基因组学与应用生物学 2018 年 第 37 卷 第 3 期 第 1248-1256 页 Genomics and Applied Biology, 2018, Vol.37, No.3, 1248-1256

研究报告

Research Report

葡萄、桃和可可基因组的进化分析

崔晓波 孙梦琦 赵康路 闫立仁 张岚* 华北理工大学生命科学学院,基因组学与计算生物学研究中心,唐山,063200 * 通讯作者,zhanglan1374@sohu.com

摘 要 本研究以葡萄、桃和可可为研究对象,基于比较基因组学,利用基因同源共线性方法对基因组内的 结构和基因组间同源信息进行比对分析,确定了物种基因组内和基因组间的同源片段。通过统计3个物种基 因组间的同源共线基因的保留情况发现,葡萄基因组的保留情况最好,桃次之(为73.4%),可可最差(为68.9%), 其丢失均可能是由于双子叶植物共有的三倍化导致基因组稳定性遭到破坏。另外,共线基因间的同义核苷酸 替换率的频数分布证实,葡萄、桃和可可仅经历过一次古老的全基因组三倍化,并未经历最近的全基因组加 倍,且可可基因组进化最快,葡萄基因组进化最保守;3个物种的分歧时间分别为:葡萄(~110 Mya)、可可 (~90 Mya)、桃(~80 Mya)。本研究将为3个物种及双子叶植物基因组的结构、功能和进化等研究提供重要的 理论依据。

关键词 葡萄,桃,可可,全基因组加倍,基因组稳定性

Evolution Analysis of Vitis vinifera, Prunus persica and Theobroma cacao Genomes

Cui Xiaobo Sun Mengqi Zhao Kanglu Yan Liren Zhang Lan * Center for Genomics and Computational Biology, College of Life Sciences, North China University of Science and Technology, Tangshan, 063200 * Corresponding author, zhanglan1374@sohu.com

DOI: 10.13417/j.gab.037.001248

Abstract In this study, *Vitis vinifera, Prunus persica* and *Theobroma cacao* were selected as the research objects. Based on the comparative genomics, the genomic structure and gene homology information were comparatively analyzed by homologous genes collinearity method to determine homologous fragments within and between selected genomes. The statistics of the retention of homologous collinear genes among three species genomes showed that *Vitis vinifera* genes retained the best, followed by *Prunus persica* better (about 73.4%) and *Theobroma cacao* came last (about 68.9%), all of which could be due to common triploidization of dicotyledons leading to the destruction of genomic stability. In addition, the frequency distribution of synonymous nucleotide substitution rates between collinear homologous genes indicated that *Vitis vinifera*, *Prunus persica* and *Theobroma cacao* genomes had only experienced an ancient whole-genome triploidization without other whole-genome polyploidization recently. It was found that *Theobroma cacao* genome evolved most fastly and *Vitis vinifera*(~110 Mya), *Theobroma cacao* (~90 Mya) and *Prunus persica* (~80 Mya). This study would provide important theoretical basis for researches on genomic structure, function and evolution of three species and dicotyledons.

Keywords Vitis vinifera, Prunus persica, Theobroma cacao, Whole genome polyploidization, Genome stability

基金项目:本研究由华北理工大学青年科学研究基金(Z201528)和华北理工大学创新性实验计划项目(X2016261)共同资助 引用格式: Cui X.B., Sun M.Q., Zhao K.L., Yan L.R., and Zhang L., 2018, Evolution analysis of *Vitis vinifera*, *Prunus persica* and *Theobroma cacao* genomes, Jiyinzuxue Yu Yingyong Shengwuxue (Genomics and Applied Biology), 37(3): 1248-1256 (崔晓波,孙梦琦, 赵康路, 闫立仁, 张岚, 2018, 葡萄、桃和可可基因组的进化分析, 基因组学与应用生物学, 37(3): 1248-1256)

葡萄(Vitis vinifera, 2n=38)、桃(Prunus persica, 2n=16)和可可(Theobroma cacao, 2n=20)是 3 种具有 极高经济价值和科学研究价值的双子叶植物。其中, 葡萄作为最古老的双子叶植物之一,对于研究开花 植物祖先的遗传性状和特征起着相当重要的作用 (Jailon et al., 2007)桃作为一种已经被种植了四千多 年并且具有高度遗传特征的树种和水果,对于基因 组学的研究至关重要(Verde et al., 2013);可可作为巧 克力的来源和重要的热带果树经济作物,在各方面 都颇有研究价值,在基因组进化方面尤为突出(Argout et al., 2011)。这 3 个物种的经济价值驱动着科学 研究的进一步实践,同时全基因组测序的完成更为 其进化和基因功能等研究提供了便利。

植物基因组进化过程中的一个重要特征 即全基 因组加倍 极大地影响了植物基因组的结构和基因 功能的变化。越来越多的研究表明几乎所有的被子 植物都经历过全基因组加倍 ,而双子叶植物也不例外 , 葡萄、桃和可可也是如此。全基因组加倍一般包括两 种:全基因组三倍化(whole genome triplication, WGT, y)和全基因组二倍化(whole genome duplication, WGD, α or β)。葡萄、桃和可可等双子叶植物的共同祖先在 130~150 百万年前(million years ago, Mya)经历过一 次古老全基因组的三倍化,成为六倍体祖先,葡萄、 桃和可可的全基因组的测序发现他们均没有经历过 最近的全基因组加倍(Jailon et al., 2007; Argout et al., 2011; Verde et al., 2013)。基因组发生加倍后极其不 稳定 会发生基因丢失、倒位、易位、染色体断裂和重 排等一系列遗传现象,同时也会产生大量的同源染色 体片段并且伴随产生新的基因(Tang et al., 2010)。葡 萄的化石证据表明其大约分化产生于 100~120 Mya (Manchester et al., 2013), 桃和可可的化石证据表明 两者大约分化产生于 60~90 Mya (图 1) (Motamayor et al., 2002; Su et al., 2015), 然而对于他们之间更为 准确的物种分化时间和基因组进化过程还有待于进 一步通过基因组比较来详细分析。

研究对葡萄、桃和可可的全基因组进行了深入 的比较基因组学分析,对3个物种的基因组进行了 同源结构的比较分析,确定了基因组分别经历的加 倍事件以及3个物种基因组内和基因组间的同源共 线片段;通过分析基因组间的同义核苷酸替换率的 频数分布情况,比较了3个物种的进化速率,大致确 定了各物种的分化时间,从而为3个物种基因组相 关的进化研究提供理论依据。



图 1 葡萄,桃和可可的系统进化关系

注: 左侧为系统进化树, 右侧 Ks 为对应进化事件的同义核酸 替换率

Figure 1 Phylogenetic evolution relationship between grape, peach and cacao

Note: The left part was the phylogenetic evolution tree and the right part was synonymous nucleotide substitution rate (Ks) corresponding to the evolution events

1 结果与分析

1.1基因组内同源结构

3 个物种基因组中存在大量的由同源基因构成 的片段(图 2)。其中,葡萄和桃的基因组中有较多的同 源片段,可可中的较少,葡萄中的同源片段多以短片段 存在,而桃中的同源片段较长。葡萄中较长的同源基因 片段分别存在于 Vv1-Vv14、Vv1-Vv17、Vv3-Vv18、 Vv4-Vv11、Vv5-Vv7、Vv6-Vv8、Vv8-Vv13、Vv9-Vv11、Vv15-Vv16,桃的基因组中较长的共线性基因片 段分别存在于 Pp1-Pp3、Pp1-Pp5、Pp1-Pp8、Pp2-Pp5、 Pp2-Pp6、Pp3-Pp5、Pp3-Pp7、Pp4-Pp4;可可基因组 中较长的共线性基因片段分别存在于 Tc01-Tc02、Tc01-Tc05、Tc01-Tc08、Tc02-Tc03、Tc04-Tc09、Tc04-Tc10、Tc06-Tc07、Tc09-Tc09、Tc09-Tc10。

上述 3 个物种基因组的同源片段中的基因是由 全基因组加倍产生的,是在进化历程中保留下来的 较好的同源片段。桃基因组的染色体片段较长,但同 源片段较少,在点图中呈现红色的线;可可基因组的 同源片段较短,且同源片段数量最少,大部分呈蓝色 的线,这些均是由葡萄、桃和可可的共同祖先加倍产 生的重复基因所造成。例如(图 2A),葡萄的每条染色 体的同源片段在基因组内呈现 1:2 的对应关系,分别 用红色矩形框和蓝色矩形框加以区分。这次发生在 约 130 Mya 的双子叶植物祖先中的三倍化事件,使 物种的每条染色体形成另外的两个同源拷贝,之后 又经历一系列的染色体片段的断裂、融合等过程才 形成现在的 19 条葡萄的染色体。桃、可可与葡萄类 似,其祖先基因组也经历了这次全基因组的三倍化 事件,故对应在同源性阵图(图 2B)上即为基因组内 的片段呈现 1:2 的同源关系。

在葡萄和桃基因组内的同源性点阵图中发现, 3个物种都经历了古老的双子叶植物祖先基因组的 三倍化事件,此后没有再发生共有的或独立的全基 因组加倍事件,这与之前葡萄、桃和可可的研究结 论一致 (Jailon et al., 2007; Argout et al., 2011; Verde et al., 2013)。





图 2 葡萄和桃各自基因组内同源性

注:每个点代表一对同源基因,根据基因对在染色体上的位置 打点;其中最相似的基因打点为红色,相似性次之的基因打点 为蓝色,其他为灰色,有色框中标出的是同源片段

Figure 2 Homologous dot plots of grape and peach genomes Note: Each dot represents a pair of homologous genes and dots were made on the chromosome in accordance with gene pairs; Red dots were the most similar gene pairs, blue dots were the less similar gene pairs and gray dots were the other similar gene pairs, The fragments marked with color frame were homologous

1.2 基因组间同源结构

葡萄的每条染色体均对应3条完整的桃染色体 片段(图3),如葡萄1号染色体对应桃的3条染色体 分别为1、3和5号,桃的每条染色体也均对应3条完 整的葡萄染色体片段,如桃4号染色体对应葡萄3条 染色体分别为10、12和19号染色体。在这两种情况 下3条同源染色体都可以明显地分为一条最好的 (红色实线框)和两组较差的(有色虚线框)。这说明两 个物种基因组间的同源片段表现为1:1的最好直系 同源关系以及1:2的次好种外旁系同源关系。这与二 者在基因组内的同源比对结果保持一致,古老的双子 叶祖先在经历过一次三倍化事件后,依次分化为葡 萄和桃,且葡萄和桃没有发生独立的全基因组加倍 事件,而可可的情况与前两者类似(未展示),也没有 经历最近的基因组加倍,仅经历了古老祖先的三倍化。

1.3 同源共线基因的推断与统计

采用以往的研究方法(李育先等, 2016),搜索并



图 3 葡萄,桃间同源基因结构

注: 红色的短线是 Blastp 比对两物种基因组序列获得的最相 似的基因对, 表示物种间直系同源染色体片段; 蓝色的线是由 次好相似基因构成, 表示种间旁系同源染色体片段

Figure 3 Homologous dot plots between grape and peach genomes Note: Red short lines meant the most similar gene pairs obtained from comparsion of two species' genomic sequences and represented orthology chromosome fragments between species; Blue short lines consisted of less similar gene pairs and represented paralogy chromosome fragments between species

Genomics and Applied Biology

获取了葡萄、桃、可可基因组内和基因组间的同源共 线片段和同源共线基因对的数量(表 1)。

在物种基因组内,利用前述方法,分别在葡萄、 桃、可可基因组中推断出 84、73、95 个旁系同源片段 (表 1) 这些同源片段分别含有 2 017、3 640、2 652 个 旁系同源基因对,各占自身基因组的7.6%、12.7%、 9.0%。其中 桃保留了较多的同源片段和同源基因, 可能是由于桃基因组测序比较完善,大部分基因都 锚定到了染色体上。葡萄基因组中 同源基因对含量 大于 10 的片段有 256 个 大于 50 的片段有 14 个 最 长的同源片段位于7号和5号染色体之间。含有70个 基因对 ;桃基因组中,同源基因对含量大于10的片 段有 261 个,大于 50 的片段有 55 个,最长的同源片 段位于1号和5号染色体之间,含有191个基因对; 而可可基因组中,同源基因对含量大于10的片段有 1 008 个 ,大于 50 的片段有 27 个 ,最长的同源片段 位于1号和3号染色体之间,含有122个基因对。以 上数据表明 3 个物种均没有发生独立的基因组加倍 现象,同源基因和同源片段数量相差并不大。

在 3 个物种基因组间,分别在葡萄、桃、可可中 搜索到了不同的直系同源片段(表 1),并且基于同源 片段的搜索与同义核苷酸置换率(Ks)的比较,区分了 种间直系同源片段(由物种分化形成)和种外旁系同 源片段 (由基因或基因组的加倍产生)。表中数据显

表 1 葡萄, 桃和可可基因组内, 间的同源片段, 基因对的含量 Table 1 Numbers of homologous fragments and gene pairs within and between grape, peach and cacao genomes

物种	葡萄	桃	<u>न</u> न	
Species	Grape	Peach	Cacao	
葡萄	84/2 017*	93/6 778	101/5 826	
Grape				
桃	108/2 947	73/3 640	102/8 740	
Peach				
可可	87/1 923	91/3 639	95/2 652	
Cacao				

注: 上表中处于主对角线的数字代表物种内旁系同源片段和 基因对的数量; 主对角线上侧的数字代表物种间直系同源片 段和基因对的数量; 主对角线下侧的数字代表物种间种外旁 系同源片段和基因对的数量; 同源片段/同源基因对 *

Note: Numbers on the main diagonal indicated paralogous fragments and gene pairs within a genome; Numbers above the main diagonal indicated orthologous fragments and gene pairs between two genomes; Numbers below the main diagonal indicated out-paralogous fragments and gene pairs between two genomes; Homologous fragments/homologous gene pairs * 示,在葡萄、桃和可可之间,有93~102个直系同源片 段,含有5826~8740个直系同源基因对,有87~108个 种外旁系同源片段,含有1923~3639个种外旁系同 源基因对。由此结果也显示,桃和可可之间的直系和 种外旁系同源基因比其他的要多,也在一定程度上 说明桃和可可的分化时间要晚一些,基因组的进化 上具有一定的相似性。

1.4 多基因组联合比对图谱的构建

利用推断的不同物种的直系和旁系同源关系, 建立了葡萄、桃和可可的多基因组联合比对列表,并 依此构建了圈状图来进行相应的展示(图 4)。

由于葡萄、桃和可可的基因组只经历了古老的 双子叶植物祖先的三倍化事件(ECT, γ 事件)而并未 经历独立的基因组加倍 ,葡萄、桃和可可的基因组均 可以分成3组旁系同源片段,其中每组的一条葡萄 染色体片段与桃和可可的一套染色体片段是直系同 源的关系。每圈葡萄(V)与其相邻的桃圈(P) 和可可 (T) 圈即是 3 个物种中最好的直系同源关系(图 4 中 的第1,2,3圈的染色体片段之间) 葡萄、桃和可可它 们自己的3圈染色体片段分别代表了物种基因组内 的旁系同源关系 (图 4 中的第 1, 4, 7 圈, 第 2, 5, 8 圈, 第3, 6, 9圈的染色体片段之间);每个物种的一圈 与其他物种的另外两圈染色体片段之间即为种外旁 系同源关系(图 4 中的第 1 圈的葡萄染色体片段与第 5,8圈的桃的染色体片段)。葡萄、桃和可可基因组 内,它们自己的旁系同源基因都发生了大规模的丢 失,然而3个物种间的直系同源基因则是保持了相 对的完整性,比如葡萄的后面两组(第4,第7圈)旁 系同源基因很明显不如第一组(第1圈)的完整。而 相对于葡萄的染色体(第1圈),桃和可可的染色体 (第2,第3圈)则保留了大部分的直系同源基因。以 上结果表明,在物种经历全基因组加倍事件之后,由 于进化的需要,可能伴随着大量基因片段的丢失,所 以物种基因组内的旁系同源基因才发生了大规模 的丢失。

1.5 共线性基因保留情况统计

葡萄、桃和可可由于只经历了古老双子叶植物 祖先的三倍化,三者之间的同源共线基因理论上应 该是相同的,但统计结果(表 2)显示是不同的,这说 明 3 个物种基因组的稳定性在三倍化后可能遭到破 坏,从而导致了各自的基因有不同程度的丢失,其中 葡萄的丢失最少,桃的次之(26.4%),可可丢失最为严 重(31.1%)。葡萄作为双子叶植物中的模式物种,其染 Ch# 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19



图 4 葡萄,桃和可可基因组的同源比对

注:上面的9个圈展示了葡萄(V),桃(P)和可可(T)的基因组的 直系同源,旁系同源,种外旁系同源的相关信息;最里边一圈 的不同颜色代表葡萄的19条染色体,每条染色体由众多短线 组成,每条短线代表一个基因,每条短线的颜色对应葡萄的特 定染色体;3个物种共同祖先基因组的三倍化导致每个物种 染色体至少有3圈,其中最里面的连线代表葡萄基因组内的 旁系同源关系,其他8圈代表葡萄与桃和可可基因间的直系 同源和种外旁系同源关系

Figure 4 Homologous alignment of grape, peach and cacao genomes

Note: Genomic paralogy, orthology and out-paralogy information within and between grape (V), peach (P) and cacao (T) were displayed in nine circles; The inner circle represented 19 grape chromosomes with different colour, Each grape chromosome consisted of many short lines and each of them represented a gene, A gene short line was colored in correspondence to its source chromosome of grape; Each chromosome was composed of at least three circles due to the whole genome triplication of three species' common ancestor, The curvy lines in the inner circle showed paralogy relation within grape genome and the other eight circles showed orthology and outparalogy relation between grape, peach and cacao genes

色体有 19 条,可能发生了多条祖先染色体的融合, 但保留了大量的祖先基因,基因组比较稳定。桃仅有 8 条染色体,但同源共线性基因保留的相对较好,可 能是由于其功能需要,也可能是由于基因组组装质量 较好。而可可的同源共线性基因的大量丢失,可能由 于基因组进化速率过快,稳定性处于较低水平所致。

1.6 推断物种发生的主要进化事件

物种发生的主要进化事件的时间和年代的推断 主要依据祖先在全基因组加倍事件产生的重复基

表 2 葡萄,桃和可可基因组间共线性基因保留情况统计

Table 2 Statistics of collinear genes retention between grape, peach and cacao genomes

物种	葡萄	桃	可可
Species	Grape	Peach	Cacao
葡萄	-	6 164 (73.4%)/	4 044 (68.9%)/
Grape		8 4021	5 872
桃	-	-	2 706 (72.0%)/
Peach			3 760
可可	-	-	-
Cacao			

注: 保留的共线性基因数(所占比例)/两物种间的共线性基因 总数

Note: The reserved collinear gene number (rate)/the total number of collinear gene between two species

因,而这里采用了以往研究中的类似方法来估计物种的主要进化事件(张琼等,2014; Wang et al.,2015b)。 通过计算葡萄、桃和可可基因组内的同义核酸替换 率(Ks)来估计各物种发生的主要进化事件,以不同物种基因组间的同义核酸替换率来为各个物种的分化 时间以及进化速率等提供依据。采用统计学方法绘制物种间和物种内的共线性基因同义核酸替换率 (Ks)频数折线图(图 5),其中的共线性基因利用的是物种间和物种内的重复基因。Ks 频数分布图显示,葡 萄和可可仅有一个最大峰值,分别对应 ECT 事件。 而桃的存在两个峰值,较大的峰值才对应 ECT 事件。古老的 ECT 事件所对应的 Ks 峰值在不同物种中的分布也不尽相同,这也是由不同物种进化速率的不同导致的。显然,葡萄在此峰值时的 Ks 约为 1.1 桃的约为 1.2,可可的约为 1.3。这也说明,由这次



图 5 物种基因组同源基因对的同义核苷酸替换率频数分布 折线

Figure 5 Distribution line graph of synonymous nucleotide substitution rate frequency of homologous gene pairs between grape, peach and cacao 古老三倍化产生的基因在葡萄中的进化速率最慢, 在可可中进化最快。

以上表明葡萄的进化速率最慢,因此选择葡萄作为判断物种进化速率的标准。结果表明,古老的 ECT 发生于 130 Mya,此后的葡萄分化产生于 110 Mya, 可可分化产生于 90 Mya 桃分化产生于 80 Mya。

2 讨论

研究对葡萄、桃和可可的全基因组进行了比较 基因组学分析,表明3个物种都仅经历了一次古老 的双子叶植物祖先的三倍化事件,而后并未经历共 有或单独的基因组加倍,这与文献(Jailon et al., 2007) 得出的结论一致;尽管与葡萄相比,桃中祖先三倍化 的染色体片段呈现碎片化,但也并未经历最近的全 基因组加倍(Verde et al., 2013);可可的 10 条染色体 是由于祖先的 21 条染色体中的 11 条发生了融合, 这说明可可也没有经历最近的基因组加倍(Argout et al., 2011)。

通过推断并统计物种基因组内和基因组间的同 源共线性片段,发现物种基因组内同源片段相差不 大,而物种基因组间直系要明显多于种外旁系同源 片段,文献(Wang et al., 2015b)也有类似结论,这可能 是由物种分化时同源基因的分离所致。同时,这一结 论在统计学上也是保守的,本研究在推断染色体同 源性时使用了严格的阈值限制:包括严格的 BLAST E-values 选取、去除大的重复基因家族、限制同源共 线性片段的大小以及显著性检验,而且,这一过程 已经被广泛应用于解析许多植物基因组的进化史 上(Wang et al., 2006;Ming et al., 2008; Paterson et al., 2009a; Paterson et al., 2009b; Wang et al., 2009; Wang et al., 2011a; 2011b; Paterson et al., 2012; Lee et al., 2013)。

从多物种基因组的联合比对和基因组间共线性 基因的保留情况统计 发现各物种基因组在双子叶植 物共同三倍化后稳定性明显降低(Wang et al., 2015b), 从而导致了大规模的基因丢失,且随着物种的不断 进化,基因丢失仍在继续。同时,各物种基因丢失的 分布也说明了该过程具有随机性。

通过分析 3 个物种同源基因对的同义核酸替换 率 3 个物种较大的 Ks 频数分布峰值也验证了上述 的 3 个物种只经历了古老双子叶植物祖先的三倍 化 同时也估计了 3 个物种的分化时间和顺序 :即依 次为葡萄(~110 Mya)、可可(~90 Mya)、桃(~80 Mya); 且可可的基因组进化速率最快 桃的基因组次之 ,葡 萄基因组进化最保守。

3 材料与方法

3.1数据来源

3 个物种葡萄、桃和可可的全基因组数据(表 3), 来源于各物种基因组测序的公共数据库。

3.2 基因组同源性分析

为展示基因组内与基因组间的同源性,利用 BLAST (Altschul et al., 1990)软件包中 blastp 子程序 分别对葡萄、桃和可可的蛋白序列进行基因组内和 基因组间的搜索比对,其中 E-value≤1e-5;同时结合 3 个物种中基因在染色体上的位置信息,编写 Perl 脚 本绘制同源结构点阵图,将基因间的最好匹配、次好 匹配以及其他匹配用不同的颜色标识出来,用以区 分直系同源和旁系同源对。

3.3 共线性基因的推断

利用共线性分析工具 MCscan (Tang et al., 2008) 和 ColinearScan (Wang et al., 2006),寻找 3 个物种基 因组内和基因组间的同源共线基因片段。通过比较同 源序列相似性和共线性基因数量,区分直系和旁系同 源基因。通过检查在全基因组水平上的 DNA 序列的同 一性和染色体之间共线性基因数量所反映出的 DNA 相似性,从而区分了直系和种外旁系同源区域。

3.4 多基因组联合比对图谱的构建

因葡萄基因组更古老,故以葡萄基因组中的共 线性基因作参考来建立3个物种基因组的联合比对 列表,然后通过搜索上述已生成的同源共线基因片 段结果,并结合物种基因组内的共线性基因对加以补 充,防止同源片段丢失。相关研究(Wang et al., 2015a; 2015b)表明,此方法比直接参考基因间的相似性来判 断直系和旁系同源基因要准确的多。通过 MCSCAN 软件包中的 Circular Comparer.py 程序并以已产生 的多重比对列表作为输入来生成多基因组联合比 对图谱。

3.5 基因组间共线性基因的保留情况统计

利用物种间双向 blastp 结果中的最好匹配基因 对 搜索上述多基因组联合比对列表 从而统计得出 基因组间共线性基因的保留和丢失情况。

3.6 同义核苷酸替换率(Ks)

通过运行 BioPerl (http://bioperl.org/index.html)生

	-			
物种	缩写	版本	数据来源	基因
Species	Abbreviation	Version	Data sources	Genes
葡萄	Vv	Genoscope (Aug 2007)	PGDD (http://chibba.agtec.uga.edu/duplication/)	26 346
Grape				
桃	Рр	Version 1.0	PGDD (http://chibba.agtec.uga.edu/duplication/)	28 689
Peach				
可可	Tc	Version 1.1	CocoaGen DB (http://cocoagendb.cirad.fr/)	29 452
Cacao				

表 3 全基因组数据来源

Table 3 The whole genome data sources

物统计学软件包中包含的 Nei-Gojobori 方法来估计 同源基因之间的同义核苷酸替换率(Ks)。

作者贡献

孙梦琦和闫立仁负责实验设计; 准晓波、赵康路 完成数据分析和初稿撰写; 张岚为研究的指导者, 指 导实验设计及论文的修改。全体作者都阅读并同意 最终的文本。

致谢

本研究由华北理工大学青年科学研究基金(Z20-1528)和华北理工大学创新性实验计划项目(X20162-61)共同资助。

参考文献

- Altschul S.F., Gish W., Miller W., Myers E.W., and Lipman D.J., 1990, Basic local alignment search tool, Journal of Molecular Biology, 215(3): 403-410
- Argout X., Salse J., Aury J.M., Guiltinan M.J., Droc G., Gouzy J., Allegre M., Chaparro C., Legavre T., Maximova S.N., Abrouk M., Murat F., Fouet O., Poulain J., Ruiz M., Roguet Y., Rodier-goud M., Barbosa-neto J.F., Sabot F., Kudrna D., Ammiraju J.S., Schuster S.C., Carlson J.E., Sallet E., Schiex T., Dievart A., Kramer M., Gelley L., Shi Z., Berard A., Viot C., Boccara M., Risterucci A.M., Guignon V., Sabau X., Axtell M.J., Ma Z., Zhang Y., Brown S., Bourge M., Golser W., Song X., Clement D., Rivallan R., Tahi M., Akaza J.M., Pitollat B., Gramacho K., D'Hont A., Brunel D., Infante D., Kebe I., Costet P., Wing R., McCombie W.R., Guiderdoni E., Quetier F., Panaud O., Wincker P., Bocs S., and Lanaud C., 2011, The genome of *Theobroma cacao*, Nat. Genet., 43(2): 101-108
- Jailon O., Aury J.M., Noel B., Policriti A., Clepet C., Casagrande A., Choisne N., Aubourg S., Vitulo N., Jubin C., Vezzi A., Legeai F., Hugueney P., Dasilva C., Horner D., Mica E.,

Jublot D., Poulain J., Bruyere C., Billault A., Segurens B., Gouyvenoux M., Ugarte E., Cattonaro F., Anthouard V., Vico V., Del Fabbro C., Alaux M., Di Gaspero G., Dumas V., Felice N., Paillard S., Juman I., Moroldo M., Scalabrin S., Canaguier A., Le Clainche I., Malacrida G., Durand E., Pesole G., Laucou V., Chatelet P., Merdinoglu D., Delledonne M., Pezzotti M., Lecharny A., Scarpelli C., Artiguenave F., Pe M.E., Valle G., Morgante M., Caboche M., Adam-blondon A.F., Weissenbach J., Quetier F., and Wincker P., 2007, The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla, Nature, 449(7161): 463-467

- Lee T.H., Tang H.B., Wang X.Y., and Paterson A.H., 2013, PGDD: a database of gene and genome duplication in plants, Nucleic Acids Res., 41(Database issue): 1152-1158
- Li Y.X., Xia R.Y., Wang J.P., and Wang X.Y., 2016, Analysis of collinearity and polyploidization process of grape and coffee genome, Huabei Ligongdaxue Xuebao (Journal of North China University of Science and Technology), 38(4): 38-44 (李 育先,夏瑞燕,王金朋,王希胤, 2016,葡萄与咖啡基因组 的多倍化过程及共线性分析,华北理工大学学报(自然科 学版), 38(4): 38-44)

Manchester S.R., Kapgate D.K., and Wen J., 2013, Oldest fruits of the grape family (Vitaceae) from the Late Cretaceous Deccan cherts of India, Am. J. Bot., 100(9): 1849-1859

Ming R., Hou S.B., Feng Y., Yu Q.Y., Dionne-Laporte A., Saw J.
H., Senin P., Wang W., Ly B.V., Lewis K.L., Salzberg S.L.,
Feng L., Jones M.R., Skelton R.L., Murray J.E., Chen C.X.,
Qian W.B., Shen J.G., Du P., Eustice M., Tong E., Tang H.
B., Lyons E., Paull R.E., Michael T.P., Wall K., Rice D.W.,
Albert H., Wang M.L., Zhu Y.J., Schatz M., Nagarajan N.,
Acob R.A., Guan P., Blas A., Wai C.M., Ackerman C.M.,
Ren Y., Liu C., Wang J.M., Wang J.P., Na J.K., Shakirov E.
V., Haas B., Thimmapuram J., Nelson D., Wang X.Y., Bowers J.E., Gschwend A.R., Delcher A.L., Singh R., Suzuki J.
Y., Tripathi S., Neupane K., Wei H.R., Irikura B., Paidi M.,
Jiang N., Zhang W.L., Presting G., Windsor A., Navajas-Pérez R., Torres M.J., Feltus F.A., Porter B., Li Y.J.,

Burroughs A.M., Luo M.C., Liu L., Christopher D.A., Mount S.M., Moore P.H., Sugimura T., Jiang J.M., Schuler M.A., Friedman V., Mitchell-Olds T., Shippen D.E., dePamphilis C.W., Palmer J.D., Freeling M., Paterson A.H., Gonsalves D., Wang L.,and Alam M., 2008, The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus), Nature, 452(7190): 991-996

- Motamayor J.C., Risterucci A.M., Lopez P.A., Ortiz C.F., Moreno A., and Lanaud C., 2002, Cacao domestication I: the origin of the cacao cultivated by the Mayas, Heredity (Edinb), 89(5): 380-386
- Paterson A.H., Bowers J.E., Bruggmann R., Dubchak I., Grimwood J., Gundlach H., Haberer G., Hellsten U., Mitros T., Poliakov A., Schmutz J., Spannagl M., Tang H.B., Wang X. Y., Wicker T., Bharti A.K., Chapman J., Feltus F.A., Gowik U., Grigoriev I.V., Lyons E., Maher C.A., Martis M., Narechania A., Otillar R.P., Penning B.W., Salamov A.A., Wang Y., Zhang L.F., Carpita N.C., Freeling M., Gingle A. R., Hash C.T., Keller B., Klein P., Kresovich S., McCann M.C., Ming R., Peterson D.G., Mehboob-ur-Rahman, Ware D., Westhoff P., Mayer K.F., Messing J., and Rokhsar D.S., 2009a, The Sorghum bicolor genome and the diversification of grasses, Nature, 457(7229): 551-556
- Paterson A.H., Bowers J.E., Feltus F.A., Tang H.B., Lin L.F., and Wang X.Y., 2009b, Comparative genomics of grasses promises a bountiful harvest, Plant Physiol., 149(1): 125-131
- Paterson A.H., Wendel J.F., Gundlach H., Guo H., Jenkins J., Jin D.C., Llewellyn D., Showmaker K.C., Shu S.Q., Udall J., Yoo M.J., Byers R., Chen W., Doron-Faigenboim A., Duke M.V., Gong L., Grimwood J., Grover C., Grupp K., Hu G.J., Lee T.H., Li J.P., Lin L.F., Liu T., Marler B.S., Page J.T., Roberts A.W., Romanel E., Sanders W.S., Szadkowski E., Tan X., Tang H.B., Xu C.M., Wang J.P., Wang Z.N., Zhang D., Zhang L., Ashrafi H., Bedon F., Bowers J.E., Brubaker C.L., Chee P.W., Das S., Gingle A.R., Haigler C.H., Harker D., Hoffmann L.V., Hovav R., Jones D.C., Lemke C., Mansoor S., ur Rahman M., Rainville L.N., Rambani A., Reddy U.K., Rong J.K., Saranga Y., Scheffler B.E., Scheffler J.A., Stelly D.M., Triplett B.A., Van Deynze A., Vaslin M.F., Waghmare V.N., Walford S.A., Wright R.J., Zaki E.A., Zhang T.Z., Dennis E.S., Mayer K.F., Peterson D.G., Rokhsar D.S., Wang X.Y., and Schmutz J., 2012, Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres, Nature, 492(7429): 423-428
- Su T., Wilf P., Huang Y.J., Zhang S.T., and Zhou Z.K., 2015, Peaches preceded humans: fossil evidence from SW China, Sci. Rep., 5: 16794-16814
- Tang H.B., Bowers J.E., Wang X.Y., and Paterson A.H., 2010, Angiosperm genome comparisons reveal early polyploidy

in the monocot lineage, Proc. Natl. Acad. Sci. USA, 107(1): 472-477

- Tang H.B., Wang X.Y., Bowers J.E., Ming R., Alam M., and Paterson A.H., 2008, Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps, Genome. Res., 18 (12): 1944-1954
- Verde I., Abbott A.G., Scalabrin S., Jung S., Shu S., Marroni F., Zhebentyayeva T., Dettori M.T., Grimwood J., Cattonaro F., Zuccolo A., Rossini L., Jenkins J., Vendramin E., Meisel L. A., Decroocq V., Sosinski B., Prochnik S., Mitros T., Policriti A., Cipriani G., Dondini L., Ficklin S., Goodstein D.M., Xuan P., Del F.C., Aramini V., Copetti D., Gonzalez S., Horner D.S., Falchi R., Lucas S., Mica E., Maldonado J., Lazzari B., Bielenberg D., Pirona R., Miculan M., Barakat A., Testolin R., Stella A., Tartarini S., Tonutti P., Arus P., Orellana A., Wells C., Main D., Vizzotto G., Silva H., Salamini F., Schmutz J., Morgante M., and Rokhsar D.S., 2013, The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution, Nat. Genet., 45(5): 487-494
- Wang X.W., Wang H.Z., Wang J., Sun R., Wu J., Liu S., Bai Y., Mun J.H., Bancroft I., Cheng F., Huang S.W., Li X.X., Hua W., Wang J.Y., Wang X.Y., Freeling M., Chris-Pires J., Paterson A.H., Chalhoub B., Wang B., Hayward A., Sharpe A. G., Park B.S., Weisshaar B., Liu B.H., Li B., Liu B., Tong C.B., Song C., Duran C., Peng C.F., Geng C.Y., Koh C, Lin C.Y., Edwards D., Mu D.S., Shen D., Soumpourou E., Li F., Fraser F., Conant G., Lassalle G., King G.J., Bonnema G., Tang H.B., Wang H.P., Belcram H., Zhou H.L., Hirakawa H., Abe H., Guo H., Wang H., Jin H, Parkin I.A., Batley J., Kim J.S., Just J., Li J.W., Xu J.H., Deng J., Kim J.A., Li J.P., Yu J.Y., Meng J.L., Wang J.P., Min J.M., Poulain J., Wang J., Hatakeyama K., Wu K., Wang L., Fang L., Trick M., Links M.G., Zhao M.X., Jin M., Ramchiary N., Drou N., Berkman P.J., Cai Q.L., Huang Q.F., Li R.Q., Tabata S., Cheng S.F., Zhang S., Zhang S.J., Huang S.M., Sato S., Sun S.L., Kwon S.J., Choi S.R., Lee T.H., Fan W., Zhao X., Tan X., Xu X., Wang Y., Qiu Y., Yin Y., Li Y.R., Du Y.C., Liao Y.C., Lim Y., Narusaka Y., Wang Y.P., Wang Z.Y., Li Z.Y., Wang Z.W., Xiong Z.Y., and Zhang Z.H., 2011a, The genome of the mesopolyploid crop species Brassica rapa, Nat. Genet., 43(10): 1035-1039
- Wang X.Y., Guo H., Wang J.P., Lei T.Y., Liu T., Wang Z.Y., Li Y.X., Lee T.H., Li J.P., Tang H.B., Jin D.C., and Paterson A.H., 2015a, Comparative genomic de-convolution of the cotton genome revealed a decaploid ancestor and widespread chromosomal fractionation, New Phytologist, 209(3): 1252-1263
- Wang X.Y., Shi X.L., Li Z., Zhu Q.H., Kong L., Tang W., Ge S., and Luo J.C., 2006, Statistical inference of chromosomal ho-

mology based on gene colinearity and applications to Arabidopsis and rice, BMC Bioinformatics, 7(1): 447-459

- Wang X.Y., Tang H.B., and Paterson A.H., 2011b, Seventy Million years of concerted evolution of a homoeologous chromosome pair, in parallel, in major poaceae lineages, Plant Cell, 23(1): 27-37
- Wang X.Y., Tang H.B., Bowers J.E., and Paterson A.H., 2009, Comparative inference of illegitimate recombination between rice and sorghum duplicated genes produced by polyploidization, Genome. Res., 19(6): 1026-1032

Wang X.Y., Wang J.P., Jin D.C., Guo H., Lee T.H., Liu T., and

Paterson A.H., 2015b, Genome alignment spanning major poaceae lineages reveals heterogeneous evolutionary rates and alters inferred dates for key evolutionary events, Mol. Plant, 8(6): 885-898

Zhang Q., Wang Z.Y., Ma X.L., Nie L.M., Wang H.L., and Wang J.P., 2014, Comparative analysis of paleopolyploidy evolution in genomes of *Setaria italica* and *Zea mays*, Henan Nongye Kexue (Journal of Henan Agricultural Sciences), 43 (6): 10-15 (张琼, 王振怡, 马雪莲, 聂林曼, 汪厚龙, 王金朋, 2014, 谷子和玉米基因组多倍化进化比较分析, 河南农业 科学, 43(6): 10-15)

International Journal of Super Species Research



International Journal of Super Species Research (ISSN 1927-6621) is an open access, peer reviewed journal published online by BioPublisher. The journal publishes all aspects of super species research, containing super species, plant species and their biological aspects including life histories population dynamics molecular biosystematics co-evolutionary networks among plants and other organisms and mechanisms of speciation; as well as summarizing the current understanding of the biology of species including systematics, distribution, fossil history, genetics, anatomy, physiology, behavior, ecology, and conservation. Email: edit@ijssr.biopublisher.ca

Web: http://ijssr.biopublisher.ca