

# New approaches to *Prunus* transcriptome analysis

Pedro Martínez-Gómez · Carlos H. Crisosto ·  
Claudio Bonghi · Manuel Rubio

Received: 14 December 2010 / Accepted: 26 April 2011 / Published online: 17 May 2011  
© Springer Science+Business Media B.V. 2011

**Abstract** The recent sequencing of the complete genome of the peach offers new opportunities for further transcriptomic studies in *Prunus* species in the called post-genomics era. First works on transcriptome analysis in *Prunus* species started in the early 2000s with the development of ESTs (expressed sequence tags) and the analysis of several candidate genes. Later, new strategies of massive analysis (high throughput) of transcriptomes have been applied, producing larger amounts of data in terms of expression of a large number of genes in a single experiment. One of these systems is massive transcriptome analysis using cDNA biochips (microarrays) to analyze thousands of genes by hybridization of mRNA labelled with fluorescence. However, the recent emergence of a massive sequencing methodology (“deep-sequencing”) of the transcriptome (RNA-Seq), based on lowering the costs of DNA (in this cases complementary, cDNA) sequencing, could be more suitable than the application of microarrays. Recent papers have described the tremendous power of this technology, both in terms of profiling coverage and quantitative accuracy in transcriptomic studies. Now this technology is being applied to plant species, including *Prunus*. In this work, we analyze the potential in using this RNA-

Seq technology in the study of *Prunus* transcriptomes and the development of genomic tools. In addition, the strengths and limitations of RNA-Seq relative to microarray profiling have been discussed.

**Keywords** *Prunus* · Genome · Transcriptome · Candidate genes · Microarray · Next generation sequencing · RNA-Seq · ESTs · SNPs · Expression level · Breeding

## Introduction

The genus *Prunus*, from subfamily *Prunoideae* inside the family *Rosaceae*, includes several species producing edible drupes of significant economic importance. In 2009, worldwide annual production of *Prunus* exceeded 37 million metric tons, including 18.6 million tons of peaches and nectarines [*P. persica* (L.) Batsch]; 10.7 million tons of prunes (*P. domestica* L.), plums (*P. salicina* Lindl), sloes (*P. spinosa* L.), and cherry plums (*P. cerasifera* Ehrh.); 3.8 million tons of apricots (*P. armeniaca* L.); 2.3 million tons of almonds (*P. amygdalus* Batsch syn. *P. dulcis* (Miller) D.A. Webb); and 2.1 million tons of sweet (*P. avium* L.) and sour cherries (*P. cerasus* L.) (<http://faostat.fao.org>).

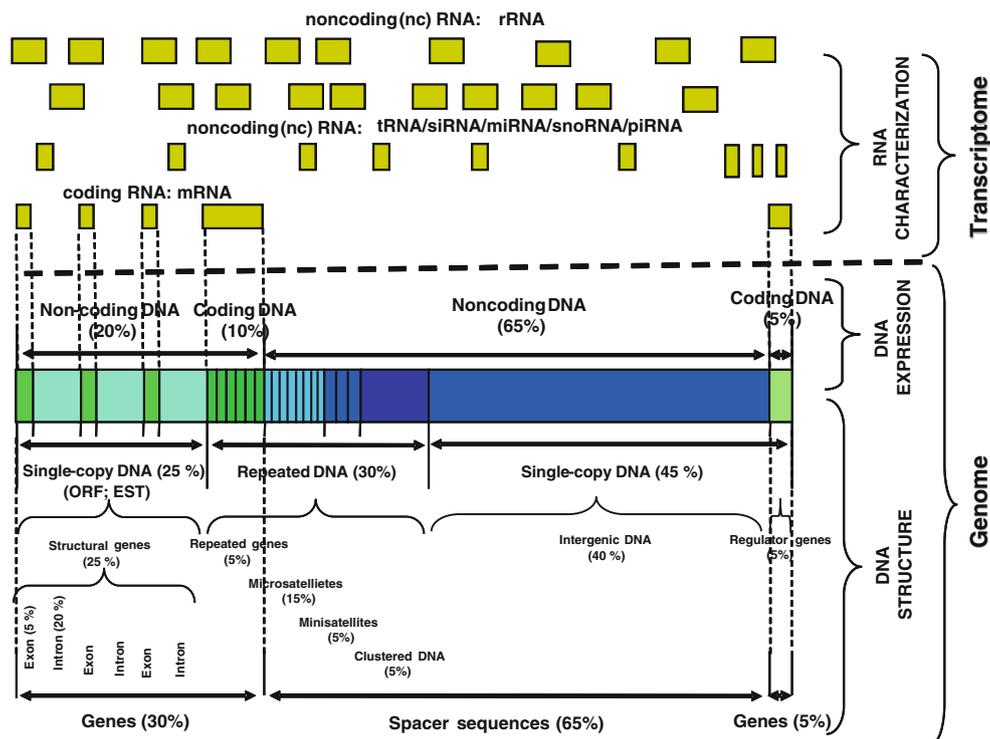
*Prunus* species are characterized by developing only one ovary in which two ovules typically form, one of which degenerates soon after anthesis. The fruit is a drupe with high commercial value in most cases, where the mature, stony endocarp together with the seed (only with a commercial value in the case of almond) forms a propagation unit comparable to a botanical seed surrounded by its protective testa. The genome size of *Prunus*, has been estimated in around 280,000,000 base pairs (bp) (Arumuganathan and Earle 1991; Baird et al. 1994), although recent discovery indicated a size of 227,000,000 bp in the case of the peach genome

---

P. Martínez-Gómez (✉) · M. Rubio  
Department of Plant Breeding, CEBAS-CSIC, PO Box 164,  
30100 Espinardo, Murcia, Spain  
e-mail: pmartinez@cebas.csic.es

C. H. Crisosto  
Department of Plant Science, University of California-Davis,  
Davis, CA 95616, USA

C. Bonghi  
Department of Environm Agron and Crop Science,  
University of Padua, 35020 Padua, Italy



**Fig. 1** Schematic representation of genome (DNA) organization and genetic expression (transcriptome, RNA) in plants

(Sosinski et al. 2010). The size of these genomes is relatively small inside the plant kingdom, where genomes range from the 115,000,000 bp of *Arabidopsis thaliana* L. Heynh (2.4 times smaller than *Prunus*) to the 5,000,000,000 bp of maize (*Zea mays* L.) and the 90,000,000,000 bp of lily (*Lilium longiflorum* Duch) (300 times greater).

Molecular studies in these *Prunus* species have been oriented mainly toward genome (DNA) studies, whereas transcriptome (RNA) studies (Fig. 1) have been more scarce (Dirlewanger et al. 2004, 2009a; Shulaev et al. 2008; Peace and Norelli 2009). These molecular studies have been performed mainly in peach (Abbott et al. 2009; Pozzi and Vecchiotti 2009), considered the model species; followed by almond (Martínez-Gómez et al. 2007; Arús et al. 2009), a very close species to peach; apricot (Folta and Gardiner 2009); cherry (Dirlewanger et al. 2009b); and plum (Esmenjaud and Dirlewanger 2007). As a result of many of these genome studies and using high-information content fingerprinting (HICF), the first physical map of *Prunus* has been generated in peach (Zhebentyayeva et al. 2008). This map has been composed of 2,138 contigs containing 15,655 BAC (Bacterial artificial chromosome) clones, and can be considered the first approach to the complete *Prunus* genome sequencing.

The recent sequencing of the complete genome of the peach offers new opportunities for further genomic and transcriptomic *Prunus* studies. The International Peach Genome Initiative (IPGI) has recently released the

complete peach genome sequence [*Prunus persica* genome (v1.0)] which is available at <http://www.rosaceae.org> (Sosinski et al. 2010). Additional advantages to use this *Prunus* reference genome include the well-established international network of cooperation among researchers and the high level of synteny between *Prunus* genomes (Dirlewanger et al. 2004; Jung et al. 2009). This synteny has also been described inside the *Rosaceae* family (Dirlewanger et al. 2004; Arús et al. 2006; Shulaev et al. 2008; Cabrera et al. 2009; Illa et al. 2011). In this new genomic context, considered the post-genomics era (Leader 2005), transcriptome analysis could be a good alternative to deep in the molecular basis and the expression of the main agronomic traits in *Prunus*.

In this work, the different approaches in the analysis of the transcriptome in *Prunus* have been described and the results compared and discussed, including EST development, candidate gene analysis and microarray application. In addition, a special emphasis is placed on the study of the application of the recent technique based on the high-throughput sequencing of RNA (RNA-Seq).

#### Early transcriptome analysis in *Prunus*: EST development and candidate gene analysis

The transcriptome can be described as the complete list of all types of RNA molecules, whether coding (messenger

RNA, mRNA) or noncoding (ncRNA) including ribosomal RNA (rRNA), transfer RNA (tRNA), short interfering RNA (siRNA), micro RNA (miRNA), small nucleolar (snoRNA), or Piwi-interacting RNA (piRNA) (Fig. 1), expressed in a whole organism. Forrest and Carninci (2009) described how an important part of the genome is transcribed, although only a limited fraction is assigned to genes. Understanding the transcriptome, both as a template for protein expression and as a regulatory molecule, is essential for interpreting the functional elements of the genome and revealing the molecular constituents of cells and tissues in a biological context (Blencowe et al. 2009).

First works on transcriptome analysis in *Prunus* species have been oriented to the development and alignments of expressed sequence tags (ESTs) and to candidate gene (complementary DNA, cDNA) analysis. Development of ESTs has been initiated in peach in 2002 (Yamamoto et al. 2002) and later has been also applied to other species such as apricot (Decroocq et al. 2003; Grimplet et al. 2005) and almond (Jiang and Ma 2003) (Table 1).

To date, a collection of more than 100,000 ESTs from different *Prunus* (mainly peach, but also almond and apricot) based on cDNA libraries has been released to public databases, and more than 25,000 putative unigenes (10,000 contigs and 15,000 singlets) have been detected. This information can be compiled in the GDR (Genome Database for Rosaceae) (<http://www.rosaceae.org>) (Jung et al. 2004; Horn et al. 2005; Cabrera et al. 2009). In addition, single nucleotide polymorphisms (SNPs) detection on unigene contigs from EST assemblies has been performed, and almost 6,000 SNPs have been identified, revealing an estimated frequency of 0.07 SNPs/100 bp (Meneses et al. 2007, <http://www.rosaceae.org>).

The above-mentioned research is complementary to the other studies regarding EST development in *Prunus* performed by different research groups in Italy as part of the work of the Italian National Consortium for Peach Genomics (<http://www.itb.cnr.it/estree/>). The ESTree database also analyzes a more reduced collection of around

80,000 sequences from different peach cDNA libraries. Mapping results for an additional 200 ESTs are presented together with peach mapping data from the GDR. In the database quantitative expression analyses in collections of ESTs from different tissues, genotypes or developmental stages are presented, according to sequence distribution in Gene Ontology (GO) classes (Vendramin et al. 2007; Lazzari et al. 2008).

On the other hand, candidate gene (CG) approaches, known as allelic association studies between an expressed gene and a DNA sequence, have been also initiated in peach with the study of several genes involved in relevant metabolic pathways for fruit growth and ripening with special emphasis in the ethylene pathway (Bonghi et al. 1998; Ruperti et al. 1998, 2001). In addition, Etienne et al. (2002) studied the associations between genes involved in some of these metabolic pathways and major genes or QTLs (Quantitative trait loci) (Table 1). These authors isolated eighteen cDNAs, encoding key proteins in sugar and organic acid metabolic pathways. Twelve candidate genes have been localized in the genetic linkage map, including those associated with ripening date, fruit development period, flesh weight, pH, titratable acidity (TA), soluble solids contents (SSC), malic acid, citric acid, quinic acid, sucrose, glucose and fructose.

These CG studies have been mainly continued in peach with the assay of different ESTs developed and the search for genes responsible for fruit traits including quality (Trainotti et al. 2003; Horn et al. 2005; Ogundiwin et al. 2009; Vecchiotti et al. 2009; Vizoso et al. 2009; Bonghi et al. 2010; Falchi et al. 2010; Le Dantec et al. 2010); development, abscission and ripening physiology (Rasori et al. 2002); and storage (González-Agüero et al. 2008; Ogundiwin et al. 2008a; Basset et al. 2009; Tittarelli et al. 2009; Falara et al., 2011). Part of these results have been biologically validated by real time quantitative PCR (qRT-PCR) (Ogundiwin et al. 2008a; Tittarelli et al. 2009). In addition, these studies in peach have been oriented to the study of flower and vegetative bud development

**Table 1** Main approaches to *Prunus* transcriptome analysis assayed

	Starting year	Number of data generated	Species studied	Agronomical traits studied
Candidate gene analysis (CGs)	1998	Hundreds (from 5 to 30 genes per experiment)	Peach, apricot, almond, plum, sweet cherry	Fruit quality, fruit development, bitterness, PPV resistance, flowering
Expressed sequence tags (ESTs)	2002	Thousands (around 100,000 identified in the databases)	Peach, apricot, almond, plum	Fruit quality, plant development, flower development
Microarray analysis	2006	Thousands (from 2,000 to 8,000 probes per microarray)	Peach, apricot, almond, plum, prune	Fruit quality, flowering, flower compatibility, response to hypoxia
Deep transcriptome sequencing (RNA-Seq) <sup>a</sup>	2010	Millions (around 20 millions of reads per line of sequencer)	Peach, apricot, plum	PPV resistance, graft incompatibility, fruit quality, flowering time

<sup>a</sup> Studies carried out by different groups although at this moment their results are not published

(Bielenberg et al. 2008). A group of *DAM* (*dormancy-associated*) *SVP*-like (*Short Vegetative Phase*) *MADS*-box genes have been described to be responsible for this absence of vegetative dormancy in peach (Li et al. 2009; Jiménez et al. 2010a, b).

In apricot, candidate genes have been assayed in the study of *Plum pox virus* (PPV) resistance including resistance gene analogs (RGA) (Dondini et al., 2004; Decroocq et al. 2005; Lalli et al. 2005), nucleotide binding site-leucine-rich repeat (NBS-LRR) (Soriano et al. 2005), eukaryotic translation initiation factor (*eIF4E*), RNA helicase SDE3 (*Cd93*) and Argonaut *AGO1* protein (Marandel et al. 2009). In addition, several works have made possible the identification of candidate gene coding for enzymes involved in apricot ripening, such as ACC oxidase (Mbégué et al. 1999); polyphenol oxidase (Chevalier et al. 1999); enzymes involved in the softening process (Mbégué et al. 2002); cell wall, sugar, lipid, organic acid, and protein metabolism (Geuna et al. 2005) and enzymes involved in apricot aroma (González-Agüero et al. 2009).

In almond, Suelves and Puigdomenech (1998) identified and sequenced a gene highly expressed in the floral organs of almond and coding for the cyanogenic enzyme (R)-(+)-mandelonitrile lyase. However, the study of the mRNA levels during seed maturation and floral development in fruit and floral samples indicated a lack of correlation between these characteristics and levels of mandelonitrile lyase mRNA and the kernel bitterness of almond cultivars. In addition, Silva et al. (2005) developed a strategy for the discovery of flowering regulatory genes in almond using almond cDNAs and *Prunus* ESTs. Finally, Sánchez-Pérez et al. (2010) assayed candidate genes involved in the amygdalin pathway, including glucosyl transferases, prunasin hydrolases and amygdalin hydrolases, and developed molecular markers linked to seed bitterness.

In plum, these studies have been focused in the search for genes responsible for fruit quality traits, ripening physiology and fruit storage (Fernández-Otero et al. 2006, 2007; El-Sharkawy et al. 2010). These authors mainly studied the genes involved in the biosynthesis and signaling of ethylene in fruit tissues.

In sweet cherry, more recently, Sooriyapathirana et al. (2010) have identified a candidate gene (*PavMYB10*), homologous to apple *MdMYB10* and *Arabidopsis AtPAP1*, located in LG 3 of the genetic linkage map. These authors have suggested that *PavMYB10* could be the major determinant of fruit skin and flesh coloration in sweet cherry.

The described *Prunus* transcriptome studies have been hampered by the lack of sufficient EST/cDNA coverage to yield significant information on the extent of regulated RNA processing events. For this reason, systems of massive analysis (high-throughput) of the transcriptome provide more rapid and reproducible information on a large

number of RNA sequences. These studies can produce larger amounts of data in terms of expression of genes and are a good alternative to candidate gene analysis and EST development for new transcriptomic studies.

### High-throughput transcriptome analysis in *Prunus*: the use of microarrays

One of these high-throughput systems is the use of cDNA biochips (microarrays) to analyze thousands of genes by hybridization of mRNA labelled with fluorescence (Schna et al. 1995). Currently, most microarray profiling systems employ glass slides containing thousands of anchored sequences of interest (Aharoni and Vorst 2001; Wullschlegler and Difazio 2003). The development of custom microarrays with probe sets designed to detect individual exons, or using combinations of probes specific to exon and splice junction sequences, has overcome many of the obstacles encountered when analyzing EST/cDNA data and thus offers a rapid means of profiling RNA expression and processing in different biological contexts (Clark et al. 2002). For expression studies using cDNA microarrays, the combination of two differently labelled samples on the same microarray is common practice in evaluating the labelled mRNA hybridized as gene expression (Aharoni and Vorst 2001).

In *Prunus* studies, microarray technology, developed in the mid-90s, has only recently been incorporated (Table 1). This technology has been used in the fabrication of different microarrays using unigene sets as probes. These studies have been performed mainly in fruit quality traits, although other important agronomic traits like flower compatibility, bud dormancy or in vitro regeneration have also been incorporated.

The  $\mu$ PEACH1.0, containing 4,806 oligonucleotide probes selected from EST collected in the ESTree repertoire, has been the first microarray developed for a large scale transcriptome analysis of peach fruit ripening (Trainotti et al. 2006). Its use allowed the identification of 269 and 109 up-regulated and down-regulated genes, respectively, during the transition from pre-climacteric to climacteric phase. The same tool has been also used for studying, more accurately, the interactions among plant hormones (Trainotti et al. 2007; Ziosi et al. 2008), and postharvest technological aspects (effect of 1-methylcyclopropene, ethylene antagonist, etc.) (Ziliotto et al. 2008).

In addition, transcriptomic studies have been carried out in apricot in the study of fruit development and ripening using the  $\mu$ PEACH1.0 (Manganaris et al. 2011). When applied to  $\mu$ PEACH1.0, apricot target cDNAs showed significant hybridization with an average of 43% of spotted probes validating the use of  $\mu$ PEACH1.0 to profile the

transcriptome of apricot fruit. Microarray analyses, carried out separately on peach and apricot fruit to profile transcriptome changes during fruit development showed that 70% of genes had the same expression pattern in both species. Such data indicate that the transcriptome (RNA level) is quite similar in apricot and peach fruit, but also highlighted the presence of species-specific transcript changes. These results confirmed the high level of synteny between *Prunus* genomes (DNA level) (Dirlewanger et al. 2004; Jung et al. 2009) above described. A similar comparative approach has been used to dissect common and/or diverse mechanisms regulating plum fruit ripening in genotypes characterized by different patterns of ethylene production (Manganaris et al. 2010).

The second microarray developed for a large scale transcriptome analysis of peach has been developed by Ogundiwin et al. (2008b) starting from a database (ChillPeach), containing 7,862 high quality ESTs (corresponding to 4,468 unigenes) obtained from mesocarp tissue of two fullsib progeny contrasting for tolerance to chilling injury (CI). The microarray analysis highlighted 399 genes differentially expressed in cold stored fruit, 287 up-regulated and 112 down-regulated. Ten of them, validated using real time quantitative PCR (qRT-PCR), have been associated to the tolerance of chilling injury being more expressed in cold-treated CI-resistant population compared to its cold-treated susceptible counterpart. More recently, Rubio-Cabetas et al. (2010) also used this ChillPeach microarray in the study of the response to hypoxia in several *Prunus* species used as rootstock. These authors found 916 genes with a different behaviour, 482 more highly expressed in the sensitive genotypes, and 434 in the tolerant genotypes.

Later, Leida et al. (2010), by combining suppression subtractive hybridization (SSH) and microarray approaches, developed a new microarray containing only 2,500 oligonucleotide probes. The obtained results evidenced the quantitative nature of the process of breaking dormancy in flower buds in peach, identifying around 100 unigenes involved in the process.

On the other hand, in the case of almond, specific microarray studies have been also performed in the identification of key genes in adventitious shoot regeneration (Santos et al. 2009). These authors using a microarray containing 3,840 probes found statistically significant differential expression for 128 cDNA clones (58 early, and 70 late), representing 92 unique gene functions. In addition, they observed different genes encoding proteins related to protein synthesis and processing, and nitrogen and carbon metabolism differentially expressed in the early stage of almond micropropagation, while genes encoding proteins involved in plant cell rescue and interaction with the environment have been mostly found in the late stage.

Part of these results obtained in peach and almond with the application of the different microarrays have been biologically validated by qRT-PCR (Ogundiwin et al. 2008b; Ziliotto et al. 2008; Santos et al. 2009; Leida et al. 2010) evidencing the key genes involved in the different process.

Finally, this technology has been used in sweet and sour cherry species in the identification of flower incompatibility alleles using other microarray containing only 80 oligonucleotide probes (can be considered a macroarray) corresponding to the first and the second intron of the *S-RNase* gene (Pasquer et al. 2008).

A major drawback of this microarray approach is that profiling coverage is strictly limited by the probe sets available for specific hybridization in each species as we have described before. For this reason and thanks to the availability of full genome sequence for some species, new tiling microarray platforms have been developed using different spaced oligonucleotides that span the entire genome of an organism (Yazaki et al. 2007). However, further contributing to the limited sensitivity and specificity of this technology is the fact that detection is indirect, measured by fluorescent signal, and is thus subjected to a variety of confounding noise variables (Bellin et al. 2009; Aharoni and Vorst 2001; Wullschlegel and Difazio 2003).

In this context, the emergence of massive (high-throughput) transcriptome sequencing techniques (RNA-Seq), made possible by lowered sequencing costs, is an alternative to the application of microarrays. Next generation DNA sequencing platforms can generate as much data in one day as several hundred traditional Sanger-type DNA capillary sequencers operated by a single person (Morozova and Marra 2008; Linnarsson 2010). In addition, van Bakel et al. (2010) indicated that tiling microarrays, in contrast to RNA-Seq, are susceptible to a high rate of false-positives in identifying transcripts with low expression levels, affecting the estimated proportion of transcription.

#### **Ultra high-throughput transcriptome analysis: transcriptome deep-sequencing (RNA-Seq)**

RNA-Seq involves direct sequencing of cDNAs using high-throughput sequencing technologies, allowing the level of transcription from a particular genomic region to be quantified from the density of corresponding reads (Flintoft 2008; Mortazavi et al. 2008). This technology represents the latest and most powerful tool for characterizing transcriptomes (Wang et al. 2009), although some specific problems were also described in comparison with the microarrays analysis including sequencing error data, system-specific error model, or reduced length of sequences that affects the possibility to unambiguously resolve

ubiquitous repeat in large genomes (Balwierz et al. 2009; Bellin et al. 2009). Several studies demonstrate that RNA-Seq provides an extremely reproducible and quantitative readout of transcript abundance in comparison with microarray analysis, although in general, there is a high degree of correspondence between the two technologies in terms of exon-level fold changes and detection (Hoen et al. 2008; Feng et al. 2010). Bradford et al. (2010) described how over 80% of exons detected as expressed in RNA-Seq have been also detected on the exon array.

The information in a single lane of sequencing data appears comparable to that in a single array in terms of enabling identification of differentially expressed genes, while allowing for additional analyses such as the detection of low-expressed genes, alternative splice variants, and novel transcripts (Marioni et al. 2008; Feng et al. 2010). In addition, because RNA-Seq is performed using tagged libraries of short cDNAs, prepared from fragmented or unfragmented RNA, it does not require prior knowledge of the sequences to be profiled. Unlike array-based approaches, RNA-Seq gives a potentially comprehensive view of the transcriptome. Another advantage of RNA-Seq is its ability to provide information on transcripts that are expressed at very low levels, limited only by the total number of reads that are generated (Flintoft 2008). In this sense, in contrast to microarrays, RNA-Seq provides a relatively unbiased and direct digital readout of cDNA sequence generated from an RNA sample (Hoen et al. 2008).

This powerful technology, available since only a couple of years, is already making substantial contributions towards the understanding of genome expression and regulation in living organisms (Parkhomchuk et al. 2009; Wilhelm and Landry 2009; Marguerat and Bähler 2010). This technology allows us to survey multiple levels of natural variation at unprecedented resolution (Gilad et al. 2009) and can be very useful in the assessment of alternative splicing and the detection of novel gene structures (Montgomery and Dermitzakis 2009). Werner (2010) indicated the great potential contribution of this technology to functional genomics with a special focus on gene regulation by transcription factor binding rates. In addition, RNA-Seq can be combined with other genomic and proteomic investigations to provide an integrated view of gene regulation (Fu et al. 2009; Hawkins et al. 2010).

The first step in RNA-Seq experiments consists of the isolation of total RNA. Later, mRNA (Poly A+) must be purified (enriched) from total RNA removing mainly rRNA which it is present in high quantities (until 85%) in the transcriptome (Fig. 1) (Haas and Zody 2010; Nagalakshmi et al. 2010; Sooknanan et al. 2010). RNA-Seq methods that do not require the purification of mRNA only are valuable for some applications, including samples with low input amounts or partial degradation (Levin et al. 2010).

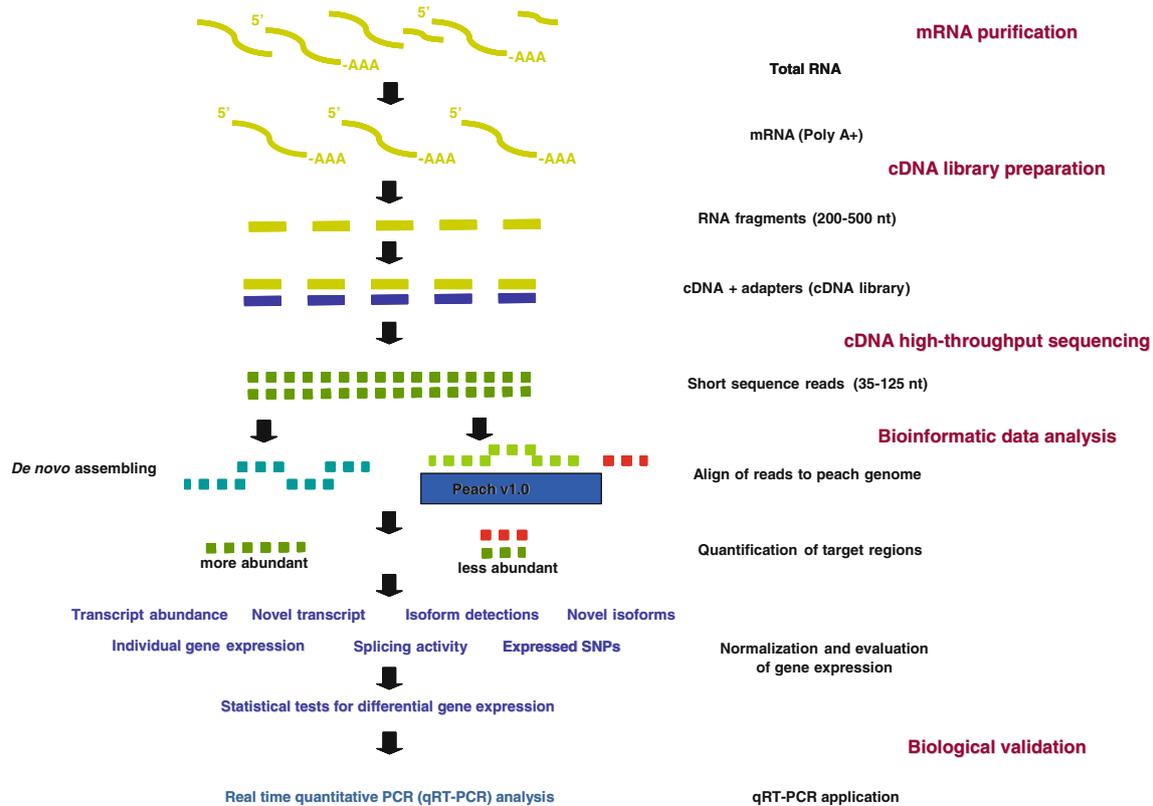
In a second step, mRNA must be fragmented into smaller pieces (200–500 bp). Later, this RNA previously fragmented is reverse-transcribed into cDNA. Then, cDNA is processed to generate a cDNA library. The goal of this step is to generate high quality, full-length cDNAs from RNA samples of interest to be fragmented and then ligated to an adapter for further amplification and sequencing (Fig. 2) (Wilhelm and Landry 2009; Nagalakshmi et al. 2010; Wang et al. 2010a).

Following a quality check, the library is denatured and transferred to a flow cell. Each molecule is then sequenced in a high-throughput manner to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing). In principle, any of the three high-throughput DNA sequencing technologies [Roche (454 Life Sciences, Branford, CT, USA), Illumina/Solexa (Genome Analyzer, San Diego, CA, USA), and ABI/Solid (Carlsbad, CA, USA)] can be used for RNA-Seq. However, currently, the most widely used system to generate RNA-Seq data is the Illumina system mainly due to the cheaper cost per base sequenced (Fig. 2).

The 454 pyro-sequencing technology (<http://www.454.com>) generates around a million readings of fragments of 200–400 nucleotides, which will account for a total of between 200 and 400 MB for each 454 run (Ando and Grumet 2010). Making an estimate of the number and average length of genes that *Prunus* may have based on similarity using those known in peach, we will obtain an approximate coverage of the transcriptome of 4×. This coverage is not enough to do comparative studies between genotypes. However, sequencing fragments obtained by 454 are suitable to be used in database homology searches and in the identification of SNPs among different genotypes despite the lower coverage of 454 sizes obtained from the ESTs (Barbazuk et al. 2007).

The Illumina Genome Analyzer platform (<http://www.illumina.com/applications.ilmn>) and the ABI/Solid technology (<http://www.appliedbiosystems.com>) can produce data sets comprising a higher number of reads, tens of millions, of a smaller size, currently of 35–125 nucleotides per read. However, in the case of ABI/Solid technology results are expressed in a colour space which must be translated to a sequence space for the further analysis.

Each Illumina/Solexa sequencing run generates about 2 Gb of sequence, which account for about 10–30 million fragments initially between 35 and 75 nucleotides (Croucher et al. 2009) and at this moment of 125 nucleotides (<http://www.illumina.com/applications.ilmn>). This will involve an estimated coverage of the transcriptome of *Prunus* of around 20×. It is known that a large number of readings are relevant in studies of RNA-Seq, as this increases the probability of capture for the mRNA sequences that are less represented and may have a role in cell physiology. In relation to this, RNA-Seq analyses



**Fig. 2** Summarized overview of RNA-Seq technology using Illumina Genome Analyzer. mRNA (*Poly A+*) purified (*enriched*) from total RNA is fragmented, ligated to specific adaptor sequences, and retrotranscribed to convert to a cDNA library. Short sequence reads are generated from the cDNA library. These reads can be mapped on a reference genome (the reference *Prunus persica* Peach genome v1.0) using efficient alignment software. Part of these reads can be aligned to previously annotated sequences in the reference genome (shown in *green*). In addition, some reads without a match to the reference map are shown in *red* (unmatched reads). Another alternative strategy consists of *de novo* assembling to produce a

new genome-scale transcriptional map. Later, a bioinformatic analysis of the data must be performed in order to describe the gene expression level, the abundance of transcripts, the presence of novel transcripts or the isoform detection. In the following steps data must be analysed in order to describe gene expression levels, to individuate novel transcripts and to perform the detection of alternative isoforms. Finally, the putative candidate genes expressed differentially can be validated by qRT-PCR. Adapted to *Prunus* analysis from Blencowe et al. (2009), Wang et al. (2009), Nagalakshmi et al. (2010), Costa et al. (2010), Haas and Zody (2010), and Wang et al. (2010a)

performed in well-characterized animal model systems have identified around 93% of the exons previously described (Mortazavi et al. 2008).

A modified protocol of the Illumina/Solexa technology called DSSS (direct strand specific sequencing) has been recently described, allowing strand-specific transcriptome sequencing (Parkhomchuk et al. 2009; Vivancos et al. 2010). DSSS enables efficient antisense detection and precisely assigns transcription start points and untranslated regions. In addition, strand-specific RNA-Seq is a powerful tool for novel transcript discovery and genome annotation because it enables the identification of the strand of origin for non-coding RNA and antisense RNA, as well as defining the ends of adjacent or overlapping transcripts transcribed in different directions (Levin et al. 2010).

Even if RNA-Seq costs for the generation of such large datasets are still considerable, the expansion of these

technologies will hopefully lead to a decrease of sequencing costs in the next future. Furthermore, when considering the price per sequenced base and the fact that RNA-Seq compared to microarray data processing and cleaning requires lower amounts of manual intervention, RNA-Seq is actually less expensive than microarray profiling methods.

### RNA-Seq data analysis

Like other high-throughput sequencing technologies, RNA-Seq faces several informatics challenges, including the development of efficient methods to store, retrieve and process large amounts of data (usually millions of reads). Furthermore, RNA-Seq is perhaps the most complex application for analysis among the different high-

throughput sequencing technologies (Blencowe et al. 2009; Auer and Doerge 2010). The unprecedented level of sensitivity and the large amount of available data produced by sequencing platforms provide clear advantages as well as new challenges and issues. To this effect, a new generation of more sophisticated algorithms and software tools is emerging to assist in the analysis of these experiments. The efficient and effective processing and analysis of RNA-Seq data is becoming the bottleneck for turning the possibilities provided by the new technology into real scientific discovery (Horner et al. 2010; Pepke et al. 2009; Wang et al. 2010b).

Information obtained in the different transcriptome analyses presents several levels of complexity (Nagalakshmi et al. 2008, 2010; Sultan et al. 2008; Wilhelm et al. 2010). The lowest level corresponds to the identification of individual genes (ESTs). These can be entered into the EST database of the National Centre for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/>) and, for what concerns *Prunus* sequences, into the GDR database. Putative SNPs that occur in different genotypes will also be recorded to complete the database of SNPs of GDR. Deep sequencing of RNA has been used to quickly generate large numbers of sequences in non model plant species, and has been also applied to SNP detection in maize transcriptomes (Barbazuk et al. 2007). The described high levels of synteny between *Prunus* genomes and inside the *Rosaceae* genomes (Dirlewanger et al. 2004; Arús et al. 2006; Shulaev et al. 2008; Jung et al. 2009) can facilitate this alignment process and the homology search for ESTs and SNPs.

At higher levels of complexity, however, we can examine the complete picture of the entire transcriptome: all the genes that are expressed in a genotype, differential splicing activity, splice isoform detection, alternative transcripts, identification of ortholog regions, location of transcription boundaries, transcription factor analysis, expressed SNPs, or novel regions of transcription (Nagalakshmi et al. 2008, 2010; Sultan et al. 2008; Chepelev et al. 2009; Zenoni et al. 2010). In addition, expression levels of specific genes, allele-specific expression of transcripts, and association of genomic regions with the known annotation regions can be accurately determined by RNA-Seq experiments in order to address many biological-related issues (Wang et al. 2009; Bashir et al. 2010; Wang et al. 2010a). This level of complexity in a holistic view of the transcriptome is not possible using systems other than massive sequencing. All these attributes are not readily achievable from previously widespread hybridization-based or tag sequence-based approaches.

A first step in the analysis of RNA-Seq data could be to identify reads that uniquely align to the reference genome and transcriptome. In this sense, to date there is no a “gold

standard” method for analyzing RNA-Seq data, and several laboratories have developed independent methods that score read alignments and counts in different ways (Pepke et al. 2009). However, with this strategy, reads that derive from non-unique sequences including transcribed retrotransposons, duplicated or paralogous genes, and repetitive splice junction sequences can be eliminated with the consequent loss of information. An example of the information that can be recovered by RNA-Seq analysis is given by Mortazavi et al. (2008), who analyzed mouse tissue transcriptomes using Illumina data sets consisting of 41–52 million 25-nt reads from three different mouse tissues. These authors identified unique zapping reads corresponding to 17,000 previously unannotated regions of known genes. Many of these sequences appeared to correspond to extended 3' or 5' untranslated regions (UTRs), indicating that these genes contain longer and more variable UTR sequences than previously believed. In addition, 145,000 distinct splice junctions have been detected (from a total of 200,000 previously annotated junctions), and alternative splicing has been detected in 3,500 genes. However, this process can remove reads from transcribed retrotransposons.

Two different strategies for reconstructing transcripts from RNA-Seq reads have been developed. RNA-Seq data analysis should be mapped to a reference genome if available, as in the case of *Prunus*, or de novo assembled to produce a genome-scale transcriptional map (Fig. 2).

There are several programs for mapping reads to the genome, including ELAND, SOAP, SeqMap, Galaxy, ERANGE, MAQ, RMAP, MIRA and Mosaik (Horner et al. 2010; Wang et al. 2009; Li and Homer 2010). In *Prunus* species, this alignment is currently performed onto the reference peach genome (v1.0) (<http://www.rosaceae.org>) (Sosinski et al. 2010). After sequences quality filtering reads are assembled to generate “contigs” and homology to known sequences inferred by BLAST (Basic Local Alignment Search Tool) versus reference databases (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) as well as *Prunus*-specific databases (<http://www.rosaceae.org>).

On the other hand, different short-read assembly programs including Velvet and ABySS have been developed to produce a genome-scale transcriptional map assembling the de novo ESTs (Horner et al. 2010; Pepke et al. 2009). This second strategy can be an alternative to the alignment of reads to the reference genome which can better resolve the problem of the reduced length of the sequences, although it is much more complex (Denoeud et al. 2008; Robertson et al. 2010). In this sense, Paszkiewicz and Studholme (2010) have described around twenty different software tools and they have implemented new algorithms specially aimed at de novo assembling reads of 30–50 nucleotides in length. One of this software tools (Oases) is

specifically designed for assembling transcribed sequences. Surget-Groba and Montoya-Burgos (2010) have developed a methodology to assemble transcriptomes where the sequence coverage of transcript is highly heterogeneous. More recently, an automated software pipeline (Rnnotator) that generates transcript models by de novo assembly of stranded RNA-Seq reads without the need for a reference genome has been developed (Martin et al. 2010) open the door to a faster de novo assembly of transcriptomes.

Browser-driven analyses are very important for visualizing the quality of the data and to interpret specific events on the basis of the available annotations and mapped reads. However, they only provide a qualitative picture of the phenomenon under investigation, and the enormous amount of data does not allow us to easily focus on the most relevant details. Hence, the second phase of most of the RNA-Seq pipeline consists of the automatic quantification of the transcriptional events across the entire genome (Costa et al. 2010; Haas and Zody 2010; Wang et al. 2010a). From this point of view, the interest lies in both quantifying known elements and detecting new transcribed regions, defined as transcribed segments of DNA not yet annotated as exons in databases. The ability to detect these unannotated regions, even though biologically relevant, is one of the main advantages of the RNA-Seq over microarray technology. Usually, the quantification step is preliminary to any differential expression approach (Pepke et al. 2009; Parkhomchuk et al. 2009).

One of the main objectives in most transcriptome studies is to quantify differences in expression across multiple samples in order to capture differential gene expression, to identify sample-specific alternative splicing isoforms and their differential abundance. The differential analysis for the clean read sequences from RNA-Seq can be analyzed using an R Package (R Development Core Team 2009) for the identification of differentially expressed genes (<http://bioinfo.au.tsinghua.edu.cn>) using the DEG-Seq (Wang et al. 2010c). These results could have a great potential application in the identification of differentially expressed genes that may be correlated with the differential response of the organism.

In differential expression assays, in order to discover biologically important changes in expression, scaling normalization can be an important step in the analysis mainly in the case of heavily expressed genes and in studies of comparison of global response of the transcriptome (Balwierz et al. 2009). These authors suggested that, across all human and mouse tissues, there is no natural expression scale that distinguishes the large number of transcription start sites (TSSs) that are expressed at very low rates—called background transcription by these authors—from the highly regulated expression of the TSSs of highly expressed genes. Background transcription of the most highly

expressed genes are just the extreme of a scale-free distribution and it is possible to normalize the expression data from different deep sequencing data-sets to estimate relative RNA production levels and clarify the information. Robinson and Oshlack (2010) have recently developed a simple and effective method for performing normalization of data which improve the results for inferring differential expression in available data sets. However, this normalization can produce a loss of biological information mainly in the case of low expressed genes. In addition, after this normalization, the quantitative nature of the gene expression is changing to a qualitative nature (fold change levels) as in the case of the microarray analysis (Bellin et al. 2009; Hoen et al. 2008; Feng et al. 2010).

On the other hand, a new software tool (ALEXA-Seq) based on gene model profiling has been improved for the analysis of alternative splice events in RNA sequencing (Cloonan and Grimmond 2010). Knowing the precise sequence of the expressed genes, it is possible to detect individual transcriptomic events. Other R software packages for different statistical analysis of RNA-Seq data from the Bioconductor Project can be found in <http://www.bioconductor.org>.

Finally, the putative candidate genes expressed differentially identified using RNA-Seq technology can be validated by real time quantitative PCR (qRT-PCR) studies mainly in the case of low expression levels (Zenoni et al. 2010). This strategy using qRT-PCR have been largely applied in studies of microarray analysis using housekeeping genes as reference genes in case of intermediate-high expressed genes (Ogundiwin et al. 2008b; Ziliotto et al. 2008). In addition, Tong et al. (2009) defined eleven reference genes specific of *Prunus* species which displayed a wide range of quantification cycle ( $C_q$ ) values using RT-PCR with SYBR green and can be used for this biological validation of results.

### RNA-Seq application in plant species including *Prunus*

RNA-Seq has mainly been applied to humans, mammals and yeast, and it is now being applied to plant species (Parkhomchuk et al. 2009; Wilhelm and Landry 2009; Marguerat and Bähler 2010). In plants, RNA-Seq offers the same potential and opportunities for transcriptome studies and gene expression analysis as in other living organisms (Bräutigam and Gowik 2010; Priest et al. 2010). In the case of *Prunus* species the application of this technology is starting (Table 1).

The first application of this methodology to plant species has been performed in the model species *Arabidopsis thaliana* L. Weber et al. (2007) initiated these studies using massive sequencing of *Arabidopsis* DNA for the detection

of transcripts, comparing their results with the developed ESTs. These authors identified 17,444 unigenes in *Arabidopsis*. These studies have been completed using the Illumina platform for ultra-high-throughput RNA sequencing (Filichkin et al. 2010). These authors confirmed a majority of annotated introns and identified thousands of novel alternatively spliced mRNA isoforms; they suggested that at least 42% of intron-containing genes in *Arabidopsis* are alternatively spliced. More recently, this technology has been applied in *Arabidopsis* in the study of protein arginine methylation in transcriptional regulation and RNA processing assays (Deng et al. 2010) and the identification of meiosis-specific genes by combining isolated meiocytes, RNA-Seq, bioinformatics and statistical analysis (Chen et al. 2010).

In the case of crop species, RNA-Seq has been applied to date in maize, cucumber, rice, soybean, tomato and grapevine. In maize (*Zea mays* L.), Vega-Arreguín et al. (2009) developed a protocol to optimize the sequencing of cDNAs using four consecutive 454 pyrosequencing runs of a cDNA library obtained from 2 week-old ‘Palomero Toluqueño’ maize plants. These 454 runs generated over 1.5 million reads, representing the largest number of sequences reported from a single plant cDNA library. A collection of 367,391 quality-filtered reads (30.09 Mb) from a single run has been sufficient to identify transcripts corresponding to 34% of public maize ESTs databases. In addition, Eveland et al. (2010), used this technique in the study of meristem development mapping 86% of nonredundant signature tags to the reference maize genome, which associated with 37,177 gene models and unannotated regions of expression and Li et al. (2010) applied RNA-Seq using Illumina Genetic Analyzer to the developmental dynamics of the maize leaf transcriptome.

On the other hand, Ando and Grumet (2010) used 454 pyrosequencing to develop a cucumber (*Cucumis sativus* L.) fruit transcriptome atlas, identifying highly expressed transcripts and characterizing key functions during exponential fruit growth. The resulting 187,406 expressed sequence tags (ESTs) have been assembled into 13,878 contigs.

In rice (*Oryza sativa* L.), Zhang et al. (2010) and Lu et al. (2010), using high-throughput paired-end RNA-Seq, described the high complexity of the transcriptome. An analysis of alternative splicing in the rice transcriptome revealed that alternative *cis*-splicing occurred in ~33% of all rice genes. In addition, these authors also identified 234 putative chimeric transcripts that seem to be produced by *trans*-splicing, indicating that transcript fusion events are more common than expected (Zhang et al. 2010).

In soybean (*Glycine max* L.), Severin et al. (2010) developed a RNA-Seq atlas that extends the analyses of previous gene expression atlases performed using

microarray technology providing an example of new methods to accommodate the increase in transcriptome data obtained from next generation sequencing. Data contained within this RNA-Seq atlas of soybean can be explored at <http://www.soybase.org/soyseq>.

In tomato (*Lycopersicon esculentum* Mill.), Francis (2010), studying the whole transcriptome of six different inbred cultivars, used Illumina II sequencing technology to generate >2.5 Gb of total sequence for each tomato cultivar. This author assembled 60 bp reads to generate 32.5 Mb of transcriptome sequence.

In tree species, Zenoni et al. (2010) reported the first use of RNA-Seq to gain insight into the wide range of transcriptional responses associated with berry development in grapevine (*Vitis vinifera* L.) cultivar ‘Corvina’. More than 59 million sequence reads, 36–44 bp in length, have been generated from three developmental stages: post setting, veraison and ripening. The sequence reads have been aligned onto the 8.4-fold draft sequence of the Pinot Noir 40024 genome and then analyzed to measure gene expression levels and to detect alternative splicing events and expressed single nucleotide polymorphisms. These authors detected 17,324 genes expressed during berry development, 6,695 of which have been expressed in a stage-specific manner, suggesting differences in expression for genes in numerous functional categories and a significant transcriptional complexity. In addition, Picardi et al. (2010) used short read from Illumina/Solexa and ABI/Solid to investigate the editing pattern in mtRNA (mitochondrial RNA) of *Vitis vinifera* providing significant support for conversions in coding regions and additional modifications in noncoding RNAs.

In the case of *Prunus* species, the sequencing of the complete genome of the model species peach (Sosinski et al. 2010) offers new opportunities for RNA-Seq technology. Currently, different groups are undertaking RNA-Seq studies (Table 1).

A recent project is being developed by our group on gene expression analysis of resistance to PPV in two apricot genotypes with similar agronomical characteristics, ‘Rojo Pasión’ and ‘Z506-7’, from the same cross [‘Orange Red’ (resistant to PPV) × ‘Currot’ (susceptible to PPV)], but one being resistant to PPV (‘Rojo Pasión’) and the other susceptible (‘Z506-7’) (Rubio et al. 2010). In these experiments an approach based on deep sequencing of transcriptome with an unreplicate design (Auer and Doerge 2010) has been assayed. First results performed in collaboration with the Ultrasequencing Unit of the Centre for Genomic Regulation of Barcelona (Spain) (<http://www.crg.es>) showed the high number of reads of 32 nucleotides (more than 20 million) obtained per line of a flow cell from Illumina Genome Analyzer Iix, and their good quality (only 2% of the reads have been discarded). In addition,

**Table 2** Summary of read number in the deep sequencing of the mRNA of ‘Rojo Pasión’ and Z506-7’ apricot genotypes

	‘Rojo Pasión’		‘Z506-7’	
	Number	%	Number	%
Total reads sequenced (32 nt)	24,697,708	100	19,456,487	100
Total reads aligning to only a single location	17,861,031	72.3	14,153,017	72.7
Unique hits with 0 mismatches <sup>a</sup>	10,173,713	41.2	8,073,698	41.5
Unique hits with 1 mismatches <sup>a</sup>	5,547,046	22.5	4,395,740	22.6
Unique hits with 2 mismatches <sup>a</sup>	2,140,272	8.7	1,683,579	8.7
Hits aligning to multiple location	1,374,180	5.6	1,050,884	5.4
Reads without a match to the reference <sup>a</sup>	4,994,949	20.2	3,867,465	19.9
Reads removed due to poor quality	467,548	1.9	385,121	2.0

<sup>a</sup> Alignment of reads (using ELAND mapper version 1.6) to the *Prunus persica* genome (v1.0) (<http://www.rosaceae.org>; Sosinski et al. 2010) used as reference

72% of these reads have been aligned to a single location to the reference peach genome (v1.0) (<http://www.rosaceae.org/peach/genome>) (Sosinski et al. 2010) with 0, 1 or 2 mismatches (Table 2). These first results confirmed the high levels of synteny between *Prunus* genomes (Jung et al. 2009) and the utility of the peach genome as reference in RNA-Seq studies in other *Prunus* species. Differential analysis of this RNA-Seq data (Wang et al. 2010b, c) is being performed with a great potential application in the identification of differentially expressed genes that may be correlated with the differential response of apricot against PPV.

Pina et al. (2010) used pyro-sequencing 454 to sequence two *Prunus* transcriptomes for the study of the molecular mechanisms of graft incompatibility. The number of reads obtained (around 600,000) has been smaller although of higher size (385 nucleotides of mean) than in the case of the sequencing using Illumina (Rubio et al. 2010) (Table 2). These authors identified 207 contigs from the graft-compatible libraries although these results are now under biological validation using qRT-PCR.

In addition, another complex biological model of key interest is vegetative and flower bud dormancy. Recent transcriptomic studies in peach using microarrays evidenced a large number (around 100) of genes involved in this process (Leida et al. 2010). RNA-Seq can complete this previous result and better elucidate the genetic control of this complex and key trait in the adaptation and productivity of *Prunus* species. In this species, Zhebentyayeva et al. (2010) have developed a comprehensive program for identifying genetic pathways and potential epigenetic mechanisms involved in control of chilling requirement and flowering time using the deep sequencing of miRNA as a complementary experimental system.

Finally, traits related to fruit quality and storage suitability are also of great interest in *Prunus* species. In this field, advanced transcriptomic studies have been performed

using microarrays, leading to significant knowledge of the ESTs involved in these complex traits (Grimplet et al. 2005; Ogundiwin et al. 2008b; Ziliotto et al. 2008). For this reason, RNA-Seq is being used as a continuation of these previous studies.

**Acknowledgments** This study has been supported by the projects “Importance, transmission and resistance sources in the main viruses affecting stone fruits in the Region of Murcia” (08672/PI/08) of the Seneca Foundation of the Region of Murcia and “Gene expression analysis of the resistance to *Plum pox virus*, PPV (Sharka) in apricot by transcriptome deep-sequencing (RNA-Seq)” of the Spanish Ministry of Science and Innovation (Project reference AGL2010-16335).

## References

- Abbott AG, Sosinski B, Orellana A (2009) Functional genomics in peach. In: Folta KM, Gardiner SE (eds) Genetics and genomics of rosaceae. Springer, Heidelberg, pp 259–275
- Aharoni A, Vorst O (2001) DNA microarray for functional plant genomics. *Plan Mol Biol* 48:99–118
- Ando K, Grumet R (2010) Transcriptional profiling of rapidly growing cucumber fruit by 454-Pyrosequencing analysis. *J Am Soc Hort Sci* 135:291–302
- Arumuganathan K, Earle DE (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9:208–218
- Arús P, Yamamoto T, Dirlwanger E, Abbott AG (2006) Synteny in the rosaceae. *Plant Breed Rev* 27:175–211
- Arús P, Gradziel TM, Oliveira M, Tao R (2009) Genomics of almond. In: Folta KM, Gardiner SE (eds) Genetics and genomics of rosaceae. Springer, Heidelberg, pp 187–242
- Auer PL, Doerge RW (2010) Statistical design and analysis of RNA-Seq data. *Genetics* 185:405–416
- Baird WV, Estager AS, Wells JK (1994) Estimating nuclear-DNA content in peach and related diploid species using laser flow-cytometry and DNA hybridization. *J Amer Soc Hort Sci* 119:1312–1316
- Balwierz PJ, Carninci P, Daub CO, Kawai J, Hayashizaki Y, Van Belle W, Beisel C, van Nimwegen E (2009) Methods for analyzing deep sequencing expression data: constructing the human and mouse promoterome with deep CAGE data. *Genome Biol* 10:R79

- Barbazuk WB, Emrichm SJ, Chen HD, Li L, Schnable P (2007) SNP discovery via 454 transcriptome sequencing. *Plant J* 51: 910–918
- Bashir A, Bansla V, Bafna V (2010) Designing deep sequencing experiments: detecting structural variation and estimating transcript abundance. *BMC Genomics* 11:385
- Basset CL, Wisniewski ME, Artlip TS, Richart G, Norelli JL, Farel RE (2009) Comparative expression and transcript initiation of three peach dehydrin genes. *Planta* 230:107–118
- Bellin D, Ferrarini A, Chimento A, Kaiser O, Levenkova N, Bouffard P, Delledonne M (2009) Combining next-generation pyrosequencing with microarray for large scale expression analysis in non-model species. *BMC Genomics* 10:555
- Bielenberg DG, Wang Y, Li Z, Zhebentyayeva T, Fan S, Reighard GL, Scorza R, Abbott AG (2008) Sequencing and annotation of the evergrowing locus in peach (*Prunus persica* (L.) Batsch] reveals a cluster of six MADS-box transcription factors as candidate genes for regulation of terminal bud formation. *Tree Genet Genomes* 4:495–507
- Blencowe BJ, Ahmad S, Lee LJ (2009) Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes. *Gen Develop* 23:1379–1386
- Bonghi C, Ferrarese L, Ruperti B, Tonutti P, Ramina A (1998) Endo- $\beta$ -1, 4-glucanases are involved in peach fruit growth and ripening, and regulated by ethylene. *Physiol Plantarum* 102: 346–352
- Bonghi C, Begheldo M, Ziliotto F, Rasori A, Trainotti L, Tosetti R, Tonutti P (2010) Transcriptome analyses and postharvest physiology of peaches and nectarines. *Acta Hort* 877:69–73
- Bradford JR, Hey Y, Yates T, Li YY, Pepper SD, Miller CJ (2010) A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. *BMC Genomics* 11:282
- Bräutigam A, Gowik U (2010) What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. *Plant Biol* 12:831–841
- Cabrera A, Koziak A, Howad W, Arús P, Jezzone AF, van der Knaap E (2009) Development and bin mapping of a Rosaceae conserved ortholog set (COS) of markers. *BMC Genomics* 10:562
- Chen CB, Farmer AD, Langley RJ, Mudge J, Crow J, May GD, Huntley J, Smith AG, Retzel EF (2010) Meiosis-specific gene discovery in plants: RNA-Seq applied to isolated *Arabidopsis* male meiocytes. *BMC Plan Biol* 10:280
- Chepelev I, Wei G, Tang Q, Zao K (2009) Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. *Nucleic Acids Res* 37:e106
- Chevalier T, Rigal D, Mbeguie-A-Mbeguie D (1999) Molecular cloning and characterization of apricot fruit polyphenol oxidase. *Plant Physiol* 119:1261–1269
- Clark TA, Sugnet CW, Ares MJ (2002) Genome wide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* 296:907–910
- Cloonan N, Grimmond SM (2010) Simplifying complexity. *Nat Methods* 7:793–794
- Costa V, Angelini C, De Feis I, Ciccodicol A (2010) Uncovering the complexity of transcriptomes with RNA-Seq. *J Biomed Biotech*. Article ID 853916
- Croucher NJ, Fookes MC, Perkins TT, Turner DJ, Marguerat SB, Keane T, Quail MA, He M, Assefa S, Bahler J, Kingsley RA, Parkhill J, Bentley SD, Dougan G, Thomson NR (2009) A simple method for directional transcriptome sequencing using Illumina technology. *Nucleic Acids Res* 37:e148
- Decroocq V, Fave MG, Hagen L, Bordenave L, Decroocq S (2003) Development and transferability of apricot and grape EST microsatellite markers across taxa. *Theor Appl Genet* 106: 912–922
- Decroocq V, Foulongne M, Lambert P, Gall OL, Mantin C, Pascal T, Schurdi-Levraud T, Kervella J (2005) Analogues of virus resistance genes map to QTLs for resistance to sharka disease in *Prunus davidiana*. *Mol Genet Genomics* 272:680–689
- Deng X, Gu L, Lu T, Lu F, Lu Z, Cui P, Pei Y, Wang B, Hu S, Cao X (2010) Arginine methylation mediated by the *Arabidopsis* homolog of PRMT5 is essential for proper pre-mRNA splicing. *Proc Natl Acad Sci USA* 107:19114–19119
- Denoeud F, Aury JM, Da Silva C, Noel B, Rogier O, Delledonne M, Morgante M, Valle G, Wincker P, Scarpelli C, Jaillon O, Artiguenave F (2008) Annotating genomes with massive-scale RNA sequencing. *Genome Biol* 9:R175
- Dirlewanger E, Graziano E, Joobeur T, Garriga-Caldre F, Cosson P, Howad W, Arús P (2004) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. *P Natl Acad Sci USA* 101:9891–9896
- Dirlewanger E, Denoyes B, Yamamoto T, Chagné D (2009a) Genomics tools across Rosaceae species. In: Folta KM, Gardiner SE (eds) *Genetics and genomics of rosaceae*. Springer, Heidelberg, pp 539–561
- Dirlewanger E, Claverie J, Jezzone AF, Wünsch A (2009b) Sweet and sour cherries: linkage maps, QTL detection and marker assisted selection. In: Folta KM, Gardiner SE (eds) *Genetics and genomics of rosaceae*. Springer, Heidelberg, pp 291–313
- Dondini L, Costa F, Tataranni G, Tataranni S, Sansavini S (2004) Cloning of apricot RGAs (resistance gene analogs) and development of molecular markers associated with Sharka (PPV) resistance. *J Hort Sci Biotech* 79:729–734
- El-Sharkawy I, Sherif S, Mila I, Bouzayen M, Jayasankar S (2010) Molecular characterization of seven encoding ethylene-responsive transcriptional factors during plum fruit development and ripening. *J Exp Bot* 60:907–922
- Esmenjaud D, Dirlewanger E (2007) Plum. In: Kole CR (ed) *Genome mapping and molecular breeding*. Fruits & nuts, vol 4. Springer, Heidelberg, pp 119–136
- Etienne C, Rothan C, Moing A, Plomion C, Bodenes C, Svanella-Dumas L, Cosson P, Pronier V, Monet R, Dirlewanger E (2002) Candidate genes and QTLs for sugar and organic content in peach [*Prunus persica* (L.) Batsch]. *Theor Appl Genet* 105: 145–159
- Eveland AL, Satoh-Nagasawa N, Goldshmidt A, Meyer S, Sakai H, Ware D, Jackson D (2010) Digital gene expression signatures for maize development. *Plant Physiol* 154:1024–1039
- Falara V, Manganari GA, Ziliotto F, Manganaris A, Bonghi C, Ramina A, Kanellis AK (2011) A  $\beta$ -D-xylosidase and a PR-4B precursor identified as genes in accounting for differences in peach cold storage tolerance. *Funct Int Gen* 11 (in press)
- Falchi R, Cipriani G, Marrazzo T, Nonis A, Vizzotto G, Ruperti B (2010) Identification and differential expression dynamics of peach small GTPases encoding genes during fruit development and ripening. *J Exp Bot* 61:2829–2842
- Feng L, Liu H, Liu Y, Lu ZK, Guo GW, Guo SP, Zheng HW, Gao YN, Cheng SJ, Wang J, Zhang Y (2010) Power of deep sequencing and agilent microarray for gene expression profiling study. *Mol Biotech* 45:101–110
- Fernández-Otero C, Matilla AJ, Rasori A, Ramina A, Bonghi C (2006) Regulation of ethylene biosynthesis in reproductive organs of damson plum (*Prunus domestica* L. Subsp Syriaca). *Plant Science* 171:74–83
- Fernández-Otero CI, Torre F, Iglesias R, Rodríguez-Gacio MC, Matilla AJ (2007) Stage- and tissue-expression of genes involved in the biosynthesis and signalling of ethylene in reproductive organs of damson plum. *Plant Physiol Bioch* 45:199–208
- Filichkin SA, Priest HD, Givan SA, Shen RK, Bryant DW, Fox SE, Wong WK, Mockler TC (2010) Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Gen Res* 20:45–58

- Flintoft L (2008) Transcriptomics: digging deep with RNA-Seq. *Nat Rev Genet* 9:413
- Folta KM, Gardiner SE (2009) Genomic-based opportunities in apricot. In: Folta KM, Gardiner SE (eds) *Genetics and genomics of rosaceae*. Springer, Heidelberg, pp 315–335
- Forrest ARR, Carninci P (2009) Whole genome transcriptome analysis. *RNA Biol* 6:107–112
- Francis D (2010) Next generation sequencing of the tomato transcriptome. In: *International horticulture conference, Lisboa (Portugal)*, p 10
- Fu X, Fu N, Guo S, Yan Z, Xu Y, Hu H, Menzel C, Chen W, Li YX, Zeng R, Khaitovich P (2009) Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC Genomics* 10:161
- Geuna F, Banfi R, Bassi D (2005) Identification and characterization of transcripts differentially expressed during development of apricot (*Prunus armeniaca* L.). *Tree Genet Genomes* 1:69–78
- Gilad Y, Pritchard JK, Thornton K (2009) Characterizing natural variation using next-generation sequencing technologies. *Trends Genet* 25:463–471
- González-Agüero M, Pavez L, Ibanez F, Pacheco I, Campos-Vargas R, Meisel LA, Orellana A, Retamales J, Silva H, Gonzalez M, Cambiazo V (2008) Identification of woolliness response genes in peach fruit after post-harvest treatments. *J Exp Bot* 59:1973–1986
- González-Agüero M, Troncoso S, Gudenschwager O, Campos-Vargas R, Moya-León MA, Defilippi BG (2009) Differential expression levels of aroma-related genes during ripening of apricot (*Prunus armeniaca* L.). *Plant Physiol Bioch* 47:435–440
- Grimplet J, Romieu C, Audergon JM, Marty I, Albagnac G, Lambert P, Bouchet JP, Terrier N (2005) Transcriptomic study of apricot fruit (*Prunus armeniaca* L.) ripening among 13, 006 expressed sequence tags. *Physiol Plant* 125:281–292
- Haas BJ, Zody MC (2010) Advancing RNA-Seq analysis. *Nat Biotech* 28:421–423
- Hawkins RD, Hon GC, Ren B (2010) Next-generation genomics: an integrative approach. *Nat Rev Genet* 11:476–486
- Hoehn PAC, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RHAM, de Menezes RX, Boer JM, van Ommen GJB, den Dunnen JT (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Res* 36:e141
- Horn R, Lecouls AC, Callahan A, Dandekar A, Garay L, McCord P, Howad W, Chan H, Verde I, Main D, Jung S, Georgi L, Forrest S, Mook J, Zhebentyayeva T, Yu YS, Kim HR, Jesudurai C, Sosinski B, Arús P, Baird V, Parfitt D, Reighard G, Scorza R, Tomkins J, Wing R, Abbott AG (2005) Candidate gene database and transcript map for peach, a model species for fruit trees. *Theor Appl Genet* 110:1419–1428
- Horner DS, Pavesi G, Castrignano T, De Meo PD, Liuni S, Sammeth M, Picardi E, Pesole G (2010) Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Briefs Bioinform* 11:181–197
- Illa E, Sargeant DJ, Lopez E, Bushara J, Cestaro A, Pindo M, Cabrera A, Iezzoni A, Gardiner S, Velasco R, Arús P, Chagne D, Troggio M (2011) Comparative analysis of rosaceous genomes and the reconstruction of a putative ancestral genome for the family. *BMC Evol Biol* 11:9
- Jiang YQ, Ma RC (2003) Generation and analysis of expressed sequence tags from almond (*Prunus dulcis* Mill.) pistils. *Sex Plant Rep* 16:197–207
- Jiménez S, Li ZG, Reighard GL, Bielenberg DG (2010a) Identification of genes associated with growth cessation and bud dormancy entrance using a dormancy-incapable tree mutant. *BMC Plant Biol* 10:25
- Jiménez S, Reighard GL, Bielenberg DG (2010b) Gene expression of *DAM 5* and *DAM6* is suppressed by chilling temperatures and inversely correlated with bud break. *Plant Mol Biol* 73:157–167
- Jung S, Jesudurai C, Staton M, Ficklin S, Cho I, Abbott A, Tomkins J, Main D (2004) GDR (Genome database for Rosaceae): integrated web resources for Rosaceae genomics and genetics research. *BMC Bioinform* 5:130
- Jung S, Jiwan D, Cho I, Abbott A, Tomkins J, Main D (2009) Synteny of *Prunus* and other model plant species. *BMC Genomics* 10:76
- Lalli DA, Decroocq V, Blenda AV, Garay L, Le Gall O, Damsteegt V, Reighard GL, Abbott AG (2005) Identification and mapping of resistance gene analogs (RGA) in *Prunus*: a resistance map for *Prunus*. *Theor Appl Genet* 111:1504–1513
- Lazzari B, Caprera A, Vecchiotti A, Merelli I, Barale F, Milanese L, Stella A, Pozzi C (2008) Version VI of the ESTree db: an improved tool for peach transcriptome analysis. *BMC Bioinform* 9:S9
- Le Dantec L, Cardinet G, Bonet J, Fouché M, Boudehri K, Monfort A, Poëssel JL, Moing A, Dirlwanger E (2010) Development and mapping of peach candidate genes involved in fruit quality and their transferability and potential use in other Rosaceae species. *Tree Genet Genomes* 6:995–1012
- Leader DJ (2005) Transcriptional analysis and functional genomics in wheat. *J Cereal Sci* 41:149–163
- Leida C, Terol J, Martí G, Agustí M, Llácer G, Badenes ML, Ríos G (2010) Identification of genes associated with bud dormancy release in *Prunus persica* by suppression subtractive hybridization. *Tree Physiol* 30:655–666
- Levin J, Adiconis X, Yassour M, Thompson D, Guttman M, Berger M, Fan L, Friedman N, Nusbaum C, Gnirke A, Regev A (2010) Development and evaluation of RNA-Seq methods. *Genome Biol* 11:P26
- Li H, Homer N (2010) A survey of sequence alignment algorithms for next generation sequencing. *Briefs Bioinform* 11:473–483
- Li ZG, Reighard GL, Abbott AG, Bielenberg DG (2009) Dormancy-associated MADS genes from the EVG locus of peach [*Prunus persica* (L.) Batsch] have distinct seasonal and photoperiodic expression patterns. *J Exp Bot* 60:3521–3530
- Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta L, Kebrom TH, Provart N, Patel R, Myers CR, Reidel EJ, Turgeon R, Liu P, Sun Q, Nelson T, Brutnell T (2010) The developmental dynamics of the maize leaf transcriptome. *Nat Genet* 42:1060–1067
- Linnarsson S (2010) Recent advances in DNA sequencing methods—general principles of sample preparation. *Exp Res* 316:1339–1343
- Lu TT, Lu GJ, Fan DL, Zhu CR, Li W, Zhao QA, Feng Q, Zhao Y, Guo YL, Li WJ, Huang XH, Han B (2010) Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-Seq. *Genet Res* 20:1238–1249
- Manganaris GA, Ziosi V, Bonghi C, Costa G, Tonutti P, Ramina A (2010) A preliminary transcriptomic approach to elucidate postharvest ripening of plum fruit. *Acta Hort* 874:99–106
- Manganaris GA, Rasori A, Bassi D, Geuna F, Ramina A, Tonutti P, Bonghi C (2011) Comparative transcript profiling of apricot (*Prunus armeniaca* L.) fruit development and on-tree ripening. *Tree Genet Genomes* 7 (in press)
- Marandel G, Pascal T, Candresse T, Decroocq V (2009) Quantitative resistance to *Plum pox virus* in *Prunus davidiana* P1908 linked to components of the eukaryotic translation initiation complex. *Plant Pathol* 58:425–435
- Marguerat S, Bähler J (2010) RNA-Seq: from technology to biology. *Cell Mol Life Sci* 67:569–579
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-Seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18:1509–1517

- Martin J, Bruno VM, Fang ZD, Meng XD, Blow M, Zhang T, Sherlock G, Snyder M, Wang Z (2010) Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC Genomics* 11:663
- Martínez-Gómez P, Sánchez-Pérez R, Dicenta F, Howad W, Arús P, Gradziel TM (2007) Almonds. In: Kole CR (ed) *Genome mapping and molecular breeding. Fruits & nuts, vol 4*. Springer, Heidelberg, pp 229–242
- Mbégué D, Chahine H, Gomez RM, Gouble B, Reich M, Audergon J-M, Souty M, Albagnac G, Fils-Lycaon B (1999) Molecular cloning and expression of a cDNA encoding 1-aminocyclopropane-1-carboxylate (ACC) oxidase from apricot fruit (*Prunus armeniaca*). *Physiol Plant* 105:294–303
- Mbégué D, Gouble B, Gomez RM, Audergon J-M, Albagnac G, Fils-Lycaon B (2002) Two expansion cDNAs from *Prunus armeniaca* expressed during fruit ripening are differently regulated by ethylene. *Plant Physiol Bioch* 40:445–452
- Meneses C, Jung S, Arús P, Aranzana MJ, Abbott A (2007) In silico analysis and first applications of SNPs from the GDR database in peach. In: XII EUCARPIA fruit section symposium, Zaragoza (Spain), p 67
- Montgomery SB, Dermitzakis ET (2009) The resolution of the genetics of gene expression. *Hum Mol Genet* 18:R211–R215
- Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92:255–264
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5:621–628
- Nagalakshmi U, Wang Z, Waern K, Shou C, Gerstein M, Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320:1344–1349
- Nagalakshmi U, Waern K, Snyder M (2010) RNA-Seq: a method for comprehensive transcriptome analysis. *Curr Prot Mol Biol*, 4.11.1–4.11.13
- Ogundiwin EA, Peace C, Nicolet CM, Rashbrook VK, Gradziel TM, Bliss FA, Parfitt D, Crisosto C (2008a) Leucoanthocyanidin dioxygenase gene (PpLDOX): a potential functional marker for cold storage browning in peach. *Tree Genet Genomes* 4:543–554
- Ogundiwin EA, Marti C, Forment J, Pons C, Granell A, Gradziel TM, Peace C, Crisosto C (2008b) Development of ChillPeah genomic tools and identification of cold-responsive genes in peach fruits. *Plant Mol Biol* 68:379–397
- Ogundiwin EA, Peace C, Gradziel TM, Bliss FA, Parfitt D, Crisosto C (2009) A fruit quality gene map of *Prunus*. *BMC Genomics* 10:587
- Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobisch S, Lehrach H, Soldatov A (2009) Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res* 37:e123
- Pasquer F, Frey B, Frey E (2008) Identification of cherry incompatibility alleles by microarray. *Plant Breed* 127:413–417
- Paszkiwicz K, Studholme DJ (2010) De novo assembly of short sequence reads. *Briefs Bioinform* 11:457–472
- Peace C, Norelli JL (2009) Genomics approaches to crop breeding improvement in Rosaceae. In: Folta KM, Gardiner SE (eds) *Genetics and genomics of Rosaceae*. Springer, Heidelberg, pp 19–53
- Pepke S, Wold B, Mortazavi A (2009) Computation for ChIP-seq and RNA-Seq studies. *Nat Methods* 6:S22–S32
- Picardi E, Horner DS, Chiara M, Valle G, Pesole G (2010) Large-scale detection and analysis of RNA editing in grape mtRNA by RNA deep-sequencing. *Nucleic Acids Res* 38:4755–4767
- Pina A, Staton M, Zhebentyayeva T, Mockaitis K, Errea P, Abbott A (2010) Studying the molecular mechanisms of graft incompatibility in *Prunus* using 454 sequencing. 5th International Rosaceae genomics conference. November 2010. Stellenbosch (South Africa), O39
- Pozzi C, Vecchiotti A (2009) Peach structural genomics. In: Folta KM, Gardiner SE (eds) *Genetics and genomics of Rosaceae*. Springer, Heidelberg, pp 235–257
- Priest HD, Fox SE, Filichkin SA, Mockler TC (2010) Utility of next-generation sequencing for analysis of horticultural crop transcriptomes. *Acta Hort* 859:283–288
- Rasori A, Ruperti B, Bonghi C, Tonutti P, Ramina A (2002) Characterisation of two putative ethylene receptor genes expressed during peach fruit development and abscission. *J Exp Bot* 53:2333–2339
- Robertson G, Schein J, Chiu R, Corbett R et al (2010) De novo assembly and analysis of RNA-Seq data. *Nat Methods* 7:909–912
- Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-Seq data. *Genome Biol* 11:R25
- Rubio M, Rodríguez-Moreno L, Martínez-Gómez P (2010) Gene expression analysis of resistance to Plum pox virus, “Sharka”, in apricot by transcriptome deep-sequencing (RNA-Seq). 5th International Rosaceae genomics conference. November 2010. Stellenbosch (South Africa), p 38
- Rubio-Cabetas MJ, Amador ML, Pons C, Marti C, Granell A (2010) A microarray analysis revealed and oxidative response genes underlying the differential response to hypoxia of two *Prunus* genotypes. 5th International Rosaceae genomics conference. November 2010. Stellenbosch (South Africa), O42
- Ruperti B, Bonghi C, Tonutti P, Ramina A (1998) Ethylene biosynthesis in peach fruitlet abscission. *Plant Cell Environ* 21:731–737
- Ruperti B, Bonghi C, Rasori A, Ramina A, Tonutti P (2001) Characterization and expression of two members of the peach 1-aminocyclopropane-1-carboxylate oxidase gene family. *Physiol Plant* 111:336–344
- R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>
- Sánchez-Pérez R, Howad W, García-Mas J, Arús P, Martínez-Gómez P, Dicenta F (2010) Molecular markers for kernel bitterness in almond. *Tree Genet Genomes* 6:237–247
- Santos AM, Oliver MJ, Sanchez AM, Payton PR, Gomes JP, Miguel C, Oliveira MM (2009) An integrated strategy to identify key genes in almond adventitious shoot regeneration. *J Exp Bot* 60:4159–4173
- Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–470
- Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE, Graham MA, Cannon SB, May GD, Vance CP, Shoemaker RC (2010) RNA-Seq atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biol* 10:160
- Shulaev V, Korban SS, Sosinski B, Abbott AG, Aldwinckle HS, Folta KM, Iezzoni A, Main D, Arús P, Dandekar AM, Lewers K, Gardiner SE, Potter D, Veilleux E (2008) Multiple models for Rosaceae genomics. *Plant Physiol* 147:985–1003
- Silva C, Garcia Mas J, Sánchez AM, Arús P, Oliveira MM (2005) Looking into flowering time in almond: the candidate gene approach. *Theor Appl Genet* 110:959–968
- Sooknanan R, Pease J, Doyle K (2010) Novel methods for rRNA removal and directional, ligation-free RNA-Seq library preparation. *Nat Methods* 7:I–II
- Sooriyapathirana SS, Khan A, Sebolt AM, Wang DC, Bushakra JM, Lin-Wang K, Allan AC, Gardiner SE, Chagne D, Iezzoni AF

- (2010) QTL analysis and candidate gene mapping for skin and flesh color in sweet cherry fruit (*Prunus avium* L.). *Tree Genet Genomes* 6:821–832
- Soriano JM, Vilanova S, Romero C, Llácer G, Badenes ML (2005) Characterization and mapping of NBS-LRR resistance gene analogs in apricot. *Theor Appl Genet* 110:980–989
- Sosinski B, Verde I, Morgante M, Rokhsar D (2010) The international peach genome initiative. A first draft of the peach genome sequence and its use for genetic diversity analysis in peach. 5th International Rosaceae genomics conference. November 2010. Stellenbosch (South Africa), O46
- Suelves M, Puigdomenech P (1998) Molecular cloning of the cDNA coding for the (R)-(+)-mandelonitrile lyase of *Prunus amygdalus*: temporal and spatial expression patterns in flowers and mature seeds. *Planta* 206:388–393
- Sultan M, Schulz MH, Richard H, Magen M, Scherf M, Borodina T, Soldatov A, Parkhomchuk D (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321:956–960
- Surget-Groba Y, Montoya-Burgos JI (2010) Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genet Res* 20:1432–1440
- Tittarelli A, Santiago M, Morales A, Meisel LA, Silva H (2009) Isolation and functional characterization of cold-regulated promoters, by digitally identifying peach fruit cold-induced genes from a large EST dataset. *BMC Plant Biol* 7:513
- Tong Z, Gao Z, Wang F, Zhou J, Zang Z (2009) Selection of reliable reference genes for gene expression studies in peach using real-time PCR. *BMC Mol Biol* 10:71
- Trainotti L, Zanin D, Casadoro G (2003) A cell wall-oriented genomic approach reveals a new and unexpected complexity of the softening in peaches. *J Exp Bot* 54:1821–1832
- Trainotti L, Bonghi C, Ziliotto F, Zanin D, Rasori A, Casadoro G, Ramina A, Tonutti P (2006) The use of microarray  $\mu$ PEACH1.0 to investigate transcriptome changes during transition from preclimacteric to climacteric phase in peach fruit. *Plant Sci* 170:606–613
- Trainotti L, Tadiello A, Casadoro G (2007) The involvement of auxin in the ripening of climacteric fruits comes of age: the hormone plays a role of its own and has an intense interplay with ethylene in ripening peaches. *J Exp Bot* 58:3299–3308
- van Bakel H, Nislow C, Blencowe BJ, Hughes TR (2010) Most ‘dark matter’ transcripts are associated with known genes. *PLoS Biol* 8:e1000371
- Vecchiotti A, Lazzari B, Ortugno C, Bianchi F, Malinverdi R, Carprear A, Mignani I, Pozzi C (2009) Comparative analysis of expressed sequence tags from tissues in ripening stages of peach. *Tree Genet Genomes* 5:377–391
- Vega-Arreguín JC, Ibarra-Laclette E, Martínez O, Vielle Cazalda P, Herrera-Estrella L, Herrera-Estrella A (2009) Deep sampling of the *Palomero* maize transcriptome by a high throughput strategy of pyrosequencing. *BMC Genomics* 10:299
- Vendramin E, Dettori MT, Giovinazzi J, Mical S, Quarte R, Verde I (2007) A set of EST-SSRs isolated from peach fruit transcriptome and their transportability across *Prunus* species. *Mol Ecol Notes* 7:307–310
- Vivancos A, Güell M, Dohm JC, Serrano L, Himmelbauer H (2010) Strand-specific deep sequencing of the transcriptome. *Genet Res* 20:989–999
- Vizoso P, Meisel LA, Lee A, Tittarelli A, Latorre M, Saba J, Caroca R, Maldonado J, Cambiazo V, Campos-Vargas R, Gonzalez M, Orellana A, Silva H (2009) Comparative EST transcript profiling of peach fruits under different post-harvest conditions reveals candidate genes associated with peach fruit quality. *BMC Genomics* 10:423
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57–63
- Wang L, Li PH, Brutnell TP, Thomas P (2010a) Exploring plant transcriptomes using ultra high-throughput sequencing. *Briefs Funct Gen* 9:118–128
- Wang X, Wang XW, Wang LK, Feng ZX, Zhang XG (2010b) A review on the processing and analysis of next generation RNA-Seq data. *Prog Biochem Biophys* 37:834–846
- Wang LK, Feng ZX, Wang X, Wang XW, Zhang XG (2010c) DEGseq: an R package for identifying differentially expressed genes from RNA-Seq data. *Bioinformatics* 26:136–138
- Weber APM, Weber KL, Carr K, Wilkerson C, Ohlrogge JB (2007) Sampling the *Arabidopsis* transcriptome with massively parallel pyrosequencing. *Plant Physiol* 144:32–42
- Werner T (2010) Next generation sequencing in functional genomics. *Briefs Bioinform* 11:499–511
- Wilhelm BT, Landry JR (2009) RNA-Seq-quantitative measurement of expression through massively parallel RNA-Sequencing. *Methods* 48:249–257
- Wilhelm BT, Marguerat S, Goodhead I, Bähler J (2010) Defining transcribed regions using RNA-Seq. *Nat Protoc* 5:255–266
- Wullschlegel M, Difazio TD (2003) Emerging use of gene expression microarrays in plant physiology. *Comp Funct Gen* 4:216–224
- Yamamoto T, Mochida K, Imai T, Shi IZ, Ogiwara I, Hayashi T (2002) Microsatellite markers in peach [*Prunus persica* (L.) Batsch] derived from an enriched genomic and cDNA libraries. *Mol Ecol Notes* 2:298–302
- Yazaki J, Gregory BD, Ecker JR (2007) Mapping the genome landscape using tiling array technology. *Curr Opin Plant Biol* 10:534–542
- Zenoni S, Ferrarini A, Giacomelli E, Xumerle L, Fasoli M, Malerba G, Bellin D, Pezzotti M, Delle Donne M (2010) Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. *Plant Physiol* 152:1787–1795
- Zhang GJ, Guo GW, Hu XD, Zhang Y, Li QY, Li RQ, Zhuang RH, Lu ZK, He ZQ, Fang XD, Chen L, Tian W, Tao Y, Kristiansen K, Zhang XQ, Li SG, Yang HM, Wang J, Wang J (2010) Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genet Res* 20:646–654
- Zhebentyayeva TN, Swire-Clark G, Georgi LL, Garay L, Jung S, Forrest A, Blackmon B, Horn R, Howad W, Arús P, Main D, Sosinski B, Baird WV, Reighard GL, Abbott AG (2008) A framework physical map for peach, a model Rosaceae species. *Tree Genet Genomes* 4:745–756
- Zhebentyayeva T, Fan S, Olukolu B, Barakat A, Leida C, Badenes ML, Bielenberg D, Reighard G, Okie W, Abbott AG (2010) From genetics to epigenetics in control of chilling requirements and flowering time in peach. 5th International Rosaceae genomics conference. November 2010. Stellenbosch (South Africa), O54
- Ziliotto F, Begheldo M, Rasori A, Bonghi C, Tonutti P (2008) Transcriptome profiling of ripening nectarine (*Prunus persica* L. Batsch) fruit treated with 1-MCP. *J Exp Bot* 59:2781–2791
- Ziosi V, Bonghi C, Bregoli AM, Trainotti L, Biondi S, Sutthiwal S, Kondo S, Costa G, Torrigiani P (2008) Jasmonate-induced transcriptional changes suggest a negative interference with the ripening syndrome in peach fruit. *J Exp Bot* 59:563–573