

# 果树全基因组测序现状与展望

黄威剑, 李 梦\*

(南京农业大学园艺学院, 南京 210095)

**摘 要:** 全基因组测序可以获得物种的基因组序列信息, 对于探索物种起源和进化过程及基因的开发利用等研究至关重要。本文中综述了 7 种重要果树(苹果、柑橘、葡萄、梨、草莓、香蕉和桃)的全基因组测序研究进展, 探讨果树全基因组测序目前存在的问题, 并对未来果树全基因组测序、DNA 测序技术的选择、测序后的研究方向、测序数据的开放性问题进行讨论。

**关键词:** 全基因组测序; 苹果; 柑橘; 葡萄; 梨; 草莓; 香蕉; 桃

**中图分类号:** S 66

**文献标志码:** A

**文章编号:** 0513-353X (2021) 04-0733-16

## Status and Prospects Whole Genome Sequencing in Fruit Trees

HUANG Weijian and LI Meng \*

(College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China)

**Abstract:** Whole genome sequencing can obtain genome sequence information of species, which is very important for exploring the origin and evolution of the species, and fundamental for the development and utilization of genes. This article reviewed the progress of genome sequencing of seven important fruit trees (apple, citrus, grape, pear, strawberry, banana and peach), discussed the currently existing problems of whole genome sequencing of fruit trees, and elaborated selection of DNA sequencing technology, openness of sequencing data, and research after sequencing and other issues in the future research.

**Keywords:** whole genome sequencing; apple; citrus; grape; pear; strawberry; banana; peach

根据联合国粮农组织统计, 2018 年世界苹果总产量 8 614.2 万 t, 梨 2 371.4 万 t, 葡萄 7 912.6 万 t, 桃 2 445.3 万 t, 草莓 833.7 万 t, 柑橘 7 541.3 万 t, 香蕉 14 142.3 万 t。自 2007 年葡萄基因组测序完成之后, 越来越多的果树完成了全基因组测序, 为果树分子生物学的研究提供数据资源。

全基因组测序 (Whole Genome Sequencing) 是指通过高通量测序平台对细胞或组织中全部基因进行测序 (Torpdahl et al., 2010)。然后通过生物信息研究得到物种的基因组序列、基因组拼装信息, 进行基因组注释、基因功能分类、比较基因组学研究、构建数据库等。果树全基因组测序为果树基因组学研究提供了生物信息基础, 通过研究果树基因组结构、功能, 了解果树性状表达的原因, 通过分析生物信息探索果树起源和进化的过程, 开展功能基因的定位、克隆等 (zhang et al., 2020)。

果树的基因组较大, 并且高度杂合, 存在种间或者种内的遗传渐渗, 遗传背景复杂, 小规模、片段化的基因测序研究难以对其进行深入了解, 因此高通量全基因组测序对果树全基因组研究、关

收稿日期: 2020-08-20; 修回日期: 2020-11-18

基金项目: 南京农业大学校级教育教学改革研究项目园艺品牌专业专项 (2019P007)

\* 通信作者 Author for correspondence (E-mail: mli@njau.edu.cn, Tel: 025-84395262)

键基因挖掘、遗传育种等具有至关重要的作用。

2014 年, 完成测序的果树基因组较少, 多数仅完成首次测序, 乔鑫等 (2014) 对已测序的果树物种基因组进行综述。而本文是基于目前重要果树物种有多个测序基因组的情况下对其总结。文中综述了测序技术从诞生至今的发展状况, 汇总讨论了目前 7 种重要果树物种全部已测序基因组的信息, 对果树全基因组测序目前存在的问题和挑战进行了分析, 以期在选择测序方法、测序品种和测序精度上为后续研究提供参考。

## 1 测序技术发展回顾

### 1.1 第一代测序技术

1977 年, Sanger 和 Coulson 发明了 Sanger 测序技术, 该技术主要基于 DNA 双脱氧链终止法 (Chain Terminator Sequencing), 并通过结合荧光标记和毛细管阵列电泳技术来实现测序自动化, 成功得到噬菌体 X174 的全部基因组序列, 标志着第一代测序技术的出现, 推动了生物学的发展 (Velculescu et al., 1997)。Sanger 测序技术的准确率高达 99.999%, 测序读长较长, 能达到 800 ~ 1 000 bp, 平均 700 bp (表 1), 且只需 1 h 即可完成测序, 但是由于其通量低且成本高昂, 目前主要应用在检验突变位点、细菌基因组测序等方面 (Niedringhaus et al., 2011; Liu et al., 2012)。

表 1 三代测序技术指标汇总  
Table 1 Three generations of sequencing technology

测序技术代数 Sequencing technology generation	DNA 扩增方式 DNA amplification method	测序平台 Sequencing platform	平均读长/bp Average read length	正确率/% Accuracy	成本/(\$·Gb <sup>-1</sup> ) Cost	测序时间/h Sequencing time
一代 1st	质粒扩增 Plasmid amplification	Sanger	700	99.9	500 000.00	1
二代 2nd	乳液 PCR	Roche 454	700	99	9 500.00	23
	Emulsion PCR	SOLiD	120	99.9	50.00	144
	桥式 PCR	Solexa	300	99.9	20.00	19 ~ 40
三代 3rd	Bridge PCR					
	单分子测序 Single molecule sequencing	PacBio RS II MinION	10 000 200 ~ 500	85 ~ 90 70 ~ 90	1 000.00 1 000.00	0.5 ~ 6 48

### 1.2 第二代测序技术

随着人们对基因遗传信息的了解和掌握, 基因检测技术不断发展和完善, 第二代测序技术应运而生。目前市场上主要有三大测序平台 (表 1), 分别是 Roche 公司的焦磷酸测序技术 Roche 454 (Margulies et al., 2005; Rothberg & Leamon, 2008), Illumina 公司的 Solexa 平台 (Bentley et al., 2008) 和 ABI 公司的 SOLiD 测序技术 (Schuster, 2008; Valouev et al., 2008)。

第二代测序技术的核心是合成和测序同时进行 (Sequencing by Synthesis), 原理是使用不同的荧光标记区分脱氧核糖核苷三磷酸, 互补链在合成时每增加一种脱氧核糖核苷三磷酸都会放出不同的荧光信号, 以此获得待测序列的信息 (Chen et al., 2008; Shendure & Ji, 2008)。第二代测序技术测序成本大大降低, 具有高通量的优势, 一次能够产生大量数据, 并且准确性高 (99.0% ~ 99.9%), 但读长较短, 增加了拼接难度, 并且 PCR 扩增的过程可能会发生错误, 不利于进一步分析。

其中 Roche454 具有读长较长 (700 bp)、测序时间较短 (23 h) 的优点, 但是其成本较高

(Niedringhaus et al., 2011)。同为乳液 PCR 扩增方式, SOLiD 的成本较低, 但是测序时间 (144 h) 是二代测序技术中最长的 (Niedringhaus et al., 2011)。Solexa 的 DNA 扩增方式是桥式 PCR, 其成本最低, 同时测序读长比 SOLiD 更长, 测序时间仅需 19 ~ 40 h, 因此被广泛应用。

### 1.3 第三代测序技术

近年来测序技术高速发展, 为更加高效地发掘 DNA 序列信息, 改善第一、二代测序技术成本、通量等问题, 研究出了第三代测序技术——单分子测序技术 (Single molecule sequencing)。目前单分子测序平台包括 Pacific Biosciences 公司的单分子实时测序技术 (PacBio RS II) 以及 Oxford Nanopore Technologies 公司的单分子纳米孔测序技术 (The single molecule nanopore DNA sequencing, MinION)。

PacBio RS II 使用单分子实时测序技术 (Single molecule real-time, SMRT), 其最大特点是不需要进行 PCR 扩增过程, 并且可以对高 GC 含量的区域进行测序 (Ardui et al., 2018), 以及超长读长 (10 kb)、测序所需时间短、直接检测表观修饰位点等优点 (Rhoads & Au, 2015)。但也存在劣势, 例如由于相机帧数赶不上碱基的掺入速度, 导致缺失错误, 通量较低, 存在插入删除错误、均聚物错误和随机性等问题 (Rhoads & Au, 2015; Ji et al., 2017)。SMRT 测序的正确率大约为 85% ~ 90% (表 1), 碱基错测率为 1%, 但是可以通过增加测序的覆盖度来提高准确度。当测序深度达到 15 $\times$ , 组装的序列准确度可达到 99% 以上 (Eid et al., 2009)。目前 SMRT 在小型基因组的从头测序 (*de novo*) 和完整组装中表现出良好的前景。

MinION: 2014 年, Oxford Nanopore Technologies (ONT) 公司发布一款商用的单分子纳米孔测序技术的测序仪——MinION (Laver et al., 2015)。MinION 测序仪可以通过测量 DNA 链通过生物孔时电导率的变化从而识别 DNA 碱基, 其质量仅 90 g, 同时具有快速、低价等优点, 因此引起了广泛的关注 (Lu et al., 2016)。

总的来说, 第三代测序技术的改进在于: 读长增加, 每次读取的读取长度从数十个碱基增加到数万个碱基; 成本降低, 所需的样品量和试剂少; 测序时间减少, 无需洗涤以及扫描的过程, 将测序时间从几天减少到几小时; 降低误差, 减少或消除 PCR 扩增引入的测序偏差。但是正确率还难以达到二代技术的水平, 所以二、三代测序技术常常结合使用。

新的测序技术正在不断发展, 但这并不意味着第一、二代测序技术的退出, 各种测序技术各有特点, 应根据测序内容选择合适的方法。现阶段中, 应用最多的是 PacBio 和 Illumina HiSeq, 这两种技术各有所长, 因此常常被结合使用。例如葡萄 ‘FPS04’ 和 ‘Chardonnay’、苹果 ‘金冠’ 双倍体 ‘GDDH13’、草莓 ‘Camarosa’、梨 ‘山西杜梨’、柑橘 ‘Hongkong Kumquat’ 等品种的测序。其中草莓 ‘Camarosa’ 的测序过程中使用了 PacBio 和 Illumina HiSeq 以及 10X Genomics 技术。

## 2 果树全基因组测序研究进展

### 2.1 苹果

2019 年, 中国苹果栽培面积和产量超过世界总面积和产量的 50%, 因此苹果的测序研究对中国农业的发展具有重要意义。

‘金冠’ 苹果 (Golden Delicious) 又名金帅、黄香蕉和黄元帅, 具有产量高、稳定、易管理的特点。2010 年, Velasco 等 (2010) 利用 Sanger 和 Roche454 测序平台对二倍体 ‘金冠’ 苹果进行了

首次测序, 组装基因组大小为 603.9 Mb(表 2), 占苹果基因组的 81.4%, 组装基因组由 1 629 个 Scaffold 搭建而成, Contig N50 仅 16.7 kb, 测序深度 16.9 $\times$ , 重复序列比例较高, 达到了 67%, 这或许是导致金冠苹果基因组较大的原因, 该组装体对于基因组非重复部分的覆盖仅占 89%, 并且 Contig N50 较短, 组装体连续性不佳。该研究发现栽培苹果的祖先是 *M. sieversii*, 提出果实发育有关基因家族的扩展或许解释了梨果的形成。

对于苹果新品种的培育而言, 高质量的基因组信息尤为重要。2016 年, Li 等(2016)利用 Illumina HiSeq 和 PacBio RS II 测序平台对二倍体‘金冠’苹果进行了从头测序, 相比于 Velasco 等(2010)的测序结果, 多组装了 27.9 Mb(表 2), 达到了 632.4 Mb, 由 1 629 个 Scaffold 组成, 占基因组的 90%, Contig N50 提升了近 7 倍, 达到了 111.6 kb, 测序深度达到了 131 $\times$ , 该组装体质量较高, 可用于 RNA 序列和重测序研究以及基因组的编辑和基因克隆等。

2017 年, Daccord 等(2017)对‘金冠’苹果双单倍体‘GDDH13’进行了从头测序, 包含 42 140 个蛋白质编码基因(表 2), 发现 93% 的注释基因具有转录证据。利用 BioNano 光学图谱进行了组装, 组装基因组由 280 个 Scaffold 构成, 大小为 649.7 Mb, Contig N50 达到了 699 kb, 测序深度在苹果中最高, 达到了 835 $\times$ , 测序深度比 Li 等(2016)的提高了 5.37 倍。该研究中通过深度的转座子注释, 预测苹果在 2 100 万年前经历过转座子(transposable elements)爆发, 而这个时间点恰好与天山山脉隆起的时间点相近, 这可能和苹果的多样化有关, 同时还研究制作了基因组甲基化图谱。

2019 年, Zhang 等(2019)结合 Illumina HiSeq、PacBio 测序平台、BioNano 光学图谱和 HI-C 染色体构象捕获技术对三倍体苹果‘寒富’的纯合子系‘HFTH1’的基因组进行测序组装, 组装全长 658.9 Mb(表 2), 占基因大小(708.5 Mb)的 93%, 对 44 677 个蛋白质编码基因进行鉴定, 组装完整性较高, Contig N50 达到 6.99 Mb, Scaffold 仅 502 个, 同时测序深度达到 552 $\times$ , 整体组装质量高, 同时检测到大量的基因组变异, 发现了苹果着色有关的核心转录激活因子。

## 2.2 柑橘

2013 年, Xu 等(2013)完成对二倍体甜橙‘Sweet orange’基因组草图的分析。该基因组的 20% 由重复序列构成(表 2), 包含 29 445 个蛋白质编码基因, 基因组大小为 367 Mb, 组装的序列覆盖了 87.3% 的基因组, 测序深度达到了 214 $\times$ , 基因组杂合度为 50%。通过测序结果和比较分析, 揭示了甜橙是柚和柑橘的回交杂种后代, 并且解释了甜橙高维生素 C 含量的原因。该基因组草图对于柑橘性状改进提供了重要的信息。

2014 年, Wu 等(2014)完成单倍体克里曼丁橘‘Clemenules’的测序。该基因组大小为 310 Mb, 由 2 931 个 Scaffold 组成, 测序深度达到了 6.94 $\times$ , 重复序列比例 44.7%, 包含 24 533 个蛋白质编码基因。该研究揭示了柑橘复杂的驯化史, 探究了柑橘起源, 明确了分类学关系。

2016 年, Zhang 等(2016)利用 Illumina HiSeq 测序技术完成了枳柚‘Swingle Citrumelo’基因组草图的测序和组装。该基因组组装大小为 280.6 Mb, 由 66 319 个 Scaffold 组成, 覆盖了 74%(380 Mb)的基因组, 重复序列占组装基因组大小的 16.8%, Contig N50 为 11.4 kb, 测序深度为 15 $\times$ 。该研究同时对柑橘枯萎根系进行了转录组分析, 通过差异基因表达分析发现了感染枯萎病对柑橘会出现淀粉积累、干旱胁迫等现象, 在分子层面对柑橘枯萎病的表型进行解释, 有利于进一步了解柑橘的抗性。

2017 年, Wang 等(2017)完成对单倍体柚‘Pummelo’的测序, 该基因组大小为 380.8 Mb, 由 1 612 个 Scaffold 组成, 组装基因组大小为 344.8 Mb, 最大的 Scaffold 达到了 14.3 Mb, Contig N50

为 2 180 kb, 对 30 123 个蛋白质编码基因进行鉴定, 平均测序深度为 427 $\times$ 。该研究中对隔离柑橘种群和自然柑橘种群的分析为柑橘的无融合生殖提供了新观点, 发现 *CitRWP* 基因会影响柑橘的无融合生殖。

同年, Shimizu 等 (2017) 利用 Illumina HiSeq 及 PacBio RS II 测序平台完成对二倍体温州蜜橘 ‘Satsuma’ 的全基因组测序。Satsuma 组装基因组由 20 876 个 Scaffold 构成 (表 2), 大小为 359 Mb, 比其他柑橘品种更大一些, 重复序列比例达到 39.5%, 并对 20 876 个蛋白质编码基因进行了注释。在该研究中发现温州蜜橘 ‘Satsuma’ 是 Kishu 的回交后代, 并且证明目前已有能力构建精确的杂合二倍体基因组。

‘莽山野柑’ 是一类独特的柑橘种质资源, 其栽培历史超过 4 000 年。2018 年, Wang 等 (2018) 完成二倍体莽山野柑 (‘Mangshan Mandarin’) 的测序, 其基因组由 42 714 个 Scaffold 组成, 组装大小为 334 Mb, Contig N50 为 24.7 kb, 平均测序深度 199.7 $\times$ , 重复序列比例为 50.1%, 并对 28 820 个蛋白质编码基因进行了鉴定。结果表明莽山野柑属于原始类型, 经过两次独立驯化事件, 导致产生 MD1 和 MD2 两个栽培种, 为野生柑橘的起源提供了遗传证据。

2019 年, Zhu 等 (2019) 完成二倍体 ‘香港金橘’ (‘Hongkong Kumquat’) 基因组的从头测序和组装, 其基因组大小为 389 Mb, 组装基因组由 900 个 Scaffold 组成, 大小约为 373.6 Mb, 相当于 96% 的基因组大小, 测序深度为 145 $\times$ , Contig N50 为 2.2 Mb, 重复序列比例 14.3%, 包含 32 257 个蛋白质编码基因。通过对蛋白质编码基因和基因组的比较基因组分析, 发现 Hongkong Kumquat 基因组中 96.9% 的基因在柑橘亚族中具有同源性, 并且由于该品种在生长期等方面的优势, 表明该种质在基因研究中具有继续挖掘的潜力。

## 2.3 葡萄

‘PN40024’ 葡萄最初为 ‘黑比诺’ (Pinot Nior) 的一个种系, 果实黑色, 是一个古老的酿酒品种。2007 年, Jaillon 等 (2007) 利用 Sanger 测序技术对高度纯合 (约 93%) 的单倍体 ‘PN40024’ 葡萄品种完成了首次测序, 因为是利用第一代测序技术, 读长较短, Contig N50 值仅为 65.9 kb (表 2), 组装的基因组由 3 514 个 Scaffold 构成, 大小为 487.1 Mb, 重复序列比例达到了 41.4%, 测序深度 8.4 $\times$ 。‘PN40024’ 的基因组没有经历最近的全基因组复制, 因此能够在该基因组中发现古老的性状, 提供宝贵的遗传资源。同时, 该结果中解释了开花植物进化过程中全基因组复制事件的时间顺序。

2016 年, Chin 等 (2016) 利用 PacBio RS II 对 ‘赤霞珠’ (‘Cabernet Sauvignon’) 进行测序, 并用 FALCON 进行组装, 组装基因组大小为 591.4 Mb, 由 1 314 个 Scaffold 构成, Contig N50 达到了 2 170 kb, 最长 Contig 为 14 Mb, 基于 FALCON 的组装体更加连续和完整, 在测序质量上有了较大的提升 (表 2)。

2018 年, Roach 等 (2018) 将 PacBio RS II、Illumina HiSeq 以及 MiSeq 测序技术结合, 对二倍体 ‘霞多丽’ (‘Chardonnay’) 进行了测序 (表 2), 组装基因组大小为 490 Mb, 并对 29 675 个蛋白质编码基因进行注释, Contig N50 达到了 935 kb, 测序深度 115 $\times$ , 经过生物信息分析, 提出 ‘霞多丽’ 基因组显示近交特征的原因是 ‘黑皮诺’ 和 ‘Gouais Blanc’ 之间存在极近的亲缘关系。

‘FPS 04’ 是 ‘霞多丽’ 葡萄的无性繁殖体。2019 年, Zhou 等 (2019) 使用了 PacBio RS II 和 Illumina HiSeq 测序平台对二倍体 ‘FPS 04’ 进行测序组装 (表 2), 该基因组包含 38 020 个蛋白质编码基因, 基因组大小为 606 Mb, 由 684 个 Scaffold 组装而成, Contig N50 达到了 1 240 kb, 测序深度达到 220 $\times$ 。在该研究中, 将 ‘FPS04’ 与已测序的 3 个葡萄品种 (‘霞多丽’ ‘赤霞珠’ 和 ‘黑

皮诺’)的基因组进行综合比较,发现葡萄基因组中存在 15%的染色体结构变异,并提出负选择对染色体结构变异类型有影响。

## 2.4 梨

2013年,Wu等(2013)利用了Illumina HiSeq测序技术对二倍体砀山酥梨进行了全基因组测序,得到高质量的基因组图谱。砀山酥梨组装基因组(表2)由2 103个Scaffold构成,大小为512 Mb,占基因组大小的97.1%,重复序列比例(53.1%)较高,达到271.9 Mb,基因组杂合度为1.99%,共包含42 812个基因,在该研究中探索了梨石细胞形成、气味释放、糖积累等生物学过程的分子机理。

继东方梨测序后,Chagne等(2014)利用Roche 454完成了单倍体西洋梨‘Bartlett’基因组序列草图的绘制,‘Bartlett’的组装基因组由142 083个Scaffold构成,总计覆盖577.3 Mb,占基因组大小(600 Mb)的96.2%,包含43 419个基因,比大多数植物的基因数量要多,Contig N50为6.53 kb(表2)。

2019年,Linsmith等(2019)利用PacBio RS II测序技术、Bionano光学图谱以及HI-C染色体构象捕获技术完成了对双单倍体品种‘BartlettDHv2.0’的测序和高质量组装。组装的基因组大小为496.9 Mb,由494个Scaffold组装而成,重复序列占组装基因组大小的49.7%(247 Mb),Contig N50相较于Chagne等(2014)的‘Bartlett’基因组测序结果提升了近千倍,达到了5 300 kb(表2)。

野生杜梨在中国北方广泛分布,对逆境的抵抗能力非常强,具备良好的嫁接亲和性,常常作为梨栽培的砧木使用,是梨属植物的优良抗性性状来源。2020年,Dong等(2020)利用Illumina HiSeq和PacBio测序平台,结合BioNano光学图谱以及Hi-C染色体构象捕获技术完成了对‘山西杜梨’的从头测序(*de novo*)。其基因组组装大小为532.7 Mb(表2),包括247.4 Mb的重复序列和59 522个蛋白质编码基因,Contig N50为1 570 kb。发现次生代谢物中存在显著富集的扩增基因,以及150个TIR-NBS-LRR(TNL)型基因,解释了‘山西杜梨’为何具有较强的环境适应性和抗病性,提出了花青素还原酶(ANR)代谢途径会影响梨果实的涩味,山梨糖醇转运蛋白(SOT)跨膜转运可能影响可溶性有机物的积累。

## 2.5 草莓

2011年,Shulaev等(2011)选利用Roche 454、Illumina Solexa以及Technologies SOLID技术对二倍体森林草莓‘Hawaii 4’的基因组进行了测序,并用Celera Asse Mbler完成组装,组装全长仅为209.8 Mb(表2)。在其基因组中重复序列比为23%,鉴定了34 809个基因,测序深度为39×,并确定了多个对草莓开花时间、口感、营养价值等方面有影响的基因。

2014年,Hirakawa等(2014)通过Roche 454以及Illumina HiSeq技术对八倍体日本草莓品种‘Reikou’进行了基因组测序,用Newbler 2.7进行组装,序列同一性达到90%,组装基因组大小为698 Mb(表2),是先前测得的‘Hawaii 4’基因组的3.3倍,由RepeatMasker识别得出的重复序列总长度为8 697 730 bp,占总序列长度的5%。

2019年,Edger等(2019)通过Illumina HiSeq和PacBio RS II技术以及10X Genomics对主流品种八倍体‘Camarosa’进行了短读和长读组合测序,覆盖基因组共计615×(表2),注释108 087个基因,与‘Reikou’相比,‘Camarosa’的组装更为完整,质量更高。

黄毛草莓(*Fragaria nilgerrensis*)‘Ruegen’起源于中国云南地区。2020年,Zhang等(2020)结合PacBio和Illumia HiSeq测序平台以及HI-C染色体构象捕获技术完成对二倍体‘Ruegen’的测

序和高质量组装。测序深度达到 152×, 基因组组装全长 270.3 Mb, 包含 28 780 个蛋白质编码基因, 总序列 Scaffold 数仅有 257 个, Contigs N50 达到 85 000 kb, 重复序列比例为 43.4%, 基因组装配的完整度较高。仅有不到 2% 的 CEG (核心真核基因) 无法检测, 95% 的基因具有完全的覆盖。该研究发现 *FnMYB10* 的低水平表达可能与草莓果实颜色相关。

## 2.6 其他果树

2012 年, D'Hont 等 (2012) 为改良香蕉的遗传基因, 整合了 Sanger、Roche 454 和 Illumina 技术对野生香蕉双单倍体 ‘DH-Pahang’ (*Musa acuminata*) 进行了全基因组测序。其组装大小为 472.2 Mb, 占基因组 (523 Mb) 大小的 90%, 其中包含了 24 425 个 Contig 和 7 513 个 Scaffold, 测序深度达到 50× (表 2), 对 36 542 个蛋白质编码基因进行鉴定后, 发现了 235 个 microRNA, 其中有 8 个 microRNA 先前被认为是禾本科植物的特征。并发现单子叶植物和真双子叶植物分化前的保守非编码序列 (Conserved non-coding sequences)。

2013 年, Davey 等 (2013) 利用 Illumina 测序平台完成野生香蕉二倍体 ‘Pisang Klutuk Wulung’ (*M. balbisiana*) 的测序, 该基因组大小为 432.7 Mb, 组装大小为 341.4 Mb, 基因数量与 ‘DH-Pahang’ 相似, 有 36 638 个, 重复序列比例 29.5%, 测序深度为 41.1× (表 2)。发现 ‘Pisang Klutuk Wulung’ 与目前主流香蕉栽培品种具有极近的亲缘关系, 该基因组的开发将有利于香蕉抗性方面的研究。

2016 年, Wu 等 (2016) 使用 Illumina HiSeq 2000 对二倍体阿宽蕉 (*M. itinerans*) 进行了测序, 并利用 SOAPdenovo2 进行了组装, 基因组大小为 615.2 Mb, 组装基因组 (462.1 Mb) 大小占 75.2%, 包含 32 456 个蛋白质编码基因, 基因组杂合度为 0.25%, 测序深度相较于 ‘DH-Pahang’ 提升较大, 达到 120.7× (表 2)。野生阿宽蕉是芭蕉属中抗寒性和抗病性最强的品种之一, 因此对阿宽蕉的测序将利于挖掘香蕉抗性基因。研究中发现纯化选择在消除有害突变和稳定 MYB 转录因子方面具有积极作用。并为芭蕉属谱系特异性基因组的进化研究提供了重要的资料。

2019 年, Wang 等 (2019) 利用 Illumina HiSeq 和 PacBio 测序平台以及 HI-C 染色体构象捕获技术完成双单倍体 ‘DH-PKW’ (*M. balbisiana*) 基因组测序组装, 测序深度达到 417×, DH-PKW 的基因组组装大小为 430 Mb, Contig N50 达到了 1 830 kb, Scaffold 数仅有 294 个, 基因组完整性加强, 重复序列比例为 55.8%, 包含 35 148 个蛋白质编码基因。根据遗传背景将香蕉划分为 *M. acuminata* (A 基因组) 和 *M. balbisiana* (B 基因组), 目前的栽培三倍体基因型香蕉主要来自与 A 基因组和 B 基因组的杂交, 通过共线性分析比较 A 基因组和 B 基因组的结构变异, 发现两个基因组之间共线性和序列相似性较高, 并发现 B 基因组中 *ACO* 基因的扩增导致香蕉成熟期乙烯生物合成加强。

桃 (*Prunus persica*) 在中国的栽培面积和总产量均排世界第一 (Li & Wang, 2020)。2013 年, Verde 等 (2013) 利用 Sanger WGS 测序技术完成了对二倍体桃品种 ‘Lovell’ 的高质量基因组草图的制作, 该基因组大小为 265 Mb (表 2), 由 202 个 Scaffold 组装而成, 组装大小为 224.6 Mb, Contig N50 为 294 kb, 预测了 27 852 个蛋白质编码基因和非编码 RNA, 测序深度为 8.47×。通过该研究加深了对桃驯化历史的了解, 发现遗传瓶颈效应对桃基因组的多样性存在影响, 指出桃没有经历最近的全基因组复制事件。

表 2 7 种已测序果树汇总  
Table 2 Summary of seven sequenced fruit trees

物种 Species	测序材料 Sequencing material	测序方法 Sequencing method	基因组/ Mb Genome size	组装全 长/Mb Assembly	基因 数量 Number of gene	重复 序列 比率 /% Repeat	总序列 骨架数 Scaffold Total	Contig N50/kb	基因组杂 合度/% Genomic Hetero- zygosity	测序深度 Sequencing depth	倍性 Ploidy	参考文献 Reference
苹果 Apple ( <i>Malus × domestica</i> )	金冠 Golden Delicious	Sanger, Roche454	742.3	603.9	57 386	67	1 629	16.7	缺失 Deficiency	16.9×	二倍体 Diploid	Velasco et al., 2010
	金冠 Golden Delicious	Illumina HiSeq, PacBio RS II	701	632.4	53 922	60	缺失 Deficiency	111.6	缺失 Deficiency	131×	二倍体 Diploid	Li et al., 2016
	金冠双单倍体 GDDH13	Illumina HiSeq, PacBio RS II	651	649.7	42 140	57.3	280	699	缺失 Deficiency	835×	双单倍体 Double haploid	Daccord et al., 2017
	寒富 HFTH1	Illumina HiSeq, PacBio RS II	708.5	658.9	44 677	59.8	502	6 990	缺失 Deficiency	552×	三倍体 Triploid	Zhang et al., 2019
柑橘 Citrus	甜橙 Sweet Orange ( <i>C. sinensis</i> )	Illumina GAII	367	320.5	29 445	20	4 811	49.89	50	214×	二倍体 Diploid	Xu et al., 2013
	克里曼丁橘 Clemenules ( <i>C. clementina</i> )	Sanger	302	310	24 533	44.7	2 931	115.9	缺失 Deficiency	6.94×	单倍体 Haploid	Wu et al., 2014
	枳柚 Swingle Citrumelo ( <i>C. paradise</i> Macf. × <i>Poncirus trifoliata</i> (L.) Raf.)	Illumina HiSeq	380	280.6	29 054	16.8	66 319	11.4	缺失 Deficiency	15×	不可用 Not Available	Zhang et al., 2016
	柚 Pummelo ( <i>C. grandis</i> )	Illumina HiSeq, PacBio RS II	380.8	344.8	30 123	45.8	1 612	2 180	缺失 Deficiency	427×	单倍体 Haploid	Wang et al., 2017
	温州蜜橘 Satsuma ( <i>C. unshiu</i> Marc.)	Illumina HiSeq, PacBio RS II	不可用 Not Available	359	29 024	39.5	20 876	缺失 Deficiency	0.44	缺失 Deficiency	二倍体 Diploid	Shimizu et al., 2017
	莽山野柑 Mangshan Mandarin ( <i>C. reticulata</i> )	Illumina HiSeq	344.3	334	28 820	50.1	42 714	24.7	缺失 Deficiency	199.7×	二倍体 Diploid	Wang et al., 2018
	香港金橘 Hongkong Kumquat ( <i>Fortunella hindsii</i> )	Illumina HiSeq, PacBio	389	373.6	32 257	14.3	900	2 200	0.8	145×	二倍体 Diploid	Zhu et al., 2019
葡萄 Grape ( <i>Vitis vinifera</i> )	黑皮诺 PN40024	Sanger	475	487.1	30 434	41.4	3 514	65.9	2.6	8.4×	单倍体 Haploid	Jaillon et al., 2007
	赤霞珠 Cabernet Sauvignon	PacBio RS II	633	591.4	36 687	51.1	1 314	2 170	缺失 Deficiency	缺失 Deficiency	二倍体 Diploid	Chin et al., 2016
	霞多丽 Chardonnay	Illumina HiSeq, PacBio RS II, MiSeq	580	490	29 675	38.7	缺失 Deficiency	935	缺失 Deficiency	115×	二倍体 Diploid	Roach et al., 2018
	霞多丽 FPS 04	Illumina HiSeq, PacBio RS II	600	606	38 020	47.3	684	1 240	缺失 Deficiency	220×	二倍体 Diploid	Zhou et al., 2019



续表 2

物种 Species	测序材料 Sequencing material	测序方法 Sequencing method	基因组/ Mb Genome size	组装全 长/Mb Assembly	基因 数量 Number of gene	重复 序列 比率 /% Repeat	总序列 骨架数 Scaffold Total	Contig N50/kb	基因组杂 合度/% Genomic Hetero- zygosity	测序深度 Sequencing depth	倍性 Ploidy	参考文献 Reference
梨 Pear	砀山酥梨 Dangshan Suli ( <i>Pyrus bretschneideri</i> Rehd.)	Illumina HiSeq	527	512	42 812	53.1	2 103	35.7	1.99	194×	二倍体 Diploid	Wu et al., 2013
	西洋梨 Bartlett ( <i>P. communis</i> L.)	Roche 454 HiSeq	600	577.3	43 419	34.1	142 083	6.53	缺失 Deficiency	11.4×	单倍体 Haploid	Chagne et al., 2014
	西洋梨 BartlettDHv2.0 ( <i>P. communis</i> L.)	PacBio SII	528	496.9	37 445	49.7	494	5 300	缺失 Deficiency	63×	双单倍体 Double haploid	Linsmith et al., 2019
	山西杜梨 Shanxi Duli ( <i>P. betuleafolia</i> )	Illumina HiSeq, PacBio	511	532.7	59 522	46.4	139	1 570	1.54	缺失 Deficiency	缺失 Deficiency	Dong et al., 2020
草莓 Strawberry	森林草莓 Hawaii 4 ( <i>Fragaria vesca</i> )	Roche 454, SOLiD, Illumina HiSeq	240	209.8	34 809	23	3 263	13 000	缺失 Deficiency	39×	二倍体 Diploid	Shulaev et al., 2011
	Reikou ( <i>F. ananassa</i> )	Roche 454, Illumina HiSeq	692	698	45 377	5	7 598	46.8	缺失 Deficiency	缺失 Deficiency	八倍体 Octoploid	Hirakawa et al., 2014
	卡麦罗莎草莓 Camarosa ( <i>F. ananassa</i> )	Illumina HiSeq , PacBio RS II	813.4	660	108 087	36	25 426	79.9	缺失 Deficiency	615×	八倍体 Octoploid	Edger et al., 2019
	黄毛草莓 Ruegen ( <i>F. nilgerrensis</i> )	Illumina HiSeq, PacBio SMRT	276	270.3	28 780	43.4	257	85 000	缺失 Deficiency	152×	二倍体 Diploid	Zhang et al., 2020
	彭亨香蕉 DH-Pahang ( <i>Musa acuminata</i> )	Sanger, Roche454, Illumina GAIIx	523	472.2	36 542	44	7 513	43.1	缺失 Deficiency	50×	双单倍体 Double haploid	D'Hont et al., 2012
香蕉 Banana	Pisang Klutuk Wulung ( <i>M. balbisiana</i> )	Illumina HiSeq	432.7	341.4	36 638	29.5	63 245	7.9	缺失 Deficiency	41.4×	二倍体 Diploid	Davey et al., 2013
	阿宽蕉 ( <i>M. itinerans</i> )	Illumina HiSeq	615.2	462.1	32 456	38.9	7 194	33.9	0.25	120.7×	二倍体 Diploid	Wu et al., 2016
	DH-PKW ( <i>M. balbisiana</i> )	Illumina HiSeq, PacBio, Sanger	492	430	35 148	55.8	294	1 830	缺失 Deficiency	417×	双单倍体 Double haploid	Wang et al., 2019
桃 Peach ( <i>Prunus persica</i> )	Lovell	Sanger	265	224.6	27 852	37.14	202	294	缺失 Deficiency	8.47×	二倍体 Diploid	Verde et al., 2013

3 果树全基因组测序研究展望

1990 年人类基因组的测序耗资 30 亿美元，随着测序技术的发展，测序成本不断降低，在保证精度、准确率、时间条件下，越低的成本意味着越多的物种会进一步的被了解。此外还有测序质量的问题，测序质量的提高不单依靠测序这一步骤，而在于整个建库、测序、比对、组装、变异检测、注释的过程。建立不同级别的库，可以用于解决数据分析时重复序列带来的问题。突破拼接组装算

法的障碍让序列更加完整、精确。而测序读长若能达到基因组全长, 则能得到完全准确的基因组数据, 无需拼接。

考虑到成本、时间、通量等问题, 目前小规模测序通常使用第一代测序技术。二代测序技术在转录组分析、测序、基因组重测序、从头测序、基因表达分析、小分子 RNA、基因型分析等领域应用广泛。第三代测序技术因其读长以及无需 PCR 扩增的优势, 可显著提高组装质量和完整性, 减少 Contig 数量以及补充 Scaffold 中 GAP 区域 (Zhang et al., 2012), 并且适用于高 GC 区域的测序, 目前在全基因组测序、甲基化研究、SNP 检测、RNA-Seq 等方面普遍应用。

### 3.1 物种间全基因组数据比较

上述 7 个物种中基因组最大的是葡萄 ‘Camarosa’ (813.4 Mb)。组装基因组最大的是草莓 ‘Reikou’ (698 Mb), 最小的是草莓 ‘Hawaii’ (209.8 Mb)。重复序列比例最高的是苹果 (67%), 导致其基因组普遍较大 (Imelfort & Edwards, 2009)。重复序列比例最低的是草莓 ‘Reikou’ (5%)。Scaffold 最多的是应用 Roche 454 组装的西洋梨 ‘Bartlett’ (142 083), 最少的是应用 PacBio 测序技术和 BioNano 光学图谱以及 HI-C 染色体构象捕获技术混合组装的 ‘山西杜梨’ (139)。Contig N50 值最低的是西洋梨 Bartlett (6.53 kb), 最高的是苹果三倍体寒富纯合子系 HFT1 (6.99 Mb)。基因组杂合度最高的是甜橙 (50%), 最低的是阿宽蕉 (*M.itinerans*) (0.25%)。测序深度最高的是苹果 ‘金冠’ 双单倍体 ‘GDDH13’ (835 $\times$ ), 最低是桃 ‘Lovell’ (8.47 $\times$ )。除去 2 个缺失染色体倍性信息的品种外, 在上述完成测序的 28 个果树品种中, 15 个品种是二倍体, 单倍体 4 个 (柑橘 2 个, 葡萄 1 个, 梨 1 个), 仅有 1 个三倍体 (苹果), 双单倍体 4 个 (香蕉 2 个, 苹果 1 个, 梨 1 个), 八倍体 2 个 (草莓 2 个)。

### 3.2 果树测序数据的分析和应用

对果树进行全基因组测序只是研究的一个起点, 可以推进物种的起源、进化研究, 也可以应用于对果树品种抗性、植物发育、性状、遗传特性、代谢等相关关键基因的研究。

#### 3.2.1 果树物种的起源和进化

‘金冠’ 苹果 (Velasco et al., 2010) 基因组草图完成的同时, 确定了栽培苹果的祖先是 *M.sieversii*, 通过对果树发育相关的基因进行研究, 可能解释了梨果的形成。‘金冠’ 双单倍体 ‘GDDH13’ (Daccord et al., 2017) 测序后, 对苹果的起源进化、DNA 甲基化与果实发育、遗传标记进行研究, 提出苹果的起源地在天山山脉, 同时提出转座子 (TE) 与物种形成及生物多样性有关。

对甜橙 (Xu et al., 2013) 测序, 同时对其起源进化、维生素 C 代谢基因进行了研究, 证明甜橙起源于柚和橘的回交杂种。Mangshan Mandarin (Wang et al., 2018) 测序后对柑橘的起源、进化以及地理起源进行了研究, 指出柑橘驯化地点在华南地区。

对 ‘霞多丽’ 葡萄 (Roach et al., 2018) 测序后发现其起源于亲本近交的证据, 对 ‘霞多丽’ 基因组中核苷酸变异模式的研究结果表明 ‘Gouais Blanc’ 和 ‘Pinot noir’ 亲缘关系极近, 同时对 ‘霞多丽’ 葡萄的驯化过程进行解析。对 ‘FPS 04’ 葡萄 (Zhou et al., 2019) 测序后发现葡萄染色体结构变异十分普遍, 并且定向选择会对结构变异造成影响, 但是在不同类型结构变异间存在差异, 这加深了对葡萄基因组进化的理解。

对 ‘Bartlett’ 梨 (Chagne et al., 2014) 测序后证明西洋梨与中国梨和苹果的基因组密切相关, 亲缘关系较近。

对草莓品种 ‘Camarosa’ (Edger et al., 2019) 测序后研究了其起源进化过程, 并进行系统发育

分析, 提供了八倍体草莓起源于北美的证据。

对香蕉品种 ‘DH-Pahang’ (D'Hont et al., 2012) 的测序加深了对禾本科植物特征的研究, 揭示了单子叶植物的系统发育关系。

对桃品种 ‘Lovell’ (Verde et al., 2013) 测序后, 研究桃基因组的进化以及遗传瓶颈, 并探索了桃的驯化历程, 发现遗传瓶颈对桃基因组多样性的影响较大。

### 3.2.2 果树功能基因组学与基因挖掘

Velasco 等 (2010) 测序后研究指出 ‘金冠’ 苹果具有较多与山梨醇代谢有关的关键基因, 例如: *A6PR*、*SDH*、*PcSOT* 等共计 71 个山梨糖醇代谢基因。对三倍体苹果 ‘寒富’ 的纯合子系 ‘HFT11’ (Zhang et al., 2019) 测序后, 对基因组变异、果皮颜色方面进行研究, 发现 *MdMYB1* 上游存在与苹果红色果皮有关的核心转录激活因子。

Xu 等 (2013) 对甜橙 (Sweet Orange) 测序后对维生素 C 代谢基因进行了研究, 指出 *GalUR* 基因与 AsA 的累积存在关联。Zhang 等 (2016) 对 Swingle Citrumelo 测序后, 对柑橘枯萎病进行研究, 通过差异基因分析发现在柑橘在枯萎病的情况下出现淀粉积累、干旱胁迫等现象, 加深了对柑橘枯萎病的病理生理学了解。

Wang 等 (2017) 对单倍体柚 ‘Pummelo’ 测序后对柑橘无融合生殖提出了新见解, 发现 *CitRWP* 在多胚品种中特异表达并且表达水平较高。Wang 等 (2018) 对莽山野柑 (‘Mangshan Mandarin’) 测序后发现 *ACO* 基因与柑橘中柠檬酸浓度调控有关, 并在栽培和野生柑橘果实中发现 928 个差异表达的基因。Zhu 等 (2019) 对 ‘香港金橘’ (‘Hongkong Kumquat’) 测序后发现该品种的 96.9% 的蛋白质编码基因与其他 8 个柑橘物种具有同源性, 表明其具有作为 “模式物种” 的研究潜力。

Jaillon 等 (2007) 对 ‘PN40024’ 葡萄测序后发现葡萄中与香味有关的酶 (Terpene synthases), 可促进与树脂、精油、香味有关的次生代谢物的合成。Roach 等 (2018) 对 ‘霞多丽’ 葡萄测序并结合 15 个主流的 ‘霞多丽’ 无性繁殖品种的重测序数据, 鉴定出了 1 620 个标记, 可通过标记对霞多丽无性繁殖品种间的遗传突变进行研究。

Wu 等 (2013) 对 ‘砀山酥梨’ 测序后, 研究影响梨肉质、香味、糖含量等的关键基因, 探究影响果实中木质素合成的途径, 并发现  $\alpha$ -亚麻酸的代谢与梨香味形成之间存在密切关系, 发现梨中 *SOT*、*SDH* 以及 *S6PDH* 高于其他非蔷薇科物种。Chagne 等 (2014) 对 ‘Bartlett’ 测序后了解到  $\alpha$  扩展蛋白 ( $\alpha$ -expansins) 可能与果实软化相关。Dong 等 (2020) 对 ‘山西杜梨’ 测序后对影响其抗性、香味、代谢等相关基因进行了研究, 发现 150 个与抗性有关的 TIR-NBS-LRR (TNL) 型基因, 花青素还原酶 (ANR) 代谢途径对梨果实涩味有影响。

Shulaev 等 (2011) 对草莓 ‘Hawaii 4’ 测序后分析影响草莓开花时间、营养价值、香味等关键基因, 发现 681 个草莓中特有的基因簇, 并证明草莓和其他植物共享一组核心的信号传导元件。Hirakawa 等 (2014) 对草莓 ‘Reikou’ 测序后研究多倍体植物的测序和组装, 并利用 632 个 SSR 标记和 964 个基因位点进行草莓系统发育分析。Edger 等 (2019) 对草莓 ‘Camarosa’ 测序后分析代谢、抗病性状有关的显性亚基因组。Zhang 等 (2020) 对黄毛草莓 ‘Ruegen’ 测序后发现草莓果实颜色与 *FnMYB10* 存在关联。

Wu 等 (2016) 对阿宽蕉测序后, 研究影响其抗病性、抗寒性的基因, 并发现 MYB 转录因子家族中存在一种进化模式, 即纯化选择会消除不良突变并稳定 MYB 转录因子的基本功能, 使其发挥核心功能。

### 3.3 果树全基因组测序存在的问题和挑战

Velasco 等 (2010) 借助 Sanger 和 Roche 454 首次完成 ‘金冠’ 苹果测序, 但是所得基因组质量不高, Contig N50 为 16 kb, 测序深度仅为 16.9 $\times$ 。Li 等 (2016) 首次在苹果中采用 Illumina 和 PacBio 技术组合进行从头组装, Contig N50 长度提升至 111.6 kb, 测序深度提升 7.8 倍, 新的基因组更加适用于研究复杂性状。为进一步提高苹果参考基因组的质量, 在 Illumina 和 PacBio 组合的基础上, Daccord 等 (2017) 首次应用光学图谱技术 (BioNano) 参与苹果基因组的组装, Scaffold 数减少至 280 个, 测序深度提高 6.4 倍, 达到 835 $\times$ 。Zhang 等 (2019) 在 Illumina 和 PacBio 测序技术和 BioNano 光学图谱的基础上, 首次将 HI-C 染色体构象捕获技术应用于苹果中, 通过 HI-C 组装的苹果基因组, Contig N50 达到 6.99 MB, 基因组完整性大大提高。利用 PacBio、Illumina、光学图谱技术和 HI-C 染色体构象捕获技术共同参与测序和组装基因组在其他果树品种中也有应用 (Linsmith et al., 2019; Dong et al., 2020)。

甜橙基因组测序过程中, 为降低基因组测序的复杂性, Xu 等 (2013) 选择从甜橙的花药培养物中获得双单倍体系 (Dihaploid), 随后使用单倍体或双单倍体植物进行基因组测序, 这一策略后来被广泛应用 (Wu et al., 2014; Wang et al., 2017, 2019; Linsmith et al., 2019)。Shimizu 等 (2017) 选择 Satsuma 的杂合二倍体进行测序, 并用 PLATANUS 进行组装, 组装的 Scaffold 的组装质量评分为 94.2%, 证明混合装配法是可靠的方法, 此后在 ‘莽山野柑’ Mangshan Mandarin (Wang et al., 2018) 上也有应用。

葡萄中, 为解决远交和杂合基因组的组装问题, Chin 等 (2016) 利用单分子实时测序 (SMRT) 的长读长以及组装工具 FALCON 应对这一挑战。选择高度杂合的 ‘赤霞珠’ 品种为对象, 所组装基因组的 Contig N50 达到 2 170 kb。

梨基因组的杂合度较高, 利用 WGS 测序后组装的基因组质量不够理想, 因此 Wu 等 (2013) 首次应用 BAC-by-BAC 策略与 Illumina 测序技术结合, 在高度重复的高杂合度基因组中进行从头组装, 确保了梨基因组更加精准的组装结果。

目前果树全基因组还存在着部分测序数据开放不够充分的问题, 影响其引用和使用效率。大量的数据产生只完成了第一步, 数据的存储和调用, 如何更好的应用为整个科学研究服务乃是更加重要的问题。现在已经有大量的专门针对某物种的数据库。如草莓数据库 (<http://strawberry-garden.kazusa.or.jp/>); 葡萄数据库 (<https://www6.inra.fr/igpp/>); 柑橘数据库 (<http://citrus.hzau.edu.cn/orange/>, <https://www.citrusgenomedb.org/>); 香蕉数据库 (<https://banana-genome-hub.southgreen.fr/>) 及整合蔷薇科 (草莓、苹果、梨、桃) 数据的数据库 (<https://www.rosaceae.org/>)。为了降低研究时在搜寻数据库和比对序列等的时间成本, 迫切需要能够整合多种园艺作物基因组数据库和使用便利的数据分析工具。因此生物信息分析和统计基因组等方法的研究日益凸显出重要性。

组装对材料纯合度的要求较高, 而大多果树都是异交植物, 纯合体果树材料不易获得, 因此纯合度往往难以达到最理想的状态。果树全基因组学测序的研究结果为转录组学、比较基因组学研究、重要性状基因挖掘利用等提供了极大地帮助, 但果树的性状大多受微效多基因控制, 目前仍然难以准确发掘。生物信息学以及测序技术在生命科学领域的重要性不言而喻, 将继续成为不可或缺的研究方法。

### References

- Ardui S, Ameer A, Vermeesch J R, Hestand M S. 2018. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Res*, 46 (5): 2159 - 2168.

- Bentley D R, Balasubramanian S, Swerdlow H P, Smith G P, Milton J, Brown C G, Hall K P, Evers D J, Barnes C L, Bignell H R, Boutell J M, Bryant J, Carter R J, Keira Cheetham R, Cox A J, Ellis D J, Flatbush M R, Gormley N A, Humphray S J, Irving L J, Karbelashvili M S, Kirk S M, Li H, Liu X, Maisinger K S, Murray L J, Obradovic B, Ost T, Parkinson M L, Pratt M R, Rasolonjatovo I M J, Reed M T, Rigatti R, Rodighiero C, Ross M T, Sabot A, Sankar S V, Scally A, Schroth G P, Smith M E, Smith V P, Spiridou A, Torrance P E, Tzonev S S, Vermaas E H, Walter K, Wu X, Zhang L, Alam M D, Anastasi C, Aniebo I C, Bailey D M D, Bancarz I R, Banerjee S, Barbour S G, Baybayan P A, Benoit V A, Benson K F, Bevis C, Black P J, Boodhun A, Brennan J S, Bridgham J A, Brown R C, Brown A A, Buermann D H, Bundu A A, Burrows J C, Carter N P, Castillo N, Chiara E, Catenazzi M, Chang S, Neil Cooley R, Crake N R, Dada O O, Diakoumakos K D, Dominguez-Fernandez B, Earnshaw D J, Egbujor U C, Elmore D W, Etchin S S, Ewan M R, Fedurco M, Fraser L J, Fuentes Fajardo K V, Scott Furey W, George D, Gietzen K J, Goddard C P, Golda G S, Granieri P A, Green D E, Gustafson D L, Hansen N F, Harnish K, Haudenschild C D, Heyer N I, Hims M M, Ho J T, Horgan A M, Hoschler K, Hurwitz S, Ivanov D V, Johnson M Q, James T, Huw Jones T A, Kang G, Kerelska T H, Kersey A D, Khrebtukova I, Kindwall A P, Kingsbury Z, Kokko-Gonzales P I, Kumar A, Laurent M A, Lawley C T, Lee S E, Lee X, Liao A K, Loch J A, Lok M, Luo S, Mammen R M, Martin J W, McCauley P G, McNitt P, Mehta P, Moon K W, Mullens J W, Newington T, Ning Z, Ling Ng B, Novo S M, O'Neill M J, Osborne M A, Osnowski A, Ostadan O, Paraschos L L, Pickering L, Pike A C, Pike A C, Chris Pinkard D, Pliskin D P, Podhasky J, Quijano V J, Racz C, Rae V H, Rawlings S R, Chiva Rodriguez A, Roe P M, Rogers J, Rogert Bacigalupo M C, Romanov N, Romieu A, Roth R K, Rourke N J, Ruediger S T, Rusman E, Sanches-Kuiper R M, Schenker M R, Seoane J M, Shaw R J, Shiver M K, Short S W, Sizto N L, Sluis J P, Smith M A, Ernest Sohna Sohna J, Spence E J, Stevens K, Sutton N, Szajkowski L, Tregidgo C L, Turcatti G, VandeVondele S, Verhovskiy Y, Virk S M, Wakelin S, Walcott G C, Wang J, Worsley G J, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin J C, Hurler M E, McCooke N J, West J S, Oaks F L, Lundberg P L, Klennerman D, Durbin R, Smith A J. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456 (7218): 53 – 59.
- Chagne D, Crowhurst R N, Pindo M, Thrimawithana A, Deng C, Ireland H, Fiers M, Dzierzon H, Cestaro A, Fontana P, Bianco L, Lu A, Storey R, Knabel M, Saeed M, Montanari S, Kim Y K, Nicolini D, Larger S, Stefani E, Allan A C, Bowen J, Harvey I, Johnston J, Malnoy M, Troggio M, Perchev L, Sawyer G, Wiedow C, Won K, Viola R, Hellens R P, Brewer L, Bus V G, Schaffer R J, Gardiner S E, Velasco R. 2014. The draft genome sequence of European pear (*Pyrus communis* L. 'Bartlett'). *PLoS ONE*, 9 (4): e92644.
- Chen W, Kalscheuer V, Tzschach A, Menzel C, Ullmann R, Schulz M H, Erdogan F, Li N, Kijas Z, Arkesteijn G, Pajares I L, Goetz-Sothmann M, Heinrich U, Rost I, Dufke A, Grasshoff U, Glaeser B, Vingron M, Ropers H H. 2008. Mapping translocation breakpoints by next-generation sequencing. *Genome Res*, 18 (7): 1143 – 1149.
- Chin C S, Peluso P, Sedlazeck F J, Nattestad M, Concepcion G T, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer G R, Delledonne M, Luo C, Ecker J R, Cantu D, Rank D R, Schatz M C. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods*, 13 (12): 1050 – 1054.
- Daccord N, Celton J M, Linsmith G, Becker C, Choise N, Schijlen E, van de Geest H, Bianco L, Micheletti D, Velasco R, Di Pierro E A, Gouzy J, Rees D, Guerif P, Muranty H, Durel C E, Laurens F, Lespinasse Y, Gaillard S, Aubourg S, Quesneville H, Weigel D, van de Weg E, Troggio M, Bucher E. 2017. High-quality *de novo* assembly of the apple genome and methylome dynamics of early fruit development. *Nat Genet*, 49 (7): 1099 – 1106.
- Davey M W, Gudimella R, Hari Krishna J A, Sin L W, Khalid N, Keulemans J. 2013. A draft *Musa balbisiana* genome sequence for molecular genetics in polyploid, inter- and intra-specific *Musa* hybrids. *BMC Genomics*, 14 (1): 683.
- D'Hont A, Denoeud F, Aury J M, Baurens F C, Carreel F, Garsmeur O, Noel B, Bocs S, Droc G, Rouard M, Da S C, Jabbari K, Cardi C, Poulain J, Souquet M, Labadie K, Jourda C, Lengelle J, Rodier-Goud M, Alberti A, Bernard M, Correa M, Ayyampalayam S, McKain M R, Leebens-Mack J, Burgess D, Freeling M, Mbeguie-A-Mbeguie D, Chabannes M, Wicker T, Panaud O, Barbosa J, Hribova E, Heslop-Harrison P, Habas R, Rivallan R, Francois P, Poirion C, Kilian A, Burthia D, Jenny C, Bakry F, Brown S, Guignon V, Kema G, Dita M, Waalwijk C, Joseph S, Dievart A, Jaillon O, Leclercq J, Argout X, Lyons E, Almeida A, Jeridi M, Dolezel J, Roux N, Risterucci A M, Weissenbach J, Ruiz M, Glaszmann J C, Quetier F, Yahiaoui N, Wincker P. 2012. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*, 488 (7410): 213 – 217.



- Dong X, Wang Z, Tian L, Zhang Y, Qi D, Huo H, Xu J, Li Z, Liao R, Shi M, Wahocho S A, Liu C, Zhang S, Tian Z, Cao Y. 2020. *De novo* assembly of a wild pear (*Pyrus betuleafolia*) genome. *Plant Biotechnol J*, 18 (2): 581 – 595.
- Edger P P, Poorten T J, VanBuren R, Hardigan M A, Colle M, McKain M R, Smith R D, Teresi S J, Nelson A, Wai C M, Alger E I, Bird K A, Yocca A E, Pumpin N, Ou S, Ben-Zvi G, Brodt A, Baruch K, Swale T, Shiue L, Acharya C B, Cole G S, Mower J P, Childs K L, Jiang N, Lyons E, Freeling M, Puzey J R, Knapp S J. 2019. Origin and evolution of the octoploid strawberry genome. *Nat Genet*, 51 (3): 541 – 547.
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. 2009. Real-time DNA sequencing from single polymerase molecules. *Science*, 323 (5910): 133 – 138.
- Hirakawa H, Shirasawa K, Kosugi S, Tashiro K, Nakayama S, Yamada M, Kohara M, Watanabe A, Kishida Y, Fujishiro T, Tsuruoka H, Minami C, Sasamoto S, Kato M, Nanri K, Komaki A, Yanagi T, Guoxin Q, Maeda F, Ishikawa M, Kuhara S, Sato S, Tabata S, Isobe S N. 2014. Dissection of the octoploid strawberry genome by deep sequencing of the genomes of *Fragaria* species. *DNA Res*, 21 (2): 169 – 181.
- Imelfort M, Edwards D. 2009. *De novo* sequencing of plant genomes using second-generation technologies. *Brief Bioinform*, 10 (6): 609 – 618.
- Jaillon O, Aury J M, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Huguency 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 F C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrini 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, Wincker P. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 449 (7161): 463 – 467.
- Ji P, Zhang Y, Wang J, Zhao F. 2017. MetaSort untangles metagenome assembly by reducing microbial community complexity. *Nat Commun*, 8: 14306.
- Laver T, Harrison J, O'Neill P A, Moore K, Farbos A, Paszkiewicz K, Studholme D J. 2015. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol Detect Quantif*, 3: 1 – 8.
- Li X, Kui L, Zhang J, Xie Y, Wang L, Yan Y, Wang N, Xu J, Li C, Wang W, van Nocker S, Dong Y, Ma F, Guan Q. 2016. Improved hybrid *de novo* genome assembly of domesticated apple (*Malus × domestica*). *Gigascience*, 5 (1): 35.
- Li Y, Wang L R. 2020. Genetic resources, reeding programs in China, and gene mining of peach: a review. *Horticultural Plant Journal*, 6 (4): 205 – 215.
- Linsmith G, Rombauts S, Montanari S, Deng C H, Celton J M, Guerif P, Liu C, Lohaus R, Zurn J D, Cestaro A, Bassil N V, Bakker L V, Schijlen E, Gardiner S E, Lespinasse Y, Durel C E, Velasco R, Neale D B, Chagne D, van de Peer Y, Troggio M, Bianco L. 2019. Pseudo-chromosome-length genome assembly of a double haploid “Bartlett” pear (*Pyrus communis* L.). *BioRxiv*, 651778.
- Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. 2012. Comparison of next-generation sequencing systems. *J Biomed Biotechnol*, 2012 (7): 251364.
- Lu H, Giordano F, Ning Z. 2016. Oxford nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics*, 14 (5): 265 – 279.
- Margulies M, Egholm M, Altman W E, Attiya S, Bader J S, Bemben L A, Berka J, Braverman M S, Chen Y J, Chen Z, Dewell S B, Du L, Fierro J M, Gomes X V, Godwin B C, He W, Helgesen S, Ho C H, Irzyk G P, Jando S C, Alenquer M L, Jarvie T P, Jirage K B, Kim J B, Knight J R, Lanza J R, Leamon J H, Lefkowitz S M, Lei M, Li J, Lohman K L, Lu H, Makhijani V B, McDade K E, McKenna M P, Myers E W, Nickerson E, Nobile J R, Plant R, Puc B P, Ronan M T, Roth G T, Sarkis G J, Simons J F, Simpson J W, Srinivasan M, Tartaro K R, Tomasz A, Vogt K A, Volkmer G A, Wang S H, Wang Y, Weiner M P, Yu P, Begley R F, Rothberg J M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437 (7057): 376 – 380.
- Niedringhaus T P, Milanova D, Kerby M B, Snyder M P, Barron A E. 2011. Landscape of next-generation sequencing technologies. *Anal Chem*,

- 83 (12): 4327 – 4341.
- Qiao Xin, Li Meng, Yin Hao, Li Lei-yan, Wu Jun, Zhang Shao-ling. 2014. Advances in whole genome sequencing of fruit trees. *Acta Horticulturae Sinica*, 41 (1): 165 – 177. (in Chinese)
- 乔 鑫, 李 梦, 殷 豪, 李雷廷, 吴 俊, 张绍铃. 2014. 果树全基因组测序研究进展. *园艺学报*, 41 (1): 165 – 177.
- Rhoads A, Au K F. 2015. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics*, 13 (5): 278 – 289.
- Roach M J, Johnson D L, Bohlmann J, van Vuuren H, Jones S, Pretorius I S, Schmidt S A, Borneman A R. 2018. Population sequencing reveals clonal diversity and ancestral inbreeding in the grapevine cultivar Chardonnay. *PLoS Genet*, 14 (11): e1007807.
- Rothberg J M, Leamon J H. 2008. The development and impact of 454 sequencing. *Nat Biotechnol*, 26 (10): 1117 – 1124.
- Schuster S C. 2008. Next-generation sequencing transforms today's biology. *Nat Methods*, 5 (1): 16 – 18.
- Shendure J, Ji H. 2008. Next-generation DNA sequencing. *Nat Biotechnol*, 26 (10): 1135 – 1145.
- Shimizu T, Tanizawa Y, Mochizuki T, Nagasaki H, Yoshioka T, Toyoda A, Fujiyama A, Kaminuma E, Nakamura Y. 2017. Draft sequencing of the heterozygous diploid genome of Satsuma (*Citrus unshiu* Marc.) using a hybrid assembly approach. *Front Genet*, 8: 180.
- Shulaev V, Sargent D J, Crowhurst R N, Mockler T C, Folkerts O, Delcher A L, Jaiswal P, Mockaitis K, Liston A, Mane S P, Burns P, Davis T M, Slovin J P, Bassil N, Hellens R P, Evans C, Harkins T, Kodira C, Desany B, Crasta O R, Jensen R V, Allan A C, Michael T P, Setubal J C, Celton J M, Rees D J, Williams K P, Holt S H, Ruiz R J, Chatterjee M, Liu B, Silva H, Meisel L, Adato A, Filichkin S A, Troglio M, Viola R, Ashman T L, Wang H, Dharmawardhana P, Elser J, Raja R, Priest H D, Bryant D J, Fox S E, Givan S A, Wilhelm L J, Naithani S, Christoffels A, Salama D Y, Carter J, Lopez G E, Zdepksi A, Wang W, Kerstetter R A, Schwab W, Korban S S, Davik J, Monfort A, Denoyes-Rothan B, Arus P, Mittler R, Flinn B, Aharoni A, Bennetzen J L, Salzberg S L, Dickerman A W, Velasco R, Borodovsky M, Veilleux R E, Foltá K M. 2011. The genome of woodland strawberry (*Fragaria vesca*). *Nat Genet*, 43 (2): 109 – 116.
- Torp Dahl M, Löfström C, Nielsen E M. 2010. Whole genome sequencing. *Methods in Molecular Biology*, 628 (6): 215.
- Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek J A, Costa G, McKernan K, Sidow A, Fire A, Johnson S M. 2008. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res*, 18 (7): 1051 – 1063.
- Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar S K, Troglio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Dal Ri A, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoy M, Micheletti D, Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, Zini E, Eldredge G, Fitzgerald L M, Gutin N, Lanchbury J, Macalma T, Mitchell J T, Reid J, Wardell B, Kodira C, Chen Z, Desany B, Niazi F, Palmer M, Koepke T, Jiwan D, Schaeffer S, Krishnan V, Wu C, Chu V T, King S T, Vick J, Tao Q, Mraz A, Stormo A, Stormo K, Bogden R, Ederle D, Stella A, Vecchiotti A, Kater M M, Masiero S, Lasserre P, Lepinasse Y, Allan A C, Bus V, Chagne D, Crowhurst R N, Gleave A P, Lavezzo E, Fawcett J A, Proost S, Rouze P, Sterck L, Toppo S, Lazzari B, Hellens R P, Durel C E, Gutin A, Bumgarner R E, Gardiner S E, Skolnick M, Egholm M, Van de Peer Y, Salamini F, Viola R. 2010. The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat Genet*, 42 (10): 833 – 839.
- Velculescu V E, Zhang L, Zhou W, Vogelstein J, Basrai M A, Bassett D J, Hieter P, Vogelstein B, Kinzler K W. 1997. Characterization of the yeast transcriptome. *Cell*, 88 (2): 243 – 251.
- Verde I, Abbott A G, Scalabrini 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, Pirone 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, 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 L, He F, Huang Y, He J, Yang S, Zeng J, Deng C, Jiang X, Fang Y, Wen S, Xu R, Yu H, Yang X, Zhong G, Chen C, Yan X, Zhou C, Zhang H, Xie Z, Larkin R M, Deng X, Xu Q. 2018. Genome of wild mandarin and domestication history of mandarin. *Mol Plant*,

- 11 (8): 1024 – 1037.
- Wang X, Xu Y, Zhang S, Cao L, Huang Y, Cheng J, Wu G, Tian S, Chen C, Liu Y, Yu H, Yang X, Lan H, Wang N, Wang L, Xu J, Jiang X, Xie Z, Tan M, Larkin R M, Chen L L, Ma B G, Ruan Y, Deng X, Xu Q. 2017. Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nat Genet*, 49 (5): 765 – 772.
- Wang Z, Miao H, Liu J, Xu B, Yao X, Xu C, Zhao S, Fang X, Jia C, Wang J, Zhang J, Li J, Xu Y, Wang J, Ma W, Wu Z, Yu L, Yang Y, Liu C, Guo Y, Sun S, Baurens F C, Martin G, Salmon F, Garsmeur O, Yahiaoui N, Hervouet C, Rouard M, Laboureau N, Habas R, Ricci S, Peng M, Guo A, Xie J, Li Y, Ding Z, Yan Y, Tie W, D'Hont A, Hu W, Jin Z. 2019. *Musa balbisiana* genome reveals subgenome evolution and functional divergence. *Nat Plants*, 5 (8): 810 – 821.
- Wu G A, Prochnik S, Jenkins J, Salse J, Hellsten U, Murat F, Perrier X, Ruiz M, Scalabrin S, Terol J, Takita M A, Labadie K, Poulain J, Couloux A, Jabbari K, Cattonaro F, Del F C, Pinosio S, Zuccolo A, Chapman J, Grimwood J, Tadeo F R, Estornell L H, Munoz-Sanz J V, Ibanez V, Herrero-Ortega A, Aleza P, Perez-Perez J, Ramon D, Brunel D, Luro F, Chen C, Farmerie W G, Desany B, Kodira C, Mohiuddin M, Harkins T, Fredrikson K, Burns P, Lomsadze A, Borodovsky M, Reforgiato G, Freitas-Astua J, Quetier F, Navarro L, Roose M, Wincker P, Schmutz J, Morgante M, Machado M A, Talon M, Jaillon O, Ollitrault P, Gmitter F, Rokhsar D. 2014. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat Biotechnol*, 32 (7): 656 – 662.
- Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, Khan M A, Tao S, Korban S S, Wang H, Chen N J, Nishio T, Xu X, Cong L, Qi K, Huang X, Wang Y, Zhao X, Wu J, Deng C, Gou C, Zhou W, Yin H, Qin G, Sha Y, Tao Y, Chen H, Yang Y, Song Y, Zhan D, Wang J, Li L, Dai M, Gu C, Wang Y, Shi D, Wang X, Zhang H, Zeng L, Zheng D, Wang C, Chen M, Wang G, Xie L, Sovero V, Sha S, Huang W, Zhang S, Zhang M, Sun J, Xu L, Li Y, Liu X, Li Q, Shen J, Wang J, Paull R E, Bennetzen J L, Wang J, Zhang S. 2013. The genome of the pear (*Pyrus bretschneideri* Rehd.) . *Genome Res*, 23 (2): 396 – 408.
- Wu W, Yang Y L, He W M, Rouard M, Li W M, Xu M, Roux N, Ge X J. 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.
- Xu Q, Chen L L, Ruan X, Chen D, Zhu A, Chen C, Bertrand D, Jiao W B, Hao B H, Lyon M P, Chen J, Gao S, Xing F, Lan H, Chang J W, Ge X, Lei Y, Hu Q, Miao Y, Wang L, Xiao S, Biswas M K, Zeng W, Guo F, Cao H, Yang X, Xu X W, Cheng Y J, Xu J, Liu J H, Luo O J, Tang Z, Guo W W, Kuang H, Zhang H Y, Roose M L, Nagarajan N, Deng X X, Ruan Y. 2013. The draft genome of sweet orange (*Citrus sinensis*) . *Nat Genet*, 45 (1): 59 – 66.
- Zhang C M, Hao Y J. 2020. Advances in genomic, transcriptomic, and metabolomic analyses of fruit quality in fruit crops. *Horticultural Plant Journal*, 6 (4): 205 – 215.
- Zhang J, Lei Y, Wang B, Li S, Yu S, Wang Y, Li H, Liu Y, Ma Y, Dai H, Wang J, Zhang Z. 2020. The high-quality genome of diploid strawberry (*Fragaria nilgerrensis*) provides new insights into anthocyanin accumulation. *Plant Biotechnol J*, 18 (9): 1908 – 1924.
- Zhang L, Hu J, Han X, Li J, Gao Y, Richards C M, Zhang C, Tian Y, Liu G, Gul H, Wang D, Tian Y, Yang C, Meng M, Yuan G, Kang G, Wu Y, Wang K, Zhang H, Wang D, Cong P. 2019. A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat Commun*, 10 (1): 1494.
- Zhang Y, Barthe G, Grosser J W, Wang N. 2016. Transcriptome analysis of root response to citrus blight based on the newly assembled Swingle citrumelo draft genome. *BMC Genomics*, 17: 485.
- Zhou Y, Minio A, Massonnet M, Solares E A, Gaut B S. 2019. Structural variants, clonal propagation, and genome evolution in grapevine (*Vitis vinifera*) . *BioRxiv*: 508119.
- Zhu C, Zheng X, Huang Y, Ye J, Chen P, Zhang C, Zhao F, Xie Z, Zhang S, Wang N, Li H, Wang L, Tang X, Chai L, Xu Q, Deng X. 2019. Genome sequencing and CRISPR/Cas9 gene editing of an early flowering Mini-Citrus (*Fortunella hindsii*) . *Plant Biotechnol J*, 17 (11): 2199 – 2210.