Using genomics to improve fruit quality

Claudio Meneses¹ and Ariel Orellana¹,²,*

¹ Centro de Biotecnología Vegetal, Facultad de Ciencias Biológicas, Universidad Andrés Bello.
² FONDAP Center for Genome Regulation.

ABSTRACT

New fruit varieties are needed to satisfy consumers, and the industry is facing new challenges in order to respond to these demands. The emergence of genomic tools is releasing information on polymorphisms that can be utilized to expedite breeding processes in species that are difficult to breed, given the long periods of time required to get new varieties. The present review describes the current stages of the ongoing efforts that are being taken to apply these technologies to obtain varieties with improved fruit quality in species of the family Rosaceae.

Key words: Polymorphisms, Marker Assisted Selection, Genomic Tools.

The market is demanding fruit with better quality. In addition to aspects important for consumers such as flavor, aroma and consistency, among others, the globalization of world has expanded the places where they are produced and sometimes fruit needs to be shipped long distances to reach its final markets. This poses a challenge, since fruit quality also has to respond to extended post-harvest storage without affecting fruit quality. Finally, the effects of climate change are perturbing the climatic conditions where fruits have been historically produced, therefore, the new varieties also need to be better adapted to grow in harsher conditions. Breeding is the main process utilized to obtain new varieties. However, new approaches are required to breed fruit varieties that respond to the emerging challenges. The identification of molecular determinants, along with the generation of selection markers for traits of interest, offer an opportunity to respond to these challenges. In recent years the genomes of some fruit species of the Rosaceae have been sequenced (Velasco et al., 2010; Shulaev et al., 2011; Verde et al., 2013; Zhang et al., 2012a; Wu, 2013), leading to a great expansion in the identification of selection markers that can be applied in breeding programs to expedite the development of varieties that respond to the emerging challenges. In the present review we focus on the new trends that are emerging in relationship to the use of genomic tools and the identification of selection markers in species of the Rosaceae.

FRUIT BREEDING AND MARKER-ASSISTED SELECTION IN ROSACEAE

The Rosaceae contains over 100 genera and 3,000 species. It is the third most economically important plant family in temperate regions (Dirlewanger et al., 2002), including species of economic importance such as apple (Malus x domestica), peach (Prunus persica), strawberry (Fragaria ananassa), plum (Prunus salicina), almond (Prunus dulcis), pear (Pyrus communis), European plum (Prunus domestica) and sweet cherry (Prunus avium), among others. The total world production of edible rosaceous fruits reached about 145 million tons, with 998,000 hectares harvested in 2011 (FAO, 2013). For this reason, initiatives to generate new fruit varieties that can be adapted to the local agronomic conditions and satisfy the requirements of the consumer have been carried out in the last 50 years. In recent years breeding programs have been dealing with increasing challenges, including the development of improved varieties for traits that have complex genetic control (quantitative traits). These programs must deliver results in the shortest possible time and they also need to be adapted to local agricultural climatic conditions. For a long time the breeding programs focused their efforts on developing varieties with superior performance in terms of production (yield, average fruit size) and commercial features (full surface color), in order to satisfy the grower’s requirements, without considering the needs of final consumers (fruit quality). At present, improving fruit quality is one of the main goals for growers and breeders in species of the Rosaceae, since the ultimate goal of the industry is to satisfy consumers. Phenotypic expression of most fruit quality traits is quantitative. They are based on complex biochemical processes that are often determined by the interaction of various genes involved in different metabolic pathways, such as soluble solids content, titratable acidity, flesh firmness and juiciness (Illa et al., 2010). The strategy used by most breeding programs is based on performing hundreds of controlled crosses to select from thousands of individuals a few exhibiting the target traits. This approach has been successful in producing most of the varieties that today are available in the market. However, it has a number of disadvantages, since this strategy is particularly time consuming and costly in fruit trees because of the length of time required to obtain fruit production (2-3 years for peach and 3-5 years for apple) and the resources to maintain the seedlings in the field during the evaluation and selection processes. In fact, at least 12-20 years are needed to obtain one new Rosaceae fruit tree variety. For this reason, the development of genomic tools to support the early selection of genotypes may be the key to improve the efficiency of breeding programs.

USING GENOMIC TOOLS TO OBTAIN MARKERS FOR SELECTION

The advances in DNA sequencing technologies speeded up the unraveling of genomes from different species; to date
several genomes of the Rosaceae have been sequenced Table 1 including Malus x domestica (apple), Fragaria vesca (strawberry), Prunus persica (peach), Prunus mume and Pyrus bretschneideri (Velasco et al., 2010; Shulaev et al., 2011; Verde et al., 2013; Zhang et al., 2012a; Wu, 2013). In addition, the sequences of Prunus amygdalus (almond) and Prunus avium (sweet cherry) have been recently made available to the research community (http://genomicsdata.wsu.edu/public_access/index.php). The availability of these genomes made easier to re-sequence the genome of other individuals from the same species and identify polymorphisms or structural variants present in their genomes, such as single nucleotide polymorphisms (SNPs), insertions/deletions (INDELs), copy number variants and translocations. Since SNPs are by far the most abundant polymorphisms present in all genomes, most of the efforts in searching for markers that may be useful in assisted breeding are targeted to this class of polymorphisms. SNPs are widely distributed in the genome; they can be located in inter- or intragenic regions. The first attempts to identify SNPs using sequencing technologies approaches began by investigating their presence in expressed genes using expressed sequence tags (ESTs). An interesting feature of SNPs located on ESTs is the possibility to establish a link between functional and structural genomics. However, there are some limitations when low coverage sequencing technologies are utilized, since SNPs discovery is dependent on the expression level, and it is difficult to obtain SNPs for low expression genes. Furthermore, some problems may also arise in discriminating paralogous genes. Despite these constraints, this approach proved to be successful for identifying SNPs in apple (Chagne et al., 2008), almond (Wu et al., 2008) and strawberry (Bombarely et al., 2010). The emergence of the so-called next generation sequencing technologies (NGS) not only improved genomic sequencing but also made it possible to perform transcriptome analyses at higher coverage (RNA-seq analyses). This methodological upgrade expanded and improved the detection of polymorphisms in expressed genes. Thus, more than 2,000 SNPs were obtained by comparing the transcriptome of Bing and Rainier, two cherry cultivars (Koecke et al., 2012), and more than 9,000 SNPs were detected in the peach transcriptome (Wang et al., 2013).

Wide genome sequencing of different individuals expanded the identification of polymorphisms. This analysis has the advantage that in theory we can obtain a broad heterogeneity within a given species; however, this is a challenging process since many individuals exhibiting the whole diversity of that species need to be represented. Moreover, using current technologies it is not possible to re-sequence the whole genome in an affordable manner. However, for breeding purposes, the use of NGS technologies to obtain genomic information from those individuals that could be used as parents is enough to provide information on polymorphisms that can be useful markers. For example, in apple the genomic sequencing of 27 cultivars led to the identification of more than 2 million SNPs (Chagne et al., 2012a); in pears, the sequencing of three European cultivars identified more than 1 thousand SNP's (Montanari et al., 2013); in peach the sequencing of two varieties obtained more than 6,000 SNPs (Ahmad et al., 2011) while a more recent effort, sequencing 56 breeding accessions, identified more than 1 million SNPs (Verde et al., 2012).

All this effort searching for SNPs has been transformed in a collection of polymorphic SNPs from different species, deposited in chip arrays that have been useful to construct high density genetic maps of populations. More important, these chip arrays are a relevant step towards the identification of genome regions that are key in defining complex agricultural traits (Verde et al., 2012; Chagne et al., 2012a; Montanari et al., 2013).

HOW TO TRANSFORM A POLYMORPHISM INTO A SELECTION MARKER?

Once the molecular markers are available, it is necessary to establish a link between the genetic markers and the fruit quality traits. To do this, there are several methodological alternatives which have been used with relative success and there are two widely accepted: quantitative trait loci (QTL) analysis and association mapping. Although these strategies are pre-NGS era, at present thousands of polymorphisms are available in several Rosaceae fruit tree species and these can be expanded using NGS platforms. All this information along with the possibility of genotyping hundreds of individuals, including segregants from bi-parental mapping populations or varieties, improves the probability to find markers linked to traits; due to a higher resolution of the genetic maps, the genome coverage is better and the genetic variability is abundant due to the number of individuals considered in the analysis. Thus genetic linkage maps, QTL analysis and linkage disequilibrium maps appear to provide tools that will allow the dissection of complex genetic traits and marker-assisted selection of economic importance traits.

Genetic Linkage Maps and QTL Analysis: A linkage map may be thought as a ‘road map’ of the chromosomes derived from two different parents (Paterson, 1996) that indicates the position and relative genetic distances considering the recombination rate between markers in chromosomes. The most important use for linkage maps is to identify chromosome regions containing genes and QTLs associated with traits of interest. Genetic map construction has three steps: i) generation of the mapping population, II) polymorphism identification and iii) linkage analysis of markers (Collard et al., 2005). QTLs are chromosome regions containing one or more genetic factors that are responsible for a fraction of the phenotypic variation of a quantitative trait (Tanksley, 1993). QTL analysis is based on the principle of detecting an association between phenotypes and the genotypes of markers distributed in the genome. Markers are used to divide the mapping population into different genotypic groups based on the presence or absence of a particular marker locus and to determine whether significant differences exist between groups in relationship to the trait being measured (Tanksley, 1993; Young, 1996).

Until recently, most of the markers used in genetic analyses were simple sequence repeats (SSRs); however, their availability is not high, therefore the resolution of the genetic maps was low. In spite of this, QTLs for flowering and ripening time, fruit quality, tree architecture and resistance to pathogens have been reported (Dirlewanger et al., 1996; Abbott et al., 1998; Dirlewanger et al., 1999; Etienne et al., 2002; Quilot et al. 2004; Dirlewanger et al., 2006; Lambert et al., 2007; Ogundiwon et al., 2008; Dirlewanger et al., 2009). In addition, QTLs for fruit weight, soluble solid content and acidity have...
been reported (Dirlewanger et al., 1999; Etienne et al., 2002; Verde et al., 2002; Dirlewanger et al., 2009), as well as chilling injury (Cantin et al., 2010). Furthermore, QTL analyses were carried out for fruit size in apple (Stoeckli et al., 2008; Kenis et al., 2008; Davoghalaei, 2012; Potts et al., 2013), acidity (Xu et al., 2012; Zhang et al., 2012b; Kenis et al., 2008; Potts et al., 2013), soluble solid content (Kenis et al., 2008; Potts et al., 2013), harvest date (Kenis et al. 2008) flowering date (Celton et al., 2012), fruit maturity (Morimoto et al. 2013), fruit number (Stoeckli et al., 2009), volatile organic compounds (Zini et al., 2005; Dunemann et al., 2009; Dunemann et al., 2012; Khan et al., 2012; Chagne et al., 2012b; Vogt et al., 2013; Costa et al., 2013), texture (King et al., 2000; King et al., 2001; Kenis et al., 2008; Costa et al., 2008; Costa et al., 2010; Longhi et al., 2012; Longhi et al., 2013), ethylene production (Costa et al., 2014), Vitamin C content (Davey et al., 2006; Mellidou et al., 2012) and flesh browning (Kenis et al., 2008; Di Guardo et al., 2013). Also, as the cost of map construction was higher, the efforts were concentrated in a few mapping populations and often they were considered as reference progenies, but these populations only segregated for a few traits. For example, the genetic reference map of Prunus (Dirlewanger et al., 2004) was built using a mapping population obtained from the interspecific cross between almond and peach. At first this population was very appropriate, because it showed a high level of polymorphism and most markers probed were polymorphic. However, despite the marker saturation, this population does not have segregating phenotypes for fruit quality.

At present, SNP discovery using NGS platforms is faster and cheaper per data point. In addition, high performance SNP genotyping systems are available. This allows the generation of saturated genetic maps and QTL detection using these technologies. In peach, QTL analysis for fruit quality traits has been undertaken by Eduardo et al. (2011), in which maps based on SSRs were initially constructed on the mapping population Bolero x Oro A. Subsequently, they saturated the Bolero and Oro A maps using SNPs contained in the Illumina 9,000 SNP array v1 for peach (Verde et al., 2013). Thus a total of 1,453 and 229 SNPs in Bolero and Oro A were mapped, respectively (Eduardo et al., 2012). Based on this approach co-localization between candidate genes and major QTLs were detected: two putative terpene synthases and one lipoxygenase (Lox) might be involved in the biosynthesis of linalool, p-menth-1-en-9-al and nonanal, respectively. The contribution of the SNP array helped to generate a saturated map and the resolution and coverage improved considerably. Similarly, QTLs for texture were detected by Longhi et al. (2012) in apples. Two maps were saturated using SNPs that were identified during an early assembly draft (4X sequencing depth) of the Golden Delicious

<table>
<thead>
<tr>
<th>Species</th>
<th>Genome size determined (Mbp)</th>
<th>Genome size estimated (Mbp)</th>
<th>SNPs</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pyrus bretschneideri</td>
<td>512</td>
<td>527</td>
<td>3402159</td>
<td>Wu et al., 2013</td>
</tr>
<tr>
<td>Prunus persica</td>
<td>224.6</td>
<td>265</td>
<td>953357</td>
<td>IPGJ, 2013</td>
</tr>
<tr>
<td>Malus x domestica</td>
<td>598.3</td>
<td>742.3</td>
<td>2113120</td>
<td>Velasco et al., 2010; Chagne et al., 2012</td>
</tr>
<tr>
<td>Prunus mume</td>
<td>237</td>
<td>280</td>
<td>200627</td>
<td>Zhang et al., 2012; Sun et al., 2013</td>
</tr>
</tbody>
</table>
out in pear (Inoue, et al., 2007; Iwata, et al., 2013) and peach (Aranzana et al., 2010), but using hundreds of SSRs. The availability of SNP chips should improve their discoveries.

FINAL REMARKS

Genomic studies are identifying an increasing number of SNPs. This information, combined with the work on segregating populations, or by using association mapping, should lead to an increasing number of markers that can be used in the early selection of varieties in breeding programs. However, in order to obtain markers that are accurate in their prediction, a great deal of effort needs to be made in phenotyping those features that are important for the consumers as well as the industry. In addition, these approaches should also lead us to the identification of genes involved in determining the molecular components involved in defining complex traits. Thus another important product of these studies will be a better understanding of the biology behind processes that reach our senses.

ACKNOWLEDGEMENTS

This work has been supported by FONDECYT # 11121396; FONDAP-CRG 15090007; ICM P10-062-F; PB 16.

LITERATURE CITED


Meneres & Orellana Biol Res 46, 2013, 347-352


identifies unique patterns of genetic diversity domestication and
VERDE I, BASSIL N, SCALABRIN S, GILMORE B, LAWELEY C, GASIC K,
MICHIELETTI D, ROSYARA U, CATTONARO F, VENDRAMIN E, et al.
(2012) Development and evaluation of a 9K SNP array for peach by
internationally coordinated SNP detection and validation in breeding
VERDE I, QUARTA R, CEDROLA C, DETTORI MT (2002) QTL analysis of
Identification of lipoxygenase (LOX) genes putatively involved in
fruit flavour formation in apple (Malus x domestica). Tree Genetics and
WANG L, ZHAO S, GU C, ZHOU Y, ZHOU H, MA J, CHENG J, HAN Y
(2013) Deep RNA-Seq uncovers the peach transcriptome landscape.
Plant Molecular Biology 83:365-377.
WU J, WANG Z, ZHI Z, ZHANG S, MING R, ZHU S, KHAN MA, TAO S,
KORBAN SS, WANG H, et al. (2013) The genome of the pear (Pyrus
bretschneideri Rehnd). Genome research 23:396-408.
High resolution melting analysis of almond SNPs derived from ESTs.
Theoretical and Applied Genetics 118:1-14.
XU K, WANG A, BROWN S (2012) Genetic characterization of the Ma locus
YOUNG ND (1996) QTL mapping and quantitative disease resistance in
ZHANG G, SERBOLT A M, SOORIYAPATHIRANA S S, WANG D, BINK M,
OLMSTEAD JW, LEZZONI AF (2010) Fruit size QTL analysis of an F1
population derived from a cross between a domesticated sweet cherry
cultivar and a wild forest sweet cherry. Tree Genetics & Genomes 6:25-36.
ZHANG Q, CHEN W, SUN L, ZHAO F, HUANG B, WANG J, YANG W,
TAO Y, YUAN Z, FAN G, XING Z, HAN C, PAN H, ZHONG X, SHI W,
LIANG X, DU D, SUN F, XU Z, HAO R, LV T, LV Y, ZHENG Z, SUN
M, LUO L, CAI M, GAO Y, WANG J, YIN Y, XU X, CHENG T, WANG J
ZHANG Q, MA B, LI H, CHANG Y, HAN Y, LI J, WEI G, ZHAO S, KHAN
M A, ZHOU Y, GU C, ZHANG X, HAN Z, KORBAN SS, LI S, HAN Y
(2012b) Identification characterization and utilization of genome-wide
simple sequence repeats to identify a QTL for acidity in Apple. BMC
Genomics 13, 537.
ZHU C, GORE M, BUCKLER E, YU J (2008) Status and Prospects of
Association Mapping in Plants. The Plant Genome 1:5-20.
ZINI E, BIASIOLI P, GASPERI F, MOTT D, APREA E, MARK TD,
PATOCCHI A, GESSLER C, KOMJANC M (2005) QTL mapping of
volatile compounds in ripe apples detected by proton transfer reaction-