Applications of Bioinformatics to Plant Biotechnology

Diego F. Gomez-Casati[1]*, María V. Busi[1], Julieta Barchiesi[1], Diego A. Peralta[1], Nicolás Hedin[1] and Vijai Bhadauria[2]

[1]Centro de Estudios Fotosintéticos y Bioquímicos (CEFOBI-CONICET), Universidad Nacional de Rosario, Suipacha, Rosario, Argentina.
[2]Crop Development Centre and Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada.

*Correspondence: gomezcasati@cefobi-conicet.gov.ar

https://doi.org/10.21775/9781910190654.06

Abstract

Bioinformatics encompasses many tools and techniques that today are essential for all areas of research in the biological sciences. New databases with a wealth of information about genomes, proteins, metabolites, and metabolic pathways appear almost daily. Particularly for scientists who carry out research in plant biology, the amount of information has multiplied exponentially due to the large number of databases available for many individual plant species. In this sense, bioinformatics together with next-generation sequencing and 'omics' approaches, can provide tools for plant breeding and the genetic engineering of plants. In addition, these technologies enable a better understanding of the processes and mechanisms that can lead to plants with increased tolerance to different abiotic stress conditions and resistance to pathogen attack, as well as the development of crop varieties with improved nutritional quality of seeds and fruits.

Introduction

In recent years, bioinformatics has emerged as a new, interdisciplinary field of science that can be described as the structuring of biological information to enable the understanding of biological phenomena (Edwards and Batley, 2004; Rhee *et al.*, 2006). Bioinformatics is essential in many areas of biology such as molecular biology, genetics and physiology, and plant biotechnology. Bioinformatics also includes the collection of biological information, storage and management of data, and the development of software for analysis of large data sets.

In order to deal with this deluge of available information, scientists need the ability to collect accurate biological data, store it in databases, and organize it in a systematic way. In addition, computer programs and mathematical tools such as algorithms and statistics are necessary for the analysis of the data.

Recent advances in genomic technologies have led to the accumulation of large amounts of data, resulting in a significant increase in the amount of biological information available for plant research and many other areas of biology, particularly the biomedical sciences. The applications of bioinformatics in plant biotechnology have expanded with the emergence of next-generation sequencing (NGS) techniques and the various omics technologies, providing data integration across the various omics platforms to reveal information about the genome, transcriptome, proteome, metabolome, and metabolic pathways of different plant species (Edwards and Batley, 2004; Rhee *et al.*, 2006).

In this review we cover some aspects of the applications of bioinformatics to problems in plant biotechnology, such as the improvement of fruit quality, the use of bioinformatics tools for plant

breeding, and the improvement of crop species in their responses to various types of stress and pathogen attack. We also describe the databases available for the study of gene expression and metabolites, especially those containing information on different plant species.

Gene expression and other omics resource databases

Transcriptomic analysis comprises the study of changes in gene expression profiles due to changes in the normal condition of a cell or tissue. In addition, transcriptomic studies can be used to assess how these changes in gene expression affect the development, growth, or phenotype of a particular biological system.

Microarrays are currently the most widely used technology for monitoring gene expression in plants. This technique allows the simultaneous determination of transcript abundance for many thousands of genes (Rhee *et al.*, 2006). Microarray data may be directly integrated into functional genomic approaches aimed at both assigning function to identified genes and studying the organization and control of genetic pathways that act concurrently to create a functional organism. The rationale behind this approach is that genes with similar expression patterns may be functionally related and subject to the same or similar genetic control mechanisms. Therefore, a common strategy undertaken in early microarray studies was to analyse data by clustering genes into groups based on their expression profiles scored over multiple experiments (Aharoni *et al.*, 2001; Brown and Botstein, 1999). Recently, RNA sequencing (RNA-seq) has been developed to directly measure mRNA profiles in a biological system using NGS technology (Morin *et al.*, 2008; Weber, 2015). This technique has the advantage in that it can generate relative measures of mRNA and also individual exon abundance while simultaneously profiling the prevalence of both annotated and novel exons and exon-splicing events. In addition, it is possible to identify known single nucleotide polymorphisms (SNPs) as well as novel single-base variants (Morin *et al.*, 2008). Microarray and RNA-seq techniques have enabled not only the identification of genes and the networks involved in different developmental processes, but have also helped to identify the downstream genes involved in such metabolic processes (Agarwal *et al.*, 2014). Fortunately, all gene expression data collected under different metabolic and environmental conditions is deposited in databases that are publicly accessible (Table 6.1). These data provide valuable information that could potentially be used to produce improved plant varieties.

On the other hand, bioinformatics is also being applied to the study and analysis of the metabolome. The omics approach known as metabolomics involves the comprehensive, high-throughput analysis of complex metabolite mixtures that, ideally, allows the identification and quantification of every individual metabolite that are the end products of cellular regulatory processes (Gomez-Casati *et al.*, 2013). Metabolomic studies, in conjunction with genomics or proteomics, can reflect changes in the phenotype and function of a particular tissue or organism. As with gene expression data, there are several databases for plant species (some of which also include useful information for biomedical researchers). Some of these databases also provide users with a powerful tool that allows them to generate an unambiguous means to identify structures. This 'tag' is known as the International Chemical Identifier (InChI) developed by the International Union of Pure and Applied Chemistry (IUPAC) and the National Institute of Standards and Technology (NIST). The InChI identifier is based on the chemical structure, and it makes it easier to perform searches on databases since the user can translate from the InChI to the structure, or back-translate. CheBI (www.ebi.ac.uk/chebi/) is one of the databases that allow users to generate InChI identifiers, and PubChem (http://pubchem.ncbi.nlm.nih.gov/) has a structure drawing tool to generate the InChI identifier. In addition, Fiehn Lab Chemical Translation Service (http://cts.fiehnlab.ucdavis.edu/) is an online tool that has the capability to generate InChI identifiers.

Some of the most well-known of the available metabolomics databases are the Golm Metabolome Database (http://gmd.mpimp-golm.mpg.de/), which contains information about the mass spectra from biologically active metabolites, and the Human Metabolome Database www.hmdb.ca/), which contains information on metabolites found in humans, including biochemical, clinical, and chemical data linking known metabolites to several genes and proteins. The Madison Metabolomics

[23]

Table 6.1 Plant-specific gene expression databases

Plant species	Database name	URL
Arabidopsis	NASC Arrays	http://affymetrix.arabidopsis.info/Arabidopsis
Arabidopsis	Genevestigator	www.genevestigator.ethz.ch/
Arabidopsis	AREX LITE	www.arexdb.org/
Arabidopsis	AraNet	www.inetbio.org/aranet/
Barley	Barleybase	www.barleybase.org
Citrus	CitrusPLEX	www.plexdb.org/plex.php?database=Citrus
Cotton	CottonPLEX	www.plexdb.org/plex.php?database=Cotton
Generic	GEO (Gene Expression Omnibus)	www.ncbi.nlm.nih.gov/projects/geo
Generic	ArrayExpress	www.ebi.ac.uk/arrayexpress
Generic	Plant Co-expression database	http://planex.plantbioinformatics.org/
Generic	Expression Atlas EMBL-EBI	www.ebi.ac.uk/gxa/home
Generic	BAR	http://bar.utoronto.ca/
Generic	PLANET	http://aranet.mpimp-golm.mpg.de/
Generic	PathoPlant	www.pathoplant.de/
Generic	Plant Promoter Database	http://linux1.softberry.com/berry.phtml?topic=plantprom&group=data&subgroup=plantprom
Generic	PLEXdb	www.barleybase.org/plexdb/html/index.php
Grape	GrapePLEX	www.plexdb.org/plex.php?database=Grape
Maize	Maize genomics and genetics database	www.maizegdb.org/
Maize	Zeamage	www.maizearray.org
Maize	MaizePLEX	www.plexdb.org/plex.php?database=Corn
Medicago	MedicagoPLEX	www.plexdb.org/plex.php?database=Medicago
Poplar	PoplarPLEX	www.plexdb.org/plex.php?database=Poplar
Rice	Rice Expression Database	http://tenor.dna.affrc.go.jp/
Rice	NSF Rice Oligo Array Project Database	www.ricearray.org
Rice	Rice expression Profile Database	http://ricexpro.dna.affrc.go.jp/
Rice	RicePLEX	www.plexdb.org/plex.php?database=Rice
Solanaceae	Solanaceae Gene Expression Database	www.tigr.org/tdb/potato
Solanaceae	SolGenomics Network	https://solgenomics.net/gem/experimental_design.pl?id=2
Soybean	Soybean Genome and Microarray Database	https://omictools.com/soybean-genome-and-microarray-database-tool
Soybean	SoyPLEX	www.plexdb.org/plex.php?database=Soybean
Sugarcane	SugarPLEX	www.plexdb.org/plex.php?database=Sugarcane
Tobacco	TOBFAC	http://infogizmo.com/tobfac_tobacco_transcription_factors.html
Tomato	Tomato Functional Genomics Database	http://ted.bti.cornell.edu
Tomato	TomPLEX	www.plexdb.org/plex.php?database=Tomato
Vitis vinifera (grape)	VvGDB	www.plantgdb.org/VvGDB/
Wheat	WheatPLEX	www.plexdb.org/plex.php?database=Wheat

Database (http://mmcd.nmrfam.wisc.edu/) contains information about metabolites determined using NMR and MS. Another integrative cross-species and cross-technique database is Metabolights (www.ebi.ac.uk/metabolights/), which contains information about metabolite structures, spectra, and their biological roles. Metabolomics at Rothamsted (MeT-RO) is a project that contains several resources that can be applied to plant and microbial metabolomics (www.metabolomics.bbsrc.ac.uk/MeT-RO.htm). The Metlin Metabolite Database contains information on about 240,000 metabolites and nearly 72,000 high-resolution MS/MS spectra and tandem MS experiments (http://metlin.scripps.edu/). PRIMe, a platform for RIKEN metabolomics, is a database that integrates genomic and metabolomics data. PRIMe contains information on metabolites obtained from GC-MS, LC-MS, CE-MS and NMR spectroscopy (http://prime.psc.riken.jp/). ReSpect for Phytochemicals (http://spectra.psc.riken.jp/menta.cgi/index) is a collection of more than 9000 literature and in-house MSn spectra records for research on plant metabolomics. MassBank (www.massbank.jp) is a shared public repository of mass spectral data that, as of this writing (September 2016), contained 41,092 spectra records. These data are useful for chemical identification of compounds detected by mass spectrometry. The MS-MS Fragment Viewer (http://webs2.kazusa.or.jp/msmsfragmentviewer/) is a metabolomics database that contains FT-, IT-MS/MS and FT-MS spectral data with predicted structures of ion fragments observed in LC-FT/ICR-MS analysis. Metabolome Express (www.metabolome-express.org) is a public place where GC/MS metabolomics datasets can be processed, interpreted, and shared. It houses both uncurated repositories as well as quality-controlled databases.

Databases specific for individual plant species have also been developed; an example is the Metabolome Tomato Database (MoTo DB) (www.ab.wur.nl/moto/), which is based on metabolites from *Solanum lycopersicum* obtained by LC-MS. Plant Metabolomics (http://plantmetabolomics.vrac.iastate.edu/ver2/) is a metabolomic and functional genomics tool for determining the roles of *Arabidopsis thaliana* genes. Terpmed (www.terpmed.eu/databases.html) is another plant related database that contains information about plant terpenoids, particularly sesquiterpene lactones and phenolic diterpenes important for use as therapeutic drugs.

There are also several resources that integrate metabolic pathways with metabolite data such as MetaCyc (http://metacyc.org/) that contains information on ≈2450 pathways from more than 2780 organisms, and BioCyc (http://biocyc.org) a collection of about 7600 pathway and genome databases. PlantCyc (http://plantcyc.org) is a comprehensive multispecies database that collects biochemical information from many plant pathways. Other related databases are RiceCyc (http://pathway.gramene.org/gramene/ricecyc.shtml), the Sol Genomics Network (http://solgenomics.net/tools/solcyc/index.pl), AraCyc (www.arabidopsis.org/biocyc/), HumanCyc (http://humancyc.org/), Mapman (http://mapman.gabipd.org/web/guest/mapman), ChemSpider (www.chemspider.com/), Reactome (www.reactome.org), the KEGG Pathway database (www.genome.jp/kegg/pathway.html), KappaView (http://kpv.kazusa.or.jp/kpv4/), and KNApSAcK (http://kanaya.naist.jp/KNApSAcK/). These databases are summarized in Table 6.2.

Applications of bioinformatics to enhance fruit development and ripening

The increasing number of NGS technologies and new bioinformatics algorithms developed since the publication of the *Arabidopsis thaliana* genome sequence in 2000 (The Arabidopsis Genome Initiative, 2000), have enabled a deeper understanding of plant biology in both model species and crops (Gapper *et al.*, 2014).

Current NGS technologies enable whole-genome shotgun DNA sequencing, and have become established as the preferred genome sequencing technology, due to longer read lengths, drastically reduced costs, and improved assembly algorithms. These advantages are reflected in the fact that the genomes sequenced in the last 10 years are mostly from crops or non-model plant species, allowing an increase in the number of applied studies in crops and fruits of interest (Feuillet *et al.*, 2011; Gapper *et al.*, 2014; Hamilton and Buell, 2012).

[24]

Table 6.2 Metabolomics resources databases

Name	URL	Information/species
ChEBI	www.ebi.ac.uk/chebi/	Dictionary of small chemical molecules
PubChem	http://pubchem.ncbi.nlm.nih.gov/	General dictionary of chemicals
Fiehn Lab Chemical	http://cts.fiehnlab.ucdavis.edu/	Chemical structures, names, synonyms and database identifiers
Translation service		
Golm metabolome database	http://gmd.mpimp-golm.mpg.de/	GS-MS
Human metabolome database	www.hmdb.ca/	Chemical and biological data of human metabolites
Madison metabolomics	http://mmcd.nmrfam.wisc.edu/	NMR and MS
Database		
Metabolights	www.ebi.ac.uk/metabolights/	Metabolite structures, spectra, function/ cross-species
Metabolomics at Rothamsted	www.metabolomics.bbsrc.ac.uk/ MeT-RO.htm	Plant and microbial metabolites
MeT-RO		
Metlin metabolite database	http://metlin.scripps.edu/	High-resolution MS/MS spectra and tandem MS experiments
PRIMe	http://prime.psc.riken.jp/	Genomic and metabolomics data, NMR spectroscopy, GC-MS, LC-MS and CE-MS
ReSpect for Phytochemicals	http://spectra.psc.riken.jp/menta. cgi/index	MSn spectra data for research on plant metabolomics
MassBank	www.massbank.jp	Database of mass spectral data
MS-MS Fragment Viewer	http://webs2.kazusa.or.jp/ msmsfragmentviewer/	FT-MS, IT- and FT-MS/MS spectral data
MetabolomeExpress	http://metabolome-express.org	GC/MS datasets
Metabolome Tomato Database	www.ab.wur.nl/moto/	Metabolites identified by LC-MS
MoTo DB		
Plantmetabolomics	http://plantmetabolomics.vrac. iastate.edu/ver2/	*Arabidopsis* and other plant species
Terpmed	www.terpmed.eu/databases.html	Plant terpenoids, natural products, secondary metabolites, therapeutic drugs
MetaCyc	http://metacyc.org/	Integration of metabolite data with metabolic pathways/2000 organisms
BioCyc	http://biocyc.org	Collection of pathway/genome database
PlantCyc	http://plantcyc.org	Multispecies plant database
RiceCyc	http://pathway.gramene.org/ gramene/ricecyc.shtml	Metabolic pathways, enzymes, metabolites
Sol Genomics Network	http://solgenomics.net/tools/solcyc/ index.pl	Pathway genome databases/Solanaceae species
AraCyc	www.arabidopsis.org/biocyc/	Metabolic pathways, compounds, *Arabidopsis*
HumanCyc	http://humancyc.org/	Metabolic pathways, genome/human
Mapman	http://mapman.gabipd.org/web/ guest/mapman	Datasets (e.g. gene expression data, metabolic pathways)
ChemSpider	www.chemspider.com/	Chemical structure database
Reactome	www.reactome.org	Free curated and peer reviewed pathway database

Table 6.2 Continued

Name	URL	Information/species
KEGG Pathway database	www.genome.jp/kegg/pathway.html	Pathways, metabolism, genetic information, cellular processes, human diseases
KappaView	http://kpv.kazusa.or.jp/kpv4/	Metabolic pathway database aimed at metabolic regulation
KNApSAcK	http://kanaya.naist.jp/KNApSAcK/	Species–metabolite relationship database

The genome sequences of 27 fruit and nut crops have been published to date; these include grape (Jaillon *et al.*, 2007), papaya (Ming *et al.*, 2008), cucumber (Huang *et al.*, 2009), apple (Velasco *et al.*, 2010), strawberry (Shulaev *et al.*, 2011), tomato (Tomato Genome Consortium, 2012), melon (Garcia-Mas *et al.*, 2012), banana (D'Hont *et al.*, 2012), Chinese plum (Zhang *et al.*, 2012), pear (Wu *et al.*, 2013), watermelon (Guo *et al.*, 2013), peach (Verde *et al.*, 2013), kiwifruit (Huang *et al.*, 2013), orange (Xu *et al.*, 2013a), mandarin (Wang *et al.*, 2015), jujube (Liu *et al.*, 2014), mulberry (He *et al.*, 2013), pineapple (Zhang *et al.*, 2014), blueberry (Gupta *et al.*, 2015), raspberry (VanBuren *et al.*, 2016), mango (Azim *et al.*, 2014), macadamia nut (Nock *et al.*, 2014), cocoa (Motamayor *et al.*, 2013), coffee (Denoeud *et al.*, 2014), hazelnut (Hampson *et al.*, 1996), aubergine (Hirakawa *et al.*, 2014) and pepper (Qin *et al.*, 2014).

[22]

Nevertheless, the additional development of physical and functional annotation and improved genome assembly is required. For instance, the assembly of the tomato genome is the most complete of all fruit species sequenced, but functional annotation of its genes is still incomplete (Gapper *et al.*, 2014).

The new genomics approaches enable the identification of genes responsible for mutant phenotypes in a way that is much faster than ever before. One example is the cloning of a GOLDEN-LIKE2 Myb superfamily transcription factor responsible for the uniform ripening mutant in tomato. This mutation has been bred into many commercial cultivars resulting in more uniform ripening (Nguyen *et al.*, 2014; Powell *et al.*, 2012).

The availability of a collection of mutants with ripening inhibition phenotypes has been essential to clarifying the transcriptional control of fruit ripening (Gapper *et al.*, 2014). Tomato has one of the best characterized mutant collections available (http://tgrc.ucdavis.edu/) (Gapper *et al.*, 2014; Seymour *et al.*, 2013). Similar collections are the European Prunus Database (www.bordeaux.inra.fr/eupeachdb/) and the Citrus Genome Database (www.citrusgenomedb.org/).

Gene expression during fruit ripening is also regulated by epigenetic modifications, which can consist of genome structure, packaging, and non-sequence-based changes such as DNA methylation, acetylation, and histone modifications (Gallusci *et al.*, 2016; Gapper *et al.*, 2014). The entire set of epigenetic modifications, known as the epigenome, can be elucidated using NGS approaches. Recently, using bisulfite sequencing of the tomato genome, Liu and colleagues demonstrated for the first time that active DNA demethylation is an absolute requirement for fruit ripening to occur, and show a direct cause and effect relationship between hypermethylation at specific promoters and repression of gene expression (Liu *et al.*, 2015). However, even in tomato, the extent and role of the epigenetic regulation of fruit ripening is still relatively poorly understood (Gallusci *et al.*, 2016).

Several transcriptomic studies have identified new regulators of fruit development and ripening. Zhu and colleagues, using RNA sequencing and functional analysis, demonstrated the regulatory role of long non-coding RNAs in tomato fruit ripening (Zhu *et al.*, 2015). In another example, Asif and collaborators examined the global changes that occurred during fruit ripening in banana, sequencing cDNA libraries from unripe and ripe banana fruit pulp using the 454-GS platform (Asif *et al.*, 2014).

Frequently used proteomics techniques are now being adapted to high throughput screening methods that allow for the increased use of annotated genome sequences and robust bioinformatics tools.

Mass spectrometry-based methods have gained importance compared with antibody-based techniques owing to their higher specificity, precision and reproducibility, and the capability to quickly analyse numerous peptide transitions in a single test (Findeisen and Neumaier, 2009; Gapper *et al.*, 2014).

Using high-throughput iTRAQ (isobaric tag for relative and absolute quantization of tryptic peptides) and high-resolution mass spectrometry, Li and colleagues performed comparative analyses of the proteome and transcriptome during pear fruit development and identified 35 differentially expressed proteins related to fruit quality, including proteins related to sugar formation, aroma synthesis, and to the formation of lignin (Li *et al.*, 2015). Studies of this type can provide resources for improvement of fruit quality in pear.

Fruits are an extremely rich source of primary and secondary metabolites. Metabolomics studies on the fruit ripening process have been undertaken in several species including peach (Lombardo *et al.*, 2011), melon (Moing *et al.*, 2011), tomato (Oms-Oliu *et al.*, 2011), apple (Rudell *et al.*, 2009), pear (Pedreschi *et al.*, 2009), avocado (Pedreschi *et al.*, 2014) and pepper (Osorio *et al.*, 2012). These studies have provided new insights into conserved and unique metabolic pathways associated with fruit maturation (Gapper *et al.*, 2014).

One interesting recent study was a comparative investigation to identify common and/or species-specific modes of regulation in sugar accumulation. For that purpose, Dai and collaborators designed a process-based mathematical framework to compare soluble sugar accumulation in three fruits: grape, tomato, and peach (Dai *et al.*, 2016). These authors demonstrated that at maturity, grape presented the highest soluble sugar concentrations, followed by peach and tomato. They also showed that the higher soluble sugar concentration in grape compared to tomato is a result of higher sugar importation, although the higher soluble sugar concentration in grape compared to peach is due to lower water dilution. On the other hand, they observed that carbon utilization for synthesis of non-soluble sugar compounds was conserved among the three fruit species. These results increase our understanding of the origins of differences in soluble sugar concentration between fleshy fruits.

Biotechnology and bioinformatics for plant breeding

At present, there exists a continuous challenge in plant breeding to improve crops that produce fodder, fuel, food, or other products through the application of crop genetics (Ray and Satya, 2014). These challenges are also related to the fact that some crop genomes are not yet fully sequenced and annotated, either because these crops have been under-researched, or due to the structural complexity of some very large genomes (Sikhakhane *et al.*, 2016).

The accessibility of whole genome sequences allows for the identification of sequence-based single nucleotide polymorphisms (SNPs) as DNA markers instead of DNA fragment-based polymorphism identification, which has substantially increased the number of informative markers that can be developed (Stapley *et al.*, 2010). In this way, although there are various strategies for plant breeding, the use of genomics-assisted breeding is an effective and economic strategy that is becoming widely used (Sikhakhane *et al.*, 2016). Moreover, the genetics of particular traits in species with large and complex genomes may now be studied in related plants with smaller genomes that share conserved regions through comparative genomics. This will potentially identify genes or quantitative trait loci (QTL) and putative SNP markers to be annotated for genome-wide association mapping. The application of NGS techniques in several crops is revolutionizing and accelerating the pace of plant breeding (Sikhakhane *et al.*, 2016). Fig. 6.1 shows a scheme for plant breeding that incorporates NGS.

Since the introduction of the Sanger sequencing technique nearly 40 years ago (Sanger *et al.*, 1977), several significant advances have been made that address the limitations of this early technology. These advances complement the development of more sophisticated DNA sequencing methodologies that allow *de novo* genome sequencing by rapidly producing huge amounts of sequence information at very low cost. Table 6.3 shows a summary of the advances made in DNA sequencing technology, from dideoxy chain termination sequencing to modern NGS technologies including 454 sequencing (Roche), Illumina sequencing-by-synthesis, SOLiD sequencing by oligonucleotide

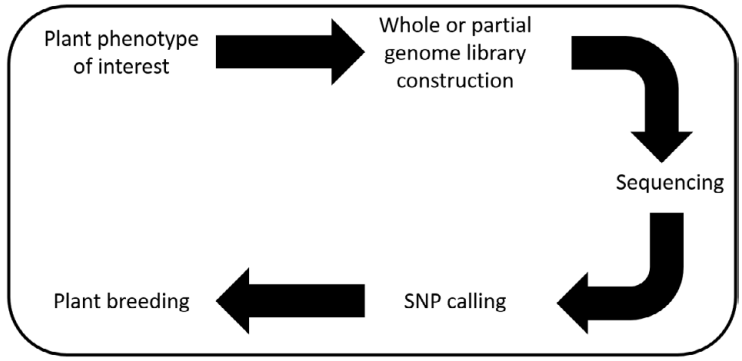

Figure 6.1 Plant breeding assisted by NGS.

[25]

Table 6.3 NGS evolution timeline

Year introduced	Method/Platform	Read length	Performance (data yield)
1977	Sanger sequencing	Up to 1000 bp	Up to 84 kb per 3 hours
2004	Roche/454	Up to 1000 bp	700 Mb per day
2006	Illumina	Up to 300 bp	130 Gb per 30 h
2006	ABI SOLiD	Up to 50 bp	20 Gb per 72 hours
2009	Helicos	35 bp (average)	35 Gb per 8 days
2010	Ion Torrent	35 to 400 bp	2 Gb per 5 hours (average)
	PacBio		

ligation and detection, single molecule sequencing (Helicos Biosciences and Pacific Biosciences), and Ion Torrent sequencing. There are available several modifications for each of these technologies and protocols are continuously being developed to overcome some of the current limitations.

Despite the beneficial prospects of the various NGS platforms, many limitations must be addressed to maximize their potential. These limitations are related to bioinformatics data analysis and the storage and integration of the large number of data generated by these technologies. This constant challenge is also related to the advent of new techniques and protocols in order to make data treatment and analysis easier. Therefore, high-throughput bioinformatics equipment, computational investments, and human resources combined with several NGS technologies will permit the data generated from different sources to be used universally.

Genotyping-by-sequencing (GBS), a widely used SNP discovery method, is feasible for species with large genomes and species with high levels of nucleotide diversity due to advances in genomic technologies that have decreased the cost of high throughput DNA sequencing. GBS is based on NGS technology and takes advantage of reducing genomic complexity to enable high-throughput genotyping of large numbers of samples at a large number of SNP marker loci (Glaubitz *et al.*, 2014). Repetitive regions of the genome may be avoided and lower copy regions targeted with high efficiency by using restriction enzymes (REs). This approach widely simplifies bioinformatic sequence alignment issues in species that possess a high level of genetic diversity (Elshire *et al.*, 2011). This robust, cost-effective, and straightforward method is widely used for genotyping and linkage mapping in many diverse plant species. Although the methods used to analyse raw GBS sequence data are constantly being improved, there are two main commonly used methods at present: genome-wide association studies (GWAS) (Cantor *et al.*, 2010; Visscher *et al.*, 2012) and TASSEL-GBS (Glaubitz *et al.*, 2014). These methods are designed for the efficient processing of raw GBS sequence data

for SNP genotyping. GWAS and TASSEL-GBS procedures largely satisfy the following design criteria: (1) the ability to run on modest computing resources, including desktop or laptop machines without a lot of RAM, (2) scalability from small to large studies (large breeding programs), and (3) applicability in an accelerated breeding context. Wide adoption of these approaches across different methods will speed up the production of useful databases and datasets for effective plant breeding (Sikhakhane *et al.*, 2016).

Bioinformatics for studying stress resistance in plants

The molecular regulatory networks involved in stress resistance and adaptation in plants can be decoded using a combination of omics studies (Unamba *et al.*, 2015). NGS technologies and potent computational pipelines have reduced the cost of whole genome and transcriptomic sequencing, allowing their application to model and non-model plants. This generates a large number of functional genomics resources, which allows us to increase our understanding of the molecular mechanisms underlying plant responses to stress conditions (Xu *et al.*, 2013b).

Plants constantly adjust their transcriptome profile in response to different stresses (Debnath *et al.*, 2011); thus, techniques such as microarrays or RNA-seq can provide a vast amount of information about differential gene expression. One example is the transcriptomic analysis of salt and extreme drought tolerance in *Ipomoea imperati*, a sweet potato relative, that was performed using 454 pyrosequencing by Solis and collaborators. This study identified several genes related to salt tolerance in this species. The information generated in this study could be useful in marker-assisted breeding to improve salt tolerance in sweet potato (Solis *et al.*, 2016).

In another report, Bhardwaj and colleagues performed the first comprehensive transcriptome study of *B. juncea* (the oilseed crop Indian mustard) under conditions of high temperature and drought stress (Bhardwaj *et al.*, 2015). They constructed three transcriptome libraries, sequenced them on the Illumina GA IIx platform, and assembled the high quality reads obtained using the SOAP *de novo* assembler. Bhardwaj *et al.* identified a subset of about 19,000 transcripts that are differentially regulated by either high temperature and/or drought stress. Most of the up-regulated transcription factor genes belonged to heat shock factors and dehydration-responsive element-binding gene families. This information will enhance the understanding of the molecular mechanisms behind the plant responses to abiotic stresses, and will enable efficient strategies to improve tolerance to high temperature and drought stress.

Considerable progress has been made towards engineering transgenic plants that express abiotic stress tolerance. Analyses of the large quantity of abiotic stress-related transcriptomic data generated in several plant species has revealed the importance of osmoprotectants in stress resistance. These osmoprotectants help protect plants against the damaging effects of osmotic and ionic stress. For example, several drought- and salt-resistant transgenic potato plants developed using this strategy have been reported. The tubers were successfully engineered to express an osmotin-like protein (Evers *et al.*, 1999), glyceraldehyde-3-phosphate dehydrogenase (Jeong *et al.*, 2001), and nucleoside diphosphate kinase 2 (Tang *et al.*, 2008).

The many efforts to improve abiotic stress tolerance in plants have resulted in important achievements (Khan *et al.*, 2015). However, due to the complexity of stress tolerance, the success of these strategies has been limited because most of the resulting transgenic plants have been tested only under controlled laboratory conditions. In such conditions, transgenic plants are evaluated at the early growth stage and are exposed to the stress condition for a short period of time. Short exposure experiments at early growth stages performed in laboratories may not predict the response of the plant under true field conditions. In addition, transgenic plants may be exposed to multiple stresses in the field, which may suppress the protective effects of the introduced transgene.

Future investigations into improving plant stress tolerance should put more emphasis on combining different approaches, such as multigenic strategies to simultaneously incorporate more than one gene in transgenic plants. For example, the genes involved in osmoprotectant synthesis could be co-expressed with other stress resistance-related genes such as ion transporters and transcription factors (Khan *et al.*, 2015).

Bioinformatics approaches to study resistance to plant pathogens

Plants have evolved a diverse range of strategies to counteract pathogen attacks that involve the modification of gene expression, activation of several metabolic pathways, and posttranslational modification of proteins (Gomez-Casati *et al.*, 2016).

The new DNA sequencing technologies, combined with sophisticated bioinformatics, are having a major influence on research in the field of plant pathology. These technologies are gaining importance in studies of the genomics, metagenomics, proteomics, metabolomics, and transcriptomics of both the host plant and the pathogen, and are also extremely useful in epidemiology and diagnostics, especially with respect to viruses (Studholme *et al.*, 2011).

Whole genome sequences are available for numerous plants and a diverse group of microbial phytopathogens that includes bacteria, viruses, fungi, and oomycetes. Many of these microbial genome sequences were obtained using the traditional Sanger sequencing method. The first plant pathogen genome sequenced using second-generation sequencing technologies was *P. syringae* pathovar *oryzae*, a bacterium that causes halo blight on rice (Reinhardt *et al.*, 2009; Studholme *et al.*, 2011). Reinhardt and colleagues sequenced the *P. syringae* pv. *oryzae* genome by combining data generated from the Roche 454 and Illumina sequencing platforms and using a new bioinformatics pipeline based on their VCAKEv1.5 assembly algorithm (Jeck *et al.*, 2007). Since that time, several other phytopathogen genomes have been sequenced (Harrison *et al.*, 2016; Studholme *et al.*, 2011; Wasukira *et al.*, 2014). These draft genome sequences are expected to contain errors and be incomplete. However, they can provide valuable information about the molecular basis for infection of plant hosts, including he sequences of potential novel virulence factors, like T3SS effectors (Studholme *et al.*, 2009).

For disease diagnostics, NGS allows the direct identification of bacteria or viruses in an unbiased way without requiring previous knowledge of the pathogen DNA/RNA sequence or antibodies (Quan *et al.*, 2008). Finding a pathogen nucleic acid sequence within host tissue does not necessarily prove that the pathogen is causing the plant disease symptoms (Studholme *et al.*, 2011). However, several published cases present concrete evidence for a relationship between the presence of disease and the pathogen found. For example, Adams and colleagues found that the presence of *Carrot yellow leaf virus* (CYLV) is strongly related to the occurrence of internal necrosis of carrots by using a metagenomic approach, which consisted of constructing an RNA-seq library and sequencing it on the Illumina Miseq instrument (Adams *et al.*, 2014).

At present, most of plant RNA viruses identified by NGS were obtained from either total nucleic acid or total double-stranded RNA extracted from infected plant tissue (Barba *et al.*, 2014). In another strategy, host nucleic acid was eliminated by hybridization to nucleic acid isolated from healthy plants in order to enrich for virus sequences prior to sequencing. Plant viruses can also be identified indirectly by detecting specific RNA molecules, called short interfering RNAs (siRNA), that the host plant generates in response to infection by RNA viruses (Mlotshwa *et al.*, 2008). NGS of siRNAs is a good approach to identify viruses infecting plants, even when they are present at very low titres and in symptomless infections.

Microbial metagenomic studies can be performed on environmental samples or individual plants (Dutta *et al.*, 2014; Roossinck *et al.*, 2015). In general, most viruses identified in wild plants cannot be associated with pathologies, although some of them can be pathogenic when infecting domesticated plants. Thus, the use of metagenomics data provides advance knowledge of viruses that could cause threats if they jump from wild species into domestic hosts.

Using high-throughput transcriptomic technologies, such as microarrays or RNA-Seq, makes it possible to obtain gene expression profiles from the host plants and to detect gene expression changes in the host-associated pathogen. On the other hand, the use of different proteomics approaches allows for the identification of many resistance proteins that plants express in the presence of specific pathogen effectors that trigger its defence mechanisms. Using this methodology, several antimicrobial proteins expressed during phytopathogenic interactions have been identified, mainly by high resolution of two-dimensional polyacrylamide gel electrophoresis coupled with mass spectrometry (Gomez-Casati *et al.*, 2016; Mehta *et al.*, 2008).

All of the omics technologies discussed here can provide a wealth of information about the pathogen life cycle, and enable the discovery of novel virulence factors in pathogens as well as their host targets (Gomez-Casati *et al.*, 2016). Both approaches are important for developing strategies to improve disease resistance in plants.

Traditionally, the control of pathogen invasion and spread has been achieved through breeding programmes that focus on increasing host resistance to particular pathogens, or by the application of chemical pesticides and fungicides. A relatively straightforward approach to generate pathogen resistant plants is through introduction of genes from other organisms to develop new cellular metabolic pathways.

Several genes involved in specific immune responses to different phytopathogens were identified using RNA-seq technology; these include genes that encode a tyrosine protein kinase (Epk1) (Pombo *et al.*, 2014), a cell wall-associated kinase (SlWAK1) (Rosli *et al.*, 2013), and a nucleotide binding leucine-rich repeat receptor (NB-LRR) (Bernoux *et al.*, 2011).

Whereas some attempts to engineer disease resistance in economically important crop plants have failed, several have been successful. Cao and colleagues reported transgenic maize lines that constitutively express a mutant *E. coli* dsRNA-specific endoribonuclease gene and show increased resistance to the dsRNA virus that causes Maize rough dwarf disease (Cao *et al.*, 2013). Mysore and collaborators showed that the overexpression of a serine/threonine kinase gene (Pto) in tomato confers protection against *P. syringae* (Mysore *et al.*, 2003). In addition, the overexpression of NPR1, a gene involved in systemic acquired resistance in plants, resulted in increased disease resistance in apple, *Malus domestica*, to two important fungal pathogens, *Venturia inaequalis* and *Gymnosporangium juniperi-virginianae* (Gomez-Casati *et al.*, 2016; Malnoy *et al.*, 2007).

On the other side, the most usual way of obtaining virus resistant plants is by expressing in them a viral gene encoding the coat protein. So, the plant produces this viral protein before the virus attacks, and when the virus tries to infect the plant, the host gene silencing pathway is already activated and degrades the virus RNA early in the infection process. All genetically modified virus resistant plants existing on the market (e.g. squash and papayas) have this resistance mechanism (Ratcliff *et al.*, 1999; Voinnet, 2001).

Conclusions

The application of bioinformatics to the study and characterization of biological phenomena represents a fundamental shift in the way that scientists study living organisms. As mentioned above, the many genomic, transcriptomic, proteomic, and metabolite databases available have the potential to accelerate the rate of functional discoveries in plant biology. Although these databases provide access to gene information, gene expression data, and metabolite profiles, as well as the experimental conditions used to generate them, additional measures are needed to complement genome annotation efforts effectively and to increase discovery of gene function and regulation. One possible course of action would be the generation of reference expression datasets for plant cells and/or tissues from specific stages of plant development. In addition, it would be highly desirable to develop databases which could contain increasingly integrated information about genomic, transcriptomic, metabolomic, as well as proteomic and phenomic data for each plant species.

Acknowledgements

This work was supported by grants from ANPCyT (PICT 2013-2188 and 2014-2184).

References

Adams, I.P., Skelton, A., Macarthur, R., Hodges, T., Hinds, H., Flint, L., Nath, P.D., Boonham, N., and Fox, A. (2014). Carrot yellow leaf virus is associated with carrot internal necrosis. PLOS ONE9, e109125. http://dx.doi.org/10.1371/journal.pone.0109125

Agarwal, P., Parida, S.K., Mahto, A., Das, S., Mathew, I.E., Malik, N., and Tyagi, A.K. (2014). Expanding frontiers in plant transcriptomics in aid of functional genomics and molecular breeding. Biotechnol. J. *9*, 1480–1492. http://dx.doi.org/10.1002/biot.201400063

Aharoni, A., De Vos, C.H., Wein, M., Sun, Z., Greco, R., Kroon, A., Mol, J.N., and O'Connell, A.P. (2001). The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. Plant J. *28*, 319–332. http://dx.doi.org/10.1046/j.1365-313X.2001.01154.x

Asif, M.H., Lakhwani, D., Pathak, S., Gupta, P., Bag, S.K., Nath, P., and Trivedi, P.K. (2014). Transcriptome analysis of ripe and unripe fruit tissue of banana identifies major metabolic networks involved in fruit

ripening process. BMC Plant Biol. *14*, 316. http://dx.doi.org/10.1186/s12870-014-0316-1

Azim, M.K., Khan, I.A., and Zhang, Y. (2014). Characterization of mango (Mangifera indica L.) transcriptome and chloroplast genome. Plant Mol. Biol.*85*, 193–208. http://dx.doi.org/10.1007/s11103-014-0179-8

Barba, M., Czosnek, H., and Hadidi, A. (2014). Historical perspective, development and applications of next-generation sequencing in plant virology. Viruses *6*, 106–136. http://dx.doi.org/10.1016/j.pbi.2011.05.005

Bernoux, M., Ellis, J.G., and Dodds, P.N. (2011). New insights in plant immunity signaling activation. Curr. Opin. Plant Biol. *14*, 512–518. http://dx.doi.org/10.1016/j.pbi.2011.05.005

Bhardwaj, A.R., Joshi, G., Kukreja, B., Malik, V., Arora, P., Pandey, R., Shukla, R.N., Bankar, K.G., Katiyar-Agarwal, S., Goel, S., *etal.* (2015). Global insights into high temperature and drought stress regulated genes by RNA-Seq in economically important oilseed crop *Brassica juncea.* BMC Plant Biol. *15*, 9. http://dx.doi.org/

Brown, P.O., and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. Nat. Genet.*21* (Suppl. 1),33–37. http://dx.doi.org/10.1038/4462

Cantor, R.M., Lange, K., and Sinsheimer, J.S. (2010). Prioritizing GWAS results: A review of statistical methods and recommendations for their application. Am. J. Hum. Genet. *86*, 6–22. http://dx.doi.org/10.1016/j.ajhg.2009.11.017

Cao, X., Lu, Y., Di, D., Zhang, Z., Liu, H., Tian, L., Zhang, A., Zhang, Y., Shi, L., Guo, B., *et al.* (2013). Enhanced virus resistance in transgenic maize expressing a dsRNA-specific endoribonuclease gene from *E. coli*. PLOS ONE *8*, e60829. http://dx.doi.org/10.1371/journal.pone.0060829

D'Hont, A., Denoeud, F., Aury, J.M., Baurens, F.C., Carreel, F., Garsmeur, O., Noel, B., Bocs, S., Droc, G., Rouard, M., *et al.* (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature *488*, 213–217. http://dx.doi.org/10.1038/nature11241

Dai, Z., Wu, H., Baldazzi, V., van Leeuwen, C., Bertin, N., Gautier, H., Wu, B., Duchêne, E., Gomès, E., Delrot, S., *et al.* (2016). Inter-species comparative analysis of components of soluble sugar concentration in fleshy fruits. Front. Plant Sci. *7*, 649. http://dx.doi.org/10.3389/fpls.2016.00649

Debnath, M., Pandey, M., and Bisen, P.S. (2011). An omics approach to understand the plant abiotic stress. OMICS *15*, 739–762. http://dx.doi.org/10.1089/omi.2010.0146

Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., Zheng, C., Alberti, A., Anthony, F., Aprea, G., *etal.* (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science *345*, 1181–1184. http://dx.doi.org/10.1126/science.1255274

Dutta, M., Sokhandan Bashir, N., Palmer, M.W., and Melcher, U. (2014). Genomic characterization of Ambrosia asymptomatic virus 1 and evidence of other Tymovirales members in the Oklahoma tallgrass prairie revealed by sequence analysis. Arch. Virol. *159*, 1755–1764. http://dx.doi.org/10.1016/j.tibtech.2004.03.002

Edwards, D., and Batley, J. (2004). Plant bioinformatics: from genome to phenome. Trends Biotechnol. *22*, 232–237. http://dx.doi.org/10.1016/j.tibtech.2004.03.002

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLOS ONE*6*, e19379. http://dx.doi.org/10.1371/journal.pone.0019379

Evers, D., Overney, S., Simon, P., Greppin, H., and Hausman, J.F. (1999). Salt tolerance of *Solanum tuberosum* L. overexpressing an heterologous osmotin-like protein. Biol. Plant. *42*, 105–112. http://dx.doi.org/10.1023/a:1002131812340

Feuillet, C., Leach, J.E., Rogers, J., Schnable, P.S., and Eversole, K. (2011). Crop genome sequencing: lessons and rationales. Trends Plant Sci.*16*, 77–88. http://dx.doi.org/10.1016/j.tplants.2010.10.005

Findeis en, P., and Neumaier, M. (2009). Mass spectrometry-based clinical proteomics profiling: current status and future directions. Expert. Rev. Proteomics*6*, 457–459. http://dx.doi.org/10.1586/epr.09.67

Gallusci, P., Hodgman, C., Teyssier, E., and Seymour, G.B. (2016). DNA Methylation and Chromatin Regulation during Fleshy Fruit Development and Ripening. Front. Plant Sci.*7*, 807. http://dx.doi.org/10.3389/fpls.2016.00807

Gapper, N.E., Giovannoni, J.J., and Watkins, C.B. (2014). Understanding development and ripening of fruit crops in an 'omics' era. Hortic. Res.*1*, 14034. http://dx.doi.org/10.1038/hortres.2014.34

Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., González, V.M., Hénaff, E., Câmara, F., Cozzuto, L., Lowy, E., *et al.* (2012). The genome of melon (Cucumis melo L.). Proc. Natl. Acad. Sci. U.S.A.*109*, 11872–11877. http://dx.doi.org/10.1073/pnas.1205415109

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., and Buckler, E.S. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLOS ONE*9*, e90346. http://dx.doi.org/10.1371/journal.pone.0090346

Gomez-Casati, D.F., Pagani, M.A., Busi, M.V., and Bhadauria, V. (2016). Omics approaches for the engineering of pathogen resistant plants. Curr. Issues. Mol. Biol.*19*, 89–98.

Gomez-Casati, D.F., Zanor, M.I., and Busi, M.V. (2013). Metabolomics in plants and humans: applications in the prevention and diagnosis of diseases. Biomed Res. Int. *2013*, 792527. http://dx.doi.org/10.1155/2013/792527

Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W.J., Zhang, H., Zheng, Y., Mao, L., Ren, Y., Wang, Z., *et al.* (2013). The draft genome of watermelon (Citrullus lanatus) and resequencing of 20 diverse accessions. Nat. Genet. *45*, 51–58. http://dx.doi.org/10.1038/ng.2470

Gupta, V., Estrada, A.D., Blakley, I., Reid, R., Patel, K., Meyer, M.D., Andersen, S.U., Brown, A.F., Lila, M.A., and Loraine, A.E. (2015). RNA-Seq analysis and annotation of a draft blueberry genome assembly identifies candidate genes involved in fruit ripening, biosynthesis of bioactive compounds, and stage-specific

alternative splicing. Gigascience *4*, 5. http://dx.doi.org/10.1186/s13742-015-0046-9

Hamilton, J.P., and Buell, C.R. (2012). Advances in plant genome sequencing. Plant J. *70*, 177–190. http://dx.doi.org/10.1111/j.1365-313X.2012.04894.x

Hampson, C.R., Coleman, G.D., and Azarenko, A.N. (1996). Does the genome of *Corylus avellana* L. contain sequences homologous to the self-incompatibility gene of Brassica? Theor. Appl. Genet. *93*, 759–764. http://dx.doi.org/10.1007/BF00224073

Harrison, J., Grant, M.R., and Studholme, D.J. (2016). Draft Genome Sequences of Two Strains of Xanthomonas arboricola pv. celebensis Isolated from Banana Plants. Genome Announc. *4*, e01705–15. http://dx.doi.org/10.1128/genomeA.01705-15

He, N., Zhang, C., Qi, X., Zhao, S., Tao, Y., Yang, G., Lee, T.H., Wang, X., Cai, Q., Li, D., *et al.* (2013). Draft genome sequence of the mulberry tree Morus notabilis. Nat. Commun. *4*, 2445. http://dx.doi.org/10.1038/ncomms3445

Hirakawa, H., Shirasawa, K., Miyatake, K., Nunome, T., Negoro, S., Ohyama, A., Yamaguchi, H., Sato, S., Isobe, S., Tabata, S., *etal.* (2014). Draft genome sequence of eggplant (*Solanum melongena* L.): the representative solanum species indigenous to the old world. DNA Res. *21*, 649–660. http://dx.doi.org/10.1093/dnares/dsu027

Huang, S., Ding, J., Deng, D., Tang, W., Sun, H., Liu, D., Zhang, L., Niu, X., Zhang, X., Meng, M., *et al.* (2013). Draft genome of the kiwifruit Actinidia chinensis. Nat. Commun. *4*, 2640. http://dx.doi.org/10.1038/ncomms3640

Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., Lucas, W.J., Wang, X., Xie, B., Ni, P., *etal.* (2009). The genome of the cucumber, Cucumis sativus L. Nat. Genet. *41*, 1275–1281. http://dx.doi.org/10.1038/ng.475

Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., *et al.* (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature *449*, 463–467.

Jeck, W.R., Reinhardt, J.A., Baltrus, D.A., Hickenbotham, M.T., Magrini, V., Mardis, E.R., Dangl, J.L., and Jones, C.D. (2007). Extending assembly of short DNA sequences to handle error. Bioinformatics *23*, 2942–2944.

Jeong, M.J., Park, S.C., and Byun, M.O. (2001). Improvement of salt tolerance in transgenic potato plants by glyceraldehyde-3 phosphate dehydrogenase gene transfer. Mol. Cells *12*, 185–189.

Khan, M.S., Ahmad, D., and Khan, M.A. (2015). Utilization of genes encoding osmoprotectants in transgenic plants for enhanced abiotic stress tolerance. Electronic J. Biotechnol. *18*, 257–266. http://dx.doi.org/10.1016/j.ejbt.2015.04.002

Li, J.M., Huang, X.S., Li, L.T., Zheng, D.M., Xue, C., Zhang, S.L., and Wu, J. (2015). Proteome analysis of pear reveals key genes associated with fruit development and quality. Planta *241*, 1363–1379. http://dx.doi.org/10.1007/s00425-015-2263-y

Liu, M.J., Zhao, J., Cai, Q.L., Liu, G.C., Wang, J.R., Zhao, Z.H., Liu, P., Dai, L., Yan, G., Wang, W.J., *et al.* (2014). The complex jujube genome provides insights into fruit tree biology. Nat. Commun. *5*, 5315. http://dx.doi.org/10.1038/ncomms6315

Liu, R., How-Kit, A., Stammitti, L., Teyssier, E., Rolin, D., Mortain-Bertrand, A., Halle, S., Liu, M., Kong, J., Wu, C., *etal.* (2015). A DEMETER-like DNA demethylase governs tomato fruit ripening. Proc. Natl. Acad. Sci. U.S.A. *112*, 10804–10809. http://dx.doi.org/10.1038/ncomms6315

Lombardo, V.A., Osorio, S., Borsani, J., Lauxmann, M.A., Bustamante, C.A., Budde, C.O., Andreo, C.S., Lara, M.V., Fernie, A.R., and Drincovich, M.F. (2011). Metabolic profiling during peach fruit development and ripening reveals the metabolic networks that underpin each developmental stage. Plant Physiol. *157*, 1696–1710. http://dx.doi.org/10.1104/pp.111.186064

Malnoy, M., Jin, Q., Borejsza-Wysocka, E.E., He, S.Y., and Aldwinckle, H.S. (2007). Overexpression of the apple MpNPR1 gene confers increased disease resistance in Malus x domestica. Mol. Plant Microbe Interact.*20*, 1568–1580. http://dx.doi.org/10.1094/MPMI-20-12-1568

Mehta, A., Brasileiro, A.C., Souza, D.S., Romano, E., Campos, M.A., Grossi-de-Sá, M.F., Silva, M.S., Franco, O.L., Fragoso, R.R., Bevitori, R., *et al.* (2008). Plant-pathogen interactions: what is proteomics telling us? FEBS J.*275*, 3731–3746. http://dx.doi.org/10.1111/j.1742-4658.2008.06528.x

Ming, R., Hou, S., Feng, Y., Yu, Q., Dionne-Laporte, A., Saw, J.H., Senin, P., Wang, W., Ly, B.V., Lewis, K.L., *et al.* (2008). The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature *452*, 991–996. http://dx.doi.org/10.1038/nature06856

Mlotshwa, S., Pruss, G.J., and Vance, V. (2008). Small RNAs in viral infection and host defense. Trends Plant Sci. *13*, 375–382. http://dx.doi.org/10.1016/j.tplants.2008.04.009

Moing, A., Aharoni, A., Biais, B., Rogachev, I., Meir, S., Brodsky, L., Allwood, J.W., Erban, A., Dunn, W.B., Kay, L., *etal.* (2011). Extensive metabolic cross-talk in melon fruit revealed by spatial and developmental combinatorial metabolomics. New Phytol. *190*, 683–696. http://dx.doi.org/10.1111/j.1469-8137.2010.03626.x

Morin, R., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T., McDonald, H., Varhol, R., Jones, S., and Marra, M. (2008). Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. BioTechniques *45*, 81–94. http://dx.doi.org/10.2144/000112900

Motamayor, J.C., Mockaitis, K., Schmutz, J., Haiminen, N., Livingstone, D., Cornejo, O., Findley, S.D., Zheng, P., Utro, F., Royaert, S., *et al.* (2013). The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. Genome Biol. *14*, r53. http://dx.doi.org/10.1186/gb-2013-14-6-r53

Mysore, K.S., D'Ascenzo, M.D., He, X., and Martin, G.B. (2003). Overexpression of the disease resistance gene Pto in tomato induces gene expression changes similar to immune responses in human and fruitfly. Plant Physiol. *132*, 1901–1912. http://dx.doi.org/10.1104/pp.103.022731

Nguyen, C.V., Vrebalov, J.T., Gapper, N.E., Zheng, Y., Zhong, S., Fei, Z., and Giovannoni, J.J. (2014). Tomato

GOLDEN2-LIKE transcription factors reveal molecular gradients that function during fruit development and ripening. Plant Cell *26*, 585–601. http://dx.doi.org/10.1105/tpc.113.118794

Nock, C.J., Baten, A., and King, G.J. (2014). Complete chloroplast genome of Macadamia integrifolia confirms the position of the Gondwanan early-diverging eudicot family Proteaceae. BMC Genomics *15 Suppl 9*, S13. http://dx.doi.org/10.1186/1471-2164-15-S9-S13

Oms-Oliu, G., Hertog, M.L.A.T.M., Van de Poel, B., Ampofo-Asiama, J., Geeraerd, A.H., and Nicolaï, B.M. (2011). Metabolic characterization of tomato fruit during preharvest development, ripening, and postharvest shelf-life. Postharv Biol. Technol. *62*, 7–16. http://dx.doi.org/10.1016/j.postharvbio.2011.04.010

Osorio, S., Alba, R., Nikoloski, Z., Kochevenko, A., Fernie, A.R., and Giovannoni, J.J. (2012). Integrative comparative analyses of transcript and metabolite profiles from pepper and tomato ripening and development stages uncovers species-specific patterns of network regulatory behavior. Plant Physiol. *159*, 1713–1729. http://dx.doi.org/10.1104/pp.112.199711

Pedreschi, R., Hertog, M., Robben, J., Lilley, K.S., Karp, N.A., Baggerman, G., Vanderleyden, J., and Nicolaï, B. (2009). Gel-based proteomics approach to the study of metabolic changes in pear tissue during storage. J. Agric. Food Chem. *57*, 6997–7004. http://dx.doi.org/10.1021/jf901432h

Pedreschi, R., Muñoz, P., Robledo, P., Becerra, C., Defilippi, B.G., van Eekelen, H., Mumm, R., Westra, E., and de Vos, R.C.H. (2014). Metabolomics analysis of postharvest ripening heterogeneity of 'Hass' avocadoes. Postharv. Biol. Technol. *92*, 172–179. http://dx.doi.org/10.1021/jf901432h

Pombo, M.A., Zheng, Y., Fernandez-Pozo, N., Dunham, D.M., Fei, Z., and Martin, G.B. (2014). Transcriptomic analysis reveals tomato genes whose expression is induced specifically during effector-triggered immunity and identifies the Epk1 protein kinase which is required for the host response to three bacterial effector proteins. Genome Biol. *15*, 492. http://dx.doi.org/10.1186/s13059-014-0492-1

Powell, A.L., Nguyen, C.V., Hill, T., Cheng, K.L., Figueroa-Balderas, R., Aktas, H., Ashrafi, H., Pons, C., Fernandez-Munoz, R., Vicente, A., *etal.* (2012). Uniform ripening encodes a Golden 2-like transcription factor regulating tomato fruit chloroplast development. Science *336*, 1711–1715. http://dx.doi.org/10.1126/science.1222218

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., Cheng, J., Zhao, S., Xu, M., Luo, Y., *etal.* (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. Proc. Natl. Acad. Sci. U.S.A. *111*, 5135–5140. http://dx.doi.org/10.1073/pnas.1400975111

Quan, P.L., Briese, T., Palacios, G., and Ian Lipkin, W. (2008). Rapid sequence-based diagnosis of viral infection. Antiviral Res. *79*, 1–5. http://dx.doi.org/10.1016/j.antiviral.2008.02.002

Ratcliff, F.G., MacFarlane, S.A., and Baulcombe, D.C. (1999). Gene silencing without DNA. rna-mediated cross-protection between viruses Plant. Cell. *11*, 1207–1216.

Ray, S., and Satya, P. (2014). Next generation sequencing technologies for next generation plant breeding. Front. Plant Sci. *5*, 367. http://dx.doi.org/10.3389/fpls.2014.00367

Reinhardt, J.A., Baltrus, D.A., Nishimura, M.T., Jeck, W.R., Jones, C.D., and Dangl, J.L. (2009). De novo assembly using low-coverage short read sequence data from the rice pathogen *Pseudomonas syringae* pv. *oryzae*. Genome Res. *19*, 294–305. http://dx.doi.org/10.1101/gr.083311.108

Rhee, S.Y., Dickerson, J., and Xu, D. (2006). Bioinformatics and its applications in plant biology. Annu. Rev. Plant Biol. *57*, 335–360. http://dx.doi.org/10.1146/annurev.arplant.56.032604.144103

Roossinck, M.J., Martin, D.P., and Roumagnac, P. (2015). Plant virus metagenomics: advances in virus discovery. Phytopathology *105*, 716–727. http://dx.doi.org/10.1094/PHYTO-12-14-0356-RVW

Rosli, H.G., Zheng, Y., Pombo, M.A., Zhong, S., Bombarely, A., Fei, Z., Collmer, A., and Martin, G.B. (2013). Transcriptomics-based screen for genes induced by flagellin and repressed by pathogen effectors identifies a cell wall-associated kinase involved in plant immunity. Genome Biol. *14*, R139. http://dx.doi.org/10.1186/gb-2013-14-12-r139

Rudell, D.R., Mattheis, J.P., and Hertog, M.L. (2009). Metabolomic change precedes apple superficial scald symptoms. J. Agric. Food Chem. *57*, 8459–8466. http://dx.doi.org/10.1021/jf901571g

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. *74*, 5463–5467.

Seymour, G.B., Østergaard, L., Chapman, N.H., Knapp, S., and Martin, C. (2013). Fruit development and ripening. Annu. Rev. Plant Biol. *64*, 219–241. http://dx.doi.org/10.1146/annurev-arplant-050312-120057

Shulaev, V., Sargent, D.J., Crowhurst, R.N., Mockler, T.C., Folkerts, O., Delcher, A.L., Jaiswal, P., Mockaitis, K., Liston, A., Mane, S.P., *et al.* (2011). The genome of woodland strawberry (*Fragaria vesca*). Nat. Genet. *43*, 109–116. http://dx.doi.org/10.1038/ng.740

Solis, J., Baisakh, N., Brandt, S.R., Villordon, A., and La Bonte, D. (2016). Transcriptome profiling of beach morning glory (*Ipomoea imperati*) under salinity and its comparative analysis with sweetpotato. PLOS ONE *11*, e0147398. http://dx.doi.org/10.1371/journal.pone.0147398

Stapley, J., Reger, J., Feulner, P.G., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P., and Slate, J. (2010). Adaptation genomics: the next generation. Trends Ecol. Evol.*25*, 705–712. http://dx.doi.org/10.1016/j.tree.2010.09.002

Studholme, D.J., Glover, R.H., and Boonham, N. (2011). Application of high-throughput DNA sequencing in phytopathology. Annu. Rev. Phytopathol. *49*, 87–105. http://dx.doi.org/10.1146/annurev-phyto-072910-095408

Studholme, D.J., Ibanez, S.G., MacLean, D., Dangl, J.L., Chang, J.H., and Rathjen, J.P. (2009). A draft genome sequence and functional screen reveals the repertoire of type III secreted proteins of *Pseudomonas syringae* pathovar *tabaci* 11528. BMC Genomics *10*, 395. http://dx.doi.org/10.1146/annurev-phyto-072910-095408

Tang, L., Kim, M.D., Yang, K.S., Kwon, S.Y., Kim, S.H., Kim, J.S., Yun, D.J., Kwak, S.S., and Lee, H.S. (2008). Enhanced tolerance of transgenic potato plants overexpressing nucleoside diphosphate kinase 2 against multiple environmental stresses. Transgenic Res. *17*, 705–715. http://dx.doi.org/10.1007/s11248-007-9155-2

Sikhakhane, T.N., Figlan, S., Mwadzingeni, L., Ortiz, R., and Tsilo, T.J. (2016). Integration of Next-generation Sequencing Technologies with Comparative Genomics in Cereals. In Agricultural and Biological Sciences 'Plant Genomics', InTech, ed. http://dx.doi.org/10.5772/61763

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature *408*, 796–815. http://dx.doi.org/10.1038/35048692

Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature *485*, 635–641. http://dx.doi.org/10.1038/nature11119

Unamba, C.I., Nag, A., and Sharma, R.K. (2015). Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants. Front. Plant Sci. *6*, 1074. http://dx.doi.org/10.3389/fpls.2015.01074

VanBuren, R., Bryant, D., Bushakra, J.M., Vining, K.J., Edger, P.P., Rowley, E.R., Priest, H.D., Michael, T.P., Lyons, E., Filichkin, S.A., *et al.* (2016). The genome of black raspberry (Rubus occidentalis). Plant J. *87*, 535–547. http://dx.doi.org/10.1111/tpj.13215

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S.K., Troggio, M., Pruss, D., *et al.* (2010). The genome of the domesticated apple (Malus × domestica Borkh.). Nat. Genet.*42*, 833–839. http://dx.doi.org/10.1038/ng.654

Verde, I., Abbott, A.G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., Zhebentyayeva, T., Dettori, M.T., Grimwood, J., Cattonaro, F., *etal.* (2013). The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nature Genet. *45*, 487–494. http://dx.doi.org/10.1038/ng.2586

Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. (2012). Five years of GWAS discovery. Am. J. Hum. Genet. *90*, 7–24. http://dx.doi.org/10.1016/j.ajhg.2011.11.029

Voinnet, O. (2001). RNA silencing as a plant immune system against viruses. Trends Genet. *17*, 449–459.

Wang, Y., Zhou, L., Li, D., Dai, L., Lawton-Rauh, A., Srimani, P.K., Duan, Y., and Luo, F. (2015). Genome-wide comparative analysis reveals similar types of NBS genes in hybrid *Citrus sinensis* genome and original *Citrus clementine* genome and provides new insights into non-TIR NBS genes. PLOS ONE*10*, e0121893. http://dx.doi.org/10.1371/journal.pone.0121893

Wasukira, A., Coulter, M., Al-Sowayeh, N., Thwaites, R., Paszkiewicz, K., Kubiriba, J., Smith, J., Grant, M., and Studholme, D.J. (2014). Genome sequencing of *Xanthomonas vasicola* Pathovar *vasculorum* reveals variation in plasmids and genes encoding lipopolysaccharide synthesis, Type-IV Pilus and Type-III secretion effectors. Pathogens *3*, 211–237. http://dx.doi.org/10.3390/pathogens3010211

Weber, A.P. (2015). Discovering new biology through sequencing of RNA. Plant Physiol. *169*, 1524–1531. http://dx.doi.org/10.1104/pp.15.01081

Wu, J., Wang, Z., Shi, Z., Zhang, S., Ming, R., Zhu, S., Khan, M.A., Tao, S., Korban, S.S., Wang, H., *et al.* (2013). The genome of the pear (*Pyrus bretschneideri* Rehd.). Genome Res. *23*, 396–408. http://dx.doi.org/10.1101/gr.144311.112

Xu, Q., Chen, L.L., Ruan, X., Chen, D., Zhu, A., Chen, C., Bertrand, D., Jiao, W.B., Hao, B.H., Lyon, M.P., *et al.* (2013a). The draft genome of sweet orange (Citrus sinensis). Nat. Genet. *45*, 59–66. http://dx.doi.org/10.1038/ng.2472

Xu, Y., Gao, S., Yang, Y., Huang, M., Cheng, L., Wei, Q., Fei, Z., Gao, J., and Hong, B. (2013b). Transcriptome sequencing and whole genome expression profiling of chrysanthemum under dehydration stress. BMC Genomics *14*, 662. http://dx.doi.org/10.1186/1471-2164-14-662

Zhang, J., Liu, J., and Ming, R. (2014). Genomic analyses of the CAM plant pineapple. J. Exp. Bot. *65*, 3395–3404. http://dx.doi.org/10.1093/jxb/eru101

Zhang, Q., Chen, W., Sun, L., Zhao, F., Huang, B., Yang, W., Tao, Y., Wang, J., Yuan, Z., Fan, G., *et al.* (2012). The genome of Prunus mume. Nat. Commun. *3*, 1318. http://dx.doi.org/10.1038/ncomms2290

Zhu, B., Yang, Y., Li, R., Fu, D., Wen, L., Luo, Y., and Zhu, H. (2015). RNA sequencing and functional analysis implicate the regulatory role of long non-coding RNAs in tomato fruit ripening. J. Exp. Bot. *66*, 4483–4495. http://dx.doi.org/10.1093/jxb/erv203