Chapter 2

Implementation of a pipeline for Expressed Sequence Tags analysis

Expressed Sequence Tag (EST) projects by pyrosequencing produce high amounts of redundant, partial sequences which need further data analysis. The processing steps are guided towards generating a biological database. The final EST database gathers the sequencing information after data quality check, assembly and annotation. After the raise of cheap next generation sequencing facilities, EST projects became available to small laboratories, therefore generating new needs of data handling.

The software pESTle (*p*ipeline for *E*ST *l*ocal *e*xploit) is a set of procedures which automatically quality checks, assembles, stores and annotates ESTs generated *via* high-throughput sequencing technologies. It uses a PostgreSQL relational database as storage system, easing the information retrieval; and well tested, third-party bioinformatics tools for assembly and functional categorization. pESTle performs different data mining procedures and annotates the sequences locally. It also produces a browseable Web page gathering the results, easing information retrieval. In order to exemplify its usefulness, the software development was paired with a laboratory experiment comprising a 454 sequencing of 600,000 reads.

pESTle addresses the challenging EST data mining procedure in a well-established database management system platform, which allows high flexibility and customization capabilities. pESTle performs the data analysis from raw data resulting on a curated set of information stored; and serves it by a user-friendly Web interface.[7]

[7]The release candidate can be requested to izaskun.mallona@upct.es under the GPL v2 terms.

2.1 Introduction

Transcriptome sequencing projects produce expressed sequence tags (ESTs), that reflect the transcribed parts of a genome. Comprehensive sets of ESTs are used for gene discovery, gene mapping and marker development. Since the spread of cheap EST sequencing projects, many analytical pipelines to deal with ESTs have been developed. Their goals are to manipulate the raw data obtained by sequencing and to integrate it in a database, which gathers further mining results on the sequences (Nagaraj *et al.*, 2007b).

The three major steps on EST processing include checking for sequencing errors and contamination, EST assembly and functional annotation. Electronic function inference can be produced by similarity searches against annotated databases, thus enriching the starting assembled sequences with putative ontologies and biochemical functions (Ayoubi *et al.*, 2002). The assembled sequences can be searched for patterns, such as repeats (Robinson *et al.*, 2004) or single nucleotide polymorphism (SNP) calling, or summarized, such as by codon usage analysis (Nakamura *et al.*, 2000).

Common EST analysis procedures, including function inference, checking for repeats and mapping to other databases, can be integrated into dataflows of consecutive steps. However, the design of these pipelines differ in some aspects, such as: whether they are executed locally or depend on external servers; the starting data format allowed (that is, chromatograms, fasta files, phd files and so) and the associated amount of effort required for pre-processing (such as vector, adaptor and low-quality bases removal or chimeric reads detection); and the manner the data is offered afterward (presence of a Web interface or not). Albeit their differences, these systems share an automated or semi-automated procedure for cleansing, assembling and annotating through comparison to public databases (Nagaraj *et al.*, 2007b). A shortlist of tools developed for EST analysis include PipeOnline 2.0 (Ayoubi *et al.*, 2002), ParPEST (D'Agostino *et al.*, 2005), ESTExplorer (Nagaraj *et al.*, 2007a), ESTpass (Lee *et al.*, 2007), EST2uni (Forment *et al.*, 2008), dCAS (Guo *et al.*, 2009), est2assembly (Papanicolaou *et al.*, 2009) and ngs_backbone (Blanca *et al.*, 2011).

Here we present a new pipeline, pESTle, which performs the EST mining locally, thus increasing data handling independence over Web-based services. As output it produces a Web-based searchable interface allowing data querying and retrieval. pESTle has three major components: first, the database; second, the scripts used for the data preprocessing and anno-

tation; and third, the scripts leading to the development of a Common Interface Gateway (CGI) searchable database. The software package is freely distributed under a GPL v2 license, and runs on a Linux-based server with Apache, python/Bioython (Chapman and Chang, 2000), PostgreSQL and EMBOSS (Rice *et al.*, 2000).

2.2 Material and methods

2.2.1 Wet lab

Methods for RNA extraction, cDNA production and sequencing were described by Mallona *et al.* (2011a).

2.2.2 Environment

pESTle is developed in Python/Biopython, C and PostgreSQL by iterative and incremental development and mostly under the object-oriented programming paradigm (Booch *et al.*, 2007), and uses an enhanced entity-relationship model (EER) to design the database (Chen, 1976).

To satisfy the computational requirements of the assembly and functional annotation, the analysis were performed on a cluster using a node of two Intel Xeon Quad-Core. Job control was performed with Torque, an open source version of the original Portable Batch System (PBS) project (Jones, 2001) developed by NASA, Ames Research Center, Lawrence Livermore National Laboratory, and Veridian Information Solutions, Inc.

The Web server offering the graphical user interface (as that present in `http://srvgen.upct.es/opuntia/database.html`, user: opuntia, password: Opuntia ficus-indica) runs with one GB RAM and a CPU at 2.80GHz under Ubuntu GNU/Linux with kernel 2.6.31-14-server.

2.3 Results

2.3.1 Data flow

The pipeline starts with the raw reads produced by 454 pyrosequencing. The output is a fully assembled, quality checked collection of clustered sequences and singletons which are annotated and accessible through a Web interface.

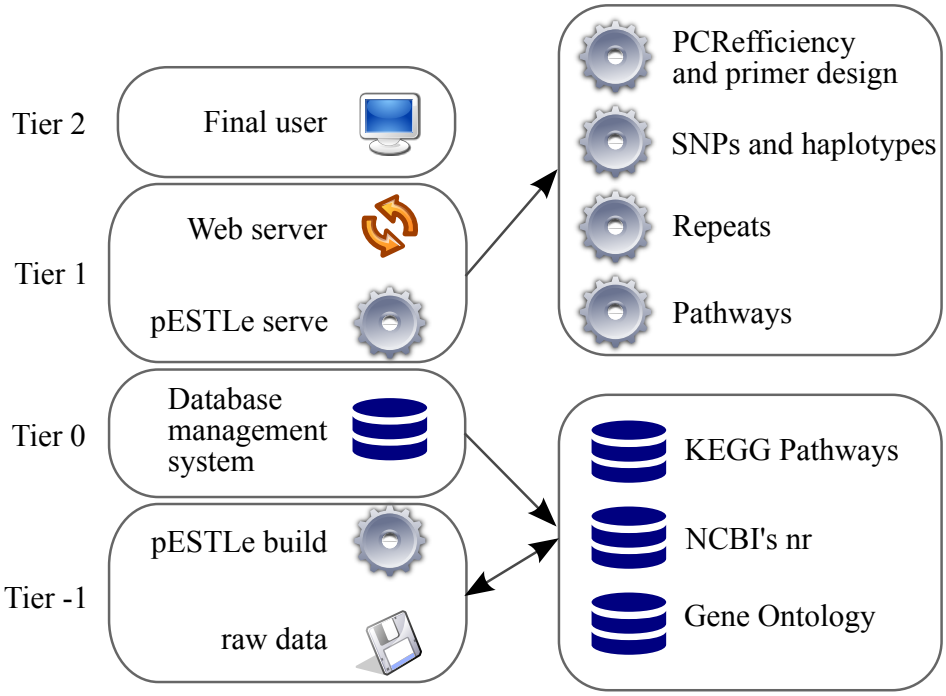

Figure 2.1: pESTle architecture. The proposed EST database comprises five tiers. Tier -1 includes the preprocessing, clustering and assembly of the raw data and interacts with previously established databases, such as Gene Ontology or NCBI's nonredundant. Tier 0 is the PostgreSQL core of database management system. Tier 1 comprises the pESTle scripts intended to facilitate the access to the database, and crosstalks with the common gateway interface and the apache Web server. Finally, tier 2 is the graphical user interface from which the final user queries the database.

The pESTle environment stores and queries the EST data through the PostgreSQL database management system. As it is a relational database system, it stores the data in interconnected tables in a module-built manner, thus easing data updating, such as adding new features o running again some of the processing steps.

The EST analysis comprises four consecutive steps: first, the preprocessing; second, the clustering and assembly; third, the structural annotation; and fourth, the functional annotation.

During the preprocessing, the raw data files are checked for vector or low-quality zones and are conveniently edited. Short sequences or sequencing artifacts are rejected. Repeat searching is conducted with RepeatMasker (Chen, 2004), and vector contamination with cross_match (Ewing *et al.*, 1998).

The structural annotation step relies on mining sequence patterns, such as SSRs or SNPs. sputnik (Robinson *et al.*, 2004) allows SSR detection and qualitysnpng, an update of qualitysnp (Tang *et al.*, 2006), the SNP recognition and haplotype number estimation[8]

The functional annotation step assigns a putative function and several categories, such as KEGG Orthologies or Gene Ontology annotations, to the ESTs selected. The ESTs are queried against well described databases, such EBI's gene association files or Pfam databases, *via* blast (Altschul *et al.*, 1990). In order to avoid electronically inferred putative misassignments, the user is asked to decide whether only human curated databases must be used; if not, electronically annotated databases are queried but descriptors containing terms such as "unknown" or "hypothetical" are skipped (Forment *et al.*, 2008).

In plants, gene and whole genome duplications occur, thus challenging sequence clustering and differentiation between alleles and paralogues. pESTle handles the data assembly with the well tested CAP3 assembler, and uses by default arguments leading to high astringency. After the alignment, the ace file is parsed and submitted to a new algorithm of single nucleotide polymorphism (SNP) and haplotype detection (data not shown).

2.3.2 Architecture

The data flow architecture (figure 2.1) is designed in four tiers over locally managed databases, thus ensuring security at several levels. pESTle op-

[8]In SNP mining, we define haplotype as a variant of a transcript; that is, a group of sequences within a cluster, discarding paralogs, which can be handled as an allele. Thus there are as many possible alleles as the ploidy of the organism.. A new algorithm

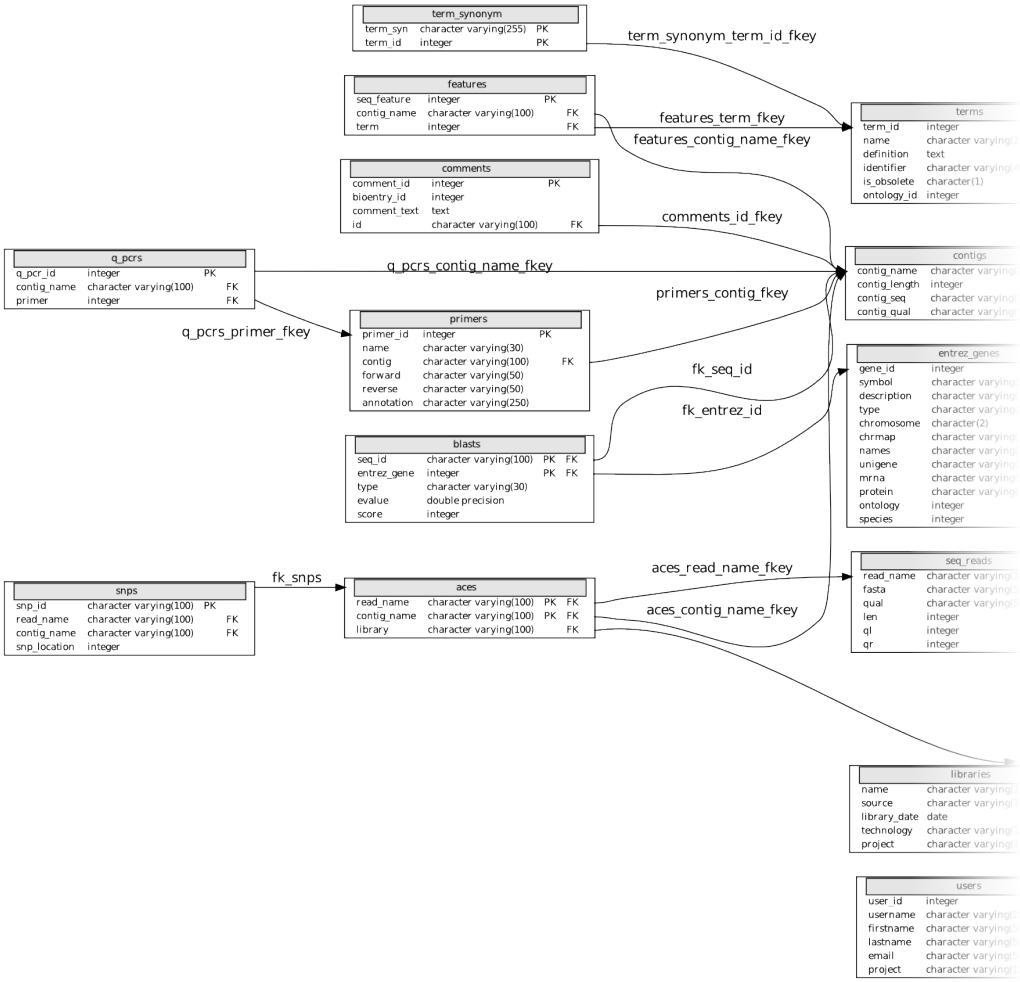

Figure 2.2: Entity Relationship Diagram of the core of pESTle, as represented by PostgreSQL autodoc (Taylor, 2007), part I.

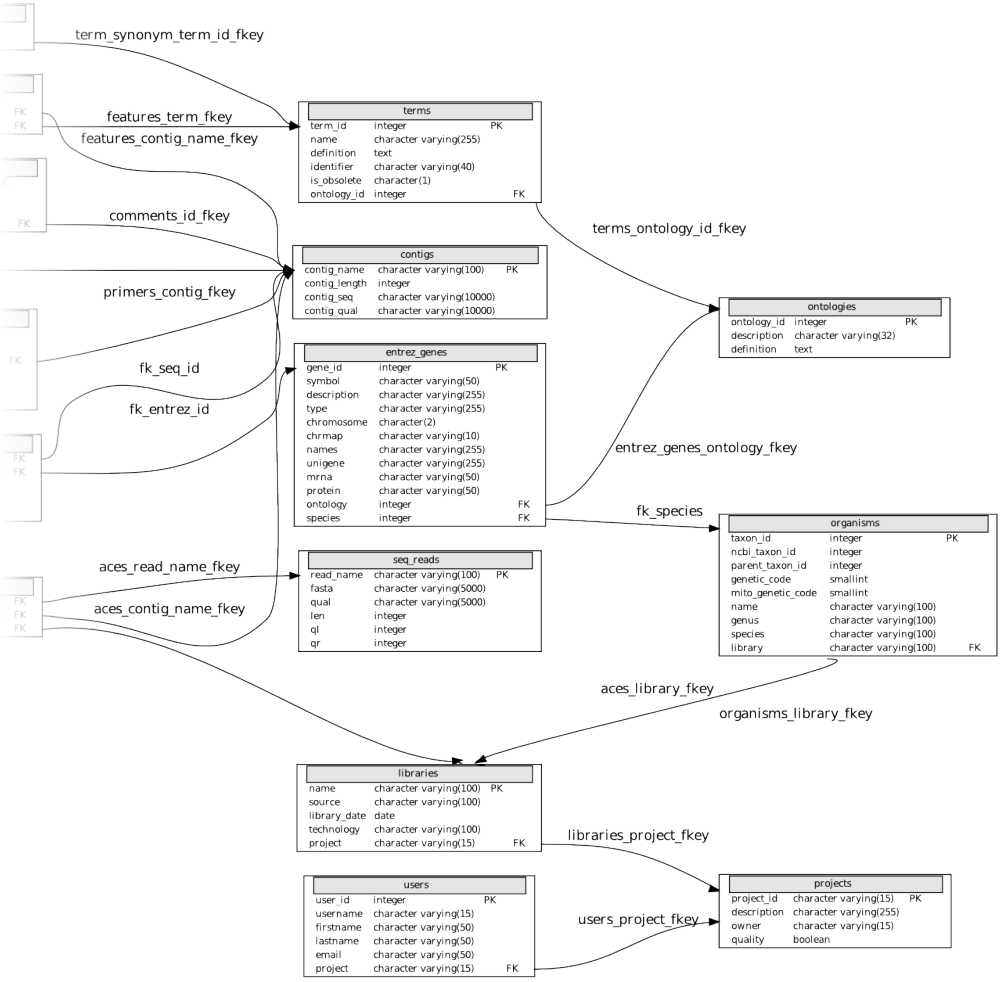

Figure 2.3: Entity Relationship Diagram of the core of pESTle, as represented by PostgreSQL autodoc (Taylor, 2007), part II.

erates in two major user cases: data retrieval from a fully assembled EST database, and database generation from raw sequencing reads.

Regarding the first case, the final user accesses the database at tier 2, whereas the core of the data is located in tier 0, with pESTle interleaving the two layers. It is interesting to note that the final user at tier 2 cannot introduce raw SQL (structured query language) sentences for database access; the graphical user interface at tier 1 restricts those to a set of allowed queries, thus adding a layer of security to the database, which is physically handled by the PostgreSQL core at tier 0. The architecture regards crosstalk between external applications such as a primer design application which estimates also PCR efficiency (Mallona *et al.*, 2011b), databases such as NCBI's nonredundant, or the result of mining the tandem repeats as computed by trfinder (Benson, 1999). The relational database structure, normalized according to the Boyce-Codd criteria (Codd, 1974), is represented by its EER diagram presented in figure 2.2.

The database build by pESTle starts with the raw data at tier -1, which communicates with reference databases (such as NCBI's nonredundant, or Gene Association files), and leads to the generation of a database accessible through the database management system at tier 0. pESTle comprises all the scripts used for communicating the quality check and assembly applications as well the annotating tool. It is worth noting that the final user has no access by graphical user interfaces to pESTle scripts used for database generation.

Data integrity is secured by the PostgreSQL management system at tier 0, which ensures: referencial and declarative checks of constraints; transaction logging and convenient rollbacking in case of unexpected errors (such as energy supply failures); concurrency control, reducing chances of interferences between concurrent applications; and triggers raising exceptions when the consistency checks do not succeed (Stinson, 2001).

To explore the data quering flexibility, a SQL query is summarized below. It selects all the contigs satisfying three conditions: first, to be annotated by blast to *A. thaliana* as involved in "calcium signalling"; second, having one or more SNPs that change "A" to "T"; and third, being longer than 500 bp. The query result contain: the contig name and length, the gene symbol of the homologous *A. thaliana* entry, the blast score for such assignation, and the number of SNPs. In case of multiple results, the

for SNP and haplotype calling has been developed and included into pESTle (data not shown).

matching objects are sorted by the blast score and the SNP location.

It is worth noting that the raw SQL searching is restricted to advanced users, as the Web graphical user interface offers pre-established SQL queries.

```
SELECT
  c.contig_name, c.contig_length, e.symbol, b.score, s.number
FROM
  blast_arabidopsis b, contig c, snps s, q_pcrs q, entrez_genes e
WHERE
  e.description LIKE '%calcium signaling%' AND
  s.snp_type AT AND
  c.contig_length >500
ORDER BY b.score, s.location DESC;
```

2.4 Discussion

A paradigm of successful biological database is Ensembl, a platform which integrates genomic information with abundant references to external databases. Although firstly focused on Chordata genomes, its scope has widened in species and in features, including regulation and function apart from raw sequences. One of the goals of the platform is to serve visual representations aiding on data interpretation, thus acting as an integrator and gatherer of knowledge (Flicek *et al.*, 2010). Moreover, data storage and mining goes beyond a mere aid on sequence documentation. Bourne (2005) reflected on the value of the database entries compared to journal papers. In his opinion, the database entry is more accessed and, under certain point of view, more important; however, the visibility of a journal is much higher. pESTle offers a flexible and easy-to-use EST database management system.

The increasing number of biological databases devoted to a species reflect that the paradigm of a single, all-including, biological database is a lacking solution. The specialization derived from the interests and expertise of a given community do enrich the panorama, as it leads to the scientific independence of the databases which, in spite of that, can be interlinked to others enabling cross-database querying (Stein *et al.*, 2003). In a context of cheap sequencing projects affordable by modest institutions, small and independent projects are viable. pESTle, as a local

pipeline, cares of data processing and serving as well permits continuous update and population *via* wet lab feedback.

Many automated EST pipelines have been developed, adapted to different sequencing technologies, performing locally, in parallel or in the cloud and using different software implementation strategies. The design of a pipeline and the management of the data requires a significant effort in software engineering. These differences on architecture rely on the different layers behind the EST pipeline (raw data handling and assembly, annotation and database). pESTle runs locally, and thus gives high control on data management, getting free of third party annotation platforms. As both the annotation and the data serving are designed according to an object-oriented paradigm and as the database structure is relational, pESTle allows easy inclusion of new modules and capabilities, therefore allowing interaction to other tools.

2.5 An application of pESTle: OpuntiaESTdb

pESTle has been used to analyze the prickly pear *Opuntia ficus-indica* transcriptome and to build the web-searchable OpuntiaESTdb (Mallona *et al.*, 2011a). This cactus species has agricultural importance as is highly efficient doing photosyntesis. As obligate Crassulacean-acid metabolism (CAM) plant, *O. ficus-indica* can take up relatively large amounts of CO_2 with respect to water loss by transpiration (4 to 10 mmol CO_2 per mol H_2O compared to 1 to 1.5 mmol in C_3 plants), and the annual above-ground drymass can be increased by 37 to 40% for *O. ficus-indica* when the CO_2 level is doubled (Cui *et al.*, 1993) (Nobel and Israel, 1994). The usage of *O. ficus-indica* is very diverse. It is used as animal crop, either fresh or as silage, usually supplemented with protein feed and minerals (Mondragón-Jacobo and Pérez-González, 2001). Further uses are as vegetable and fruit crop (Saenz, 2000; Feugang *et al.*, 2006). Other usages include the mucilage from cladodes and fruit peels, a complex polysaccharide that can absorb large amounts of water, for alimentary, medical and cosmetic purposes as well as for improvement of the infiltration of the water into soil (Gardiner *et al.*, 1999) or as agent for clarifying drinking water (Saenz *et al.*, 2004). *O. ficus-indica* was shown to be hexaploid, even so it has also been reported as heptaploid (Pinkava, 2002).

The sequencing of cDNA derived from RNA pools of various tissues generated ESTs with an average read length of 344 bp (table 2.1). Cluster-

Table 2.1: pESTle usage example. *O. ficus-indica* EST characteristics determined before and after assembly. Each contig was compared with the NCBI's nonredundant database with the blastx software; contigs matching CAM or Arabidopsis are defined as those with their highest blastx hit scoring an accession of that species.

Sequences before assembly		
	Total Number of Reads	604,176
	Total Number of Bases w/o keys, tags and bad quality bases	208,173,730
	Average Read Length w/o keys, tags and bad quality bases	344
Sequences after assembly		
	Total Number	43,066 contigs and 407,253 singlets
	Contigs average length ± standard deviation	611.585±151.5864
	Singlets average length ± standard deviation	1384.955±196.6052
	Contigs matching CAM	29,835
	Contigs matching Arabidopsis	1015

ing of these ESTs produced a total of 43,066 contigs and 407,253 unassembled singletons with an average length of 612 and 1385 bp, respectively. Annotation against the NCBI's nonredundant database showed that 29,835 contigs produced the lowest blast hit scores against sequences from a CAM species (69.3%) as listed by Sayed (2001) whereas 1015 were closer to those of *A. thaliana* (2.4%).

The OpuntiaESTdb web interface includes a searchable database through blastn, tblastn, blastx, tblastx and blastp assembled contigs and singletons. Preexistent ESTs from CAM species were recovered from TIGR (The Institute for Genomic Research) assemblies and included in the database to allow for more comprehensive searches. Contig sequence recovery includes functional annotation fetching, the annotations obtained from its RefSeq/UniProtKB putative orthologs. Thus, Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthologs and Pathways, PubMed publications, ExPaSy, InterPro and GO annotations, Tandem Repeats, and UniProtKB cellular locations and keywords are retrieved if present. KEGG Pathways including *O. ficus-indica* contigs are offered as highlighted maps. Additional information on data analysis and functional categorization can be found in Mallona *et al.* (2011a), including a flowchart of database construction and its applications; the comparative distribution of a selection of functional categories between the classified genes from the Arabidopsis genome and the *O. ficus-indica* EST clusters as detected by WEGO; the distribution of functional categories among contigs showing the 10 most common GO terms in the EST database for each of the three ontology domains, molecular function, cellular component and biological process; and the KEGG pathway for carbon fixation in photosynthetic organisms with highlighted orthologs from the OpuntiaESTdb.

2.6 Conclusions

A new 454-based EST analysis tool capable of handling both EST preprocessing, clustering and assembly and data serving has been produced. Designed to run locally, it ensures data security and integrity, as well presents a user-oriented web interface for easy-to-use data mining. pESTle is fully automatic and modular, thus easing the incorporation of new capabilities and the interaction with third party software. As a open-sourced project, pESTle offers high evolvability, as its functions could be reused by the bioinformatics community and new capabilities can be incorporated.

2.7 Authors' contributions

IM, ME and JW carried out the design of the study. JW performed the *O. ficus-indica* lab handling and participated in data collection. IM developed and conducted the data analysis strategy, designed and coded the software. IM wrote the manuscript. All authors read and approved the final manuscript.

2.8 Acknowledgements

We thank Luis Pedro García and SEDIC for aid in conducting the computational-intensive part of this work and Perla Gómez di Marco for comments on the manuscript. Funding by Fundación Séneca is acknowledged (Project ESP-SOL). IM is granted by Fundación Séneca.

References

Abràmoff, M., Magalhães, P. and Ram, S. (2004). Image processing with ImageJ. *Biophotonics International*, 11(7):36–42. (Cited on page 85.)

Adams, S., Pearson, S., Hadley, P. and Patefield, W. (1999). The Effects of Temperature and Light Integral on the Phases of Photoperiod Sensitivity in Petunia× hybrida. *Annals of Botany*, 83(3):263. (Cited on page 6.)

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723. (Cited on page 46.)

Akamatsu, T., Hanzawa, Y., Ohtake, Y., Takahashi, T., Nishitani, K. and Komeda, Y. (1999). Expression of endoxyloglucan transferase genes inacaulis mutants of Arabidopsis. *Plant Physiology*, 121(3):715–722. (Cited on page 97.)

Albert, N., Lewis, D., Zhang, H., Irving, L., Jameson, P. and Davies, K. (2009). Light-induced vegetative anthocyanin pigmentation in Petunia. *Journal of Experimental Botany*, 60(7):2191–2202. (Cited on page 4.)

Alizadeh, A., Eisen, M., Davis, R., Ma, C., Lossos, I., Rosenwald, A., Boldrick, J., Sabet, H., Tran, T., Yu, X. *et al.* (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, 403(6769):503–511. (Cited on page 17.)

Allison, D., Cui, X., Page, G. and Sabripour, M. (2006). Microarray data analysis: from disarray to consolidation and consensus. *Nature Reviews Genetics*, 7(1):55–65. (Cited on page 11.)

Altschul, S., Gish, W., Miller, W., Myers, E. and Lipman, D. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410. (Cited on page 29.)

Anastasiou, E., Kenz, S., Gerstung, M., MacLean, D., Timmer, J., Fleck, C. and Lenhard, M. (2007). Control of Plant Organ Size by KLUH CYP78A5-Dependent Intercellular Signaling. *Developmental Cell*, 13(6):843–856. (Cited on pages 20 and 80.)

Andersen, C., Jensen, J. and Orntoft, T. (2004). Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research*, 64(15):5245–5250. (Cited on pages 56, 62 and 75.)

Ando, T., Nomura, M., Tsukahara, J., Watanabe, H., Kokubun, H., Tsukamoto, T., Hashimoto, G., Marchesi, E. and Kitching, I. (2001). Reproductive isolation in a native population of Petunia sensu Jussieu (Solanaceae). *Annals of Botany*, 88(3):403. (Cited on page 3.)

Andreson, R., Mols, T. and Remm, M. (2008). Predicting failure rate of PCR in large genomes. *Nucleic Acids Research*, 36(11):e66–. (Cited on page 48.)

Andrew, F. (1999). RNA-triggered gene silencing. *Trends in Genetics*, 15(9):358 – 363. (Cited on pages 114 and 120.)

Ausubel, F., Bahnsen, K. and Hanson, M. (1980). Cell and tissue culture of haploid and diploid Petunia 'Mitchell'. *Plant Molecular Biology Newsletter*, 1:26–32. (Cited on page 4.)

Autran, D., Jonak, C., Belcram, K., Beemster, G., Kronenberger, J., Grandjean, O., Inze, D. and Traas, J. (2002). Cell numbers and leaf development in Arabidopsis: a functional analysis of the STRUWWELPETER gene. *The EMBO Journal*, 21(22):6036–6049. (Cited on page 75.)

Ayoubi, P., Jin, X., Leite, S., Liu, X., Martajaja, J., Abduraham, A., Wan, Q., Yan, W., Misawa, E. and Prade, R. (2002). PipeOnline 2.0: automated EST processing and functional data sorting. *Nucleic Acids Research*, 30(21):4761–4769. (Cited on page 26.)

Baer, R., Bankier, A., Biggin, M., Deininger, P., Farrell, P., Gibson, T., Hatfull, G., Hudson, G., Satchwell, S., Seguin, C. *et al.* (1984). DNA sequence and expression of the B95-8 Epstein—Barr virus genome. *Nature*, 310:207–211. (Cited on page 12.)

von Balthazar, M. and Endress, P. (2002). Development of inflorescences and flowers in Buxaceae and the problem of perianth interpretation. *International Journal of Plant Sciences*, 163(6):847–876. (Cited on page 20.)

Barnes, N. (2010). Publish your computer code: it is good enough. *Nature*, 467(7317):753–753. (Cited on page 16.)

Baskaran, N., Kandpal, R., Bhargava, A., Glynn, M., Bale, A. and Weissman, S. (1996). Uniform amplification of a mixture of deoxyribonucleic acids with varying GC content. *Genome Research*, 6(7):633. (Cited on page 49.)

Bassett Jr, D., Eisen, M. and Boguski, M. (1999). Gene expression informatics—it's all in your mine. *Nature Genetics*, 21:51. (Cited on page 2.)

Beemster, G., Vercruysse, S., De Veylder, L., Kuiper, M. and Inze, D. (2006). The Arabidopsis leaf as a model system for investigating the role of cell cycle regulation in organ growth. *Journal of Plant Research*, 119(1):43–50. (Cited on page 75.)

Ben-Nissan, G., Lee, J., Borohov, A. and Weiss, D. (2004). GIP, a Petunia hybrida GA-induced cysteine-rich protein: a possible role in shoot elongation and transition to flowering. *The Plant Journal*, 37(2):229–238. (Cited on page 6.)

Benita, Y., Oosting, R., Lok, M., Wise, M. and Humphery-Smith, I. (2003). Regionalized GC content of template DNA as a predictor of PCR success. *Nucleic Acids Research*, 31(16):e99. (Cited on pages 48 and 49.)

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, 27(2):573–580. (Cited on page 32.)

Bey, M., Stuber, K., Fellenberg, K., Schwarz-Sommers, Z., Sommer, H., Saedler, H. and Zachgo, S. (2004). Characterization of Antirrhinum petal development and identification of target genes of the class B MADS box gene DEFICIENS. *The Plant Cell*, 16(12):3197–3215. (Cited on page 61.)

Blanca, J., Pascual, L., Ziarsolo, P., Nuez, F. and Cañizares, J. (2011). ngs_backbone: a pipeline for read cleaning, mapping and SNP calling using Next Generation Sequence. *BMC Genomics*, 12(1):285. (Cited on page 26.)

Blattner, F., Plunkett III, G., Bloch, C., Perna, N., Burland, V., Riley, M., Collado-Vides, J., Glasner, J., Rode, C., Mayhew, G. *et al.* (1997). The complete genome sequence of Escherichia coli K-12. *Science*, 277(5331):1453–1462. (Cited on page 12.)

Boatright, J., Negre, F., Chen, X., Kish, C., Wood, B., Peel, G., Orlova, I., Gang, D., Rhodes, D. and Dudareva, N. (2004). Understanding in vivo benzenoid metabolism in Petunia petal tissue. *Plant Physiology*, 135(4):1993–2011. (Cited on page 57.)

Booch, G., Maksimchuk, R., Engle, M., Young, B., Conallen, J. and Houston, K. (2007). *Object-oriented analysis and design with applications*. Addison-Wesley Professional. (Cited on page 27.)

Bourne, P. (2005). Will a biological database be different from a biological journal? *PLoS Computational Biology*, 1(3):e34. (Cited on page 33.)

Bowman, J., Alvarez, J., Weigel, D., Meyerowitz, E. and Smyth, D. (1993). Control of flower development in Arabidopsis thaliana by APETALA1 and interacting genes. *Development*, 119(3):721–743. (Cited on page 6.)

Bowman, J., Drews, G. and Meyerowitz, E. (1991). Expression of the Arabidopsis floral homeotic gene AGAMOUS is restricted to specific cell types late in flower development. *The Plant Cell*, 3(8):749–758. (Cited on page 6.)

Boyes, D., Zayed, A., Ascenzi, R., McCaskill, A., Hoffman, N., Davis, K. and Görlach, J. (2001). Growth stage–based phenotypic analysis of Arabidopsis: A model for high throughput functional genomics in plants. *The Plant Cell*, 13(7):1499. (Cited on page 19.)

Bradley, D., Carpenter, R., Sommer, H., Hartley, N. and Coen, E. (1993). Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the plena locus of Antirrhinum. *Cell*, 72(1):85–95. (Cited on pages 6 and 20.)

Bradley, J., Davies, K., Deroles, S., Bloor, S. and Lewis, D. (1998). The maize Lc regulatory gene up-regulates the flavonoid biosynthetic pathway of Petunia. *The Plant Journal*, 13(3):381–392. (Cited on page 4.)

Breslauer, K., Frank, R., Blöcker, H. and Marky, L. (1986). Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences*, 83(11):3746. (Cited on page 41.)

Britton, N. and Brown, A. (1913). *An Illustrated Flora of the Northern United States, Canada and the British Possessions*. Charles Scribner's Sons. (Cited on page ii.)

Brunner, A., Yakovlev, I. and Strauss, S. (2004). Validating internal controls for quantitative plant gene expression studies. *BMC Plant Biology*, 4:14. (Cited on page 75.)

Burge, C., Karlin, S. *et al.* (1997). Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology*, 268(1):78–94. (Cited on page 83.)

Bustin, S., Benes, V., Garson, J., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M., Shipley, G. *et al.* (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, 55(4):611. (Cited on pages 10, 18 and 40.)

Castel, R., Kusters, E. and Koes, R. (2010). Inflorescence development in petunia: through the maze of botanical terminology. *Journal of Experimental Botany*, 61(9):2235. (Cited on page 3.)

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17(4):540–552. (Cited on page 83.)

Chapman, B. and Chang, J. (2000). Biopython: Python tools for computational biology. *ACM SIGBIO Newsletter*, 20(2):19. (Cited on pages 27 and 48.)

Chavali, S., Mahajan, A., Tabassum, R., Maiti, S. and Bharadwaj, D. (2005). Oligonucleotide properties determination and primer designing: a critical examination of predictions. *Bioinformatics*, 21(20):3918. (Cited on page 48.)

Chen, J., Jiang, C., Gookin, T., Hunter, D., Clark, D. and Reid, M. (2004). Chalcone synthase as a reporter in virus-induced gene silencing studies of flower senescence. *Plant Molecular Biology*, 55(4):521–530. (Cited on pages 116 and 122.)

Chen, L., Ren, Y., Endress, P., Tian, X. and Zhang, X. (2007). Floral organogenesis in Tetracentron sinense (Trochodendraceae) and its sys-

tematic significance. *Plant Systematics and Evolution*, 264(3):183–193. (Cited on page 20.)

Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*. (Cited on page 29.)

Chen, P. (1976). The entity-relationship model—toward a unified view of data. *ACM Transactions on Database Systems (TODS)*, 1(1):9–36. (Cited on page 27.)

Choe, S., Schmitz, R., Fujioka, S., Takatsuto, S., Lee, M., Yoshida, S., Feldmann, K. and Tax, F. (2002). Arabidopsis Brassinosteroid-Insensitive dwarf12Mutants Are Semidominant and Defective in a Glycogen Synthase Kinase 3β-Like Kinase. *Plant Physiology*, 130(3):1506–1515. (Cited on page 98.)

Clough, S. and Bent, A. (1998). Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana. *The Plant Journal*, 16(6):735–743. (Cited on page 19.)

Cnudde, F., Moretti, C., Porceddu, A., Pezzotti, M. and Gerats, T. (2003). Transcript profiling on developing Petunia hybrida floral organs. *Sexual Plant Reproduction*, 16(2):77–85. (Cited on page 57.)

Cock, P., Antao, T., Chang, J., Chapman, B., Cox, C., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B. and de Hoon, M. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423. (Cited on page 16.)

Codd, E. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387. (Cited on page 17.)

Codd, E. (1974). Recent investigations in relational data base systems. In: *Proceedings of the International Federation for Information Processing Congress*, volume 74, pp. 1017–1021. (Cited on page 32.)

Coen, E. and Meyerowitz, E. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature*, 353(6339):31–37. (Cited on pages 6 and 19.)

Coker, J. and Davies, E. (2003). Selection of candidate housekeeping controls in tomato plants using EST data. *Biotechniques*, 35(4):740–742,744–746. (Cited on page 56.)

Colombo, L., van Tunen, A., Dons, H. and Angenent, G. (1997). Molecular control of flower development in Petunia hybrida. *Advances in Botanical Research*, 26:229–250. (Cited on page 3.)

Colquhoun, T., Verdonk, J., Schimmel, B., Tieman, D., Underwood, B. and Clark, D. (2010). Petunia floral volatile benzenoid/phenylpropanoid genes are regulated in a similar manner. *Phytochemistry*, 71(2-3):158–167. (Cited on page 8.)

Crepet, W. (2000). Progress in understanding angiosperm history, success, and relationships: Darwin's abominably "perplexing phenomenon". *Proceedings of the National Academy of Sciences*, 97(24):12939. (Cited on page 80.)

Cruz, F., Kalaoun, S., Nobile, P., Colombo, C., Almeida, J., Barros, L., Romano, E., Grossi-de Sa, M., Vaslin, M. and Alves-Ferreira, M. (2009). Evaluation of coffee reference genes for relative expression studies by quantitative real-time RT-PCR. *Molecular Breeding*, 23(4):607–616. (Cited on page 76.)

Cui, M., Miller, P. and Nobel, P. (1993). CO2 exchange and growth of the crassulacean acid metabolism plant Opuntia ficus-indica under elevated CO2 in open-top chambers. *Plant Physiology*, 103(2):519. (Cited on page 34.)

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. and Scheible, W. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiology*, 139(1):5–17. (Cited on pages 56 and 61.)

Da Wei Huang, B. and Lempicki, R. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1):44–57. (Cited on page 82.)

D'Agostino, N., Aversano, M. and Chiusano, M. (2005). ParPEST: a pipeline for EST data analysis based on parallel computing. *BMC bioinformatics*, 6(Suppl 4):S9. (Cited on page 26.)

Dal Santo, S., Fasoli, M., Cavallini, E., Tornielli, G., Pezzotti, M. and Zenoni, S. (2011). PhEXPA1, a Petunia hybrida expansin, is involved in cell wall metabolism and in plant architecture specification. *Plant Signaling & Behavior*, 6(12). (Cited on pages 20 and 80.)

Danna, K. and Nathans, D. (1971). Specific cleavage of simian virus 40 DNA by restriction endonuclease of Hemophilus influenzae. *Proceedings of the National Academy of Sciences*, 68(12):2913. (Cited on page 12.)

Davies, B., Egea-Cortines, M., de Andrade Silva, E., Saedler, H. and Sommer, H. (1996). Multiple interactions amongst floral homeotic MADS box proteins. *The EMBO Journal*, 15(16):4330. (Cited on page 19.)

Davis, J. and Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. In: *Proceedings of the 23rd international conference on Machine Learning*, pp. 233–240. ACM. (Cited on page 47.)

De Folter, S., Immink, R., Kieffer, M., Parenicova, L., Henz, S., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M. *et al.* (2005). Comprehensive interaction map of the Arabidopsis MADS box transcription factors. *The Plant Cell*, 17(5):1424–1433. (Cited on page 94.)

Delgado-Benarroch, L., Causier, B., Weiss, J. and Egea-Cortines, M. (2009a). FORMOSA controls cell division and expansion during floral development in Antirrhinum majus. *Planta*, 229:1219–1229. (Cited on page 61.)

Delgado-Benarroch, L., Weiss, J. and Egea-Cortines, M. (2009b). The mutants compacta ähnlich, Nitida and Grandiflora define developmental compartments and a compensation mechanism in floral development in Antirrhinum majus. *Journal of Plant Research*, 122(5):559–569. (Cited on page 43.)

Demet, S., Michael, W., Florian, W. and Jürgen, P. (2010). Prediction and analysis of the modular structure of cytochrome P450 monooxygenases. *BMC Structural Biology*, 10. (Cited on page 83.)

Disch, S., Anastasiou, E., Sharma, V., Laux, T., Fletcher, J. and Lenhard, M. (2006). The E3 Ubiquitin Ligase BIG BROTHER Controls¡ i¿ Arabidopsis¡/i¿ Organ Size in a Dosage-Dependent Manner. *Current Biology*, 16(3):272–279. (Cited on pages 20 and 80.)

Ditta, G., Pinyopich, A., Robles, P., Pelaz, S. and Yanofsky, M. (2004). The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. *Current Biology*, 14(21):1935–1940. (Cited on page 6.)

Dudareva, N. and Pichersky, E. (2008). Metabolic engineering of plant volatiles. *Current Opinion in Biotechnology*, 19(2):181–189. (Cited on page 57.)

Ecker, J. and Davis, R. (1986). Inhibition of gene expression in plant cells by expression of antisense RNA. *Proceedings of the National Academy of Sciences*, 83(15):5372. (Cited on page 12.)

Edgar, R. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–1797. (Cited on page 83.)

Egea-Cortines, M., Saedler, H. and Sommer, H. (1999). Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in Antirrhinum majus. *The EMBO Journal*, 18(19):5370–5379. (Cited on page 19.)

Ekblom, R. and Galindo, J. (2010). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, 107(1):1–15. (Cited on page 12.)

Endress, P. (1999). Symmetry in flowers: diversity and evolution. *International Journal of Plant Sciences*, pp. 3–23. (Cited on page 7.)

Endress, P. (2006). Angiosperm floral evolution: morphological developmental framework. *Advances in Botanical Research*, 44:1–61. (Cited on pages 5 and 7.)

Etzold, T., Ulyanov, A. and Argos, P. (1996). SRS: Information retrieval system for molecular biology data banks. *Methods in Enzymology*, 266:114–128. (Cited on page 17.)

Ewing, B., Hillier, L., Wendl, M. and Green, P. (1998). Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Research*, 8(3):175. (Cited on page 29.)

Exposito-Rodriguez, M., Borges, A., Borges-Perez, A. and Perez, J.

(2008). Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. *BMC Plant Biology*, 8(1):131. (Cited on pages 56, 75 and 76.)

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, pp. 783–791. (Cited on page 83.)

Feugang, J., Konarski, P., Zou, D., Stintzing, F., Zou, C. *et al.* (2006). Nutritional and medicinal use of cactus pear (Opuntia spp.) cladodes and fruits. *Front Biosci*, 11(1):2574–2589. (Cited on page 34.)

Fischer, M., Knoll, M., Sirim, D., Wagner, F., Funke, S. and Pleiss, J. (2007). The Cytochrome P450 Engineering Database: a navigation and prediction tool for the cytochrome P450 protein family. *Bioinformatics*, 23(15):2015–2017. (Cited on page 83.)

Fischer, M., Thai, Q., Grieb, M. and Pleiss, J. (2006). DWARF–a data warehouse system for analyzing protein families. *BMC bBioinformatics*, 7(1):495. (Cited on page 83.)

Flicek, P., Aken, B., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S. *et al.* (2010). Ensembl's 10th year. *Nucleic Acids Research*, 38(suppl 1):D557–D562. (Cited on page 33.)

Forment, J., Gilabert, F., Robles, A., Conejero, V., Nuez, F. and Blanca, J. (2008). EST2uni: an open, parallel tool for automated EST analysis and database creation, with a data mining web interface and microarray expression data integration. *BMC Bioinformatics*, 9(1):5. (Cited on pages 26 and 29.)

Fraley, R., Rogers, S., Horsch, R., Sanders, P., Flick, J., Adams, S., Bittner, M., Brand, L., Fink, C., Fry, J. *et al.* (1983). Expression of bacterial genes in plant cells. *Proceedings of the National Academy of Sciences*, 80(15):4803. (Cited on page 12.)

Freeman, W., Walker, S. and Vrana, K. (1999). Quantitative RT-PCR: pitfalls and potential. *Biotechniques*, 26:112–125. (Cited on pages 113 and 120.)

Fujibuchi, W., Goto, S., Migimatsu, H., Uchiyama, I., Ogiwara, A., Akiyama, Y. and Kanehisa, M. (1998). DBGET/LinkDB: an integrated database retrieval system. In: *Pacific Symposium on Biocomputing*, volume 98, pp. 683–694. (Cited on page 17.)

Furutani, M., Kajiwara, T., Kato, T., Treml, B., Stockum, C., Torres-Ruiz, R. and Tasaka, M. (2007). The gene MACCHI-BOU 4/ENHANCER OF PINOID encodes a NPH3-like protein and reveals similarities between organogenesis and phototropism at the molecular level. *Development*, 134(21):3849–3859. (Cited on page 95.)

Furutani, M., Sakamoto, N., Yoshida, S., Kajiwara, T., Robert, H., Friml, J. and Tasaka, M. (2011). Polar-localized NPH3-like proteins regulate polarity and endocytosis of PIN-FORMED auxin efflux carriers. *Development*, 138(10):2069–2078. (Cited on page 95.)

Gabrielsson, B., Olofsson, L., Sjogren, A., Jernas, M., Elander, A., Lonn, M., Rudemo, M. and Carlsson, L. (2005). Evaluation of reference genes for studies of gene expression in human adipose tissue. *Obesity Research*, 13:649–652. (Cited on page 75.)

Gagneux, P. (2003). Gene Regulation: A Eukaryotic Perspective. *Journal of Heredity*, 94(6):528. (Cited on page 10.)

Gardiner, D., Felker, P. and Carr, T. (1999). Cactus extract increases water infiltration rates in two soils. *Communications in Soil Science & Plant Analysis*, 30(11-12):1707–1712. (Cited on page 34.)

Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution*, 14(7):685–695. (Cited on page 83.)

Gautier, L., Cope, L., Bolstad, B. and Irizarry, R. (2004). affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, 20(3):307–315. (Cited on page 12.)

Gerats, T. (2009). Identification and Exploitation of Petunia Transposable Elements: a brief history. *Petunia*, pp. 365–379. (Cited on page 9.)

Gerats, T. and Strommer, J. (2009). *Petunia. Evolutionary, developmental and physiological genetics*. New York: Springer. (Cited on page 56.)

Gerats, T. and Vandenbussche, M. (2005). A model system for comparative research: Petunia. *Trends in Plant Science*, 10(5):251 – 256. Plant Model Systems. (Cited on pages 2, 3, 7, 8, 56, 113 and 120.)

Gilbert, S. (2001). Ecological Developmental Biology: Developmental Biol-

ogy Meets the Real World. *Developmental Biology*, 233(1):1 – 12. (Cited on page 5.)

Gilbert, W. and Maxam, A. (1973). The nucleotide sequence of the lac operator. *Proceedings of the National Academy of Sciences*, 70(12):3581. (Cited on page 12.)

Goffeau, A., Aert, R., Agostini-Carbone, M., Ahmed, A., Aigle, M., Alberghina, L., Albermann, K., Albers, M., Aldea, M., Alexandraki, D. *et al.* (1997). The yeast genome directory. *Nature*, 387(6632):5–6. (Cited on page 12.)

Golub, T., Slonim, D., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J., Coller, H., Loh, M., Downing, J., Caligiuri, M. *et al.* (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286(5439):531–537. (Cited on page 17.)

Gonzalez-Verdejo, C., Die, J., Nadal, S., Jimenez-Marin, A., Moreno, M. and Roman, B. (2008). Selection of housekeeping genes for normalization by real-time RT-PCR: analysis of Or-MYB1 gene expression in Orobanche ramosa development. *Analytical Biochemistry*, 379(2):176–181. (Cited on page 56.)

Goto, K. and Meyerowitz, E. (1994). Function and regulation of the Arabidopsis floral homeotic gene PISTILLATA. *Genes and Development*, 8(13):1548–1560. (Cited on page 6.)

Goto, N., Prins, P., Nakao, M., Bonnal, R., Aerts, J. and Katayama, T. (2010). BioRuby: Bioinformatics software for the Ruby programming language. *Bioinformatics*, 26(20):2617. (Cited on page 16.)

Goulao, L., Fortunato, A. and C. Ramalho, J. (2011). Selection of Reference Genes for Normalizing Quantitative Real-Time PCR Gene Expression Data with Multiple Variables in Coffea spp. *Plant Molecular Biology Reporter*, pp. 1–19. (Cited on pages 113 and 119.)

Gouy, M., Guindon, S. and Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, 27(2):221–224. (Cited on page 83.)

Greer, S., Wen, M., Bird, D., Wu, X., Samuels, L., Kunst, L. and Jetter, R. (2007). The cytochrome P450 enzyme CYP96A15 is the midchain alkane

hydroxylase responsible for formation of secondary alcohols and ketones in stem cuticular wax of Arabidopsis. *Plant Physiology*, 145(3):653–667. (Cited on page 92.)

Gübitz, T., Hoballah, M., Dell'Olivo, A. and Kuhlemeier, C. (2009). Petunia as a model system for the genetics and evolution of pollination syndromes. *Petunia*, pp. 29–49. (Cited on page 8.)

Guescini, M., Sisti, D., Rocchi, M., Stocchi, L. and Stocchi, V. (2008). A new real-time PCR method to overcome significant quantitative inaccuracy due to slight amplification inhibition. *BMC Bioinformatics*, 9(1):326. (Cited on page 84.)

Guo, Y., Ribeiro, J., Anderson, J. and Bour, S. (2009). dCAS: a desktop application for cDNA sequence annotation. *Bioinformatics*, 25(9):1195–1196. (Cited on page 26.)

Gutierrez, L., Mauriat, M., Guenin, S., Pelloux, J., Lefebvre, J., Louvet, R., Rusterucci, C., Moritz, T., Guerineau, F. and Bellini, C. (2008). The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. *Plant Biotechnology Journal*, 6(6):609–618. (Cited on page 54.)

Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J., Snesrud, E., Lee, N. and Quackenbush, J. (2000). A concise guide to cDNA microarray analysis. *Biotechniques*, 29(3):548–563. (Cited on page 11.)

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. and Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology*, 8(2):R19. (Cited on pages 56, 62 and 63.)

Helliwell, C. and Waterhouse, P. (2003). Constructs and methods for high-throughput gene silencing in plants. *Methods*, 30(4):289–295. (Cited on pages 14, 81 and 83.)

Herskowitz, I. (1987). Functional inactivation of genes by dominant negative mutations. *Nature*, 329(6136):219–22. (Cited on page 13.)

Higuchi, R., Dollinger, G., Walsh, P. and Griffith, R. (1992). Simultaneous

amplification and detection of specific DNA sequences. *Biotechnology*, 10(4):413–417. (Cited on page 10.)

Higuchi, R., Fockler, C., Dollinger, G. and Watson, R. (1993). Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Nature Biotechnology*, 11(9):1026–1030. (Cited on page 39.)

Hilson, P., Allemeersch, J., Altmann, T., Aubourg, S., Avon, A., Beynon, J., Bhalerao, R., Bitton, F., Caboche, M., Cannoot, B. *et al.* (2004). Versatile gene-specific sequence tags for Arabidopsis functional genomics: transcript profiling and reverse genetics applications. *Genome Research*, 14(10b):2176. (Cited on pages 2, 13, 14, 83, 89 and 123.)

Himeno, M., Neriya, Y., Minato, N., Miura, C., Sugawara, K., Ishii, Y., Yamaji, Y., Kakizawa, S., Oshima, K. and Namba, S. (2011). Unique morphological changes in plant pathogenic phytoplasma-infected petunia flowers are related to transcriptional regulation of floral homeotic genes in an organ-specific manner. *The Plant Journal*. (Cited on page 94.)

Hoballah, M., Gübitz, T., Stuurman, J., Broger, L., Barone, M., Mandel, T., Dell'Olivo, A., Arnold, M. and Kuhlemeier, C. (2007). Single gene–mediated shift in pollinator attraction in Petunia. *The Plant Cell*, 19(3):779–790. (Cited on page 3.)

Holland, R.C.G., Down, T.A., Pocock, M., Prlić, A., Huen, D., James, K., Foisy, S., Dräger, A., Yates, A., Heuer, M. and Schreiber, M.J. (2008). BioJava: an open-source framework for bioinformatics. *Bioinformatics*, 24(18):2096–2097. (Cited on page 16.)

Hong, S., Seo, P., Yang, M., Xiang, F. and Park, C. (2008). Exploring valid reference genes for gene expression studies in Brachypodium distachyon by real-time PCR. *BMC Plant Biology*, 8:112. (Cited on page 56.)

Honma, T. and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*, 409(6819):525–529. (Cited on page 6.)

Horiguchi, G., Fujikura, U., Ferjani, A., Ishikawa, N. and Tsukaya, H. (2006). Large-scale histological analysis of leaf mutants using two simple leaf observation methods: identification of novel genetic pathways governing the size and shape of leaves. *The Plant Journal*, 48(4):638–644. (Cited on page 85.)

Horner, D., Pavesi, G., Castrignanò, T., De Meo, P., Liuni, S., Sammeth, M., Picardi, E. and Pesole, G. (2010). Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Briefings in Bioinformatics*, 11(2):181–197. (Cited on page 12.)

Horsch, R., Fry, J., Hoffmann, N., Eichholtz, D., Rogers, S. and Fraley, R. (1985). A simple and general method for transferring genes into plants. *Science*, 227:1229–1231. (Cited on page 83.)

Hothorn, T., Hornik, K., van de Wiel, M. and Zeileis, A. (2006). coin: Conditional Inference Procedures in a Permutation Test Framework. *URL http://CRAN. R-project. org, R package version 0.6-6.* (Cited on pages 42 and 44.)

Howley, P., Israel, M., Law, M. and Martin, M. (1979). A rapid method for detecting and mapping homology between heterologous DNAs. Evaluation of polyomavirus genomes. *Journal of Biological Chemistry*, 254(11):4876. (Cited on page 41.)

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W. and Zimmermann, P. (2008). Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. *Advances in Bioinformatics*, 2008. (Cited on page 82.)

Hu, Y., Xie, Q. and Chua, N. (2003). The Arabidopsis auxin-inducible gene ARGOS controls lateral organ size. *The Plant Cell*, 15(9):1951–1961. (Cited on pages 20 and 80.)

Huang, D., Sherman, B. and Lempicki, R. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, 37(1):1. (Cited on page 82.)

Imaishi, H. and Petkova-Andonova, M. (2007). Molecular cloning of CYP76B9, a cytochrome P450 from Petunia hybrida, catalyzing the ω-hydroxylation of capric acid and lauric acid. *Bioscience, Biotechnology, and Biochemistry*, (0):612070200. (Cited on page 94.)

Immink, R., Hannapel, D., Ferrario, S., Busscher, M., Franken, J., Campagne, M. and Angenent, G. (1999). A petunia MADS box gene involved in the transition from vegetative to reproductive development. *Development*, 126(22):5117–5126. (Cited on pages 89, 94 and 128.)

Irish, V. and Litt, A. (2005). Flower development and evolution: gene duplication, diversification and redeployment. *Current Opinion in Genetics & development*, 15(4):454–460. (Cited on page 19.)

Irish, V. and Sussex, I. (1990). Function of the apetala-1 gene during Arabidopsis floral development. *The Plant Cell*, 2(8):741–753. (Cited on page 6.)

Irizarry, R., Hobbs, B., Collin, F., Beazer-Barclay, Y., Antonellis, K., Scherf, U. and Speed, T. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4(2):249. (Cited on page 82.)

Jack, T., Brockman, L. and Meyerowitz, E. (1992). The homeotic gene APETALA3 of Arabidopsis thaliana encodes a MADS box and is expressed in petals and stamens. *Cell*, 68(4):683–697. (Cited on page 6.)

Jain, M., Nijhawan, A., Tyagi, A. and Khurana, J. (2006). Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochemical and Biophysical Research Communications*, 345(2):646–651. (Cited on page 56.)

Jain, S. and Brar, D. (2010). *Molecular techniques in crop improvement*. Springer Verlag. (Cited on page 15.)

Jian, B., Liu, B., Bi, Y., Hou, W., Wu, C. and Han, T. (2008). Validation of internal control for gene expression study in soybean by quantitative real-time PCR. *BMC Molecular Biology*, 9. (Cited on pages 56, 73 and 76.)

Jiao, Y., Tausta, S., Gandotra, N., Sun, N., Liu, T., Clay, N., Ceserani, T., Chen, M., Ma, L. and Holford, M. (2009). A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nature Genetics*, 41(2):258–263. (Cited on page 54.)

Jofuku, K., den Boer, B., Montagu, M.V. and Okamuro, J. (1994). Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. *The Plant Cell*, 6(9):1211–1225. (Cited on page 6.)

Jones, J. (2001). PBS: portable batch system. In: *Beowulf cluster computing with Linux*, pp. 369–390. MIT Press. (Cited on page 27.)

Jou, W., Haegeman, G., Ysebaert, M. and Fiers, W. (1972). Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature*, 237:82–88. (Cited on page 12.)

Kamei, C., Boruc, J., Vandepoele, K., Van den Daele, H., Maes, S., Russinova, E., Inzé, D. and De Veylder, L. (2008). The PRA1 gene family in Arabidopsis. *Plant Physiology*, 147(4):1735–1749. (Cited on page 98.)

Kang, S., Kang, K., Lee, K. and Back, K. (2007). Characterization of tryptamine 5-hydroxylase and serotonin synthesis in rice plants. *Plant cell reports*, 26(11):2009–2015. (Cited on page 20.)

Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. Wiley-IEEE Press. (Cited on pages 16 and 17.)

Ke, L., Chen, Z. and Yung, W. (2000). A reliability test of standard-based quantitative PCR: exogenous vs endogenous standards. *Molecular and Cellular Probes*, 14(2):127–135. (Cited on page 76.)

Kell, D. and Oliver, S. (2004). Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays*, 26(1):99–105. (Cited on page 21.)

Kim, I., Yang, D., Tang, X. and Carroll, J. (2011). Reference gene validation for qPCR in rat carotid body during postnatal development. *BMC Research Notes*, 4(1):440. (Cited on pages 113 and 119.)

Kim, S., Soltis, D., Soltis, P., Zanis, M. and Suh, Y. (2004). Phylogenetic relationships among early-diverging eudicots based on four genes: were the eudicots ancestrally woody? *Molecular Phylogenetics and Evolution*, 31(1):16–30. (Cited on page 20.)

Kimura, M. and Kagawa, T. (2006). Phototropin and light-signaling in phototropism. *Current Opinion in Plant Biology*, 9(5):503–508. (Cited on page 95.)

Klie, M. and Debener, T. (2011). Identification of superior reference genes for data normalisation of expression studies via quantitative PCR in hybrid roses (Rosa hybrida). *BMC Research Notes*, 4(1):518. (Cited on pages 113 and 119.)

Knapp, S., Bohs, L., Nee, M. and Spooner, D. (2004). Solanaceae –a model for linking genomics with biodiversity. *Comparative and Functional Genomics*, 5(3):285–291. (Cited on page 7.)

Koes, R., Spelt, C. and Mol, J. (1989). The chalcone synthase multigene family of Petunia hybrida (V30): differential, light-regulated expression during flower development and UV light induction. *Plant Molecular Biology*, 12(2):213–225. (Cited on page 4.)

Koes, R., Spelt, C., Reif, H., Vandenelzen, P., Veltkamp, E. and Mol, J. (1986a). floral tissue of Petunia hybrida (v30) expresses only one member of the chalcone synthase multigene family. *Nucleic Acids Research*, 14(13):5229–5239. (Cited on page 4.)

Koes, R., Spelt, C., Reif, H., Vandenelzen, P., Veltkamp, E. and Mol, J. (1986b). Floral Tissue of Petunia-Hybrida (V30) Expresses Only One Member of the Chalcone Synthase Multigene Family. *Nucleic Acids Research*, 14(13):5229–5239. (Cited on page 57.)

Köhler, J., Philippi, S. and Lange, M. (2003). SEMEDA: ontology based semantic integration of biological databases. *Bioinformatics*, 19(18):2420–2427. (Cited on page 17.)

Koltunow, A.M., Truettner, J., Cox, K.H., Wallroth, M. and Goldberg, R.B. (1990). Different Temporal and Spatial Gene Expression Patterns Occur during Anther Development. *The Plant Cell*, 2(12):1201–1224. (Cited on page 10.)

Koornneef, M., Van Eden, J., Hanhart, C., Stam, P., Braaksma, F. and Feenstra, W. (1983). Linkage map of Arabidopsis thaliana. *Journal of Heredity*, 74(11):265–272. (Cited on page 18.)

Kotakis, C., Vrettos, N., Kotsis, D., Tsagris, M., Kotzabasis, K. and Kalantidis, K. (2010). Light intensity affects RNA silencing of a transgene in Nicotiana benthamiana plants. *BMC plant biology*, 10(1):220. (Cited on page 14.)

Kramer, E., Dorit, R. and Irish, V. (1998). Molecular Evolution of Genes Controlling Petal and Stamen Development: Duplication and Divergence Within the APETALA3 and PISTILLATA MADS-Box Gene Lineages. *Genetics*, 149(2):765–783. (Cited on page 19.)

Kramer, E. and Irish, V. (2000). Evolution of the petal and stamen developmental programs: evidence from comparative studies of the lower eudicots and basal angiosperms. *International Journal of Plant Sciences*, 161(S6):S29–S40. (Cited on page 80.)

Kramer, E. and Zimmer, E. (2006). Gene duplication and floral developmental genetics of basal eudicots. *Advances in Botanical Research*, 44:353–384. (Cited on page 20.)

Van der Krol, A., Mur, L., Beld, M., Mol, J. and Stuitje, A. (1990). Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *The Plant Cell*, 2(4):291–299. (Cited on pages 8 and 13.)

Kulcheski, F., Muschner, V., Lorenz-Lemke, A., Stehmann, J., Bonatto, S., Salzano, F. and Freitas, L. (2006). Molecular phylogenetic analysis of Petunia juss.(Solanaceae). *Genetica*, 126(1):3–14. (Cited on page 3.)

Kunst, L., Klenz, J., Martinez-Zapater, J. and Haughn, G. (1989). AP2 Gene Determines the Identity of Perianth Organs in Flowers of Arabidopsis thaliana. *The Plant Cell*, 1(12):1195–1208. (Cited on page 6.)

Kuromori, T., Wada, T., Kamiya, A., Yuguchi, M., Yokouchi, T., Imura, Y., Takabe, H., Sakurai, T., Akiyama, K., Hirayama, T. *et al.* (2006). A trial of phenome analysis using 4000 Ds-insertional mutants in gene-coding regions of Arabidopsis. *The Plant Journal*, 47(4):640–651. (Cited on page 97.)

Lally, D., Ingmire, P., Tong, H. and He, Z. (2001). Antisense expression of a cell wall–associated protein kinase, WAK4, inhibits cell elongation and alters morphology. *The Plant Cell*, 13(6):1317–1332. (Cited on page 98.)

Lee, B., Hong, T., Byun, S., Woo, T. and Choi, Y. (2007). ESTpass: a web-based server for processing and annotating expressed sequence tag (EST) sequences. *Nucleic Acids Research*, 35(suppl 2):W159–W162. (Cited on page 26.)

Lee, R. (1977). Effects of organic acids on the loss of ions from barley roots. *Journal of Experimental Botany*, 28(3):578–587. (Cited on page 94.)

Lefever, S., Hellemans, J., Pattyn, F., Przybylski, D., Taylor, C., Geurts, R., Untergasser, A., Vandesompele, J. *et al.* (2009). RDML: structured

language and reporting guidelines for real-time quantitative PCR data. *Nucleic Acids Research*, 37(7):2065. (Cited on page 10.)

Levin, J., de Framond, A., Tuttle, A., Bauer, M. and Heifetz, P. (2000). Methods of double-stranded RNA-mediated gene inactivation in Arabidopsis and their use to define an essential gene in methionine biosynthesis. *Plant Molecular Biology*, 44(6):759–775. (Cited on page 13.)

Li, S., Lauri, A., Ziemann, M., Busch, A., Bhave, M. and Zachgo, S. (2009). Nuclear activity of ROXY1, a glutaredoxin interacting with TGA factors, is required for petal development in Arabidopsis thaliana. *The Plant Cell*, 21(2):429–441. (Cited on pages 20 and 80.)

Liu, F., Jenssen, T., Trimarchi, J., Punzo, C., Cepko, C., Ohno-Machado, L., Hovig, E. and Kuo, W. (2007). Comparison of hybridization-based and sequencing-based gene expression technologies on biological replicates. *BMC Genomics*, 8(1):153. (Cited on page 9.)

Liu, W. and Saint, D. (2002). A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. *Analytical Biochemistry*, 302(1):52–59. (Cited on pages 42 and 61.)

Livak, K. and Schmittgen, T. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2 (-Delta Delta C (T)) method. *Methods*, 25(4):402–408. (Cited on pages 40 and 41.)

Lloyd, J. and Meinke, D. (2012). A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis thaliana. *Plant Physiology*. (Cited on pages 2, 92, 107 and 108.)

Luscombe, N., Greenbaum, D. and Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine*, 40(4):346–358. (Cited on page 16.)

Luu-The, V., Paquet, N., Calvo, E. and Cumps, J. (2005). Improved real-time RT-PCR method for high-throughput measurements using second derivative calculation and double correction. *Biotechniques*, 38(2):287–293. (Cited on page 40.)

Mallona, I., Egea-Cortines, M. and Weiss, J. (2011a). Conserved and Divergent Rhythms of Crassulacean Acid Metabolism-Related and Core

Clock Gene Expression in the Cactus Opuntia ficus-indica. *Plant Physiology*, 156(4):1978–1989. (Cited on pages 27, 34, 36 and 43.)

Mallona, I., Lischewski, S., Weiss, J., Hause, B. and Egea-Cortines, M. (2010). Validation of reference genes for quantitative real-time PCR during leaf and flower development in Petunia hybrida. *BMC Plant Biology*, 10(1):4. (Cited on pages 42, 43, 53, 65, 84, 103 and 106.)

Mallona, I., Weiss, J. and Egea-Cortines, M. (2011b). pcrEfficiency: a Web tool for PCR amplification efficiency prediction. *BMC Bioinformatics*, 12(1):404. (Cited on pages 32 and 40.)

Mallory, A., Bartel, D. and Bartel, B. (2005). MicroRNA-directed regulation of Arabidopsis AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. *The Plant Cell*, 17(5):1360–1375. (Cited on pages 20 and 80.)

Manchado-Rojo, M., Weiss, J. and Egea-Cortines, M. (2008). Using 23 rDNA to identify contaminations of Escherichia coli in Agrobacterium tumefaciens cultures. *Analytical Biochemistry*, 372:253–254. (Cited on pages 43 and 60.)

Mandel, M., Gustafson-Brown, C., Savidge, B. and Yanofsky, M. (1992). Molecular characterization of the Arabidopsis floral homeotic gene APETALA1. *Nature*, 360(6401):273–277. (Cited on page 6.)

Maniatis, T. (1989). *Molecular cloning: a laboratory manual/J. Sambrook, EF Fritsch, T. Maniatis*. New York: Cold Spring Harbor Laboratory Press. (Cited on pages 123 and 124.)

Mann, T., Humbert, R., Dorschner, M., Stamatoyannopoulos, J. and W.S.Noble (2009). A thermodynamic approach to PCR primer design. *Nucleic Acids Research*, 37(13):e95–. (Cited on page 48.)

Marmur, J. and Doty, P. (1962). Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *Journal of Molecular Biology*, 5:109. (Cited on page 41.)

Martí, C., Orzáez, D., Ellul, P., Moreno, V., Carbonell, J. and Granell, A. (2007). Silencing of DELLA induces facultative parthenocarpy in tomato fruits. *The Plant Journal*, 52(5):865–876. (Cited on pages 20 and 80.)

Martienssen, R. (1998). Functional genomics: probing plant gene function and expression with transposons. *Proceedings of the National Academy of Sciences*, 95(5):2021. (Cited on page 92.)

Matzke, M., Matzke, A. and Kooter, J. (2001). RNA: guiding gene silencing. *Science*, 293(5532):1080. (Cited on pages 13 and 14.)

Maxam, A. and Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, 74(2):560. (Cited on page 12.)

McClintock, B. (1953). Induction of instability at selected loci in maize. *Genetics*, 38(6):579. (Cited on page 2.)

McClintock, B. (1983). The significance of responses of the genome to challenge. *Physiology Or Medicine Literature Peace Economic Sciences*, p. 180. (Cited on page 9.)

Meinke, D., Meinke, L., Showalter, T., Schissel, A., Mueller, L. and Tzafrir, I. (2003). A sequence-based map of Arabidopsis genes with mutant phenotypes. *Plant Physiology*, 131(2):409–418. (Cited on page 108.)

Meissner, R., Jin, H., Cominelli, E., Denekamp, M., Fuertes, A., Greco, R., Kranz, H., Penfield, S., Petroni, K., Urzainqui, A. *et al.* (1999). Function search in a large transcription factor gene family in Arabidopsis: assessing the potential of reverse genetics to identify insertional mutations in R2R3 MYB genes. *The Plant Cell*, 11(10):1827–1840. (Cited on page 107.)

Melzer, R. and Theißen, G. (2009). Reconstitution of 'floral quartets' in vitro involving class B and class E floral homeotic proteins. *Nucleic Acids Research*, 37(8):2723–2736. (Cited on page 6.)

Merali, Z. (2010). Computational science: Error, why scientific programming does not compute. *Nature*, 467(7317):775–777. (Cited on page 16.)

Meyerowitz, E. (1994). *Arabidopsis*, volume 27. Cold Spring Harbor Laboratory Pr. (Cited on page 18.)

Meyerowitz, E., Smyth, D. and Bowman, J. (1989). Abnormal flowers and pattern formation in floral. *Development*, 106(2):209. (Cited on page 18.)

Mitchell, A., Hanson, M., Skvirsky, R. and Ausubel, F. (1980). Anther culture of Petunia: Genotypes with high frequency of callus, root, or plantlet formation. *Zeitschrift für Pflanzenzuchtung*, 100:131–146. (Cited on pages 4, 56, 61 and 84.)

Mondragón-Jacobo, C. and Pérez-González, S. (2001). *Cactus (Opuntia spp.) as forage*, volume 169. FAO. (Cited on page 34.)

Monod, J. (1949). The growth of bacterial cultures. *Annual Reviews in Microbiology*, 3(1):371–394. (Cited on pages 10 and 40.)

Morrison, T., Weis, J. and Wittwer, C. (1998). Quantification of low-copy transcripts by continuous SYBR® Green I monitoring during amplification. *Biotechniques*, 24(6):954–962. (Cited on page 40.)

Murashige, T. and Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia plantarum*, 15(3):473–497. (Cited on page 126.)

Myouga, F., Hosoda, C., Umezawa, T., Iizumi, H., Kuromori, T., Motohashi, R., Shono, Y., Nagata, N., Ikeuchi, M. and Shinozaki, K. (2008). A heterocomplex of iron superoxide dismutases defends chloroplast nucleoids against oxidative stress and is essential for chloroplast development in Arabidopsis. *The Plant Cell*, 20(11):3148–3162. (Cited on page 92.)

Nagaraj, S., Deshpande, N., Gasser, R. and Ranganathan, S. (2007a). ESTExplorer: an expressed sequence tag (EST) assembly and annotation platform. *Nucleic Acids Research*, 35(suppl 2):W143–W147. (Cited on page 26.)

Nagaraj, S., Gasser, R. and Ranganathan, S. (2007b). A hitchhiker's guide to expressed sequence tag (EST) analysis. *Briefings in Bioinformatics*, 8(1):6. (Cited on pages 10 and 26.)

Nakamura, Y., Gojobori, T. and Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic acids research*, 28(1):292–292. (Cited on page 26.)

Napoli, C., Lemieux, C. and Jorgensen, R. (1990). Introduction of a chimeric chalcone synthase gene into petunia results in reversible cosuppression of homologous genes in trans. *The Plant Cell*, 2(4):279–289. (Cited on pages 8 and 13.)

Nasir, K., Takahashi, Y., Ito, A., Saitoh, H., Matsumura, H., Kanzaki, H., Shimizu, T., Ito, M., Fujisawa, S., Sharma, P. *et al.* (2005). High-throughput in planta expression screening identifies a class II ethylene-responsive element binding factor-like protein that regulates plant cell death and non-host resistance. *The Plant Journal*, 43(4):491–505. (Cited on page 107.)

Nath, U., Crawford, B., Carpenter, R. and Coen, E. (2003). Genetic control of surface curvature. *Science's STKE*, 299(5611):1404. (Cited on page 98.)

NCAR, R.A.P. (2010). *verification: Forecast verification utilities.* R package version 1.31. (Cited on page 42.)

Nelson, D. and Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. *The Plant Journal*, 66(1):194–211. (Cited on page 109.)

Nelson, D. (2006). Plant cytochrome P450s from moss to poplar. *Phytochemistry Reviews*, 5(2):193–204. (Cited on page 109.)

Nicot, N., Hausman, J., Hoffmann, L. and Evers, D. (2005). Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *Journal of Experimental Botany*, 56(421):2907–2914. (Cited on pages 56, 61, 75 and 76.)

Nishihara, M. and Nakatsuka, T. (2010). Genetic engineering of novel flower colors in floricultural plants: Recent advances via transgenic approaches. *Methods in Molecular Biology*, 589(Part 2):325–347. (Cited on page 8.)

Nobel, P. and Israel, A. (1994). Cladode development, environmental responses of CO2 uptake, and productivity for Opuntia ficus-indica under elevated CO2. *Journal of Experimental Botany*, 45(3):295. (Cited on page 34.)

Olsvik, P., Lie, K., Jordal, A., Nilsen, T. and Hordvik, I. (2005). Evaluation of potential reference genes in real-time RT-PCR studies of Atlantic salmon. *BMC Molecular Biology*, 6(1):21. (Cited on page 76.)

Otsu, N. *et al.* (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on systems, Man, and Cybernetics*, 9(1):62–66. (Cited on page 85.)

Paolacci, A., Tanzarella, O., Porceddu, E. and Ciaffi, M. (2009). Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. *BMC Molecular Biology*, 10(1):11. (Cited on page 76.)

Papanicolaou, A., Stierli, R. *et al.* (2009). Next generation transcriptomes for next generation genomes using est2assembly. *BMC bioinformatics*, 10(1):447. (Cited on page 26.)

Pelaz, S., Ditta, G., Baumann, E., Wisman, E. and Yanofsky, M. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature*, 405(6783):200–203. (Cited on pages 6 and 19.)

Pelaz, S., Gustafson-Brown, C., Kohalmi, S., Crosby, W. and Yanofsky, M. (2001). APETALA1 and SEPALLATA3 interact to promote flower development. *The Plant Journal*, 26(4):385–394. (Cited on page 6.)

Pfaffl, M., Horgan, G. and Dempfle, L. (2002). Relative expression software tool (REST(C)) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Research*, 30(9):e36. (Cited on pages 56 and 85.)

Pfaffl, M., Tichopad, A., Prgomet, C. and Neuvians, T. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. *Biotechnology Letters*, 26(6):509–515. (Cited on pages 56 and 76.)

Pierce, K., Sanchez, J., Rice, J. and Wangh, L. (2005). Linear-After-The-Exponential (LATE)-PCR: Primer design criteria for high yields of specific single-stranded DNA and improved real-time detection. *Proceedings of the National Academy of Sciences of the United States of America*, 102(24):8609. (Cited on page 49.)

Pihur, V., Datta, S. and Datta, S. (2009). RankAggreg, an R package for weighted rank aggregation. *BMC Bioinformatics*, 10(1):62. (Cited on pages 55, 61 and 63.)

Pinkava, D. (2002). On the evolution of the continental North American Opuntioideae. *Studies in the Opuntioideae (Cactaceae). Milborne Port, Dorset: David Hunt (Succulent plant research*, 6:59–98. (Cited on page 34.)

Pinyopich, A., Ditta, G., Savidge, B., Liljegren, S., Baumann, E., Wisman, E. and Yanofsky, M. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature*, 424(6944):85–88. (Cited on page 6.)

Pnueli, L., Carmel-Goren, L., Hareven, D., Gutfinger, T., Alvarez, J., Ganal, M., Zamir, D. and Lifschitz, E. (1998). The SELF-PRUNING gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of CEN and TFL1. *Development*, 125(11):1979–1989. (Cited on page 3.)

Qu, W., Shen, Z., Zhao, D., Yang, Y. and Zhang, C. (2009). MFEprimer: multiple factor evaluation of the specificity of PCR primers. *Bioinformatics*, 25(2):276. (Cited on page 40.)

Quattrocchio, F., Wing, J., Leppen, H., Mol, J. and Koes, R. (1993). Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes. *The Plant Cell*, 5(11):1497–1512. (Cited on page 4.)

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. (Cited on page 42.)

Ramakers, C., Ruijter, J., Deprez, R. and Moorman, A. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *NeuroScience Letters*, 339(1):62–66. (Cited on page 40.)

Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J. *et al.* (2001). Multiclass cancer diagnosis using tumor gene expression signatures. *Proceedings of the National Academy of Sciences*, 98(26):15149. (Cited on page 17.)

Ratanasut, K., Wongkhamprai, B. and Maknoi, S. (2011). Expression of a CYP76AB1 correlates with the sequential white-blue-white colour transition of Vanda coerulea petals. *Biologia Plantarum*, 55(2):353–356. (Cited on page 94.)

Raymond, E. (1999). The cathedral and the bazaar. *Knowledge, Technology & Policy*, 12(3):23–49. (Cited on page 16.)

Re, A.D. (2010). *compute.es: Compute Effect Sizes*. R package version 0.2. (Cited on page 42.)

Reale, L., Porceddu, A., Lanfaloni, L., Moretti, C., Zenoni, S., Pezzotti, M., Romano, B. and Ferranti, F. (2002). Patterns of cell division and expansion in developing petals of Petunia hybrida. *Sexual Plant Reproduction*, 15(3):123–132. (Cited on page 73.)

Reboul, J., Vaglio, P., Rual, J., Lamesch, P., Martinez, M., Armstrong, C., Li, S., Jacotot, L., Bertin, N. *et al.* (2003). C. elegans ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nature Genetics*, 34(1):35–41. (Cited on page 21.)

Reid, K., Olsson, N., Schlosser, J., Peng, F. and Lund, S. (2006). An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC Plant Biology*, 6:27. (Cited on pages 56 and 75.)

Rhee, S., Beavis, W., Berardini, T., Chen, G., Dixon, D., Doyle, A., Garcia-Hernandez, M., Huala, E., Lander, G., Montoya, M. *et al.* (2003). The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Research*, 31(1):224–228. (Cited on page 82.)

Rice, P., Longden, I. and et al., A.B. (2000). EMBOSS: the European molecular biology open software suite. *Trends in Genetics*, 16(6):276–277. (Cited on pages 27 and 48.)

Rijpkema, A., Gerats, T. and Vandenbussche, M. (2006a). Genetics of Floral Development in Petunia. *Advances in Botanical Research*, 44:237–278. (Cited on page 3.)

Rijpkema, A., Royaert, S., Zethof, J., van der Weerden, G., Gerats, T. and Vandenbussche, M. (2006b). Analysis of the Petunia TM6 MADS box gene reveals functional divergence within the DEF/AP3 lineage. *The Plant Cell*, 18(8):1819–1832. (Cited on page 80.)

Ritz, C. and Spiess, A. (2008). qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics*, 24(13):1549. (Cited on pages 42, 84 and 85.)

Robertson, D. (2004). VIGS vectors for gene silencing: many targets, many tools. *Annu. Rev. Plant Biol.*, 55:495–519. (Cited on page 13.)

Robinson, A., Love, C., Batley, J., Barker, G. and Edwards, D. (2004). Simple sequence repeat marker loci discovery using SSR primer. *Bioinformatics*, 20(9):1475. (Cited on pages 26 and 29.)

Rounsley, S. and Last, R. (2010). Shotguns and SNPs: how fast and cheap sequencing is revolutionizing plant biology. *The Plant Journal*, 61(6):922–927. (Cited on page 2.)

Rozen, S. and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology*, 132(3):365–386. (Cited on pages 48 and 83.)

Rual, J., Hirozane-Kishikawa, T., Hao, T., Bertin, N., Li, S., Dricot, A., Li, N., Rosenberg, J., Lamesch, P., Vidalain, P. *et al.* (2004). Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Research*, 14(10b):2128. (Cited on page 21.)

Saenz, C. (2000). Processing technologies: an alternative for cactus pear (¡ i¿ Opuntia¡/i¿ spp.) fruits and cladodes. *Journal of Arid Environments*, 46(3):209–225. (Cited on page 34.)

Saenz, C., Sepulveda, E. and Matsuhiro, B. (2004). Opuntia spp mucilage's: a functional component with industrial perspectives. *Journal of arid environments*, 57(3):275–290. (Cited on page 34.)

Saiki, R., Gelfand, D., Stoffel, S., Scharf, S., Higuchi, R., Horn, G., Mullis, K. and Erlich, H. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239(4839):487–491. (Cited on pages 9 and 12.)

Sanger, F., Nicklen, S. and Coulson, A. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12):5463. (Cited on page 12.)

Sayed, O. (2001). Crassulacean acid metabolism 1975–2000, a check list. *Photosynthetica*, 39(3):339–352. (Cited on page 36.)

Schefe, J., Lehmann, K., Buschmann, I., Unger, T. and Funke-Kaiser, H. (2006). Quantitative real-time RT-PCR data analysis: current concepts

and the novel "gene expression's CT difference" formula. *Journal of Molecular Medicine*, 84(11):901–910. (Cited on page 40.)

Schmidt, N., Merker, M. and Becker, D. (2012). Novel high-throughput RNAi vectors for plant biotechnology. *Plant Breeding*. (Cited on page 15.)

Schulze, A. and Downward, J. (2001). Navigating gene expression using microarrays-a technology review. *Nature Cell Biology*, 3(8):190–195. (Cited on page 11.)

Schwarz-Sommer, Z., Hue, I., Huijser, P., Flor, P., Hansen, R., Tetens, F., Lonnig, W., Saedler, H. and Sommer, H. (1992). Characterization of the Antirrhinum floral homeotic MADS-box gene deficiens: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. *The EMBO Journal*, 11(1):251–263. (Cited on page 6.)

Sclep, G., Allemeersch, J., Liechti, R., De Meyer, B., Beynon, J., Bhalerao, R., Moreau, Y., Nietfeld, W., Renou, J., Reymond, P. *et al.* (2007). CATMA, a comprehensive genome-scale resource for silencing and transcript profiling of Arabidopsis genes. *BMC Bioinformatics*, 8(1):400. (Cited on page 14.)

Sheahan, J., Cheong, H. and Rechnitz, G. (1998). The colorless flavonoids of Arabidopsis thaliana (Brassicaceae). I. A model system to study the orthodihydroxy structure. *American journal of botany*, 85(4):467–467. (Cited on page 109.)

Sindelka, R., Ferjentsik, Z. and Jonak, J. (2006). Developmental expression profiles of Xenopus laevis reference genes. *Developmental Dynamics*, 235(3):11. (Cited on page 76.)

Sing, T., Sander, O., Beerenwinkel, N. and Lengauer, T. (2009). *ROCR: Visualizing the performance of scoring classifiers.* R package version 1.0-4. (Cited on page 42.)

Sink, K. (1984). Petunia. *Monographs on Theoretical and Applied Genetics*, 9. (Cited on page 8.)

Sirim, D., Wagner, F., Lisitsa, A. and Pleiss, J. (2009). The Cytochrome P450 Engineering Database: integration of biochemical properties. *BMC Biochemistry*, 10(1):27. (Cited on page 83.)

Smaczniak, C., Immink, R., Muiño, J., Blanvillain, R., Busscher, M., Busscher-Lange, J., Dinh, Q., Liu, S., Westphal, A., Boeren, S. *et al.* (2012). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proceedings of the National Academy of Sciences*, 109(5):1560–1565. (Cited on page 94.)

Small, I. (2007). RNAi for revealing and engineering plant gene functions. *Current Opinion in Biotechnology*, 18(2):148–153. (Cited on page 81.)

Smith, H., Wilcox, K. *et al.* (1970). A restriction enzyme from Haemophilus influenzae. I. Purification and general properties. *Journal of Molecular Biology*, 51(2):379. (Cited on page 12.)

Smith, N., Singh, S., Wang, M., Stoutjesdijk, P., Green, A. and Waterhouse, P. (2000). Gene expression: Total silencing by intron-spliced hairpin RNAs. *Nature*, 407(6802):319–320. (Cited on pages 13 and 15.)

Smyth, D.R. (2005). Morphogenesis of Flowers: Our Evolving View. *The Plant Cell*, 17(2):330–341. (Cited on pages 18 and 19.)

Smyth, D., Bowman, J. and Meyerowitz, E. (1990). Early flower development in Arabidopsis. *The Plant Cell*, 2(8):755–767. (Cited on pages 81 and 86.)

Soltis, D., Senters, A., Zanis, M., Kim, S., Thompson, J., Soltis, P., Ronse De Craene, L., Endress, P. and Farris, J. (2003). Gunnerales are sister to other core eudicots: implications for the evolution of pentamery. *American Journal of Botany*, 90(3):461–470. (Cited on page 20.)

Somerville, C. and Koornneef, M. (2002). A fortunate choice: the history of Arabidopsis as a model plant. *Nature Reviews Genetics*, 3(11):883–889. (Cited on page 2.)

Sommer, H., Beltran, J., Huijser, P., Pape, H., Lonnig, W., Saedler, H. and Schwarz-Sommer, Z. (1990). Deficiens, a homeotic gene involved in the control of flower morphogenesis in Antirrhinum majus: the protein shows homology to transcription factors. *The EMBO Journal*, 9(3):605–613. (Cited on pages 6 and 19.)

Souer, E., van der Krol, A., Kloos, D., Spelt, C., Bliek, M., Mol, J. and Koes, R. (1998). Genetic control of branching pattern and floral identity during Petunia inflorescence development. *Development*, 125(4):733–742. (Cited on page 4.)

Spitzer, B., Ben Zvi, M., Ovadis, M., Marhevka, E., Barkai, O., Edelbaum, O., Marton, I., Masci, T., Alon, M. and Morin, S. (2007a). Reverse genetics of floral scent: Application of tobacco rattle virus-based gene silencing in Petunia. *Plant Physiology*, 145(4):1241–1250. (Cited on page 57.)

Spitzer, B., Zvi, M., Ovadis, M., Marhevka, E., Barkai, O., Edelbaum, O., Marton, I., Masci, T., Alon, M., Morin, S. *et al.* (2007b). Reverse genetics of floral scent: application of tobacco rattle virus-based gene silencing in petunia. *Plant Physiology*, 145(4):1241–1250. (Cited on page 107.)

Srivastava, G., Guo, J., Shi, H. and Xu, D. (2008). PRIMEGENS-v2: genome-wide primer design for analyzing DNA methylation patterns of CpG islands. *Bioinformatics*, 24(17):1837. (Cited on page 83.)

Stajich, J., Block, D., Boulez, K., Brenner, S., Chervitz, S., Dagdigian, C., Fuellen, G., Gilbert, J., Korf, I., Lapp, H. *et al.* (2002). The Bioperl toolkit: Perl modules for the life sciences. *Genome Research*, 12(10):1611. (Cited on page 16.)

Steeves, T. and Sussex, I. (1989). *Patterns in plant development*. Cambridge University Press. (Cited on pages 5 and 6.)

Stehmann, J., Lorenz-Lemke, A., Freitas, L. and Semir, J. (2009). The genus Petunia. *Petunia*, pp. 1–28. (Cited on page 3.)

Stein, L. (2002). Creating a bioinformatics nation. *Nature*, 417(6885):119–120. (Cited on page 16.)

Stein, L. *et al.* (2003). Integrating biological databases. *Nature Reviews Genetics*, 4(5):337–345. (Cited on pages 18 and 33.)

Stevens, R., Goble, C., Baker, P. and Brass, A. (2001). A classification of tasks in bioinformatics. *Bioinformatics*, 17(2):180–188. (Cited on page 17.)

Stinson, B. (2001). *PostgreSQL essential reference*. Sams. (Cited on page 32.)

Stubbe, H. (1966). Genetik und Zytologie von Antirrhinum L. sect. *Antirrhinum. Jena, Germany: VEB Gustav Fischer Verlag.* (Cited on page 18.)

Stuurman, J., Hoballah, M., Broger, L., Moore, J., Basten, C. and Kuhlemeier, C. (2004). Dissection of floral pollination syndromes in Petunia. *Genetics*, 168(3):1585. (Cited on pages 3 and 8.)

Suzuki, T., Higgins, P. and Crawford, D. (2000). Control selection for RNA quantitation. *Biotechniques*, 29:332–337. (Cited on page 76.)

Szécsi, J., Joly, C., Bordji, K., Varaud, E., Cock, J., Dumas, C. and Bendahmane, M. (2006). BIGPETALp, a bHLH transcription factor is involved in the control of Arabidopsis petal size. *The EMBO Journal*, 25(16):3912–3920. (Cited on pages 20 and 80.)

Tamaki, K., Imaishi, H., Ohkawa, H., Oono, K. and Sugimoto, M. (2005). Cloning, Expression in Yeast, and Functional Characterization of CYP76A4, a Novel Cytochrome P450 of Petunia That Catalyzes (!-1)-Hydroxylation of Lauric Acid. *Bioscience, Biotechnolology and Biochemistry*, 69(2):406–409. (Cited on page 94.)

Tang, J., Vosman, B., Voorrips, R., Van Der Linden, C. and Leunissen, J. (2006). QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species. *BMC Bioinformatics*, 7(1):438. (Cited on page 29.)

Taylor, R. (2007). PostgreSQL autodoc. (Cited on pages 30 and 31.)

Thareau, V., Déhais, P., Serizet, C., Hilson, P., Rouzé, P. and Aubourg, S. (2003). Automatic design of gene-specific sequence tags for genome-wide functional studies. *Bioinformatics*, 19(17):2191–2198. (Cited on page 14.)

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, 408(6814):796. (Cited on pages 12 and 18.)

The Petunia Platform (2004). Petunia Platform. `http://www.petuniaplatform.net/`. (Cited on pages 8 and 114.)

Theißen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J., Münster, T., Winter, K. and Saedler, H. (2000). A short history of MADS-box genes in plants. *Plant Molecular Biology*, 42(1):115–149. (Cited on page 19.)

Theißen, G., Kim, J. and Saedler, H. (1996). Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box

gene subfamilies in the morphological evolution of eukaryotes. *Journal of Molecular Evolution*, 43(5):484–516. (Cited on page 19.)

Theißen, G. and Saedler, H. (2001). Plant biology. Floral quartets. *Nature*, 409(6819):469–471. (Cited on page 6.)

Tichopad, A., Dilger, M., Schwarz, G. and Pfaffl, M. (2003). Standardized determination of real-time PCR efficiency from a single reaction set-up. *Nucleic Acids Research*, 31(20):e122. (Cited on page 40.)

Till, B., Reynolds, S., Greene, E., Codomo, C., Enns, L., Johnson, J., Burtner, C., Odden, A., Young, K., Taylor, N. *et al.* (2003). Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Research*, 13(3):524–530. (Cited on page 15.)

Toguri, T., Kobayashi, O. and Umemoto, N. (1993). The cloning of eggplant seedling cDNAs encoding proteins from a novel cytochrome P-450 family (CYP76). *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 1216(1):165–169. (Cited on page 94.)

Tornielli, G., Koes, R. and Quattrocchio, F. (2009). The genetics of flower color. *Petunia*, pp. 269–299. (Cited on page 8.)

Trobner, W., Ramirez, L., Motte, P., Hue, I., Huijser, P., Lonnig, W., Saedler, H., Sommer, H. and Schwarz-Sommer, Z. (1992). GLOBOSA: a homeotic gene which interacts with DEFICIENS in the control of Antirrhinum floral organogenesis. *The EMBO Journal*, 11(13):4693–4704. (Cited on pages 6 and 20.)

van Tunen, A., Mur, L., Brouns, G., Rienstra, J., Koes, R. and Mol, J. (1990). Pollen-and anther-specific chi promoters from petunia: tandem promoter regulation of the chiA gene. *The Plant Cell*, 2(5):393–401. (Cited on page 4.)

Tuomi, J., Voorbraak, F., Jones, D. and Ruijter, J. (2010). Bias in the Cq value observed with hydrolysis probe based quantitative PCR can be corrected with the estimated PCR efficiency value. *Methods*, 50(4):313–322. (Cited on page 49.)

Van Moerkercke, A., Galván-Ampudia, C., Verdonk, J., Haring, M. and Schuurink, R. (2012). Regulators of floral fragrance production and their target genes in petunia are not exclusively active in the epidermal cells of petals. *Journal of Experimental Botany*. (Cited on page 4.)

Vandenbussche, M., Janssen, A., Zethof, J., Van Orsouw, N., Peters, J., Van Eijk, M., Rijpkema, A., Schneiders, H., Santhanam, P., De Been, M. *et al.* (2008). Generation of a 3D indexed Petunia insertion database for reverse genetics. *The Plant Journal*, 54(6):1105–1114. (Cited on pages 8, 15, 116 and 121.)

Vandenbussche, M., Zethof, J., Royaert, S., Weterings, K. and Gerats, T. (2004). The duplicated B-class heterodimer model: whorl-specific effects and complex genetic interactions in Petunia hybrida flower development. *The Plant Cell*, 16(3):741–754. (Cited on pages 8 and 80.)

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N. and De Paepe, A. (2002). Accurate normalization of realtime quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, 3:research0034.1–research0034.11. (Cited on pages 56, 62, 68, 75 and 77.)

Varadaraj, K. and Skinner, D. (1994). Denaturants or cosolvents improve the specificity of PCR amplification of a G+ C-rich DNA using genetically engineered DNA polymerases. *Gene(Amsterdam)*, 140(1):1–5. (Cited on page 49.)

Varshney, R. and Tuberosa, R. (2007). Genomics-assisted crop improvement: an overview. *Genomics-assisted Crop Improvement*, pp. 1–12. (Cited on page 15.)

Velázquez, K., Renovell, A., Comellas, M., Serra, P., García, M., Pina, J., Navarro, L., Moreno, P. and Guerri, J. (2010). Effect of temperature on RNA silencing of a negative-stranded RNA plant virus: Citrus psorosis virus. *Plant Pathology*, 59(5):982–990. (Cited on page 14.)

Wallace, R., Shaffer, J., Murphy, R., Bonner, J., Hirose, T. and Itakura, K. (1979). Hybridization of synthetic oligodeoxyribonucleotides to ΦX 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Research*, 6(11):3543. (Cited on page 41.)

Wang, M.B. and Waterhouse, P.M. (2002). Application of gene silencing in plants. *Current Opinion in Plant Biology*, 5(2):146 –150. (Cited on page 13.)

Wanntorp, L. and De Craene, L. (2007). Flower development of Meliosma (Sabiaceae): Evidence for multiple origins of pentamery in the eudicots. *American Journal of Botany*, 94(11):1828–1836. (Cited on page 20.)

Waterhouse, P., Graham, M. and Wang, M. (1998). Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proceedings of the National Academy of Sciences*, 95(23):13959. (Cited on page 15.)

Weigel, D. (1995). The APETALA2 domain is related to a novel type of DNA binding domain. *The Plant Cell*, 7(4):388. (Cited on page 19.)

Weigel, D. and Meyerowitz, E. (1994). The ABCs of floral homeotic genes. *Cell*, 78(2):203–209. (Cited on page 6.)

Weiss, J. and Egea-Cortines, M. (2009). Transcriptomic analysis of cold response in tomato fruits identifies dehydrin as a marker of cold stress. *Journal of Applied Genetics*, 50(4):311–319. (Cited on page 43.)

Wellesen, K., Durst, F., Pinot, F., Benveniste, I., Nettesheim, K., Wisman, E., Steiner-Lange, S., Saedler, H. and Yephremov, A. (2001). Functional analysis of the LACERATA gene of Arabidopsis provides evidence for different roles of fatty acid ω-hydroxylation in development. *Proceedings of the National Academy of Sciences*, 98(17):9694. (Cited on pages 106 and 109.)

Weng, H., Molina, I., Shockey, J. and Browse, J. (2010). Organ fusion and defective cuticle function in a lacs1lacs2 double mutant of Arabidopsis. *Planta*, 231(5):1089–1100. (Cited on page 109.)

Wijsman, H. (1983). On the interrelationships of certain species of Petunia: 2. Experimental data: crosses between different taxa. *Acta Botanica Neerlandica*, 32(1/2):97–107. (Cited on page 3.)

Witten, I., Frank, E. and Hall, M. (2011). *Data Mining: Practical Machine Learning tools and techniques*. Morgan Kaufmann. (Cited on page 16.)

Wollmann, H., Mica, E., Todesco, M., Long, J.A. and Weigel, D. (2010). On reconciling the interactions between APETALA2, miR172 and AGAMOUS with the ABC model of flower development. *Development*, 137(21):3633–3642. (Cited on page 20.)

Wood, S. (2001). mgcv: GAMs and generalized ridge regression for R. *Future*, 1:20. (Cited on pages 42 and 46.)

Wood, S. (2004). Stable and efficient multiple smoothing parameter estima-

tion for generalized additive models. *Journal of the American Statistical Association*, 99(467):673–686. (Cited on page 46.)

Yamaguchi, A., Wu, M., Yang, L., Wu, G., Poethig, R. and Wagner, D. (2009). The MicroRNA-Regulated SBP-Box Transcription Factor SPL3 Is a Direct Upstream Activator of LEAFY, FRUITFULL, APETALA1. *Developmental Cell*, 17(2):268–278. (Cited on page 94.)

Yanofsky, M., Ma, H., Bowman, J., Drews, G., Feldmann, K. and Meyerowitz, E. (1990). The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. *Nature*, 346(6279):35–39. (Cited on pages 6 and 19.)

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L. *et al.* (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research*, 34(Suppl 2):W293. (Cited on page 82.)

Young, R. and Center, N. (2000). Biomedical Discovery Review with DNA Arrays. *Cell*, 102:9–15. (Cited on page 2.)

Yuan, J., Reed, A., Chen, F. and Stewart, C. (2006). Statistical analysis of real-time PCR data. *BMC Bioinformatics*, 7(1):85. (Cited on page 11.)

Zambryski, P., Joos, H., Genetello, C., Leemans, J., Van Montagu, M. and Schell, J. (1983). Ti plasmid vector for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. *The EMBO Journal*, 2(12):2143. (Cited on page 12.)

Zenoni, S., Fasoli, M., Tornielli, G., Dal Santo, S., Sanson, A., de Groot, P., Sordo, S., Citterio, S., Monti, F. and Pezzotti, M. (2011a). Overexpression of PhEXPA1 increases cell size, modifies cell wall polymer composition and affects the timing of axillary meristem development in Petunia hybrida. *New Phytologist*. (Cited on pages 20 and 80.)

Zenoni, S., Reale, L., Tornielli, G., Lanfaloni, L., Porceddu, A., Ferrarini, A., Moretti, C., Zamboni, A., Speghini, S. and Ferranti, F. (2004). Downregulation of the Petunia hybrida alpha-expansin gene PhEXP1 reduces the amount of crystalline cellulose in cell walls and leads to phenotypic changes in petal limbs. *The Plant Cell*, 16(2):295–308. (Cited on pages 20, 62 and 80.)

Zenoni, S., D'Agostino, N., Tornielli, G.B., Quattrocchio, F., Chiusano, M.L., Koes, R., Zethof, J., Guzzo, F., Delledonne, M., Frusciante, L., Gerats, T. and Pezzotti, M. (2011b). Revealing impaired pathways in the an11 mutant by high-throughput characterization of Petunia axillaris and Petunia inflata transcriptomes. *The Plant Journal*, 68(1):11–27. (Cited on pages 113 and 120.)

Zhi-Liang, H., Bao, J. and Reecy, J. (2008). CateGOrizer: a web-based program to batch analyze gene ontology classification categories. *Online Journal of Bioinformatics*, 9:108–112. (Cited on page 82.)