Recent advances in fruit crop genomics

Qiang XU, Chaoyang LIU, Manosh Kumar BISWAS, Zhiyong PAN, Xiuxin DENG (✉)

Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan 430070, China

Abstract In recent years, dramatic progress has been made in the genomics of fruit crops. The publication of a dozen fruit crop genomes represents a milestone for both functional genomics and breeding programs in fruit crops. Rapid advances in high-throughput sequencing technology have revolutionized the manner and scale of genomics in fruit crops. Research on fruit crops is encompassing a wide range of biological questions which are unique and cannot be addressed in a model plant such as Arabidopsis. This review summarizes recent achievements of research on the genome, transcriptome, proteome, miRNAs and epigenome of fruit crops.

Keywords fruit crops, fruit biology, gene function, genomics

1 Introduction

Fruit crops include a wide range of plant species that can be classified into several groups, including evergreen (e.g. orange and papaya), deciduous (e.g. apple and peach), herbaceous (e.g., banana and strawberry), and vine (e.g., grapevine and kiwifruit) species. Fruit crops are of great economic importance; the gross world production reached 263 billion tonnes, accounting for about 20% of all agricultural products in 2011 (FAO statistics). Fruit is also an important part of the human diet and contains abundant components beneficial to human health.

Numerous efforts have been made to improve the agronomic and economic traits with the aim of increasing production and enhancing the quality, resistances to the stresses, postharvest and nutritional properties. Manipulation at the gene level is a potential tool to achieve these goals, and increasing efforts have been directed to this area, including the genetic improvement of target traits via gene targeting. To this end, identification of key genes and elucidation of gene function is essential. Classical approaches depend mainly on forward genetics, and identify the key genes (loci) based on genetic maps produced from large genetic populations. Traditional genetic experiments are rather complicated in most fruit crops due to their biologic characteristics, such as long life cycle, large tree canopy or asexual reproduction (apomixis). As a result, such genetic studies are long and expensive, often requiring large areas of land. Over the past decade, the field of fruit crop genomics has advanced rapidly and provided an alternative way to identify potential key genes, particularly those controlling the interesting and specific traits in fruits that cannot be studied in model plants such as Arabidopsis and rice. These aims have been rapidly facilitated by recent emergence of next generation (Next-Gen) sequencing technology (Illumina, SOLiD and 454 platforms) [1]. This sequencing-based approach has attracted fruit crop researchers since it is convenient, efficient, and fast, and requires relatively little space compared with classic genetic studies. We believe that the sequencing-based approach will become more popular for basic research on fruit crops, particularly as a direct step to the gene level, which is a necessary step for understanding the molecular basis of traits of interest. In the near future, multiple omics-based research strategies will form the basis for the next wave of advancements in fruit crop research, particularly as the number of sequenced genomes is increasing rapidly. This review summarizes recent studies of the genomes, transcriptomes, proteomes, miRNAs and epigenomes of fruit crops (Fig. 1).

2 Genomic research on fruit crops

2.1 The genome

To date, the genomes of 12 fruit crops have been sequenced (Table 1). The first fruit crop genome to be published was that of grapevine in 2007. This was also the fourth flowering crop to be sequenced [11]. Sequencing of a transgenic genotype of papaya followed in 2008 [2]. These genomes were sequenced using Sanger sequencing technology, an accurate but expensive and time-consuming process. The emergence of Next-Gen sequencing plat-
forms (e.g., Illumina; SoLID, Life Technologies; 454, Roche), and the recent reduced cost and improved efficiency of such technology, has dramatically accelerated the sequencing of fruit crops. The apple genome was assembled using a combination of Sanger and 454 sequencing reads [5] and such combinatorial strategies were also implemented in the de novo genome assemblies of strawberry [4], banana [6] and pear [9]. There are also several instances of de novo genome assembly that employed Next-Gen data alone, such as sequencing of sweet orange [3] and kiwifruit [10].

Evolution is one of the interesting areas that can be investigated by de novo genome analysis. Following interpretation of the genome-scale data within the grapevine genome, and comparison of this data with those of Arabidopsis, rice and poplar, Jaillon et al. [11] postulated the existence of an ancient whole genome triplication event in all eudicot genomes. This model is widely accepted; thus, the grapevine genome provides a reference for whole genome duplication (WGD) analysis. Furthermore, a recent WGD event was identified and the followed chromosome fusion events contributed to the speciation of apple [5]. Whole genome data are also powerful for analysis of speciation or species origin. Sequencing of the sweet orange genome and re-sequencing of progenitors, including pummelo and mandarin, revealed the origin of cultivated orange [3].

Biological questions can also be addressed from the genome data. From the genomes of papaya and date palm it was possible to identify sex-linked chromosome or genomic regions. These are important, not only for the understanding of sex differentiation in plants, but are also beneficial for breeding, because female fruit-producing trees are generally of greater agricultural significance than male trees [2,7]. Research on sweet orange demonstrated that the high content of vitamin C in fruit is due to the high expression of genes involved in the galacturonate pathway, particularly the rate-limiting enzyme of the galacturonate pathway, GalUR. The recent expansion of GalUR gene family may provide a genomic basis for regulating vitamin C production [3].

2.2 Molecular markers

Simple sequence repeat (SSR) markers have been increasingly used for the analysis of fruit crops and have contributed significantly to fruit tree breeding programs. The most common uses of SSR markers in fruit tree research include characterization of germplasm evaluation, genetic diversity, cultivar identification and linkage map construction [12,13]. Due to the requirement for prior sequence information, use of SSRs has benefited from the availability of rapidly accumulating genomic resources including expressed sequence tags (EST), whole genome sequences and BAC end sequences. These resources allow data mining by using pipelines composed of SSR

Table 1 Published de novo genomes of fruit crops

<table>
<thead>
<tr>
<th>Species</th>
<th>Common name</th>
<th>Genome size/Mb</th>
<th>Gene number</th>
<th>TE(^1) percentage</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carica papaya</td>
<td>Papaya</td>
<td>372</td>
<td>24746</td>
<td>52.0</td>
<td>[2]</td>
</tr>
<tr>
<td>Citrus sinensis</td>
<td>Sweet orange</td>
<td>367</td>
<td>29655</td>
<td>20.5</td>
<td>[3]</td>
</tr>
<tr>
<td>Fragaria vesca</td>
<td>Strawberry</td>
<td>240</td>
<td>34809</td>
<td>22.0</td>
<td>[4]</td>
</tr>
<tr>
<td>Malus domestica</td>
<td>Apple</td>
<td>742</td>
<td>57386</td>
<td>42.4</td>
<td>[5]</td>
</tr>
<tr>
<td>Musa acuminate</td>
<td>Banana</td>
<td>523</td>
<td>36542</td>
<td>43.7</td>
<td>[6]</td>
</tr>
<tr>
<td>Phoenix dactylifera</td>
<td>Date palm</td>
<td>658</td>
<td>25059</td>
<td>–</td>
<td>[7]</td>
</tr>
<tr>
<td>Prunus persica</td>
<td>Peach</td>
<td>265</td>
<td>27852</td>
<td>29.6</td>
<td>[8]</td>
</tr>
<tr>
<td>Pyrus bretschneideri</td>
<td>Pear</td>
<td>527</td>
<td>42812</td>
<td>53.1</td>
<td>[9]</td>
</tr>
<tr>
<td>Actinidia chinensis</td>
<td>Kiwifruit</td>
<td>758</td>
<td>39040</td>
<td>36.0</td>
<td>[10]</td>
</tr>
<tr>
<td>Vitis vinifera</td>
<td>Grapevine</td>
<td>490</td>
<td>33514</td>
<td>41.4</td>
<td>[11]</td>
</tr>
</tbody>
</table>

Note: \(^1\)TE, transposable element.
SNPs (single nucleotide polymorphisms) have made the most impact as molecular markers in recent years and have revolutionized the scale, number and efficiency of molecular markers in many ways. SNPs are highly abundant in the genome and are amenable to marker development and high-throughput genotyping. The reduced cost of Next-Gen sequencing provides an opportunity for the development of genome wide SNP markers, which have wide applications in genomic research. SNP markers were generated for phylogenetic and population structure determination of citrus germplasm [13] and in grape domestication [18]. Three peach genomes were sequenced and mined for high-density SNP markers to be used in genotyping, map construction and genomics were sequenced and mined for high-density SNP markers [13] and in grape domestication [18]. Three peach genomes were sequenced and mined for high-density SNP markers to be used in genotyping, map construction and genomics were sequenced and mined for high-density SNP markers [13] and in grape domestication [18]. Three peach genomes were sequenced and mined for high-density SNP markers to be used in genotyping, map construction and genomics were sequenced and mined for high-density SNP markers [13] and in grape domestication [18]. Three peach genomes were sequenced and mined for high-density SNP markers to be used in genotyping, map construction and genomics were sequenced and mined for high-density SNP markers [13] and in grape domestication [18].

2.3 The transcriptome

Following the determination of the genome sequence, the next stage in the investigation of gene function is transcriptome analysis. The hybridization- or sequence-based approaches constituted the most two popular platforms to deduce and quantify gene expression. In contrast to hybridization-based methods, sequence-based approaches directly determine the cDNA sequence and its abundance. The Next-Gen sequencing methods have provided a new approach (RNA-seq or RNA-Seq) for both mapping and quantifying transcriptomes, which has clear advantages over existing approaches. Fruit crops have complex genetic backgrounds and are highly heterozygous by nature. By reducing this complexity, in combination with exquisite experimental design, transcriptome analysis can provide information about traits of interest.

RNA-Seq is rapidly becoming the primary approach for transcriptome research since it provides a more thorough and scientific representation of the absolute transcript population, and is more sensitive to genes expressed at lower levels. Moreover, prior gene sequence information is not required. Exploration of RNA-Seq in fruit crops demonstrated its advantages over other techniques [22,23]. RNA-seq has been widely applied for the dissection of various traits, such as disease resistance [24], fruit development [25] and the columnar feature of certain trees [26].

2.4 The proteome

Proteomics yields information about the expression and function of proteins that is important for understanding the molecular basis of metabolic and biochemical behavior in plants.

Proteomic analysis has been utilized to study the molecular mechanisms underlying fruit development and ripening. Citrus provides a representative non-climacteric fruit crop. Comprehensive proteomic analysis with LC-MS/MS identified 1400 proteins from different fractions of fruit juice cells [27], and a similar technique was used to explore the possible molecular mechanism underlying carotenoid accumulation [28] and anthocyanin accumulation [29] in citrus fruits at different ripening stages.

Information gained from the proteome is most powerful when it is combined with that from other omics tools. For example, in grape, the abundance of proteins involved in sugar (such as fructose) and organic acid (such as malate) was found to correlate with metabolite levels [30]. In citrus fruit, storage at low temperature was found to result in the upregulation of stress-responsive genes, arrested signal transduction, and inhibited primary metabolism, secondary metabolism, and the transportation of metabolites [31].

Proteomics has also been used to dissect the proteins involved in biotic and abiotic stress responses. In citrus, a plant natriuretic peptide-like molecule of the pathogen Xanthomonas axonopodis pv. citri induced rapid changes in host photosynthesis at the protein expression [32]. In papaya, two-dimensional gel electrophoresis, difference gel electrophoresis and label-free quantitative proteomics were used to investigate the influence of the sticky disease pathogen (Papaya meleira virus) on leaves and papaya latex [33]. Other abiotic stresses have also been widely investigated by proteomic approaches, such as salinity [34].

2.5 miRNA

microRNAs (miRNAs), which are RNA molecules composed of 20–24 nucleotides (nt), regulate gene expression by cleaving target mRNAs or by inhibiting their translation. Various miRNAs act as negative regulators in plant development and stress responses, including miR156, miR159 and miR172, which are involved in flower development, particularly with respect to flowering time regulation and the phase change from vegetative growth to reproductive growth [35].

miRNAs were discovered using cloning and sequencing methods, and computational prediction of miRNA precursors, pri-miRNAs, from ESTs or whole genome sequences were performed in a wide range of fruit crops.
recent noteworthy study showed that the methylation in the involved in controlling resveratrol biosynthesis [50]. In Vitis amurensis [49].

In MYB10 example, the promoter of the anthocyanin regulator gene genetic regulation of fruit plant secondary metabolism. For that DNA methylation also plays an important role in the adult phase [48]. In addition, several studies have shown results suggest that demethylation may occur during the and dormancy shoots [47], and in crab apple seeding trees juvenile tree shoots is much lower than that of mature-tree Castanea sativa aging has also been reported for several species; in [46]. A change in the DNA methylation pattern during tree micropropagation or long-term cryopreservation [46]. A change in the DNA methylation pattern during tree aging has also been reported for several species; in Castanea sativa, the global DNA methylation level of juvenile tree shoots is much lower than that of mature-tree and dormancy shoots [47], and in crab apple seeding trees results suggest that demethylation may occur during the adult phase [48]. In addition, several studies have shown that DNA methylation also plays an important role in the genetic regulation of fruit plant secondary metabolism. For example, the promoter of the anthocyanin regulator gene MYB10 is highly methylated in green strips of apple peel [49]. In Vitis amurensis, DNA methylation may be involved in controlling resveratrol biosynthesis [50]. A recent noteworthy study showed that the methylation in the promoter region of the MYB transcription factor in pear is associated with green-skinned mutation and involved in the regulation of anthocyanin production in the outer pericarp [51].

3 Future perspectives

Dramatic advancements in genomics are likely to drive a wave of fruit crop research in the near future. A further increases in the power and capacity of Next-Gen sequencing technology is likely to markedly reduce the cost of sequencing from the current 300 RMB per gigabase. We anticipate that on one hand, many de novo genomes will be revealed, such as chestnut, Chinese date, litchi and walnut among others. Heterozygosity should be carefully considered as one of the challenging factors in such project. Heterozygosity can be significantly reduced by prior work including successive inbreeding or double-haploid induction. The latter provide homozygous genotypes, which greatly reduce genome complexity and facilitate genome assembly. On the other hand, reference-genome based studies will stimulate burgeoning research, e.g., epigenome, and ultimately, genome re-sequencing analysis will become a common approach to dissect agronomic traits in fruit crops (summarized in Fig. 1). Large segregant and natural populations will be needed for fast and efficient genomic sequencing to identify the genes controlling agronomic/biological traits and to elucidate the underlying genetic architecture and mechanisms.

Omnics-based studies will provide more specific results and may even yield information at the single cell level. Most current genetic studies consider one individual with homogenous DNA; however, recent studies have revealed heterogeneity in the composition of variants even within individual. This observation is particularly true for human cancer genomics [52]. Consequently, to reduce the heterogeneity and increase the chance of identification of causal genetic events, the sampling strategy should be as specific as possible to facilitate the pinpointing of the trait of interest. For long-lived perennial fruit trees, a relatively high level of somatic mutations accumulate during cultivation; with bud or limb mutations observed frequently in the vegetatively propagated trees. Therefore, researchers should be careful to avoid casual sampling even within a single tree. Samples for genomic studies should be restricted to the smallest possible area.

Biological replicates are required for transcriptome, proteome and small RNA studies; however, most published results only included two biological replicates. We suggest that future experiments include at least three biological replicates so that the statistical analysis of genomic data can be performed with greater power and confidence. As the cost of Next-Gen sequencing has dropped considerably, more replicates can be analyzed, which is increasingly required for the publication of
genomic studies. A recent study of confidence-based somatic mutation evaluation demonstrated that inclusion of genome sequencing replicates will increase the confidence statistical comparisons and the identification of somatic mutations [53].

We anticipate that many important agronomic traits will be investigated by genomics and functional studies. Biological questions that are unique to fruit crops and cannot be addressed in model plants such as Arabidopsis, will be particularly worthy of investigation. The trait of somatic mutations [53].

Acknowledgements The National Natural Science Foundation of China (Grant No. 31330066, 31221062 and 31222047) provided financial support for this study. Jidi Xu, Yuanlong Liu and Keqin Yu are thanked for helpful discussion on the epigenetics, miRNA and transcriptome sections, respectively.

Compliance with ethics guidelines Qiang Xu, Chaoyang Liu, Manosh Kumar Biswas, Zhiyong Pan and Xiuxin Deng declare that they have no conflict of interest or financial conflicts to disclose.

This article is a review and does not contain any studies with human or animal subjects performed by the any of the authors.

References


