

李毅丰,姜悦畅,孙中锋,等. 基于重测序的短节间番茄品系变异检测分析[J]. 江苏农业科学,2022,50(13):58-63.

doi:10.15889/j.issn.1002-1302.2022.13.010

基于重测序的短节间番茄品系变异检测分析

李毅丰¹,姜悦畅¹,孙中锋¹,王 帅^{1,2},唐贝贝¹,师海林¹,张 宁^{1,2},王玉斌³,毛秀杰^{1,2}

(1. 河北科技师范学院园艺科技学院,河北秦皇岛 066004;

2. 河北省特色园艺种质挖掘与创新利用重点实验室,河北秦皇岛 066004; 3. 河北省承德市农业农村局,河北承德 067000)

摘要:为揭示短节间番茄的基因变异类型及基因变异对代谢通路的影响,以短节间番茄品系 CH 和普通节间长度番茄品系 DH 进行农艺性状分析以及全基因组重测序,对单核苷酸多态性(single nucleotide polymorphism, SNP)、小片段插入和缺失(insertion-deletion, InDel)、拷贝数变异(copy number variation, CNV)及染色体结构变异(structural variation, SV)4种变异类型进行检测注释。结果表明,在2个样本中共检测到1 086 531个SNP、277 973个InDel、25 797个CNV、33 516个SV,在代谢通路中主要集中于嘌呤生物碱生物合成、二萜生物合成、莽草酸途径生物碱的生物合成中发现。以SNP以及InDel中的基因作为切入点,挖掘到3个变异基因,其在CH中的相对表达量均低于DH。研究结果明确了变异位点信息,为进一步明确调控番茄节间长度相关基因的研究奠定了基础。

关键词:番茄;节间;重测序;变异检测

中图分类号:S641.201

文献标志码:A

文章编号:1002-1302(2022)13-0058-06

番茄(*Solanum lycopersicum* L.)是重要的蔬菜作物和模式植物。实际生产应用中,集约化的育苗方式增加了番茄育苗密度,植株间相互遮阴,促使幼苗发生徒长,因此番茄节间长度的研究受到广泛关注^[1]。番茄节间长度是影响番茄产量的重要农艺性状之一,不仅会对番茄栽培密度产生影响,还对番茄的叶面积指数、光能利用率、干物质积累以及果实品质产生影响^[2-4]。番茄节间长度缩短不仅能够有效降低植株高度,抑制徒长,增加土地利用率,减少土壤中营养物质的消耗,还能降低苗期管理和果实采收的人工成本^[5],因此,对短节间番茄进行研究具有重要意义。

随着测序技术的迅速发展,通过高通量测序技术能够对许多作物基因进行研究。周世奇等对选育的航天突变体烟草NC89-M与野生型NC89进行全基因组重测序,在NC89中检测获得了1 848 013个单核苷酸多态性(single nucleotide polymorphism, SNP)、398 922个小片段插入和缺失(insertion-deletion, InDel)、41 969个染色体结构变异

(structural variation, SV),在NC89-M中检测到1 876 219个SNP、402 011个InDel、42 699个SV,采用生物信息学分析方法得出代谢通路和次生代谢产物合成相关基因突变数目最多,通过变异基因功能注释获得MADS-box基因和KNOX1基因,分别调控开花时间和侧生器官发育与叶缘形状^[6]。杜海东等对3个自封顶番茄品系进行基因组重测序,通过变异检测分析3个样本检测到5 968 501个SNP和485 114个InDel,与参考基因组比对后共发生33 473个变异基因^[7]。对CDS区域的变异基因进行KEGG数据库比对,发现主要集中在基础代谢和玉米素的生物合成,通过多序列比对获得了16个调控封顶花序数的关键基因。

目前,关于番茄节间长度的研究,多集中于对长节间番茄的研究^[8],而对短节间番茄突变体的研究鲜见报道。因此,本研究对2个节间长度不同的番茄品系进行全基因组重测序,并对其SNP、InDel、拷贝数变异(copy number variation, CNV)、SV等4种变异类型进行深度挖掘,以及基因变异对代谢通路的影响,从而为分子水平改良番茄节间长度的研究奠定基础。

1 材料与方法

1.1 试验材料

本试验以短节间番茄品系CH和普通节间番茄

收稿日期:2021-09-11

基金项目:河北省重点研发计划(编号:20326325D)。

作者简介:李毅丰(1996—),男,河北保定人,硕士,主要从事番茄遗传育种与分子生物学研究。E-mail:1658529139@qq.com。

通信作者:毛秀杰,博士,教授,主要从事番茄遗传育种与分子生物学研究。E-mail:maoxiujie@126.com。

品系 DH 为试验材料,均为无限生长型。试验在河北科技师范学院园艺园林试验站 7 号温室中进行,于 2021 年 1 月 16 日进行播种育苗,待幼苗长至 5 ~ 6 张真叶时进行移栽。采用日光温室栽培,大行距为 70 cm,小行距为 40 cm,株距为 30 cm,应用滴灌方式进行水肥管理。在开花结果期取幼嫩组织用于基因组重测序分析。

1.2 试验方法

1.2.1 全基因组 DNA 提取 通过十六烷基三甲基溴化铵(CTAB)法对样本茎段进行 DNA 的提取;用 0.8% 琼脂糖凝胶电泳对 DNA 质量进行检测;采用核酸蛋白分析仪对提取的 DNA 进行定量。

1.2.2 全基因组重测序 对质检合格的 DNA 采用超声进行随机打断。采用 TruSeq DNA PCR-Free Prep kit 的标准建库流程制备测序文库,对序列末端进行修复,3'端添加多聚腺苷酸(PloyA)、5'端添加含有文库特异性标签、纯化、测序文库模板富集等步骤完成测序文库的制备。对完成的测序文库在 Agilent Bioanalyzer 上进行质检;采用 Quant-iT PicoGreen dsDNA Assay Kit 对文库进行定量(文库浓度大于 2 nmol 质量合格);根据所需测序量对合格的各上机测序文库梯度稀释后按相应比例混合使用;应用 NovaSeq 测序仪进行双端测序。

1.2.3 数据统计 采用 BWA(0.7.12-r1039) mem 程序将经过过滤后得到的高质量数据比对到参考基因组上,比对的参数均按照 bwamem 的默认参数^[9]。采用 ANNOVAR 软件对 SNP 位点和 InDel 位点进行注释^[10]。采用 GATK(<https://www.broadinstitute.org/gatk/>)(RealnTimes = 1)进行 CNV 检测^[11],使用 BreakDancer 进行结构变异(SV)检测^[12]。

2 结果与分析

2.1 CH 与 DH 农艺性状分析

在开花结果期对 2 个番茄品系的株高、总节间长度、平均节间长度进行分析(表 1、图 1)。结果表明,CH 株高均极显著高于 DH;CH 与 DH 相比总节间长度不存在显著差异;CH 与 DH 之间的平均节间长度均存在极显著差异。

2.2 变异检测分析

2.2.1 CH 与 DH 基因组重测序结果与分析 由表 2 可见,CH 和 DH 这 2 个样本的 Reads 总数分别为 168 148 334、152 257 248 个,过滤后 Reads 数量及

表 1 CH 与 DH 开花结果期农艺性状分析

材料	株高 (cm)	总节间长度 (cm)	平均节间长度 (cm)
DH	90.3 ± 6.1bB	78.0 ± 4.7aA	5.5 ± 0.22aA
CH	103.1 ± 5.4aA	84.8 ± 8.4aA	4.82 ± 0.41bB

注:同列数据后不同小写字母表示差异显著($P < 0.05$),不同大写字母表示差异极显著($P < 0.01$)。



左侧植株为短节间番茄品系 CH, 右侧植株为长节间番茄品系 DH

图 1 CH 与 DH 开花结果期植株表型

过滤后 Reads 数量占原始 Reads 数量的百分比分别为 160 322 604 个(95.35%)、145 443 866 个(95.53%);碱基总数分别为 25 222 250 100、22 838 587 200 bp,过滤后 Reads 碱基数及高质量 Reads 碱基数占原始碱基总数的百分比分别为 23 522 020 839 bp(93.26%)、21 329 453 341 bp(93.39%)。GC 含量在 36.83 ~ 36.98 之间;Q20 ≥ 96.82%、Q30 ≥ 91.85%。

2.2.2 CH 和 DH 与参考基因组对比情况 以 Heinz 1706 为参考基因组。2 个样本的比对率在 99.61% ~ 99.89% 之间(表 3)。平均测序深度为 23X,1X 覆盖度在 96.63% ~ 99.74% 之间,4X 覆盖度在 95.62% ~ 99.52% 之间,10X 覆盖度在 93.19% ~ 96.97% 之间,20X 覆盖度在 46.47% ~ 52.31% 之间(表 4)。综上,比对结果正常,可用于后续相关分析。

2.2.3 SNP 检测与注释 依据与参考基因组的比对结果(表 5)可知,CH 和 DH 的总 SNP 变异个数相同,为 3 893 609 个,CH 和 DH 的纯合基因型数分别

表 2 CH 与 DH 的碱基数据统计

样本名	Reads 数量			碱基			GC 含量	Q20 (%)	Q30 (%)
	总数 (个)	过滤后数量 (个)	过滤后比例 (%)	总数 (bp)	过滤后数量 (bp)	过滤后比例 (%)			
DH	152 257 248	145 443 866	95.53	22 838 587 200	21 329 453 341	93.39	36.98	96.84	91.95
CH	168 148 334	160 322 604	95.35	25 222 250 100	23 522 020 839	93.26	36.83	96.82	91.85

表 3 CH 与 DH 序列比对结果统计

样本名	所有 Reads 数 (个)	与参考基因组比对 Reads 数(个)	比对率 (%)
DH	146 460 907	146 295 155	99.89
CH	162 490 595	161 864 192	99.61

表 4 CH 与 DH 比对测序深度和覆盖度结果统计

样本名	平均测序 深度(X)	1X 覆盖度 (%)	4X 覆盖度 (%)	10X 覆盖度 (%)	20X 覆盖度 (%)
DH	22.02	99.74	99.52	96.97	46.47
CH	23.70	96.63	95.62	93.19	52.31

表 5 CH 与 DH SNP 检测结果统计

样品名	纯合基因 型数(个)	杂合基因 型数(个)	未知基因 型数	与参考基因组不一致 的纯合基因型数
DH	3 506 620	251 093	2 994	132 902
CH	186 147	1 106 264	36 007	2 565 191

为 186 147、3 506 620 个,CH 中纯合基因型低于 DH;杂合基因型数分别为 1 106 264、251 093 个;未知基因型数分别为 36 007、2 994 个;与参考基因组不一致的纯合基因型数分别为 2 565 191、132 902 个。由 SNP 在染色体上的分布(图 2)可知,SNP 在 9 号染色体的 24.89 ~ 27.79 Mb、35.32 ~ 38.21 Mb 区域内富集,初步判断在此区域内发生基因突变,其余 SNP 均匀分布在 6 号染色体的 1.74 ~ 8.11 Mb、12.16 ~ 31.26 Mb 和 9 号染色体的 7.53 ~ 18.53 Mb、21.42 ~ 62.53 Mb 之间。

依据参考基因组比对结果,对 CH 和 DH 中的 SNP 进行比较(表 6),结果表明,CH 和 DH 中共检测到 1 086 531 个 SNP。对 2 个样本之间的全基因组 SNP 变异进行注释,发现 SNP 变异主要集中在基因间区,其比例约占总数的 78.03%;其次是发生于内含子区域内,其比例约占总数的 8.48%;而发生于 CDS 区域的非同义突变比例约占 CDS 区域总数的 60.85%。

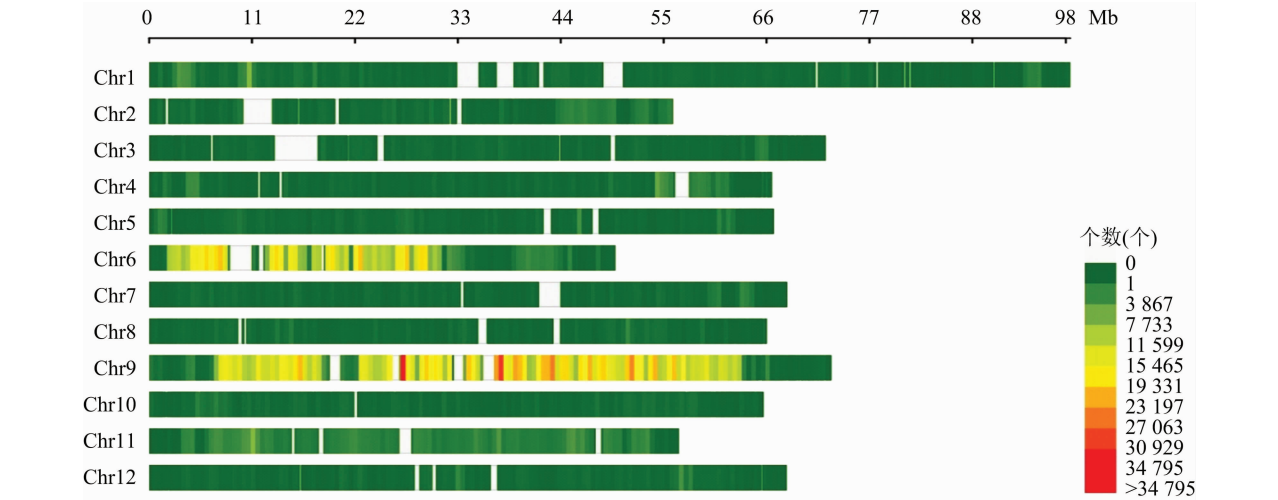


图2 CH 与 DH SNP 在染色体上的分布

基因组 SNP 突变可以分成 6 类,分别为 T:A > C;G、T:A < C;G、T:A < G;C、T:A > A;T、C;G > G;C、C;G > T:A。SNP 频谱分析如图 3 所示,由图 3 可知,T:A > C;G 和 C;G > T:A 为主要的 SNP 突

变型。
2.2.4 InDel 检测及注释 为了定位目标性状,每组性状相关样本一起 call 群体 InDel,经过滤对最终得到的群体 SNP 在每个样本中的数目做统计见表

表 6 CH 与 DH SNP 注释结果

变异类型	CH 与 DH SNP 数目 (个)	区域
外显子区域	23 919	
终止密码子获得	1 108	CDS
终止密码子丢失	227	CDS
同义突变	19 040	CDS
非同义突变	31 666	CDS
splicing junction 2 bp 的区域	190	
内含子区域	92 148	
基因间区域	847 873	
转录起始位点上游 1 kb 的区域	37 387	
转录终止位点下游 1 kb 的区域	28 405	
转录起始位点上游和下游的区域	4 568	
SNP 位点总数	1 086 531	

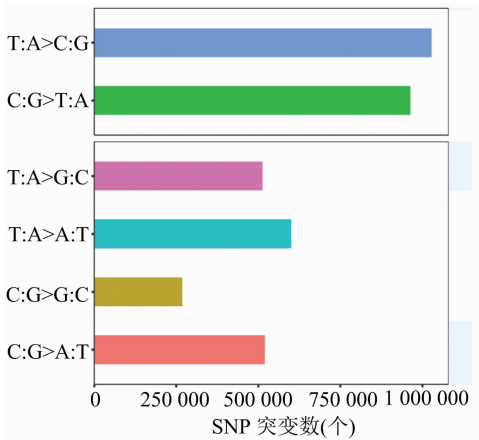


图3 CH 与 DH SNP 突变频谱分析

7。由表 7 可知,CH 和 DH 的纯合基因型数分别为 15 327、259 512 个,CH 中纯合基因型数小于 DH;杂合基因型数分别为 76 637、29 143 个,CH 中杂合基因型数多于 DH;未知基因型数分别为 3 180、733 个;与参考基因组不一致的纯合基因型数分别为 265 384、71 140 个。

表 7 CH 与 DH InDel 检测结果统计

样品名	纯合基因型数(个)	杂合基因型数(个)	未知基因型数(个)	与参考基因组不一致的纯合基因型数(个)
DH	259 512	29 143	733	71 140
CH	15 327	76 637	3 180	265 384

为鉴定 CH 与 DH 之间的 InDel 变异,分别进行比较(表 8)。结果表明,CH 和 DH 之间共检测到 277 973 个 InDel。对 2 个样本之间的全基因组 InDel 变异进行注释,发现 InDel 变异主要集中在基因间区域,其比例约占总数的 76.53%;其次是发生

于内含子区域内,其比例约占总数的 11.67%;而发生于 CDS 区域的移码突变比例约占总数的 0.55%。
2.2.5 CH 与 DH CNV 变异分析 采用 CNVnator v0.2.7 检测 3 个样本的全基因组中存在的 CNV,通过 CNV 统计(表 9)可知,CH 中检测出的 CNV 为 12 756 个,拷贝数缺失 10 970(85.9%),拷贝数增加 1 786(14.1%);在 DH 中检测出的 CNV 为 13 041 个,拷贝数缺失 11 576(88.8%),拷贝数增加 1 465(11.2%)。因此得出,CH 和 DH 的 CNV 变异主要是发生拷贝数缺失。

表 8 CH 与 DH InDel 注释结果统计

变异类型	CH 与 DH SNP 数目 (个)	区域
外显子区域	2 317	
终止密码子获得	61	CDS
终止密码子丢失	14	CDS
移码缺失	812	CDS
移码插入	706	CDS
非移码缺失	349	
非移码插入	375	
splicing junction 2 bp 的区域	98	
内含子区域	32 426	
基因间区域	212 744	
转录起始位点上游 1 kb 的区域	14 858	
转录终止位点下游 1 kb 的区域	11 165	
转录起始位点上游和下游的区域	2 048	
InDel 位点总数	277 973	

表 9 CH 与 DH CNV 变异分析

样品名	拷贝数缺失 (个)	拷贝数增加 (个)	CNV 数 (个)
DH	11 576	1 465	13 041
CH	10 970	1 786	12 756

2.2.6 CH 与 DH SV 变异分析 染色体的结构变异主要包括缺失(deletion, DEL)、插入(insertion, INS)、倒位(inversion, INV)、染色体内易位(intra-chromosomal translocation, ITX)、染色体间易位(inter-chromosomal translocation, CTX)。采用 Breakdancer1.3.7 版本检测染色体结构变异,5 种结构变异的数量统计见表 10。CH 的 SV 数量最多为 20 501 个,其中染色体缺失的数量最多为 8 176 个,占总变异的 39.9%,染色体间易位和染色体内易位分别占总变异的 24.6%、20.0%,插入和倒位分别占总变异的 5.5%、10.0%。DH 的 SV 数量最少为 13 015 个,其中变异最多的 2 种类型为染色体间易

位和缺失分别为 4 668 个、4 468 个,占总变异的 35.9%、34.3%,插入、倒位、染色体内部易位分别占总

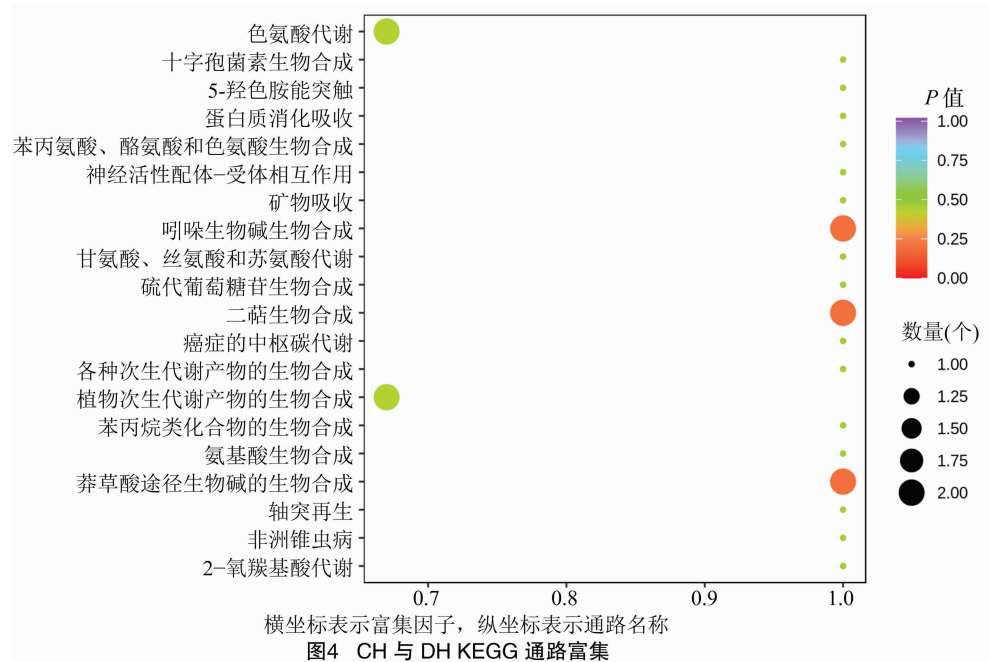
变异的 3.7%、6.5%、19.6%。CH 和 DH 的 SV 变异主要发生的变异类型为缺失。

表 10 CH 与 DH SV 变异分析

样品名	染色体间易位(个)	缺失(个)	插入(个)	倒位(个)	染色体内部易位(个)	SV 数量(个)
DH	4 668	4 468	481	852	2 546	13 015
CH	5 040	8 176	1 130	2 060	4 095	20 501

2.2.7 CH 与 DH 代谢水平的差异分析 为了解基因变异引起代谢物质产生差异,因此对 CH 与 DH 节间中的代谢物质进行了分析。根据差异代谢物结果,对 CH 与 DH 的 KEGG 通路进行比较,获得

KEGG 通路富集图(图 4)。由图 4 可知,在 CH 和 DH 中共注释到 20 个代谢通路,其中有 3 个通路被显著富集,包含吡啶生物碱生物合成、二萜生物合成、草莽酸途径生物碱的生物合成。



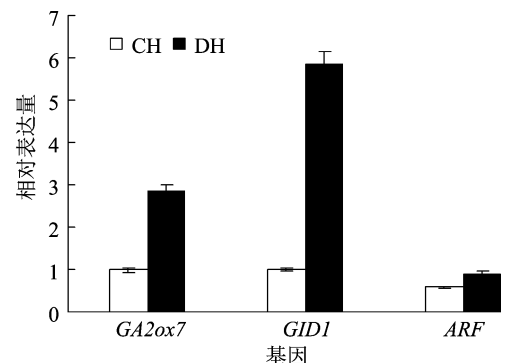
2.2.8 番茄短节间形成相关基因挖掘 为挖掘参与调控番茄短节间形成的相关基因,以 SNP 以及 InDel 中的基因作为切入点,进行变异基因挖掘。通过代谢通路富集情况,筛选出 3 个与节间长度相关联的变异基因,各基因功能注释、变异类型见表 11,根据功能注释获得了控制赤霉素氧化酶、赤霉素受体、生长素响应因子的基因。

表 11 变异基因功能注释

基因名称	功能注释	变异类型
<i>GA2ox7</i>	赤霉素 2-β-双加氧酶	SNP
<i>GID1</i>	赤霉素受体	SNP
<i>ARF</i>	生长素响应因子	SNP

2.2.9 候选基因表达分析 基于 qRT-PCR 技术检测开花结果期中与节间长度相关联的变异基因,并进行相对表达量分析。由图 5 可得,3 个变异基因在 2 个试验材料中均表现出差异。*GA2ox7* 基因

在 DH 中的表达量是 CH 中的 2.85 倍;*GID1* 基因在 DH 中的表达量是 CH 中的 5.87 倍;*ARF* 基因在 DH 中的表达量是 CH 中的 1.58 倍。在 DH 中 3 个基因的表达量明显高于 CH,以此推断 *GA2ox7*、*GID1*、*ARF* 在番茄节间长度的调控中发挥着至关重要的作用。



3 讨论与结论

番茄是重要的蔬菜作物,其节间长度受多种因素影响,包括环境因子、植物激素、遗传因子等^[13-15]。目前已有植物节间长度分子研究的报道,刘根忠等通过对 226 份番茄核心种质资源进行全基因组关联分析,鉴定出 7 个与节间长度相关的数量性状基因座(QTL),通过对显著的 SNP 位点进行候选基因分析,挖掘出控制番茄节间长度的主效基因 *IL10* 和 *IL11*^[16];刘忠祥等通过对玉米进行外源赤霉素(GA)的喷施,确定玉米对 GA 的响应正常,在利用分子标记技术将控制株高的主效 QTL qPH3.2 共定位在第 3 染色体上^[17]。研究发现,不同的土壤环境会对番茄内源激素含量产生影响,从而对植株的生长发育产生影响^[18];控制节间发育相关的基因通过影响代谢通路来控制植物节间发育^[19]。前人研究发现,分子标记辅助育种能够提高植物优良性状的选择效率,对植物遗传育种具有重要意义^[20]。随着番茄全基因组测序的完成以及二代测序的迅速发展,使番茄基因组的变异检测分析成为可能^[21]。

本研究通过对短节间番茄品系 CH 和普通节间番茄品系 DH 进行全基因组重测序,与参考基因组 Heinz 1706 番茄相比,在 DH 中检测到 3 893 609 个 SNP、360 528 个 InDel、13 041 个 CNV、13 015 个 SV;在 CH 中检测到 3 893 609 个 SNP、360 528 个 InDel、12 756 个 CNV、20 501 个 SV。结果表明,2 个番茄材料主要发生的变异类型为单核苷酸多态性(SNP)和小片段插入和缺失(InDel),与油菜^[22]、烟草^[6]和水稻^[23]的研究结果相吻合。对 2 个样本的 KEGG 通路进行比较,富集到的代谢通路包括吡啶生物碱生物合成、二萜生物合成、莽草酸途径生物碱的生物合成。为后续对变异基因分子标记的开发、分子标记辅助育种、基因克隆以及基因功能验证的研究奠定重要基础。

参考文献:

- [1] 王红飞,尚庆茂. 蔬菜徒长苗的形态及生理特征研究进展[J]. 中国蔬菜,2017(7):22-28.
- [2] 杨冬艳,桑 婷,冯海萍,等. 种植密度对日光温室东西垄向栽培番茄产量构成及光环境的影响[J]. 河南农业科学,2021,50(5):99-106.
- [3] Fang Q, Zhang X Y, Chen S Y, et al. Selecting traits to reduce seasonal yield variation of summer maize in the North China plain [J]. Agronomy Journal,2019,111(1):343-353.

- [4] Gao Z, Feng H Y, Liang X G, et al. Limits to maize productivity in the North China Plain: a comparison analysis for spring and summer maize[J]. Field Crops Research,2018,228:39-47.
- [5] 卢 阳,李 娜,王 星,等. 茄科作物激素调控的矮化突变体研究进展[J]. 植物遗传资源学报,2021,22(3):609-615.
- [6] 周世奇,刘东阳,潘旭浩,等. 烟草航天诱变突变体变异检测及分析[J]. 植物遗传资源学报,2019,20(2):377-386.
- [7] 杜海东,游 茜,李毅丰,等. 基于重测序的 3 个自封顶番茄品系基因组变异检测分析[J]. 分子植物育种,2022,20(3):756-764.
- [8] Sun X R, Shu J S, Ali Mohamed A M, et al. Identification and characterization of EI (elongated internode) gene in tomato (*Solanum lycopersicum*) [J]. International Journal of Molecular Sciences,2019,20(9):2204.
- [9] Li H, Handsaker B, Wysoker A, et al. 1000 genome project data processing subgroup. The sequence alignment/map (SAM) format and SAMtools [J]. Bioinformatics,2009,25(16):2078-2079.
- [10] Wang K, Li M Y, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data [J]. Nucleic Acids Research,2010,38(16):e164.
- [11] 马婷玉. 基于多组学解析黄花蒿核心种质及青蒿素生物合成调控机制[D]. 北京:中国中医科学院,2020.
- [12] Chen K, Wallis J W, McLellan M D, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation[J]. Nature Methods,2009,6(9):677-681.
- [13] 杨亚娜,樊小雪,徐 刚,等. 不同红蓝 LED 光照强度和灌溉量交互作用对番茄幼苗生长的影响[J]. 浙江农业学报,2019,31(5):737-745.
- [14] 孙晓荣. 番茄长节间基因 *EI* (Elongated Internode) 的图位克隆与功能分析[D]. 北京:中国农业科学院,2019.
- [15] 范业庚,丘立杭,黄 杏,等. 甘蔗节间伸长过程赤霉素生物合成关键基因的表达及相关植物激素动态变化[J]. 植物学报,2019,54(4):486-496.
- [16] 刘根忠,石春美,于会洋,等. 番茄节间长度基因 *IL10* 和 *IL11* 的鉴定和功能验证[J]. 园艺学报,2021,48(7):1340-1348.
- [17] 刘忠祥,杨 梅,殷鹏程,等. 玉米株高主效 QTL qPH3.2 精细定位及遗传效应分析[J]. 作物学报,2018,44(9):1357-1366.
- [18] 陈为京,李润芳,杨焕明,等. 黄绿木霉 T1010 对亚低温胁迫下日光温室番茄内源激素的调控效应[J]. 西南农业学报,2010,23(3):814-819.
- [19] 魏 强,丁雨龙. 矢竹地下茎转录组测序及节间生长相关基因表达分析[J]. 南京林业大学学报(自然科学版),2017,41(5):42-48.
- [20] 王艺蓉. 分子标记辅助樱桃番茄优良自交系及杂交组合的选育[D]. 武汉:华中农业大学,2019.
- [21] 高宝祯,刘 博,李石开,等. 白菜类作物开花时间的全基因组关联分析[J]. 中国农业科学,2017,50(17):3375-3390.
- [22] 胡 鸣,姚圣黎,程晓晖,等. 基于高深度重测序的春性、半冬性和冬性甘蓝型油菜基因组遗传变异分析[J]. 中国油料作物学报,2018,40(4):469-478.
- [23] 张志勇,蒲志刚,王 平,等. 水稻航天诱变突变体全基因组测序研究[J]. 西南农业学报,2014,27(2):469-475.