Departamento de Ciencia y Tecnología Agraria
Instituto de Biotecnología Vegetal

Identification of novel genes involved in Petunia flower development using transcript profiling and reverse genetics

Izaskun Mallona González

Universidad Politécnica de Cartagena
2012

Supervisors

Marcos Egea, Ph.D., Universidad Politécnica de Cartagena
Julia Weiss, Ph.D., Universidad Politécnica de Cartagena

Thesis tribunal

Verónica Truniger Rietmann, Ph.D., Chair
Juan Pablo Fernández Trujillo, Ph.D, Secretary
Catalina Egea Gilabert, Ph.D., Member
Juan Capel Salinas, Ph.D., Member
Manuel Acosta Echevarría, Ph.D., Member

Contact information

Institute of Plant Biotechnology
Department of Agrarian Science and Technology
I+D+i Building
Plaza del Hospital, s/n
30202 Cartagena, Spain

Email address: ibv@upct.es
URL: `http://www.upct.es/~ibvupct/`
Telephone: +34 868 07 1066

Cover

Petunia violacea Lindl. from "An Illustrated Flora of the Northern United States, Canada, and the British Possessions" (Britton and Brown, 1913)

Identification of novel genes involved in Petunia flower development using transcript profiling and reverse genetics

Izaskun Mallona González

Genetics
Institute of Plant Biotechnology
Plaza del Hospital, s/n, Cartagena, Spain
izaskun.mallona@upct.es, izaskun.mallona@gmail.com
`http://www.upct.es/~genetica`

Abstract

Petals are a key element on plant life cycle as, in many species, they attract pollinators, thus aiding to reproduction. Furthermore, they have economic importance in ornamental crops. In the present study, petal transcriptional patterns were compared within the flower organs in *Arabidopsis thaliana*. It was found that catalytic molecular functions were overrepresented in petals. A shortlist comprising the top ten differentially expressed genes in petals were mapped to the model species with industrial value *Petunia hybrida*, and further downregulated by RNAi. The silencing phenotypes found permitted to assign functions in petal development to seven novel genes: when silenced, they triggered alterations on flower size and shape (*PhCYP76*, *PhNPH3*, *PhFeSOD*, *PhXTH*, *PhCYP96* and *PhWAK*), petal smoothness (*PhPRA*), color (*PhNPH3* and *PhWAK*) and symmetry (*PhCYP76*). Pleiotropic phenotypes were found, such as changes in root morphology and leaf color (*PhCYP76*), flower number, capsule and seed morphology (*PhCYP96*) and plant height (*PhCYP76* and *PhCYP96*).

To accomplish the experimental design, three methods were developed. First, the "pESTle" management system that assembles, annotates, stores and serves expressed sequence tag data. Second, a reference gene selection for real time PCR experiments that includes a new method for stability estimation based on rank aggregation of published algorithms, and concludes that a normalization factor with two members of *EF1α*, *SAND*, *CYP* or *RAN1* is stable enough under most conditions. And third, a PCR efficiency estimator based on amplicon characteristics which allows efficiency-driven primer design in a Web tool.

General Terms:
petunia, petal, silencing, bioinformatics, Q-PCR, transcriptomics

Additional Key Words and Phrases:
garden petunia, petal development, post-transcriptional gene silencing, reverse genetics, gene expression, polymerase chain reaction, real-time PCR, reference genes, PCR efficiency, expressed sequence tags, EST pipeline

Identification of novel genes involved in Petunia flower development using transcript profiling and reverse genetics

Izaskun Mallona González

Genética
Instituto de Biotecnología Vegetal
Plaza del Hospital, s/n, Cartagena, Spain
izaskun.mallona@upct.es, izaskun.mallona@gmail.com
`http://www.upct.es/~genetica`

Resumen

La corola es un elemento básico en el ciclo de vida vegetal puesto que, en muchos casos, atrae a polinizadores que intervienen en su propagación. Además, posee un valor económico en especies ornamentales. En el presente trabajo se realizó un análisis comparativo contrastando el transcriptoma de pétalos frente al resto de órganos florales de *Arabidopsis thaliana*. Como resultado se observó la preponderancia de genes involucrados en funciones catalíticas. A fin de dilucidar el rol de estos genes, se escogieron aquellos nueve con mayor expresión diferencial y se silenciaron de forma estable mediante ARN de interferencia. De esta forma, el análisis fenotípico ha permitido asignar un papel a siete nuevos genes: al ser silenciados, provocan alteraciones en la forma y tamaño (*PhCYP76*, *PhNPH3*, *PhFeSOD*, *PhXTH*, *PhCYP96* y *PhWAK*), textura (*PhPRA*), color (*PhNPH3* y *PhWAK*) y simetría (*PhCYP76*) de los pétalos; así como el color de las hojas y la morfología radicular (*PhCYP76*), el número de flores y la altura de la planta (*PhCYP76* y *PhCYP96*).

A fin de respaldar el diseño experimental se desarrollaron tres métodos. En primer lugar, el gestor «pESTle», que aúna el ensamblaje, anotación, almacenamiento y difusión de ESTs. En segundo lugar, la selección de genes de referencia para PCR en tiempo real, que reúne un nuevo método para su evaluación basado en la agregación de rangos; y la descripción de la pertinencia de emplear un factor de normalización con dos genes de entre *EF1α*, *SAND*, *CYP* y *RAN1* en la mayor parte de los casos. Finalmente, se produjo el estimador «pcrEfficiency», que proporciona un entorno de diseño de cebadores que simultáneamente predice su eficiencia.

Palabras clave
petunia, petal, silencing, bioinformatics, Q-PCR, transcriptomics

Lemas adicionales
garden petunia, petal development, post-transcriptional gene silencing, reverse genetics, gene expression, polymerase chain reaction, real-time PCR, reference genes, PCR efficiency, expressed sequence tags, EST pipeline

Universidad Politécnica de Cartagena

DT-13

CONFORMIDAD DE DEPOSITO DE TESIS DOCTORAL POR LA COMISIÓN ACADÉMICA DEL PROGRAMA

D/Dª. Francisco Artés Hernández, Presidente/a de la Comisión Académica del Programa "Técnicas Avanzadas en Investigación y Desarrollo Agrario y Alimentario".

INFORMA:

Que la Tesis Doctoral titulada, "Identification of novel genes involved in Petunia flower development using transcript profiling and reverse genetics" ha sido realizada por D/Dª. Izaskun Mallona González, bajo la dirección y supervisión de los doctores Marcos Egea Gutiérrez-Cortines y Julia Weiss, dando su conformidad a la misma la Comisión Académica, con la finalidad de que sea presentada ante la Comisión de Doctorado.

La Rama de conocimiento por la que esta tesis ha sido desarrollada es:

☑ Ciencias
☐ Ciencias Sociales y Jurídicas
☐ Ingeniería y Arquitectura

En Cartagena, a 17 de mayo de 2012

EL PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA

Fdo. Francisco Artés Hernández

COMISIÓN DE DOCTORADO

Conformidad de solicitud de autorización de depósito de tesis doctoral por los directores de la tesis

Dr. Marcos Egea Gutiérrez-Cortines y Dra. Julia Weiss, directores de la tesis doctoral "Identification of novel genes involved in Petunia flower development using transcript profiling and reverse genetics"
INFORMAN:
Que la referida Tesis Doctoral ha sido realizada por Izaskun Mallona González, dando nuestra conformidad para que sea presentada ante la Comisión de Doctorado.
La rama de conocimiento por la que esta tesis ha sido desarrollada es:

- Ciencias

En Cartagena, a de de

Marcos Egea Gutiérrez-Cortines

Julia Weiss

Original Publications of the Thesis

- Chapter 3: Izaskun Mallona, Julia Weiss and Marcos Egea-Cortines (2011). pcrEfficiency: a Web tool for PCR amplification efficiency prediction. *BMC Bioinformatics*, 12:404.
- Chapter 4: Izaskun Mallona, Sandra Lischewski, Julia Weiss, Bettina Hause, and Marcos Egea-Cortines (2010). Validation of reference genes for quantitative real-time PCR during leaf and flower development in *Petunia hybrida*. *BMC Plant Biology*, 10(1):4.
- International Congresses
  - Chapter 4: Izaskun Mallona, Sandra Lischewski, Julia Weiss, Bettina Hause, and Marcos Egea-Cortines (2009). Selection of referece genes for real time RT-PCR in two varieties of petunia (2009). Xth Petunia Days International Congress. Cartagena, Spain.
  - Chapter 4: Izaskun Mallona, Sandra Lischewski, Julia Weiss, Bettina Hause, and Marcos Egea-Cortines (2010). Validation of reference genes for quantitative real-time PCR during leaf and flower development in *Petunia hybrida*. XVII Congress of the Federation of European Societies of Plant Biology. Valencia, Spain.
  - Chapter 5: Izaskun Mallona, Julia Weiss and Marcos Egea-Cortines (2012). Identification of new genes involved in floral development in *Petunia hybrida*. 12th Petunia Days International Congress. Erfurt, Germany.

Contents

List of Figures

List of Tables

Nomenclature

ACT ACTin 11

AG AGamous

AN2 ANthocyanin 2

AN4 ANthocyanin 4

AP APetala

BL BLind

BLAST Basic Local Alignment Search Tool

BLOSUM Blocks Substitution Matrices

cDNA Complementary DeoxyriboNucleic Acid

CGI Common Gateway Interface

Cq Cycle of Quantification

C_T Cycle Threshold

CV Coefficient of Variation

CYP CYcloPhilin

DEF DEFiciens

DNA DeoxyriboNucleic Acid

DNase DeoxyriboNucleASE

dTph1 Defective Transposable element Petunia Hybrida 1

EBI European Bioinformatics Institute

EC Enzyme Commission

EF1α Elongation Factor 1-α

EMBL European Molecular Biology Laboratory

EMBOSS European Molecular Biology Open Software Suite

EST Expressed Sequence Tag

FUL FrUitfuLl

GAM Generalized Additive Model

GAPDH GlycerAldehyde-3-Phosphate DeHydrogenase

GLO GLObosa

GO Gene Ontology

GPL GNU Public License

GST Gene-Specific Tag

HTML Hypertext Markup Language

IQR Interquartile Range

JDBC Java Database Connectivity

KEGG Kyoto Encyclopedia of Genes and Genomes

M1 *P. hybrida* strain W138

MIQE Minimum Information for publication of Quantitative real-time PCR Experiments

mRNA Messenger RNA

M-MuLV Moloney Murine Leukemia Virus

NGS Next Generation Sequencing

NRQ Normalized Relative Quantity

ODBC Open Database Connectivity

ORF Open Reading Frame

PCA Principal Component Analysis

PCR Polymerase Chain Reaction

PDB Protein Data Bank

PFG Petunia Flowering Gene

PhCYP76 Petunia Hybrida CYtochrome P450 member from superfamily 76

PhCYP96 Petunia Hybrida CYtochrome P450 member from superfamily 96

pESTle Pipeline for Expressed Sequence Tags Local Exploit)

PhFeSOD Petunia Hybrida Fe Superoxide Dismutase-like

PhNPH3 Petunia Hybrida Non-Phototropic Hypocotyl 3-like

PhPRA Petunia Hybrida Prenilated Rab Acceptor-like

PhRBK Petunia Hybrida Rop Binding protein Kinases-like

PhTCP Petunia Hybrida TCP-like

PhXTH Petunia Hybrida Xyloglucan endo-Transglycosylase Hydrolase-like

PhWAK Petunia Hybrida Wall-Associated Kinase-like

PIR/PSD Protein Information Resource / Protein Sequence Database

PI PIstillata

PR Precision and Recall

PTGS Post-Transcriptional Gene Silencing

PV Pairwise Variation

Q-PCR Quantitative PCR

RAN1 RAs-related Nuclear protein 1

RDML Real-time pcr Data Markup Language

RDMS Relational Database Management System

REST Relative Expression Software Tool

RG Reference Gene

RNA Ribonucleic Acid

RNAi RNA interference

RNase RiboNucleASE

ROC Receiver Operator Characteristic

RPS13 Ribosomal Potein S13

RQ Relative Quantity

RT-PCR Real Time PCR

SAM Shoot Apical Meristem

SAND SAND family protein

SEP SEPallata

SGN Sol Genomics Network

SNP Single Nucleotide Polymorphism

SQL Structured Query Language

SSR Short Sequence Repeat

T-DNA Transfer DNA

Ti Tumor-inducing plasmid

TM6 Tomato Mads box gene 6

TIGR The Institute for Genomic Research

TILLING Targeting Induced Local Lesions in Genomes

Tm Temperature of Melting

TUB β-Tubulin 6 chain

UBQ Polyubiquitin

V30 *P. hybrida* Violet 30

VIGS Virus-Induced Gene Silencing

W115 *P. hybrida* Mitchell diploid

W138 *P. hybrida* M1 line

WT Wild Type

WWW World Wide Web

Chapter 1

Gene function discovery through transcript profiling

Projects for gene function elucidation by reverse genetics are outnumbering the classical forward approaches. The raise of cheap sequencing projects has supported the gene function identification by its overexpression or silencing. Genome and transcriptome databases, linked to the microarray platforms, have permitted the *–omics* mining in any imaginable condition, such as plant-pathogen interaction, nutrient deficits, time course or anatomical parts.

This approach requires model organisms satisfying some requirements, such as sequence availability and genetic manipulation capabilities, like easy transformation or tag-identified mutagenesis. *Petunia hybrida*, as a model system but also as an ornamental crop, outperforms the single gene function assignment through reverse genetics.

Reverse genetics coupled with transcript profiling also needs a developed context of computational biology tools, such as the statistical methods for microarray analysis, the bioinformatics scaffolding system for sequence analysis, the data mining of the usually huge amounts of data produced, and the astringent quality check methods during all the procedure.

This chapter focuses on a new approach to novel gene identification in a developmental biology context: the flower architecture in *P. hybrida*. It builds the background to the dissertation discussing the *P. hybrida* model system, the facilities for gene expression analysis and the data analysis by data mining and bioinformatics, and concludes stressing the aims and outlook of the present PhD thesis.

1.1 Introduction

Developmental biology is aimed to elucidate the mechanisms underlying the making-off an organism at any stage, as well to analyze the cross-talk with environmental factors and evolutionary aspects. Its discipline

developmental genetics focuses on the molecular level and studies patterns of inheritance and variability, using molecular biology tools.

As a strategy to face a genetic study three aspects must be cared of: first, the selection of a good model organism, with widespread availability and ease of maintenance (Somerville and Koornneef, 2002; Gerats and Vandenbussche, 2005); second, the development of descriptive systems allowing systematic and quantitative data retrieval (Young and Center, 2000); and third, the data mining and or modeling of the observed results, thus providing not only data interpretation but prediction (Bassett Jr *et al.*, 1999). Exemplary projects using this strategies include the classical works of Mendel and McClintock. Mendel was able to extract a model of basic gene inheritance working with peas and simplifying the expectancies to simple ratios, and McClintock unveiled the controlling elements in maize through statistical analysis (McClintock, 1953).

The availability of fast and cheap DNA sequencing technologies has meant a revolution on genotype-phenotype interaction analysis in plants. A key point to this usefulness has been the sharing of the data on public accessible repositories, which has been resourceful for both the academic community as well the industry, as the scientists can broad their analysis and the companies can benefit mining the academic literature (Rounsley and Last, 2010). This enormous increase in data availability has boosted the development of tools able to handle the data in an integrative manner, and has supported the increase on reverse genetics strategies. An example of such impact is summarized with the overall tendency in *Arabidopsis thaliana* research during the decade of 2000-2010; that is, to substitute the forward, classical experimental design by a reverse genetics approach. The trend is explainable by its efficacy: over this period it unveiled 90% of the *A. thaliana* genes whose alteration triggers a mutant phenotype (Lloyd and Meinke, 2012).

The bioinformatics behind the reverse genetics approach is crucial, as it, in some cases, produces the hypothesis. For instance, the Agrikola intended to create a collection of *A. thaliana* lines silencing all its gene-specific tags (GSTs): the project cloned 21,500 GSTs representing at least the same amount of *A. thaliana* genes, subcloned them into binary hairpin vectors and created a collection of silencing *A. thaliana* lines (Hilson *et al.*, 2004). The silencing in such a large-scale analysis generates hypothesis after genotype-phenotype linking, as the triggered phenotype suggests the GST role *in vivo*.

P. hybrida, as a model species, has both the availability of EST (ex-

pressed sequence tag) information (that is, the *–omic* sequences to mine) and the facilities of easy plant transformation, maintenance and handling. Apart of that, it is a major crop in bedding plant industry and new varieties are continuously been developed.

1.2 Petunia as a model system on development

1.2.1 Botanical description

Petunia (*Petunia* × *hybrida* Hort. ex Vilm., also presented as *P. hybrida*) is a hybrid of ornamental interest which has been widely used as a model system due to the availability of molecular genetics tools (Gerats and Vandenbussche, 2005). Its parents are possibly *P. axillaris* and *P. integrifolia*, which can be artificially cross-fertilized (Wijsman, 1983; Ando *et al.*, 2001). The differences between both species are regarded as a result of their pollination syndromes (Stuurman *et al.*, 2004): *P. integrifolia* is pollinated by bees and has small, purple, almost scentless flowers (Ando *et al.*, 2001); whereas the *P. axillaris* flowers are fragrant, white, and have a long corolla tube that resembles in length the provoscis of its nocturnal pollinators, the hawkmoths (Hoballah *et al.*, 2007).

Petunia species are herbs with non-woody stems and are usually annual. The genus is endemic to South America, with subtropical distribution (Kulcheski *et al.*, 2006). Vegetative shoots in *P. hybrida* are monopodial (that is, grow upward in a single main stem), whereas the inflorescences are sympodial. This type of inflorescence patterning occurs as the apical meristem forms a terminal flower and the inflorescence stem elongates at the lateral meristems (Pnueli *et al.*, 1998). Hence, *P. hybrida* inflorescences are scorpioid cymes (Castel *et al.*, 2010), as depicted in figure 1.1.

P. hybrida leaf shape is ovate, obovate or elliptic, from sesile to petiolate, with flat margin and pinnate venation pattern. Depending on the line, color and trichome density vary (Stehmann *et al.*, 2009).

Wildtype *Petunia* flowers are typically purple and comprise four concentric whorls of five sepals, five fused petals, five stamens partially fused to the corolla, and two carpels. The zygomorphy is maintained in all four whorls from early organ differentiation onward (Rijpkema *et al.*, 2006a). The fruits, which can be harvested about four weeks after pollination, are dehiscent capsules surrounded by the enlarged calyx (Colombo *et al.*, 1997). Seeds are small (less than 1.4 mm), foveolated, with variable seed coat morphology (Stehmann *et al.*, 2009).

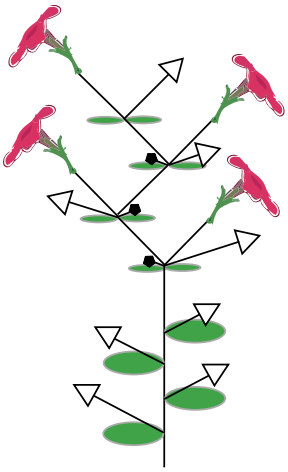


Figure 1.1: Model for Petunia branching. The diagram shows the branching pattern in wildtype. Leaves are indicated by green ovals. Vegetative axillary meristems are represented by open triangles, and dormant meristems, by pentagons. Note the presence of zones with suppressed branching before the zone that leads to the terminal flowers (Souer *et al.*, 1998).

Mitchell diploid, also known as W115, is a easily transformable *P. hybrida* line which carries mutated alleles for regulatory *loci* associated to pigmentation, *AN4* and *AN2*. Therefore, it is white-flowered. *AN2* regulates anthocyanin synthesis in the flower limb, whereas *AN4* modulates anther pigmentation and the petal tube (Quattrocchio *et al.*, 1993; Bradley *et al.*, 1998). Its origin is regeneration after anther culture of hybrids of *P. axillaris* × (*P. axillaris* × *P. hybrida* cv. 'Rose of Heaven') (Mitchell *et al.*, 1980; Ausubel *et al.*, 1980; Albert *et al.*, 2009).

V30, also named Violet 30, is a line of *P. hybrida* recalcitrant to transformation (van Tunen *et al.*, 1990). Having a bright violet color, it has been used in chalcone synthase gene family studies (Koes *et al.*, 1986a, 1989). The progeny of its crossing with M1 (another *P. hybrida* standard line) is easily transformable (Van Moerkercke *et al.*, 2012). A photograph depicting its differences in color with Mitchell diploid is presented as figure 1.2.

Figure 1.2: Mitchell and V30 flower shape. Mitchell diploid (on the left) shows white petals, while V30 (on the right) presents violet ones. Limb and tube morphology and calyx morphology and pigmentation vary as well.

1.2.2 Developmental biology

Developmental biology regards the manner in which organisms grow and define their intrinsic characteristics. These characteristics, or phenotype, are specified by differential gene expression, which is regulated by internal mechanisms, such as oscillators; and by the integration of environmental cues, such as temperature, population density or photoperiod (Gilbert, 2001). Its topics include organ identity, final size, symmetry, floral phyllotaxis, synorganization and the delimitation of the ecological influence on morphogenesis (Endress, 2006).

Vascular plant embryos show two distinctive zones which retain the capacity of continued growth, even at adult stage. These regions are set apart and comprise: first, the shoot apical meristem, which will produce the shoot and give rise to aerial organs; and second, the root apical meristem, which will produce the root. The implications of having an uninterrupted proliferating zone is that the shoot growth is indeterminate; that is the number of leaves is indefinite (Steeves and Sussex, 1989).

Plant tissues are composed of cells surrounded by a rigid, tightly cemented cell wall which limits the cell mobility and thus impairs cell migration replacement as a developmental variable. Instead, plants grow by addition of new cells at the primary meristems (tip growth of shoots and roots) or secondary meristems (lateral growth). So the phenomena that directly participate in the plant body formation could be classified into two groups: growth and differentiation (Steeves and Sussex, 1989).

Growth is the irreversible increase in size and is the result of cell division and cell enlargement. Differentiation is the result of cell changes about distinctiveness, which could bring different histological patterns; and, coordinated with growth, to organ shaping or organogenesis. Coordinated cell differentiation and growth build up the structural patterns that define a vascular plant (Steeves and Sussex, 1989).

The transition from the vegetative to reproductive phase implies the induction and development of an inflorescence meristem, which will generate floral meristems. Both endogenous (number of leaves) and environmental factors (photoperiod) contribute to that shift; in *P. hybrida*, gibberellins (Ben-Nissan *et al.*, 2004) as well photoperiod (Adams *et al.*, 1999) are known to be involved.

There are two major types of floral phyllotaxis, or patterns of organ arrangement in the flower: spiral and whorled. Spiral phyllotaxis is characterized by organ formation at constant time intervals, called plastochrons, with maintained angles to the predecesor organ, known as divergence angles. Whorled phyllotaxis have arrangements of organs which appear almost synchronously inside a ring shaped geometry; organs within a whorl are equidistant whereas a shift appears between consecutive whorls.

Organ identity determination in whorled flowers, such those from *A. thaliana*, Snapdragon (*Antirrhinum majus*) or *P. hybrida*, can be explained by the ABC model, as depicted in figure 1.3. It formulates that the combined expression of some of the members of a set of five functional classes of genes (A, B, C, D and E) determines the final category of a floral organ. D and E genes are the latest additions to the model. D-function genes are specifically involved in the formation of ovule and seed development. E genes are required for all the floral organs especification; they also regulate flower meristem determinacy. Summarizing the classical ABC model we could distinguish the four types of whorls according to an expression pattern of few genes as follows: A-function genes alone specify sepals; A- plus B- function genes, petals; B- plus C genes, stamens; and finally C genes alone specify carpels (Coen and Meyerowitz, 1991; Weigel and Meyerowitz, 1994; Mandel *et al.*, 1992; Bowman *et al.*, 1993; Sommer *et al.*, 1990; Jack *et al.*, 1992; Schwarz-Sommer *et al.*, 1992; Trobner *et al.*, 1992; Goto and Meyerowitz, 1994; Yanofsky *et al.*, 1990; Bradley *et al.*, 1993; Irish and Sussex, 1990; Jofuku *et al.*, 1994; Bowman *et al.*, 1991; Kunst *et al.*, 1989; Melzer and Theißen, 2009; Pelaz *et al.*, 2000, 2001; Theißen and Saedler, 2001; Pinyopich *et al.*, 2003; Ditta *et al.*, 2004; Honma and Goto, 2001).

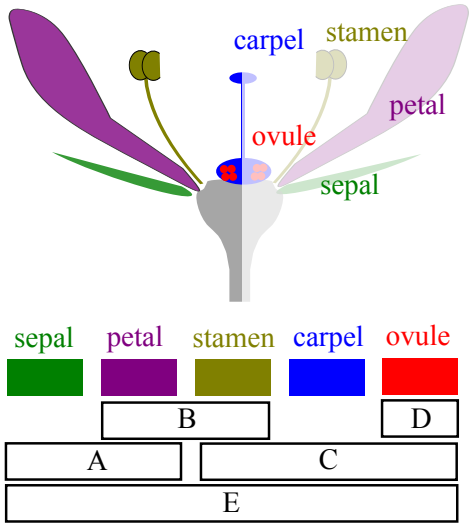


Figure 1.3: Diagram of the ABCDE model of flower development.

Most flowers have either a polysymmetric pattern of organ arrangement (actinomorphic) or a single plane of symmetry (monosymmetric, zygomorphic) (Endress, 1999). Among the Solanaceae there is some diversity at the zygomorphy degree: it is high in *Schianthus pinnatus*; slight in *Hyoscyamus niger*; and very small in *Solanum peruvianum*. The basal Solanaceae clades have strongly zygomorphic flowers, and the *P. hybrida* clade maintains the caracter although reduced. Monosymmetric flowers have apparently evolved several times independently. The character is whorl-specific in the family (Knapp *et al.*, 2004).

Whorled floral phyllotaxis allows synorganization, which is the coherence of two or more floral structures to produce a new structure or hyperorgan. It occurs either by local coordination or congenital or postgenital fusion. It happens frequently among organs of the same whorl, giving rise to fused petals in the perianth; or rarely between whorls (Apocynaceae, Orchidaceae) (Endress, 2006).

As a plant model, *P. hybrida* has as advantages the easy tissue culture, transformation and regeneration. In contrast with other widely used plant models, like *A. thaliana*, *P. hybrida* is more complex ecologically considering its niches (Gerats and Vandenbussche, 2005); even morphologically it has high diversity due to its varying pollination syndromes (such hawkmoth in *P. axillaris* and bee in *P. integrifolia*). These pollination

syndromes have been characterized by Stuurman *et al.* (2004). Based on its importance as model organism, EST and genomic facilities have been developed by its community, the Petunia Platform (The Petunia Platform, 2004)[1].

1.2.3 Petunia tools and impact

Among other emerging model systems, *P. hybrida* has desirable traits aiding genetics analysis that include the ease of culturing, short generation time, ease to transform and facilities for transposon tagging analysis. Because of that, it has been used for pollination syndromes elucidation. Its phenotypic traits that natural vary include the color, scent and nectar production, and flower morphology architecture determination, among others (Gübitz *et al.*, 2009). The genetics behind the flower color has been extensively analyzed by genetic engineering: genes encoding biosynthetic enzymes as well as regulators and vacuolar components have been isolated and characterized (Tornielli *et al.*, 2009); even more, novel flower colors have been achieved *via* transgenic approaches (Nishihara and Nakatsuka, 2010). Scent production, and specially benzenoid/phenylpropanoid volatiles, has been subjected to chemical elucidation and functional genomics comparison; its rhythmic emission and biosynthesis have been unveiled (Colquhoun *et al.*, 2010). The ABC-model extension to *P. hybrida* has been proposed, with an interesting complexity of B-class genes, which have undergone duplication and are represented by two major groups, the *DEF*/*AP3* and the *GLO*/*PI* lineages (Vandenbussche *et al.*, 2004).

The Petunia community excels in single-gene expression modification by reverse genetics. Both the easy transformation protocols and the availability of systems for permanent and transient expression analysis, and the the usage of the endogenous transposable system, permits gene function assignment (Gerats and Vandenbussche, 2005). The *dTph1* transposon system paired with a mapping facility has allowed to recover mutants in a cheap and effective way (Vandenbussche *et al.*, 2008). This transposable system, very active in some lines, has been suggested to be triggered by the high interspecific hybridization that produced the varieties and cultivars of

[1] Significant research has been performed in *Petunia* species for a long while; moreover, the pioneer observation of post-transcriptional gene silencing is attributed to Napoli *et al.* (1990); Van der Krol *et al.* (1990), who worked on *P. hybrida*. However, the first *P. hybrida* monograph was published in 1984 (Sink, 1984), and the Petunia Platform's website gathers the Petunia community since 2004 (The Petunia Platform, 2004).

P. hybrida; in that sense, *P. hybrida* is an example of the "genomic shock" proposed by McClintock (1983)(Gerats, 2009).

1.3 Gene expression analysis

Two major strategies are used for gene expression analysis, the hybridization-based and the amplification-based. Three main techniques come from these strategies: first, the pure hybridization methods, such as microarrays; second, the flexible probe-based quantitative real-time PCR (Q-PCR or RT-RT-PCR)[2] platforms, which in fact mixes both strategies; and third, the sequencing processes based on pure amplification methods. The amplification-based methods include techniques allowing DNA sequencing, and rely on the *in vitro* DNA synthesis *de novo* (Saiki *et al.*, 1988). The hybridization-based methods include the microarray analysis and the probe-based Q-PCR; both cases start with a predefined set of probes whose sequence is known. The sequencing-based methods, such as the EST projects, are more complex as they have the potential of identifying previously unknown sequences. The amplification-based technologies apart from the sequencing, such as the SYBR Green methods for Q-PCR, are very versatile (Liu *et al.*, 2007).

1.3.1 Expressed Sequence Tags sequencing

EST sequencing projects aim to obtain high amounts of redundant, partial sequences which need further data analysis. The processing steps are guided towards generating a biological database. The final processed EST database gathers the sequencing information after data quality check, assembly and annotation. After the raise of cheap next generation sequencing facilities, gathering of thousands of sequences became available to small laboratories, and therefore EST sequencing projects have arisen for many species.

Roughly, the classical procedure for EST sequencing could be summarized as follows. The messenger RNAs (mRNA) provide insight of the part of the genome being transcribed, and can be reverse transcribed to complementary DNA (cDNA) and further sequenced. The sequencing produces

[2]Through the dissertation RT-PCR and Q-PCR are treated as synonyms, thus focusing on the *R*eal-*T*ime kinetic data acquisition of the technique, and hence its remarkably handy statistical analysis (in contrast to northern blotting, in example). Hence, a reverse transcription-based assay using RT-PCR will be named *RT-RT-PCR*.

reads of variable length, depending on the technology used, that typically have lower base calling in the ends, thus producing a reliable sequencing in the middle. Taking into account the quality of each position and comparing to the rest of cDNAs sequences, the reads are clustered and assembled, therefore transformed into consensus EST sequences. After this step, the consensus sequences could be electronically translated and compared to protein databases; or compared by nucleotide database similarity searches. According to the function assigned to reference nucleotide or protein database matching entries, a functional annotation is assigned. Finally, a computer application serves the information and allows data visualization and interpretation (Nagaraj *et al.*, 2007b).

1.3.2 Quantitative, real-time PCR

Patterns of temporal and spatial gene expression trigger organ development (Koltunow *et al.*, 1990). Gene expression is regulated in many ways (Gagneux, 2003), both transcriptionally and post-transcriptionally, but the most common approach to its quantification is to measure the amount of the transcripts of interest. A fast, accurate and sensitive mRNA quantifying technique is the Q-PCR (Higuchi *et al.*, 1992).

Q-PCR allows continuous monitoring of the amplification kinetics of a PCR in real time measuring the specific fluorescence signal in each cycle; thus it produces a reaction kinetics which resembles the bacterial growth formulae (Monod, 1949). Each primer-sample combination will produce two parameters which will be the basis for data analysis: the C_T (cycle threshold; also called quantification cycle or Cq by Bustin *et al.* (2009)[3]) and the efficiency. The C_T is the fractional number of cycles in which the kinetic curve reaches a threshold amount of fluorescence. The amplification efficiency could be visualized in a half-logarithmic plot in which log transformed fluorescence values are plotted against the time (cycle number). Roughly, the efficiency represents the yield of the reaction.

[3] Although Bustin *et al.* (2009) presents the term Cq as the RDML (the proposed standard for storaging and exchanging data, the XML-based Real-Time PCR Data Markup Language)(Lefever *et al.*, 2009) standard arguing that the plethora of "quantitation cycle" terms refer to the same parameter, it is worth noting that depending on the author and, specially, the thermocycler's manufacturer, the parameter meaning varies. For example, the Corbett's RotorGene series distinguish between C_T and take-off: the first being calculated on a fractional threshold of total fluorescence basis; and the second according to the method of modified second derivative. Therefore, a definition of the sense used in each chapter is specified in their materials and method section.

Q-PCR data can be quantified absolutely and relatively. Absolute quantification employs a standard curve of an internal or external calibrator in order to extrapolate the input template copy number. Relative quantification measures the expression of the target gene and one or more reference genes, and thus returns a comparison of their expression; this relative amount is recorded for a target sample and a control sample (Yuan *et al.*, 2006). However, the relative quantification accuracy depends on the pattern of expression of the reference genes, which must be almost identical between samples. Thus the gene expression stability of the reference genes must be assessed prior to performing relative quantitation.

Several methods of reference gene expression stability estimating tools have been published. Roughly, they test two hypothesis: first, whether the variance of the gene expression is higher within biological replicates than between them (the lower the variance, the better the stability); and second, given a list of candidate reference genes, whether the pairwise variation between them is low or high (the lower the variation, the better the stability).

1.3.3 Microarray analysis

As opposed to Q-PCR, in which a few transcripts are analyzed at a time, microarray analysis is based on obtaining all the mRNA of a given sample, processing, labelling and hybridizing it in parallel to a grand number of DNA sequences which are supposed to represent an high percentage of the possible transcripts of that sample. This allows detecting simultaneously the expression level of thousands of transcripts (Schulze and Downward, 2001).

The expression profiling using the microarray technology is accomplished by two major approaches: first, the one-channel dye system, which hybridizes a set of transcripts to the chip; and second, the two-channel system, which co-hybridizes two or more fluorescent labeled probes and thus compares the relative expression levels of these probes (Hegde *et al.*, 2000). The analysis of changes in gene expression is conducted through a strategy that involves three stages: first, the array design and fabrication; second, the probe handling and hybridization; and third, the data collection, normalization and statistical testing (Allison *et al.*, 2006).

There are very successful microarray platforms for model species, such as the Affymetrix family for *A. thaliana* transcriptomics. However, it is possible to design and build a homemade microarray design. In both cases,

probe uniqueness and internal controls are challenges to meet during the data analysis. The splitting of an EST on different probes forces a summarization step in data analysis, which integrates them all. The mismatch controls give insight on hybridization specificity, even though some researches directly ignore the mismatching probes (Gautier *et al.*, 2004).

1.4 Reverse genetics

The availability of recombinant DNA technology, which allows the DNA transfer and stabilization into cloning vectors (Smith *et al.*, 1970; Danna and Nathans, 1971), allowed to conceive methods for stable plant transformation. Zambryski *et al.* (1983) explotated the oncogenic capabilities of T-DNA to raise a DNA-driven engineering golden era. Regarding *P. hybrida*, Monsanto was able to develop a transformation for the Mitchell diploid line introducing chimeric tumor-inducing (Ti) plasmids by *in vitro* transformation (Fraley *et al.*, 1983). Hence, since the early eighties the molecular genetics has the technology to select, clone and deliver pieces of DNA to plants. However, the utility of this techniques relies on the ability to gather gene sequences to deliver.

The first sequencing methods for RNA (Jou *et al.*, 1972) and DNA polymers (Gilbert and Maxam, 1973; Maxam and Gilbert, 1977; Sanger *et al.*, 1977) were developed in the seventies, giving input to the scientific community in a manner focused on single genes of interests. Moreover, big consortiums were able to accomplish genome sequencing projects (Baer *et al.*, 1984; Goffeau *et al.*, 1997; Blattner *et al.*, 1997; The Arabidopsis Genome Initiative, 2000). These methods, along with the discovery of the DNA polymerase-based DNA amplification (PCR) (Saiki *et al.*, 1988), provided a new approach to gene function discovery.

The raise of cheap and high-throughput methods, called next generation sequencing (NGS), permitted genome or transcriptome sequencing projects to small and medium sized laboratories. The scope also widened to non-model organisms (Ekblom and Galindo, 2010). Along with this technologies, the necessity of bioinformatics platforms for scientists without high specialization on computer science has arisen (Horner *et al.*, 2010).

This technical advances on plant transformation, molecular cloning, sequencing, PCR and bioinformatics are key to the current capacity to afford gene function garnering through reverse genetics. Knocking-out gene expression driven by antisense RNA was described by Ecker and Davis

(1986). Overexpression experiments in *P. hybrida* lead to the surprising discovery of sense supression or co-supression, phenomenon which later was ascribed to post-transcriptional gene silencing (PTGS) (Napoli *et al.*, 1990; Van der Krol *et al.*, 1990).

1.4.1 Gene silencing

Genetic engineering-driven gene silencing has been described as an effective way of down regulating gene expression, giving rise to phenotypes equivalent to the mutants carrying null alleles (Wang and Waterhouse, 2002). Gene constructs featuring intron-spliced transgenes which after transcription generate hairpin structures can induce gene silencing very efficiently (Smith *et al.*, 2000). The block of expression has been assigned to two phenomena: first, the elicitation of the cleavage of complementary transcripts (PTGS); and second, the impairment of the gene transcription through genomic DNA methylation, thus blocking transcript production (transcriptional gene-silencing or TGS) (Matzke *et al.*, 2001). Technology to knockdown genes by RNA interference (RNAi) elicits both PTGS and TGS phenomena.

Discovery of gene function could be achieved by molecular impairment of the genetic information and the subsequent loss-of-function phenotype analysis; some approaches include T-DNA or transposons insertions, or radiation- or chemically-induced mutations. However, these techniques lead to a limitation in plant recovery: genes which are essential will not produce homozygous mutated individuals, and in some cases they will even block the heterozygous viability (Herskowitz, 1987). This could be overcome either with constructs leaded by inducible promoters or with RNAi, because engineered plants with post-transcriptionally silenced genes show a gradient of phenotypes which offer a valuable advantage against this drawback, allowing to reduce the chance of non-recovery (Levin *et al.*, 2000).

A given species can undergo a RNAi assay if a successful construct delivery procedure has been developed (as *Agrobacterium*-based transformation, transient expression or even virus-based infection (Robertson, 2004)). Established such a method, a comprehensive gene catalog which provides a shortlist of targets to silence (Hilson *et al.*, 2004) is needed in order to establish RNAi strategies. The choice of adequate sequences to clone is influenced by the availability of public data to assess the transcript behavior (that is, microarray transcript profiling) but also by the major aim of

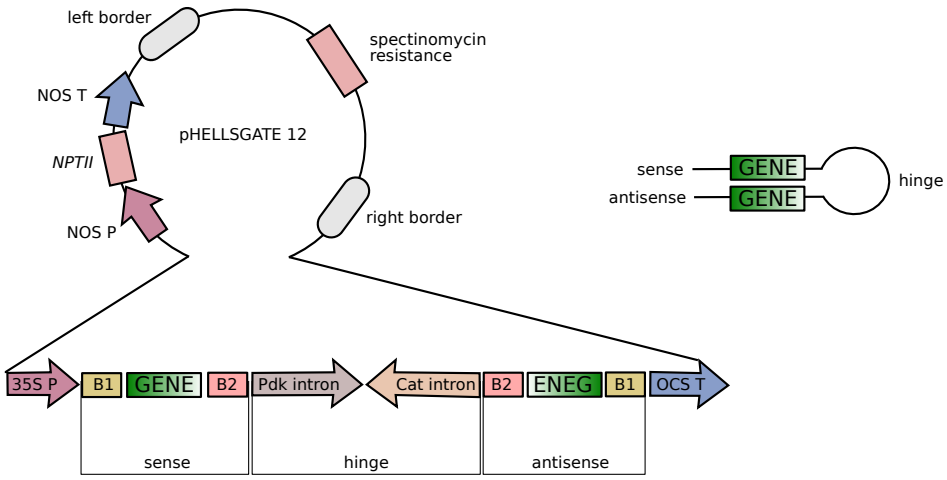


Figure 1.4: pHellsgate 12 vector map. The expression vector is designed to produce hairpin-like structures with a sense and antisense perfect matching sequences separated by a hinge composed by two introns. The vector construct is performed by Gateway directional recombination in a two step procedure with a first cloning into the donor vector pDONR 221 (Invitrogen). Final vectors have *B1*- and *B2*-flanked sequences that conform two identical inverted repeats. The hairpin's hinge is composed by the *Flaveria Pdk* gene and the intron of the castor bean *Cat* gene. Empty pHellsgate12 is 17,681 bp long. Spectinomycin is used for selection in bacteria and kanamycin (*nptII*) in *P. hybrida* (Helliwell and Waterhouse, 2003)

the study, generally whether a very specific silencing is needed or whether a group of similarly-structured transcripts can be inhibited (Hilson *et al.*, 2004).

Designing a RNAi experiment deals with two major issues: first, the specificity; and second, the efficiency. The specificity depends on the uniqueness of the sequence cloned and introduced into the silencing-producing vector, and can be increased or lowered depending on whether a single gene is intended to be silenced or not (Matzke *et al.*, 2001). The silencing efficacy is poorly understood, although some algorithms have been developed to aid in target design (Thareau *et al.*, 2003; Sclep *et al.*, 2007). Finally, this efficiency is very dependent upon environmental conditions, specially temperature (Velázquez *et al.*, 2010) and light (Kotakis *et al.*, 2010).

An efficient apporach to gene silencing construct design is the hairpin-

generating structure: an expression vector containing a promoter-controlled cassette containing two cloning sites, for sense and antisense sequences, separated by an intron (Waterhouse *et al.*, 1998; Smith *et al.*, 2000). However, the hairpin-generating structures have the inconvenience of difficult cloning due to the long homologies, so new methods have been proposed containing only a sense copy of the sequence to silence flanked by two promoters in inverse orientation (Schmidt *et al.*, 2012). Figure 1.4 depicts the structure of the hairpin-generating vector pHellsgate12, which contains the classical "promoter-sense-intron-antisense-terminator" cassette.

1.4.2 Reverse genetics on ornamental plants

Both genetic engineering and molecular genomics provide tools that have been used on plant breeding and plant biotechnology. Genetic engineering examples are the transgenics with high agronomic value that have been developed for maize, rice, soybean and tomato among others (Varshney and Tuberosa, 2007). Molecular genomics tools, derived from genome and transcriptome sequencing projects, lead to the consequent marker-assisted selection or marker-driven diversity checks of natural populations (Jain and Brar, 2010).

The application of transformation-based reverse genetics methods is limited because its cost, the ethical issues and the possibility of transient disruption instead of permanent blocking (Varshney and Tuberosa, 2007). Given the difficulties to introduce genetically modified plants in the ornamental market, transposon tagged or mutagenesis-based methods present an interesting alternative. Mutagenesis-based methods such as the targeting induced local lesions in genomes (TILLING) permit to induce variability in a transformation-independent manner, as well to identify mutations in any target *via* heteroduplex analysis (Till *et al.*, 2003). The *dTph1* transposon system present in *P. hybrida*, easily accesable by blast, allows to gather new alleles of a gene of interest to complement a reverse genetics-based gene function elucidation (Vandenbussche *et al.*, 2008). A simple plan of crossing with a background lacking the active *dTph1* transposase permits allele fixation and commercialization.

1.5 Data analysis

Because of the availability of the above-mentioned genetic resources regarding EST or genomic sequences, the convenient data analysis tools have

been developed.

1.5.1 Bioinformatics

Since the advent of cheap *–omics* tools, biological data handling, including the comprehensive data storage, mining and integration, has become an invaluable tool besides basic and applied research, but also as an subject itself: new algorithms and tools are being developed continuously.

Computational biology and bionformatics aims are: first, to allow efficient data storage, access and update; second, to develop tools and resources to analyze that data; and third, the use of such tools to integrate data and to interpret results in a biologically meaningful way (Luscombe *et al.*, 2001). The three aims show a gradient in scientific skills: the first one must be accomplished by computer scientists and the last one requires biological expertise. In order to facilitate the access to bioinformatics front-ends to life scientists, some tools are offered as Web services (Stein, 2002).

The development of new software in bioinformatics has been eased thanks to the availability of open source application programming interfaces (APIs), such as BioPerl (Stajich *et al.*, 2002), BioPython (Cock *et al.*, 2009), BioRuby (Goto *et al.*, 2010) or BioJava (Holland *et al.*, 2008), among others. These APIs aid to the development of robust, well-tested applications (Raymond, 1999); moreover, the release of such code as open source software allows scientific reproducibility (Barnes, 2010; Merali, 2010).

1.5.2 Data mining and modeling

The storage of *–omic* data in public repositories has permitted the usage of data mining techniques. Data mining is the computer science discipline which handles databases in order to gain knowledge detecting data patterns (Witten *et al.*, 2011). It is an iterative process of automatic data selection, preparation and manipulation in order to obtain nontrivial and valuable outcomes. Its goals are two: first, to describe huge datasets thus translating only machine-analyzable data into human-readable information; and second; to model and therefore build predictive systems using some variables or the interaction between variables present within the database (Kantardzic, 2011).

A data mining procedure starts with an hypothesis and the subsequent data collection, which is mainly performed by database management sys-

tems (the relational databases according to the paradigm drafted by Codd (1970) are common), and the preprocessing, which includes the outlier detection, scaling and feature selection. The corpus of the data mining, however, include a set of distinct tasks, such as classification, regression, clustering, summarization, dependency modeling and change and deviation detection. After the succesful application of some of the techniques available to fullfil the tasks, a model validation step must be applied. If the quality check supports the model validity, model interpretation and conclusion drawing apply. If necessary, the starting hypothesis or data selection and preprocessing can change and the data mining lifecycle starts over (Kantardzic, 2011).

The application of data mining procedures to genetics is increasing, specially to handle the huge amounts of *–omics* data. Specially, it has been proved insightful in marker prediction from microarray data in cancer research (Alizadeh *et al.*, 2000; Golub *et al.*, 1999; Ramaswamy *et al.*, 2001). However, its usage in plant developmental genetics itself is scarce.

1.5.3 Biological databases

Data storage is the natural step after data acquisition. In some cases, the aim of handling data is only to search and analyze it locally and by a single user. However, the amount of data or the necessity to provide concurrent access (that is, by multiple users) call for specific solutions. Most data storage techniques for the least demanding situations are based on flat files, which are usually integrated through indexing (Fujibuchi *et al.*, 1998; Etzold *et al.*, 1996). The usage of database management systems, however, provide more analytical power and checks for internal consistency. Biological databases are mostly implemented on relational database management systems (RDMS). Leading RDMS are Oracle, IBM, PostgreSQL or MySQL, often queried through interfaces such as ODBC and JDBC. The integration of biological RDMS-driven biological databases is tightly related with the data mining and integration (Köhler *et al.*, 2003; Stevens *et al.*, 2001).

The integrative capabilities of the biological databases rely on their modular design: that is, the possibility to add content to increase its informativeness. A typical scenario is the creation of a blastable database after cDNA sequencing in order to facilitate similarity searches. This hypothetical database can be further enriched through bioinformatics or data mining procedures; the typical follow-up to the sequencing are functional

annotation or polymorphism detection. Given a gene of interest inside the database, a common task is to design primers and perform gene expression analysis through Q-PCR. An application aiding primers design accessing the database could produce a set of primers to feedback it; thus, the database is enriched with a new entry linking the primers to the starting gene of interest. After the Q-PCR, results can be saved according to a data handling standard, such as MIQE (Bustin *et al.*, 2009), and attached to the gene of interest and primer combination. In that manner, a knowledge gathering is performed and the database populated, increasing its data mining potential as the content increases. This tendency to feedback different data in a relational manner is a challenging field on bioinformatics (Stein *et al.*, 2003).

1.6 Thesis scope

In this section we place the contents of the thesis within the current scientific background, providing a scaffold to the dissertation and clarifying its aim.

1.6.1 State of the art

The discovery of the genetic control underlying the floral architecture was tightly related with the mutant analysis. The defective plants with morphogenetic abnormalities were analyzed in the early eighties by gene mapping and cloning, thus allowing to disclose the underlying gene-to-function link (Smyth, 2005). So the homeotic mutants were used as tools for elucidating the normal organogenetic process (Meyerowitz *et al.*, 1989).

The source and characterization strategy of those aberrant phenotypes depended on the species used. Plant developmental biology was at first mainly conducted with two species, *A. thaliana* and Snapdragon. In the eighties, *A. thaliana* had a convenient linkage map (Koornneef *et al.*, 1983) and Snapdragon had their transposable element system already cloned and a large stock of mutants (Stubbe, 1966).

The success of *A. thaliana* as plant model was evident even in 1994, six years before the sequencing of its genome (The Arabidopsis Genome Initiative, 2000; Meyerowitz, 1994). The advantages for being in such position were its fast life cycle, its easy insertional mutagenesis and transforma-

tion[4], the development of efficient crossing procedures and the availability of chromosome maps and molecular markers. The large amount of seed produced each generation, which could be maintained effortless as genetic stocks, eased the retrieval and handling of an extensive amount of mutants (Boyes *et al.*, 2001).

In spite of that, the first plant homeotic gene cloned was *deficiens* from Snapdragon (Sommer *et al.*, 1990). Later additional homeotic genes were cloned and found to belong to transcription factors, known as MADS-box family members[5]. As soon these, as *agamous* of *A. thaliana* among others, were cloned (Yanofsky *et al.*, 1990), the ABC model of floral patterning was proposed (Coen and Meyerowitz, 1991). Such MADS proteins were shown to function as multimers (Davies *et al.*, 1996; Egea-Cortines *et al.*, 1999), and the model was widened ten years later when a set of MADS genes, including *sepallata*, were found to be interacting with the ABC genes in organ determination of *A. thaliana* floral primordia (Pelaz *et al.*, 2000).

The extensive floral diversity has given rise to comparative flower development, a field which integrates the new knowledge gathered from different species (Smyth, 2005). However, traditional genetic approaches are not feasible in many species: forward genetic approaches, like mutagenesis, requires fast outcross and progeny retrieval; and reverse genetic approaches, like persistent post-transcriptional gene silencing, depends on regeneration capabilities of tissue cultures of *Agrobacterium*-mediated transformed material (Irish and Litt, 2005).

Duplication of homeotic genes, such as those related with the lineage *apetala3/pistillata*, has occurred multiple times (Kramer *et al.*, 1998). Duplicate genes can diverge in function and structure (that is, transcription pattern and coding sequence) and it has been suggested that this functional divergence could be the basis of the current floral morphology diversity (Theißen *et al.*, 1996, 2000).

Accordingly, research on developmental genetics of the flower is being conducted in several species apart from *A. thaliana* and Snapdragon. A short list of model species in the field are: first, basal angiosperms, such

[4]Specially, the vacuum-infiltration and the floral dip procedures shortened the plant transformation protocols, making possible to recover *A. thaliana* transgenics without plant tissue culture or regeneration (Clough and Bent, 1998).

[5]Besides, some homeotic genes were found not to be MADS-box family members. *apetala2*, a gene that regulates some aspects of meristem and floral organ identity, was ascribed to a novel type of DNA binding domain (Weigel, 1995).

the basal eudicots Ranunculales[6]; second, monocots; third, *Gerbera*, an Asteraceae; and finally *P. hybrida*, close to tomato.

The increase in number of model species studied led to find some weakness of the ABC model. Even in Arabidopsis, A-function weak alleles do not trigger full homeotic conversion, and strong mutants show flowers which often lack the second whorl. Since the model proposed that the A and C function genes have a cadastral function (meaning that interact in a mutually antagonistic manner), the only A-mutant-like phenotype described in Snapdragon is produced by overexpression of C function in whorls 1 and 2 (Bradley *et al.*, 1993). Furthermore, the mRNA of *apetala2*, an A class gene, has been shown to complement the spatial expression of the miR172, a microRNA, with a transient overlap in the whorls 2 and 3 of the flower primordia; so the model has been updated proposing a cadastral action of miR172 over *apetala2* (Wollmann *et al.*, 2010).

The criticism over the ABC model also affects the B and C genes, which, when mutated, alter aspects apart from the organ identity. Loss-of-C-function mutants increase the number of floral organs thus indicating that C genes also affect the flower determinacy. Snapdragon loss-of-B-function mutants lack organs in the fourth whorl, thus pointing that B-function genes do not affect only whorls 2 and 3 (Trobner *et al.*, 1992).

Apart from the classical ABC genes, a plethora of genes have been described to affect flower development. Obviously, mutants with impaired plant hormone metabolism or signaling do show aberrant phenotypes (Mallory *et al.*, 2005; Hu *et al.*, 2003; Martí *et al.*, 2007). Strikingly, the serotonin biosynthesis is implicated in flower development also (Kang *et al.*, 2007). Expansins, cell wall proteins involved in cell wall loosening while enlarging, alter petal size in a location-specific manner (Zenoni *et al.*, 2004; Dal Santo *et al.*, 2011; Zenoni *et al.*, 2011a). Among the diversity of genes altering flower development, glutaredoxins (Li *et al.*, 2009), cytochromes P450 (Anastasiou *et al.*, 2007), bHlH transcription factors (Szécsi *et al.*, 2006) or E3 ubiquitin ligases (Disch *et al.*, 2006) have been described.

Gene function discovery through the available reverse genetics tools

[6]Apart of Ranunculales, the basal eudicots *taxa* Trochodendraceae (Chen *et al.*, 2007), Buxaceae (von Balthazar and Endress, 2002) and Sabiaceae (Wanntorp and De Craene, 2007), among others, have been subjected to many evolutionary and developmental studies; the basal eudicots represent a critical stage of angiosperm evolution, and the Ranunculid clade (*sense* Soltis *et al.* (2003); Kim *et al.* (2004)) can clarify the ancestral MADS-box gene functions as well the implications of the duplication in their lineages (Kramer and Zimmer, 2006).

became a major approach in the post-genomic era. Hairpin-based gene silencing expression vectors, such as the pHellsgate12, have proved to be valid in Solanaceous species (a map of the vector is available as figure 1.4). The Agrikola project project impact has been aforementioned, and roughly the same approach has been employed on different systems. It is clear that the genotype causes the phenotype but also that the genotype could be inferred from the phenotype. The increase of the genomic data gathered from in-depth studied species such as human or *Caenorhabditis elegans* lead to somehow the same large-scale analysis, based in this case in massive cloning of open reading frames (ORFs) and the subsequent protein production in heterologous systems; the approach lead to the creation of the *ORFeome* (Reboul *et al.*, 2003; Rual *et al.*, 2004).

Hence the research executed at the Agrikola project was performed in a hypothesis-free manner, in which the synthetic, inductive approaches were conducted starting from data. Such approach, in opposition to the analytic, hypothesis driven, deductive strategy based on the mining of an idea which leads to some data production, has been described as epistemologically valid (Kell and Oliver, 2004).

1.6.2 Aim

The objective of the present dissertation is to enlarge the current knowledge on flower development by means of the application of reverse genetics techniques in a hypothesis-free manner. An overview of the experimental approach is depicted in figure 1.5.

The availability of techniques and literature at the start of the research in 2008 included: an efficient protocol of plant transformation (based on leaf coculture with *A. tumefaciens*); a unannotated, non-public set of EST sequences of mainly *P. hybrida* roots, albeit floral and inflorescence organs were also included (which was uploaded afterwards to NCBI's with accession UNILIB:50472); and a coherent work on flower genetics (specially regarding MADS box genes).

The pitfalls included: the lack of genomic sequences; the relatively long life cycle of the species; and the inherent diversity between *P. hybrida* lines, as the most used varieties are hybrids with unknown combinations of *P. axillaris* and *P. integrifolia*.

Thus the work outline conceived the reverse genetics approach in a hypothesis-generating manner: according to a bioinformatics parameter (that is, the likeliness of being differentially expressed in petals), we chose

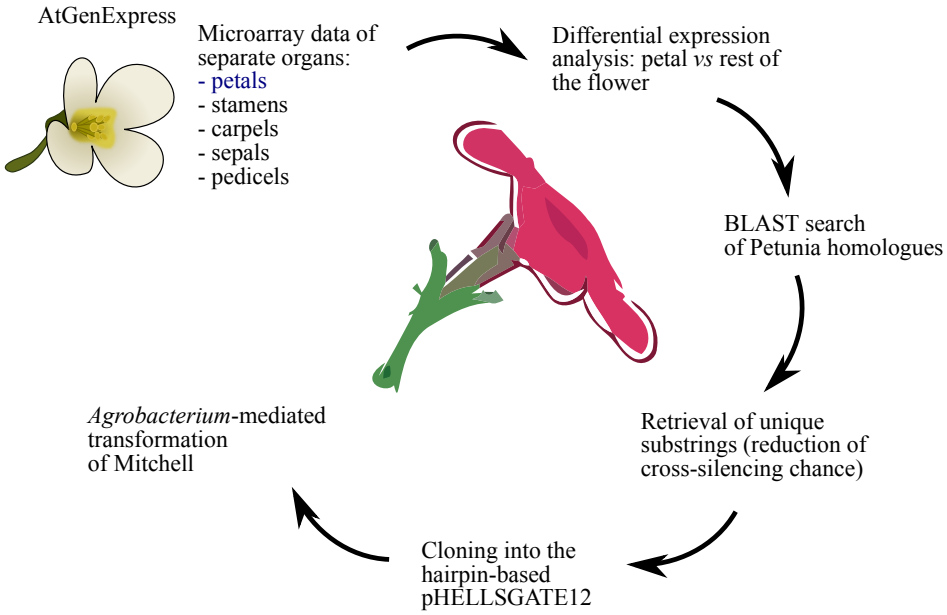


Figure 1.5: Experimental design and approach to the main objective of the present dissertation. First, the microarray data were retrieved from AtGenExpress; second, a linear model was fitted to check the differential expression on petals; third, the ranking of target transcripts was mapped to a local EST database gathering *P. hybrida* sequences; fourth, an alignment-based method was applied to select the parts of the target EST which appear only once at the database; fourth, such unique substrings were cloned, and subcloned to the pHellsgate12 vector; and fifth, a transformation protocol was applied using *in vitro* leaf explant coculture with the construct-carrying *A. tumefaciens*.

nine transcripts and silenced them independently. Chapter 5 summarizes this. The reason for such an approach is the trend of finding non ABC-based genes altering flower architecture; in this sense, a statistically-based hypothesis generating seemed innovating. The focus on silencing rather than overexpression is due to the lack of an exhaustive biological database containing full coding sequences, as opposed to the abundance of EST information on partial mRNAs.

Along with this aim, we developed some tools easing or allowing the main objective achievement. These tools have a basis on computational biology. First of all, an EST database handling platform was produced, as described in chapter 2. Secondly, as the PCR efficiency is a cornerstone on gene expression quantification, we modelized it and integrated a predictor into a primer design software (chapter 3). Finally, and focusing again on developing an accurate gene expression platform and thus precision phenotyping, we identified the suitable reference genes for real time PCR in our experimental conditions, as described in chapter 4.

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.

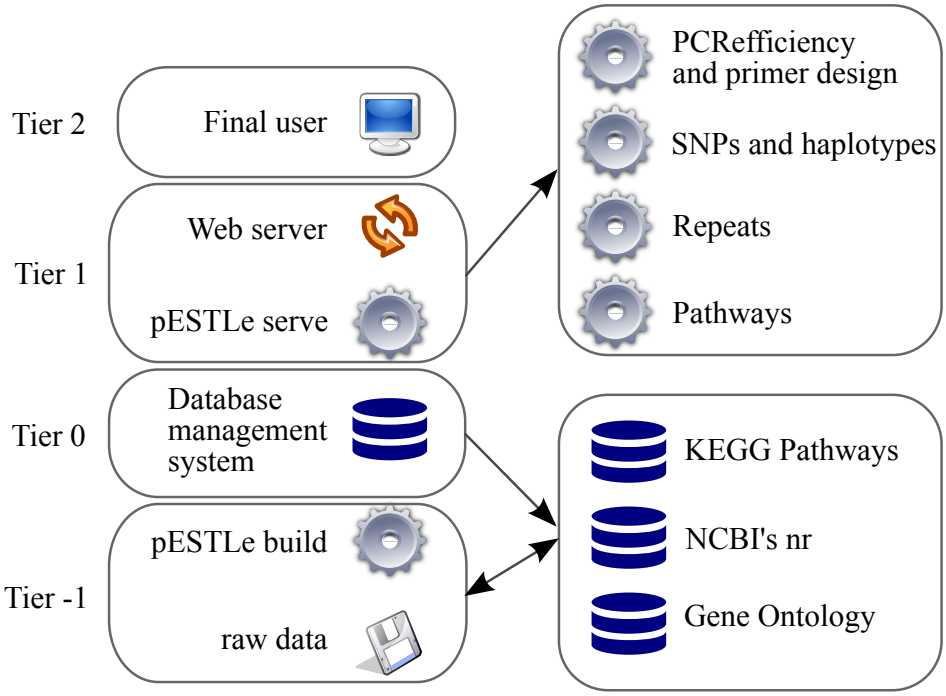


Figure 2.1: pESTle architecture. The proposed EST database comprises five tiers. Tier -1 includes the preprocessing, clustering and assembly of the raw data and interacts with previously established databases, such as Gene Ontology or NCBI's nonredundant. Tier 0 is the PostgreSQL core of database management 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

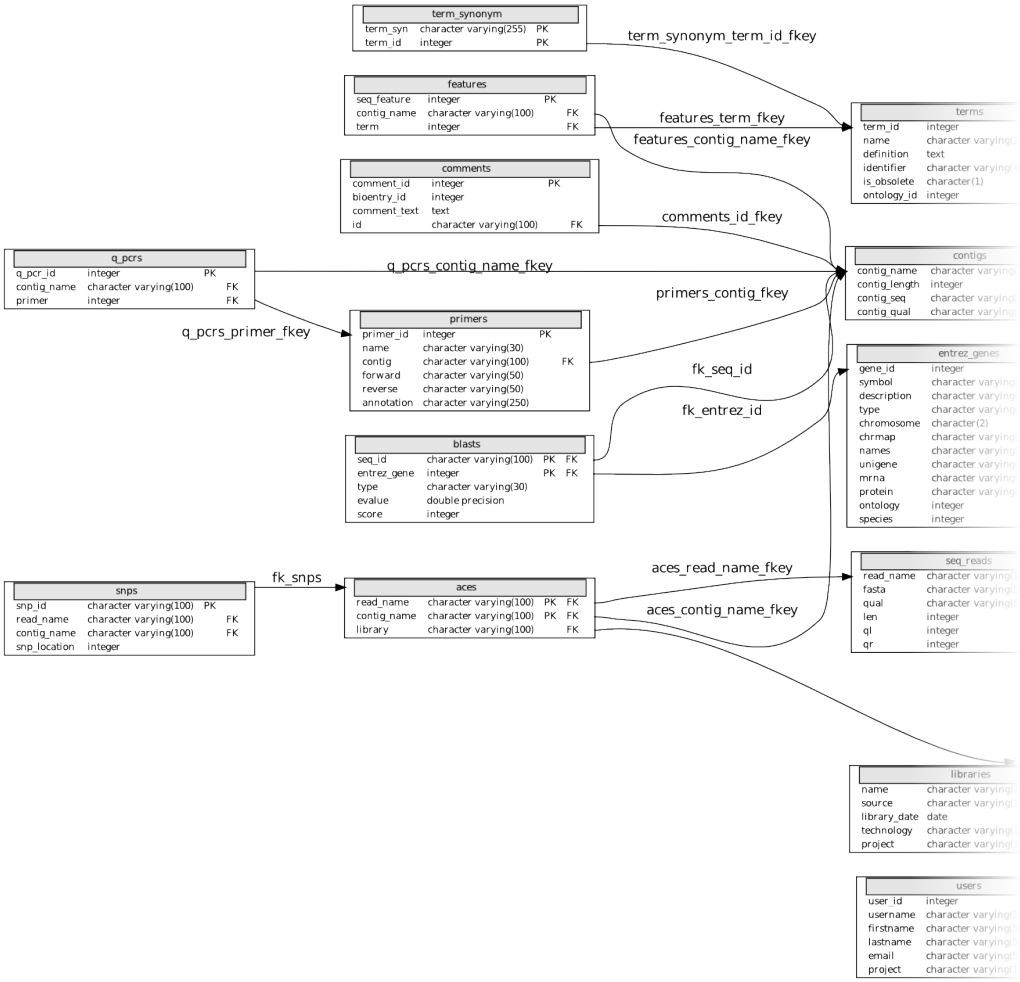


Figure 2.2: Entity Relationship Diagram of the core of pESTle, as represented by PostgreSQL autodoc (Taylor, 2007), part I.

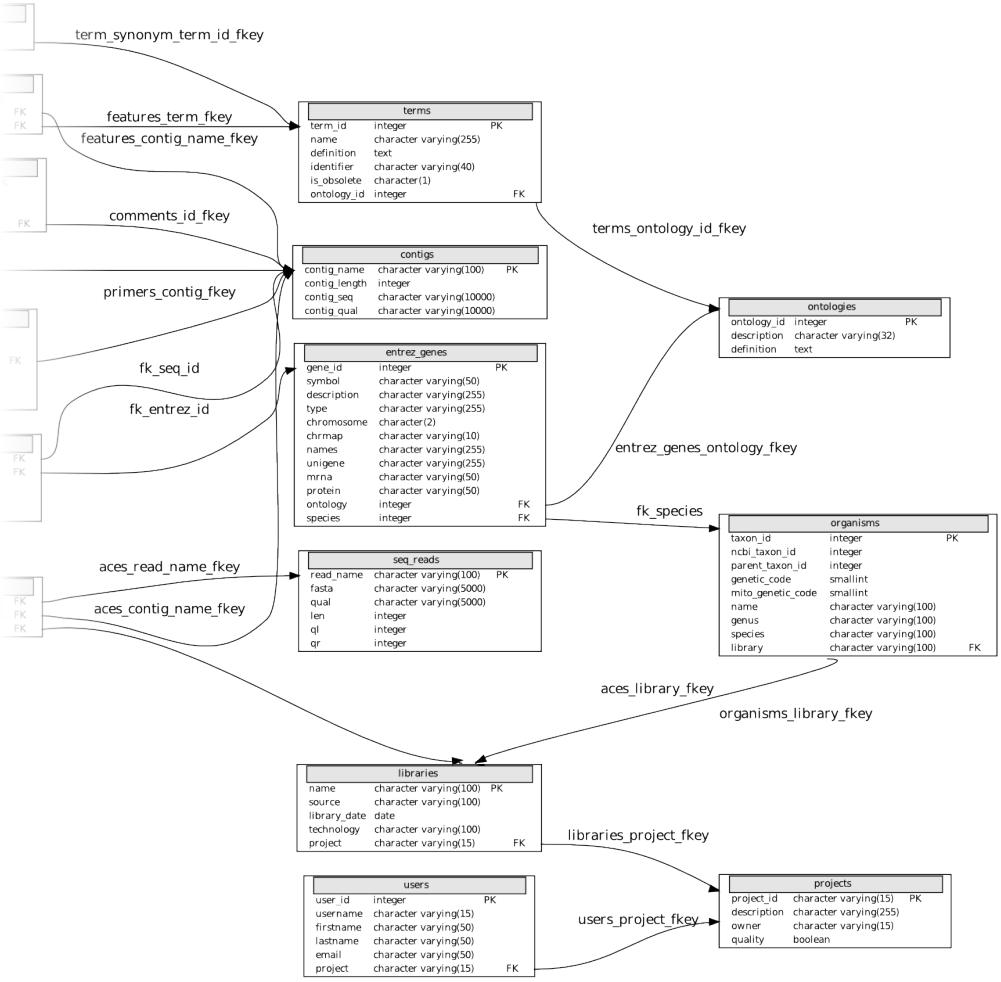


Figure 2.3: Entity Relationship Diagram of the core of pESTle, as represented by PostgreSQL autodoc (Taylor, 2007), part II.

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.

Chapter 3

pcrEfficiency: A Web tool for PCR amplification efficiency prediction

Relative calculation of differential gene expression in quantitative PCR reactions requires comparison between amplification experiments that include reference genes and genes under study. Ignoring the differences between their efficiencies may lead to miscalculation of gene expression even with the same starting amount of template. Although there are several tools performing PCR primer design, there is no tool available that predicts PCR efficiency for a given amplicon and primer pair.

We have used a statistical approach based on 90 primer pair combinations amplifying templates from bacteria, yeast, plants and humans, ranging in size between 74 and 907 bp to identify the parameters that affect PCR efficiency. We developed a generalized additive model fitting the data and constructed an open source Web interface that allows the obtention of oligonucleotides optimized for PCR with predicted amplification efficiencies starting from a given sequence.

pcrEfficiency provides an easy-to-use web interface allowing the prediction of PCR efficiencies prior to web lab experiments thus easing quantitative real-time PCR set-up. A web-based service as well the source code are provided freely at `http://srvgen.upct.es/efficiency.html` under the GPL v2 license[9]

3.1 Introduction

Since the development of quantitative PCR (Q-PCR) in the early nineties (Higuchi *et al.*, 1993), it has become an increasingly important method

[9]Izaskun Mallona, Marcos Egea-Cortines and Julia Weiss. pcrEfficiency: a Web tool for PCR amplification efficiency prediction. *BMC Bioinformatics*, 12:404, 2011.

for gene expression quantification. Its aim is to amplify a specific DNA sequence under monitoring and measuring conditions that allow stepwise quantification of product accumulation. Product quantification has fostered the development of analysis techniques and tools. These data mining strategies focus on the cycle in which fluorescence reaches a defined threshold (value called *quantification cycle* or *Cq*) (Luu-The *et al.*, 2005; Bustin *et al.*, 2009); with the *Cq* parameter, quantification could be addressed following two approaches: (i) the standard curve method (Morrison *et al.*, 1998) and (ii) the $\Delta\Delta Cq$ method (Livak and Schmittgen, 2001).

It is worth noting that these classical quantification methods assume that amplification efficiency is constant or even equal to 100%. An efficiency value of 100% implies that during the exponential phase of the Q-PCR reaction, two copies are generated from every available template. But it has been shown that these assumptions are not supported by experimental evidences (Ramakers *et al.*, 2003). With the aim of estimating PCR efficiency, and thus to include it in further analysis procedures, two strategies have been developed: (i) kinetics-based calculation and (ii) standard curve assessment.

Taking into account the reaction kinetics, which is basically equivalent to the bacterial growth formulae (Monod, 1949), amplification efficiency could be visualized in a half-logarithmic plot in which log transformed fluorescence values are plotted against the time (cycle number). In these type of graphic representations, the phase of exponential amplification is linear and the slope of this line is the reaction efficiency (Schefe *et al.*, 2006). Empirical determinations of amplification efficiencies show that ranges lay between 1.65 and 1.90 (65% and 90%) (Tichopad *et al.*, 2003). Standard curve-based calculation method relies on repeating the PCR reaction with known amounts of template. *Cq* values *versus* template (that is, reverse transcribed total RNA) concentration input are plotted to calculate the slope. Laboratories where few genes are analyzed for diagnostic may develop standard curves but they are in most cases out of scope for research projects where tens-hundreds of genes will be tested for changes in gene expression.

Several aspects influence PCR yield and specificity, including reagents concentration, primer and amplicon length, template and primer secondary structure, or G+C content (Qu *et al.*, 2009). The goals of a PCR assay design are: (i) obtaining the desired product without mispriming and (ii)

(Mallona *et al.*, 2011b).

rising yield towards optimum. In most cases, the sequence to amplify is a fixed entity, so setting up an efficient reaction involves changes in reagents concentrations (salts, primers, enzyme) and specifically an optimal primer design. Thus a plethora of primer designing tools have been published, regarding as little as G+C content for T_m calculation (Marmur and Doty, 1962; Wallace *et al.*, 1979), evaluating salt composition (Howley *et al.*, 1979) or even employing Nearest Neighbor modules, which consider primer and salt concentrations (Breslauer *et al.*, 1986).

Efficiency values are essential elements in the $\Delta\Delta Cq$ method and its variants: relative quantities are calculated using the efficiency value as the base in an exponential equation in which the exponent depends on the Cq. Thus efficiency strongly influences the relative quantities calculation, which are required to estimate gene expression ratios (Livak and Schmittgen, 2001).

In this work, we analysed Q-PCR efficiency values from roughly 4,000 single PCR runs with the aim of elucidating the major variables involved in PCR efficiency. With this data we developed a generalized additive model (GAM), which relies on nonlinear regression analysis, and implemented it in a open, free online Web tool allowing efficiency prediction.

3.2 Methods

3.2.1 DNA templates

We used a variety of DNA templates to obtain data for efficiencies including genomic DNA from bacteria, yeasts, plants and humans, and plasmid DNA. Samples using cDNA as template were produced from isolated mRNA from different sources. Synthesis of first strand cDNA was performed from DNAase treated mRNA, using the Maxima kit from Fermentas as described in the protocol. Samples amplified by whole genome amplification using the ϕ29 DNA polymerase were performed with the Genomiphi kit (GE-Healthcare) according to manufacturers manual. A summary of the data is available as Additional file 1 at the published version of the manuscript.

3.2.2 Real time PCR

PCR reactions used were carried out with the SYBR Premix Ex Taq (TaKaRa Biotechnology, Dalian, Jiangsu, China) in a Rotor-Gene 2000

thermocycler (Corbett Research, Sydney, Australia) and analysed with Rotor-Gene analysis software v.6.0 as described before (Mallona *et al.*, 2010). A second set of reactions was performed with a Mx3000P machine (Stratagene, Amsterdam) and analyzed with the qpcR R package (Ritz and Spiess, 2008). Reaction profiles used were 40 cycles at 95 °C for 30 s, an amplicon-specific annealing temperature for 20 s and amplification at 72 °C. In order to ensure the specificity of the reaction, uniqueness of Q-PCR products were checked by melting analysis (data not shown). Fluorescence data acquisitions during the cycling steps were collected at 72 °C step, temperature at which eventual primer-dimers should be melted, thus avoiding artifactual contribution to the fluorescence measure. Once finished, analysis was followed by a melting curve whose ramp was delimited between the annealing temperature and 95 °C. Reaction volume was 15 μL and each primer was 240 nM.

Reaction efficiency was calculated using the amplification curve fluorescence, analyzing each PCR reaction (tube) separately as before (Liu and Saint, 2002). Efficiency value (E) was defined as $E = \frac{F_n}{F_{n-1}}$, in which n is determined as the 20% value of the fluorescence at the maximum of the second derivative curve. Efficiency calculations were performed with the qpcR R package (Ritz and Spiess, 2008). Curves were formed by 40 points, each one representing a fluorescence measure in each amplification cycle. The Rotor-Gene 2000-based runs were baseline corrected either by standard normalization (substraction of the fluorescence present in the first five cycles of each sample) or by "dynamic tube" normalization (which uses the second derivative of each sample trace to determine the take-off, thus asigning a threshold separately to each reaction), whereas the Mx3000P were by "adaptive baseline" correction (which assigns a threshold independently to each sample).

3.2.3 Data mining

Data mining was performed with the R statistical environment v2.7.1 and v2.10.1 (R Development Core Team, 2008) with the following libraries: coin v1.0-4, (Hothorn *et al.*, 2006) mgcv v1.7-5 (Wood, 2001), ROCR (Sing *et al.*, 2009) v1.0-4, compute.es v0.2 (Re, 2010) and verification v1.31 (NCAR, 2010). The final model was implemented in a CGI server-side set of Python/BioPython scripts interacting with the web browser requests. Source code of both modeling procedures and the server-side application are available at the website.

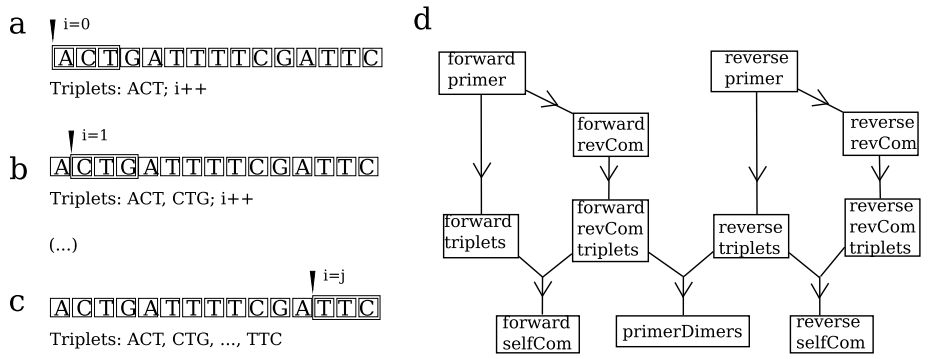


Figure 3.1: Algorithms for primer self-complementarity (primerSelfcom) and cross hybridization (primerDimers) computing. **(a)**, **(b)** and **(c)** show three stages of the sliding window triplet extraction step. All the DNA string is reduced into overlapping triplets. **(d)** reflects the general overview of the algorithm. As a first step, triplets are extracted for each primer. Then, primers are reverse complemented and thereafter splitted into overlapping triplets. Comparison between triplets allows the generation of an estimate of similarity, which is employed as a hybridization predictor.

3.3 Results

3.3.1 Data overview

Our data were generated from 90 different amplification products that included four *Escherichia coli* strains, three *Agrobacterium tumefaciens* strains (Manchado-Rojo *et al.*, 2008), three tomato varieties (Weiss and Egea-Cortines, 2009), three *Petunia hybrida* lines (Mallona *et al.*, 2010), one *Antirrhinum* line (Delgado-Benarroch *et al.*, 2009b), one *Opuntia ficus-indica* genotype (Mallona *et al.*, 2011a) and human liquid cytology samples used to test for *Prostate Serum Antigen* presence. Efficiencies ranged between 1 (no amplification) and 2 (perfect exponential duplication). We wrote an R script (see Materials and Methods) to extract the numerical inputs for further statistical analysis. The script regarded complete amplicon length, primer sequence, G+C content of amplicon and primers, presence of repetitions in the amplicon (N6 or above), primer melting temperature and the 3' terminal, last two nucleotides of each primer and primers ten-

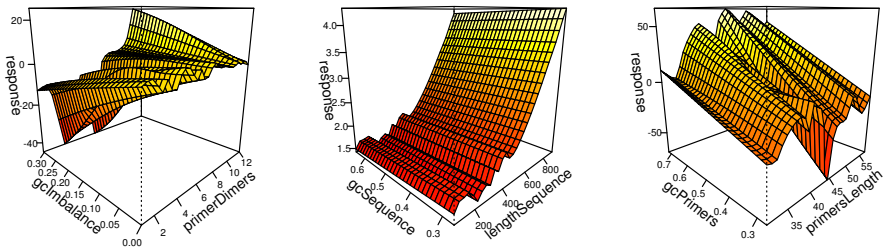


Figure 3.2: Perspective plot views of the GAM. Results of the best-fitting smooths for the variables included in the model. The interaction between the two variables is presented as a surface; the z-axis shows the response and the relative importance of each variable is presented in the x- and y-axis.

dency to hybridize (figure 3.1). Metadata for each PCR reaction included sample origin (that is, genomic or cDNA), operator involved, species and line or variety, exact sequence amplified, primer length and PCR efficiency; a summary of the data is shown as Additional File 1 at the published version of the manuscript.

Efficiency dataset had highly repetitive data (presence of ties). Since some of our comparisons intended to assess relationships between two quantitative variables, we applied Spearman tests*Spearman's!test*. However, analysis of quantitative versus qualitative data was performed employing Kruskal-Wallis tests. Due to the presence of ties, which hampers the application of rank-based tests, asymptotic tests were applied. p-values were approximated *via* its asymptotic distribution and ties were adjusted *via* random rank averaging. Both procedures are implemented in the coin R package (Hothorn *et al.*, 2006). We chose a value of 0.05 as cut-off of statistical signification. A summary is shown in table 3.1. *Post-hoc* asymptotic Wilcoxon Mann-Whitney rank sum test were performed to discriminate the contribution of some of the categorical variables, as well to analyzed effect sizes; the statistical outputs are shown as table 3.3.

Table 3.1: Statistical results for univariate analysis. In the case of asymptotic Spearman tests, Z value are shown; χ^2 value is written for asymptotic Kruskal-Wallis tests. An asterisk over the p-value reflects a significant influence (p-value ≤ 0.05). Sperman's correlation analysis shows the ρ statistic as well the Cohen's d and *Log Odds* estimates of effect size. For *post-hoc* asymptotic Wilcoxon Mann-Whitney rank sum tests and its effect size estimators see 3.3.

Variable	value	df	ρ	d	Log Odds	p-value
Primers length	Z = -7.4398	–	-0.118	-0.239	-0.433	1.008e-13*
Sequence length	Z = -5.423	–	-0.086	-0.173	-0.314	5.86e-08*
Sequence G+C content	Z = -10.2664	–	-0.163	-0.331	-0.601	<2.2e-16*
A repeats	Z = 2.1004	–	0.033	0.067	0.121	0.03569*
T repeats	Z = 3.9818	–	0.063	0.127	0.230	6.84e-05*
C repeats	Z = -5.294	–	-0.084	-0.169	-0.307	1.196e-07*
G repeats	Z = -7.1808	–	-0.114	-0.230	-0.418	6.929e-13*
Primers T_m	Z = 1.4653	–	0.023	0.047	0.085	0.1428
Primers self complementarity	Z = 11.9002	–	0.190	0.386	0.700	<2.2e-16*
Primer dimers	Z = 4.4161	–	0.070	0.141	0.256	1.005e-05*
Primer GC imbalance	Z = 11.1367	–	0.177	0.360	0.654	< 2.2e-16*
Primers GC content	Z = 4.5921	–	0.073	0.147	0.266	4.388e-06*
Sequence palindromes	Z = -3.4951	–	-0.056	-0.111	-0.202	0.0004738*
Species	χ^2 = 585.616	9	–	–	–	<2.2e-16*
Template	χ^2 = 1241.562	12	–	–	–	<2.2e-16*
Variety or line	χ^2 =585.8386	18	–	–	–	<2.2e-16*
Template source	χ^2 = 940.8915	24	–	–	–	<2.2e-16*
Operator	χ^2 = 727.4887	8	–	–	–	<2.2e-16*
Primer's 3' last two nucleotides	χ^2 = 237.911	15	–	–	–	<2.2e-16*

Table 3.2: GAM summary as estimated by the *gam* function of the mgcv R package. Model formula corresponds to: $efficiency\ s(lengthSequence, gcSequence) + s(primersLength, gcPrimers) + s(gcImbalance, primerDimers)$.

GAM analysis				
	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	1.73825	0.00153	1136	<2e-16
Approximate significance of smooth terms				
	edf	Ref.df	F	p-value
s(lengthSequence,gcSequence)	17.58	18.08	2.332	0.00114
s(primersLength,gcPrimers)	27.96	28.46	22.950	<2e-16
s(gcImbalance,primerDimers)	28.83	29.33	16.717	<2e-16
R-sq.(adj) = 0.41 Deviance explained = 42.1% GCV score = 0.0094091 Scale est. = 0.0092293 n = 3944				

In order to build up a predictive, statistical model we used a generalized additive modelling procedure as an effective technique for conducting nonlinear regression analysis in which factors were modeled using nonparametric smooth functions. GAM function was implemented by the R project *mgcv* package (Wood, 2001), according to the formulation described in (Wood, 2004). Data fitting is shown in table 3.2. Figure 3.2 shows perspective plot views of the GAM; predicted efficiency is plotted as a response surface defined by the values of two interacting variables.

3.3.2 Statistical modelling

Model selection was performed according to the Akaike's Information Criterion, a penalized log-likelihood system addressing model goodness and denying low parsimony (Akaike, 1974). We selected a GAM based on the interaction of the length and G+C content of the sequence as well, independently, of the primers; and the interaction of the G+C imbalance between primers with an estimation of the tendency to produce primer dimers. The R squared parameter of the model is 0.41, whereas the deviance explained is 42.1 %.

As the model intends to estimate the PCR efficiency of a set composed by a given amplicon and a given set of oligo primer pairs, it could be validated in terms of ranking performance. We defined a threshold

Table 3.3: *Post-hoc* categorical variable analysis. The variables regarding PCR template (*GD*, genomic; *CD*, cDNA; *plasmid*, *Escherichia coli* plasmid; *yGD*, yeast genomic) and 3' primer *termini* (*U*, purine; *Y*, pyrimidine) were analyzed by asymptotic Wilcoxon Mann-Whitney rank sum tests (*Z*, Z value; *p*, p-value), and the effect sizes estimated by the two tailed p-value (Cohen's *d*, mean difference; Hedge's *g*, unbiased estimate of *d*; *r*, correlation coefficient; and *n*, the total sample size, which is twice the effective sample size when the *termini* of the two oligos are analyzed).

Template	Z	p-value	d	g	r	n
CD *vs.* GN	-31.16	<2.2e-16	0.27	0.27	0.13	3851
CD *vs.* plasmid	-12.06	<2.2e-16	0.55	0.55	0.16	2548
GD *vs.* plasmid	-9.92	<2.2e-16	0.57	0.57	0.19	1785
yGD *vs.* plasmid	8.88	<2.2e-16	0.77	0.77	0.36	480
yGD *vs.* CD	3.95	7.506e-05	0.26	0.26	0.07	2546
yGD *vs.* GD	9.72	<2.2e-16	0.57	0.57	0.19	1783
UU *vs.* YY	-8.38	<2.2e-16	0.27	0.27	0.13	3600
UU *vs.* UY	-6.26	3.695e-10	0.19	0.19	0.09	4179
UU *vs.* YU	-8.74	<2.2e-16	0.27	0.27	0.13	3517
UY *vs.* YU	-3.15	0.001596	0.09	0.09	0.04	4288
UY *vs.* YY	-2.27	0.02308	0.06	0.06	0.03	4371
YU *vs.* UY	-3.15	0.001596	0.09	0.09	0.04	4288

of experimental efficiency measured during the Q-PCR obtaining a decision criterion of adequate PCR performance. We took results below 1.60, 1.70, 1.80 and 1.90 (that is 60 %, 70 % and so on) of efficiency as fails, thus this threshold acts as a binary classifier of success. Receiver Operator Characteristic (ROC) curves are commonly used to analyze how the number of correctly classified positive cases whose predicted efficiency is over the threshold change with the number of incorrectly classified negative examples whose predicted efficiency is below that threshold (Davis and Goadrich, 2006). This representation is complemented with the precision and recall (PR) curves, which evaluate the relationship between the precision (the ability of presenting only relevant items) and recall (true positive rate) (Davis and Goadrich, 2006). ROC and PR curves are shown in figure 3.3; ROC rises rapidly to the the upper-left-hand corner thus reflecting that the false-positive and false-negative rates are low, whereas the

PR curve locates at the upper-right-hand corner and thereafter indicates that most of the items classified as positive are true positives.

3.3.3 Implementation

In order to ease PCR efficiency prediction prior to wet-lab PCR set-up, we developed a user-friendly, freely available web tool for assessing primer suitability before and during primer design. For this purpose, we wrote a set of Python/Biopython scripts (Chapman and Chang, 2000) requested through a Common Gateway Interface (CGI). These scripts were developed to call to the Primer3 software (Rozen and Skaletsky, 2000), working inside the EMBOSS package (Rice *et al.*, 2000). The web tool called PCR efficiency calculator allows primer design starting from a DNA fragment producing a set of theoretical PCR efficiency values. It also predicts PCR efficiency values for preexisting primers and DNA template combinations.

3.4 Discussion

Several tools have been developed to assess primer design procedures. Most of them consider hairpin structure formation avoidance, selection of nucleotides in 3' *termini*, primer melting temperature, etc. (Chavali *et al.*, 2005). However, intrinsic amplicon characteristics are not contemplated in primer design. The work we present includes this important parameter as amplification was found to be highly dependent on template structure (tables 3.1 and 3.2). Indeed PCR specificity or PCR failure have been found to be dependent on sequence similarity between primers and template, lack of mismatches, or number of priming sites (Mann *et al.*, 2009; Andreson *et al.*, 2008). Using logistic regression analysis, Benita *et al.* (2003) found that PCR success is highly dependent on regionalized G+C content in the template thus showing the importance of template structure as a second step in PCR optimization. Generally, PCR success is evaluated as a dichotomy by presence or absence of product. However in Q-PCR experiments amplification efficiency becomes an important parameter to perform proper statistical analysis that should yield the actual differences in expression between several transcripts. Thus PCR efficiency becomes as important as *Cq* values to determine differential gene expression. Our model showed that G+C content in the amplicon plays a key role in PCR efficiency confirming previous work and including it inside as a predictor of PCR efficiency.

Multiple parameters, such sequence palindrome abundance or the nucleotide at the 3' primer *termini*, were found to significantly contribute to the PCR efficiency when analyzed separately. However, they were not significant when included to a multiple component GAM. The model presented in this work estimates an efficiency value *per* PCR reaction regarding three parameters, each of one represents the interaction between two independent variables: the interaction between G+C content of the amplicon with its length; the interaction between G+C content of the primers with its length; and the interaction between G+C content imbalance between primers (gcImbalance; their difference in G+C) with their tendency to hybridize and thus to form primer dimers (primerDimers).

Our model gives a high influence to the difference of G+C content between primers. Previous works noted that PCR using unequal primer concentrations have better efficiencies when their melting temperatures differ in ≥ 5 °C (Pierce *et al.*, 2005). However, when our tool is piped to the Primer3 primer-design workflow, this difference is restricted by the Primer3 algorithms, thus avoiding design of highly unequal primers. Very high or very low amplicon G+C content affects amplification success (Baskaran *et al.*, 1996; Varadaraj and Skinner, 1994). Specially, regionalized G+C-content has been shown to be relevant in PCR success prediction (Benita *et al.*, 2003).

The comparison of the model performance in the ROC space discriminates 1.60 as the classifier threshold which leads to the worst model behaviour, but shows only minor differences for the other cut-offs. The analysis of PR curves allows further comparisons and highlights that 1.80 shows the highest degree of resolution. Tuomi *et al.* (2010) described 1.80 as boundary for optimized PCR reactions.

It is worth noting that the tool developed aids in primer design prior to the wet lab experiments. Since it remains clear that there are physical constraints which establish the maximum PCR efficiency of a given set of one amplicon and a pair of oligos, bias is introduced in many ways (pipetting, reactives, PCR machine, etc.). We would like to point out that our work does not intend to substitute the experimental efficiency calculation nor modify the quantification settings; its aim is to partner primer design and thus allow obtaining results for normalization that do not require primer combination testings.

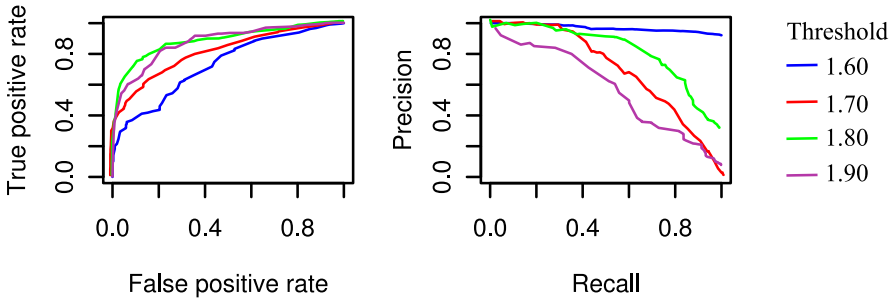


Figure 3.3: ROC and PR curves. ROC and PR curves are plotted for various experimental efficiency thresholds, which define the decision criteria of succesful PCR performance. A good behaviour in ROC space is to be in the upper-left-hand corner, whereas in PR space the goal is to locate at the upper-right-hand corner.

3.5 Conclusions

Using a wide range of amplicons and PCR set-ups, we statistically modelized the response of the PCR efficiency value, a parameter affecting PCR success and involved in effective gene expression quantitation. In order to ease PCR primer design for Q-PCR experiments, the efficiencypredicting model was included in the Primer3 design pipeline and freely provided as a web tool. This tool should help to generate primer combinations with similar theoretical efficiencies to well established PCR primers or to ease multiplex PCR reactions where efficiencies should be similar among templates.

3.6 Availability and requirements

- Project name: pcrEfficiency, a web tool for PCR efficiency estimation
- Project home page: `http://srvgen.upct.es/efficiency.html`
- Operating system(s): Platform independent
- Programming language: R, python
- Other requirements: None
- License: GNU GPL v2.

- Any restrictions to use by non-academics: None.

3.7 Authors' contributions

IM, ME and JW carried out the design of the study. IM and JW performed the Q-PCR experiments and participated in data collection. IM designed the algorithms, developed the model and programmed the application. IM and ME wrote the manuscript. All authors read and approved the final manuscript.

3.8 Acknowledgements

This work was funded by Fundación Séneca 11895/PI/09 and MICINN BFU-2010-15843. IM received a grant from Fundación Séneca. The authors would like to thank Luis Pedro García and SEDIC for aid in conducting the computational-intensive part of this work, as well the web server set-up. Luciana Delgado-Benarroch is acknowledged for web usability suggestions and Marta Pawluczyk for comments on the manuscript.

Chapter 4

Validation of reference genes for quantitative real-time PCR during leaf and flower development in *Petunia hybrida*

Identification of genes with invariant levels of gene expression is a prerequisite for validating transcriptomic changes accompanying development. Ideally expression of these genes should be independent of the morphogenetic process or environmental condition tested as well as the methods used for RNA purification and analysis.

In an effort to identify endogenous genes meeting these criteria nine reference genes (RG) were tested in two Petunia lines (Mitchell and V30). Growth conditions differed in Mitchell and V30, and different methods were used for RNA isolation and analysis. Four different software tools were employed to analyze the data. We merged the four outputs by means of a non-weighted unsupervised rank aggregation method. The genes identified as optimal for transcriptomic analysis of Mitchell and V30 were *EF1α* in Mitchell and *CYP* in V30, whereas the least suitable gene was *GAPDH* in both lines.

The least adequate gene turned out to be *GAPDH* indicating that it should be rejected as reference gene in Petunia. The absence of correspondence of the best-suited genes suggests that assessing reference gene stability is needed when performing normalization of data from transcriptomic analysis of flower and leaf development.[10]

[10]Izaskun Mallona, Sandra Lischewski, Julia Weiss, Bettina Hause, and Marcos Egea–Cortines. Validation of reference genes for quantitative real-time PCR during leaf and flower development in *Petunia hybrida*. *BMC Plant Biology*, 10(1):4, 2010. (Mallona *et al.*, 2010).

Figure 4.1: Developmental stages of leaves and flowers used for RNA extractions. Representative photographs of leaves and flowers of *P. hybrida* lines Mitchell (a, c) and V30 (b, d) are shown. The leaf stages are young, small leaf (leaf of the left in a, b) and fully expanded leaf (leaf of the right in a, b). Flowers at four different developmental stages are shown (c, d). From left to right they range from young flower bud (stage A, 1-1.5 cm), over-elongated bud (stage B, 2.5-3 cm) and pre-anthesis (stage C, 3.5-4.5 cm) to fully developed flower (stage D, open flower). Bar: 1 cm.

4.1 Introduction

The general aims of transcriptomic analysis are identification of genes differentially expressed and measurement of the relative levels of their transcripts. Transcriptomic analysis like that relying on microarray techniques reveals an underlying expression dynamic that changes between tissues and over time (Jiao *et al.*, 2009). Results must then be validated by other means in order to obtain robust data that will support working hypotheses directed at a better understanding of development or environmental responsiveness. Since the advent of quantitative PCR, it has become the method of choice to validate gene expression data. However, data obtained by qPCR can be strongly affected by the properties of the starting material, RNA extraction procedures, and cDNA synthesis. Therefore, relative quantification procedures require comparison of the gene of interest to an internal control, based on a normalization factor derived from one or more genes that can be argued to be equally active in the relevant cell types. This requires the previous identification of such genes, which can then be reliably used to normalise relative expression of genes of interest.

Identification of candidate genes useful for normalization has become a major task, as it has been shown that normalization errors are probably the most common mistake, resulting in significant artefacts that can lead to erroneous conclusions (Gutierrez *et al.*, 2008). Several software tools have

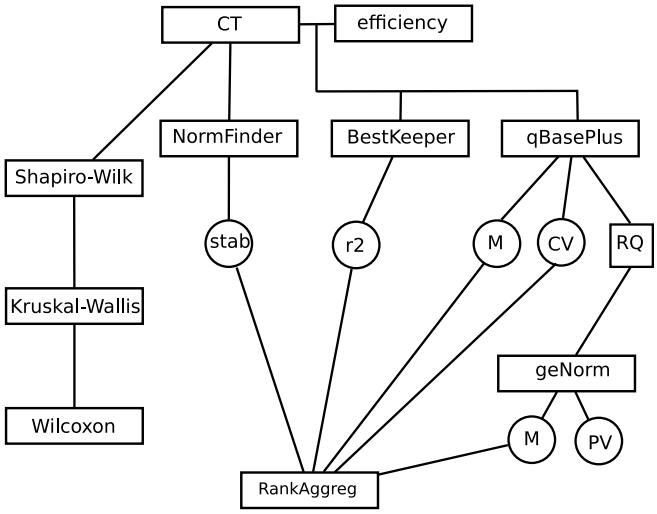


Figure 4.2: Data analysis flow chart. CT (cycle threshold) values were calculated using different thresholds depending on the variety. Efficiency value taken for line Mitchell was 2; for line V30, there was one value for each tube. Circles indicate statistical results to be merged with RankAggreg (Pihur *et al.*, 2009). Relative quantities (RQ) were scaled to the sample with lowest CT value (flower stage C). CT data were checked for normality (Shapiro-Wilk test) and, due to non-normality, they were analysed by non-parametrical tests (Kruskal and Wallis). Since CT values showed non-equal distributions according to the organ from which RNA was extracted, they were further tested using pairwise Wilcoxon tests with Bonferroni's correction with the aim of solving pairwise significant variations. A significance threshold of 0.05 was used. Abbreviations: PV, pairwise variation; M, classical stability value; stab, NormFinder stability value; CV, variation coefficient; r2, determination coefficient - regression to BestKeeper; RQ, relative quantities.

been developed to compute relative levels of specific transcripts (commonly referred to as "gene expression", although obviously transcript stability is also an important factor contributing to transcript levels) based on group-wise comparisons between a gene of interest and another endogenous gene (Pfaffl *et al.*, 2002). However identification of genes with stable patterns of gene expression requires pairwise testing of several genes with each other. Among the software programs developed toward this end are geNorm (Vandesompele *et al.*, 2002), BestKeeper (Pfaffl *et al.*, 2004), NormFinder (Andersen *et al.*, 2004) or qBasePlus (Hellemans *et al.*, 2007). The programs geNorm and qBasePlus use pairwise comparisons and geometric averaging across a matrix of reference genes. qBasePlus also calculates a coefficient of variation (CV) for each gene as a stability measurement. BestKeeper uses pairwise correlation analysis of each internal gene to an optimal normalization factor that merges data from all of them. Finally, NormFinder fits data to a mathematical model, which allows comparison of intra- and intergroup variation and calculation of expression stability.

Using the programs described above researchers have identified genes suitable for use as normalization controls in Arabidopsis (Czechowski *et al.*, 2005), rice (Jain *et al.*, 2006), potato leaves (Nicot *et al.*, 2005), the parasitic plant *Orobanche ramosa* (Gonzalez-Verdejo *et al.*, 2008), *Brachypodium distachyon* (Hong *et al.*, 2008) and grape (Reid *et al.*, 2006). In the Solanaceae, candidate genes for normalization have been determined based on EST abundance (Coker and Davies, 2003), and qPCR followed by statistical analysis using the tools described above have been reported (Exposito-Rodriguez *et al.*, 2008).

A feature shared amongst these studies, and a large number of additional publications describing human, animal and plant systems, is the identification of genes specific for a certain tissue, developmental stage or environmental condition. This is a logical experimental design, as individual research programs tend to be focused, and the number of appropriate genes can be expected to be inversely related to the number of cell types or conditions under investigation. Recent studies that included different cultivars of soybean (Jian *et al.*, 2008), underscore how the characteristics of the plant and the types of organs studied must drive the experimental approach to transcriptomic analysis.

The garden Petunia (*Petunia hybrida*) has been extensively used as a model for developmental biology (Gerats and Strommer, 2009; Gerats and Vandenbussche, 2005). Amongst the inbred Petunia lines used in research, the white-flowered Mitchell (Mitchell *et al.*, 1980), also known as W115,

is routinely exploited for transformation and scent studies (Dudareva and Pichersky, 2008; Boatright *et al.*, 2004; Spitzer *et al.*, 2007a). The genetics of flower pigmentation has been intensively studied in lines such as V30 (Koes *et al.*, 1986b). Mitchell and V30 are genetically dissimilar, as demonstrated in mapping studies, and vary in a number of other ways, including growth habit and amenability to propagation in culture. Here we have used multiple developmental stages of flowers and leaves of these two Petunia lines to identify genes that show reliable robustness as candidates for use in normalization of relative transcript abundance. The experiments were carried out in two different laboratories, with different PCR machines and different purification and amplification conditions. We found that the final shortlist of valuable genes was different between lines suggesting the necessity of performing reference gene stability measurements as part of the experimental design where differences in gene expression in Petunia is tested.

4.2 Material and methods

4.2.1 Plant material

P. hybrida lines Mitchell and V30 were grown in growth chambers. Mitchell plants were grown on ED73 Optifer (Patzer) under a 10 h light/14 h dark cycle, with a constant temperature of 22 °C (60% humidity). V30 plants were germinated in vermiculite and grown in a vermiculite-perlite-turf-coconut fiber mixture (2:1:2:2). Plants were kept under a long day photoperiod (16L: 8 D) with 25 °C in L and 18 °C in D.

Flowers were classified into four developmental stages: flower buds (stage A, 1-1.5 cm), elongated buds (stage B, 2,5-3 cm), pre-anthesis (stage C, 3.5 -4.5 cm) and fully opened flowers shortly before anthesis (stage D) according to Cnudde *et al.* (2003). Leaves were harvested at two different stages, stage A corresponded to young, small leaves and stage C to fully expanded ones. Three independent samples of each of the developmental stages of flowers and leaves were taken.[11]

[11] Each independent sample comprised nine organs (in example, nine flower buds), and three independent samples were taken for each developmental stage.

Table 4.1: Genes, primers and amplicon characteristics. Selected candidate reference genes accessions are shown as SGN identifiers and Arabidopsis TAIR databases (in brackets). Homologous Arabidopsis genes were determined on the basis of tblastx e-values. Amplicon length is presented as bp.

Gene name	Molecular function	Accesion (SGN, TAIR and tblastx e-value)	Sequence	Length
ACT	Actin 11	SGN-U208507 (At3g12110.1, 2e-110)	TGCACTCCCACATGCTATCCT / TCAGCCGAAGTGGTGAAA-GAG	114
CYP	Cyclophilin	SGN-U207595 (At2g21130.1, 1.9e-75)	AGGCTCATCATTCCACCGTGT / TCATCTGCGAACTTAG-CACCG	111
EF1α	Elongation factor 1-alpha	SGN-U207468 (At5g60390.1, 0)	CCTGGTCAAATTGGAAACGG / CAGATCGCCTGT-CAATCTTGG	103
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	SGN-U209515 (At1g42970.1, 9.2e-79)	AACAACTCACTCCTACACCGG / GGTAGCACTAGAGA-CACAGCCTT	135
RPS13	Ribosomal protein S13	SGN-U208260 (At4g00100.1, 4e-77)	CAGGCAGGTTAAGGCAAAGC / CTAGCAAGGTACA-GAAACGGC	114
RAN1	GTPase nuclear protein	SGN-U207968 (At5g20010.1, 1e-119)	AAGCTCCCACCTGTCTGGAAA / AACAGATTGCCGGAAGCCA	103
SAND	SAND family protein	SGN-U210443 (At2g28390.1, 8.2e-76)	CTTACGACGAGTTCAGATGCC / TAAGTCCTCAACACG-CATGC	135
TUB	Tubulin beta-6 chain	SGN-U207876 (At5g12250.1, 6e-147)	TGGAAACTCAACCTCCATCCA / TTTCGTCCATTCCTTCAC-CTG	114
UBQ	Polyubiquitin	SGN-U207515 (At4g02890.2, 8e-107)	TGGAGGATGGAAGGACTTTGG / CAGGACGACAACAAG-CAACAG	153

4.2.2 RNA isolation and cDNA synthesis

Mitchell material

Total RNA was isolated from 100 mg homogenized plant material using an RNeasy Mini Kit (Qiagen, Hilden, Germany). Putative genomic DNA contamination was eliminated by treatment with recombinant DNase I (Qiagen) as recommended by the vendor. RNA concentration and purity was estimated from the ratio of absorbance readings at 260 and 280 nm and the RNA integrity was tested by gel electrophoresis. cDNA synthesis was performed using M-MLV reverse transcriptase (Promega, Mannheim, Germany) starting with 1 μg of total RNA in a volume of 20 μL with oligo(dT)19 primer at 42 °C for 50 min.

V30 material

Samples were homogenized in liquid nitrogen with a mortar and pestle. Total RNA was isolated using the NucleoSpin RNA Plant (Macherey-Nagel, Düren, Germany) according to the manufacturer's protocol. This RNA isolation kit contains DNaseI in the extraction buffer, added to the column once RNA is bound to the spin column. RNA was measured by photometry at 260 nm and quality-controlled on denaturing agarose gels. Total RNA (0.8 μg) was transcribed using the SuperScript III (Invitrogen Corp., Carlsbad, CA) and oligodT20 employing 10 μL 2x RT reaction mix, 2 μL RT enzyme mix and 8 μL RNA. Reverse transcription was performed on a GeneAmp Perkin-Elmer 9700 thermocycler (Perkin Elmer, Norwalk, CT, USA) by using the following programme: 10 min at 25 °C, 30 min at 50 °C and 5 min at 85 °C; addition of 1 u of *Escherichia coli* RNAse H, and incubation for 2 h at 15 °C.

4.2.3 PCR optimisation

We selected nine genes to be tested as reference transcripts (*ACT*, *CYP*, *EF1α*, *GAPDH*, *RAN1*, *RPS13*, *SAND* and *UBQ*) based on previous descriptions (table 4.1). PCR conditions were optimised using cDNA from leaves (stage A) in a Robocycler gradient 96 (Stratagene, La Jolla, CA) and GoTaq Flexi DNA polymerase (Promega) in a 25 μL reaction containing: 2 μL of cDNA, 2 mM MgCl2, 0.2 mM each dNTP, 0.4 μL of each primer and 1.25 U enzyme.

4.2.4 Real-time PCR

Mitchell

Real-time PCR was performed in an Mx 3005P QPCR system (Stratagene, La Jolla, CA) using a SYBR Green based PCR assay (with ROX as the optional reference dye; Power SYBR Green PCR Mastermix, Applied Biosystems, Foster City, CA). A master mix containing enzymes and primers was added individually per well. Each reaction mix containing a 15 ng RNA equivalent of cDNA and 1 pM gene-specific primers (table 4.1) was subjected to the following protocol: 95 °C for 10 min followed by 50 cycles of 95 °C for 30 sec, 60 °C for 1 min and 72 °C for 30 sec, and a subsequent standard dissociation protocol. As a control for genomic DNA contamination, 15 ng of total non-transcribed RNA was used under the same conditions as described above. All assays were performed in three technical replicates, as well three biological replicates [12].

V30

Reactions were carried out with the SYBR Premix Ex Taq (TaKaRa Biotechnology, Dalian, Jiangsu, China) in a Rotor-Gene 2000 thermocycler (Corbett Research, Sydney, Australia) and analysed with Rotor-Gene analysis software v6.0 as described before (Manchado-Rojo *et al.*, 2008) with the following modifications: reaction profiles used were 40 cycles of 95 °C for 30 s, 55 °C or 60 °C for 20 s, 72 °C for 15 s, and 80 °C for 15 s, followed by melting at 50-95 °C employing the following protocol: 2 μL RNA equivalent of cDNA, 7.5 μL SYBR Premix Ex Taq 2x, 0.36 μL of each primer at 10 μM and 4.78 μL distilled water. Annealing temperature was 55 °C (*TUB*, *CYP*, *ACT*, *EF1α*, *GAPDH*, and *SAND*) or 60 °C (*RPS13*, *UBQ*, *RAN1*) according to the previous optimisation. In order to reduce pipetting variability, we performed reaction batches containing primer pairs, and templates were added in the end. Three technical replicates for each reaction and non-template controls were included, as well three biological replicates.

4.2.5 Bioinformatics and statistical analysis

Data analysis strategy is described in detail in results. Reaction efficiency calculus was done using the amplification curve fluorescence, analyzing

[12] Runs were baseline corrected either by standard normalization (substraction of the

each tube separately as described by Liu and Saint (2002). It was calculated as shown in equation 4.1 :

$$Efficiency = \frac{F_n}{F_{n-1}} \tag{4.1}$$

in which n is defined as the 20% value of the fluorescence at the maximum of the second derivative curve. Curve was defined by one measure in each amplification cycle. We used only the exponential phase of the amplification reaction. Software packages included geNorm v3.4, the excel add-in of NormFinder v0.953 , BestKeeper v1 and qBasePlus v1.2 . Other statistical procedures were performed with the R environment v2.7.1 with the packages stats v2.7.1 , multcompView v0.1-0 and RankAggreg v0.3-1 (Pihur *et al.*, 2009).

4.3 Results

4.3.1 Petunia lines, developmental stages and selection of genes for normalization

Two very different Petunia lines were used for the analyses. Mitchell, also known as W115 , is a doubled haploid line obtained from anther culture of an interspecific Petunia hybrid (Mitchell *et al.*, 1980); it is characterized by vigorous growth, exceptional fertility, strong fragrance and white flowers. V30 is an inbred line of modest growth habit and fertility featuring deep purple petals and pollen. From each line we harvested flowers representing four developmental stages, from young flower buds to open flowers shortly before anthesis, and two leaf developmental stages, young and full-sized (figure 4.1).

Potentially useful RG were selected based on review of the relevant literature, from which we identified genes previously used for normalization or routinely used as controls for northern blots or RT-PCR. From the original list we developed a short list of nine (table 4.1), including genes encoding *Actin-11* (*ACT*), *Cyclophilin-2* (*CYP*)(Nicot *et al.*, 2005), *Elongation factor 1α* (*EF1α*), *Ubiquitin* (*UBQ*) *Glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*), *GTP-binding protein RAN1* (*RAN1*), *SAND protein* (*SAND*) (Czechowski *et al.*, 2005; Bey *et al.*, 2004; Delgado-Benarroch

fluorescence present in the first five cycles of each sample) or by "dynamic tube" normalization.

et al., 2009a), *Ribosomal protein S13* (*RPS13*) (Andersen *et al.*, 2004) and *β-Tubulin 6* (*TUB*) (Zenoni *et al.*, 2004). The products of these genes are associated with a wide variety of biological functions. Moreover, these genes are described as not co-regulated, a prerequisite for using one of the algorithms to identify stably expressed genes (geNorm) reliably (Vandesompele *et al.*, 2002).

4.3.2 Strategy for data mining and statistical analysis

The genes described above were selected to test for stability of transcript levels through leaf and flower development in two Petunia lines, Mitchell and V30. As the aim of the present work is to find if we could obtain a similar rank of genes irrespective of the Petunia line, growth conditions or sample processing, we developed all the data mining procedures separately for each line. Cycle threshold (CT) values were determined and expression stability, i.e., the constancy of transcript levels, ranked. As a strategy for calculating relative expression quantities (RQ) we applied the qBasePlus software [13], taking into account for each reaction its specific PCR efficiency. Rescaling of normalized quantities employed the sample with the lowest CT value (see materials and methods in section 4.2 and figure 4.2). With qBasePlus we measured expression stability (M values) [14]

[13] The exact definition of the normalized, relative quantities (NRQ) is as depicted in equation 4.2:

$$NRQ = \frac{E_{goi}^{\Delta Ct,goi}}{\sqrt[f]{\prod_o^f E_{ref_o}^{\Delta Ct,ref_o}}} \tag{4.2}$$

The formula is an improvement of the classic ΔΔ-Ct method published by Hellemans *et al.* (2007). Note that both the efficiency (E) and the crossing point (Ct) are included; moreover, several reference genes (ref) can be used simultaneously to extract the relative quantities of a gene of interest (goi).

[14] According to Vandesompele *et al.* (2002), starting with a reference gene called j and another k, the M value computation relies on log-transforming all the expression rations a_{ij}/a_{ik} as depicted in equation 4.3. Then, denoting s_{jk} the standard deviation of such collection of expression ratios A_{jk}, the gene stability measure M_j is defined as the arithmetic mean of the pairwise variations s_{jk} obtained from log-expression ratios (equation 4.4).

$$A_{jk} = \left\{ log_2(\frac{a_{1j}}{a_{1k}}), log_2(\frac{a_{2j}}{a_{2k}}), \cdots, log_2(\frac{a_{mj}}{a_{mk}}) \right\} \tag{4.3}$$

$$M_j = \frac{\sum_{k=1}^{n} s_{jk}}{n-1} \tag{4.4}$$

and coefficients of variation [15](CV values). Relative quantities were transferred to geNorm for computing M stability values. It is worth noting that the procedure for computing M values differs between geNorm and qBasePlus. Finally, we used the combined stability measurements produced by geNorm, NormFinder, BestKeeper and qBasePlus to establish a consensus rank of genes by applying RankAggreg (Pihur *et al.*, 2009). The input to this statistical package was a matrix of rank-ordered genes according to the different stability measurements previously computed. RankAggreg calculated Spearman footrule distances and the software reformatted this distance matrix into an ordered list that matched each initial order as closely as possible. This consensus rank list was obtained by means of the Cross-Entropy Monte Carlo algorithm present in the software.

4.3.3 CT values and variability between organs and developmental stages in Mitchell and V30

Real-time PCR reactions were performed on the six cDNA samples obtained from each Petunia line with the nine primer pairs representing the candidate RG. In order to assess run reliability non-template controls were added and three technical repetitions were included for each biological replicate. CT values were defined as the number of cycles required for normalized fluorescence to reach a manually set threshold of 20% total fluorescence. Product melting analysis and/or gel electrophoresis allowed for the discarding of non-specific products. Moreover, we considered only CT technical repetitions differing by less than one cycle.

The CT values obtained for all the genes under study differed between the two Petunia lines (figure 4.3). The range of values was consistently narrower in Mitchell than in V30. This could indicate that gene expression in general is less variable in Mitchell than in V30, however these data correspond to averages derived from all the samples and further analysis showed that in fact V30 exhibited more constant levels of tested transcripts at the single organ level or developmental stage (see below).

For Mitchell samples *UBQ* was the most highly expressed gene overall, with a CT of 14.8, and *SAND* the lowest, with a CT of 21.2. In contrast,

[15]Calculation of the coefficient of variation (CV) of a given reference gene p across a number of samples is represented as equation 4.5 (Hellemans *et al.*, 2007). SE stands for standard error.

$$CV_p = \frac{SE\,(NRQ_p)}{\overline{NRQ}_p} \tag{4.5}$$

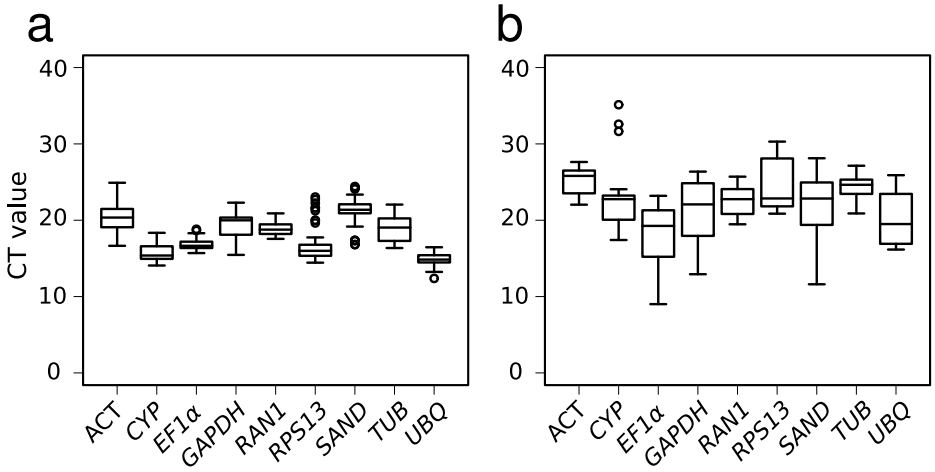


Figure 4.3: Expression profiling of reference genes in different organs and Petunia lines. Expression profiling of reference genes in different organs and Petunia lines. CT values are inversely proportional to the amount of template. Global expression levels (CT values) in the different lines tested are shown as 25th and 75th quantiles (horizontal lines), median (emphasized horizontal line) and whiskers. Whiskers go from the minimal to maximal value or, if the distance from the first quartile to the minimum value is more than 1.5 times the interquartile range (IQR) , from the smallest value included within the IQR to the first quartile. Circles indicate outliers, the values smaller or larger than 1.5 times the IQR. (a) Mitchell, (b) V30.

the highest and lowest expressed genes in V30 were *EF1α* and *ACT*, with CTs of 18.3 and 25.1, respectively.

Analysis of variance of CT values between organs was performed separately for Mitchell and V30 samples. Since CT values were not normally distributed, we calculated Kruskall-Wallis and a post-hoc Pairwise Rank Sum Wilcoxon test, both non-parametrical, using a Bonferroni correction and a significance cut-off of 0.05. In Mitchell the genes *RAN1*, *RPS13* and *UBQ* showed significant differences in transcript levels between developmental stages (see the additional file 1 of the published manuscript). *RAN1* transcript levels differed significantly between leaf A and flowers C and D, *RPS13* differed in flower D from the rest of floral stages analysed, and *UBQ* transcript levels differed significantly between leaf A and flower D. For V30, the overall CT variability was higher than that seen in Mitchell; in fact, expression of all the genes analysed showed significant differences between one or more sets of organs and/or developmental stages. Expression of the genes *GAPDH* and *TUB* differed between leaves A and C, while levels of other measured transcripts were essentially the same in the two leaf stages. In contrast, during flower development, we could distinguish genes that showed two levels of significantly different CT values (*GAPDH* and *TUB*), those that showed three (*ACT*, *CYP*, *EF1α* and *RPS13*) and others that differed at each developmental stage analysed (*RAN1*, *SAND* and *UBQ*).

4.3.4 Stability of gene expression in Mitchell and V30

Data from each of the two chosen Petunia lines were analyzed separately. As a first approach, we applied data as a unique population and transferred it to NormFinder, BestKeeper, geNorm and qBasePlus according to the flowchart plotted in figure 4.2. In a second approach, we subdivided data into several subpopulations, corresponding to unique developmental stages (i.e., flower C or leaf A), then, piped this data into the qBasePlus and geNorm tools. The results of both sets of analyses are presented in tables 4.2 and 4.3 (additional files are available as supplemental material of the published version (Mallona *et al.*, 2010)).

Table 4.2: M values computed by geNorm and qBasePlus allow to rank optimal reference genes. For each organ and mix of organs the two top-ranked genes are shown. The number of genes required for a reliable quantification is established using a Pairwise Variation (PV) cut-off of 0.15; n is the the minimum number of control genes required NA means that no one pairwise variation was under the proposed cut-off.

Mitchell								
Statistic			Flower				Leaf	
	A	B	C	D	A+B+C+D	A	C	A+C
M geNorm	*EF1α*	*SAND*	*EF1α*	*RAN1*	*RAN1*	*EF1α*	*RAN1*	*EF1α*
	(0.05)	(0.14)	(0.13)	(0.08)	(0.47)	(0.11)	(0.14)	(0.37)
	RPS13	*UBQ*	*RPS13*	*RPS13*	*SAND*	*RPS13*	*SAND*	*RPS13*
	(0.05)	(0.14)	(0.13)	(0.08)	(0.47)	(0.11)	(0.14)	(0.37)
M qBasePlus	*ACT*	*EF1α*	*CYP*	*EF1α*	*EF1α*	*RPS13*	*SAND*	*EF1α*
	(0.55)	(0.60)	(0.60)	(0.30)	(0.80)	(0.77)	(0.93)	(0.85)
	RPS13	*RPS13*	*SAND*	*RAN1*	*SAND*	*EF1α*	*RAN1*	*RAN1*
	(0.56)	(0.60)	(0.64)	(0.34)	(0.82)	(0.78)	(0.93)	(0.92)
CV qBasePlus	*ACT*	*RPS13*	*CYP*	*EF1α*	*EF1α*	*RPS13*	*SAND*	*RAN1*
	(0.05)	(0.07) *CYP*	(0.09)	(0.05) *TUB*	(0.25)	(0.14)	(0.12)	(0.20)
	SAND	(0.25)	*SAND*	(0.14)	*SAND*	*RAN1*	*RAN1*	*EF1α*
	(0.15)		(0.11)		(0.28)	(0.18)	(0.18)	(0.21)
Min. number	2 (0.04)	2 (0.07)	2 (0.12)	2 (0.05)	4(0.15)	2 (0.10)	2 (0.13)	2 (0.12)
V30								
Statistic			Flower				Leaf	
	A	B	C	D	A+B+C+D	A	C	A+C
M geNorm	*RAN1*	*TUB*	*RPS13*	*ACT*	*RAN1*	*TUB*	*RPS13*	*RAN1*
	(0.11)	(0.12) *CYP*	(0.23)	(0.02) *CYP*	(0.45)	(0.07)	(0.09) *TUB*	(0.20)
	UBQ	(0.12)	*UBQ*	(0.02)	*ACT*	*RAN1*	(0.09)	*UBQ*
	(0.11)		(0.23)		(0.45)	(0.07)		(0.20)
M qBasePlus	*CYP*	*CYP*	*SAND*	*RPS13*	*ACT*	*UBQ*	*SAND*	*UBQ*
	(0.30)	(0.66) *TUB*	(0.63) *CYP*	(0.29)	(2.27)	(0.06) *TUB*	(0.27)	(0.49)
	RAN1	(0.69)	(0.63)	*EF1α*	*RAN1*	(0.09)	*RPS13*	*RPS13*
	(0.33)			(0.30)	(2.44)		(0.29)	(0.51)
CV qBasePlus	*CYP*	*RAN1*	*SAND*	*EF1α*	*ACT*	*UBQ*	*SAND*	*UBQ*
	(0.05) *TUB*	(0.09) *CYP*	(0.10) *ACT*	(0.06) *CYP*	(0.70)	(0.06) *TUB*	(0.03)	(0.14)
	(0.10)	(0.11)	(0.25)	(0.06)	*RAN1*	(0.09)	*RPS13*	*RPS13*
					(0.82)		(0.06)	(0.16)
Min. number	2 (0.10)	2 (0.07)	2 (0.09)	2 (0.09)	NA	2 (0.04)	2 (0.03)	2 (0.07)

CT values were log-transformed and used as input for the NormFinder tool, which fitted this data into a mathematical model based on six independent groups corresponding to single developmental stages. Estimates for stability of gene expression are based on the comparison between inter- and intra-group variability. In the Mitchell line, the gene exhibiting the most stable level of expression was *EF1α* (stability value of 0.018) and *CYP* and *EF1α* represented the best combination (0.017). In V30, NormFinder estimated *UBQ* (0.053) as the most stably expressed gene, and *RAN1* and *UBQ* (0.069) as the best combination of two genes.

CT values and one efficiency value for each primer pair served as input for the BestKeeper package. This program was intended to establish the best-suited standards out of the nine RG candidates, and to merge them in a normalization factor called the BestKeeper index. Because BestKeeper software is designed to determine a reliable normalization factor but not to compute the goodness of each RG independently, we took as the stability-of-expression value the coefficient of determination of each gene to the BestKeeper index. BestKeeper calculated the highest reliability for *CYP* in line Mitchell and V30 finding *GAPDH* as the least suitable gene in Mitchell and *TUB* in V30.

qBasePlus and geNorm calculate M stability values by a slightly different procedure. This parameter is defined as the average pair-wise variation in the level of transcripts from one gene with that of all other reference genes in a given group of samples; it is inversely related to expression stability. However, because the inclusion of a gene with highly variable expression can alter the estimation of the rest, geNorm (but not qBasePlus) performs a stepwise exclusion of the least stably expressed genes. Taking into account the entire dataset from Mitchell with geNorm, *RAN1* and *SAND* were calculated to be the most stably expressed genes (M value 0.5), *GAPDH* the least (1.15). In V30, *RPS13* and *UBQ* were calculated to be the genes of least variable expression (0.64), whereas *GAPDH* was the most variable (2.61). In terms of qBasePlus M values, *EF1α* was valued as the best gene for Mitchell (0.85) and *GAPDH* the worst (1.76); for V30, *ACT* was ranked as the most valuable gene (2.11) and *GAPDH* was the worst (3.66).

Considering each developmental stage separately, we found that M values were consistently higher in Mitchell than in V30, suggesting more variable levels of RG expression in Mitchell. Flower stage D exhibited the most stable expression pattern in both lines (figure 4.4). It is noteworthy that stability of transcript levels between reproductive and vegetative modules

differed in the two lines. In general, M values calculated with qBasePlus, were higher in flowers stage C and D than in leaves from Mitchell, whereas V30 showed an opposite trend. A remarkable case was *GAPDH*, with an M value four times higher in Mitchell than in V30 at leaf stage C, whereas it was three times lower in Mitchell compared to V30 at flower stage A (see table 4.2).

Mean CV value, a measurement of the variation of relative quantities of RNA for a normalized reference gene, showed little difference between lines, with a value of 0.42 in Mitchell and 0.44 in V30, for data analysed as a whole.

4.3.5 Determination of the number of genes for normalization

Quantification of gene expression relative to multiple reference genes implies the calculation of a normalization factor (NF) that merges data from several internal genes. Determination of the minimal number of its components is estimated by computing the pairwise variation (PV) of two sequential NFs ($\frac{V_n}{V_{n+1}}$) as the standard deviation of the logarithmically transformed $\frac{NF_n}{NF_{n+1}}$ ratios, reflecting the effect of including an additional gene (Vandesompele *et al.*, 2002). If the pairwise variation value for n genes is below a cut-off of 0.15, additional genes are considered not to improve normalization. The number of genes required for normalization was determined to be two for both Mitchell and V30, except when either different floral developmental stages or vegetative and reproductive stages were mixed (see table 4.2).

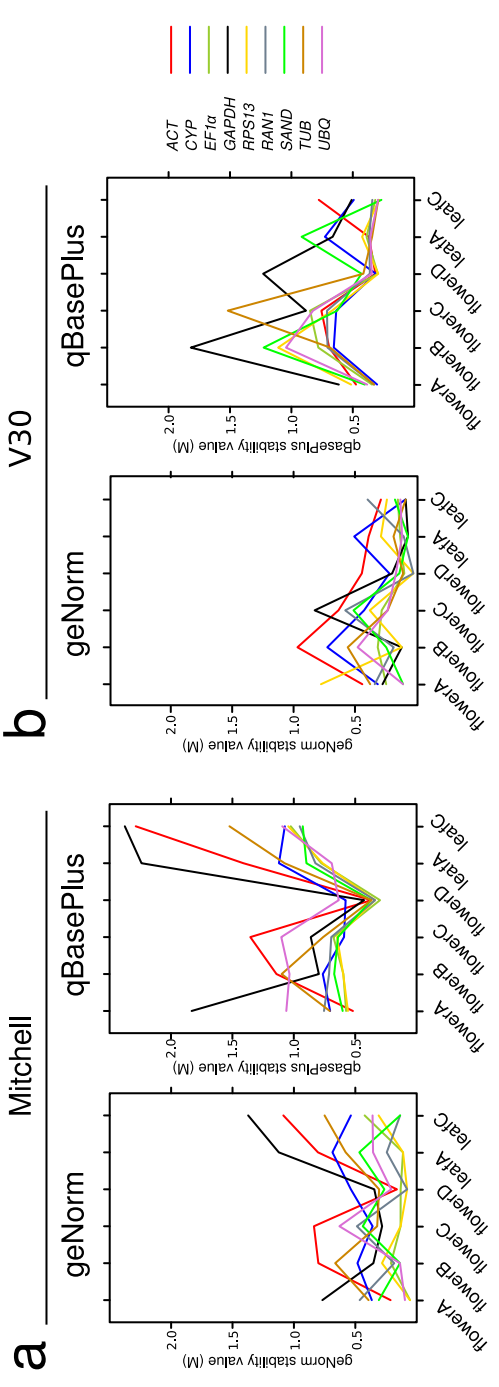


Figure 4.4: geNorm and qBasePlus average expression stability measure values for reference genes in individual samples. Average expression stability values (M values) are an inverse proportion measure of expression stability. GeNorm computes M values by stepwise exclusion of the least stable gene. (a) geNorm and qBasePlus output of Mitchell samples. (b) geNorm and qBasePlus output of V30 samples.

The PV values showed the same trend as that seen for stability measurements, i.e., the developmental stage with the lowest average PV was flower stage D, both in Mitchell and V30. In contrast, gene expression in leaves of Mitchell showed more variability, with higher PV values, than those of V30 (figure 4.5).

4.3.6 Consensus list of similarities between lines

The different software programs used to determine gene suitability for normalization of gene expression give slightly different results and statistical stability values for each gene. We arranged the internal genes in five lists according to the rank positions generated by each of the five statistical approaches, M values by geNorm and qBasePlus, NormFinder stability value, coefficient of determination to BestKeeper and CV of qBasePlus. These lists were used to create an aggregate order, with the aim of obtaining an optimal list of genes for each Petunia line. The results of the merged data revealed that the most adequate of the genes tested for normalization in Mitchell are *EF1α*, *SAND* and *RPS13*; the three showing the lowest reliability are *TUB*, *ACT* and *GAPDH* (figure 4.6). For V30, the best candidate genes are *CYP*, *RAN1* and *ACT*, while the three lowest ranking are *EF1α*, *SAND* and *GAPDH*. Thus none of the genes found as highly reliable coincide between the lines. Despite of that, *GAPDH* was highly unstable in both lines.

4.4 Discussion

4.4.1 Identification of robust normalization genes for Petunia

We have attempted to identify a set of genes suitable for normalization of transcript levels in *P. hybrida*. Since several Petunia lines are used for research, we based this work on two that are extensively used for different purposes. In an effort to reflect different growth environments typical of distinct lab setups, plants of each line were grown in a set of conditions, differing in photoperiod, thermoperiod and growth substrate between lines (see methods). RNA was isolated using different RNA extraction kits, and amplifications were carried out using different reagents and PCR machines. The experimental design aimed to maximize potential variability in transcript abundance for the putative RG under study. Highly contrasting

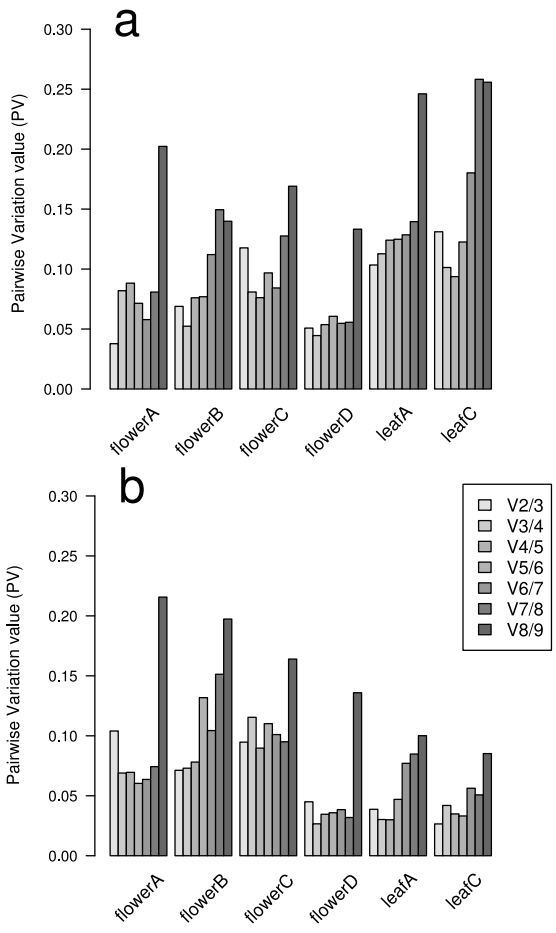


Figure 4.5: Minimum number of genes necessary for reliable and accurate normalization. GeNorm pairwise variation values (PV values) are computed by an algorithm which measures pairwise variation (v) between two sequential normalization factors NF_n and NF_{n+1}, where n is the number of genes involved in the normalization factor. (a) refers to Mitchell line and (b) to V30.

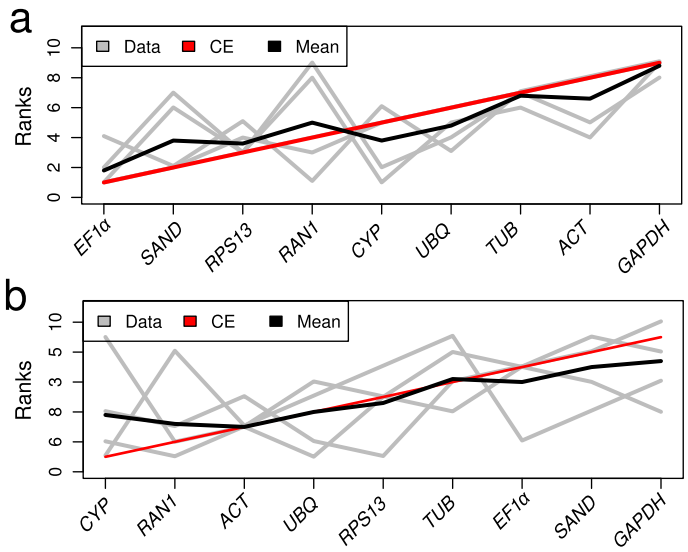


Figure 4.6: Rank aggregation of gene lists using the Monte Carlo algorithm. Visual representation of rank aggregation using Monte Carlo algorithm with the Spearman footrule distances . (a) refers to Mitchell line and (b) to V30. The solution of the rank aggregation is shown in a plot in which genes are ordered based on their rank position according to each stability measurement (grey lines). Mean rank position of each gene is shown in black, as well the model computed by the Monte Carlo algorithm (red line).

results would suggest that every laboratory do a pilot experiment to identify genes suitable for use in normalization; similar results between the two systems would point to a set of genes reliable for broad application, minimally for the lines and developmental stages described.

Our findings in terms of line-associated variability were not in accordance with the results from a soybean study comparing different cultivars. Results of that study suggested no highly relevant cultivar influence on RG suitability (Jian *et al.*, 2008). A similar study has been reported in coffee, for which average M stability values for leaves from different cultivars were lower than that for different organs of a single cultivar. Our result suggests that there are differences in gene expression between same tissues from different lines as well as different tissues from the same line.

4.4.2 Noise in gene expression patterns

Development of petals, like that of many tissues and organs in Petunia, is characterized by a spatial and temporal gradient of cell division that is eventually replaced by cell expansion (Reale *et al.*, 2002). However the experiments described here used whole flower tissues including full petals along with sepals, stamens and carpels. This imposes a general requirement that any gene emerging as robust be differentially regulated to a huge extent neither in the various tissues analyzed together nor in these tissues at different stages of maturation. One interesting aspect of our findings was the identification of flower stage C as a particularly noisy developmental stage compared to early or fully developed flowers. The transition between cell division and expansion in petals, or other flower tissues during this developmental stage, might explain the increased noise. An alternative non-exclusive explanation is that the intermediate stages of flower development are generally less tightly defined than the open flower stage.

Table 4.3: Gene suitability rankings for the whole dataset. Gene expression data were analyzed using five statistical parameters in both Petunia lines. Each column refers to a gene suitability ranking computed by one statistical tool, taking into account all data of a Petunia line.

Rank position	NormFinder		BestKeeper		qBasePlus M		qBasePlus CV		geNorm M		Consensus	
	Mitchell	V30	Mitchell	V30	Mitchell	V30	Mitchell	V30	Mitchell	V30	Mitchell	V30
1	*EF1α*	*UBQ*	*CYP*	*CYP*	*EF1α*	*ACT*	*EF1α*	*RAN1*	*RAN1*	*RPS13*	*EF1α*	*CYP*
2	*CYP*	*RAN1*	*EF1α*	*EF1α*	*SAND*	*RAN1*	*SAND*	*CYP*	*SAND*	*UBQ*	*SAND*	*RAN1*
3	*RPS13*	*ACT*	*RPS13*	*ACT*	*RAN1*	*CYP*	*RPS13*	*ACT*	*UBQ*	*RAN1*	*RPS13*	*ACT*
4	*UBQ*	*GAPDH*	*ACT*	*SAND*	*RPS13*	*TUB*	*RAN1*	*TUB*	*EF1α*	*CYP*	*RAN1*	*UBQ*
5	*ACT*	*RPS13*	*UBQ*	*UBQ*	*CYP*	*RPS13*	*CYP*	*RPS13*	*RPS13*	*ACT*	*CYP*	*RPS13*
6	*SAND*	*SAND*	*TUB*	*GAPDH*	*UBQ*	*UBQ*	*UBQ*	*UBQ*	*CYP*	*TUB*	*UBQ*	*TUB*
7	*TUB*	*EF1α*	*SAND*	*RPS13*	*TUB*	*EF1α*	*TUB*	*EF1α*	*TUB*	*EF1α*	*TUB*	*EF1α*
8	*GAPDH*	*TUB*	*RAN1*	*RAN1*	*ACT*	*SAND*	*ACT*	*GAPDH*	*ACT*	*SAND*	*ACT*	*SAND*
9	*RAN1*	*CYP*	*GAPDH*	*TUB*	*GAPDH*	*GAPDH*	*GAPDH*	*SAND*	*GAPDH*	*GAPDH*	*GAPDH*	*GAPDH*

Leaf development similarly consists of cell growth followed with cell expansion (Beemster *et al.*, 2006). However, an important difference between floral and leaf development is that leaves perform their essential function, e.g., photosynthesis, from a very early stage such that developing leaf tissue is always a mixture of at least three processes: growth, cell morphogenesis and differentiated cell function. This combination of processes might account for the increased gene expression noise observed.

4.4.3 Number of genes required for normalization of gene expression in Petunia

Gathering data from several RG into a normalization factor is currently an accepted method of accurate relative quantification of gene expression (Autran *et al.*, 2002). Moreover, this method has been statistically and empirically validated (Reid *et al.*, 2006; Gabrielsson *et al.*, 2005). Ideally the number of genes required should be low enough to make experimental procedures affordable, and high enough to merit confidence in the conclusions. The PV value obtained for both Mitchell and V30 was very low. Although the value tended to be higher in Mitchell, the number of genes deemed necessary for normalization was the same for both lines: using the proposed cut-off of 0.15 and comparing single developmental stages, the required number was two for Mitchell and V30. The requirement for only two genes is low compared to the results reported for other phylogenetically related species (Nicot *et al.*, 2005; Exposito-Rodriguez *et al.*, 2008; Brunner *et al.*, 2004) and will require significantly less work than the previously suggested minimum of three genes (Vandesompele *et al.*, 2002).

4.4.4 Data mining strategies and consensus list of genes for normalization

The present research aims to identify the control genes best suited for use in gene expression studies in several organs of two Petunia lines. The candidate RG combined classical and recently identified genes. Since each software package can introduce bias, we employed several tools in our analysis. As discussed by other authors, geNorm bases its stability measurement on pairwise comparisons of relative expression quantities of all the panel of genes in the material of interest requiring a suite of non-coregulated RG (Andersen *et al.*, 2004). BestKeeper and NormFinder examine primarily CT values, whereas qBasePlus and geNorm evaluate RQ, a consequence

of which is that PCR efficiency dissimilarities can affect stability measurements (Jian *et al.*, 2008). Nevertheless, some of these algorithms are intrinsically biased because they assume that data are normally distributed. For instance BestKeeper is based on Pearson correlation analysis, which requires normally distributed and variance homogeneous data. The author described this problem and suggested further versions of the software in which Spearman and Kendall Tau correlation should be used (Pfaffl *et al.*, 2004). However, those versions are currently not available.

Our plant material diverged in the variability of statistical outputs amongst lines. V30 showed a high variability in terms of raw expression data (CT values) and low in terms of expression stability measurements, whereas Mitchell showed the opposite responses. Our global analysis merged different statistics, some of which are CT-based and others RQ-based, with the aim of counteracting this biasing influence.

Summarizing the results of our entire dataset analysis, geNorm recommended use of *RAN1* and *SAND* genes for Mitchell and *RPS13* and *UBQ* for V30 and discouraged use of *GAPDH* for both lines. Non-suitability of *GAPDH* has been described by several authors (Suzuki *et al.*, 2000; Ke *et al.*, 2000). Regarding to Solanaceae, its unsuitability has been confirmed in tomato (Exposito-Rodriguez *et al.*, 2008) but it was selected as a stable RG in coffee (Cruz *et al.*, 2009). Due to its sequential exclusion of the least stable gene in the M value calculus algorithm, geNorm M values can differ from those of qBasePlus. qBasePlus corresponded with geNorm, evaluating *EF1α* as the most reliable gene in line Mitchell but differed in line V30, recommending *ACT* as the best candidate. *EF1α* suitability has been confirmed in potato during biotic and abiotic stress (Nicot *et al.*, 2005), atlantic salmon (Olsvik *et al.*, 2005) and several developmental stages of *Xenopus laevis* (Sindelka *et al.*, 2006). Expression of *ACT* genes differs depending on the family member. *ACT*2/7 has been reported as a stably expressed gene whereas *ACT11* was reported as unstable (Jian *et al.*, 2008; Paolacci *et al.*, 2009). It is worth noting that the *ACT* gene used in this study corresponds to an *ACT11*.

4.5 Conclusions

Altogether, there were strong similarities between the different programs but the coincidence in assigning best and worst genes was not absolute. The fact that each program identified slightly different genes as best suited

for normalization prompted us to merge the data in an unsupervised way and giving identical weight to the output of the different programs. We used the RankAggreg program for this purpose. Our results show that *GAPDH* was the worst gene to use in normalization in both lines. In contrast, the suggested genes did not coincide and were *EF1α* and *SAND* in Mitchell, whilst *CYP* and *RAN1* were the genes of choice in V30. In conclusion, we provide a list of genes in discrete developmental stages that show M values below 0.5 (table 4.2) (Vandesompele *et al.*, 2002). A normalization factor including two genes should be enough for reliable quantification. Nevertheless we propose a reference gene stability test when performing gene expression studies in Petunia.

4.6 Authors' contributions

IM, BH, JW and MEC designed the experiments. IM and SL performed the experiments. IM performed data analysis and table and figure drawing. MEC wrote the first draft, and IM, BH, SL, JW and MEC corrected and approved the manuscript. JW, BH and MEC wrote grant applications.

4.7 Ackwnoledgements

Work performed in the lab of MEC and JW was funded by BIOCARM (Project Bananasai) and MEC (Project AGL2007-61384). IM obtained a PhD fellowship from the Fundación Séneca. This work was performed in partial fulfilment of the PhD degree of IM in the framework of the MSc-PhD program with Quality mention from the Spanish Ministry of Education MCD-2005-00339. Work performed in the lab of BH was funded by the "Pact for Research and Innovation" of the Leibniz Society, Germany. Thanks to Ronald Koes and Francesca Quatroccio for providing seeds of line V30, and Tom Gerats for seeds of line Mitchell. Michiel Vandenbussche is acknowledged for primers of *GAPDH*. Thanks to Luciana Delgado-Benarroch, Juana María Gómez Ballester and María Manchado-Rojo for comments on the manuscript. Our special thanks to Judith Strommer for helping with the edition of the manuscript and advice.

Chapter 5

Identification of novel genes involved in Petunia flower development using transcript profiling and reverse genetics

Petal development provides insight on a complex process of physiological, ecological and economical importance. Molecular mechanisms underlying the process comprise ABC model members and, increasingly, novel gene products affecting a plethora of molecular functions. This work presents a novel approach based on the bioinformatics mining of *Arabidopsis thaliana* expression data that resulted in a shortlist of nine novel genes presumably affecting petal development followed by a wet lab validation by RNAi on Petunia. The statistical contrasts of target genes selection were performed within the flower organs, thus focusing on the fine-tuned genes within whorls.

Among the regenerant lines phenotypes included: changes on flower shape and size (*PhCYP76*, *PhNPH3*, *PhFeSOD*, *PhXTH*, *PhCYP96* and *PhWAK*), petal smoothness (*PhPRA*), color (*PhNPH3* and *PhWAK*) and symmetry (*PhCYP76*); as well changes on leaf pigmentation and root morphology (*PhCYP76*), number of flowers and average plant height (*PhCYP76* and *PhCYP96*). After the batch phenotyping, we focused on the progeny of lines silencing cytochromes P450, either *PhCYP76* or *PhCYP96*, as they presented stable changes on petal size along other characters.

Therefore, a data flow aiding the discovery of novel genes affecting petal development is described. The post-transcriptional gene silencing confirmed the role in petal characteristics determination of seven from nine target genes, and included alterations on flower size, color, shape as well in vegetative parameters.

5.1 Introduction

Petals, as involved in plant reproduction, represent a plant compartment crucial in plant life. Petals are one of the keys of flower evolution, as they are the plant partner of the successful interaction with the pollen-carrying insects (Crepet, 2000). Novel petal characteristics are also the aim of breeding strategies, as new varieties usually contain changes on corolla size, shape and color. At the industry level, the approach to gather variability on petals is usually pursued through loss-of-function strategies, such as batch mutagenesis.

However, the classical knowledge on flower architecture is based on the extension of the ABC model described in *Arabidopsis thaliana*; consequently, literature mostly reports new petal characteristics according to changes in architectural genes expression. This fundamental research has been complemented with new genes that alter petal development despite its non implication in the ABC model. Expansins (Zenoni *et al.*, 2004; Dal Santo *et al.*, 2011; Zenoni *et al.*, 2011a), glutaredoxins (Li *et al.*, 2009), cytochromes P450 (Anastasiou *et al.*, 2007), bHlH transcription factors (Szécsi *et al.*, 2006) and E3 ubiquitin ligases (Disch *et al.*, 2006) among others. Some of them are the key to interlink or even trigger the signal transduction pathways associated to the modulation of phytohormones (Mallory *et al.*, 2005; Hu *et al.*, 2003; Martí *et al.*, 2007).

In Petunia (*Petunia hybrida*), the molecular mechanisms underlying floral organ specification have been throughfully studied. The ABC model applies to this species with modifications. B-class genes, which are present in petals and stamens, are represented in solanaceous species, such as Petunia, by two major groups: the *DEF/AP3* and the *GLO/PI* lineages. The *DEF/AP3* comprises two paralogous groups in Petunia (Kramer and Irish, 2000) with separated functions (Rijpkema *et al.*, 2006b): first, the eu*AP3* member Petunia *DEFICIENS* (*PhDEF*), which is present in eudicot species and whose mutants *phdef* convert petals to sepals but do not affect stamens; and second, the Petunia *TM6* (*PhTM6*), which is a paleo*AP3* that completes the sepal conversion to carpels as double mutant *phdef phtm6*. It is worth noting that the *PhTM6* is retained in solanaceous species but not in *A. thaliana* nor Snapdragon. The *GLO/PI* lineage is represented by two members whose function is redundant but specialized in a spatial fashion. The double mutants *phglo1 phglo2* show full conversion of petal to sepals, whereas *phglo1* mutants are less conspicous and show green petal midveins and unfused stamen filaments (Vandenbussche

et al., 2004).

The experimental design behind the biotechnology based on a single gene manipulation could be divided into three steps (Small, 2007): first, the identification of the target gene, step that could be accomplished through bioinformatics; second, the wet lab validation after modification of its expression level; and finally, the inclusion of the biotechnological improvement into the genetical background (that is, the production of modified germoplasm). The major challenges of the gene-expression-disrupted population phenotype scoring are: first, the genome structure of the species in term of genetic redundancy (in example, the mutant phenotype can require co-silencing of a putative paralog to appear); second, the chance of plant fitness jeopardy if the transcript is essential (which can even lead to non-recovery of null-alleles); and third, the availability of appropriate methods to recognize the mutant phenotype.

In this study we selected petal-specific target genes using a dataflow starting with the *A. thaliana* microarray data publicly available. We obtained differentially expressed genes in petals compared to other floral organs, retrieved the homologous ESTs from a *P. hybrida* database and selected specific subsequences or GSTs to clone in a hairpin-generating based expression construct (Helliwell and Waterhouse, 2003). We have obtained independent T_0 and T_1 stable transgenic lines that display petal phenotypes, most of them not described in Arabidopsis, indicating that Arabidopsis gene expression data can be used as a hypothesis to identify genes involved in processes using transgenic approaches.

5.2 Material and methods

5.2.1 Data retrieval and modeling

Transcriptomic data from *A. thaliana* was obtained from AtGenExpress ExpressionSet number 1006710873. This accession corresponds to a data set referring to different organs (petals, sepals, stamens, carpels and pedicels) and distinct developmental stages (12 and 15, as described by Smyth *et al.* (1990)) in wild type Columbia (Col-0) whose RNA was hybridized to an Affymetrix 15k platform. This type of slides accommodates 22814 *in situ* synthesized probes placed in a glass surface, and includes internal controls. The hybridization design comprised three biological replicates for

each treatment [16].

Robust Multichip Averaging (RMA) (Irizarry *et al.*, 2003) as a normalization and summarization method was used. A contrast matrix was built-up allowing the estimation of five parameters of a linear model containing the five triplicates of each organ. This matrix was further used to fit the linear model. The design matrix employed compared petals against the rest of floral whorls or pedicels. Microarray analysis was conducted with the packages *affy*, *limma* and *QualityMethods* from Bioconductor version 2.3 and R version 2.7.1.

The expression profiling focusing on whole plant anatomy and development were conducted with the Genevestigator v3 tool (Hruz *et al.*, 2008).

5.2.2 Functional annotation

The Affymetrix probe identifiers were mapped to the TAIR (Rhee *et al.*, 2003) AGI accessions and thereafter assigned to Gene Ontology (GO) terms using DAVID (Da Wei Huang and Lempicki, 2008; Huang *et al.*, 2009). GO abundance was compared with the *A. thaliana* background and represented with WEGO (Ye *et al.*, 2006). GO abundance was summarized using CateGOrizer (Zhi-Liang *et al.*, 2008) with a plant GOslim allowing all the possible paths between the ancestral terms and child terms.

[16] The authors of the dataset (Detlef Weigel, Jan Lohmann, and Markus Schmid) describe it as follows:

> The activity of genes and their encoded products can be regulated in several ways, but transcription is the primary level, since all other modes of regulation (RNA splicing, RNA and protein stability, etc.) are dependent on a gene being transcribed in the first place. The importance of transcriptional regulation has been underscored by the recent flood of global expression analysis, which have confirmed that transcriptional co-regulation of genes that act together is the norm, not the exception. Moreover, many studies suggest that evolutionary change is driven in large part by modifications of transcriptional programs. An essential first step toward deciphering the transcriptional code is to determine the expression pattern of all genes. With this goal in mind, an international effort to develop a gene expression atlas of Arabidopsis has been underway since fall 2003. This project, dubbed AtGenExpress, is funded by the DFG, and will provide the Arabidopsis community with access to a large set of Affymetrix microarray data. As part of this collaboration, we have generated expression data from 80 biologicaly different samples in triplicate. The focus of this data set is on different tissues and different developmental stages in wild type Columbia (Col-0) and various mutants.

5.2.3 Phylogenetics

Sequences for phylogenetic analysis were retrieved by tblastx comparison against the NCBI's nonredundant, TAIR's transcripts and SGN's Solanaceous unigene databases. These polipeptides plus the starting *P. hybrida* were aligned by MUSCLE (Edgar, 2004) and futher processed with Gblocks (Castresana, 2000) to extract the blocks of conserved, and phylogenetically informative, sites. Phylogenetic tree reconstruction was based on Poisson distances and clustered with the neighbor joining BioNJ algorithm (Gascuel, 1997). Node support was assessed by bootstraping 1000 times (Felsenstein, 1985). Alignment and tree building were conducted with Seaview v.4 (Gouy *et al.*, 2010).

The cytochrome P450 family analysis was conducted by protein sequence comparison to the P450 Engineering Database (Fischer *et al.*, 2006, 2007; Sirim *et al.*, 2009; Demet *et al.*, 2010) followed by the phylogenetic reconstruction method exposed before. In order to dissect all the cytochrome motives required for the superfamily and family assignment, full length coding sequences were retrieved by comparison to both the *P. inflata* and *P. integrifolia* draft genomic assemblies and predicted for gene structures with Genescan (Burge *et al.*, 1997).

5.2.4 Cloning and transformation

In order to build a library of sequences to be inserted in the hairpin-producer construct, we designed primers (appendix A.3) to amplify specific products of approximately 100 pb using the application PRIMEGENS_v2 (Srivastava *et al.*, 2008). This tool ensures unique hybridization within the Petunia database through two steps: first, the design of a primer pair fulfilling the lack of cross-hybridization or internal secondary structures, as implemented in Primer3 (Rozen and Skaletsky, 2000); and second, the check of non-specific hybridization conducted *via* gapless alignment.

Target amplicons were amplified from cDNA and recombined by Gateway technology into the pDonr221 vector and subsequently subcloned into the pHellsgate12 silencing vector (Helliwell and Waterhouse, 2003). Entry clones were checked by PCR and restriction analysis, and expression vectors were checked by PCR as described by Hilson *et al.* (2004). Expression clones were introduced in the *Agrobacterium tumefaciens* LBA4404 strain.

Leaf disc transformation was performed *via* explant coculture with the *A. tumefaciens*-plasmid-carrying strain as described by Horsch *et al.*

(1985) on wild type *P. hybrida* Mitchell diploid (Mitchell *et al.*, 1980). The detailed protocol is provided as appendix A.1.3.

5.2.5 Growth conditions

Plant material was grown in a substrate mix of perlite-coconut fibre-floral substrate in a ratio of 1:1:1 and watered as required. After *ex vitro* adaptation, T_0 plants were transferred to greenhouse conditions and crossed (selfed or back crossed to the Mitchell background) using the transgenic closed flowers as pollen acceptors. The subsequent T_1 generations were grown in Sanyo MRL350 growth chambers with a light regime of 16 hours light and 8 hours darkness and a thermoperiod of 24 °C and 18 °C, respectively, and a photosynthetically active photon flux density of 250 $mEs^{-1}m^{-2}$.

For *in vitro* germination and root morphology determinations, seeds from wildtype Mitchell diploid and from segregating populations of T_1 were sown on square plastic Petri dishes of 12 cm width with Murashige and Skoog media without sugars. After image acquisition, plantlets were transferred from growth chambers to greenhouse conditions.

5.2.6 Genotyping and gene expression analysis

Segregant T_1 populations were checked for the T-DNA insertion by PCR with attB1 and attB2 primers and Promega GoTaq Green Master Mix (Promega, Madison, Wisconsin, USA) according to the following protocol: 1 min at 95 °C; 30 cycles of 10 s at 95 °C and 30 s at 72 °C.

Q-PCRs were carried out with the SYBR Premix Ex Taq (TaKaRa Biotechnology, Dalian, Jiangsu, China) according to the manufacturer's protocol in either a Rotor-Gene Q (Corbett Research, Sydney, Australia) or a Mx3000P (Stratagene, Amsterdam) thermocycler. Cycle threshold (C_T) and efficiencies were extracted using the qpcR (Ritz and Spiess, 2008) implementation of the method published by Guescini *et al.* (2008).

Differential expression was tested using a relative quantification approach of each of the target genes against two internal controls whose stability was measured as described by Mallona *et al.* (2010). The statistical test applied relies on permutation tests (which are assumption-free on data distributions): the test is based on the reallocation of expression values regardless whether they belong to the treated group or to the control, and the computation of the probability of the random allocations. The

statistical cut-off was set as $p = 0.05$ in a two-sided test. The analysis were carried out with REST (Relative Expression Software Tool) v2.0.7 (Pfaffl *et al.*, 2002) and the R package qpcR (Ritz and Spiess, 2008).

5.2.7 Microscopy

Fully opened flowers were cut with a razor blade into pieces of 0.25 cm^2 and fixed overnight in a phosphate-buffered 2% glutaraldehyde solution. Samples were dehydrated through graded ethanol series, immersed in absolute acetone, dried with carbon dioxide in a critical point dryer (Bal-Tec CPD 030), placed on a double adhesive carbon conductive tape of 20 mm width (Ted Pella, California) and covered with gold using a metal sputter coater (Polaron SC 7610) with the following conditions: 90 s and 7 Pa of vacuum for 20 mA current. The scanning electron microscopy (SEM) imaging was conducted on a Hitachi SEM S-3500N at an acceleration voltage of 5 kV, 115 μA of emmision current and 4500 μm of working distance.

The dried-gel method presented by Horiguchi *et al.* (2006) was applied to the inner part of petal limbs. Fully opened flowers were cut into pieces of 1 cm^2 and placed over a melted 2% agarose drop supplemented with bromophenol blue; after gelification, the petal was peeled off and the gel cast observed with a bright field microscope.

5.2.8 Image processing and acquisition

Image acquisition was conducted with a Nikon D3000 15/55 mm, a HP ScanJet 4200C or a Leica MZFLIII coupled to a Leica DC300F device.

Built-in ImageJ (Abràmoff *et al.*, 2004) functions were used to automatize the segmentation procedures and cell area measurements using a macro. The acquired files were transformed into 16 bits images and thresholded with the Otsu method (Otsu *et al.*, 1979).

Flowers, as acquired with a computer-driven office scanner, "in vitro" root assays and scanning microscopy cell image analysis were performed detecting particle boundaries and compiling their area, perimeter and circularity. A bounding rectangle (the smallest rectangle enclosing a particle) was used for width and its orthogonal height measurement.

5.3 Results

5.3.1 Transcript profiling

To approach the differential expression of petal transcripts in *A. thaliana*, gene expression data was gathered from the four floral whorls plus the pedicels on developmental stages 12 and 15 (Smyth *et al.*, 1990). The data is publicly available as AtGenExpress ExpressionSet number 1006710873, and corresponds to wild type Columbia (Col-0) with the hybridization platform Affymetrix 15k. Each treatment contained three replicas and each slide accommodates 22814 *in situ* synthesized probes. After the RMA normalization procedure, all the slides passed the quality check comprising MA-plots, homogeneity on signal intensity, correct replicate clustering of raw data and RNA degradation. The differential expression analysis carried out compared the chips hybridized with petal samples against the rest of treatments, and produced a list of genes sorted by the likeliness of actual differential expression. Model fitting is presented as a volcano plot (figure 5.1A).

As the comparisons to detect the petal differential expression were performed against the rest of the flower organs, a meta-analysis profiling the target expression on the whole plant was conducted. Excepting AT3G15030 (which maps to the Petunia *PhTCP*), all the genes chosen for

---

Figure 5.1 *(following page)*: Transcript profiling in *A. thaliana*. **A** represents the volcano plot after fitting the linear model that results in the target list selection. Each dot represents a probe whose coordinates are given by its significance (log odds) and its log fold change (that is, the degree of down- and upregulation). The point labels highlight those with homologous in Petunia. Accessions with silencing germoplasm availability are colored in black; those who lead to non-recovery either in red (that is, their silencing lead to lethality) or blue (successful transformation but decreased fertility, and therefore lack of germoplasm); and those already described in *P. hybrida* are printed in green.
B and **C** depict the expression profile in different plant organs and tissues (**B**) or developmental stages (**C**) of *A. thaliana* as offered by the Genevestigator tool. Color intensity indicates the expression level. The number of samples used to build the heatmap is depicted in either row or column.

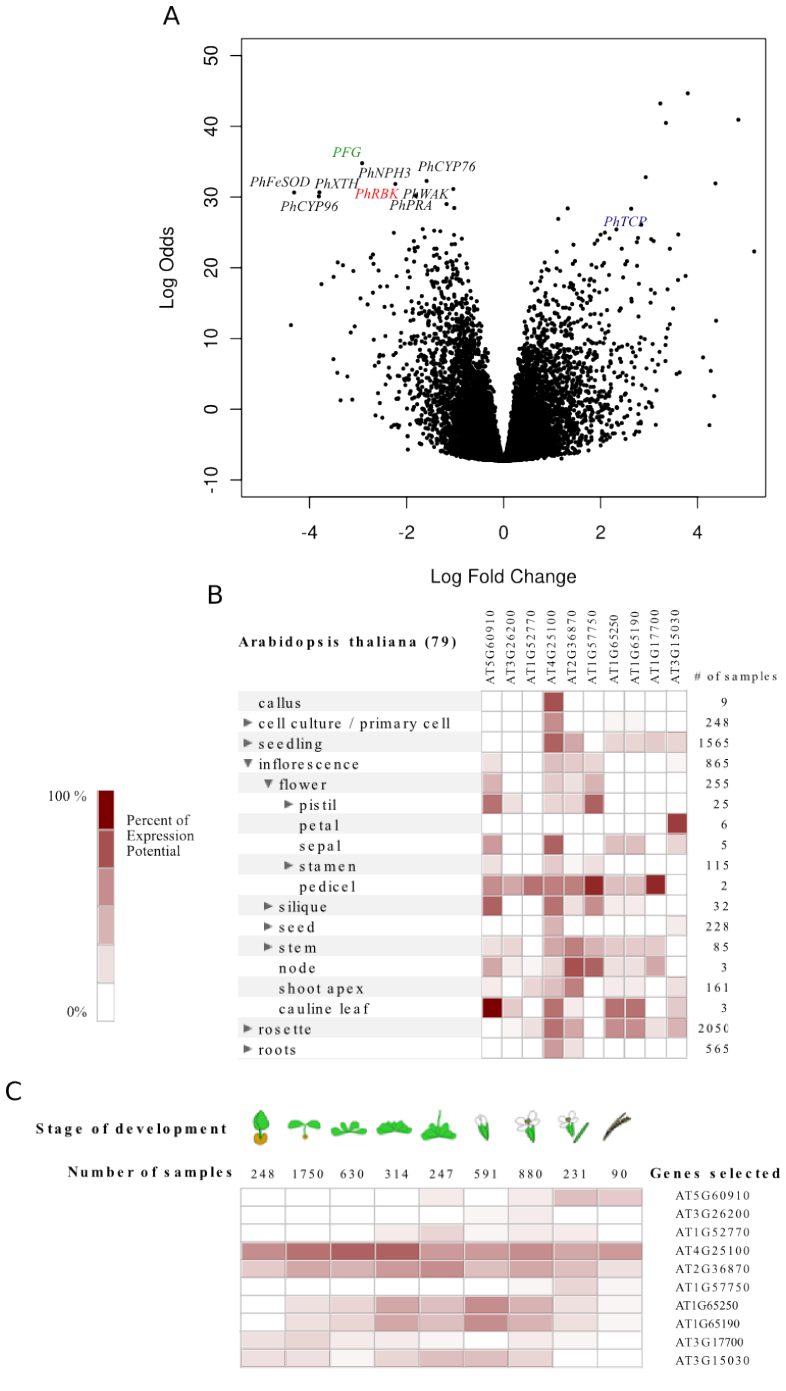

A
Log Odds
Log Fold Change
PFG
PhNPH3
PhCYP76
PhFeSOD
PhXTH
PhRBK
PhWAK
PhCYP96
PhPRA
PhTCP
B
Arabidopsis thaliana (79)
AT5G60910
AT3G26200
AT1G52770
AT4G25100
AT2G36870
AT1G57750
AT1G65250
AT1G65190
AT1G17700
AT3G15030
# of samples
callus 9
cell culture / primary cell 248
seedling 1565
inflorescence 865
flower 255
pistil 25
petal 6
sepal 5
stamen 115
pedicel 2
silique 32
seed 228
stem 85
node 3
shoot apex 161
cauline leaf 3
rosette 2050
roots 565
100 %
0%
Percent of Expression Potential
C
Stage of development
Number of samples
248 1750 630 314 247 591 880 231 90
Genes selected
AT5G60910
AT3G26200
AT1G52770
AT4G25100
AT2G36870
AT1G57750
AT1G65250
AT1G65190
AT3G17700
AT3G15030

Table 5.1: Differential expression analysis and *A. thaliana*-*P. hybrida* mapping. The 10 top-scoring *A. thaliana* identificators (represented as $AGIs$) were sorted by its Bayesian estimator of differential expression ($B-value$) and mapped by tblastx to the closest homologs to a local *P. hybrida* EST database (blast significance showed as $e-value$).

Petunia	AGI	e-value	B-value
PFG	AT5G60910	2E-84	34.78
PhCYP76	AT3G26200	2E-47	32.27
PhNPH3	AT1G52770	2E-44	31.14
PhFeSOD	AT4G25100	2E-49	30.66
PhXTH	AT2G36870	2E-43	30.17
PhCYP96	AT1G57750	2E-61	30.10
PhWAK	AT1G65250	3E-15	29.00
PhRBK	AT1G65190	1E-13	29.00
PhPRA	AT1G17700	2E-27	28.45
PhTCP	AT3G15030	4E-25	26.09

silencing were under-regulated on petals (figure 5.1A and 5.1B). According to the expression patterns in *A. thaliana* tissues and organs, as depicted in figure 5.1B, AT4G25100 is expressed in almost all the tissues except petals; AT5G60910, AT3G26200, AT2G36870, AT1G57750, AT1G65250, AT1G65190 have a wide expression range being present in five or more tissues or organs; and AT1G52770 and AT1G17700 have the narrowest expression ranges. During flower development (figure 5.1C) the expression pattern differ: AT3G26200, AT1G52770, AT1G17700 and AT3G15030 reduce their expression during or after fruit production, and AT5G60910, AT1G57750 and AT1G17700 have low expression in early bud stages.

The unmapped 100 top differentially expressed genes in petals were submitted to functional annotation and analyzed in two manners: first, they were summarized into categories allowing to describe the overall molecular functions, cell locations and biological processes; and second, they were compared to the *A. thaliana* EST background. The overview of Gene Ontology annotation is depicted as figure 5.2, whereas the significantly different categories are summarized as figure 5.3. GO categories revealed a preponderant binding and catalytic molecular functions, location within the cell and involvement on metabolic processes. According to

the comparison regarding molecular functions, antioxidant, catalytic, electron carriers, transcription regulators and transporters were overexpressed in petal transcripts. These results matched with the biological processes upregulated, such as regulation in a wide sense, metabolism, developmental processes, growth, reproduction and reproductive processes among others.

The list of petal transcripts was mapped through tblastx against a Petunia EST database (NCBI accession UNILIB:50472). Matching Petunia sequences with a hit e-score below 10^{-10} were visually inspected and eventually included into the final target list represented as table 5.1. Apart from the accessions of the TAIR and NCBI's databases, each target was named according to resemblance to similar sequences through phylogenetic analysis.

5.3.2 Phylogenetics and functional annotation

We made a phylogenetic analysis coupled to RNAi-induced gene silencing of the aforementioned genes in order to identify possible orthologs and, in case they would be available, to have a ground for phenotypic comparison in plants coming from T_0 transformation.

The diferentially expressed in petals *A. thaliana* accessions were mapped to *P. hybrida* in order to conduct a hairpin-based gene silencing with *P. hybrida* genes in the Mitchell diploid background; therefore, the T_0 individuals were expected to present altered petals. To avoid cross-silencing, each of the target genes was aligned to the *P. hybrida* EST database in order to extract a unique part of roughly 100 bp (see material and methods). After that step, the unique regions were independently cloned to the hairpin-based pHellsgate12 silencing vector and stably introduced into a wildtype Petunia background. As a proof of concept, batch phenotyping was performed on T_0 regenerants, taking as reliable those phenotypes shared among independent transformation events for a given gene.

Target *PFG* was the only gene in the shortlist which was previously described in *P. hybrida* (Immink *et al.*, 1999), and therefore was discarded. From a total of nine genes used in RNAi-induced loss of function, we obtained seven with floral phenotypes differing from wild type; one was not recovered from *in vitro* culture.

As the differential expression analysis prior to the mapping produced a shortlist of *A. thaliana* accessions, we checked the available databases for previously described *A. thaliana* mutants. The Agrikola database of gene-by-gene silencing (Hilson *et al.*, 2004) did not return phenotypes

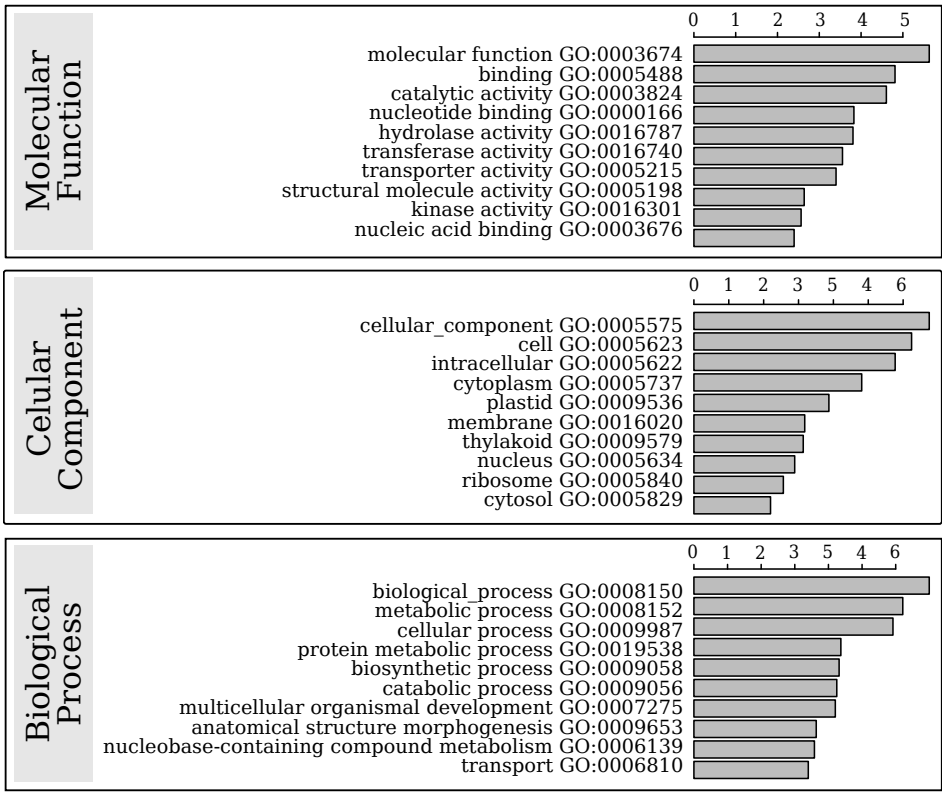


Figure 5.2: Overall Gene Ontology analysis. The ten most common Gene Ontology (GO) terms found in the top 100 differentially expressed probes. The plant GOslim was applied and all the possible paths between the ancestral term and the child term were counted. GO term abundance is represented logarithmically.

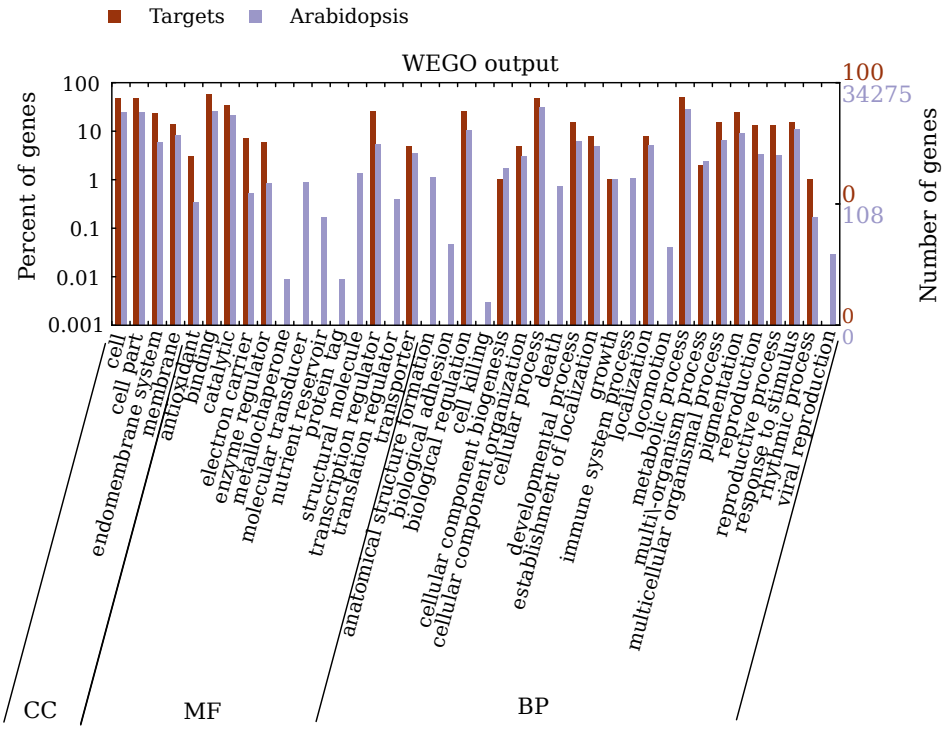


Figure 5.3: WEGO-based functional classification of the top 100 differentially expressed probes. The differentially expressed petal transcripts are depicted in red and the *A. thaliana* overall background in gray. GO term-related abundance comparison with *A. thaliana* was performed via the Pearson Chi-Square test. The presented functional categories present significant differences with the *A. thaliana* background ($p<0.05$). *CC*, cell component; *MF*, molecular function; *BP*, biological process.

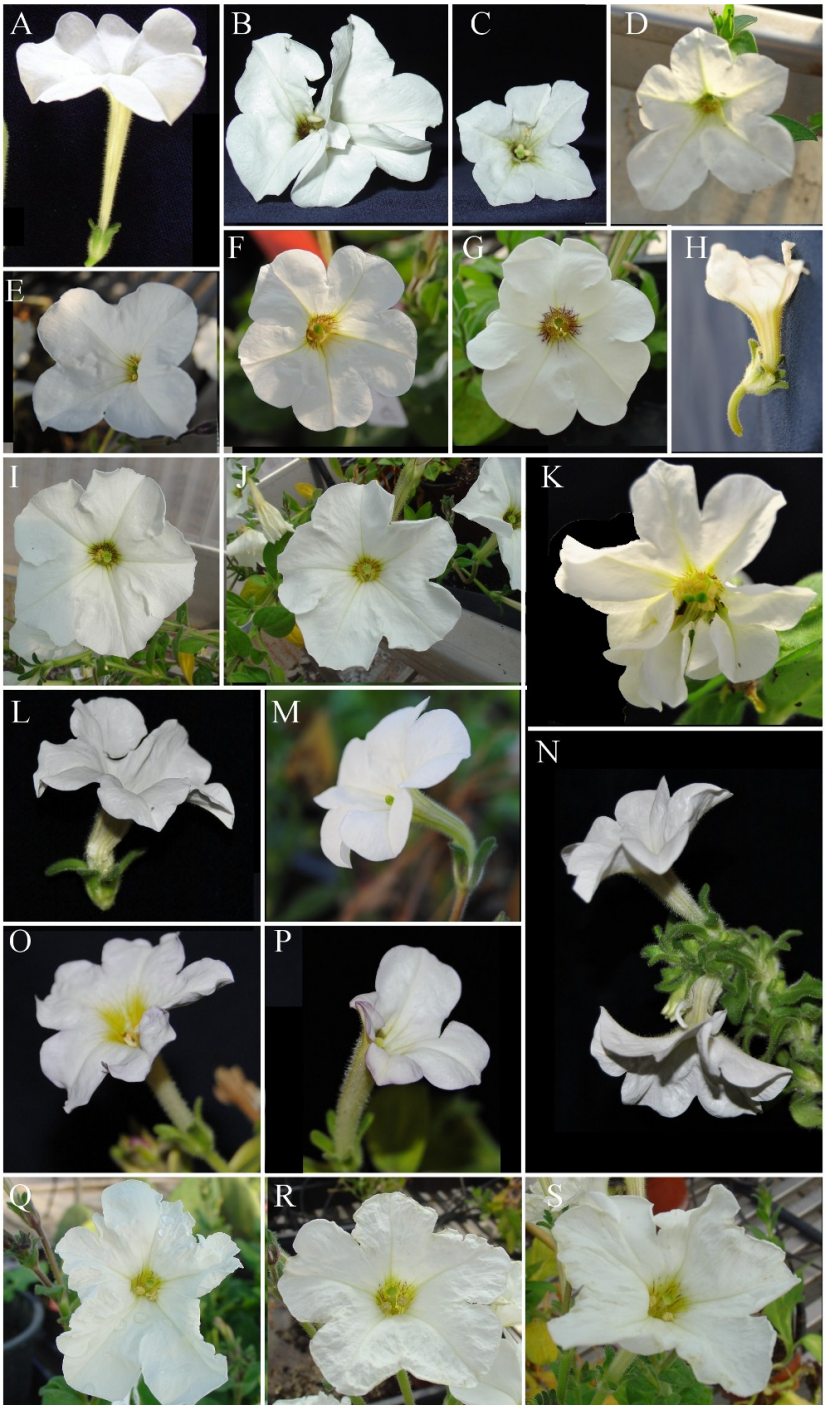
for any of the targets. The compendium of mutants published Lloyd and Meinke (2012) recovered information for three sequences whose mutants have a loss-of-function phenotype: AT5G60910 produce short siliques that are crowded with small seeds and reduced fertility (Martienssen, 1998); AT4G25100 produce pale green seedlings (Myouga *et al.*, 2008); and AT1G57750 produce low secondary alcohol and ketone levels in stem wax (Greer *et al.*, 2007).

PFG

The transcriptomics analysis ranked the *A. thaliana* petal transcripts according to its differential expression. The ranking leader, AT5G60910, is

---

Figure 5.4 *(following page)*: Diversity on petal phenotypes recovered in T_0 and T_1 individuals. Independent lines have different names (inside quotation marks).
A, Mitchell diploid wildtype.
B to **E**, *PhCYP76*-silencing lines: **B** ("2.E") and **C** ("2.4") reflect tearing of the dorsal petal due to filament fusion (the tearing ends at the sesile anther); **D** and **E**, "2.E" show alterations on symmetry, with a strong alteration on lobe proportions (**D**) and a complete lack of a lobe leading to a four-petaled flower (**E**).
F and **G**, *PhNPH3*-silencing lines: **F** represent a small flower ("3.j4"); **G**, "3.j5" shows a purple ring between petal tube and limb.
H, *PhFeSOD*-silencing line: a side view of "4.2" indicate stunted flowers with aberrant petals.
I and **J**, *PhXTH*-silencing lines: **I**, "5.j2" and **J**, "5.j9", show wider corolla limbs with distal petal fusion anomalies.
K to **N**, *PhCYP96*-silencing lines: **K** ("6.h1") reflect corolla fusion of two close flowers, resulting in a ten-petaled flower (modified image that darkens the background on the left); **L** ("6.h1") and **M** ("6.3") show anomalies on the flower tube, which bends in angle, and a narrower limb; **N** reflects shorter internodes and a pair of unfused external petals (lower flower, "6.h1").
O and **P**, *PhWAK*-silencing lines: "7.h1" shows small flowers with a yellow ring in the inner part of the limb (**O**) and slight pink color in petals (**P**).
Q to **S**, *PhPRA*-silencing lines: independent lines (**Q**, "9.j1"; **R**, "9.j2"; **S**, "9.j14") show big flowers with wrinkly corolla.


A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S

annotated as an agamous-like MADS-box protein (AGL8, FRUITFULL). FRUITFULL (FUL), next to SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1), LEAFY (LFY), AGAMOUS-LIKE 42 (AGL42), and APETALA1 (AP1), is involved in the vegetative to reproductive phase transition; this transition is regulated by the SBP-Box transcription factor SPL3 (Yamaguchi *et al.*, 2009). According to the interactome, FUL is present in flower induction and flower organ networks and interact with SOC1, SEP1 and AGL6 and AGL8, thus suggesting interaction with floral organ identity proteins (De Folter *et al.*, 2005). The interaction FUL-SEP has been further confirmed by Smaczniak *et al.* (2012). The expression of AT5G60910 as represented by Genevestigator, although, show very low expression on petals but an increase during flower development (figure 5.1B and C).

However, the mapping of the *A. thaliana* accession to the Petunia EST database identified *PFG* as described by Immink *et al.* (1999), so this gene was excluded from the shortlist to silence. According to Immink *et al.* (1999), *PFG* has a role on transition from vegetative to reproductive development. Its silencing leads to nonflowering phenotypes and down-regulation of B-class genes. Interestingly, its expression decreases under phytoplasma infection conditions, as well as that of *PhGLO1* and *FBP7* (Himeno *et al.*, 2011).

PhCYP76

The *A. thaliana* AT3G26200 mapped to the cytochrome P450 *PhCYP76*, which is assigned to the family CYP76A as depicted in figure 5.5. The cytochrome family CYP76A was described in *S. melongena* by Toguri *et al.* (1993). In *P. hybrida*, its member CYP76A4 has a role in lauric acid hydroxylation, although the physiological function of this activity has not been clarified yet (Tamaki *et al.*, 2005). In *Vanda coerulea* petals, the expression of *CYP76AB1* is correlated with the appearance of blue color in petals just before pollination (Ratanasut *et al.*, 2011), therefore indicating that this gene may play a role in petals, although the biochemical function of the protein remains unknown. The function of the *P. hybrida* CYP76B9 involves metabolism of medium-chain fatty acids such as capric and lauric acids (Imaishi and Petkova-Andonova, 2007). The oxydized forms ω-hydroxy capric acid and hydroxy lauric acid inhibit root development in *A. thaliana* (Imaishi and Petkova-Andonova, 2007); and the aliphatic oleic acid inhibits root growth in *H. vulgare* as well (Lee, 1977).

According to the phylogenetic analysis showed in figure A.2, *PhCYP76* shares sequence similarity with grape, eggplant and *P. trichocarpa* proteins and do not branch with the CYP76C cluster of *A. thaliana* nor the Solanaceous branch of CYP75B-like. *PhCYP76*-silencing lines showed smaller flowers with altered symmetry, as depicted in figure 5.4B, C, D and E.

PhNPH3

The *A. thaliana* phototropic-responsive NPH3-like protein (AT1G52770) mapped to the *P. hybrida PhNPH3*. The *A. thaliana* genome encodes 31 NPH3-like proteins, most of which have not been well characterized (Kimura and Kagawa, 2006). The best characterized gene in the family is *MACCHI-BOU 4/ENHANCER OF PINOID* (*MAB4/ENP*) involved in lateral relocalization of PIN protein (Furutani *et al.*, 2007, 2011). A phylogenetic analysis of *PhNPH3* showed closest homology to a tomato *NPH3* gene and two genes from soybean (figure A.3). As *PhNPH3* does not cluster close to *MAB4/ENP*, we conclude that *PhNPH3* is probably a paralog to *MAB4/ENP*.

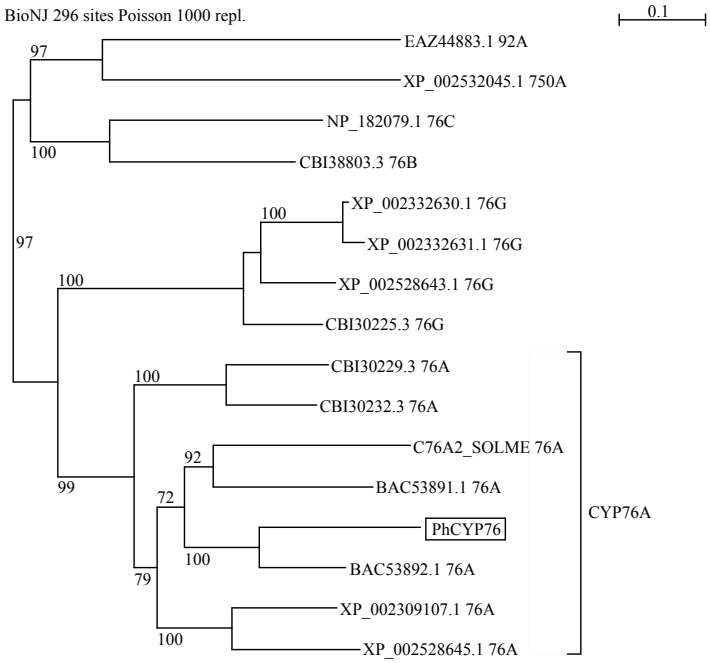
PhNPH3-silencing phenotypes included smaller flowers and a purple colored ring between tube and limb (figure 5.4F and G). This results indicate a possible role of *PhNPH3* in petal growth and pigmentation.

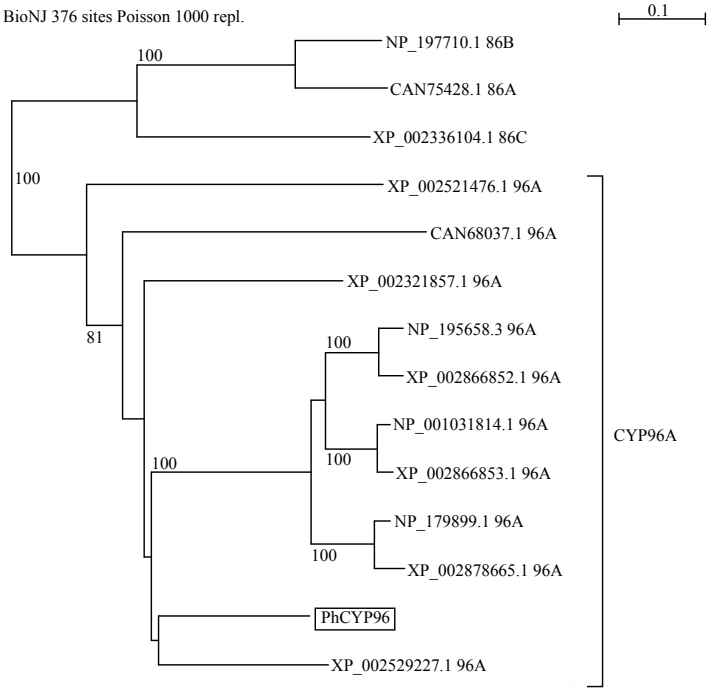
PhFeSOD

The *A. thaliana* AT4G25100 mapped to the *P. hybrida PhFeSOD*. According to its protein sequence, *PhFeSOD* is a superoxide dismutase. The phylogenetic analysis shown as figure A.4 indicates that the sequences sharing polipeptide similarity are divided into Fe/Mn superoxide dismutases, iron superoxide dismutases (FeSOD) from Solanaceae or FeSOD from

---

Figure 5.5 *(following page)*: Homologous family analysis of *PhCYP76* and *PhCYP96*. After protein to protein comparison against the CYP450 engineering database, a phylogenetic tree was reconstructed reflecting the sequence similarities to previously described proteins. A neighbor joining tree of the PhCYP76 (up) and PhCYP96 (down) is presented. Local boostrap probabilities above 70% label the nodes.


BioNJ 296 sites Poisson 1000 repl.
0.1
EAZ44883.1 92A
XP_002532045.1 750A
NP_182079.1 76C
CBI38803.3 76B
XP_002332630.1 76G
XP_002332631.1 76G
XP_002528643.1 76G
CBI30225.3 76G
CBI30229.3 76A
CBI30232.3 76A
C76A2_SOLME 76A
BAC53891.1 76A
PhCYP76
BAC53892.1 76A
XP_002309107.1 76A
XP_002528645.1 76A
CYP76A
97
100
97
100
100
100
92
99
72
100
79
100


BioNJ 376 sites Poisson 1000 repl.
0.1
NP_197710.1 86B
CAN75428.1 86A
XP_002336104.1 86C
XP_002521476.1 96A
CAN68037.1 96A
XP_002321857.1 96A
NP_195658.3 96A
XP_002866852.1 96A
NP_001031814.1 96A
XP_002866853.1 96A
NP_179899.1 96A
XP_002878665.1 96A
PhCYP96
XP_002529227.1 96A
CYP96A
100
100
81
100
100
100
100
100

A. thaliana, *Brassica oleracea*, *Lotus japonicus* and *R. communis*, among others. *PhFeSOD* clusters within the FeSOD Solanaceous branch, thus suggesting that the gene investigated could be specific for the Solanaceae.

A *Ds* insertion in AT4G25100.1 results in pale-green seedling (Kuromori *et al.*, 2006), but there is no detailed phenotypic description available. *PhFeSOD*-silencing lines had a strong floral phenotype including a stunted tube, shorter and underdeveloped limb and curly sepals (figure 5.4H).

PhXTH

PhXTH was retrieved as the AT2G36870 homolog. The phylogenetic analysis presented as figure A.5 reflects that PhXTH shares polipeptide sequence similarity to abundant endoxylogucan transferases from tomato. The tree did not offer a resolutive separation of XTH genes, although it placed with high bootstrapping confidence the PhXTH accession within a branch mainly composed by Solanaceous species but separated from the XTH9-like proteins from tomato and potato.

Expression of endo-xyloglucan transferases is strongly downregulated in *acaulis* mutants from *A. thaliana*, characterized by reduced cell length (Akamatsu *et al.*, 1999). Independent transformation events of *PhXTH*-silencing plants had wider corolla limbs with abnormal distal petal fusion figure 5.4I and J).

PhCYP96

The sixth more differentially expressed gene in petals according to our analysis, AT1G57750, mapped to *PhCYP96*. According to the cytochrome-centric phylogenetic analysis showed as figure 5.5, PhCYP96 is a CYP96 member. According to the batch phylogenetic analysis depicted as figure A.6, PhCYP96 clusters with *P. trichocarpa* and grape CYP96 proteins and not with the *A. thaliana* members nor the *A. thaliana* CYP86 called *lacerata*. The phenotype of the *PhCYP96*-silencing plants consisted in flowers with double corollas (5.4K), wide and bent tubes (figure 5.4L and M) and extra unfused petals (5.4N).

PhWAK

The wall-asociated protein kinase *PhWAK* of *P. hybrida* showed the closest homology to AT1G65250, which has unknown function in *A. thaliana* . The phylogenetic analysis depicted in figure A.7 defined a close homology

of PhWAK to a branch of predicted proteins from tobacco and WAK20 from grape, branch that is neatly separated but close to WAK4 and WAK5 from *A. thaliana*, among others. Silencing of *WAK4* in *A. thaliana* impairs cell elongation and affects plant development (Lally *et al.*, 2001).

Although most of the *PhWAK*-silencing lines were wildtype, a line had smaller flowers with defective petal growth, a slight purple color on the limb and a conspicuous yellow ring between tube and limb (figure 5.4O and P).

PhRBK

PhRBK is the protein kinase of *P. hybrida* mapping the *A. thaliana* AT1G65190 gene. It clusters within RBK-like proteins as well near to the *R. communis* BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 (figure A.8). We were not able to recover transgenic lines silencing *PhRBK*. Similarly, a lethal phenotype was described in *A. thaliana*, where loss of function alleles in this gene could not be recovered (Choe *et al.*, 2002).

PhPRA

The *A. thaliana* AT1G17700 gene maps to the *P. hybrida PhPRA*, which is electronically inferred to be a Prenylated Rab acceptor (PRA). According to the phylogenetic tree shown in figure A.9, *PhPRA* shows close homology to PRA from tomato and *Medicago truncatula*. PRA-1 proteins form a small gene family of 19 members in *A. thaliana* (Kamei *et al.*, 2008).

The *PhPRA*-silencing lines showed an extreme wrinkled petal phenotype (figure 5.4Q, R and S). This petal phenotype resembles other phenotypes described for TCP transcriptions factor such as *cincinnata*, involved in cell differentiation and growth (Nath *et al.*, 2003), suggesting a role for *PhPRA* in this complex pathway.

PhTCP

PhTCP is a homolog of the *A. thaliana* AT3G15030 gene. According to the phylogenetic analysis, *PhTCP* clusters with TCP-family members such as TCP24 of *A. thaliana* and tomato. The phenotypes of the *PhTCP*-silencing lines were wildtype, indicating that the silencing of *PhTCP* does not affect flower development in our growth conditions.

We focused on lines *PhCYP76* and *PhCYP96* as they had, as T_0, apparent modifications on petal identity. Segregant populations of T_1 plants were checked for further fine detail phenotyping.

5.3.3 Silencing of *PhCYP96* impairs B-class gene expression and produces pleiotropic effects

Four independent lines silencing *PhCYP96* showed smaller flowers. The T_1 populations used in this study were checked either for the *nptII* or attB cassette marker presence (progenies derive from self-crossing except when otherwise noted). According to the *nptII* amplification, line "6.4" had a single insertion (χ^2= 0.2252, df=1, p-value = 0.635>0.05; 10 absent and 29 present). Line "6.3" was checked using the attB1 and attB2 primers and did not have a pattern of single insertion inheritance (χ^2= 8.963, df=1, p-value = 0.003<0.05; 10 absent and 8 present). Two progenies of "6.5" were sown: a selfing and a backcross to Mitchell: the selfing did not show a pattern of single insertion inheritance (χ^2= 18.5128, df = 1, p-value =1.688e-05<0.05, 16 absent and 10 present); however, this disturbance was not found in the backcross-derived individuals (χ^2=2.3333, df=1, p-value=0.126> 0.05, 14 absent and 7 present).

Quantitative PCR analysis for transcript abundance showed that *PhCYP96* expression was significantly reduced to 7.40 % of the wildtype values ($p = 0.000 < 0.05$, figure 5.6Q).

As petal size was decreased in *PhCYP96*-silencing plants, a scanning electron microscopy analysis was conducted to survey cell behavior in sepals and the inner and outer part of both petal tube and limb (figure 5.7). Under moderated light deprivation conditions, the *PhCYP96*-line named "6.h1" showed a high rate of flower abortion that affected both the macroscopic and cellular phenotype. As depicted in figure 5.6, the petals of the aborted flowers were smaller, and its cells tinier and undifferentiated. Nearly isommetrical and still dividing petal cells of the aborted flowers were up to 20 times reduced at the outer part of the petal limbs as compared to the wildtype.

We used the progeny of the line with a stronger phenotype for further analysis (named "6.h1"). They showed small flowers with anomalies in the petal tube, including bigger diameter and bending (figure 5.6B, F and K), and extra petals (figure 5.6K). Strinkingly, under ligth deprivation conditions, the upper flowers of the inflorescence showed aborted petals (figure 5.6B, G) that featured high trichomata density in the midvein (figure 5.6C

and D). The inflorescence showed reduced internode length (figure 5.6G) that, in some cases, led to the fusion of two consecutive corollas into a 10-petaled flower (figure 5.6H and I) which, after dissection, showed in-

---

Figure 5.6 *(following page)*: *PhCYP96*-silencing lines phenotyping. All the photographs correspond to the "6.h1" progeny. The paraffin cylinder appearing in some images has a diameter of 0.5 mm.
A, wildtype Mitchell diploid (front view).
B, front view phenotypes of both the small (on the left) and aborted flowers (on the right).
C and **D**, electron scanning microscopy of the second whorl organ of the flower abortions, showing the high density of trichomata in the midvein.
E, wildtype Mitchell diploid (side view).
F, side view of the moderately small phenotype of *PhCYP96*-silencing flowers.
G, inflorescence architecture showing reduced internode length and the small and aborted flowers.
H to **J**, petal fusions leading to ten-petaled flowers; **J** was manually peeled off to reflect absence of gynoecia fusion and the curved styles.
K, inflorescence architecture showing reduced internode length and extra unfused petals.
L, Mitchell diploid wildtype capsule.
M, *PhCYP96*-silencing capsules reflecting abnormal placentation and valves.
N, Mitchell diploid wildtype seeds.
O, *PhCYP96*-silencing seeds with altered shapes and size.
P, quantitative data on flower roundness and area in a frontal view, plant height and number of flowers per plant. Data of $n = 94$ plants (flower roundness and area) or $n = 65$ plants (plant height and number of flowers by plant) from *PhCYP96* ("6.3", "6.4" and "6.5") progenies). Mann-Whitney one-sided test's p-values under the cut-off $p = 0.05$ are highlighted with an asterisk and denote significant differences.
Q, real-time PCR relative expression analysis of fully developed petals (discarding flower abortions) on $n = 5$ plants; a relative expression of one means no change to the wildtype. Significant differences are highlighted with an asterisk. *Actin* and *RPS13* were used as reference genes and showed stable expression.


A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
wildtype
PhCYP96
Round.
Area
Height
Num. flowers
Flower
Plant
Q
PhCYP96
PhDEF
PhGLO1
PhGLO2
PhTM6
Normalization:
ACT
RPS13
Gene expression to wildtype's

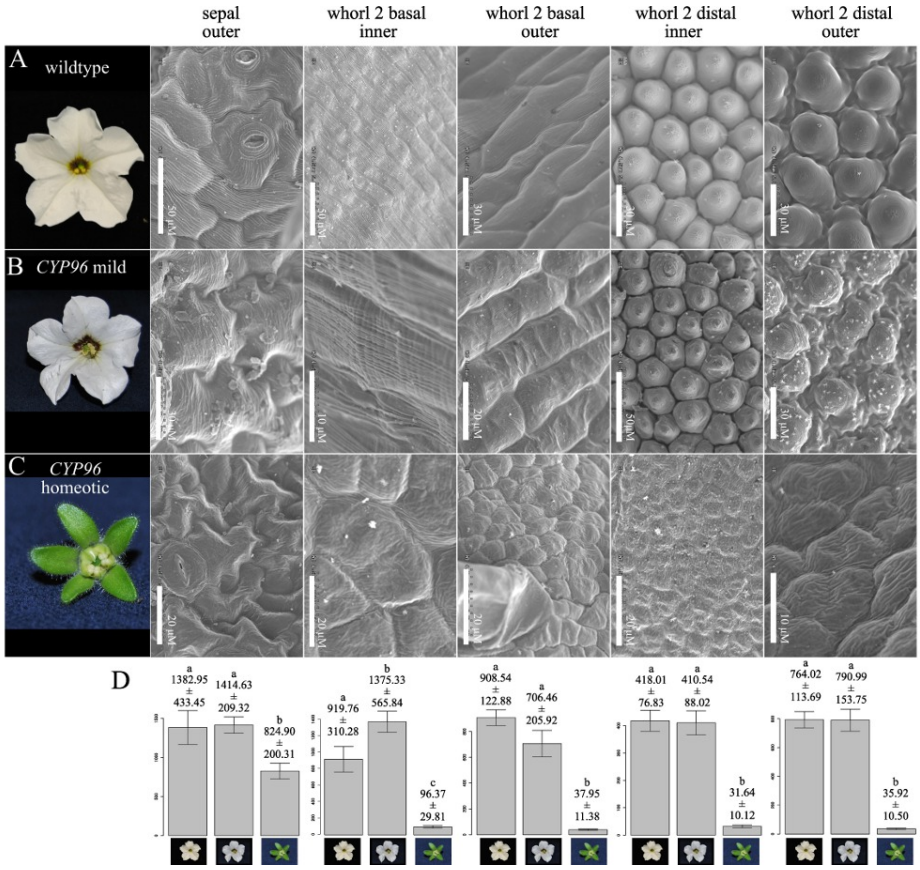


Figure 5.7:
Scanning electron microscopy of the epidermis of individuals of a T_1 segregant population of *PhCYP96*-silencing plants named "6.h1".
A, wildtype Mitchell diploid.
B, mild construct-carrying *PhCYP96* phenotype.
C, homeotic transform *PhCYP96* phenotype showing aborted second whorl organs.
D, data comparison on $n = 190$ cells obtained from a microscopy preparation of flowers from three "6.h1" construct-carrying plants and two azygous; detected differences (with significant Holm-corrected Pairwise Wilcoxon Rank Sum tests, $\alpha = 0.05$) are depicted with distinct letters. Areas are represented in μm^2.

dependent gynoecia (figure 5.6J). These abnormalities also were found in capsules and seeds. *P. hybrida* fruits have two valves, whereas *PhCYP96*-silencing lines sometimes produced aberrant fruits with higher number of valves and altered placentation (figure 5.6L and M). Compared to the wild-type, fruits and seeds appeared smaller. Seed form was affected, showing wedge or concave shapes as compared to the near perfect spheres typical of wild type (figure 5.6N and O).

As the aberrant petals suggested some decrease in petal identity, we analyzed further for changes in the expression of B-class gene expression on petals, discarding the flower abortions (figure 5.6Q). According to randomization tests, *PhGLO2* and *PhTM6* reduced its expression to 29.7% and 10.4%, respectively ($p = 0.033 < 0.05$ and $p = 0.009 < 0.05$). As an internal control, the expression of *PhCYP96* was found to be reduced to the 7.40 % of the wildtype's ($p = 0.000 < 0.05$). Neither *PhDEF* (relative expression of 45 %, $p = 0.066 > 0.05$) nor *PhGLO1* (72 %, $p = 0.316 > 0.05$) showed detectable downregulation. *Actin* and *RPS13* were used as internal controls and found to be stably expressed, as described by Mallona *et al.* (2010).

To assess the cellular phenotype of the smaller flowers, a scanning microscopy analysis was conducted (figure 5.7). Middle sized flowers with a mild floral phenotype had significantly larger cells in the inner part of the basal petal whereas the rest of the cellular sizes were similar to wild-type. In flowers with the strongest phenotypes cells in all the first and second whorl tissues analyzed were significantly smaller than in wildtype plants. However the cell types displayed in the apparently homeotically transformed organs do not coincide with wild type looking cells. Indeed we could not identify conical cells in the distal inner part of the second whorl as would be expected, and the nanoridges that characterize the petal inner cell surface were almost absent in the aborted flowers. These data indicate a strong change in the identity of the organs as would be expected also from the down regulation of the *PhGLO2* and *PhTM6* genes.

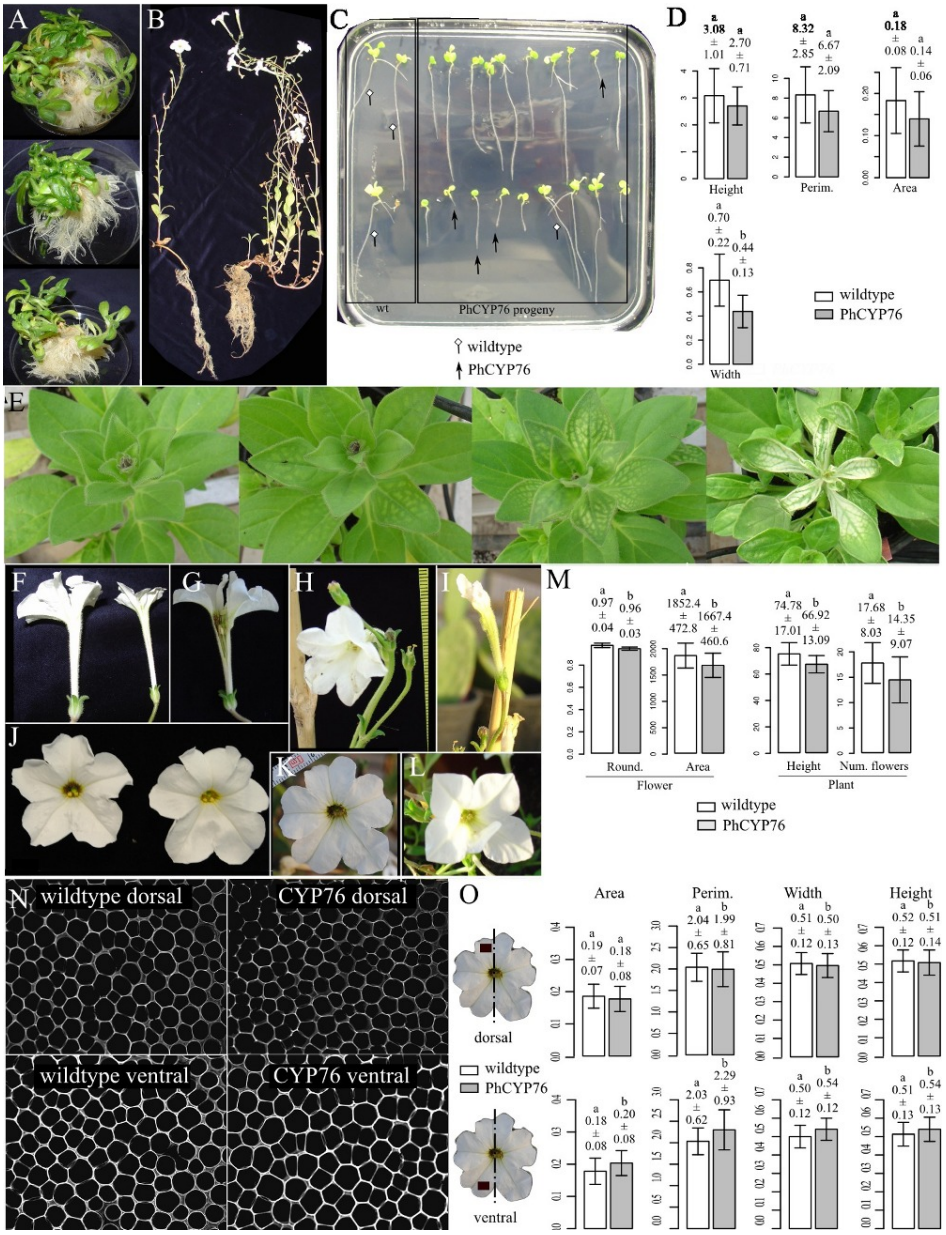
5.3.4 Silencing of *PhCYP76* produces pleiotropic effects

Three T_1 populations silencing *PhCYP76* were checked for the *nptII* marker presence (progenies derive from self-crossing except when otherwise noted). Line "2.1" had a single insertion ($\chi^2 = 0.1333$, df=1, p-value = 0.715>0.05; 11 absent and 29 present); line "2.4" did not have a single insertion ($\chi^2 = 7.5843$, df = 1, p-value = 0.006<0.05, 11 absent and 78 present); and "2.E"

had a single insertion (backcrossing to Mitchell; $\chi^2 = 0.4211 > 0.05$, df = 1, p-value = $0.516 > 0.05$, 17 absent and 21 present). Effective downregula-

---

Figure 5.8 *(following page)*: *PhCYP76*-silencing lines phenotyping.
A, independent regenerants before *ex vitro* adaption show increased root biomass.
B, line "2.A" (on the right) mantains as T_0 a bigger root system than the wildtype (on the left).
C, the germination assay on a T_1 segregant population show wildtype (diamond-shaped arrows) as well small root systems (black arrows).
D, data analysis on the root systems of the germination assay. On $n = 51$ plantlets, a bounding rectangle was fit to the root system and its height, perimeter, width and area were calculated. Detected statistical differences according to one-sided Wilcoxon tests are represented with distinct letters.
E, four siblings of a T_1 population with silenced *PhCYP76* (line "2.E") show a graded albino phenotype. On the left: a green, azygous wildtype plant; rest, construct-carrying plants with variable degree of chlorosis and dwarfism.
F, reduced petal limb expansion and tearing on "2.E" (on the left) and "2.a4" (on the right) flowers.
G, flowers of "2.a4" show complete filament fusion to the petal resulting on a sesile anther that tears the petal (the frontal cut was manually performed).
H to **I**, "2.E" show hight variability on flower shapes, from complete lack of petal development to petal limb dwarfism.
J to **L**, increased petal asymmetry in *PhCYP76*-silencing progenies. Wildtype (**J**, on the left) shows less bilaterality than "2.1" (**J**, on the right), "'2.4" (**K**) or "2.E" (**L**).
M, flower and plant macroscopic phenotyping. Flower roundness (a circle has roundness of 1) and area in a frontal view (mm^2) were measured from $n = 102$ plants of *PhCYP76* lines ("2.1", "2.4" and "2.E" progenies). Significant differences according to the Mann-Whitney one-sided test's are represented with different letters.
N, dried-gel casts of conical cells of the dorsal and ventral petal limbs.
O, statistical analysis on the dried-gel casts images. The symmetry axis and the sampling location are highlighted in the flower frontal view pictograms. $n = 2823$ cells of 15 plants, significant differences according to the one-sided Wilcoxon tests are represented with different letters.


A
B
C
wt
PhCYP76 progeny
wildtype
PhCYP76
D
Height
Perim.
Area
Width
wildtype
PhCYP76
E
F
G
H
I
J
K
L
M
Round.
Area
Flower
Height
Num. flowers
Plant
wildtype
PhCYP76
N
wildtype dorsal
CYP76 dorsal
wildtype ventral
CYP76 ventral
O
dorsal
ventral
wildtype
PhCYP76
Area
Perim.
Width
Height

tion of *PhCYP76* was checked in the progeny of the line with strongest phenotype, "2.E". According to randomization tests, the mean decrease factor of such a population was 64.7 % relative to the wildtype's expression (thus detecting efective *PhCYP76* downregulation, $p = 0.018 < 0.05$). The minimal expression factor detected in a sibling was 55.3%. *EF1* and *RPS13* were used as internal controls and found to be stably expressed, as described by Mallona *et al.* (2010)

As the T_0 individuals showed increased root biomass *in vitro* and after *ex vitro* adaption (Figure 5.8A, B, C and E), a root morphology assay was performed on T_1 individuals of lines "2.1", "2.4" and "2.E". The germination assay was conducted on *in vitro* conditions without supplementation of carbon sources. Root measures, such as length or width, were compared between the construct-carrying and the azygous. Plantlets of T_1 regenerants had significantly narrower root systems, assimilable to less secondary roots, as depicted in figure 5.8D and F, but no differences were detectable in area, perimeter nor height.

In one of the lines, "2.E", a degree of albino phenotypes and dwarfism could be observed among the T_1 siblings (figure 5.8E). In order to check whether the construct-carrying plants (homozigous or heterozygous) had the phenotype and the wildtypes (azygous) not, a contingecy table-based test was performed. As shown in table 5.2, let G_0 be the number of individuals without the construct and G_1 those harbouring it; and let P_0 be the number of individuals without phenotype and P_1 with. The binomial test design is as follows: the sample size is $n = P_0 + P_1 + G_0 + G_1 = 18$; the succesful event is the albino phenotype, hence we expect as many albinoes as construct-carrying plants, which are $e = (G_1 \cap P_0) \cup (G_1 \cap P_1) = 16$; and the observed successes are $o = (G_0 \cap P_1) \cup (G_1 \cap P_1) = 10$. The binomial test reflects that there is no detectable difference between the observed and the expected (10 successes, 18 trials, p-value $= 0.8145 > 0.05$), so the albino phenotype is associated to the silencing of *PhCYP76*. It is worth noting that none of the azygous plants showed changes in leaf color. Apart of the photobleaching, leaf shape of *PhCYP76*-silencing plants resembles the *lacerata* (*lcr*) mutants of *A. thaliana*, which include irregular leaf shapes tearing of leaf tissues, among postgenital organ fusion. The *lcr* mutants are defective in a cytochrome P450 involved in fatty acid hydroxylation (Wellesen *et al.*, 2001).

T_0 and T_1 plants silencing *PhCYP76* showed a distortion in petal lobe expansion (figure 5.8F–L). In some cases we found just by simple inspection, large differences between lower and upper petal lobe expansion thus

Table 5.2: Binomial test of the relationship between the albino phenotype and the genotype. A segregant population of T_1 plants carrying the *Ph-CYP76* silencing construct were phenotyped regarding to changes in leaf pigmentation.

		Albino phenotype	
		Absent (P_0)	Present (P_1)
Genotype	Azygous (G_0)	2	0
	Construct-carrying (G_1)	6	10

increasing the flower bilateral symmetry and leading in some cases to four-petaled flowers (figure 5.8J–L). A quantitative analysis performed on segregant T_1 populations confirmed the differences on flower roundness and total petal area (figure 5.8M). Plant height and floral number were significantly reduced.

Measuring conical cells in dorsal and ventral petal limbs (figure 5.8N) we detected a significant increase of cell area in ventral petal cells but not in the dorsal side (figure 5.8O). In fact all cellular parameters (perimeter, width and height) were significantly smaller in dorsal cells and bigger in ventral cells.

5.4 Discussion

The overall tendency in *A. thaliana* research during the decade of 1990-2010 is to substitute the forward, classical by reverse genetics approach. The work over this decade unveiled the 90% of the Arabidopsis genes whose alteration triggers a mutant phenotype (Lloyd and Meinke, 2012). However, the reverse genetics gene function addressing method relies on two aspects: first, the ability to knockout a gene; and second, the success on plant recovery and detection of corresponding phenotypes by a properly designed screening procedure (phenotyping). Identification of candidate genes for systematic gain or loss of function experiments have been performed and all have in common a previous selection of the genes based on differential expression (Nasir *et al.*, 2005), known gene function (Spitzer *et al.*, 2007b), or gene families (Meissner *et al.*, 1999). Additional steps towards identification of the gene function requires a genetic screen to identify those that modify a trait under study.

One of the apparently best studied processes in plant development is petal morphogenesis. In spite of the work performed in that has been done in identifying the organ identity genes, few genes have been described that affect petal growth. In our work we identified genes differentially expressed in Arabidopsis petal tissue against the rest of the flower and identified their corresponding orthologs in Petunia. We assumed that although there are important differences between petunia and Arabidopsis in terms of petal development, several conserved mechanisms should underlie the commonalities found in both tissues.

Meinke *et al.* (2003) proposed that at least 10 % of the protein coding sequences in *A. thaliana* would trigger a detectable mutant phenotype when silenced. After a comprehensive literature curation in *A. thaliana*, Lloyd and Meinke (2012) published a list of nearly 3,000 genes with loss-of-function mutant phenotypes, 400 of them requiring the co-silencing of a redundant paralog. The gene set represents roughly 10 % of the protein-coding sequences in the genome. Our results, based on the incorporation of a dataflow of gene selection in a hypothesis-free manner, show that one out of nine target genes resulted on lethality, thus being a phenotype of essentiality; and further seven genes (*PhCYP76*, *PhNPH3*, *PhFeSOD*, *PhXTH*, *PhCYP96*, *PhWAK* and *PhPRA*) showed floral phenotypes when silenced. This 89 % of phenotype-retrieving success differs substantially from the 10 % expectancy described in *A. thaliana*. Therefore, the identification of genes in *A. thaliana* by criteria of expression profiles and their testing in a different species can be indeed successful.

Silencing of *PhCYP96* produced a stable phenotype of smaller corollas with unfused petals. The expression analysis of non-aborted flowers revealed that some B-class genes are downregulated in transgenic plants, thus suggesting some interaction between them (figure 5.6Q). In *P. hybrida*, *PhGLO1* and *PhGLO2* act redundantly specifying petal identity, along with the A-class gene *blind* which appears to act as a repressor of the C-class gene expression in the first and second whorls. The family *DEF/AP3*, whose members are *PhDEF* and *PhTM6*, however, appear to act quite differently. *PhDEF* is required for petal as well stamen identity, but *PhTM6* only appears to be involved in stamen identity specification. The fact that in *PhCYP96*-silencing lines, all B-function genes showed some degree of downregulation albeit only *PhTM6* and *PhGLO2* were significant indicates a possible effect of *PhCYP96* in organ identity in a way that requires further work.

Seed and fruit morphology was heavily disturbed in *PhCYP96*-silencing

plants, including seed size, morphology and fruit placentation (figure 5.6M-O). The sepal retention in fully maturate capsules as well the poor seed quality is consistent with previous reports on mutants such as *lacs1 lacs2* (Weng *et al.*, 2010). The *lacs1 lacs2* has be found to be involved in cuticle and wax biosynthesis, and produces abnormal flower development as well organ fusion.

As the functional annotation of AT1G57750 suggests involvement on cutitular wax biosynthesis and the silencing of its homolog *PhCYP96* in *P. hybrida* resembles the phenotypes of the *lacs1 lacs2* mutant and altogether produces petal cells with reduced nanoridges, *PhCYP96* may be involved in cutin and cuticular wax biosynthesis. Nelson (2006) suggests that CYP86 members are an ancient group of cytochromes as they are involved in protection against water loss but; as CYP96 is a member of the CYP86 clan and is absent in moss and ferns but present in angiosperms, specially in rice, *A. thaliana* and poplar, is proposed that CYP96 is the most recent radiation of the clan (Nelson and Werck-Reichhart, 2011).

PhCYP76-silencing plants had peculiar flower and root morphology, plant height, number of flowers and leaf color. The disruption of several components of the biochemical machinery has been found to be related with albino phenotypes. An *A. thaliana* flavonoid-related mutant has been described to show a graded series of phenotypes with chlorotic leaves in heterozygous plants, albinoes homozygous and green wild-types (Sheahan *et al.*, 1998). Apart of the photobleaching, leaf shape of *PhCYP76*-silencing plants resembles the *lacerata* (*lcr*) mutants of Arabidopsis, which include irregular leaf shapes tearing of leaf tissues, among postgenital organ fusion. The *lcr* mutants are defective in a cytochrome P450 involved in fatty acid hydroxylation (Wellesen *et al.*, 2001).

Among root morphology modifications in *PhCYP76*-silencing plants, significant changes on the root system width were found. As depicted in figure 5.8D, the number of secondary roots was greatly reduced but the overall root area remained unchanged. This result is opposed to the finding of T_0 regenerants with bigger root systems (Figure 5.8 A, B, C and E). We suggest that this discrepancy could be produced by the differences on sugar supplementation between the rooting medium with high levels of carbon source and the germination assay media devoid of exogenous carbon.

PhCYP76-silencing lines showed a difference on flower symmetry as estimated simple inspection of the flowers. A statistical analysis showed significant differences towards a zygomorhpic flower as the degree of roundness decreased, and cell measurements in dorsal and ventral petals showed

opposite tendencies. Our results indicate a possible effect of *PhCYP76* in the maintenance of floral symmetry in *P. hybrida*.

5.5 Conclusions

A hypothesis-free system for systematic discovery of new genes altering petal development has been proposed and applied to garden Petunia. The post-transcriptional gene silencing confirmed the role in petal characteristics determination of seven from nine target genes, and included alterations on flower size, color, shape as well in vegetative parameters.

5.6 Author contributions

IM, ME and JW carried out the design of the study. IM performed the experimental procedures and data analysis. IM wrote the manuscript. All authors read and approved the final manuscript.

5.7 Acknowledgements

We thank María José Roca (SAIT) for her excellent help on electron microscopy. Fundación Séneca granted IM and funded the project 11895/PI/09.

Chapter 6

Conclusions

This chapter summarizes the concluding remarks, the practical applications and the future prospects of the present work.

6.1 General conclusions

As the dissertation is structured in independent parts, the discussed results of each chapter lead to partial, self-contained conclusions. However, as stated at the introductory section 1.6.2, the dissertation nucleus is chapter 5, which regards the selection and silencing of the target genes presumably affecting petal development; and the chapters 2, 3 and 4 are focused on the methods which support the main flow.

6.1.1 Chapter 2

This chapter presents a lightweight, local pipeline for EST analysis. The software presented here comprises the two major steps in EST analysis: the consensus sequence build procedure, the annotation and its storage in a relational database; and the collection of scripts used for data querying, visualization and interpretation. pESTle focuses on data security and integrity as all the procedure of assembly, clustering and annotation is produced locally. The pESTle architecture is arranged in four tiers: the central tier, Tier 0, is the database managed by PostgreSQL; pESTle surrounds either filtering the accesses to it and serving the data to a web interface (Tier 1), or building the database from raw data (Tier -1); finally, the final user client is located in Tier 2 and the raw data and reference databases are located in Tier -2.

6.1.2 Chapter 3

This chapter shows both an univariate analysis and a generalized additive model which links the PCR efficiency reaction with some parameters of the reaction, such as the G+C content of the template or the primers length.

The univariate analysis reflects that several variables contribute significantly to the PCR efficiency (in example, the amplicon length reduces it). The model regarding multiple interacting variables show that the 42% of the deviance found could be explained in a model which gathers modifiable parameters, such as lengths and purine contents. Accordingly, there is possible to improve the expected efficiency prior to wet lab experiments.

6.1.3 Chapter 4

In this chapter the suitable reference genes for developmental studies in *Petunia hybrida* flowers and leaves is performed.

GAPDH is not adequate for normalization purposes in *P. hybrida*. *EF1α* and *SAND* are valid in Mitchell, whilst *CYP* and *RAN1* are the genes of choice in V30. A normalization factor including two genes should be enough for reliable quantification. Nevertheless we propose a reference gene stability test when performing gene expression studies in *P. hybrida*.

As a novelty, this work insist on the inconsistencies found between different stability-assessing algorithms and proposes for first time the rank aggregation-based melding of the data.

6.1.4 Chapter 5

In this chapter a systematic reverse genetics approach is applied to discover novel genes involved in petal development. The statistics behind target gene selection is based on data mining of public transcriptomic *Arabidopsis thaliana* data and an EST database from *P. hybrida*.

Results assign a function on petal development to seven out of nine target genes. The silencing lines show altered flower parameters such as: size and shape (*PhCYP76*, *PhNPH3*, *PhFeSOD*, *PhXTH*, *PhCYP96* and *PhWAK*), petal smoothness (*PhPRA*), color (*PhNPH3* and *PhWAK*) and symmetry (*PhCYP76*). Pleiotropic phenotypes were found, such as root morphology and leaf color (*PhCYP76*), flower number, capsule and seed morphology (*PhCYP96*) and plant height (*PhCYP76* and *PhCYP96*). Therefore, the identification of novel genes by criteria of expression profil-

ing can indeed be successful; in this case, a proof of concept was applied to *P. hybrida* petal development.

6.2 Practical applications

Three of the chapters presented in this dissertation have a strong methodological background. The first one, chapter 4, offers a new method for estimating the reference gene stability by merging several previously described algorithms. Chapter 3 and 2 are in fact software tools, which are direct, out-of-the-lab resources: an user-friendly web tool allowing PCR efficiency estimation; and a new conceptual model for EST handling, comprising a selection of sequence (pre-)processing procedures, a relational database design, a frontend for annotation tools, and ready-to-use web user interface.

It is worth noting that the chapter 4 had 20 cites on May 2012; three citing articles apply the rank aggregation proposed in *Coffea* (Goulao *et al.*, 2011), rat carotid (Kim *et al.*, 2011) and hybrid roses (Klie and Debener, 2011).

6.3 Summary of contributions

Since *P. hybrida* started to be regarded as a model system its research topics widened, covering very different fields such as flavonoid metabolism, epigenetics, volatiles and so on (Gerats and Vandenbussche, 2005). These studies were in many cases carried out looking at the gene expression levels *via* quantitative or semiquantitative PCR. However, the first validation of the suitability of the reference genes for these studies, included in the present thesis as chapter 4, was published in 2010. The findings of the chapter were corroborated by Zenoni *et al.* (2011b) even in different *Petunia* species through microarray analysis.

Quantitative real-time has been one of the most used technologies in gene expression assessment. However, it was described that a common pitfall on its data analysis is to obviate the PCR reaction efficiency (Freeman *et al.*, 1999). As the underlying basics determining such parameter were poorly known, the experimental set-up could not be modified in order to uniformize or maximize the reaction efficiency. The chapter 3 of the present thesis offers a novel application which predicts PCR efficiency given a sequence to amplify and a primer pair. The application is acces-

sible through a web interface and thus is platform independent; is freely available and could be employed either in predicting the efficiency of a given amplicon and primers or in primer design (when only the desired amplicon is queried).

Gene function elucidation by reverse genetics is a standard approach. Post-transcriptional gene silencing has been found to be a valuable information source even if only a partial loss-of-function is generated (Andrew, 1999). Since the *-omic* information available at the start of the present thesis were ESTs, we decided to conduct an inductive, synthetic hypothesis search based on microarray analysis of previously published data. As described in chapter 5, we gathered data from *A. thaliana* expression sets and localized transcripts expressed characteristically in petals, and then examined their homologs in Petunia. After such identification, we silenced those transcripts and analyzed the phenotypes.

Roughly, post-transcriptional gene silencing can be summarized as the introduction of exogenous RNA within the cell and consequently inhibit gene expression whose sequence is equivalent to that of the RNA. DNA amplification by PCR, technologies based on recombinant DNA and plant transformation are determined by, again, DNA sequence. Thus, sequence handling is a key point on genetics research. Because of that we start the present dissertation by chapter 2 offering an EST handling platform (*P. hybrida* sequences are mostly ESTs, until the genome project finishes); this EST pipeline offers an integrative solution to EST manipulation and provides the basis to the rest of the analysis.

The aim of the present dissertation was to discover novel genes playing a role on petals in *P. hybrida*. Chapter 5 show an experimental design for accomplishing that aim as well its results. As the methods are as different as transcript profiling, plant transformation or phenotyping, a variety of techniques have been required during the PhD thesis. A summary of such a methodology is presented as figure 6.1, reflecting the interactions and dependencies between chapters.

6.4 Future research

In 2012 the Petunia Platform (The Petunia Platform, 2004) organized a genome project. The availability of the assemblies will facilitate designs involving full-length coding sequences, such as overexpression analysis; and reporter assays, such as GUS-mediated promoter spacial regulation anal-

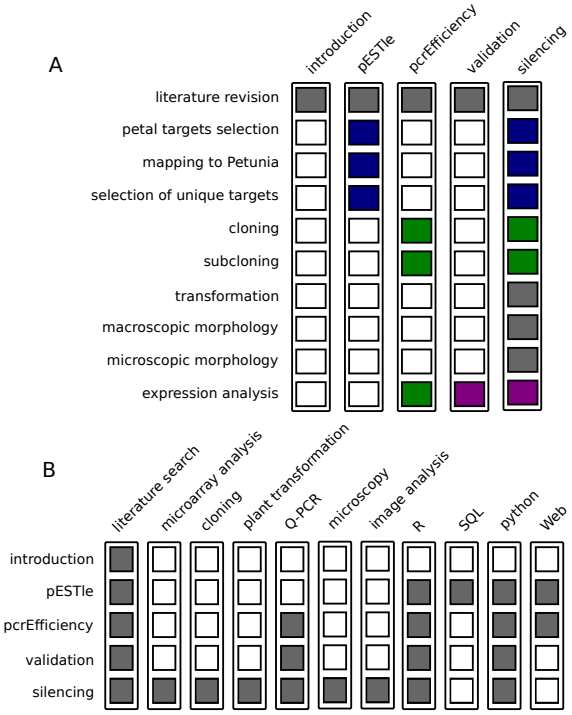


Figure 6.1: Overview of the methodology used and dependencies between chapters. Label code the chapters' number as follows: introduction, 1; pESTle, 2; pcrEfficiency, 3; validation, 4; and silencing, 5. **A**, contributions of each chapter (columns) to the dissertation. **B**, summary of the major tools employed in each chapter (rows).

ysis. Incidentally, the understanding of phenotypes associated to the silencing of *PhCYP76* and *PhCYP96* could be greatly improved by overexpression analysis; and more independent lines with impaired expression could be fetched from the Vandenbussche's and coworkers pool described in Vandenbussche *et al.* (2008). The availability of a new *dTph1*-tagged insertional mutagenesis population can enrich the number of independent lines of the genes tested in this dissertation. Even more, as the shortlist of target genes presumably involved in petal identity was longer than the ten genes scrutinized in this dissertation, new phenotypic information can be achieved from the rest of genes with unknown function.

As this dissertation presents new genes affecting the atypical class B gene *PhTM6* expression, crosses with the well-established pool of *P. hybrida* mutants will be insightful.

As the stable plant transformation is a labor-consuming protocol, transient expression silencing assays can provide an interesting alternative. Virus-induced gene silencing (VIGS) provides the tools to even silence multiple genes at once (Chen *et al.*, 2004).

Finally, microarray techniques could be applied to the some of the transgenics lines obtained. This approach enlarge the data analysis power, as changes in the transcriptomic level can be understood globally.

6.4.1 Final reflection

The large community of researchers studying model species such as *Caenorhabditis elegans*, *Drosophila melanogaster* or *A. thaliana*, among others, have unveiled abundant and striking mechanisms in developmental biology. However, some of this information in a certain way lacks universality.

The "-omics" integration could be insightful in handling this diversity, specially in terms of data acquisition from different experimental conditions as well from consolidated and emerging model species. Thus, given the facilities developed for "-omic" data handling, data mining could give some insight on both conserved and divergent patterns. This data analysis could have a direct application on fundamental biology, such as developmental regulation or evo-devo understanding; but also to industry, such as the plant breeding.

Capítulo 7

Conclusiones

El presente capítulo resume las conclusiones generales del trabajo, así como sus aplicaciones y perspectivas.

7.1 Conclusiones generales

Dado que la presente tesis está estructurada a manera de artículos científicos, cada capítulo posee su propia conclusión. No obstante y tal y como queda descrito en la introducción (sección 1.6.2), el núcleo de la tesis está representado en el trabajo de selección y silenciamiento de genes presumiblemente involucrados en la identidad de pétalo (capítulo 5); de este modo, los capítulos 2, 3 y 4 representan el bagaje metodológico que ha sido necesario desarrollar para la consecución de los objetivos del capítulo 5.

7.1.1 Capítulo 2

Contiene un nuevo sistema de gestión de ESTs denominado ≪pESTle≫. Consiste en un conjunto de aplicaciones informáticas que efectúan las dos tareas fundamentales de un entorno de almacenamiento de ESTs: en primer lugar, el procesamiento de los datos crudos con el fin de obtener secuencias consenso, anotadas y convenientemente dispuestas en un sistema gestor de bases de datos; y en segundo lugar, la difusión de estos resultados procesados, facilitando su visualización e interpretación. pESTle está diseñado con el fin de garantizar la seguridad de los datos; por ello, se ejecuta localmente y posee mecanismos de garantía de integridad. De este modo, pESTle organiza la base de datos biológica en cuatro capas: una central, la capa 0, que contiene los datos procesados convenientemente gestionados

por el núcleo de PostgreSQL; dos capas intermedias, una filtrando las consultas y modificaciones de los datos (capa 1) y otra realizando las tareas de construcción de la base de datos (capa -1); y finalmente las capas externas, que ofrecen la interfaz del usuario final (capa 2), y la que contiene los datos crudos y las bases de datos de referencia (capa -2).

7.1.2 Capítulo 3

Se describe un análisis univariante y un modelo aditivo generalizado que relacionan la eficiencia de reacción de la PCR y ciertos parámetros, como el contenido G+C del producto de amplificación o la longitud de los cebadores.

El análisis univariante refleja que la eficiencia de PCR responde a variables de distinta índole (como la longitud del producto de amplificación). El modelo multivariante reduce el número de variables a tener en cuenta y explica el 42 % de la devianza. De esta forma, el modelo reúne datos simples, como por ejemplo la longitud de los cebadores y su contenido G+C, y predice la eficiencia final. Por tanto, la herramienta permite realizar predicciones sobre cebadores preexistentes pero también ofrece una interfaz de diseño de cebadores orientado hacia la obtención de eficiencias óptimas.

7.1.3 Capítulo 4

Este capítulo analiza la idoneidad de genes de referencia para PCR en tiempo real bajo las condiciones experimentales de la tesis.

GAPDH muestra una gran inestabilidad y queda descartado como gen de normalización en *P. hybrida*. *EF1α* y *SAND* presentan el comportamiento óptimo en la línea Mitchell, mientras que *CYP* y *RAN1* lo hacen en V30. No obstante, salvo para el caso de *GAPDH*, se detecta una normalización adecuada empleando dos genes prácticamente en cualquier condición experimental. De cualquier modo, se postula la necesidad de efectuar análisis de estabilidad de expresión de los genes de referencia bajo cada condición experimental.

El trabajo resalta que las estimaciones de estabilidad producidas mediante distintos métodos son inconsistentes. Como novedad propone realizar una aglutinación de los resultados que se basa en una agregación de *rankings*.

7.1.4 Capítulo 5

Este capítulo emplea la minería de datos transcriptómicos como método para descubrir nuevos genes con presumible participación en la determinación de características de pétalo. Para ello parte de datos públicos de la especie modelo *A. thaliana* y de una base de datos local de EST de *P. hybrida*, obteniendo como resultado líneas transgénicas con una menor expresión de cada uno de los genes candidatos.

Como resultado, se describen siete nuevos genes cuyo silenciamiento altera de hecho características florales tan variadas como: la forma y tamaño (*PhCYP76*, *PhNPH3*, *PhFeSOD*, *PhXTH*, *PhCYP96* y *PhWAK*), textura (*PhPRA*), color (*PhNPH3* y *PhWAK*) y simetría (*PhCYP76*) de los pétalos. Además, se detectan cambios en otros parámetros, como el color de las hojas y la morfología radicular (*PhCYP76*), el número de flores, la morfología de las cápsulas y semillas (*PhCYP96*) y la altura de la planta (*PhCYP76* y *PhCYP96*). Por tanto, un enfoque puramente bioinformático permite recabar nuevos genes cuya función se circunscriba a un determinado ámbito; en este caso, el desarrollo del pétalo de *P. hybrida*.

7.2 Aplicaciones

Tres de los capítulos aquí presentados poseen una vocación metodológica y, por tanto, práctica. Uno de ellos, el capítulo 4, ofrece un nuevo método de estimación de idoneidad de genes de referencia. Los capítulos 3 y 2 son de hecho aplicaciones informáticas utilizables directamente en sí mismas: concretamente, el primero proporciona una interfaz web que apoya el diseño de cebadores en conjunción con un estimador de la PCR; y el segundo ofrece una nueva estrategia de análisis y almacenamiento de ESTs que redunda en la difusión de los resultados mediante una interfaz web de uso sencillo, a la vez que posee un diseño sólido basado en un gestor de bases de datos relacional.

En cuanto al impacto real del presente trabajo cabe destacar que el capítulo 4 poseía 20 citas en mayo de 2012; tres de los artículos incorporan la metodología propuesta en material tan variopinto como el cafeto (Goulao *et al.*, 2011), la carótida de rata (Kim *et al.*, 2011) o la rosa híbrida (Klie and Debener, 2011).

7.3 Contribución científica

Los estudios empleando *P. hybrida* como organismo modelo son variados e incluyen el metabolismo de flavonoides, la epigenética, la emisión de volátiles entre otros (Gerats and Vandenbussche, 2005). El análisis de la expresión génica realizado en buena parte de ellos se ha basado en la técnica de PCR cuantitativa (o semicuantitativa) mediante comparación con genes de referencia. No obstante, la primera validación de los genes de referencia, incluida como capítulo 4 en la presente tesis, data de 2010. Sus resultados han sido confirmados mediante un estudio de micromatrices de ADN realizado por Zenoni *et al.* (2011b); cabe destacar la consistencia de resultados pese a que las condiciones experimentales entre ambos estudios difieren, así como la categoría taxonómica de las petunias estudiadas.

La PCR cuantitativa en tiempo real es una de las técnicas más empleadas para evaluar la expresión génica. Existe literatura científica resaltando que una fuente importante de errores procede de obviar la eficiencia de la reacción (Freeman *et al.*, 1999). No obstante, se desconoce buena parte del fundamento bioquímico que afecta a tal eficiencia; más aún, no existe una receta clara para maximizarla o uniformizarla. El capítulo 3 de la presente tesis ofrece una aplicación que predice la eficiencia de acuerdo a características de los cebadores y de la secuencia a amplificar. Tal aplicación es libre y accesible a través de una interfaz web, por lo que es independiente de plataforma. Ofrece la la predicción de eficiencias de PCR a partir de la combinación de un juego de cebadores y un producto de amplificación, así como un diseño de cebadores dirigido que evalúa la eficiencia predicha.

La asignación de funciones a un determinado gen mediante técnicas de genética inversa se ha convertido en un estándar. El silenciamiento génico permite inhibir la expresión de forma informativa aun cuando el silenciamiento no es total (Andrew, 1999). Puesto que la información *ómica* presente al inicio de la tesis consistía en ESTs, el enfoque escogido fue sintético e inductivo: el diseño experimental se basa en una búsqueda bioinformática de datos depositados en repositorios públicos que genera la hipótesis a evaluar. Tal y como se describe en el capítulo 5, se obtuvieron datos transcriptómicos procedentes de los cuatro verticilos florales y los pedicelos de una base de datos de *A. thaliana*; después, se realizó un análisis estadístico que detectara aquellos genes con una pauta de expresión distinta en pétalos; finalmente, se obtuvieron los genes homólogos en *P. hybrida* y se silenciaron a fin de caracterizar anomalías en la corola.

A grandes rasgos, el silenciamiento génico post-transcripcional se basa

en introducir un ARN exógeno de doble cadena en la célula y disminuir con ello la expresión de genes que codifican la secuencia de este ARN. Las técnicas de amplificación de ADN mediante PCR, del ADN recombinante y por ende la transformación génica, nuevamente, dependen de la manipulación de secuencias. Por tanto, la gestión de las secuencias es un punto crucial en la investigación genética. Por esta razón el capítulo 2 abre el presente documento de tesis ofreciendo una solución integradora para la gestión de ESTs (las secuencias existentes de *P. hybrida* a falta de la publicación definitiva de su proyecto genómico) que sustenta el resto de análisis.

El objetivo de la presente tesis es la búsqueda de nuevos genes característicos de pétalo en *P. hybrida*. El capítulo 5 plantea y ejecuta un diseño experimental para este fin. Como los métodos empleados son diversos, incluyendo el análisis transcriptómico, la generación de plantas transgénicas o el fenotipado, se han empleado distintas técnicas a lo largo de la tesis. La figura 6.1 esquematiza el peso específico de cada capítulo en la consecución del objetivo final; del mismo modo, resume las técnicas empleadas y señala las dependencias entre unas partes y otras.

7.4 Perspectivas futuras

El proyecto de secuenciación del genoma de petunia difundió el primer ensamblado de *P. inflata* y *P. integrifolia* en 2012. Su disponibilidad favorecerá una aproximación de sobreexpresión o de expresión con promotores específicos, acomodando secuencias codificantes completas. La comprensión de los fenotipos *PhCYP76* y *PhCYP96* se verá facilitada mediante estos ensayos.

Al mismo tiempo, sería interesante realizar una búsqueda de germoplasma en el banco de mutantes de inserción por transposición descrito por Vandenbussche *et al.* (2008). El fenotipado podría extenderse a más líneas de las descritas en esta tesis: puesto que se infirió una expresión diferencial en el pétalo para más genes, podría extenderse el proyecto con una criba fenotípica de aquellos que estén presentes en las líneas de inserción y que, al mismo tiempo, tengan una función desconocida.

Puesto que en el capítulo 5 se muestra una interacción entre *PhCYP96* y *PhTM6*, los cruzamientos entre ambos genotipos podrían arrojar luz sobre el efecto de este gen de clase B atípico al que, de hecho, se le conoce una pauta de expresión de gen de tipo C.

Las técnicas de silenciamiento transitorio mediado por virus (VIGS) podrían ofrecer una alternativa rápida al silenciamiento estable. Además, ofrecen la posibilidad de silenciar múltiples genes simultáneamente (Chen *et al.*, 2004).

Finalmente, la caracterización de las líneas transgénicas podría realizarse mediante técnicas transcriptómicas, como las basadas en micromatrices de ADN. De este modo sería posible obtener una visión global de los todos cambios desencadenados durante el proceso de silenciamiento.

7.4.1 Reflexión final

La biología del desarrollo ha avanzado enormemente gracias al estudio de especies modelo tan dispares como *A. thaliana*, *Caenorhabditis elegans* o *Drosophila melanogaster*, entre otros. No obstante, algunos de los resultados obtenidos distan de ser universales.

Por esta razón, sería de interés ahondar en perspectivas integradoras. Esta estrategia de aglutinación podría basarse en recabar datos procedentes de distintas condiciones experimentales tanto de las especies modelo consolidadas como de las emergentes. De esta forma, y dada la facilidad de manejo de la información *-ómica*, podría realizarse una minería de datos exhaustiva que incidiese tanto en las pautas conservadas como en las divergentes. Este conocimiento sería extrapolable a cuestiones puramente fundamentales, como la regulación del desarrollo o el evo-devo; y a aproximaciones prácticas, como la mejora de caracteres de interés agronómico.

Appendix A

Supplemental information

A.1 Supplemental methods

A.1.1 Fast plasmid DNA isolation from *Escherichia coli* minicultures

A protocol for fast colony check was adapted from the Tom Gerats' lab (Nijmegen). It is an one-day, high throughput method based on 96-well plates bacterial mini-cultures. A sample of 3 μL of the last step can be used as template on a 15 μL PCR.

1. Colony picking (8 colonies per construct).
2. Inoculation of the 100 μL LB (plus the antimicrobial agent, as described by Maniatis (1989)) on a microtiter plate.
3. 37 °C incubation for four hours.
4. Sampling of 5 μL of the grown culture and diluting to 100 μL with water.
5. 94 °C incubation for five minutes (on a thermocycler).
6. Centrifugation (one minute at 1000 rpm).

A.1.2 Fast genomic DNA isolation from Petunia leaves

A protocol for DNA extraction was adapted from the Hilson *et al.* (2004)'s protocol for Arabidopsis RNAi lines genotyping (AGRIKOLA consortium).

Table A.1: Fast plant DNA extraction buffer composition

Compound	Volume
Tris HCl pH 7.5	8 mL
NaCl 5 M	2 mL
SDS 20 %	2 mL
H_2O	to 40 mL

Using young Petunia plantlets as plant material, the distal half of a young leaf was sampled and processed with the following protocol (a sample of 3 μL of the last step can be used as template on a 20 μL PCR):

1. Grind liquid nitrogen-frozen samples stored on 1.5 mL eppendorf tubes by manual pistile mecanical pressure and vortexing with stainless-steel balls.
2. Add 400 μL of the extraction buffer whose composition is described in table A.1.
3. Incubate at 50 °C for 15 min.
4. 15 min centrifugation at 4200 rpm and 4 °C.
5. Sample 150 μL, which will be transferred to a new tube; discard of the starting tube with the pellet.
6. Add 150 μL of isopropanol.
7. Vortex and incubate for 20 min at room temperature.
8. 30 min centrifugation at 4200 rpm and 4 °C.
9. Remove supernatant.
10. Add 200 μL of ethanol 70 %.
11. 15 min centrifugation at 4200 rpm and 4 °C.
12. Remove supernatant.
13. Resuspend on 100 μL of TE pH 8 (Maniatis, 1989).

A.1.3 Petunia transformation

The transformation procedure is a slightly modified version of the protocol gently provided by the Tom Gerats' lab (Nijmegen). It gathers the experience of Peter de Groot (Nijmegen) and Dave Clark (University of Florida). If not stated otherwise, the "in vitro" culture was under a light regime of 16 h L and 8 h D and a termoperiod of 23 °C L and 18 °C D. The protocol is approximately three months long.

1. Seed sterilization of Mitchell diploid (W115).
   - Place 50 μL of Petunia seeds in a 1.5 mL tube filled with ethanol 70 %. Incubate one minute.
   - Rinse with sterile water three times.
   - Soak on domestos at 20 % for ten minutes.
   - Rinse with sterile water until no foam is produced.
   - Sow in halfstrenght MS medium and grow for 6 weeks.
2. Grow the *Agrobacterium tumefaciens* culture overnight in 4 mL YEB medium plus rifampicine and spectinomycine.
3. Dilute the culture 1:10 in water.
4. Cover five Petri dishes with excised, young Petunia leaves (lacking petiole; each leaf yields 2–4 fragments).
5. Soak each Petri dish with 10 mL of the diluted *A. tumefaciens* culture and allow to sit 15 minutes.
6. Remove the *A. tumefaciens* culture.
7. Cultivate two days in dark conditions.
8. Transfer to selective medium plates (8 explants per plate).
9. After 3 weeks the *calli* will regenerate shoots. The *calli* must be subcultivated every 4 weeks, and the shoots must be excised and transferred to hormone free medium.
10. After 2 weeks and conspicous rooting, remove the medium from the roots and transfer to a substrate of vermiculite-perlite-turf-coconut fiber mixture (2:1:2:2) and cover the pot with a plastic film.

11. Cultivate for a week.

12. The *ex vitro* adaption requires another week of gradual removing (that is, rupturing) the plastic film.

13. Genotype.

The standard plant medium (based on the basal recipe described by Murashige and Skoog (1962)) contains (per litre): 30 g of sacarose and 4.4 g of Murashige and Skoog microsalts with Gamborg B5 vitamins. The pH is adjusted to approximately 5.8 with NaOH. Phytagel is used as solidifying agent at 4 g/L. The different variants used are the following:

- Halfstrength medium. Standard plant medium but with 2.2 g/L of MS salts.

- Co-cultivation medium. Contains 2 mg/L of 6-benzylaminopurine and 0.1 mg/L of 1-naphthaleneacetic acid.

- Selective medium. Contains 2 mg/L of 6-benzylaminopurine, 0.1 mg/L of 1-naphthaleneacetic acid, 500 mg/L carbenicillin and 300 mg/L kanamycin.

- Root medium. Contains 500 mg/L carbenicillin and, if desired, 50 mg/L (the kanamycin difficults root development and could be avoided without increasing the number of escapes).

A.2 Germoplasm availability

The fertile seeds available as either T_1 and T_2 are summarized in table A.2.

A.3 Sequence information

The substrings subjected to reverse genetics (cloned and thereafter used to trigger the post-transcriptional gene silencing) are shaded in lightgray. Bootstrap values greater than 70% are printed reflecting node support of the neighbor-joining trees.

Table A.2: List of germoplasm carrying a silencing construct. Although 10 independent lines silencing *PhTCP* were achieved, their propagation was not possible due to their low fertility, being "10.a5" the only line available as T_1.

Sequence silenced	Independent lines	Count	Comment
PhCYP76	2.1, 2.4, 2.6, 2.A4, 2.E, 2.Ea, 2.j1	7	
PhNPH3	3.j10, 3.j13, 3.j14, 3.j15, 3.j2, 3.j4, 3.j6, 3.j7, 3.j8, 3.j9, 3.p3, 3.p4	12	
PhFeSOD	4.2, 4.a1, 4.h1	3	
PhXTH	5.j1, 5.j2, 5.j4, 5.j6, 5.j7, 5.j9, 5.p2	7	
PhCYP96	6.2, 6.3, 6.4, 6.5, 6.a7, 6.h1, 6.j5	7	
PhWAK	7.h1, 7.j10, 7.j11, 7.j3, 7.j5, 7.j6, 7.j7, 7.j8, 7.j9	9	
PhRBK	–	–	lethal
PhPRA	9.h1, 9.h3, 9.h4, 9.j12, 9.j13, 9.j14, 9.j15, 9.j2, 9.j28, 9.j6, 9.j7, 9.j8, 9.j9, 9.ñ1, 9.p1, 9.w1	16	
PhTCP	10.a5	1	fertility

A.3.1 *PFG*

The Clone identificator "dr004P0009L15_F.ab1 2007-08-10" has "64995225" as NCBI's dbEST accesion, "FN009980" as EST name and GenBank Accession and "227589170" as GenBank gi. Immink *et al.* (1999) characterized it as *PFG*. The primers used for its amplification are: dr004P0009L15_F.ab1_attB1 TACAAAAAAGCAGGCTTGGAATCTGGAACATGCAAA, dr004P0009L15_F.ab1_attB2 CAAGAAAGCTGGGTAAGGGAAGAATCCAGCTGTTG.

```
   1 GAGAGAATAT ACATTTACTC TCTCTCTATA GTATAGTTTA ATTTATTCTG CACTATACTT
  61 TTTTGTTAGA CAAAAATGGG AAGAGGAAGA GTGCAGATGA AGAGAATTGA GAATAAAATT
 121 AATAGACAAG TTACTTTTTC AAAACGTCGA TCTGGATTAT TGAAGAAAGC TCATGAAATC
 181 TCTGTGCTTT GTGATGCTGA AGTTGGTTTA ATTGTTTTTT CTACTAAAGG CAAACTCTTT
 241 GAGTATGCTA CTGATTCTTG CATGGAGAGG ATTCTTGAAA GATATGAAAG ATACTCATAT
 301 GCTGAGAGGC AGCTTGTTTC TACTGATCAT AGCTCCCCGG GAAGCTGGAA TCTGGAACAT
 361 GCAAAACTTA AGGCCAGAAT TGAGGTTGTG CAGAGAAACC AAAGGCATTA TATGGGAGAA
 421 GATTTGGACT CGTTAAGTAT GAAAGAACTT CAGAATTTAG AACAACAGCT GGATTCTTCC
 481 CTTAAACACA TTCGATCAAG AAAGAACCAA TTGATGCATG AGTCCATTTC TGAGCTTCAA
 541 AAAAAGGACA AATCATTGCA AGAGCAAAAC AACCTTCTT
```

A.3.2 *PhCYP76*

The clone identificator "drs12P0022M07_R.ab1 2007-08-10" has "65024650" as NCBI's dbEST accesion, "FN030308" as EST name and GenBank Accession and "227617963" as GenBank gi. According to the phylogenetic analysis described in section 5.2.3, a tree-based reconstruct of sequence similarity can be found as figure A.2. The primers used for its amplification are: drs12P0022M07_R.ab1_attB1 TACAAAAAAGCAGGCTTCCAATTGCCCACACATCTA, drs12P0022M07_R.ab1_attB2 CAAGAAAGCTGGGTAGTTTGCATCCACCACTTCCT.

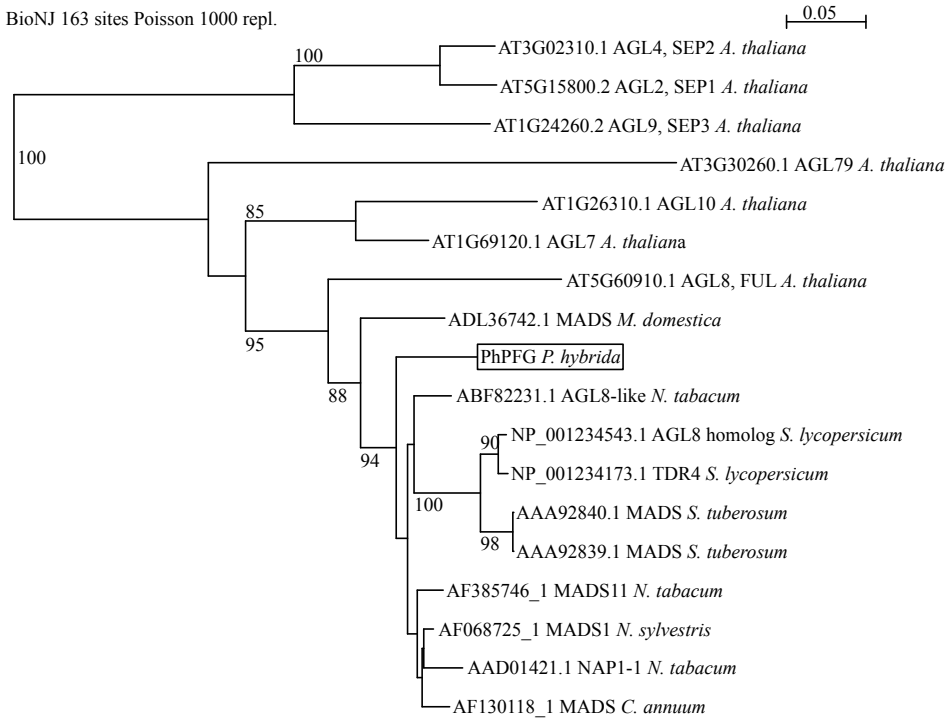


Figure A.1: Phylogenetic analysis of *PFG* (also known as dr004P0009L15_F.ab1).

```
  1 GCTTTCAAGG TTGGAGCTTT CTCACAGTCG CTCCCAACCT CTCCATCATG TCCATGGATT
 61 TAGGAGTGAC ATTATCAGGA AGCTCCCAGT TAAATGCATG AAGCAATGAG CCCAAAAGAA
121 GGTGCAACAT ACGATTACCT AAGGGAAGGC CTGCACACAT TCTTCTGCCA GCACCAAATG
181 GTATGAACTC AAAGTTTTGC CCCTTGTAAT CAGTTTTTGA ATTGAGGAAC CTCTCTGGCT
241 TGAATGCCAA AGGGTCTTCC CAGCATTCAG GGTCTCTTCC AATTGCCCAC ACATCTACAA
301 GAACTTGAGT GTCTTCTGGT ATATCATATC CCATGAAACT GGTGTCTTGA ATTGCTCTTC
361 TTGGAATTAA TAAAGGAAGT GGTGGATGCA AACGCAGCGT TTCCTTGACA ACAGCTTGCA
421 TATAATGTAA GTTATCAATG TCACTCTCCT CAAACTTCCT GTTTGGTCCT ACTACCTCGA
481 TAATCTCTGC TTTCACTCTG GTCATTGCTT CAGGATGGCG CAGGAGCTCA GTTAGTGCCC
541 ATTCGACACT GCTACTAGTT GTCTCTGAAC CACCTAGAAA CATTTCCAGT ATGAATATGT
601 TAATCTGATG TTCTGATAAA CTAGCTGGTT CATCTTTTCC AGTGCCCTCA AATTCAAGCA
661 ACACCT
```

A.3.3 *PhNPH3*

The Clone identificator "dr001P0004E24_F.ab1 2007-08-10" has "64994961" as NCBI's dbEST accesion, "FN001167" as EST name and GenBank Accession and "227590346" as GenBank gi. According to the phylogenetic analysis described in section 5.2.3, a tree-based reconstruct of sequence similarity can be found as figure A.3. The primers used for its amplification are: dr001P0004E24_F.ab1_attB1 TACAAAAAAGCAGGCTAGAGAATGGCATTGCAGCTT, dr001P0004E24_F.ab1_attB2 CAAGAAAGCTGGGTATGATGATACGTTGCACGGTT.

```
  1 GGTTGCTTGC AGGCTTGACT TGGAGAAGAG AATGGCATTG CAGCTTGGAC AGGCTGTATT
 61 AGATGATCTG TTAATTCCTT CATATTCCTT CACAGGGGAC ACATTGTTTG ATGTTGAAAC
121 CGTGCAACGT ATCATCATGA ACTTCCTTGA CAATGAAATG GATGGAAACC GATTAATGGG
181 AGATGAAGAC TATGTTTCCC CTTCATTAAG CGACATGGAG CGTGTTGGGA AACTTATGGA
241 AAATTACCTC GCTGAAATAG CCTCTGACCG TAATCTATCT GTTTCGAAGT TCATTAATCT
301 TGCTGAAATT ATCCCGGAGC AAGCAAGGAT TACTGAAGAT GGGATGTACA GGTCCATTGA
361 CATTTATTTG AAGGCACATC CAGTTCTAAG CGACTTGGAA AGGAAAAAGG TTTGCAGTGT
421 AATGGACTGT CAAAAGCTAT CTAGAGAAGC TTGTGCTCAT GCTGCTCAAA ATGATAGGCT
481 TCCTGTTCAG ACAGTTGTTC AAGTACTTTA TTACGAGNCA CAACGCCTTC GCGATGTCAT
541 GAGTAACGGG GGTAGCCAGC TTGTAGCAAC TCCTGAACCT CCCGCTGCTC TAATTCCT
```

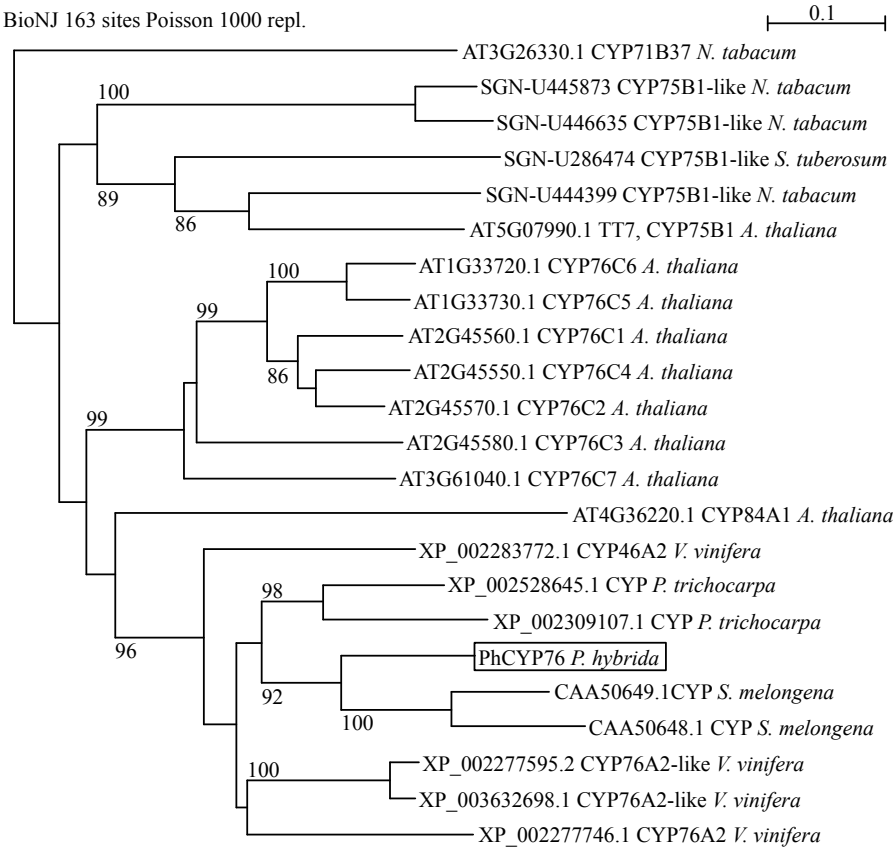


Figure A.2: Phylogenetic analysis of *PhCYP76* (also known as drs12P0022M07_R.ab1).

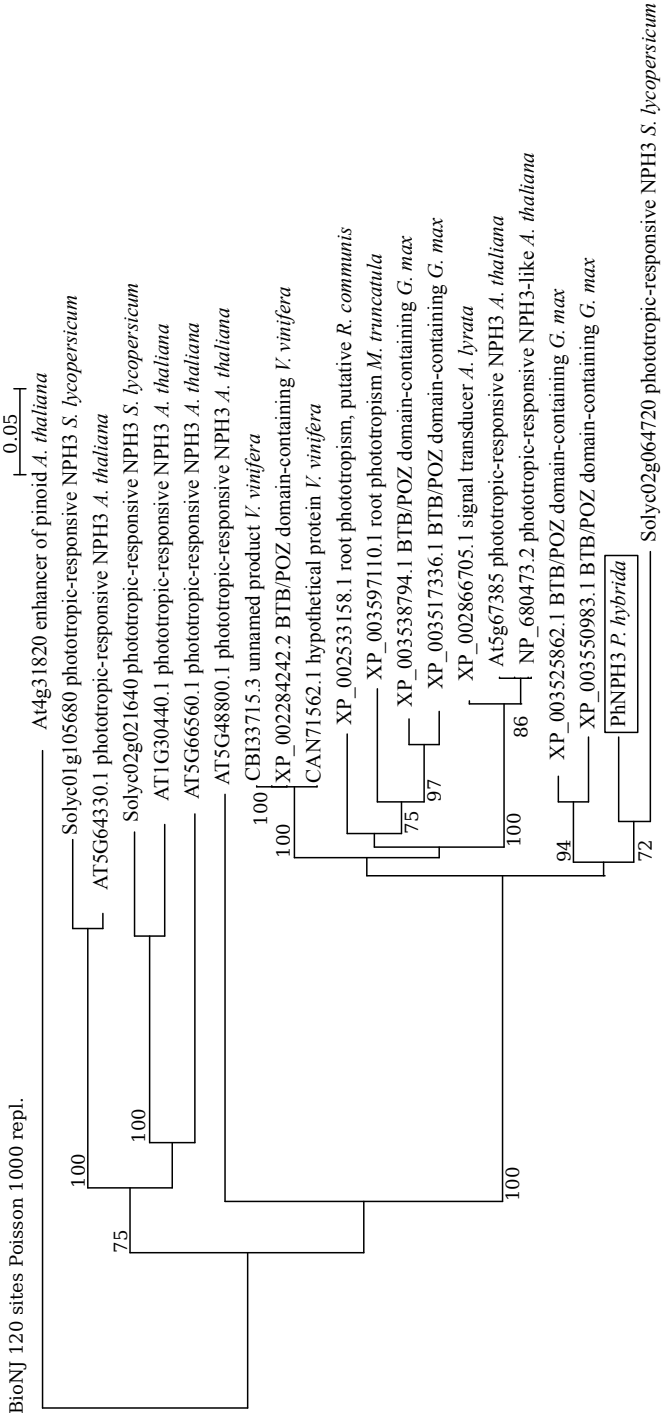


Figure A.3: Phylogenetic analysis of *PhNPH3* (also known as dr001P0004E24_F.ab1).

A.3.4 *PhFeSOD*

The Clone identificator "dr004P0002O22_F.ab1 2007-08-10" has "65016385" as NCBI's dbEST accesion, "FN007598" as EST name and GenBank Accession and "227608720" as GenBank gi. According to the phylogenetic analysis described in section 5.2.3, a tree-based reconstruct of sequence similarity can be found as figure A.4. The primers used for its amplification are: dr004P0002O22_F.ab1_attB1 TACAAAAAAGCAGGCTTAACAATGGTGCTCCCCTTC, dr004P0002O22_F.ab1_attB2 CAAGAAAGCTGGGTACTCTCCTCCTCCATTCGGTT.

```
  1 TTTGCTTTAT TTCAACTTGT CAGTCACCTT TTACACACAA GTCACAACTG ATCGATTATG
 61 GCTGCTGTTA CTGCTTCACT GACATCTGCT TTTCTTCCTC GTCAAGGATT TAATGGATCA
121 TCTCAGAGCC TACAATGGAA ATCCCAAAAG AAACAGTTAG CAAGAAAAGC TGGTCATGGT
181 ACAGTAACAG CTAAATTGGA ACTCCAGCCT CCTCCTTATC CGATGGATGC TTTGGAGCCT
241 CATATGAGTA GTAGAACATT CGAATTTCAC TGGGGGAAGC ATCACAGGGC TTATGTCGAC
301 AACTTAAACA AGCAAATAAA TGGAACAGAA CTAGATGGAA AGACACTAGA AGACTTAATT
361 CTTGTTACAT ATAACAATGG TGCTCCCCTT CCAGCATTCA ACAATGCTGC TCAGGCCTGG
421 AATCATCAGT TCTTTTGGGA ATCTATGAAA CCGAATGGAG GAGGAGACC GGCTGGTGAA
481 CTATTAGAAC TAATCAACAG AGACTTTGGT TCCTTTGATG CATTTGTTAA AGAGTTTAAG
541 GCAGCTGCAG CAACACAATT TGGCTCTGGT TGGGCCTGGC TTGCATACAA ACCCGAAGAC
601 AAAAAGCTTG CAAT
```

A.3.5 *PhXTH*

The Clone identificator "dr004P0023M11_F.ab1 2007-08-10" has "65011425" as NCBI's dbEST accesion, "FN014937" as EST name and GenBank Accession and "227606149" as GenBank gi. According to the phylogenetic analysis described in section 5.2.3, a tree-based reconstruct of sequence similarity can be found as figure A.5. The primers used for its amplification are: dr004P0023M11_F.ab1_attB1 TACAAAAAAGCAGGCTAAAGCCCCATTTACTGCCTT, dr004P0023M11_F.ab1_attB2 CAAGAAAGCTGGGTAACCATCTCATGCCTTTGGAG.

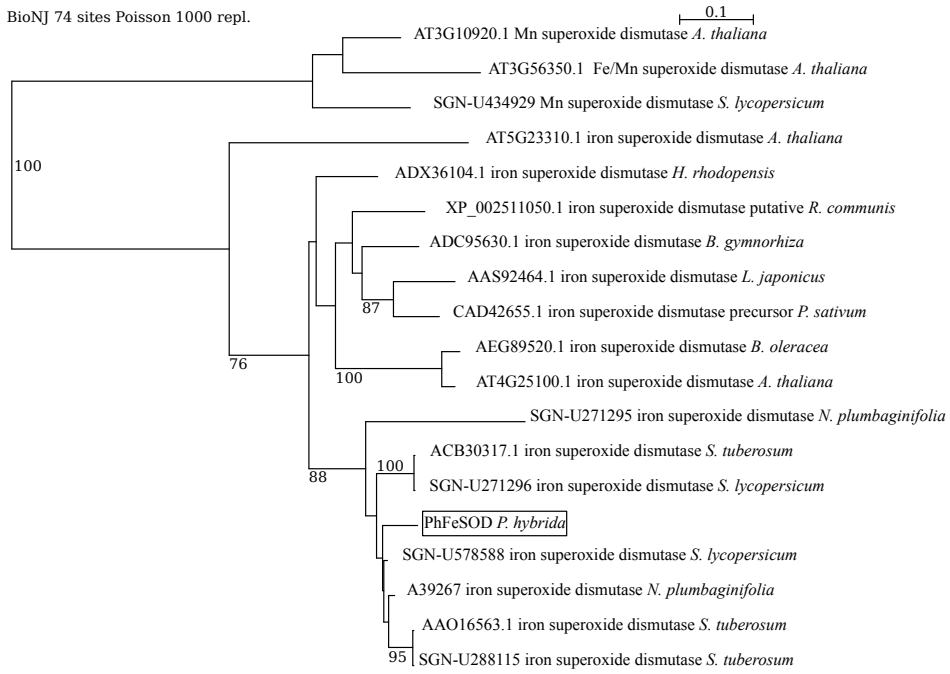


Figure A.4: Phylogenetic analysis of *PhFeSOD* (also known as dr004P0002O22_F.ab1).

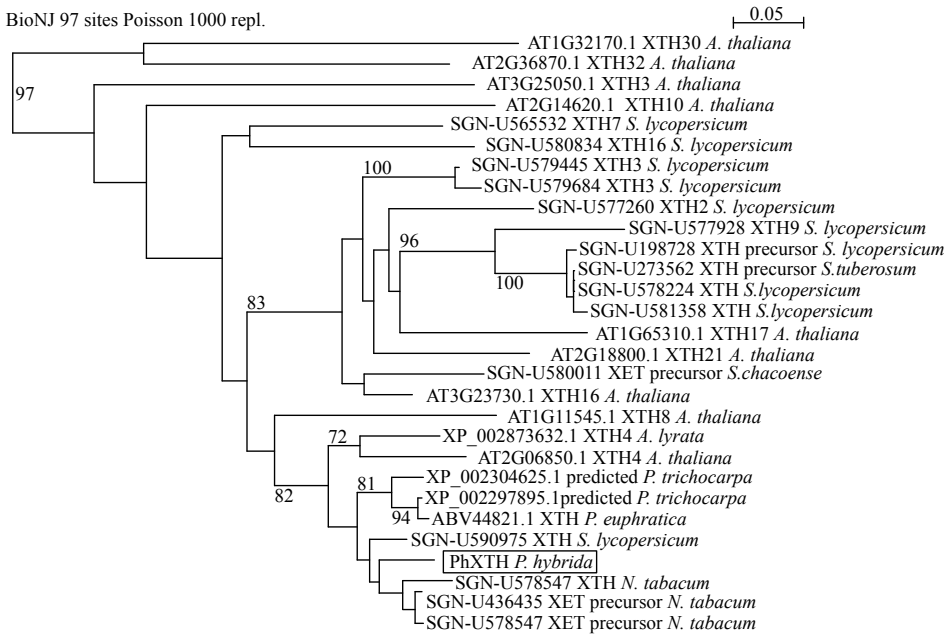


Figure A.5: Phylogenetic analysis of *PhXTH* (also known as dr004P0023M11_F.ab1).

```
  1 CTGTTTGGAC CTTCAGCATG AGAATGAAGC TTGTTGGAGG AGATTCAGCT GGTGTTGTCA
 61 CTGCATTTTA CCTGTCATCA AACAATGCAG AGCACGATGA GATAGATTTT GAGTTCTTGG
121 GTAACAGGAC AGGTCAGCCA TATATATTGC AAACAAATGT GTTCACGGGA GGAAAAGGAG
181 ACAGGGAACA AAGAGTATAT CTCTGGTTTG ACCCAACCAA GGGCTTCCAT GATTATTCCG
241 TCCTCTGGAA TACCTACCAG ATTGTGATTT TTGTGGATGA TGTCCCAATT AGAGCATTCA
301 AGAACTCCAA AGACCTTGGT GTAAAATTCC CATTCAATCA GCCCATGAAG ATATACTCAA
361 GCCTTTGGGA TGCAGATGAC TGGGCCACAA GAGGTGGTTT GGAGAAAACT GACTGGTCCA
421 AAGCCCCATT TACTGCCTTA TACACTTCAT TCCACGTAGA TGGTTGTGAA GCTGCAAACC
481 CACAAGAAGT CCAAGTTTGC AACTCCAAAG GCATGAGATG GTGGGATCAA AAGGCTTTCC
541 AAGATTTAGA TGCTCTACAA TACAGGAGAC TTCGTTGGGT TCGCCAAAAA TACACTATCT
601 ATAACTACCG CACTGACAGG ACGAGGTACC CTACCCTTCC ACCAGAATGC ACCCAGGATA
661 GAGATATATA AATTTGAGAG GGTATTTTAT GAGAACTTAT ATGTGAGTGT ATCTATGGGT
721 GTACTATAAC ATCCGAGAGA ACCCTAACAA GAGTTGTAAG CTCTATATTC CCTATCTT
```

A.3.6 *PhCYP96*

The clone identificator “drs31P0010G01_R.ab1 2007-08-10” has “65021539” as NCBI’s dbEST accesion, “FN022015” as EST name and GenBank Ac-

cession and "227617536" as GenBank gi. According to the phylogenetic analysis described in section 5.2.3, a tree-based reconstruct of sequence similarity can be found as figure A.6. The primers used for its amplification are: drs31P0010G01_R.ab1_attB1 TACAAAAAAGCAGGCTGGATTTTCAACGCGGACTTA, drs31P0010G01_R.ab1_attB2 CAAGAAAGCTGGGTATCGATCACTGGAATTAGCCC.

```
  1 CATTCAGGGG GCACTTTTCT ATTCAAGGTC CTTGGTTTGC TAACATGGAC ATGTTATCAA
 61 CAGTCGATCC AGCCAATATT CATTACATCA TGAGTTCAAA TTTCATGAAT TTTCCCAAAG
121 GCCCTAAATT CAAGGAGATG TTTGATGTTT TAGGAAATGG GATTTTCAAC GCGGACTTAG
181 ATTTGTGGAA AATTCAGAGG AAGACAACGC GAGCTTTGAT CACTCATCAG CAGTTTTACA
241 AGTTTTTAGT CAAGACTAGT CGCGACAAGG TAGAAAAAGG GCTAATTCCA GTGATCGATC
301 ATATATGCCA GACGGGTTCA ATAGTGGACT TGCAAGATTT GTTCCAAAGG TTCACTTTTG
361 ATACTACTTG TATTTTAGTC ACTGGATATG ATCCTGGATG TGTATCTATA GATTTCCCTG
421 ATGTTCCTTT CTCGAAAGCA ATGGATGATG CTGAAGAAGC CATTCTTTTC CGCCATGCAT
481 TGCCCGAGAT TATTTGGAAG TTGCAAAGAT GGCTTAGAAT TGGAGAAGAA AAGAAGATGA
541 TTAAAGCTTG GGAGACGTTG GACTATACTA TAGGTAATTA CATATC
```

A.3.7 *PhWAK*

The Clone identificator "dr001P0003O03_F.ab1 2007-08-10" has "64994813" as NCBI's dbEST accesion, "FN001019" as EST name and GenBank Accession and "227588278" as GenBank gi. According to the phylogenetic analysis described in section 5.2.3, a tree-based reconstruct of sequence similarity can be found as figure A.7. The primers used for its amplification are: dr001P0003O03_F.ab1_attB1 TACAAAAAAGCAGGCTACGAGTATGTCCCCAATGGA, dr001P0003O03_F.ab1_attB2 CAAGAAAGCTGGGTAGATGGTAAATGGGGGGAACT.

```
  1 TGTTAGTGTG CCAAGCTAGG CAATACTAAA GGCACTGATC AAGTTCTCAA CGAGGTCCGA
 61 ATACTTTGTC AAGTTCACCA CAAGAACCTC TTGCGCATTT TAGGCTGTTG TGTTGAGCTT
121 GAACAGCCAT TGCTGGTTTA CGAGTATGTC CCCAATGGAA CTCTCTTCGA TCATTTGCAA
181 GGTCCAAATC GCAAACTTCT CACTTGGGAT TGTCGACTCA ACATCGCCCA TGCAACAGCA
241 GAGGGACTTG CTTACCTACA TTTTTCTGCA GTTCCCCCCA TTTACCATCG CGATGTCAAG
301 TCCAGCAACA TATTGCTCGA CGAGAAGTTG AACGCTAAGG TCTCAGACTT TGGTCTTTCA
361 CGGTTGGCGC ATGCAGATCT AAGTCATGTG TCCACCTGTG CTCAGGGTAC CCTTGGTTAT
421 CTTGATCCTG AGTATTACAG GAACTATCAA TTGACGGACA AGAGTGATGT GTACAGTTTC
481 GGGGTTGTGT TGTTGGAACT CTTGACATCC CAAAGAGCAA TAGACTTTGA TAGAGCACAG
541 GATGATGTGA ACTTGGCTGT ATACG
```

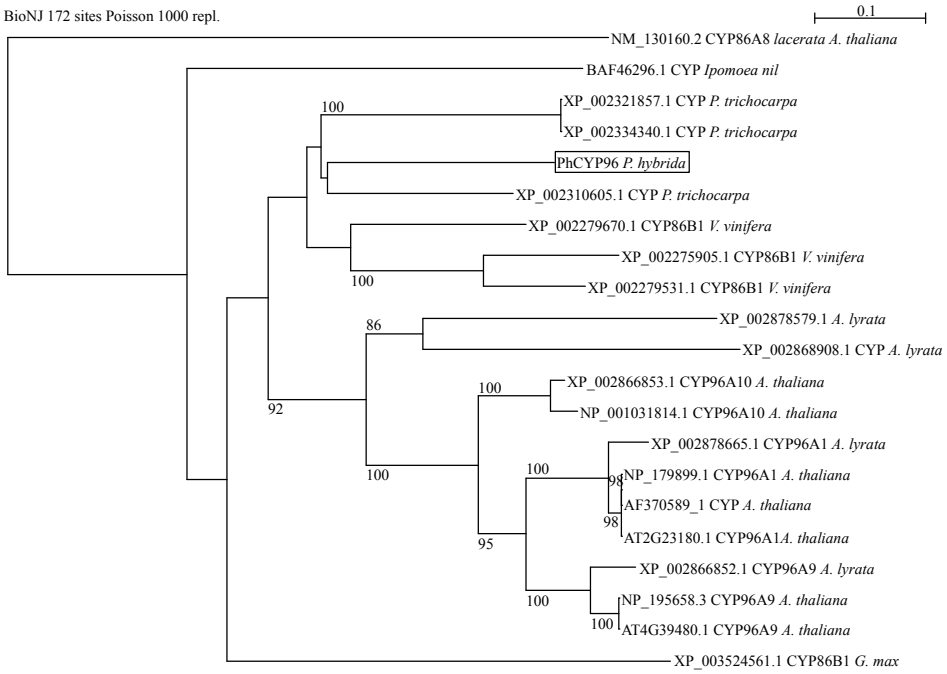


Figure A.6: Phylogenetic analysis of *PhCYP96* (also known as drs31P0010G01_R.ab1).

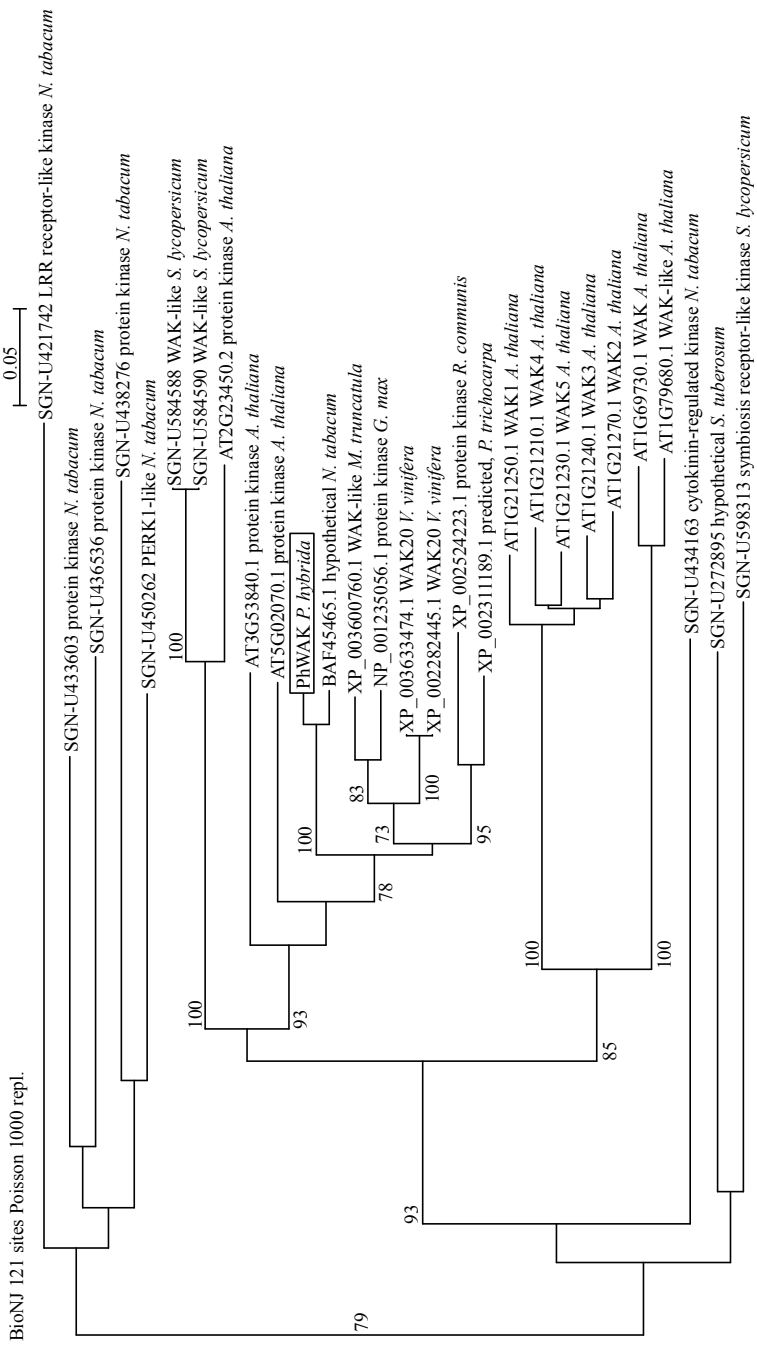


Figure A.7: Phylogenetic analysis of *PhWAK* (also known as dr001P0003O03_F.ab1).

A.3.8 *PhRBK*

The Clone identificator "dr001P0014N20_F.ab1 2007-08-10" has "64989706" as NCBI's dbEST accesion, "FN004829" as EST name and GenBank Accession and "227586588" as GenBank gi. According to the phylogenetic analysis described in section 5.2.3, a tree-based reconstruct of sequence similarity can be found as figure A.8. The primers used for its amplification are: dr001P0014N20_F.ab1_attB1 TACAAAAAAGCAGGCTCAATACTTGCACCATGGCAG, dr001P0014N20_F.ab1_attB2 CAAGAAAGCTGGGTAATTACGCGTCTTTGGCAATC.

```
  1 TTTGCCTGAT GGTCAAGTTG TCGCAGTCAA GAAAATAACC AAGAAAGAAA AGAATGATGA
 61 GGACAGAGTT GGAGATTTCT TATCTGAGTT AGGAATCATT GCTCACATCA ACAATCCAAA
121 TGCTGTTAAG TTAATTGGTT TTAGTGTTGA CAGTGGTTTG CACCTTGTTC TTCAATACTT
181 GCACCATGGC AGCCTGGCTT CTGTATTACA CGGTACTGAA GTATGCCTTG AATGGAAAAT
241 AAGATATAAA GTGGCAGTTG GGGTGGCTGA AGGACTGCGT TATCTTCATT GTGATTGCCA
301 AAGACGCGTA ATCCATAGAG ATATTACAGC CTCGAACATT CTTCTAACTG AAGAATACGA
361 ACCTCAGATA TCTGATTTTG GACTAGCAAA GTGGCTGCCA GAAAAATGGG TACATCATGT
421 TGTTTCCCCT ATAGAAGGAA CTTTTGGATA TATGGCACCA GAGTACTTTA TGCACGGAAT
481 TGTCCATGAG AAGACTGATG TTTTCGCCTT TGGAGTTCTG CTGTTGGAGC TTATAACTGG
541 TCGCCGTGCG GTTGATTCAT CTCGACAGAG TCTTGTGATG TGGGCAAAAC CACTGTTGNA
601 AAAGAGCAAC ATCAAGGAAT T
```

A.3.9 *PhPRA*

The Clone identificator "dr004P0022H01_F.ab1 2007-08-10" has "65004935" as NCBI's dbEST accesion, "FN014456" as EST name and GenBank Accession and "227598868" as GenBank gi. According to the phylogenetic analysis described in section 5.2.3, a tree-based reconstruct of sequence similarity can be found as figure A.9. The primers used for its amplification are: dr004P0022H01_F.ab1_attB1 TACAAAAAAGCAGGCTTGGTTTTGATTCATGCTTCG, dr004P0022H01_F.ab1_attB2 CAAGAAAGCTGGGTAACCAACCAATCCGAGATCAG.

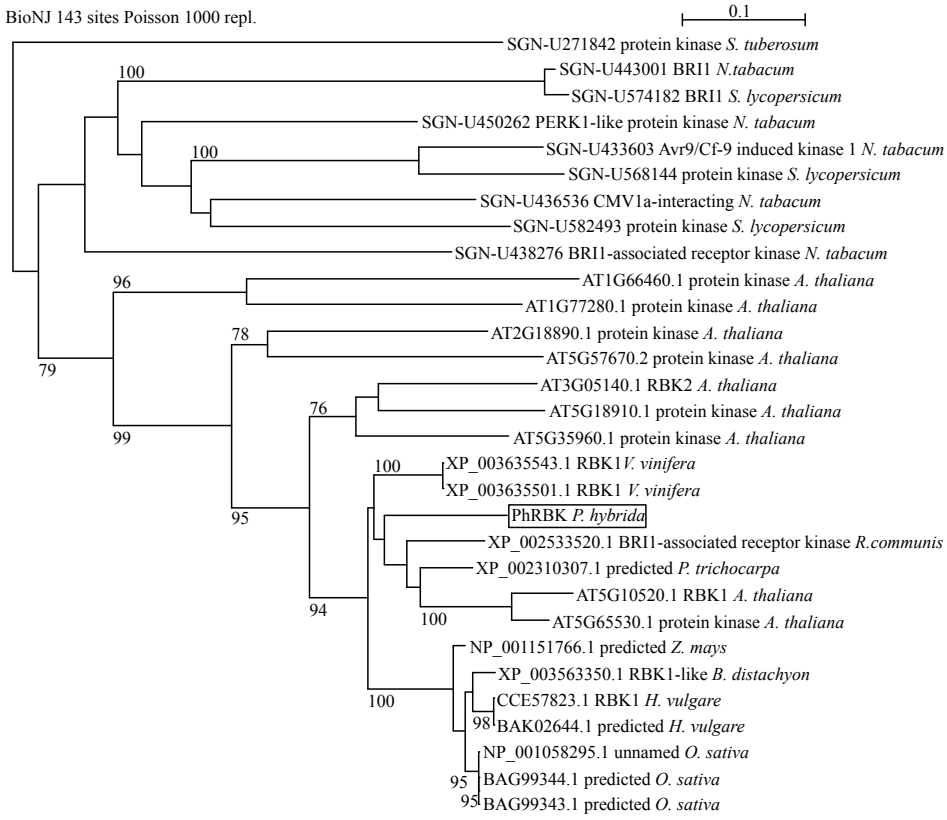


Figure A.8: Phylogenetic analysis of *PhRBK* (also known as dr001P0014N20_F.ab1).

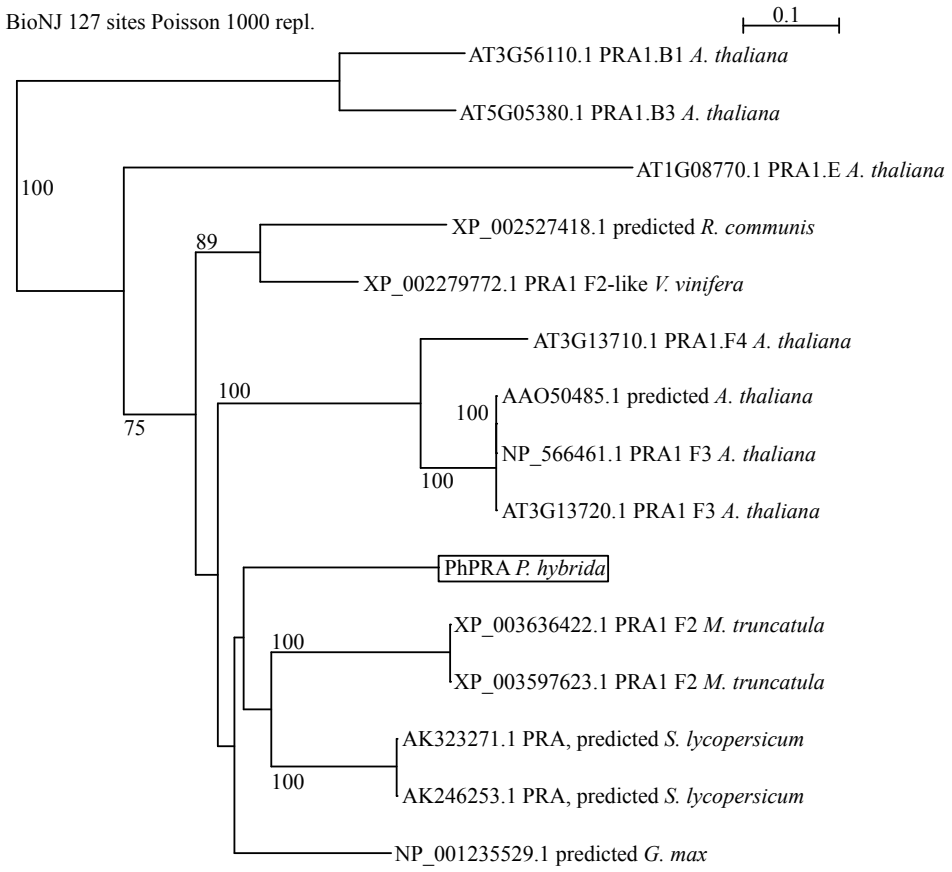


Figure A.9: Phylogenetic analysis of *PhPRA* (also known as lcl—dr004P0022H01_F.ab1).

```
  1 AGCAAAGAAC GAATCTATGA AGGTTTAGGT ACACGTCGAC CATGGAAAGA GATGTTTAAT
 61 TACCAGTCGT TAAAGTTTCC ACATGGGGTC AACGATGCAT TTTCAAGAAT CAAACACAAC
121 ATCTCTTATT TCCAAATGAA CTATGCTATT ATTGTATTAC TTATATTGTT TGTCAGTCTT
181 ATCTGGCACC CCATCTCTCT TATTGTGTTC ATTGTTTTGA TGGGTGTTTG GCTGTTTCTT
241 TATTTTCTTC GTGATGATCC TTTGGTGATT TTTGGCCGTT TGATTTCTGA TCGTGTGGTT
301 CTTATTGGTC TTGCGATCAT TACAATTGGA TTGCTCATGT TGACTCGTGC CACGTCCAAT
361 ATATTGATAG CGTTGTTGGT TGGATTTGTT GTGGTTTTGA TTCATGCTTC GCTTAGGAAA
421 ACTGATGATT TGTATGATGA AGAAGCTGCT GCTGAGTTCG TGAGTCCTTC GTCTTAGTTT
481 CGTTGGGTGC GGTCCTTCCT CGAATTCTGA TCTCGGATTG GTTGGTTCGT GCAATGGGCT
541 GTCCATTATT GTGTGTCAAT CAATGTAGTT TATTTTTAAT TTATTTATTT TGATAACTTG
601 TTTGATTGAT TGTTGTAATC ACAATACGCT
```

A.3.10 *PhTCP*

The Clone identificator "drs13P0013N08_R.ab1 2007-08-10" has "65023270" as NCBI's dbEST accesion, "FN036047" as EST name and GenBank Accession and "227614700" as GenBank gi. According to the phylogenetic analysis described in section 5.2.3, a tree-based reconstruct of sequence similarity can be found as figure A.10. The primers used for its amplification are: drs13P0013N08_R.ab1_attB1 TACAAAAAAGCAGGCTGGCGTGTTCGTCTTTCAGTT, drs13P0013N08_R.ab1_attB2 CAAGAAAGCTGGGTAGAGGCGGAAGCTCAGAGATA.

```
  1 GATGATGAAG AAGATGTAGG AGAAGTGAAA AAGAATAGTC TTTGTGATGT AGCTAGGCTT
 61 TGTGGTTGGC CCTCATCAAG AATTGTTAGA GTGTCCCGAG CATCAGGAGG GAAAGATAGG
121 CATAGTAAAG TATGGACTTC GAAGGGACTA AGAGATAGGC GTGTTCGTCT TTCAGTTAAT
181 ACAGCTATTC AGTTTTATGA TTTGCAAGAC CGGCTCGGCT ATGATCAGCC GAGCAAGGCT
241 GTGGAGTGGC TGCTAAAAGC AGCCGCTCCT TCTATCTCTG AGCTTCCGCC TCTTGAGGCA
301 TTTCCAGATA CTATGCAGCT CAGTGATGAG AAAAGGTCTA GTGCTGGAAC CGAGCCCGGG
361 TTTGATTCAG CTGATGTTGA AATGGATAAT GATGATGACC CAAATTACAA CTATCAGCAA
421 CAACAACAAC AACAACCTTG TAGTAGCAAT TCTGAGACTA GTAAAGGTTC TGGATTGTCA
481 CTTTCTAGGT CAGAAAGTCG GATCAAAGCG CGAGAAAGAG CAAAGGANAG AGCTGTAGAG
541 AAGGAAAAGG AGAAAGAAAA TGAGTCTACT GTTGTTGCTC ATCACCAAAA TATTCACCCT
601 AGCTCTTCTT TCACTGAGCT ATTGACAGGT GGTATGAACA ACAACAACAA CAACAACAAC
661 ACGAGTCCCG TTCACCAAAA CACCCAAAGG CCATGGTCTA ATAATCCTTT GGAATACTTT
721 ACCTCAGGAT TATTAGGTCC ATCACTACTC GTGGA
```

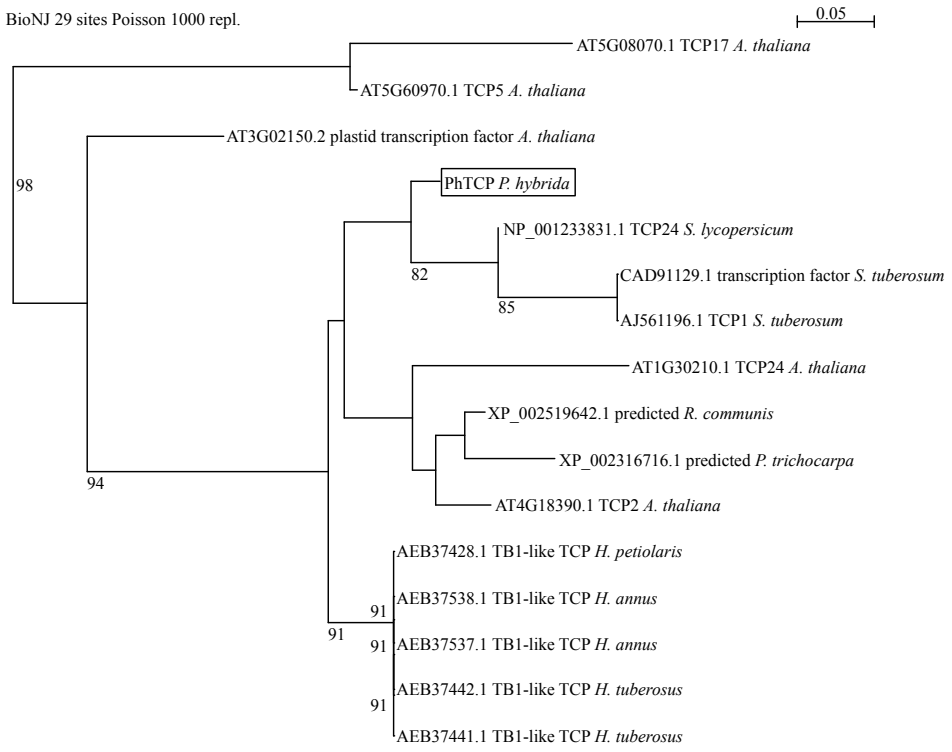


Figure A.10: Phylogenetic analysis of *PhTCP* (also known as drs13P0013N08_R.ab1).

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.)

Index