 (in Chinese with English abstract) [张远森, 秦晓萌, 严金平等(2013). 香蕉MADS-Box基因家族的生物信息学分析. *植物生理学报*, 49 (10): 1030–1040]

Zhang YW, Li W, Zhang LF, et al (2016). Genome wide variations of soybean cultivar Qihuang 34 by whole genome re-sequencing. *Chin J Oil Crop Sci*, 38 (2): 150–158 (in

Chinese with English abstract) [张彦威, 李伟, 张礼凤等 (2016). 基于重测序的大豆新品种齐黄34的全基因组变异挖掘. *中国油料作物学报*, 38 (2): 150–158]

Genome-wide variations of triploid banana (AAA group) ‘Grand Nain’ by whole-genome resequencing

SHAO Xiu-Hong^{1,2,#}, HU Chun-Hua^{2,#}, SHENG Ou², BI Fang-Cheng², DENG Gui-Ming², YANG Qiao-Song^{2,*}, YI Gan-Jun^{2,*}

¹Horticulture and Landscape College, Hunan Agricultural University, Changsha 410128, China

²Institute of Fruit Tree Research, Guangdong Academy of Agricultural Sciences; Key Laboratory of South Subtropical Fruit Biology and Genetic Resource Utilization, Ministry of Agriculture; Guangdong Key Laboratory of Tropical and Subtropical Fruit Tree Research, Guangzhou 510640, China

Abstract: For an overall understanding of the whole genome of important banana (*Musa* sp.) cultivar Grand Nain, high-throughput sequencing technology was used by resequencing with 53.79 X depth to detect its variations. Results revealed 4 598 633 single nucleotide polymorphisms (SNP), 484 752 small insertion-deletions (Indels) and 57 047 structural variations (SV). They confer 36 277 DNA variations. Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses of variations showed that 442 genes of plant hormone signal transduction had DNA variations. These variations include two main enzymes, 1-aminocyclopropane-1-carboxylate oxidase (ACO) and 1-aminocyclopropane-1-carboxylate synthase (ACS), in ethylene biosynthesis pathway. In addition, several variations in ethylene signal transduction such as *ethylene-insensitive 3-like (EIN3/EILs)*, *ethylene response sensor (ERS)*, *response-to-antagonist (RAN)*, *ethylene-insensitive 3-binding F-box (EBF)*, and *EIN2* were also detected in ‘Grand Nain’, compared to the reference genome. The results of DNA and protein sequencing analyses of *MaACO1* showed that, there were two nucleotide variations in DNA sequence, leading to changes in protein sequence, and the two nucleotide variations in *MaACO1* were also detected in A and B genomes of banana. These results would improve genetic engineering and forward genetic accomplishments with an aim to molecular breeding in banana.

Key words: banana; whole-genome resequencing; single nucleotide polymorphism; small Indel; structural variation; gene variation

Received 2017-09-11 Accepted 2018-03-14

This work was supported by the National Natural Science Foundation of China (31372018, 31772267 and 31601720), 948 Project from Ministry of Agriculture of China (2011-G16 and 2016-X22), the National Banana Industry and Technology System Project (CARS-32-01), the National Commonweal Research Project (201403036), Science and Technology Plan Project of Guangdong Province (2012A020100007, 2014B050502007, 2016B020233002 and 2017A030303044), the National Science and Technology Program in Rural Areas During the Twelfth Five-year Plan Period (2013BAD02B02), and Science and Technology Program of Guangzhou (201707010154).

#Co-first authors.

*Co-corresponding authors: Yi GJ (yiganjun@vip.163.com), Yang QS (soyang@hotmail.com).