Zhihong Gao   *Editor*

The Prunus mume Genome

Springer

Compendium of Plant Genomes

Whole-genome sequencing is at the cutting edge of life sciences in the new millennium. Since the first genome sequencing of the model plant Arabidopsis thaliana in 2000, whole genomes of about 100 plant species have been sequenced and genome sequences of several other plants are in the pipeline. Research publications on these genome initiatives are scattered on dedicated web sites and in journals with all too brief descriptions. The individual volumes elucidate the background history of the national and international genome initiatives; public and private partners involved; strategies and genomic resources and tools utilized; enumeration on the sequences and their assembly; repetitive sequences; gene annotation and genome duplication. In addition, synteny with other sequences, comparison of gene families and most importantly potential of the genome sequence information for gene pool characterization and genetic improvement of crop plants are described.

Interested in editing a volume on a crop or model plant?

Please contact Dr. Kole, Series Editor, at ckoleorg@gmail.com

More information about this series at http://www.springer.com/series/11805

Zhihong Gao

Editor

The Prunus mume Genome

Editor
Zhihong Gao
College of Horticulture
Nanjing Agricultural University
Nanjing, China

This book series is dedicated to my wife Phullara, and our children Sourav, and Devleena

Chittaranjan Kole

Preface to the Series

Genome sequencing has emerged as the leading discipline in the plant sciences coinciding with the start of the new century. For much of the twentieth century, plant geneticists were only successful in delineating putative chromosomal location, function, and changes in genes indirectly through the use of a number of "markers" physically linked to them. These included visible or morphological, cytological, protein, and molecular or DNA markers. Among them, the first DNA marker, the RFLPs, introduced a revolutionary change in plant genetics and breeding in the mid-1980s, mainly because of their infinite number and thus potential to cover maximum chromosomal regions, phenotypic neutrality, absence of epistasis, and codominant nature. An array of other hybridization-based markers, PCR-based markers, and markers based on both facilitated construction of genetic linkage maps, mapping of genes controlling simply inherited traits, and even gene clusters (QTLs) controlling polygenic traits in a large number of model and crop plants. During this period, a number of new mapping populations beyond F_2 were utilized and a number of computer programs were developed for map construction, mapping of genes, and for mapping of polygenic clusters or QTLs. Molecular markers were also used in the studies of evolution and phylogenetic relationship, genetic diversity, DNA fingerprinting, and map-based cloning. Markers tightly linked to the genes were used in crop improvement employing the so-called marker-assisted selection. These strategies of molecular genetic mapping and molecular breeding made a spectacular impact during the last one and a half decades of the twentieth century. But still they remained "indirect" approaches for elucidation and utilization of plant genomes since much of the chromosomes remained unknown and the complete chemical depiction of them was yet to be unraveled.

Physical mapping of genomes was the obvious consequence that facilitated the development of the "genomic resources" including BAC and YAC libraries to develop physical maps in some plant genomes. Subsequently, integrated genetic–physical maps were also developed in many plants. This led to the concept of structural genomics. Later on, emphasis was laid on EST and transcriptome analysis to decipher the function of the active gene sequences leading to another concept defined as functional genomics. The advent of techniques of bacteriophage gene and DNA sequencing in the 1970s was extended to facilitate sequencing of these genomic resources in the last decade of the twentieth century.

As expected, sequencing of chromosomal regions would have led to too much data to store, characterize, and utilize with the-then available computer software could handle. But the development of information technology made the life of biologists easier by leading to a swift and sweet marriage of biology and informatics, and a new subject was born—bioinformatics.

Thus, the evolution of the concepts, strategies, and tools of sequencing and bioinformatics reinforced the subject of genomics—structural and functional. Today, genome sequencing has traveled much beyond biology and involves biophysics, biochemistry, and bioinformatics!

Thanks to the efforts of both public and private agencies, genome sequencing strategies are evolving very fast, leading to cheaper, quicker, and automated techniques right from clone-by-clone and whole-genome shotgun approaches to a succession of second-generation sequencing methods. The development of software of different generations facilitated this genome sequencing. At the same time, newer concepts and strategies were emerging to handle sequencing of the complex genomes, particularly the polyploids.

It became a reality to chemically—and so directly—define plant genomes, popularly called whole-genome sequencing or simply genome sequencing.

The history of plant genome sequencing will always cite the sequencing of the genome of the model plant Arabidopsis thaliana in 2000 that was followed by sequencing the genome of the crop and model plant rice in 2002. Since then, the number of sequenced genomes of higher plants has been increasing exponentially, mainly due to the development of cheaper and quicker genomic techniques and, most importantly, the development of collaborative platforms such as national and international consortia involving partners from public and/or private agencies.

As I write this preface for the first volume of the new series "Compendium of Plant Genomes," a net search tells me that complete or nearly complete whole-genome sequencing of 45 crop plants, eight crop and model plants, eight model plants, 15 crop progenitors and relatives, and 3 basal plants is accomplished, the majority of which are in the public domain. This means that we nowadays know many of our model and crop plants chemically, i.e., directly, and we may depict them and utilize them precisely better than ever. Genome sequencing has covered all groups of crop plants. Hence, information on the precise depiction of plant genomes and the scope of their utilization are growing rapidly every day. However, the information is scattered in research articles and review papers in journals and dedicated Web pages of the consortia and databases. There is no compilation of plant genomes and the opportunity of using the information in sequence-assisted breeding or further genomic studies. This is the underlying rationale for starting this book series, with each volume dedicated to a particular plant.

Plant genome science has emerged as an important subject in academia, and the present compendium of plant genomes will be highly useful both to students and teaching faculties. Most importantly, research scientists involved in genomics research will have access to systematic deliberations on the plant genomes of their interest. Elucidation of plant genomes is of interest not only for the geneticists and breeders, but also for practitioners of an array of plant science disciplines, such as taxonomy, evolution, cytology,

physiology, pathology, entomology, nematology, crop production, biochemistry, and obviously bioinformatics. It must be mentioned that information regarding each plant genome is ever-growing. The contents of the volumes of this compendium are, therefore, focusing on the basic aspects of the genomes and their utility. They include information on the academic and/or economic importance of the plants, description of their genomes from a molecular genetic and cytogenetic point of view, and the genomic resources developed. Detailed deliberations focus on the background history of the national and international genome initiatives, public and private partners involved, strategies and genomic resources and tools utilized, enumeration on the sequences and their assembly, repetitive sequences, gene annotation, and genome duplication. In addition, synteny with other sequences, comparison of gene families, and, most importantly, the potential of the genome sequence information for gene pool characterization through genotyping by sequencing (GBS) and genetic improvement of crop plants have been described. As expected, there is a lot of variation of these topics in the volumes based on the information available on the crop, model, or reference plants.

I must confess that as the series editor, it has been a daunting task for me to work on such a huge and broad knowledge base that spans so many diverse plant species. However, pioneering scientists with lifetime experience and expertise on the particular crops did excellent jobs editing the respective volumes. I myself have been a small science worker on plant genomes since the mid-1980s and that provided me the opportunity to personally know several stalwarts of plant genomics from all over the globe. Most, if not all, of the volume editors are my longtime friends and colleagues. It has been highly comfortable and enriching for me to work with them on this book series. To be honest, while working on this series I have been and will remain a student first, a science worker second, and a series editor last. And I must express my gratitude to the volume editors and the chapter authors for providing me the opportunity to work with them on this compendium.

I also wish to mention here my thanks and gratitude to the Springer staff, Dr. Christina Eckey and Dr. Jutta Lindenborn in particular, for all their constant and cordial support right from the inception of the idea.

I always had to set aside additional hours to edit books beside my professional and personal commitments—hours I could and should have given to my wife, Phullara, and our kids, Sourav, and Devleena. I must mention that they not only allowed me the freedom to take away those hours from them but also offered their support in the editing job itself. I am really not sure whether my dedication of this compendium to them will suffice to do justice to their sacrifices for the interest of science and the science community.

Kalyani, India   Chittaranjan Kole

Preface

Prunus mume (*Prunus mume* Sieb. et Zucc), commonly known as Japanese apricot, is an ornamental and fruit tree plant. It originated in China, where it was domesticated more than 3000 years ago, and its genome is one of the first genomes among the *Prunus* subfamily of the Rosaceae family to be sequenced. The genome of *Prunus mume* was fully sequenced in 2012. A 280-Mb genome was assembled combining 101-fold next-generation sequencing and optical mapping data. Scaffolds of 83.9% were further anchored to eight chromosomes, with the genetic map being constructed by restriction-site-associated DNA sequencing. Combining the *P. mume* genome with available data, scientists succeeded in reconstructing the nine ancestral chromosomes of the Rosaceae family, as well as in depicting the chromosome fusion, fission and duplication history in three major subfamilies. The transcriptome of various tissues and genome-wide analysis revealed the characteristics of *P. mume*, including the regulation of early blooming in endo-dormancy, the immune response against bacterial infection and the biosynthesis of the flower scent. The *P. mume* genome sequence increases our understanding of Rosaceae evolution and provides important data for the improvement of fruit trees in this family.

This book aims at reviewing of whole-genome sequencing, including the nuclear genome, the chloroplast genome, the functional genome, the molecular biology, molecular markers, epigenetics and the genetic relationship with other Rosaceae species.

11 scientists from two countries have authored 14 chapters of this book to illustrate the strategies of whole-genome sequencing and advanced breeding techniques. Experts from China and Japan, involved in *Prunus mume* research, investigated the different aspects of *Prunus mume* genetics and how genome sequencing could affect biological explanation, breeding and production in *Prunus mume*.

The book will be a guide for those who are interested in gene discovery, comparative genomics as well as molecular and advanced breeding techniques. It will be particularly useful for scientists, breeders and students involved in research related to the development of the citrus industry for updating the amount of knowledge generated in recent years.

In Chap. 1, the production, origin and the present distribution, as well as the economic importance and medical value of *Prunus mume*, are described. The botanical description is covered in Chap. 2, which provides essential information about genetics and molecular mapping. In Chap. 4, the focus is

on taxonomy and germplasm. The highlight of this book is Chap. 5, which introduces whole-genome sequencing as well as gene annotation and genome evolution. Molecular mapping, systems biology, small RNA, transcriptomics and the molecular and developmental biology on self-incompatibility and pistil abortion are described in the other chapters. In addition, genome sequence-based marker development, with the knowledge of the genome, will be discussed.

Nanjing, China   Zhihong Gao
February 2019

Contents

Contributors

Wenxing Chen Graduate School of Agriculture, Kyoto University, Kyoto, Japan

Zhihong Gao College of Horticulture, Nanjing Agricultural University, Nanjing, People's Republic of China

Shahid Iqbal College of Horticulture, Nanjing Agricultural University, Nanjing, People's Republic of China

Yuto Kitamura Japanese Apricot Laboratory, Wakayama Fruit Experimental Station, Wakayama, Japan

Wenjie Luo College of Horticulture, Nanjing Agricultural University, Nanjing, People's Republic of China

Xiaopeng Ni College of Horticulture, Nanjing Agricultural University, Nanjing, People's Republic of China

Ting Shi College of Horticulture, Nanjing Agricultural University, Nanjing, People's Republic of China

Lidan Sun Beijing Key Laboratory of Ornamental Plants Germplasm Innovation & Molecular Breeding, National Engineering Research Center for Floriculture, Beijing Laboratory of Urban and Rural Ecological Environment, Engineering Research Center of Landscape Environment of Ministry of Education, Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants of Ministry of Education, School of Landscape Architecture, Beijing Forestry University, Beijing, China

Ryutaro Tao Graduate School of Agriculture, Kyoto University, Kyoto, Japan

Hisayo Yamane Graduate School of Agriculture, Kyoto University, Kyoto, Japan

Qixiang Zhang Beijing Key Laboratory of Ornamental Plants Germplasm Innovation & Molecular Breeding, National Engineering Research Center for Floriculture, Beijing Laboratory of Urban and Rural Ecological Environment, Engineering Research Center of Landscape Environment of Ministry of Education, Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants of Ministry of Education, School of Landscape Architecture, Beijing Forestry University, Beijing, China

Abbreviations

2-DE	Two-dimensional gel electrophoresis
ABA	Abscisic acid
ACC	1-aminocyclopropane-1-carboxylic acid
AFLP	Amplified fragment length polymorphism
AGL24	*AGAMOUS-LIKE24*
AP2/ERF	APETALA 2/ethylene-responsive element binding factor
AS-PCR	Allele-specific polymerase chain reaction
At	*Arabidopsis thaliana*
BAC	Bacterial artificial chromosomes
BD	Blooming date
BEAT	Benzyl alcohol acetyltransferase
BR	Bud break
CBF	C-repeat binding transcription factor
cDNA	Complementary DNA
CEN	Centroradialis
CO	CONSTANS
cpDNA	Chloroplast DNA
CR	Chilling requirement
CRL	Chilling requirement of leaf buds
CS	Cell structure
CUL1s	Cullin1-likes
DAM	Dormancy-associated MADS-box transcription factors
DCL3	*dicer-like 3*
DGE	Digital gene expression
DUBs	Deubiquitinating enzymes
EM	Energy metabolism
EST	Expressed sequence tag
FAB1	Fatty acid biosynthesis 1
FAD5	Fatty acid desaturase 5
FB	Flower buds
FT	Flowering Locus T
Fv	*Fragaria vesca*
GA	Gibberellin
GBS	Genotyping-by-sequencing
GI	General inhibitor
GO	Gene ontology

GPX	Glutathione peroxidase
GS	Genomic selection
GSI	Gametophytic self-incompatibility
GWAS	Genome-wide association study
HR	Heat requirement
HRL	Heat requirement for bud break of leaf buds
HVR	Hypervariable region
IF	Imperfect
IPCC	Intergovernmental Panel on Climate Change
IR	Inverted repeat
IRMP	International Rosaceae mapping project
JADB	Japanese apricot dormant bud EST database
LAMP	Loop-mediated isothermal amplification
LC–MS/MS	Liquid chromatography and tandem mass spectrometry
LD	Leafing date
LEA	Late embryogenesis-abundant
LFY	LEAFY
LG	Linkage group
LRR	Leucine repeat sequence
LSC	Large single copy region
LTP3	Lipid transfer protein 3
LTR	Long terminal repeat
MAB	Marker-assisted breeding
MALDI-TOF/TOF	Matrix-assisted laser desorption/ionization time of flight/time of flight
MAS	Marker-assisted selection
Md	*Malus × domestics*
miRNA	Micro RNA
MOS	Mirror orientation selection
mRNA	Messenger RNA
NBS	Nuclear binding site
NGS	Next-generation sequencing
OR	Oxidation–reduction
Os	*Oryza sativa*
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEs	Paired ends
PF	Perfect
PM	Protein metabolism
Pm	*Prunus mume*
PME	Pectin methyl esterase
PPME1	Pectin lyase-like superfamily protein
Pt	*Populus trichocarpa*
PVE	Phenotypic variance explained
QTLs	Quantitative trait locus
R	Resistance
RACE	Rapid amplification of the cDNA ends
RAD	Restriction-site-associated DNA

RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
RGA	Resistance gene analogues
RGS	Reference genome sequence
RNase	Ribonuclease
RNA-Seq	Transcriptome sequencing
ROS	Reactive oxygen species
rRNA	Ribosomal RNA
RT-PCR	Reverse transcription-polymerase chain reaction
SC	Self-compatible
SD	Stress and defence
SF	Self-fruitfulness
SFB	S haplotype-specific F-box protein gene
SFBB	S-locus F-box brothers
SI	Self-incompatibility
SLAF-seq	Specific locus amplified fragment sequencing
SLF	S-locus F-box
SMRM	Single-molecule restriction map
SMRT	Single-molecule real time
SNP	Single-nucleotide polymorphism
snRNA	Small nuclear RNA
SOC1	Suppressor of overexpression of CONSTANS 1
SSC	Small single copy region
SSH	Suppression subtractive hybridization
SSR	Simple sequence repeat
SU	Self-unfruitfulness
SVP	*Short Vegetative Phase*
TCP2	Teosinte Branched/Cycloidea/PCF
TE	Transposable element
TFL1	Terminal Flower1
tRNA	Transfer RNA
VB	Vegetative buds
Vv	*Vitis vinifera*
WGD	Whole-genome duplication
WGM	Whole-genome mapping
WGS	Whole-genome sequencing
XTH2	Xyloglucan endotransglucosylase/hydrolase 2

Production and Academic Importance of *Prunus mume*

Zhihong Gao

Abstract

Prunus mume has a great ornamental and nutritional value. The species is considered as an important fruit crop due to its special values. It is distributed throughout Eastern Asia, while the largest production region is southern China. The fruits are used in Chinese traditional medicine.

1.1 *Prunus mume* Production in Eastern Asia

The flowers of *Prunus mume* (Japanese apricot) are white for most of the fruiting varieties and red for the flowering varieties. They appear during the winter on bare branches and are fragrant, which adds to the uniqueness of the tree's character.

P. mume is spread all over the globe, especially in Eastern Asia including China, Japan and Korea (Chu 1999). The main distribution areas in China are Yunnan, Guangdong, Fujian, Zhejiang and Jiangsu Province. The total production area is about 80,000 ha in China and 10,000 ha in Japan, where the fruits are mostly being produced in the centre and the southern region (Fig. 1.1). The fruits are mainly used for the production of candy and mume wine.

1.2 Medicinal Values of *Prunus mume*

P. mume fruits are consumed as preserved fruits and have traditionally been used as herbal medicine for alleviating some common disorders in China (Terada and Sakabe 1988). In the recent years, scientists have started to pay more attention to the medicinal properties of *P. mume*. For example, the concentrated fruit juice improves human blood fluidity in vitro (Yoshikawa et al. 2002), while the methanol extract of *P. mume* flowers exhibits inhibitory activity on aldose reductase and platelet aggregation in vitro (Shi et al. 2009). The extract of *P. mume* fruits can inhibit the growth of some cancer cell lines, such as breast, lung, liver, colon and pancreatic cancer (Adachi et al. 2007; Nakagawa et al. 2007; Okada et al. 2007, 2008; Mori et al. 2007). There are several kinds of organic acids present in *P. mume* fruits (Chen et al. 2006); however, the antifungal activities of *P. mume* fruits and their organic acids have not been investigated so far.

The potential of effective substances as antifungal agents has been reported in recent years (Pawar and Thaker 2007). For example, acetone is used as a universal solvent due to its lower toxicity for plant pathogens and its medium polarity. Since ancient times, *P. mume* has successfully been used in herbal medicines for alleviating

Z. Gao (✉)

College of Horticulture, Nanjing Agricultural University, No. 1 Weigang, Nanjing 210095, People's Republic of China
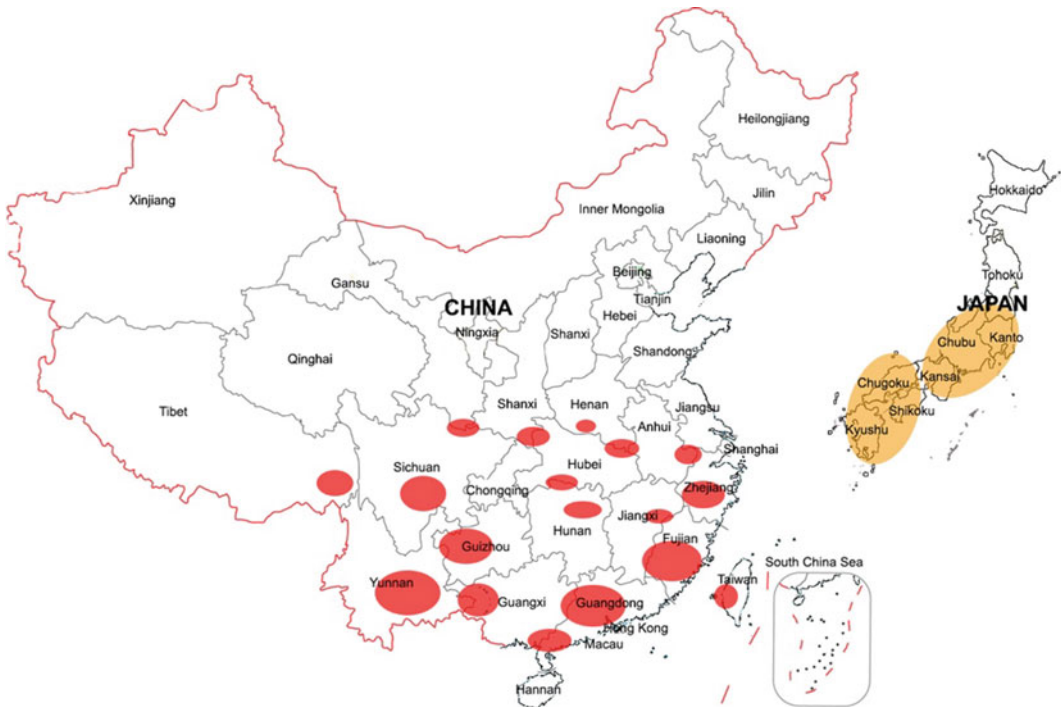e-mail: gaozhihong@njau.edu.cn

Fig. 1.1 Main distribution and production areas of Japanese apricot

fever, cough and intestinal disorders (Matsuda et al. 2003), and the fruits of *P. mume* are an important source of antioxidants, anticancer and antibacterial agents (Jo et al. 2006; Otsuka et al. 2005). *P. mume* fruits can also be used as antibacterials against *Streptococcus mutans*, *Streptococcus mitis*, *Streptococcus sanguis*, *Porphyromonas gingivalis*, *Bordetella bronchiseptica* and *Helicobacter pylori* (Wong et al. 2010; Jung et al. 2010; Miyazawa et al. 2006; Enomoto et al. 2010). The extract of *P. mume* seeds could inhibit the growth of pathogenic bacteria, including *Staphylococcus aureus*, *Escherichia coli*, *Salmonella enterica*, *Vibrio parahaemolyticus*, *Candida albicans*, *Saccharomyces cerevisiae* and *Aspergillus niger*, and the antibacterial activity could be attributed to the phenolic compounds (Xia et al. 2011). Researchers from Nanjing Agricultural University have investigated the antifungal activity of the acetone extract of *P. mume* fruit against the pathogenic fungi including *Phytophthora capsici* Leonian, *Fusarium solani* (Mart.) Sacc., *Fusarium graminearum*

Schwl, *Botryosphaeria dothidea* (Moug.) Ces. et De Not., which are the pathogenic fungi in herbaceous and woody plant. The extract of *P. mume* is a rich source of organic acids, including oxalic, maleic and acetic acid, potentially affecting the mycelial growth of plant pathogenic fungi and showing a strong antifungal action against plant pathogenic fungi due to the presence of oxalic and acetic acid. This suggests that *P. mume* fruit extract can potentially be used in agriculture as a natural fungicide. The organic acids significantly contribute to the antifungal activity, and further studies are needed to determine the specific components responsible for the high antifungal activity.

References

Adachi M, Suzuki Y, Mizuta T, Osawa T, Adachi T, Osaka K, Suzuki K, Shiojima K, Arai Y, Masuda K (2007) The "*Prunus mume* Sieb. et Zucc" (Ume) is a rich natural source of novel anti-cancer substance. Int J Food Prop 10(2):375–384

Chen JY, Zhang H, Matsunaga R (2006) Rapid determination of the main organic acid composition of raw Japanese apricot fruit juices using near-infrared spectroscopy. J Agric Food Chem 54(26):9652–9657. https://doi.org/10.1021/jf061461s

Chu M (1999) Chinese fruit tree: *Prunus mume*. China Forestry Publishing House, Beijing (in Chinese)

Enomoto S, Yanaoka K, Utsunomiya H, Niwa T, Inada K, Deguchi H, Ueda K, Mukoubayashi C, Inoue I, Maekita T (2010) Inhibitory effects of Japanese apricot (*Prunus mume* Siebold et Zucc.; Ume) on *Helicobacter pylori*-related chronic gastritis. Eur J Clin Nutr 64 (7):714

Jo SC, Nam KC, Min BR, Ahn DU, Cho SH, Park WP, Lee SC (2006) Antioxidant activity of *Prunus mume* extract in cooked chicken breast meat. Int J Food Sci Technol 41(s1):15–19. https://doi.org/10.1111/j.1365-2621.2006.01234.x

Jung BG, Ko JH, Cho SJ, Koh HB, Yoon SR, Han DU, Lee BJ (2010) Immune-enhancing effect of fermented Maesil (*Prunus mume* Siebold & Zucc.) with probiotics against *Bordetella bronchiseptica* in mice. J Vet Med Sci 72(9):1195–1202

Matsuda H, Morikawa T, Ishiwada T, Managi H, Kagawa M, Higashi Y, Yoshikawa M (2003) Medicinal flowers. VIII. Radical scavenging constituents from the flowers of Prunus mume: structure of prunose III. Chem Pharm Bull (Tokyo) 51(4):440–443

Miyazawa M, Utsunomiya H, K-i Inada, Yamada T, Okuno Y, Tanaka H, Tatematsu M (2006) Inhibition of *Helicobacter pylori* motility by (+)-Syringaresinol from unripe Japanese apricot. Biol Pharm Bull 29 (1):172–173

Mori S, Sawada T, Okada T, Ohsawa T, Adachi M, Keiichi K (2007) New anti-proliferative agent, MK615, from Japanese apricot "*Prunus mume*" induces striking autophagy in colon cancer cells in vitro. World J Gastroenterol 13(48):6512

Nakagawa A, Sawada T, Okada T, Ohsawa T, Adachi M, Kubota K (2007) New antineoplastic agent, MK615, from UME (a variety of) Japanese apricot inhibits growth of breast cancer cells in vitro. Breast J 13 (1):44–49

Okada T, Sawada T, Osawa T, Adachi M, Kubota K (2007) A novel anti-cancer substance, MK615, from ume, a variety of Japanese apricot, inhibits growth of hepatocellular carcinoma cells by suppressing Aurora A kinase activity. Hepatogastroenterology 54 (78):1770–1774

Okada T, Sawada T, Osawa T, Adachi M, Kubota K (2008) MK615 inhibits pancreatic cancer cell growth by dual inhibition of Aurora A and B kinases. World J Gastroenterol 14(9):1378–1382

Otsuka T, Tsukamoto T, Tanaka H, Inada K, Utsunomiya H, Mizoshita T, Kumagai T, Katsuyama T, Miki K, Tatematsu M (2005) Suppressive effects of fruit-juice concentrate of *Prunus mume* Sieb. et Zucc. (Japanese apricot, Ume) on *Helicobacter pylori*-induced glandular stomach lesions in Mongolian gerbils. Asian Pac J Cancer Prev 6(3):337

Pawar V, Thaker V (2007) Evaluation of the anti-*Fusarium oxysporum* f. sp *cicer* and anti-*Alternaria porri* effects of some essential oils. World J Microbiol Biotechnol 23(8):1099–1106

Shi J, Gong J, Liu J, Wu X, Zhang Y (2009) Antioxidant capacity of extract from edible flowers of *Prunus mume* in China and its active components. LWT Food Sci Technol 42(2):477–482

Terada H, Sakabe Y (1988) High-performance liquid chromatographic determination of amygdalin in ume extract. Eisei kagaku 34(1):36–40. https://doi.org/10.1248/jhs1956.34.36

Wong R, Hägg U, Samaranayake L, Yuen M, Seneviratne C, Kao R (2010) Antimicrobial activity of Chinese medicine herbs against common bacteria in oral biofilm. A pilot study. Int J Oral Maxillofac Surg 39(6):599–605. https://doi.org/10.1016/j.ijom.2010.02.024

Xia D, Wu X, Shi J, Yang Q, Zhang Y (2011) Phenolic compounds from the edible seeds extract of Chinese Mei (*Prunus mume* Sieb. et Zucc) and their antimicrobial activity. LWT Food Sci Technol 44(1):347–349

Yoshikawa M, Murakami T, Ishiwada T, Morikawa T, Kagawa M, Higashi Y, Matsuda H (2002) New flavonol oligoglycosides and polyacylated sucroses with inhibitory effects on aldose reductase and platelet aggregation from the flowers of *Prunus mume*. J Nat Prod 65(8):1151–1155

Origin and Evolution of *Prunus mume*

2

Zhihong Gao and Wenjie Luo

Abstract

Prunus mume (*P. mume* Sieb. et Zucc., $2n = 2x = 16$) is a popular ornamental plant widely cultivated in East Asia, where it was domesticated over 3000 years ago. As an important member of the genus *Prunus*, it plays a pivotal role in phylogenetic studies of the Rosaceae family. To understand the origin of *P. mume*, it is important to know how the proto-species of *P. mume* did form and where it has its origin. Specimens of wild *P. mume* have been discovered in Yunnan Sichuan, Hubei, Guizhou and Tibet, China.

2.1 Origin of *Prunus mume*

Prunus mume has two origins: one is a gradual change from one original species, and the other is the hybrid pathway of natural hybridization between two species. It is commonly believed that the relationship between *P. mume* and *Prunus armeniaca* is the closest, but there are two different views on the origin of *P. mume*. Some scholars assume that *P. mume* is a subspecies of *P. armeniaca*, (Kaneko et al. 1986; Chu 1988),

while it is also believed that the species is a hybrid between *P. armeniaca* and *Prunus salicina*, and *Prunus persica* and *Prunus sibirica* might also be involved in this process (Chen 1995). However, it is commonly accepted that *P. mume* and *P. armeniaca* have the highest cross-compatibility and the closest relationship compared to other Rosaceous plants (QiXiang 1987; Chu 1988). There are several natural and artificial hybrids between *P. mume* and the related species (Liu 1996), such as *P. mume* var. bungo, with the earliest record in the 《Mei Pu》 (an ancient Chinese book mentioning some varieties of *P. mume*, written by Fan Chengda in 1186). In 1994, Shimada et al. used analysis with random amplified polymorphic DNA (RAPD) markers and reported that *P. mume* var. bungo Makino is a hybrid of *P. mume* and *P. armeniaca* (QiXiang 1987). In addition, there are some artificial hybrids as the result of hybridization between *P. mume* and *P. armeniaca* cv 'Large Yellow', *P. davidiana* and other *Prunus* species (QiXiang 1987; Boonprakob and Byrne 1995; Bao 1992).

Previous researchers have studied the hybridization of *P. mume* in the Hunan province of China and observed inter- and intraspecific introgression (Bao 1992). Introgression of *P. mume* is considered as one of the ways to improve genetic diversity in the evolution and explains the good cross-compatibility and hybrid fertility of hybrids with *P. armeniaca*, *P. persica*, and *P. salicina*.

Z. Gao (✉) · W. Luo
College of Horticulture, Nanjing Agricultural University, No. 1 Weigang, Nanjing 210095, People's Republic of China
e-mail: gaozhihong@njau.edu.cn

2.2 Origin of *Prunus mume*

The Chinese professor Junyu Chen provided a detailed explanation of the origin of Japanese apricot, also called Mei (*P. mume* Sieb. et Zucc). According to his theory, there are three disagreements about the origin of *P. mume*. First, it is still unclear whether the species originated from China, Japan, or Korea, and to answer this question, Prof Chen quoted different scriptures. Apparently, the scientific name *P. mume* was established in 1835 by P. F. Von Siebhold and J. G. Zuccarini, based on specimens of *P. mume* from Japan. In the early stages, *P. mume* trees were introduced from Japan to Europe, and therefore, the tree was named "Japanese apricot," based on the assumption that its sole origin is Japan. However, the species actually originated from China, according to reports of plant collectors and to ancient books published in China and Japan. The earliest wild *P. mume* trees found by plant collects were two wild varieties, *P. mume* var. pallescens Frunch and *P. mume* var cernua Franch. They were found in Dali, a city in Yunnan Province, China, in 1887 and 1910, respectively (Zhang et al. 2018). Later, the wild type of *P. mume* was found in Yichang, Hubei, China (1907) and west of Wenchuan County, Sichuan, China (1908 and 1910) (Liansen et al. 1993). Therefore, it is clear that *P. mume* originated in China. According to records from Japan, wild *P. mume* trees mostly grow in the valleys, and Japanese trees have been described as shrubby, with numerous withy and prickly branches and a creeping growth; the flowers are small, single-lobe and white; the fruits are small and have no flesh. These traits are significantly different from those of currently cultivated types. Although there is no evidence to prove whether Japan is the origin area of *P. mume*, according to this record, Mei was introduced into Japan from China at about the second to the eighth century B. C. Prosser Chen and other scholar, such as Masao Yoshida, still believe that the wild *P. mume* of Japan originated from China, and it is suggested that these wild plants are the progenies of seeds are brought to the mountains and valleys by humans or livestock (Yoshida 1999).

However, in the Neolithic age, Mei was more frequently distributed in the northern areas, at latitudes including North Korea and Japan. Also, the *P. mume* var. cernua Franch has its natural distribution in the north of Laos and Vietnam (Chu 1999). The last question concerns the natural distribution centers of *P. mume* in China. For a plant, its natural distribution range is related to its occurrence history, adaptability, spreading ability and condition, as well as to the obstacles and its distribution scale. Generally, it is considered that the center of diversity is the origin area of a species. Nevertheless, this situation may not be true in other regions because of the large-scale changes in the continental plate.

Based on previous studies (Ou et al. 1993, 1999; Fan et al. 1995; Daqing 1997; Weiguo et al. 2002), specimens of wild *P. mume* were discovered successively in Yunnan Sichuan, Hubei, Guizhou and Tibet, China. In the early twentieth century, E. H. Wilson discovered numerous wild specimens in Sichuan and Hubei, according to the specimen in the herbarium of the Beijing Institute of Plant Research, Chinese Academy of Sciences, who found the wild *P. mume* successively in Guizhou, Jiangxi, Fujian, Zhejiang, Guangdong, Guangxi, Yunnan and Sichuan. Based on the Chinese horticulturist Manzhu Bao, during three surveys during 1989–1991 in Yunnan, Sichuan, and Tibet, a wild *P. mume* wood was found in Tongmai Tibet and Eryuan Yunan, and a large portion of wild *P. mume* trees was sporadically distributed in Muli Sichuan, mainly near the Hengduan Mountains and the Yunnan-Guizhou Plateau, at the juncture of the southwestern Sichuan province, northwest Yunnan and southeast Tibet. In this range, the variation of the species is relatively high, with numerous mutations and a high genetic diversity, including almost all variants. Therefore, Hengduan Mountain can be considered the center of the natural distribution and genetic diversity of *P. mume* (Liansen et al. 1993). The species is, however, also widely distributed in other parts of China, including the Yangtze River Basin, the Pearl River Basin, southwest China and Taiwan. The distribution ranges from Tongmai and Linzhi Tibet in the

west to areas in the north and the south. The northern line spans from Tongmai to the northeast to Sichuan, further to the northeast to southwestern Gansu and Shaanxi, then extending eastward to Hubei, southern Henan, Huangshan in Anhui Province, Yixing Jiangsu, Linan Zhejiang and finally to the coast of the East China Sea. The southern line also begins in Tongmai and extends eastward to Deqin and Lincang Yunnan, then further to Guangxi, Guangdong, Fujian, and Taiwan.

2.3 The Evolution of *Prunus mume*

The evolution of *P. mume* can be summarized as the development from wild Mei to fruit mume and further to the floral *P. mume*, resulting in trees used for flowers and fruits. Based on historical records, the species was domesticated for its fruits and its medicinal properties, as recorded in the book of 《Shi Jing》, the earliest poetry anthology in China. In the period of the Western Han Dynasty (202 BC–8 AD), people were more and more attracted to the flowers of this species and started to hybridize the floral Mei to obtain the *P. mume* var. typical maxim and *P. mume* var. alphandi, used as ornament plants. The breeding of flowering varieties was developed in the Ming and Qing Dynasties. In the Ming Dynasty, there were 19 varieties of flowering mume.

2.4 Conclusions

The origin of *P. mume* is still a subject of debate. According to some scholars, *P. mume* is a branch of *P. armeniaca*, while others believe that *P. mume* is a hybrid between *P. armeniaca* and *P. salicina*. Recent genetic studies have shown that *P. mume* is the closest to *P. armeniaca*. *P. mume* originated from China, with southwestern China as the natural distribution center, namely the Hengduan Mountains.

References

Bao M (1992) Comparison studies on pollen morphology between different types of *P. mume* and *P. persica*, *P. salicina* and *P. armeniaca*. J Beijing Univ 14 (4):70–74

Boonprakob U, Byrne DH (1995) Genetic relationships of diploid plums based on RAPD polymorphisms. HortScience 30(4):763

Chen J (1995) Some Aspects on Chinese Mei flower research. J Beijing Forest Uni 17(S1):1–7

Chu M (1988) Studies on isozymes of Japanese apricot (*Armeniaca mume* Sieb.). J Fruit Sci 4(4):155–157

Chu M (1999) Chinese fruit tree: *Prunus mume*. China Forestry Publishing House, Beijing (in Chinese)

Daqing L (1997) The preliminary studied on the wild fruit resources at the Yaluzangbujiang Daguaiwan district. Resour Dev Market 6:10

Fan E, Wang G, Ou M, Yang S (1995) An investigation on wild resources of Japanese apricot and its ecological factors in Libo, Guizhou province. SW China J Agri Sci 1:017

Kaneko T, Terachi T, Tsunewaki K (1986) Studies on the origin of crop species by restriction endonuclease analysis of organellar DNA. Jpn J Genet 61(2):157–168. https://doi.org/10.1266/jjg.61.157

Liansen L, Shanwen H, Meihong L (1993) A preliminary study on the status of hybridization of cultivar resources of *Prunus mume* in Hunan. Acta Hort Sin 20(4):225–230

Liu Q (1996) Discussion on the origin and cultivar evolution of Mei. J Beijing Univ 2:78–82

Ou M, Fan E, Tang L, Wang G (1993) An investigation on wide Prum at Libo county of Guizhou. Guizhou Agric Sci 1:55–57

Ou M, Wang G, Fan E, Zhou Y (1999) Distribution, ecological adaptability and horticultural characteristics of wide plum in Libo, Guizhou province. SW China J Agri Sci 3:12

QiXiang Z (1987) The interspecific crossing of Mei flower and cold hardiness breeding. J Beijing Univ 1:69–79

Weiguo F, Weifan Z, Enpu F (2002) Germplasm resources of wild fruit tree in Guizhou province. GuiZhou Daxue Xuebao (China)

Yoshida M (1999) Fruit mume breeding in Japan. J Beijing Univ 2:38–42

Zhang Q, Zhang H, Sun L, Fan G, Ye M, Jiang L et al (2018) The genetic architecture of floral traits in the woody plant *Prunus mume*. Nat Commun 9(1):1702. https://doi.org/10.1038/s41467-018-04093-z

Botanical Description of *Prunus mume*

3

Zhihong Gao and Ting Shi

Abstract

Prunus mume is a small deciduous arbour tree species with a weak trunk and a shallow and broad root system. The flower buds of *P. mume* are covered with scales, and usually, one bud has only one flower. Branching is weak and gradually weakens with increasing tree age; the majority of the long twigs are located in the middle and upper parts of the branches. Adaxial leaf surfaces in *P. mume* bear hairs on young leaves, which gradually fall as the leaves mature. The fruiting branches of *P. mume* trees can be divided into six categories: water sprouts, long branches, middle branches, short branches, tiny branches and needle branches. According to the characteristics of morphological differentiation and biochemical changes, and based on the differentiation period of the sepal, in the stage of morphological differentiation, we distinguish flower primordium differentiation and the differentiation of flower organs. The pistil of *P. mume* consists of a single carpel, and the ovary is superior. There are generally three peak fruit-drop periods in the early stage,
and some varieties show fruit-drop before harvest. The phenological period of *P. mume* varies with region and variety. Its main feature is early flowering, first flowering and then leafing; the fruit growth and development periods are short, and leaf drop is early.

3.1 Growth Characteristics

3.1.1 Tree Characteristics

Prunus mume is a small deciduous arbour tree species with a weak trunk, which is, however, slightly stronger than that of the peach tree (Chu 1999). Under natural growth conditions, *P. mume* trees often show an evacuated layered shape or a multi-main branch. Under general cultivation conditions, adult grafted *P. mume* trees are 2.5–4 m high, with a crown width of 3.5–5 m. However, in the southern region with good growth conditions, *P. mume* usually has a higher stem. For example, in Yunnan Province, China, the height can reach 5–9 m, with a crown width of 4–8 m. Maximum crown width is 10.9–12.2 m, at a height of 8.9 m and a trunk circumference of 2.7 m. The trees start to fruit at 6–7 years. Grafted trees begin to bear fruit after three years and enter the fruiting period after 8–10 years. A *P. mume* orchard has an economic life of 40–50 years and of 20 years under inadequate management. Well-managed orchards can

Z. Gao (✉) · T. Shi
College of Horticulture, Nanjing Agricultural University, No. 1 Weigang, Nanjing 210095, People's Republic of China
e-mail: gaozhihong@njau.edu.cn

produce fruits up to 70 or 80 years. The trees have a long lifespan, and several old specimens have been found in the wild. For example, in the Guoqing temple of the Sui Dynasty, Tiantai County, Zhejiang Province, there is a 1400-year-old *P. mume* tree (Miao 2014), with a trunk circumference of 141 cm (Fig. 3.1).

3.1.2 Roots

The root system of *P. mume* is shallow and broad. Most of the roots in a *P. mume* orchard are distributed in the soil layer within a depth of 40 cm, and the dense layers are concentrated at a depth of 10–20 cm. On mountains, the root system is relatively deep. The horizontal distribution of *P. mume* roots is larger than that of the canopy. The germination period of the new root of *P. mume* is earlier than that of other stone fruit trees. In Hangzhou, from late December to early January, when the soil temperature reaches 4–5 °C, new roots begin to germinate, and root germination reaches a peak from January to March. In Guangdong, the roots of *P. mume* germinate earlier, usually starting in mid-December, but growth is most abundant in autumn (Chu 1999).

3.1.3 Buds, Branches and Leaves

3.1.3.1 Buds

The buds of *P. mume* are covered with scales and are pure flower buds; generally, one bud develops into one flower (Fig. 3.2). At the top of the new shoots, self-shearing occurs, mainly due to the formation of the separation layer in the lower part of stem tip after the new shoots stopping growing, thus self-picking and forming a pseudo-terminal bud. The axillary buds have single and multiple buds, although they are all compound buds because they form accessory buds on both sides of the main buds. The accessory buds of most of the axillary buds in the upper part and lower parts of the new shoots often form dormant buds hidden in the leaf axil, with a single-bud appearance. This single bud may be a flower bud or a leaf bud. The accessory buds of the axillary buds in the middle of the new shoot tend to develop into flower buds or leaf buds, or all of them are flower buds (Guo and Wujing 1995). The buds of *P. mume* have obvious heterogeneity. Usually, buds in the upper part of the branches are the largest, and the buds of the basal two to three nodes are not fully developed and become blind nodes. The germination rate of *P. mume* is as high as 95%, and

Fig. 3.1 Ancient *P. mume* tree of in the Guoqing temple of the Sui Dynasty (http:// andonglaowang.blog.163. com/blog/static/ 8448753220156402145406/)

there is no significant difference between varieties. The dormant buds have a long lifespan. Under favourable conditions, they can bloom. Once stimulated, they are easy to germinate and take out long branches. This is also the main reason why *P. mume* trees are easy to renew and rejuvenate.

3.1.3.2   Branches

The branches of *P. mume* are weak and gradually weaken with increasing age; the majority of the long twigs are located in the middle and upper parts of the branches (Fig. 3.3). Due to short internodes, the twigs are concentrated, and branches are easy to form, causing the phenomenon of a 'snap neck'. The middle and lower parts of the branches usually produce middle or short twigs, more near the base of branches that produces the shorter twigs; the blind nodes at the base do not germinate, forming a bare belt. On the farm in the Shanghang County, Fujian Province, the saplings grow fast, and the bark of 2-year-old trees naturally splits. Flower buds form on the secondary branches after the pinching of new branches on 1-year-old trees, blooming in the following year. The taper of the branches is small, allowing bending. After the branches drop, the buds on the back of the arched branches or the base buds are easy to take the water sprout/shoot disturbed the tree shape, resulting in crown closure. The branches of adult *P. mume* trees do not easily produce adventitious roots. However, 1-year-old branches of young trees can be treated with auxin, the results showed that softwood cutting produces the adventitious roots from the phloem parenchyma cells, but hardwood cutting produces the adventitious roots from the cambium (Jiang et al. 2014). There is a significant positive correlation between rooting rate and soluble sugar content in the medium, but a significant negative correlation between rooting rate and starch content.

3.1.3.3   Leaves

Adaxial leaf surfaces in *P. mume* bear hairs on young leaves, which gradually fall as the leaves mature. When the young leaves gradually unfold, the stipules cling to the branches and then gradually unfold and fall. Compared with leaves of apricot and peaches, the corneum and upper epidermis of the leaves of *P. mume* are thicker, with a high stomata density; the calibre is small, and the total stomatal area is large. The stomata are distributed in the lower epidermis, which are formed by two kidney-shaped guard cells. It has a narrow elliptical shape, and no accessory cells are present. The epidermal cells on the adaxial leave surface are significantly larger than those

Fig. 3.2 Flower buds of *P. mume*

on the abaxial leave surface. The greater the thickness/leaf thickness ratio (PR value) of the palisade of the *P. mume* leaf, the stronger the cold resistance of the variety, and the higher the ratio of palisade tissue/sponge tissue thickness, the stronger the drought resistance. The daily variation in the respiration rate of mature leaves of *P. mume* trees during the full fruiting period is small, and the respiration rate of the leaves during the day is highest at noon. However, the photosynthesis rate greatly varies throughout the day and is relatively low before 8:00 and after 17:00, reaching its peak at around 10:00. The 'noon break' of photosynthesis is followed by a gradual increase, with another peak at 15:00, followed by a sharp decline (Chu 1999).

3.2 Fruiting Characteristics

3.2.1 Fruiting Branch Types

The fruit branches of *P. mume* trees can be divided into six categories: water sprouts, long fruit branches, medium fruit branches, short fruit branches, tiny fruit branches and needle branches (Jiang et al. 2014).

The growth and fruiting characteristics of various types of fruit branches are different.

3.2.1.1 Water Sprout/Shoot

The length of the shoot is more than 40 cm, with autumn twigs occurring in the upper part of the parent branch. The growth is booming, and the proportion of flower buds and complete flowers is very low, with an extremely poor fruiting capacity.

3.2.1.2 Long Fruit Branch

With a length of 20–40 cm, these branches occur in the upper part of the parent branch. The growth potential is relatively high, the proportion of complete flowers is low, and the fruiting capability is poor. However, a strong fruiting branch can easily form a new fruiting branch group in the following year, with a strong continuous fruiting capability.

3.2.1.3 Medium Fruit Branch

With a length of 10–20 cm, these branches occur in the upper part of the parent branch. The growth potential is moderate with more complex flower buds, a high flower ratio and a strong fruiting capability.

3.2.1.4 Short Branch

With a length of 3–10 cm, short branches occur in the middle and lower parts of the parent branch. The growth potential is moderate, the

union is full, the total proportion of flowers is the highest and the fruiting capability is the strongest, with a certain continuous fruiting capability.

3.2.1.5 Ultrashort or Tiny Fruit Branch

Tiny fruit branches can reach a length of 3 cm and occur in the middle and lower parts of the parent branch; they generally have a weak growth potential. Almost all of the buds are single-flower buds, except for the leaf buds in the shoot apex.

The proportion of complete flowers is high, and the fruiting capability is strong, but after fruiting, only weak branches are produced, and the continuous fruiting capability is weak. In particular, the extremely short fruit branches below 1 cm often wither after fruiting.

3.2.1.6 Needle Branch

At a length below 10 cm and with a needle-like apex, needle branches mostly occur in wild *P. mume* trees, at the base of young trees and water sprouts under the condition of heavy pruning or excessive fertilisation. They only contain flower buds and no leaf buds, and they have no ability to generate new shoots. At strong needle branches, there are only thin leaf buds but

fewer flower buds, and the thin branches will die off in the following 1–2 years. The proportions and fruiting performances of various fruiting branches of *P. mume* trees vary among the different cultivars. Most of the fruit varieties mainly have short and ultrashort fruit branches, but there are also a few varieties in which mainly have medium and long branches. The proportion of various fruit branches also differs with tree age and potential. Saplings contain numerous long branches, while old and weak trees have more short branches. Branches generally continuously fruit for 5–6 years, and the fruiting capability is highest in the second and third years.

3.2.2 Flower Dud Differentiation

As in other fruit trees, *P. mume* bud differentiation can be divided into three stages: physiological differentiation, morphological differentiation, and sexual cell formation (Jiang et al. 2014). In early May, the short branches of *P. mume* trees stop growing for 15–20 days and enter the physiological differentiation stage of flower buds, which lasts about 1.5 months. At the end of June and in early July, the short
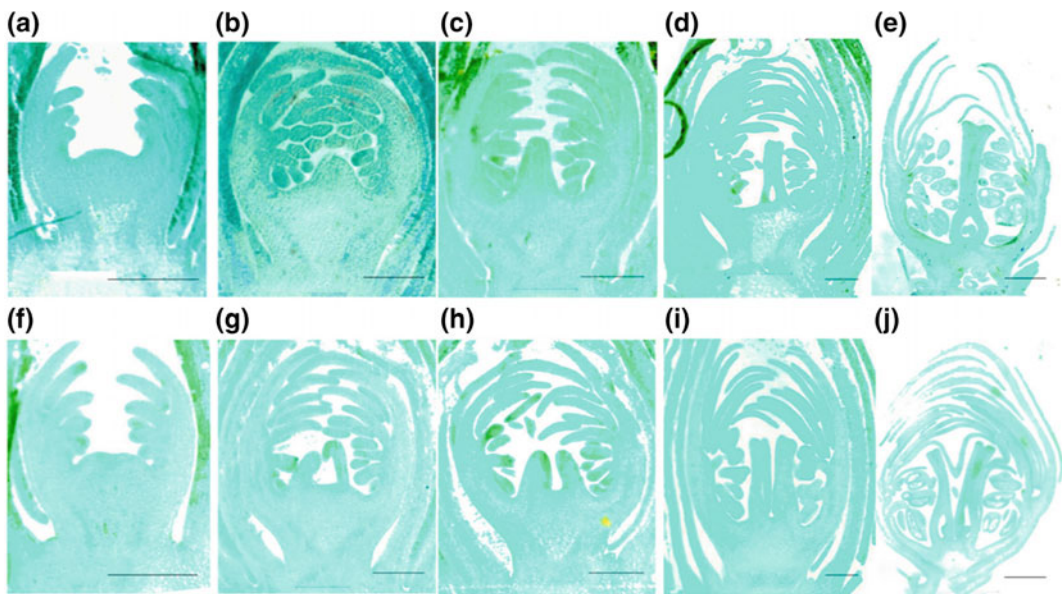

Fig. 3.4 Pistil differentiation process of *P. mume* flower buds

branches enter the stage of flower bud morphogenesis (Fig. 3.4).

According to the characteristics of morphological differentiation and biochemical changes, and based on the differentiation period of the sepal, the stage of morphological differentiation of *P. mume* trees can be divided into flower primordium differentiation (I, II, III in the initial stage of morphological differentiation) and the differentiation of flower organs (sepal, petal, stamen and pistil). From the formation of most flower primordia (July 5) to the differentiation of the pistil (October 4), the flower bud morphology development in *P. mume* takes more than three months (Shi et al. 2011b; Xu et al. 1992).

The stamen primordium begins to appear in late September. Subsequently, the stamen primordia are elongated, and the ends are slightly expanded to present the embryos of the anthers and filaments. In mid-December, the secondary sporulation cells further develop into pollen mother cells, which subsequently undergo meiosis and form a tetrad in mid- and late December. The formation of microspore tetrads is simultaneous in a tetrahedral arrangement and is co-surrounded by thicker callose layers. Shortly thereafter, the callose layers dissolve and the microspores are released. At the end of December, the coating on the outer wall of the pollen grains is formed, and the germination holes are visible. In mid- and late January of the following year, the pollen grains undergo a mitosis to form a vegetative cell and a germ cell. In early February, the pollen is mature; the pollen sac ruptures and the pollen grains are released at flowering in early March (Xu et al. 1992).

After the pistil primordium has been formed, its internal structure continues to differentiate. This period can be divided into two phases. The size of flower buds continues to increase in the early stage, the outer bud scales are loose, the basal extension of the stamens in the buds was slightly larger, and sporulation cells appear. The primordium of the pistil also elongates correspondingly, and the side of the carpel involutes and gradually heals. Later, flower buds continue to increase, and the bud scales are looser. At this stage, the enlarged anthers and short thin filaments of the stamens are clearly visible, and pollen mother cells and even pollen have formed in the pollen sac. However, the internal differentiation of the pistil is relatively slow, and the morphological differentiation of the stigma, style and ovary occurs after the carpel completely heals. An embryonic primordium is formed on the inside of the base of the ovary (Xu et al. 1995).

The differentiation peak stage of pistil primordia is in early October, about 1 week later than the peak stage of stamen primordia differentiation. In late December, the stigma, style and ovary have begun to differentiate and produce a two-ovule primordium in the ovary. From the end of February to the beginning of March of the following year, the embryo sac differentiates into a sputum-type embryo sac. The ovules gradually turn into hemitropous ovules during development. Pollen grains germinate on the stigma one day after fertilisation during flowering in early March. Part of the pollen tube is guided into the ovary cavity 2–3 days after pollination. Double fertilisation has been completed in most of the embryo sacs after around five days. When fertilisation occurs, the two sperms tend to the egg and the secondary nucleus, and one sperm enters the egg cell.

Later, a male nucleolus, smaller than the female nucleolus, appears in the egg nucleus, and finally, two equal female and male nucleoli merge into one large nucleolus. As a result, the haploid sperm and the haploid egg combine to form a diploid zygote, while the other sperm and the secondary nuclei form a primary endosperm nucleus and complete double fertilisation. After 8 days, the free endosperm nucleus begins to appear, and the proembryo starts to form after 13 days. The embryonic development of *P. mume* can be classified as 'toaster type'. Comparison of self- and cross-pollination shows that the pollen germination of cross-pollination is better than that of self-pollination, with a higher proportion of fertilised embryos.

During the flower bud differentiation of *P. mume* trees, nutrient metabolism constantly changes. During the physiological differentiation stage of the flower buds, the amino acid content

in the xylem fluid, the short branches and the leaves increases rapidly and decreases after the beginning of the morphological differentiation. The amino acid content increases again and decreases before and after flower primordium differentiation. Prior to the morphological differentiation, carbohydrates accumulate in the branches and leaves and are quickly consumed after the start of the morphological differentiation until the second peak in the primordial differentiation stage of the sepal primordia, with a high consumption during flower bud differentiation. This indicates that the flower bud differentiation of *P. mume* trees requires large amounts of carbohydrates and nitrogen for flower organ establishment.

The proteins are distributed in the floral primordium, sporulation cells, pollen mother cells and unicellular pollen grains. Starch granules only appear at the base of the bud, away from the growth cone, or in young bracts and petals formed through elongation of the primordium. The development of stamens and pistils has starch storage response when the pollen grains and embryo sacs appear. This phenomenon is consistent with the metabolism of carbon and nitrogen during flower bud differentiation, providing a basis for cell anatomy for the role of nutrients in flower bud differentiation (Xu et al. 1995).

Floral bud differentiation is closely related to the changes in endogenous hormones. During the physiological differentiation of *P. mume* buds, the activity of gibberellin in the xylem fluid decreases, the level of cytokinin is higher and the ratio of cytokinin/gibberellin sharply increases and reaches the maximum. The gibberellin content peaks at the morphological differentiation of flower buds, the cytokinin content increases, and subsequently, the cytokinin/gibberellin ratio does not change (Chu and Huang 1995). This indicates that the low concentration of gibberellin is beneficial to initiate the morphological differentiation of *P. mume* buds, while the differentiation of flower primordia requires more gibberellin. The equilibrium relationship between cytokinin and gibberellin has a direct control effect on

flower bud differentiation, along with a sequence effect. At the same time, gibberellin with strong polarity in the middle of flower bud physiological differentiation is the most active, and the activity of gibberellin, which is low before the formation of the flower primordium, increases rapidly, indicating that different kinds of gibberellin differently regulate flower bud differentiation. Moreover, there is a positive relationship between calmodulin and flower bud differentiation (Chu and Huang 1995; Cai et al. 2014).

3.2.3 Flowering and Fruiting

P. mume is an important deciduous fruit tree species (Zhuang et al. 2016). The flowering period varies greatly among different varieties. Early- and late-flowering varieties can be separated by 30 days. The flowering period of the same variety can be as long as 15–30 days, and the early-flowering variety has a longer flowering period. Due to different weather conditions before flowering, the flowering period of the same variety in the same region can vary by 40 days in different years. Early flowering in early years results in a long flowering period, while late flowering is associated with a short flowering period. The short fruit branches of the same tree bloom earlier, while the long fruit branches bloom relatively late (Deng 1998).

Regardless of length of the branch, the same fruit branch first opens its flowers above the middle of the *P. mume* from the small bud period to the complete blooming, which generally takes 3–5 days, and a few flowers only need one day. During the blooming period of *P. mume*, a few of the central anthers are cracked which leads to complete cracking within 2–3 days, and the filaments become brown and dry. After two days, the petals fall. The pollen emission period generally lasts 2–3 days, and the stigma can maintain its pollination and fertilisation ability within 3–4 days of flowering; after this, it gradually becomes brown and withered. The lifespan of a single flower is about 5–10 days. If the temperature during the flowering period is high, the

lifespan is short, while at lower temperatures, it is relatively long. With the early opening, the flower lifespan is relatively long (Xu et al. 1995).

The development of incomplete floral organs is common (Gao et al. 2012), such as hollow flowers, pistil absence, ovary withered, styles shortened or curved, and ovate-small deformed flowers (Chu 1999), which do not have the ability to fertilise and fall after blooming. There are different degrees of development between the lengths of the pistil, from complete degeneration to normal development. The length of the pistil in incomplete flower is about 0–0.6 cm. The critical length of the pistil of normal flowers is about 0.61–0.8 cm. The development tends to be normal when the pistil is longer than 0.8 cm, and the normal pistil can reach up to 1.35 cm. Each variety of *P. mume* has different degrees of incomplete flowers, and the difference among the varieties is large.

The proportion of incomplete flowers is also different for the same variety in different years (Shi et al. 2012a) and is closely related to the length of the fruit branches, the strength of the trees and the flowering period. Most varieties with long fruit branches have more incomplete flowers, while short fruit branches have flowers that are more complete. If the tree is too weak or too strong, its proportion of incomplete flowers is also high. Early-flowering varieties also have more incomplete flowers, while late-flowering varieties have flowers that are more complete. In the same tree of the same variety in the same year, there are more incomplete flowers in the early-flowering stage than in the full bloom stage (Shi et al. 2012b). Pistil malformation in the primordium differentiation period is significantly lower than in the flowering period. This further illustrates that the incomplete flower is mainly formed during the pistil structural differentiation period and due to abnormal pistil development.

The pollen number varies among different varieties and years. The pollen number of *P. mume* is lower than those of apricot, apple and other fruit tree species; therefore, the ratio of main varieties to pollination varieties should be $2\sim3{:}1$. The pollen germination rate also varies among years and varieties, but the trend of each variety is relatively stable. There is a clear diversity among the pollen quantity of different cultivars, with a coverage from 0 to 1265; also, there is an obvious diversity among the rate of pollen germination of different cultivars, with the highest rate being 84.76% and the lowest rate being 0. There is also an obvious correlativity between pollen quantity and the rate of pollen germination (Shi et al. 2011a). Therefore, most varieties can be used as pollination varieties. There is no significant difference in pollen germination rate between hollow flowers (without pistils) and complete flowers, suggesting that during flower organ development in the same variety, the degree of pistil development is different, but the difference between the stamens is small. Determination of the free proline content in anthers and pistils shows that the free proline content in anthers is significantly higher than that in pistils, and there is no significant positive correlation between free proline content and pollen number in anthers. However, there is a significant positive correlation with pollen germination rate, indicating that the free proline content in the anthers is closely related to pollen fertility.

3.2.4 Fruit Development

The pistil of *P. mume* consists of a single carpel with a superior ovary. Two hemitropous ovules are born in the ovary, and one of the ovules begins to degenerate after flowering and completely degrades and withers after 10–20 days. When the *P. mume* stigma is treated with gibberellin at the flowering stage, both ovules can develop normally, and a double embryo occurs (Chu 1999).

At the initial stage of flowering, the embryo sac contains 4–8 nuclei, and the fertilisation process is completed in 5–7 days after flowering and subsequently, primary and primordial endosperm nucleus appears. The primary endosperm nucleus begins to divide early, and the free endosperm nucleus appears 8–10 days after flowering and turns into endosperm cells at about 40 days after flowering. The zygote begins to

divide for the first time around 18 days after flowering, undergoing globular embryo and heart-shaped embryo development. It develops into embryos with obvious cotyledon germ, hypocotyl and radicle differentiation at 65 days after flowering. During this period, the endosperm gradually disintegrates, and the embryo continuously absorbs the nutrients of the endosperm to develop. The ovary wall cells divide actively from the flowering stage. About 30–35 days after flowering, the endocarp cells stop dividing, and the stomata of the epidermis and the epidermis of the exocarp are differentiated. The mesocarp cells begin to increase after flowering, and the growth rate increases at the late stage of fruit development. The endodermal cells undergo a short-term increase after the cessation of division, and lignification occurs from about 40–45 days after flowering, entering the hard nucleus stage (Xu et al. 1995).

The fresh and dry weight of *P. mume* fruits follow a double sigmoid growth curve. According to the changes in the growth rate, the fruit development process can be divided into three periods, namely the rapid growth period, the slow growth period, also known as hard-core period, and the second rapid growth period. The length of the first phase varies from species to species, and the early-maturing varieties are shorter or even less obvious. Weight gain mainly occurs in the second stage. In this period, fresh weight increase accounts for 67.4% of the total increase, and the dry weight accounts for 69.9% of the total weight. The second period is the most important period for fruit development. The increased weight of the pulp is 74.1% and 71.6% of the total fresh weight (Chu 1999).

The growth rate of the fruit longitudinal diameter is greater than that of the cross diameter in the first phase, but smaller than the cross diameter in the third phase so that the fruit shape index is changed from large to small, and the fruit develops from a long shape to a flat shape. The fresh weight of *P. mume* seeds and the vertical and horizontal diameter growth show a single sigmoid growth curve, and the longitudinal and cross diameters reach the maximum in the first stage. With fruit development, the titratable acid content in the pulp continuously increases, while the pH value decreases. The vitamin C content decreases linearly in the first stage of fruit development, and the first and second phases are relatively stable. The contents of soluble sugar and starch decrease gradually in the first and second phases of fruit development and gradually increase in the first half of the third phase. In the latter half of the period, the starch hydrolyses to soluble sugar, the starch content decreases rapidly, and the soluble sugar content increases sharply. The content of insoluble pectin in the pulp increases until 1 week before the fruit turns green and then decreases rapidly. The soluble pectin content decreases slowly in the first phase and in the first half of the second phase and then slowly increases, with a sharp increase from the mature green stage onwards (Deng 1998).

Asphaltic acid is the main free amino acid species in *P. mume* (Chu 1999). The total free amino acid content decreases rapidly in the first phase and increases rapidly in the second and third phase, with a sharp increase thereafter. The content of gibberellin in the seeds is higher than that in the pulp. As the fruit matures, the gibberellin content in pulp and seeds decreases gradually, the content of abscisic acid increases significantly, and the ratio of gibberellin/abscisic acid decreases rapidly (Zhong et al. 2013; Zhuang et al. 2013). There is no significant correlation between the contents of endogenous hormones and the growth of fruits and seeds that might be related to seed dormancy. The edible portion of the pulp increases with the weight of the fruit after the end of the hard-core period. The moisture content of most varieties is the highest at the end of the hard-core period, and the absolute water content increases with increasing fruit weight. The concentrations of acids and vitamin C are high, with a high proportion of edible pulp, and the quality, fruit weight and water content are higher than those of peach, with a high early yield. Harvesting maturity should be determined according to the processing requirements, which is essential for improving the quality and yield of the processed products (Lin et al. 2014).

There is a significant correlation between leaf weight, leaf area, average leaf width, maximum leaf width, fruit weight, core weight, fruit height and fruit width, while leaf length is negatively correlated with the average leaf width, that is, the longer the leaves, the narrower the average leaf width. The correlations between fruit weight and seed weight and between fruit height and fruit width are also significant. The heavier the fruit, the greater the fruit width. Correlations between the traits of fruit trees are complex, but fruit weight is one of the important factors affecting the yield of *P. mume*.

3.2.5   Fruit Drop

There are generally three peak periods for fruit drop, and some varieties drop their fruits before harvest.

3.2.5.1   First Fruit Drop

This occurs 6–15 days after the flowering. The main loss is due to incomplete flowering with normal shape but abnormal internal ovule development. According to a previous study, two ovules are degraded, the nucleus disintegrates prematurely, and there is no embryo. The differentiation of the capsule or embryo sac is stagnant. In this period, a separation layer is produced at the base of the flower stalk, and the flower is detached from the handle. Incomplete flower bud differentiation and floral organ dysplasia are the main reasons for fruit drop during this period (Xu et al. 1995).

3.2.5.2   Second Fruit Drop

This occurs 20–30 days after the flowering. The main loss is due to unfertilised flowers and young fruits with poor fertilisation. Detached flowers and young fruits are observed or begin to degenerate before double fertilisation in the embryo sac; in some cases, only the secondary nucleus is fertilised in the embryo sac, and the egg cells are not fertilised, resulting in the absence of pre-embryo formation. Also, although egg cells are fertilised, the zygote has been degraded before division, or the zygote is dysfunctional, forming an abnormal original embryo (Xu et al. 1995). There are two stages in this period: in the early stage, the ovary enlargement is not significant, and the fruit stalk is detached; in the later stage, the slightly expanded young fruit drops without the fruit stalk. Lack of pollination conditions or incomplete fertilisation and imperfect embryo development are the main causes of a physiological decline in this period. In addition, in this period, before and after leaf expansion, if the storage of nutrients is insufficient or the leaves are spread shortly after flowering, the young fruit will not obtain enough nutrients or will be harmed by low temperatures, which may also lead to fruit drop (Chu 1999).

3.2.5.3   Third Fruit Drop

This occurs 50–60 days after full bloom, which is fruit development stage II, also known as the hard-core stage fruit drop. The endosperm has been in the free nucleate stage within 35 days after fertilisation. After 40 days (about late April), the free nucleus of the endosperm is transformed into endosperm cells, and the endocarp no longer increases its volume. Secondary thickening of the cell wall, this time with fruit shedding, or the inner bead is prematurely disintegrated, or the suspensor and endosperm degenerate prematurely. When fruit development stops for 4–6 days, the layer is separated from the base of the fruit. Finally, the fruit stops growing for 12–14 days. In this period, the fruit separates from its base, and only the fruit falls off. At this time, fruit development is highly sensitive to external environmental conditions (Chu 1999) such as unfavourable weather conditions or unstable soil moisture. Excessive fertilisation results in the growth of new shoots, aggravating fruit drop. In addition, dry and hot winds will cause sudden fruit drop.

Three types of early fruit drop of *P. mume* are related to nutrient levels. The contents of soluble sugars, total sugars and boron in the flowers are closely related to the first fruit drop of the plum tree, while the concentrations of soluble sugars, starch, total sugars and nitrogen in the fruit significantly impact the second fruit drop. Starch

and total sugar consumption of the leaves before and after fruit drop is related to the third fruit drop. The *P. mume* tree enters the stamen and pistil structure around the deciduous, and *P. mume* tree is the first flowering plant. The nutrients required for the late pistil development, flowering and young fruit generation are mainly obtained from the nutrient stocks of the tree, suggesting that the tree body (Chu 1999) highly depends on nutrients. Therefore, strengthening the trees in autumn and winter to prevent abnormal leaves, extend the functional period of photosynthesis and improve nutrient storage is important for improving the fruit setting rate of *P. mume* trees, especially reducing first and second fruit drops.

In the pre-harvest fruit drop (late fruit drop) of some varieties, fruits fall from the greening period to the ripening period. Therefore, most of the *P. mume* fruits are harvested before or during the greening period.

3.3 Phenological Period

The phenological period of *P. mume* varies with region and variety. Its main feature is early flowering, first flowering and then leafing. The fruit growth and development periods are short, and leave drop is early (Zhuang et al. 2015).

3.3.1 Flowering Period

The flowering period of *P. mume* trees is relatively early. Flower buds in Jiangsu, Zhejiang and Anhui generally begin to expand in mid-December, with an early bloom in mid-February and a full bloom in mid-March; blooming generally finishes in late March. In Guangdong, Guangxi, Fujian, Taiwan, Yunnan and Guizhou, due to the higher temperatures in winter and spring, the flowering period is earlier, generally from December to January. In some areas of Guangxi, due to delayed flowering, flowers and leaves often appear at the same time, and flowers even occur after the first leaves,

causing serious flowering and fruit losses and reduced production. The flowering of *P. mume* trees is closely related to the pre-flowering temperatures. In Nanjing, from the beginning of February, when the effective accumulated temperature higher than 5 °C reaches 47.6 °C, the trees enter the early stage of flowering, and it was considered that the calculation method was based on the flower (Zhuang et al. 2015, 2016). This method of calculation is more stable and reliable than the calculation based on the accumulated temperature of 50–60 days before the flower, which is greater than zero.

3.3.2 Leaf-Opening Period

The leaf buds differentiate into two anterior leaves in early April, forming 14–15 scales from April to October and differentiating into five to six inner bract leaves from late October to November. From late November to April, about 11–12 leaf primordia have the fastest differentiation rate between February and April. Leaf bud formation takes a full year. It is believed that the differentiation of leaf bud scales requires a warm environment, while the differentiation of leaf primordia requires a cold condition (Jiang et al. 2014). Breaking the dormancy of leaf buds requires at least a temperature below 5 °C over more than 1 month.

In Nanjing, Jiangsu Province, after the flowering stage, the leaf buds begin to sprout at the end of March and in early April, and the new shoots begin to elongate relatively slowly. In mid-April, the leaves enter the rapid growth period. At the beginning of May, the fruits enter the hard-core stage, and the new shoots reach the highest peak. At the end of the hard-core period in mid-May, the growth rate of the new shoots gradually decreases and the shoots enter the slow growth stage. The top buds begin to self-prune and stop growing. New shoots of different lengths stop growing as the top buds self-prune and may continue until late June or later. The young tree can be sucked 2–3 times a year, and adult trees generally no longer produce secondary shoots.

The growth of *P. mume* shoots in Shanghang, Fujian Province is earlier than that in southern Jiangsu. In general, leaf buds germinate from the end of January to the beginning of February. From late February to mid-March, the new shoots grow rapidly and tend to stop growing in late April.

3.3.3 Abscission and Dormancy Period

In Guangdong, abscission generally occurs in September. If the leaves fall too late, the number of flowers will be reduced in the following year. The normal abscission period of *P. mume* in Jiangsu and Zhejiang is November. After defoliation, the trees enter the winter dormancy period, starting with a natural dormancy. After a certain low-temperature effect, the flower buds begin to expand in mid-December, ending natural dormancy and entering forced dormancy. If the temperature is high in this period, the forced dormancy period will be short.

After entering the winter dormancy period, the physiological activity of *P. mume* trees is very low, but a series of physiological and biochemical changes still occur. At the beginning of natural dormancy, the abscisic acid contents of the short branches and the flower buds are high, while the content of gibberellic acid is low. By the end of the natural dormancy, the abscisic acid content is minimised, and the gibberellin acid content is highest. Soluble protein, total nucleic acid and ribonucleic acid concentrations decrease rapidly after entering natural dormancy and increase rapidly near the end of the natural dormancy. The relative changes in endogenous hormones and the synchronous changes of protein and genetic material indicate that during the natural dormancy of *P. mume*, inhibition of genetic material is relieved due to the decrease in the abscisic acid content and the increase in the gibberellin content. The synthesis of ribonucleic acid enhances protein synthesis, thus ending the natural dormancy of *P. mume* trees. At the same time, during the entire dormancy process, the contents of soluble and reducing sugars in the short branches and the flower buds increase, while the content of starch decreases continuously during natural dormancy, and starch is gradually hydrolysed into soluble sugars for energy, facilitating germination in early spring (Wen et al. 2017).

3.4 Requirements in Terms of External Environmental Conditions

3.4.1 Temperature

P. mume is native to southern China and requires a warm climate. Annual average temperatures in the main fruit-producing areas in China are 13–23 °C, facilitating the economic cultivation of *P. mume*.

Winter temperature is a limiting factor for *P. mume* distribution. The roots of *P. mume* roots are weaker than those of apricots, and freeze damage occurs at −8 to −5 °C. When the root tip is subjected to temperatures below −5 °C, it turns from white to brown or water-stained. The cortex is kept separate from the xylem. The thicker roots are located in one. Freezing damage can occur at −8 °C, irrespective of the variety. The cold resistance of the roots of *P. mume* and apricot hybrids is biased towards apricot, with up to −13 °C, and the root tip can be resistant to −8 °C. *P. mume* branches can withstand winter temperatures of −25 °C, and the medium *P. mume* varieties are stronger than the pure *P. mume*. At winter temperatures below −25 °C, *P. mume* cannot safely overwinter, which limits the distribution of *P. mume* to the northern parts of the country (Deng 1998).

The flowering temperature is extremely important in the economic cultivation of *P. mume*. The tolerance of different varieties of *P. mume* blossoms to low temperatures has been altered. The cold resistance of the flowers decreases with the phenological period, and flower buds are stronger than the open flowers. The cold resistance of pure *P. mume* flowers is stronger than that of apricot flowers. Early-flowering varieties have a higher cold

resistance than late-flowering varieties. The minimum temperature threshold of the young fruit is −3 to −2 °C.

Flowering temperature also affects the germination of the pollen and the activity of vector insects, thereby affecting fertilisation of *P. mume*. The pollen of *P. mume* does not germinate at 0 °C, and the germination rate increases linearly with increasing temperatures in the range of 0–20 °C. There is no significant difference between 20 and 25 °C in terms of germination rate.

The activity of vector insects starts at 10 °C and increases with increasing temperature. A low-temperature resistant insect, *Syrphus confrater* Wiedemann, has been found in Lishui, Jiangsu Province. It has been observed at 4 °C, is very active at 7.5–8 °C, and is most active at 9.5–10 °C. It plays a potential role in pollination at low temperatures below 10 °C, supplementing pollination by bees. Therefore, because the *P. mume* flower shows a strong cold resistance, its pollen can adapt to low temperatures. Even at temperatures of 4–10 °C, it can be pollinated normally. From 5 to 25 °C, the flowering period of *P. mume* increases with increasing temperatures; 5–15 °C is the temperature range affecting the flowering stage. Coldness is conducive to ovary development, which warmth is conducive to pollen germination and elongation.

Although during the *P. mume* flowering period, low temperatures are unfavourable, extremely high temperatures are also not suitable. In the Puning area of Guangdong, *P. mume* flowers around January. If the temperature is around 15 °C, the flowering is good, resulting in a high fruit setting rate. On the contrary, if it is too warm in December and January, flowering is suboptimal. Temperatures above 20 °C in December are not conducive to *P. mume* blossoming.

The differentiation of *P. mume* flower and leaf buds also requires a suitable temperature. At the initial stage of morphological differentiation of *P. mume* buds, treatment with a temperature of about 15 °C can promote flower bud differentiation and increase flowering rates. Higher temperatures can accelerate the early process of flower bud differentiation, which affects the late differentiation process of flower buds; at least a

low-temperature environment below 15 °C is required for the formation of ovules and pollen. Lower temperatures (15–20 °C) are conducive to the differentiation of leaf primordia.

3.4.2   Precipitation

P. mume is a fruit tree of the summer wet zone and requires more water during its growth. In the main producing areas of China, the annual average precipitation is about 600–2200 mm, which is suitable for the normal growth of *P. mume*.

The main meteorological factors, such as temperature, do not affect fruit setting in this area, but the amount of rain during the flowering period is crucial. It not only affects the activity of pollinators, but also the cracking of the anthers and pollen spread. Although *P. mume* can adapt to different moisture regimes, a dry wind during the flowering season will affect fertilisation due to desiccation of the stigma. Excessively humidity changes after flowering will also increase fruit drop (Jiang et al. 2014).

In some years, in southern Jiangsu, temperatures are extremely high in the middle of May. At temperatures above 28 °C for more than 3 days, the fruits drop suddenly, seriously affecting yields. This sudden fruit drop is not caused by a lack of soil water, but rather by atmospheric drought, water loss during transpiration and a low fruit water level; this physiological drought causes the endosperm to shrink and disintegrate in the seed, resulting in significant fruit drop. This period is important for fruit quality, and the fruits are highly sensitive to environmental conditions.

The *P. mume* roots are relatively shallow, and drought resistance is therefore weak. The leaves of *P. mume* trees in Jiangsu, Zhejiang and Guangdong are generally curled in autumn droughts and might even fall in severe cases, affecting flower bud differentiation, nutrient storage, flowering and fruit set in the coming year.

Comparing the differences in anatomical structure and water status among *P. mume*, apricot and peach, the drought resistance of

P. mume is considerably lower than that of apricot and peach, mainly because the structures for water transport and consumption are inconsistent. Porosity adjustment ability, water holding capacity and drought resistance are inferior. As a result, under water deficiency, turbidity is lost and free abscisic acid is produced in large quantities, leading to upward curling of the leaves and even leaf shedding, with differences among varieties and cultivation conditions. The *P. mume* leaves are curled up because the epidermal cells on the leaves are extra-large and the leaves are thick. At water deficit conditions, the upper epidermal cells lose more water and shrink in volume, so they appear to curl upward rather than to wilt. This behaviour represents an adaptation to adverse environmental conditions.

The roots of *P. mume* trees are sensitive to water and have well aerobatic. Root respiration is higher than in other fruit trees. Poor resistance to sputum, avoiding water wet, the soil moisture content on the 5th to 7th day after rain is 30–40%, with optimum root growth, and there is a lot of deciduous or root rot to death, especially the peach anvil *P. mume* tree.

3.4.3 Light

P. mume has a high light requirement. The light compensation point and saturation point of photosynthesis of mature leaves of *P. mume* are 2000–3000 lx and 35,000–40,000 lx, respectively (Jiang et al. 2014). When the light intensity exceeds 30,000 lx, the photosynthetic speed slowly increases. During long sunny days, the natural light intensity is greater, and the light saturation point can reach 6–7 h. The photosynthetic production capacity of *P. mume* leaves is directly restricted through light intensity and affected by high temperatures and high relative air humidity. According to light distribution of the natural leaf curtain of *P. mume* trees, the thickness of the leaf curtain layer of *P. mume* tree to maintain high photosynthetic production efficiency that is more than 1.5 m, and the effective leaf area coefficient is about 5, which is

beneficial for improving the photosynthetic capacity of the whole tree.

There is a strong relationship between light energy distribution in the crown of *P. mume* trees and tree growth. Daily average relative light and germination rate of *P. mume* branches, the ratio of thick and short branches, flower bud rate, dry weight of flower buds and the branch group are closely related. There is also a significant positive correlation between fruit setting rate and fruit vitamin C content. Under favourite light conditions, the germination rate of the branch group is high, the hair branch is compact and thick, and a good fruiting branch group is formed, which reduces the belt of the inner sac. Improving the light conditions of the canopy can increase the quality and quantity of the flower bud differentiation of *P. mume* trees, thereby enhancing fruiting ability and improving fruit quality. When the relative light intensity is 55–75%, the *P. mume* tree has the strongest result, less than 25%, and the fruit setting rate drops significantly. At a light intensity below 10%, the normal firming ability is almost lost. Because *P. mume* fruits are harvested earlier, the sugar and acid contents of the pulp are not impacted by light conditions.

The length of light period does not affect bud differentiation, but at low temperatures and short sunshine durations, the number of leaf primordia can increase. When there is less rain during the flowering period, with sufficient sunshine and rapidly increasing temperatures, pollination and fertilisation are facilitated, resulting in increased fruit setting rates. The annual sunshine hours are positively related to the yield.

3.4.4 Soil

P. mume prefers slightly acidic soil at a pH around 6. When the pH is less than 4.5 or greater than 7.5, the trees show poor growth and even die (Jiang et al. 2014). The Japanese *P. mume* variety Bai Jiahe, growing in the calcareous purple soil of the Sichuan Basin (pH 7.0), is obviously unsuitable to these conditions and

shows iron deficiency mosaic. Therefore, in Beijing, Yan'an, Shanxi, Inner Mongolia and other places, *P. mume* is planted on apricot rootstocks and mountain peach rootstocks, which can withstand alkaline soil with a pH of 8.6.

Deep soil layers, low groundwater levels (below 1 m), well-drained clay loam and sandy loam are the most suitable conditions for *P. mume* growth. If the soil layer is shallow or the soil is too sandy, *P. mume* is susceptible to drought; heavy clay will cause long leaves, low yield, poor fruit quality as well as pest and disease infestation.

3.4.5 Other Environmental Conditions

Most Chinese gardens containing *P. mume* have built-in hilly areas, facilitation drainage during the rainy season. However, attention should be paid to water diversion irrigation and to the choice of slope direction and location. In areas where winter and spring are prone to frost, the choice of the southern or eastern slopes can avoid direct attacks from the northwestern cold current. In areas where the flowering temperature is too low and the change is fierce, soil and air temperatures during spring on the northward slopes are low, the water conditions are good, and *P. mume* growth is better than on the sunny slopes. In high mountainous areas, mountain tops are vulnerable to wind damage; therefore, when establishing a garden, it is recommended to choose the mid-section of the hillside.

Elevation is an important factor in regard to temperature and humidity. The optimum elevation for the growth of Yunnan *P. mume* is about 1600–2400 m, and profitable cultivation of fruitless *P. mume* can be performed in areas more than 3000 m above sea level and below 1100 m. Taiwan's *P. mume* gardens are mostly concentrated on slopes at 300–1000 m above sea level, on the western side of the Central Mountain Range. This is the region where Taiwan has the least variation from winter to early summer. Wind is also an important ecological factor, with a significant impact on *P. mume* trees. During the

flowering or fruiting period, when subjected to strong winds, pollinating insects are inactive, resulting in considerable flower drop. Therefore, when building a garden, wind damage should be prevented, and adequate measures should be taken, such as the creation of shelter forests.

P. mume is extremely sensitive to atmospheric pollution; especially high sulphur dioxide and fluorine contents; fluorine in the air directly threatens the normal growth of *P. mume*. When the atmospheric fluorine content reaches 1–1.2 μg per cm^3 per day, the cellulose activity of *P. mume* leaves is significantly increased, leading to early defoliation. Different varieties show different tolerances to fluoride. Among the numerous varieties in Zhejiang, soft-red *P. mume* is the strongest one. At the same time, *P. mume* trees are also extremely sensitive to pesticides such as dimethoate. Even low concentrations of dimethoate can cause significant damage and even result in the death of the tree.

References

Cai B, Zhuang W, Gao Z, Zhang Z (2014) Research advances on the mechanism of gibberellins releasing seasonal dormancy in woody plants. Acta Bot Bor Occi Sin 34(10):2145–2152. https://doi.org/10.7606/j.issn.1000-4025.2014.10.2145 (in Chinese)

Chu M (1999) Chinese fruit tree: *Prunus mume*. China Forestry Publishing House, Beijing (in Chinese)

Chu M, Huang J (1995) Study on flower bud differentiation and metabolism of *Prunus mume*. J Beijing For Univ S1:68–74 (in Chinese)

Deng L (1998) Effect of temperature on flowering and growth traits of *Prunus mume*. J Jiangsu For Sci Technol 25(Supple):26–31 (in Chinese)

Gao Z, Shi T, Luo X, Zhang Z, Zhuang W, Wang L (2012) High-throughput sequencing of small RNAs and analysis of differentially expressed microRNAs associated with pistil development in Japanese apricot. BMC Genom 13(1):371. https://doi.org/10.1186/1471-2164-13-371

Guo Z, Wujing Z (1995) Effect of temperature and day length on flower bud formation in *Prunus mume*. J Beijing For Univ S1:62–67 (in Chinese)

Jiang B, Fang Z, Fang J, He Y, Zheng B (2014) Recent advances of the flowering and its regulation for plum (*Prunus mume*). Bull Sci Technol 30(3):43–47. https://doi.org/10.13774/j.cnki.kjtb.2014.03.010

Lin Y, Yang Y, Yang X, Lu S, Xia Q (2014) Effects of ripe stage on nutritional composition and flavor of

green mume juice. Acta Agric Zhejiang 26(4):1049–1054

Miao S (2014) Exporing mei of Sui dynasty in Guoqing temple. Ancient Famouse Trees 11(3):32. https://doi.org/10.1152/ajplegacy.1975.229.3.570 (in Chinese)

Shi T, Gao Z, Zhang Z, Zhuang W (2011a) Comparison of biological traits of flowers and the rate of pollen germination among 47 *Prunus mume* cultivars. Chin Agric Sci Bull 27(04):227–232

Shi T, Zhang Q, Gao Z, Zhang Z, Zhuang W (2011b) Analyses on pistil differentiation process and related biochemical indexes of two cultivars of *Prunus mume*. J Plant Res Environ 20(4):3–9 (in Chinese)

Shi T, Gao Z, Wang L, Zhang Z, Zhuang W, Sun H et al (2012a) Identification of differentially-expressed genes associated with pistil abortion in Japanese apricot by genome-wide transcriptional analysis. PLoS One 7(10):e47810. https://doi.org/10.1371/journal.pone.0047810

Shi T, Zhuang W, Zhang Z, Sun H, Wang L, Gao Z (2012b) Comparative proteomic analysis of pistil abortion in Japanese apricot (*Prunus mume* Sieb. et Zucc). J Plant Physiol 169(13):1301–1310. https://doi.org/10.1016/j.jplph.2012.05.009

Wen L, Gao Z, Huo X, Zhuang W, Lv L (2017) Signal effect of H$_2$O$_2$ on seasonal dormancy release induced by exogenous GA4 in Japanese apricots. Acta Hortic Sin 44(2):255–267. https://doi.org/10.1038/hortres.2017.6

Xu H, Hu J, Wang D, Huang Q (1992) Developmental anatomy of Japanese apricot bud formation. J Beijing For Univ 14(4):18–22 (in Chinese)

Xu H, Wang Q, Hu J, Huang Q (1995) Studies on the development of pistil and fertilization in *Prunus mume*. Acta Bot Yunna 17(1):61–66 (in Chinese)

Zhong W, Gao Z, Zhuang W, Shi T, Zhang Z, Ni Z (2013) Genome-wide expression profiles of seasonal bud dormancy at four critical stages in Japanese apricot. Plant Mol Biol 83(3):247–264. https://doi.org/10.1007/s11103-013-0086-4

Zhuang W, Gao Z, Wang L, Zhong W, Ni Z, Zhang Z (2013) Comparative proteomic and transcriptomic approaches to address the active role of GA4 in Japanese apricot flower bud dormancy release. J Exp Bot 64(16):4953–4966. https://doi.org/10.1093/jxb/ert284

Zhuang W, Gao Z, Wen L, Huo X, Cai B, Zhang Z (2015) Metabolic changes upon flower bud break in Japanese apricot are enhanced by exogenous GA4. Hortic Res 2:15046. https://doi.org/10.1038/hortres.2015.46

Zhuang WB, Cai BH, Gao ZH, Zhang Z (2016) Determination of chilling and heat requirements of 69 Japanese apricot cultivars. Eur J Agron 74:68–74. https://doi.org/10.1016/j.eja.2015.10.006

Taxonomy and Germplasm of *Prunus mume*

4

Zhihong Gao and Ting Shi

Abstract

In this chapter, the standard and basis for *Prunus mume* taxonomy are being discussed. According to its morphological characteristics and distribution, *P. mume* can be classified into two ecological categories: high-elevation and low-elevation types. Within *P. mume* species, there are seven varieties and one variant.

4.1 Overview, Criteria and Basis of Classification in *Prunus Mume*

4.1.1 Overview of Classification

Since the denomination of *Prunus mume* in 1835, more than 45 varieties have been generated (Bao and Chen 1992). Although *P. mume* is native to China, it has mainly been studied by foreign scholars. In fact, some belonged to the problem of variation range; some to the collection site and seasons; some to natural variation; some to cultivation variation; and some belonged to

theoretical basis and some are part of the sampling range. The original variety was defined as a variant, which is particularly evident in the *P. mume* species. The special plants cultivated by nature are different from the classification and basis of naturally growing plants (Chen and Bao 1992a). The classification of cultivated species is not based on naturally occurring variations, but on the significance of each species for humans. The researchers classifying *P. mume* are plant taxonomists, while some are horticulturists and therefore have a different focus, resulting in inconsistencies (Bao and Chen 1992; Chen and Chen 2009).

Over the years, with multidisciplinary studies on *P. mume*, we have developed a deep understanding of the nature of the species as well as its varieties and variants (Bao and Chen 1994).

4.1.2 Criteria and Basis of Classification

There are two major concepts for species, namely taxonomic and biological concepts.

Taxonomic or morphological species are mainly morphological indicators, considering the geographical distribution of the species, which is the morphological-geographic method of traditional taxonomy. Taxonomists regard a sample of wax leaf as a sampling of the natural group, and morphological discontinuity exists in variation patterns displayed by the 'group' as the basis for

Z. Gao (✉) · T. Shi
College of Horticulture, Nanjing Agricultural University, No 1 Weigang, Nanjing 210095, People's Republic of China
e-mail: gaozhihong@njau.edu.cn

classifying species. The range of the group is judged according to the geographical distribution of the sample, although modern taxonomy has become increasingly important in recent years. In many cases, taxonomic species often coincide with natural breeding populations, and in fact contain the genetic content of reproductive isolation. External morphology is also the result of physiological and biochemical processes in the development of individuals. The inherent changes in plant species, such as heredity and evolution, physiological and biochemical differences, are related. However, it is necessary to make the classification more representative of the objective nature of the species *P. mume* to clarify the specific mechanism of its evolution and variation and to break through previous levels of indirect inference of origin and evolution by morphological-geographic data (Darlington 1940).

The concept of biological species is an ideal kind of concept. With 'reproductive isolation' as the sub-species standard, there are often some difficulties in theory and practice (Sokal and Crovello 1970). Genetic variation leads to morphological differentiation, not positively correlated with changes in sterility; hybrids are common in flowering plant species. Previous studies and hybridisation experiments have shown good affinity and fertility between *P. mume* and peach and between *P. mume* and apricot (Chu and Huang 1995).

Variants have morphological variations within species, the variation is relatively stable, and the range (or region) of distribution may be much smaller. Therefore, some researchers suggest variant is a kind of 'local race'. It is very common in traditional plant taxonomy and is often the case that a specimen with a few minor important traits in the same species is defined as a variant, regardless of the state of its distribution. The variant is more geographically strong; it can be geographical, ecological, cellular or a combination of the above. Several variants can be included within the same species (Doi et al. 2010; Hayashi et al. 2008).

A variant is a morphological variation within a species but does not show a certain distribution area, rather a sporadic distribution of individuals. Variants are often considered as an occasional sporadic variant of a population, distinguished by a single or a few interrelated traits. A variant is often the result of repeated pairings of recessive genes in a population with limited incidence. Genetically determined distortions are also treated as variants (Doi et al. 2010).

Intra-group variation is an adaptive characteristic of a group, and heterozygosity within an individual implies a benefit of choice. Due to the wide distribution of *P. mume*, the history of its use or of the cultivation of semi-wild or semi-cultivated plants is relatively long, the variation pattern is more complicated, wild and cultivated plants tend to be different, and there are other kinds of interventions, making the judgment of wild and cultivated *P. mume* become as ancestral signs, species, and other features are often intertwined. To this end, in terms of the determination of the varieties and variants of *P. mume*, the original one is based on its natural origin variants, and second is the variants identified before 1959 International Cultivated Plant Nomenclature Regulations (Brickell et al. 2004), although they may be closer to the origin of cultivation, but are still treated as variants according to the regulations.

4.2 Classification of *Prunus mume*

4.2.1 *Prunus mume* Sieb. et Zucc

Small arbour tree, thin shrub, 4–10 m high. The bark is grey and longitudinally cracked; 1-year-old branches without ridges are green and glabrous. Lateral buds are compound buds or solitary buds, and buds in the middle are usually leaf buds, smaller and ovate-conical; buds ovate on both sides, larger, 2.5–3.5 mm, width 2–3 mm, dark red, no hair. Bud scale 10–13 pieces, no hair on both sides or edgy hair. The leaves are broadly ovate, 4–10 cm long, 2–5 cm wide, apex long with sharp serrated teeth, glands at the dentate end, often larger at the base glandular gland, sometimes glandular at the upper end of the petiole (under light microscope), pubescent

on both sides when young, gradual detachment when growing or only pubescent between veins; petiole length 0.8–2 cm, young hair when old (Chu 1999).

Flower diameter 2–3 cm, flowers scented, open before leaves; flowers sessile or on short stalk, length 1–3 mm, often hairless. Calyx usually purple, but some varieties of flower buds are green or green-purple; sepals glabrous or sometimes pubescent; five to six sepals, ovate or sub-orbulate, lanceolate. Sometimes two to seven (centralised skin), even absent, ovary upper position, from the lower part of the style, densely pilose, style short or slightly longer than stamens. Drupe subglobose, 1.3–4 cm in diameter, yellow or greenish yellow to greenish white with dense short hair. It has a sour taste, a stone with a near ovoid or elliptical shape, a rounded tip with a small pointed tip, a base that tapers into a wedge shape, slightly flattened on both sides, a slightly blunt abdomen and a distinct longitudinal ridge on the ventral and back ribs, and the surface has a honeycomb cavity. Flowering period from December to March, fruit-ripening period 4–7 months, rarely in August.

Chinese *P. mume* are widely distributed in northern China from Shanxi Hanzhong, Henan civil rights, south to Taishan in Guangdong, from the west of Tongmai, Tibet Bomi, east to Taiwan, all cultivated or wild-type species of *P. mume*. Guangdong, Guangxi, Taiwan, Yunnan, Fujian, Zhejiang, Jiangsu and other places are the main production areas (Chen and Bao 1992a).

The wild-type *P. mume* can be divided into two distribution patterns or types: (1) high-altitude feature types; in high-altitude areas such as Tibet, Yunnan and Sichuan, the main features are multi-branched thorns, small flower diameters and early flowering periods, such as Yu Ming Xiao *P. mume*, Wuyuan Xishan wild *P. mume* and Tibet wild *P. mume*. Tibet wild *P. mume* is distributed in the southwest slope or river valley of Tongji County, Nyingchi County, Tibet Autonomous Region, with a small fruit and numerous branches; (2) low-altitude characteristics, in the middle and low mountains of east and central China, (170–1800 m above sea level) Wild *P. mume* contains numerous thorns, flower

diameter is large, and flowering period is earlier, such as in Jingdeye *P. mume* and Huangshan wild *P. mume*. These two patterns can be regarded as the results and reflections of different climates affecting wild *P. mume* at different elevations, which have been relatively fixed; overall characteristics and distribution variation patterns of wild *P. mume* can also be seen.

There are several variations of *P. mume* in areas at high elevations; previous studies have summarised about 20 varieties and variants. Chen Junyu (1996) systematically sorted the various types of *P. mume* and divided them into seven varieties and one variant (Chen and Chen 2009).

4.2.2 Subspecies Group

The subspecies of *P. mume* includes seven varieties and one variant.

4.2.2.1 *Prunus mume* Sieb. et Zucc. Var. *pleiocar Pa* Maxim

There are two to seven carpels in a flower, most of which have three to five pieces. One flower can count several fruits and vary in size. Japan calls it the seat of the *P. mume* (or the eight-room *P. mume*), with its flowers arranged like a seat. Its main features are as follows: one flower has two to seven carpels and has several types of results: one stem two fruit, fruit top separated, middle-connected double sets of *P. mume*; two fruit parallel and separated *P. mume*; three stems *P. mume* with one stem three small fruits; Wuzi *P. mume* with one stem three to five fruits, etc. Within the variant, it is a type of fruit and ornamental (Chu, 1999).

4.2.2.2 *Prunus Mume* Sieb. et Zucc, Var. *Pallescens* Franch

Branches transformed into thorns, leaves resemble hard paper and are thicker, and flowers are smaller. This variant is the main type of wild *P. mume*.

It is produced from western Sichuan to western Yunnan, in the hillside forest or by the stream, at an elevation of 1700–3100 m. The

French missionary AJM Delavay collected specimens in Dapingzi, Dali, Yunnan. The French botanist A. Franchet named the shrub or small tree in 1889. Dry brownish to brown, more lenticels from the base, which is branching the half of the blush, half of the green, the base is at a right angle, the branches turn into thorn leaves near leathery. Leaf blade dark green, ovoid, apex caudate, margin with heavy serrations, base cuneate. Slightly sprinkled powder, corolla dish; flower diameter 26 mm, stem length 2 mm, stamens numerous, inner ring neatly erect, outer ring radial, initial flowering period in mid-January, fruit-ripening period June to July, fruit spherical, yellow-green when cooked, sour and bitter (Chen and Bao 1992a).

4.2.2.3 *Prunus mume* Sieb. et Zucc. Var. *cernua* Franch

The peduncle is very long, about 1 cm, and drooping when fruiting; leaves are ovate, apex acuminate.

Produced in the west to northwest of Yunnan, at the hillside, by streams or under the sparse forest at an elevation of 1900–2600 m. The species is also distributed in northern Vietnam and northern Laos.

The slope is westward, the branchlets are dizzy and green from one side; the dry skin is brown with more lenticels. Leaf resembles thick paper, green, elliptic, apex tip, leaf margin with teeth, base wedge, hairless, leaf 4.0–7.5 cm and width 1.8–3.4 cm. Flowers are single-petal with white colour, flowering from early January to early February. Fruit-ripening period in late June, fruit spherical, yellow-green when cooked, bitter in taste, stalk length 0.8 cm; stone honeycomb-like with small holes.

4.2.2.4 *Prunus Mume* Sieb. et Zucc. Var. *microcarpa* Makino

Leaves, flowers and fruits are smaller than in other types, the stone is small and round, branches are thornier, and the leaves are lighter; fruits are smaller.

Distributed in the semi-wild state in the cities of Wuyuan and Dali in northwestern Yunnan and in the Muli Tibetan Autonomous County of

Sichuan Province (Jianying 1997). On the edge of ditches or fields, at an elevation of 1900–2400 m, dry brown with more lenticels and branches. Leaf resembles paper, light green, ovate, long apex, margin partially serrate. Leaf base is wedge-shaped, with long rust between the veins of the leaf, and leaf length is 2.2–5.5 cm, with a width of about 1.5–3.0 cm. Flowers are white, with single petals, flower diameter is 20 mm; peduncle 1–2 mm long, stamens numerous, erect, five sepals, red-to-red-green; early flowering period in late December, full bloom period in mid-January. Fruit-ripening period from June to July, fruit shape round, significantly smaller than other types; the core is small and round, about 1.5 cm long, width about 1.3 cm; honeycomb with small holes.

4.2.2.5 *Prunus mume* Sieb. et Zucc. Var. *bungo* Makino

Suture is deep, nuclear honeycomb pores are shallow, leaves are large and the base is nearly heart-shaped. The dried branches are brown with more lenticels, the twigs are purple from the front and green from the back. Leaf near leathery, dark green, egg-shaped, tail tip, leaf margin has partial heavy serration, leaf base near heart-shaped or near-truncated shape, less hair, larger leaf, 4.2–7.7 cm long and 2.5–4.2 cm wide. Flower single petal, white colour, flowering in January; fruit-ripening period from mid-June to August, fruit round or slightly round, slightly apical at the top, yellow-green when cooked; fruit ovate oblate, honeycomb.

4.2.2.6 *Prunus mume* Sieb. et Zucc. Var. *goethartiana* Kochne

The back part of the leaf has yellow-brown hair; the petiole is densely villous, and the peduncle is extremely short and hairy. The outer part of the tube is densely villous, the bracts are hairy at the edges, and the outer part is densely villous; some of the branches are also hairy.

The branchlets of Yunnan's prunes are purple and red; branches are brown with more lenticels. Leaf resembles thick paper, dark green, ovate-lanceolate, apex, apically acutely lanceolate or subulate. The back of the leaf and the

petiole are densely rusted, which is less pro-nounced on the upper surface. The leaf is 3.4–5.7 cm long and 1.6–2.7 cm wide. Flower single petal, white in colour, flowering from December to February. Ripening period is July; fruits are red-yellow when ripe, round and spherical, the stalk is 3 mm long, the stone is oval, and the honeycomb surface is obvious.

4.2.2.7 *Prunus mume* Sieb. et Zucc. Var. *pallidus* Bao et Chen

Both sides of the leaf have a white powder with white wax. Distributed at the southwestern slopes or in the river valleys at an elevation of 2100–2300 m of Linzhi, Tongmai, Tibet. The trunk has numerous lenticels; branches are purple from the front, green underneath and with more thorns. Fruit is spherical, slightly sessile, small, suture deep, fruit tip; nuclear honeycomb with small holes, except for the leaves on both sides of the wax white, with typical *P. mume* characteristics. In the place of origin and for wild *P. mume* in a mixed state, in the distance, the *P. mume* bush appears grey, flashing between the green branches; high ornamental value, but so far, no cultivation. In 1992, it was designated as a variant.

4.2.2.8 *Prunus Mume* Sieb. et Zucc. F. *sempervirens* Bao et Chen

Leaves are thicker and darker green and gradu-ally fall at the end of the year. Distributed in the *P. mume* production area at an elevation of 1900–2400 m in Yunnan and Sichuan. Leaves are thick, dark green, ovate, apex tip, heavily serrated, 70 mm long, 40 mm wide, leaf base wedge-shaped, with less hair on both sides with more branches. Flower diameter 27 mm, five white petals, seven purple sepals. Early flowering period is in mid- or late December, the fruit-ripening period is June; fruits are large, and the stone has a small hole in the armpit. This type requires a warm climate and should be called 'semi-evergreen' *P. mume*. In its local environ-ment, evergreen *P. mume* accounts for a certain proportion and exists in the entire *P. mume* population. In 1992, it was designated as a new variant of *P. mume* (Bao and Chen 1994).

4.3 *Prunus mume* Varieties and Excellent Strains

P. mume is native to China and has abundant resources. There are numerous local cultivars, but it has not been comprehensively investigated. Since the 1980s, the production of *P. mume* fruit has been developing rapidly. At the same time, breeding and introduction programs have been carried out, and many new varieties and excellent singles have been obtained. The strain (hereafter referred to as 'the excellent strain') has achieved great results and enriched the resources of fruit and *P. mume* varieties (Chen and Bao 1992b).

The classification of *P. mume* varieties is based on the application of production practice and its adaptation, which is based on the classi-fication of the fruit colour of Zeng Mian. We have added some traits related to this kind and have divided the varieties of *P. mume*. It is a major category of white, red and green *P. mume* (Jiang et al. 2017; Chen and Bao 1992b).

White *P. mume*: The fruit is generally small- or medium-sized, nearly round, unripe and ripe fruits are is light green to yellow-white; 1-year-old branches are pale green, young shoots and leaves are yellow-green and the leaves are smaller. The flesh is thinner, slightly thicker and bitter, with a large core. Yield is generally low; maturity period is the earliest one. In Fujian, fruits are harvested in the greening period around mid-April, while in the Hangzhou area, harvest is around May 20; fruits are gen-erally not suitable for raw consumption and can be dried or used as rootstock.

Green *P. mume*: Large fruit, almost round, unripe and ripe fruit green or dark green to yel-low, occasional red halo in the sun, ripening different from white *P. mume*. Young shoots and leaves are mostly yellow-green to dark green or green. The flesh is thick, crisp and juicy and has no bitter taste. The stone is nearly circular and less obvious, the nuclear point is thick, deep and less, and the stone surface is uneven, with a high amount of edible fruit. High and stable yield ability between red and white *P. mume*. Suitable for green processed products such as sugar green *P. mume*, green *P. mume* wine and *P. mume*

sauce. A few varieties contain low acid levels and can be eaten fresh.

Red *P. mume*: Fruit is large, multi-elliptic or peach-shaped. The young fruit is red, the unripe and ripe fruit is yellow-green, the sunny surface is purple-red or red, and this colour accounts for 30% when ripe, sometimes even for 70% or for the whole fruit. The 1-year-old branches are light green, the sunny side is light red or purple, young shoots and leaves are red or purple to light green. The flesh is fine, with a bitter taste and less juice, and the quality is excellent. The stone is relatively flat, the top is slightly narrow, yellow-brown, with many nuclear points, shallow and small, and the stone surface is relatively flat. It is the most productive and stable type and suitable for making *P. mume*, Chen Pi *P. mume*, Wu *P. mume*, among others.

References

Bao M, Chen J (1992) Research status and prospects of *Prunus mume (In Chinese)*. J Beijing For Univ S4:74–82

Bao M, Chen J (1994) Studies on the variation and distribution of Chinese Mei. Acta Hort Sin 21(1):81–86

Brickell CD, Baum B, Hetterscheid W, Leslie A, McNeill J, Trehane P et al (2004) International code of nomenclature for cultivated plants: introductory pages. Acta Hortic 647:7–20. https://doi.org/10.17660/ActaHortic.2004.647.1

Chen J, Bao M (1992a) Botanical classification and horticultural classification of Chinese mei (*Prunus mume*) resources. J Zhejiang For Coll 2:2

Chen J, Bao M (1992b) Studies on the classification of species (varieties, hybrids) and cultivars of Chinese mei (*Prunus mume* Sieb. et Zucc.). J Beijing For Univ S4:1–6

Chen J (1996) Chinese mei flowers (In Chinese). Hainan Publishing House, Haikou

Chen J, Chen R (2009) A new system for classifying China reference to developing superiorities groups Mei cultivar groups, with special of interspecific hybrid originated. Acta Hort Sin 36(5):693–700

Chu M (1999) Chinese fruit tree: Prunus mume (In Chinese). China Forestry Publishing House, Beijing

Chu M, Huang J (1995) Study on flower bud differentiation and metabolism of *Prunus mume*. J Beijing For Univ S1:68–74 (In Chinese)

Darlington CD (1940) Taxonomic species and genetic systems. New Syst 137-160

Doi H, Takahashi M, Katano I (2010) Genetic diversity increases regional variation in phenological dates in response to climate change. Global Change Biol 16(1):373–379. https://doi.org/10.1111/j.1365-2486.2009.01993.x

Hayashi K, Shimazu K, Yaegaki H, Yamaguchi M, Iketani H, Yamamoto T (2008) Genetic diversity in fruiting and flower-ornamental Japanese apricot (*Prunus mume*) germplasms assessed by SSR markers. Breed Sci 58(4):401–410

Jiang C, Ye X, Fang Z, Zhou D, Pan S (2017) Research progress on Japanese apricot (*Prunus mume*) in China (In Chinese). Southeast Gardening 5(5):26–31

Jianying ZXG (1997) Study on ecotypes of Prunus mume sieb. et zucc. in sichuan province. J Sichuan Agric Univ 2:148–150

Sokal RR, Crovello TJ (1970) The biological species concept: a critical evaluation. Am Nat 104(936):127–153

The Genome of *Prunus mume*

5

Qixiang Zhang and Lidan Sun

Abstract

The genome of *Prunus mume* (*mei*), which was domesticated in China more than 3000 years ago as an important fruit and ornamental plant, was one of the first sequenced genomes among the *Prunus* sub-families of Rosaceae. In this study, the 280 M genome was assembled into scaffolds by combining 101-fold NGS data and optical mapping data; 83.9% of these scaffolds were further anchored to eight chromosomes in a genetic map constructed by restriction site-associated DNA (RAD) sequencing. Combining the *P. mume* genome data with other available genome data, we reconstructed nine ancestral chromosomes of the Rosaceae family, depicting chromosome fusion, fission and duplication history in the three major Rosaceae subfamilies. We sequenced the transcriptome of various tissues and performed a genome-wide analysis to reveal the
characteristics of *P. mume*, including its regulation of early blooming in endodormancy and biosynthesis of flower scent. The *P. mume* genome sequence adds to our understanding of Rosaceae evolution and provides important data for the improvement of fruit trees.

5.1 Plant Material

Wild samples of *P. mume* were sequenced by the Illumina Genome Analyzer II in Tongmai, Bomi County, Tibet, China, which is the western-end region of the origin of domesticated *P. mume* (Fig. 5.1).

The genetic maps that were used to develop the integrated map for anchoring the scaffolds were derived from F_1 populations, totalling 260 individuals from the cross between 'Fenban' and 'Kouzi Yudie' from Qingdao Meiyuan.

5.2 Methods

5.2.1 DNA Preparation and Whole-Genome Shotgun Sequencing

We used a whole-genome shotgun sequencing strategy with the Illumina Genome Analyzer. Total DNA was extracted from fresh young leaves of wild specimens from Tongmai Town, Tibet,

Q. Zhang (✉) · L. Sun
Beijing Key Laboratory of Ornamental Plants Germplasm Innovation & Molecular Breeding, National Engineering Research Center for Floriculture, Beijing Laboratory of Urban and Rural Ecological Environment, Engineering Research Center of Landscape Environment of Ministry of Education, Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants of Ministry of Education, School of Landscape Architecture, Beijing Forestry University, Beijing 100083, China
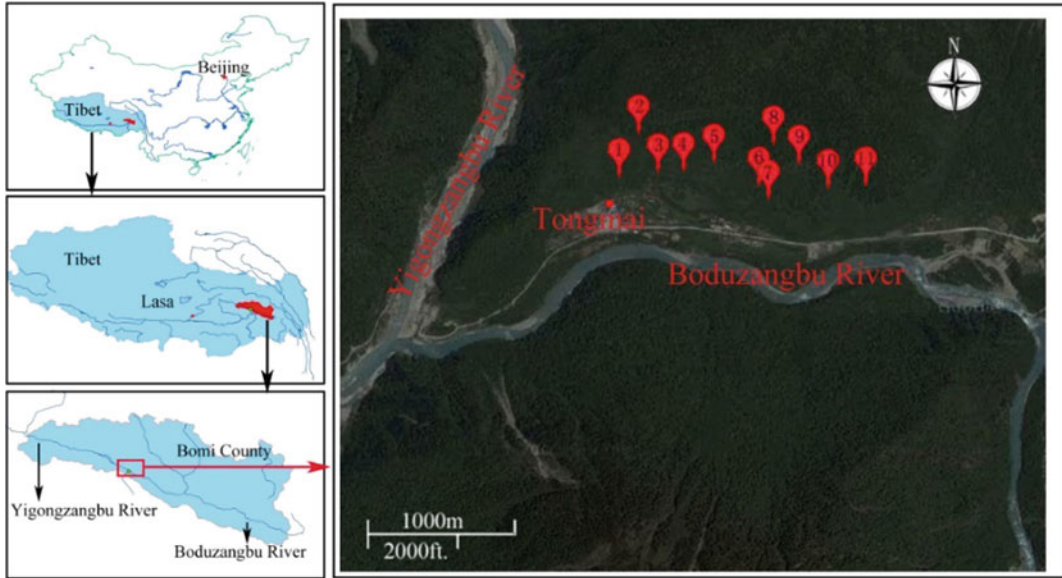e-mail: zqx@bjfu.edu.cn

Fig. 5.1 Sample distribution map of GPS for *P. mume* in Tongmai town, Tibet. Number 4 represents the location of the sample used for *P. mume* sequencing

China, using the plant genomic DNA extraction kit (TIANGEN, Beijing, China) in accordance with the manufacturer's instructions. To obtain sufficient DNA to construct and sequence the library, we performed whole-genome amplification. We constructed eight paired-end sequencing libraries with insert sizes of approximately 180 base pairs (bp), 500 bp, 800 bp, 2 kb, 5 kb, 10 kb, 20 kb and 40 kb. In total, we generated 50.4-Gb data of paired ends, ranging from 50 to 100 bp. These data were checked and filtered on the reads that were generated previously (Li, et al. 2010). We filtered the low-quality reads using the following filtering criteria: reads had an '*N*' over 10% of their length; reads contained more than 40-bp low-quality bases (quality score $\leq$ 5); reads contained more than 10-bp adapter sequences (allowing $\leq$ 2-bp mismatches); small insert size paired-end reads that were overlapped ($\geq$ 10 bp); Read1 and Read2 of paired-end reads were completely identical, which means that these paired-end reads were determined to be artefacts of the PCR experiment. Using stringent criteria, 28.4-Gb high-quality data were used for de novo genome assembly.

5.2.2 K-mer Analysis

We determined the relationship between sequencing depth and the copy number of a certain K-mer (refers to a sequence with *K* base pairs e.g. 17-mer), and when ignoring the sequence error rate, heterozygosis rate and repeat rate of the genome, the K-mer of the distribution should follow the Poisson theoretical distribution. The size of the genome was estimated using the total length of the sequence reads divided by the sequencing depth; the peak value of the frequency curve represents the overall sequencing depth. We estimated the genome size as ($N \times ($ $L - K + 1) - B)/D = G$, where *N* is the total number of sequence reads, *L* is the average length of sequence reads and *K* is the K-mer length, defined as 17 bp. Here, *B* is the total number of low-frequency (frequency $\leq$ 1 in this analysis) K-mers, while *G* is the genome size and *D* is the overall depth, estimated from the K-mer distribution. It must be pointed out that as the K-mer of the distribution should approximate the Poisson distribution, not all low-frequency

K-mers will be errors. This might lead to an underestimate of the genome size, especially at low sequencing depths.

5.2.3 Genome Assembly

We performed a whole-genome assembly using SOAP*denovo* (Version 1.05) (Li et al. 2010) with high-quality reads. Reads were loaded into the computer memory, and de Bruijn graph data structure was used to represent the overlap among the reads. The graph was simplified by removing erroneous connections and solving tiny repeats by read path. On the simplified graph, we broke the connections at repeat boundaries and output the unambiguous sequence fragments as contigs. Before generating scaffolds, in the 'map' step, SOAP*denovo* realigned all usable reads to the contig sequences and obtained aligned paired ends (PEs). Subsequently, the software calculated the number of shared PE relationships between each pair of contigs, weighed the rate of consistent and conflicting PEs and constructed the scaffolds step by step, from short-insert to long-insert PEs. To close the gaps inside the constructed scaffolds, which were composed primarily of repeats that were masked before scaffold construction, we used the PE information to retrieve the read pairs that had one end mapped onto the unique contig and the other located in the gap region and performed local assembly using GapCloser (Version 1.12) with these collected reads.

5.2.4 Estimation of Heterozygosity Rate

The heterozygosity rate was calculated by calling the heterozygous SNPs. All high-quality reads were mapped onto the genome assembly using the software SOAP2 (http://soap.genomics.org. cn/soapaligner.html) with a cut-off of less than five mismatches. Subsequently, the alignment results were analysed for SNP mining using the SOAPsnp (http://soap.genomics.org.cn/soapsnp. html). The sites that met the following criteria

were searched and termed criterion-effective sites: (a) quality score of consensus genotype in the SNP mining result is greater than 20; (b) counts of all the mapped best and second best base are supported by at least four unique reads; (c) sequencing depth is more than 10X; (d) SNPs are at least 5 bp away from each other, with an additional requirement to the criterion-effective sites that the number of reads-supported best base is smaller than four times the number of reads-supported second best base (reads-supported best base/reads-supported second best base < 4) were identified as heterozygosis sites. Finally, the heterozygosity rate was estimated as the number of heterozygosis sites divided by the number of criterion-effective sites.

5.2.5 Whole-Genome Mapping

High-quality DNA (high molecular weight, >200 kb), specific for whole-genome mapping, was prepared from fresh *P. mume* leaves. Whole-genome shotgun single-molecule restriction maps were generated with the automated Argus system (OpGen Inc., Maryland, USA). First, DNA molecules were deposited onto silane-derivatised glass surfaces in Map-Cards (OpGen Inc., MD, USA) and digested with *Nhe* I or *Bam*H I for 20 min. Then, the DNA was stained with JOJO fluorescent dye (Invitrogen, CA, USA) and imaged with the Argus system. Overall, 243,174 single-molecule restriction maps (SMRMs) (>250 kb) with an average size of 344 kb were generated. Total size was approximately 83.6 Gb.

When using whole-genome mapping data to extend scaffolds, the original scaffold sequences were first digested in silico to generate corresponding restriction maps for each scaffold. Then, the in silico restriction maps were used as seeds to identify single molecules by length-based alignment using the Genome-Builder[TM] software package (http://www.opgen.com/ products-and-services/softwares/genome-builder). These single-molecule maps were assembled together to extend the scaffolds with consensus restriction maps. Meanwhile, the scaffolds with

low coverage regions at both ends were trimmed, and only the high-quality extensions remained.

To extend sufficient scaffolds, this alignment-assembly process was iterated four times. All pairwise alignments were considered initial candidates for scaffold connection. The alignments with the highest scores remained when conflicts occurred. Super-scaffolds were constructed using the adjacent overlapping relationship between scaffolds; simultaneously, the orientation between each pair of scaffolds could be determined. The details on scaffold alignment and orientation can be obtained from the manufacturer of OpGen.

5.2.6 Identification of RAD Markers

The genetic maps that were used to develop the integrated map for anchoring the scaffolds were derived from F_1 populations, totalling 260 individuals from the cross between 'Fenban' and 'Kouzi Yudie' from Qingdao Meiyuan. Young leaves of these *P. mume* seedlings and their parents were collected for DNA extraction. Genomic DNA was isolated from the leaves using the Plant Genomics DNA kit (TIANGEN, Beijing, China) according to the manufacturer's recommendations.

The RAD protocols were the same as in Chutimanitsakun et al. (2011), except we used *Eco*R I (recognition site: 5′G^AATTC3′). All 24 F_1 plants were pooled into one sequencing library with nucleotide multiplex identifiers (4, 6 and 8 bp). Approximately, 830 Mb of 50-bp reads (3.1 Mb of reads data for each progeny on average) were generated on the NGS platform HiSeq 2000. The SNP calling process was performed using the SOAP2 + SOAPsnp pipeline.

5.2.7 Genetic Map Construction and Scaffold Anchoring

A total of 260 F_1 seedlings of the cross between 'Fenban' and 'Kouzi Yudie' were used to construct the linkage map. Linkage analysis was performed using JoinMap version 3.0 (Van

Ooijen and Voorrips 2001). The RAD-based SNP markers were first tested against the expected segregation ratio. Two heterozygous SNP alleles between two parents were expected to segregate at a 1:2:1 ratio. One heterozygous and one homozygous SNP allele between two parents were expected to segregate at a 1:1 ratio. Distorted markers ($p < 0.01$) were filtered to construct a genetic map by the chi-square test. Subsequently, reads that contained SNP markers were aligned to the scaffolds. Only unique aligned SNPs with a cut-off of 87.5% identity remained per Blat (coverage ≥ 0.90) (Kent 2002). An LOD score of 12.0 was initially set as the linkage threshold for linkage group identification. Eight linkage groups that had the same number of *P. mume* chromosomes were formed at an LOD threshold of 12.0. All SNP markers were used to construct the *P. mume* consensus map with the CP population model in JoinMap, version 3.0.

To reduce the complex of scaffolds that were anchored to hundreds of SNP markers, a tag SNP was selected from each scaffold with multiple SNPs. We calculated the recombination fractions between all pairs of SNPs in a scaffold and chose the SNP that had the minimum recombination fraction in the sum. Tag SNPs were used to identify the order of scaffolds. Subsequently, two marginal SNPs were used to orient the scaffolds. Scaffolds with only one SNP marker could be anchored but not oriented due to a lack of markers. One hundred scaffolds (18% of 567 scaffolds) were labelled 'uncertain orientation'.

5.2.8 Identification of Repetitive Elements

There are two main types of repeats in the genome, tandem repeats and interspersed repeats. We used Tandem Repeats Finder (Version 4.04) (Benson 1999) and Repbase (composed of numerous transposable elements, Versions 15.01) to identify interspersed repeats in the *P. mume* genome. We identified transposable elements in the genome at the DNA and protein levels. For the former, RepeatMasker (Version

3.2.7) was applied using a custom library (a combination of Repbase, a de novo transposable element library of the *P. mume* genome). For the latter, RepeatProteinMask, an updated tool in the RepeatMasker package, was used to conduct RM-BlastX searches against the transposable element protein database (Jurka et al. 2005). Identified repeats were classified into various categories.

5.2.9   Gene Prediction

To predict genes, four approaches were used: de novo prediction, the homology-based method, the EST-based method and transcript-to-genome sequences. For de novo prediction, Augustus (Stanke et al. 2006), GENSCAN (Salamov and Solovyev 2000) and GlimmerHMM (Majoros et al. 2004) were used with parameters trained on *Arabidopsis thaliana*. For the homology search, we mapped the protein sequences of four sequenced plants (*Cucumis sativus*, *Carica papaya*, *Fragaria vesca* and *A. thaliana*) onto the *P. mume* genome using TBLASTN, with an E-value cut-off of $1e^{-5}$; homologous genomic sequences were aligned against matching proteins using GeneWise (Birney et al. 2004) for accurate spliced alignments.

In the EST-based prediction, 4699 ESTs of *P. mume* were aligned against the *P. mume* genome using BLAT (identity ≥ 0.95, coverage ≥ 0.90) to generate spliced alignments. The de novo set (28,610–36,095), four homology-based results (24,277–29,586) and the EST-based gene set (2001) were combined by GLEAN (Elsik et al. 2007) to integrate a consensus gene set. Short genes (CDS length < 150 bp) and low-quality genes (gaps in more than 10% of the coding region) were filtered. To finalise the gene set, we aligned RNA-Seq data from buds, fruits, leaves, roots and stems to the genome using TopHat (Version 1.2.0, implemented with bowtie1 Version 0.12.5) (Trapnell et al. 2009), and the alignments were used as input for Cufflinks (Trapnell et al. 2010) (Version 0.93) with default parameters. Open reading frames (ORFs) of those transcripts were predicted using structure parameters trained on perfect genes from homology-based predictions. In the end, based on their coordinates on the genome sequences, we manually combined the GLEAN gene set and ORFs of transcripts to form the final gene set, which contained 31,390 genes.

5.2.10   RNA-seq Data Generation

Using TRIzol (Invitrogen), RNA was purified from five fresh tissues (bud, fruit, leaf, root and stem). The RNA sequencing libraries were constructed using the mRNA-Seq Prep kit (Illumina, San Diego, USA). Briefly, first-strand cDNA synthesis was performed with oligo-T primer and Superscript II reverse transcriptase (Invitrogen). The second strand was synthesised with *Escherichia coli* DNA Pol I (Invitrogen). Double-stranded cDNA was purified with a Qiaquick PCR purification kit (Qiagen) and sheared with a nebuliser (Invitrogen) to 100–500-bp fragments. After end repair and addition of a 3′-dA overhang, the cDNA was ligated to Illumina PE adapter oligo mix (Illumina) and size selected to 200 ± 20-bp fragments by gel purification. After 15 cycles of PCR amplification, the 200-bp paired-end libraries were sequenced using the paired-end sequencing module (90 bp at each end) of the Illumina HiSeq 2000 platform.

5.2.11   Gene Annotation

Genes were aligned to the Swiss-Prot (release 2011.6) and TrEMBL (release 2011.6) databases using BLASTP ($1e^{-5}$) to determine the best match of the alignments. InterProScan (Version 4.5) motifs and domains of the genes were identified against protein databases of Pfam (release 24.0), PRINTS (release 40.0), PROSITE (release 20.52), ProDom (release 2006.1) and SMART (release 6.0). Gene ontology IDs for each gene were obtained by the corresponding InterPro entry. The genes were aligned against KEGG proteins (release 58), and the matches were used to establish the KEGG pathway.

5.2.12 Identification of Noncoding RNA Genes

The tRNA genes were predicted by tRNAscan-SE (Version 1.23) (Lowe and Eddy 1997). For rRNA identification, the rRNA template sequences (e.g. *A. thaliana* and rice) were aligned against the *P. mume* genome using BLASTN to identify possible rRNAs. Other noncoding RNAs, including miRNA and snRNA, were identified using INFERNAL (Version 0.81) by searching against the Rfam database (Release 9.1).

5.2.13 Comparative Genome Analysis

Paralogous and orthologous genes were identified by BLASTP search (E-value cut-off $1e^{-5}$). After removing self-matches, syntenic blocks (≥ 5 genes per block) were identified based on MCscan (Tang et al. 2008). The aligned results were used to generate dot plots; for self-aligned results, each block represents the paralogous region that arose from genome duplication, and for inter-aligned results, each block represents the orthologous region that was derived from a common ancestor. We calculated 4DTv (the number of transversions at fourfold degenerate sites) for each gene pair in the block and drew the distribution of 4DTv values to estimate the speciation between species or WGD events.

5.2.14 Identification of Duplicate and Syntenic Regions

Three new parameters were used to identify paralogous and orthologous relationships between *P. mume*, *Malus* $\times$ *domestica*, *F. ve*sca and *Vitis vinifera* by BLASTN. Paralogous gene pairs that were identified during duplication analysis in *P. mume* and *M.* $\times$ *domestica*, respectively, and orthologous gene pairs that were identified by colinearity analysis between *P. mume* and *M.* $\times$ *domestica*, *F. ve*sca and *M.* $\times$ *domestica*, *P. mume* and *V. vinifera* were validated by CloseUp (Hampson et al. 2004) analysis. Based on the

syntenic and duplication relationships, Rosaceae ancestral chromosomes were reconstructed, and the paleo-history was analysed.

5.2.15 Identification of *CBF* and *BEAT* Genes

The *CBF* genes of *P. mume* were identified with *A. thaliana CBF* genes using BLASTP (E-value < $1e^{-10}$, identity > 30% and coverage > 70%). The *BEAT* genes of *P. mume* were identified with *BEAT* genes (Gene Bank ID: AF043464) using BLASTP (E-value < $1e^{-10}$, identity > 30% and coverage > 70%).

5.3 Results

5.3.1 Sequencing and Assembly

To construct the reference genome of *P. mume*, we initially sequenced two domesticated *P. mume* samples using the Illumina Genome Analyzer II. Both samples had a high heterozygous rate, as estimated using K-mer statistics (Fig. 5.2a, b), and the de novo assembly results by current NGS algorithms did not meet the reference quality standards. We then considered wild samples due to their lack of artificial grafting and asexual reproduction. We collected wild samples of *P. mume* in Tongmai, Bomi County, Tibet, China, which is the western-end region of the origin area of domesticated *P. mume* (Xing et al. 2009). In a previous study (Chen 1995; Xing et al. 2009), we determined that the origin area of *P. mume* was confined to an area of approximately 0.7 km^2, and that the primary distribution area was less than 0.3 km^2, below an elevation of 2230 m. Samples from this region were highly homozygous due to generations of self-fertilisation in a hermetically sealed geographic environment. We chose one such wild sample for sequencing, and its low heterozygosity was confirmed by sequencing and K-mer statistics (Fig. 5.2c).

We generated 50.4 Gb of sequencing data for this wild sample, using the Illumina Genome
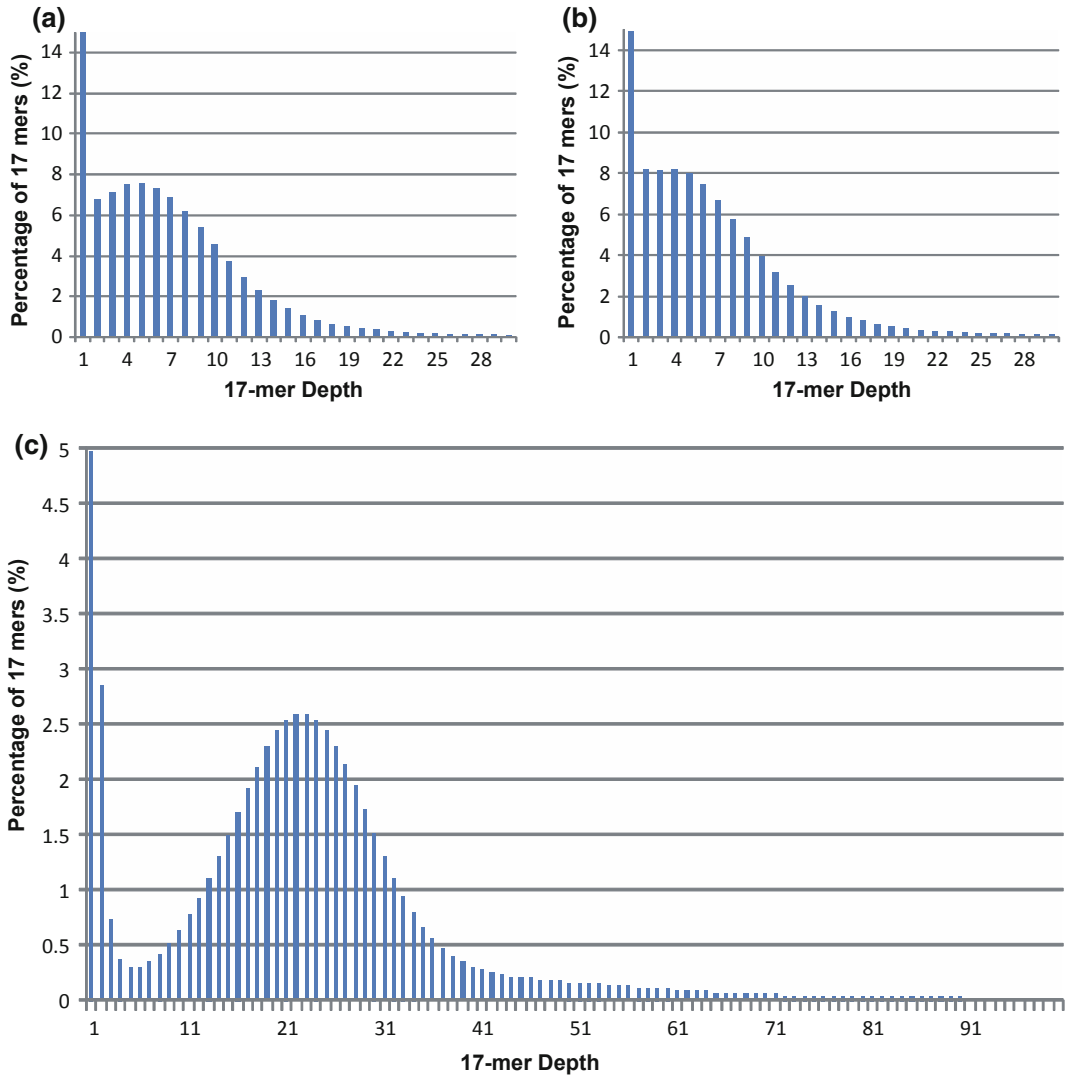
Fig. 5.2 K-mer analysis. **a**, **b** Estimating the domesticated samples, **c** Estimating the wild *P. mume* sample used for genome assembly. The *x*-axis represents depth (*X*); the *y*-axis is the proportion that represents the percentage at that depth. (Without consideration of the sequence error rate, heterozygosis rate and repeat rate of the genome, the 17-mer distribution should obey the Poisson theoretical distribution. In the actual data, due to the sequence error, the low depth of 17-mer will take up a large proportion.)

Analyzer II, including three short-insert (180–800 bp) pair-end libraries and five large-insert (2–40 Kb) mate-pair libraries (Table 5.1). The SOAP*denovo* (Li et al. 2010), a genome assembler algorithm that is based on the *de Bruijn* graph, was used to assemble the *P. mume* genome. Based on K-mer statistics, the *P. mume* genome was estimated to be 280 Mb (Fig. 5.2c). Approximately, 84.6% (237 Mb) of its genome

were assembled. The contig N50 of the assembled sequence was 31.8 Kb (longest, 201.1 Kb), and scaffold N50 was 577.8 Kb (longest, 2.87 Mb) (Table 5.2). By mapping raw reads back to the draft genome, we observed a heterozygosity rate of 0.03% in this wild sample, supporting our assumption of low heterozygosity in the wild sample. We identified 125,383-bp and 19,897-bp sequences in assembly similar to

Table 5.1 Construction of libraries, generation and filtering of sequencing data for genome assembly used

Library insert size (bp)	Read length (bp)	Raw data			Filtered data		
		Total data (Gb)	Sequence depth (X)	Physical depth (X)	Total data (Gb)	Sequence depth (X)[a]	Physical depth (X)[a]
180	100 PE	6.6	23.6	21.2	6.1	21.8	19.6
500	150 PE	10.2	36.4	60.8	7.1	25.4	42.2
800	100 PE	3.7	13.2	52.8	3.0	10.7	42.9
2000	45 PE	2.8	10.0	222.2	2.5	8.9	198.4
5000	45 PE	3.0	10.7	595.3	2.5	8.9	496.1
10,000	90 PE	11.4	40.7	2261.9	4.4	15.7	873.0
20,000	90 PE	4.7	16.8	1865.0	0.8	2.9	317.5
40,000	50 PE	8.0	28.6	11,442.9	2.0	7.1	2871.4
Total		50.4	180.0	16,522.1	28.4	101.4	4861.1

[a]Assumed genome size is 280 Mb

Table 5.2 Statistics of *P. mume* genome assembly

	Contig		Scaffold		Whole-genome mapping	
	Size (bp)	Number	Size (bp)	Number	Size (bp)	Number
N90	5769	7803	85,987	482	85,987	361
N80	12,180	5272	217,085	316	224,931	195
N70	18,473	3815	339,338	229	432,540	118
N60	24,813	2791	443,973	168	711,996	75
N50	31,772	2009	577,822	120	1085,026	48
Longest	201,075		2871,019		15,622,157	
Total number (>100 bp)		45,592		29,989		29,868
Total number (>2 Kb)		10,894		1449		1328
Total	219,917,886	45,811	237,149,662	29,989	237,166,662	29,868

chloroplast and mitochondrial sequences, respectively. Although most of the regions were shorter than 1000 bp, we found a ∼25,000-kb chloroplast-similar region which might be the plasmid sequence or a plastid-transferred nuclear fragment.

To improve the assembly of the *P. mume* genome, we performed whole-genome mapping (WGM), an automated high-throughput optical mapping method (Zhou et al. 2004), to generate an entire genomic map (Fig. 5.3a). We constructed the WGM map with a 300-fold whole-genome depth using *Bam*H I and *Nhe* I independently. Through an iterative assembly strategy, combined with WGM and sequence data, 170 scaffolds, that were assembled by NGS, were grouped into 49 large scaffolds; thus, the scaffold N50 improved significantly from 578 Kb to 1.09 Mb (Table 5.2).

Subsequently, we constructed a high-density genetic map by applying restriction site-associated DNA (RAD) marker strategy (Chutimanitsakun et al. 2011; Baxter et al. 2011; Miller et al. 2007) in a segregating F_1 population. The consensus genetic map contained eight linkage groups, consisting of a set of 1484 high-quality SNP markers (co-dominate markers), 779 of which were used in anchoring and
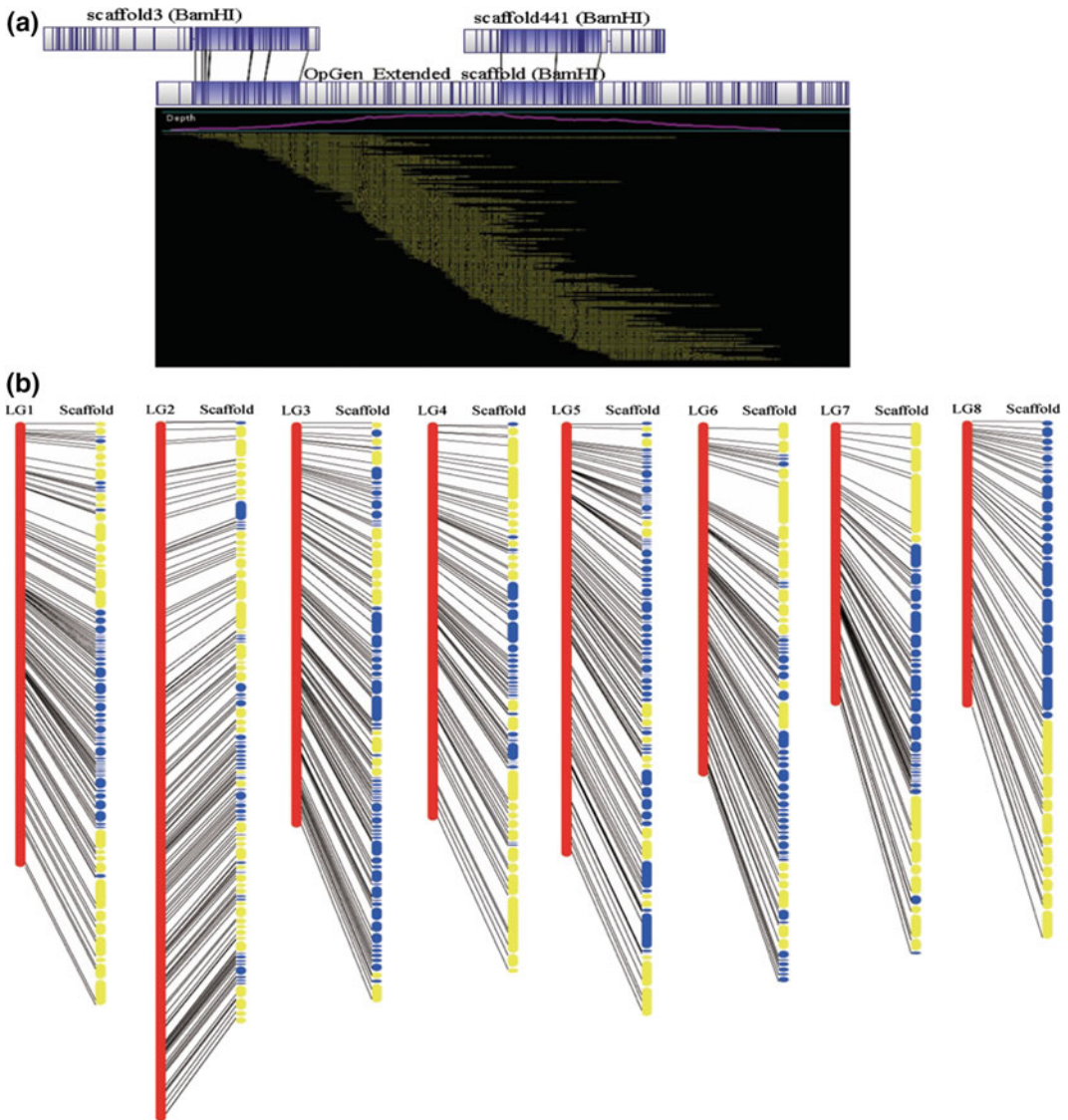
Fig. 5.3 Whole-genome mapping assembly and anchoring of the *P. mume* genome. **a** Assembly of *P. mume* genome by whole-genome mapping. **b** Anchoring of the *P. mume* genome into eight linkage groups using 779 high-quality SNP markers. Yellow scaffolds were anchored by whole-genome mapping and SNP markers, whereas the blue scaffolds were anchored by SNP markers

orienting scaffolds (Fig. 5.3b). The genetic map improves the quality of the reference and would be useful in map-based cloning and further marker-assisted molecular breeding.

5.3.2 Genome Annotation

We annotated 106.8 Mb (45.0% of the assembled genome) of repetitive sequences (Table 5.3)
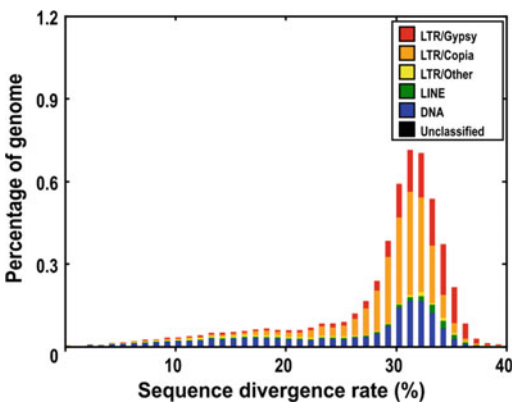
Table 5.3 Statistics of repeats in the *P. mume* genome

Type	Repeat size (Mb)	% of genome
Protein mask	17.32	7.29
RepeatMasker	12.36	5.20
Trf	10.58	4.45
Denovo	103.15	43.41
Total	106.75	44.92

in the *P. mume* genome by integrating the de novo and homology-based approaches. Transposable elements (TEs) were the predominant components, constituting 97.9% of all repetitive sequences. The long terminal repeat (LTR) *Copia* family and the *Gypsy* family are the most abundant TEs in the *P. mume* genome. The TE divergence rate suggested a lack of recent amplification (Fig. 5.4); these were conserved in *M.* × *domestica* and *F. vesca*, while *M.* × *domestica* had a much higher proportion in the *Gypsy* family than *P. mume* and *F. vesca* (Table 5.4).

To improve gene annotation, we generated 11.3 Gb of RNA-Seq data from five major tissues: bud, fruit, leaf, root and stem (Table 5.5). Using these data, integrated with ab initio homology prediction methods, we annotated 31,390 protein-coding genes, comparable with the value in *F. vesca* (34,809) (Shulaev et al.

2011) and less than that found for *M.* × *domestica* (57,386) (Velasco et al. 2010), as expected. Average transcript length in *P. mume* was 2514 bp, with 4.6 exons per gene (Table 5.6). Approximately 82.5% of all genes could be functionally annotated following a consensus method of either known homologous or predictive sequence signatures using Swiss-Prot, GO, TrEMBL (Bairoch and Apweiler 2000), InterPro (Zdobnov and Apweiler 2001) or KEGG (Kanehisa and Goto 2000), and we considered 98.3% of these annotations were high confidence (Table 5.7). We noted that 25,854 (82.6%) gene models were expressed, of which 768, 308, 240, 762 and 179 genes were expressed solely in buds, fruits, leaves, roots and stems, respectively. We further integrated 82,832 peach EST sequences; together with RNA-seq data, 85.1% of gene models represented transcripts. We also annotated the noncoding RNA genes in the current assembly, including 508 tRNA, 209 miRNA, 125 rRNA and 287 snRNA (Table 5.8).

5.3.3 Genome Evolution

By genome self-alignment via MCscan (Tang et al. 2008) (Fig. 5.5) and 4DTv (the number of transversions at fourfold degenerate sites) distribution of duplicated pairs, there was no recent whole-genome duplication after *P. mume* species differentiation from *M.* × *domestica* (Fig. 5.6a).

We examined the paleo-history in *Prunus* and found that there was a triplicated arrangement (ancestral γ event). We aligned 27,819 gene models to the seven paleo-hexaploid ancestor chromosome groups in grape (Jaillon et al. 2007) and identified that the colinearity blocks

Fig. 5.4 Divergence rates of the transposable elements in the assembled scaffolds. The divergence rate was calculated based on the alignment between the RepeatMasker-annotated repeat copies and the consensus sequence in the repeat library

Table 5.4 Occurrence of transposable elements in sequenced Rosaceae genomes

Classification	*P. mume*			*M.* × *domestica*			*F. vesca*		
	Total length	TE coverage	Total genome coverage	Total length	TE coverage	Total genome coverage	Total length	TE coverage	Total genome coverage
	(Mb)	(%)	(%)	(Mb)	(%)	(%)	(Mb)	(%)	(%)
LTR/Copia	23.8	22.8	10	40.6	12.9	5.5	10.8	22.5	5.3
LTR/Gypsy	20.4	19.5	8.6	187.1	59.5	25.2	12.9	26.8	6.4
LTR/Other	21.8	20.8	9.2	3.2	1	0.4	8.5	17.7	4.2
LINE	3.1	3	1.3	48.1	15.3	6.5	0.7	1.5	0.3
SINE	0.9	0.9	0.4	-	-	-	0.2	0.4	0.1
DNA transposons	20.2	19.3	8.5	6.6	2.1	0.9	12.9	26.8	6.4
Other	1.1	1.1	0.5	-	-	-	2.1	4.4	1
Unknown	13.3	12.7	5.6	28.9	9.2	3.9	-	-	-
Total	104.6	100	44.1	314.5	100	42.4	48.1	100	

contained 2772 orthologs; the extent of these blocks covered 78.1% of the *P. mume* genome.

Further, we aligned 27,819 *P. mume* gene models to themselves and identified seven major blocks of duplication, which corresponded to 194 gene pairs, covering 38.5% of the anchored genome. The chromosome-to-chromosome relationships P5-P7, P2-P4-P8, P1-P2-P4-P6, P1-P5, P2-P5-P8, P2-P4-P7 and P3-P4 (Fig. 5.6b) suggested that triplicated arrangement (γ event) marks remained in the *P. mume* genome. Based on the evidence of paleo-hexaploidisation (γ event) and lineage-specific duplications in eudicots, it was possible to examine chromosomal changes during the evolution of *P. mume* and other Rosaceae species.

5.3.4 Reconstruction of Ancestral Chromosomes of Rosaceae

We reconstructed nine ancestral chromosomes of the Rosaceae family and determined the history of chromosome fusion, fission and duplication in the three major Rosaceae subfamilies. Previous studies have reported the eudicots ancestor with seven proto-chromosomes (Jaillon et al. 2007) and the grass ancestor with five proto-chromosomes

(Salse et al. 2009). Using *M.* × *domestica* as a reference, we analysed the syntenic relationships between the sequenced Rosaceae genomes of *P. mume* (P), *F. vesca* (F) and *M.* × *domestica* (M).

We identified 151 blocks that contained 4546 orthologous genes, covering 96.9% of the anchor *P. mume* genome, between *P. mume* and *M.* × *domestica* versus 132 blocks with 2031 orthologous genes, covering 88.8% of the *F. vesca* genome, between *F. vesca* and *M.* × *domestica*. The chromosome-to-chromosome orthologous relationships are shown in Fig. 5.7.

Combining intergenomic and intragenomic analysis of the Rosaceae genomes, we noted the following primary chromosome pair combinations: M5-M10/P3-P6/F2-F3, M3-M11/P1-P3-P6/F2-F3-F5, M9-M17/P2-P4/F1-F6, M13-M16/P2/F4, M4-M12-M14/P1-P2-P5-P8/F5-F6, M5-M14/P7/F5, M1-M2-M7/P5/F7, M8-M15/P2/F2-F5 and M1-M2-M15/P1-P2-P3-P8/F1 (Fig. 5.7). We reconstructed a putative ancestral genome of Rosaceae and proposed an evolutionary scenario of *P. mume*, *M.* × *domestica* and *F. vesca* from the putative nine-chromosome ancestor (Fig. 5.8).

In *P. mume*, chromosomes 4, 5 and 7 did not undergo rearrangement, coming directly from ancient chromosomes III, VII and VI,

Table 5.5 List of tissues and reads for whole-transcriptome sequencing mapped onto the *P. mume* genome

Tissue type	Bud	Fruit	Leaf	Root	Stem
Perfect match read no.	1,24,99,437	1,77,05,866	1,62,37,146	1,83,73,940	1,21,46,280
≤ 5 bp mismatch read no.	34,75,572	57,89,365	51,87,087	63,26,385	42,86,634
Unique match read no.	1,54,05,362	2,24,04,254	2,04,69,481	2,35,28,150	1,57,83,948
Multi-position match read no.	5,69,647	10,90,977	9,54,752	11,72,175	6,48,966
Total mapped reads no.	1,59,75,009	2,34,95,231	2,14,24,233	2,47,00,325	1,64,32,914
Total unmapped reads no.	35,23,231	53,38,431	45,09,397	66,38,567	38,28,296
Total reads no.	1,94,98,240	2,88,33,662	2,59,33,630	3,13,38,892	2,02,61,210
Total base pairs (bp)	1,75,48,41,600	2,59,50,29,580	2,33,40,26,700	2,82,05,00,280	1,82,35,08,900

respectively. Chromosome 1 came from ancestral chromosomes II, V and IX. Chromosome 2 originated from ancient chromosomes IV and VIII, into which some of chromosomes III and V were inserted. Chromosome 3 came from chromosomes I, II and IX; chromosome 6 came from I, II, and V, and chromosome 8 originated from ancient chromosomes V and IX. Thus, we hypothesised that at least eleven fissions and eleven fusions occurred in *P. mume* from the nine common ancestral chromosomes. For *M.* × *domestica*, at least one whole-genome duplication (WGD) and five fusions took place to reach the 17-chromosome structure, compared with 15 fusions for *F. vesca* to affect the 7-chromosome structure.

5.3.5 Early Blooming of P. Mume

P. mume is nearly the first tree that blooms in early spring, blooming even at temperatures below 0 °C. Thus, *P. mume* has a specific mechanism to acclimate to cold weather and to release itself from dormancy. The (DAM) dormancy-associated MADS-box transcription factors family, which is related to dormancy induction and release (Sasaki et al. 2011), was identified in the *P. mume* assembly, and all six *DAM* genes were noted in the arrayed tandem (Fig. 5.9a, Table 5.9). In a previous study, the authors identified six *DAM* genes, all of which were transcriptional repressors in *P. mume* (Sasaki et al. 2011). They also found that all *DAM* genes were repressed during prolonged exposure to cold and maintained at low levels until endodormancy release. The authors of another study observed that expression of *DAM3*, *DAM5* and *DAM6* was suppressed by chilly temperatures, bottoming on bud break in *Prunus persica* (Jiménez et al. 2010). These findings suggest that DAM inhibits dormancy release and that its expression is suppressed during prolonged exposure to cold to allow bud release from dormancy.

We hypothesised that the *DAM* genes explained the early dormancy release in *P. mume*. To this end, we examined the

Table 5.6 General statistics of predicted protein-coding genes

Gene set	EST	Protein homology search				Gene finder			GLEAN	RNA-Seq	Combine
		C. sativus	*C. papaya*	*F. vesca*	*A. thaliana*	Augustus	GENSCAN	GlimmerHMM			
Number	4,699	24,277	27,200	29,586	25,414	32,479	28,610	36,095	30,012	21,585	31,390
Average length of transcribed region (bp)	2001	2533	2022	2642	2412	2442	5211	2032	2523	2454	2514
Average length of CDS (bp)	562	1053	913	1043	1008	1175	1315	964	1164	1074	1146
#Exons per gene	3.1	4.2	3.7	4	4.2	5.1	6	3.9	4.7	4.4	4.6
Average length of exon (bp)	184	253	247	257	241	229	217	245	249	245	249
Average length of intron (bp)	701	469	411	521	441	307	772	364	369	409	380

**Table 5.7** Functional annotation of predicted genes with homology or functional classification by each method

	Database	Number	Percent (%)
Annotated	Swiss-Prot	19,696	62.8
	InterPro	21,236	67.7
	GO	16,822	53.6
	KEGG	15,504	49.4
	Trembl	25,650	81.7
	Total	25,905[a]	82.5
Unannotated		5485	17.5
Total		31,390	100

[a]449 annotations were hits to hypothetical or uncharacterised proteins

Table 5.8 Noncoding RNA gene fragments in the current *P. mume* assembly

ncRNA Type	Copy#	Average length (bp)	Total length (bp)	% of genome
miRNA	209	120.65	25,216	0.0106
tRNA	508	75.21	38,209	0.0012
rRNA	125	196.89	24,611	0.0103
28S	46	348.98	16,053	0.0067
18S	17	111.29	1892	0.0008
5.8S	11	112.55	1238	0.0005
5S	51	106.43	5428	0.0022
snRNA	287	118.09	33,891	0.0142
CD-box	158	98.08	15,497	0.0065
HACA-box	21	118.14	2481	0.001
slicing	108	147.34	15,913	0.0067

Fig. 5.5 Whole-genome duplication in the *P. mume* genome mapped using gene collinear order information. Syntenic blocks are formed by red or blue dots, representing best hits across any two chromosomes in the same or opposite direction, respectively
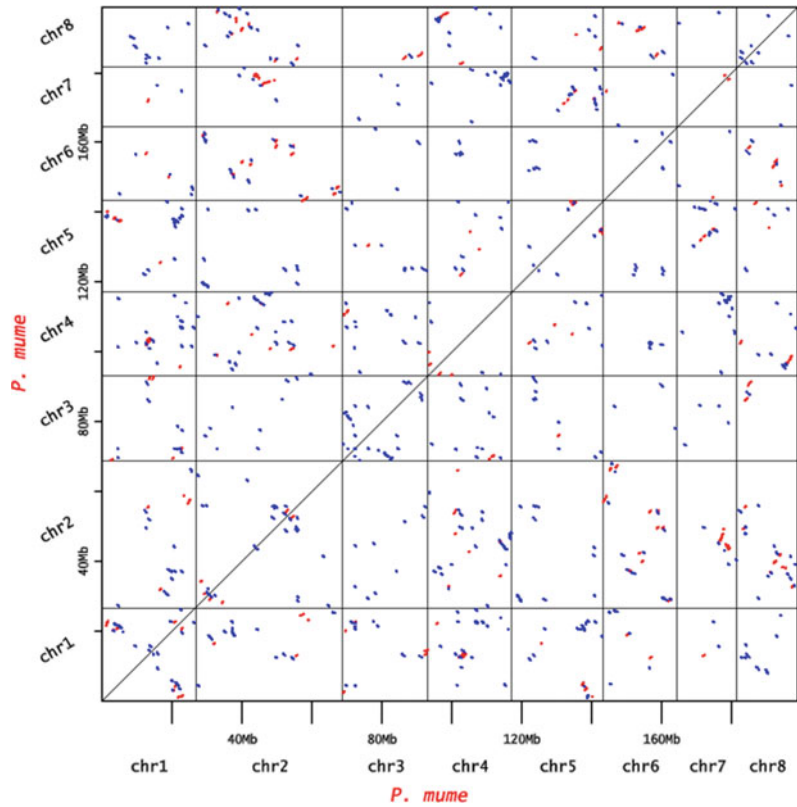
phylogenetic relationships between *DAM* genes in *P. mume*, using the PHYML 3.0 software. The molecular evolution models of the six tandem *DAM* genes in *P. mume* suggested that they were derived from serial duplication events in the following order: *PmDAM1*, *PmDAM3*, *PmDAM2*, *PmDAM5*, *PmDAM4* and *PmDAM6* (Fig. 5.10). This model is consistent with previous studies in the peach genome (Jiménez et al. 2009), and we suggest that these duplication events are unique in the *Prunus* subfamily—we did not find these tandem *DAM* genes in *M.* × *domestica* or *F. vesca*. The application of two other programs, MEGA 4 (Tamura et al. 2007) and PAML4 (Yang 2007), obtained similar results. By estimation of pairwise dN and dS rates using MEGA 4, there was significant purifying selection and no significant positive

selection of the six sequences. In an analysis of *P. mume DAM* genes, most sites were highly conserved, with a dN/dS rate ratio near 0 or nearly neutral.

In addition to *DAM*, the *C-repeat binding transcription factor* (*CBF*) mediates the establishment of early dormancy release—overexpression of *PpCBF1* in *M.* × *domestica* results in a strong sensitivity to short day lengths and induced dormancy at optimal growth temperatures (Wisniewski et al. 2011). Horvath et al. (2010) noted that *EeDAM1* was cold stress-responsive and contains putative CBF-binding sites, which are cis-regulating motifs that are targeted by the cold/drought stress CBF regulon in the 2000-bp region upstream of the *EeDAM1* translation start codon. This finding suggests that *CBF* controls the cold-responsive *EeDAM1* gene (Horvath et al.
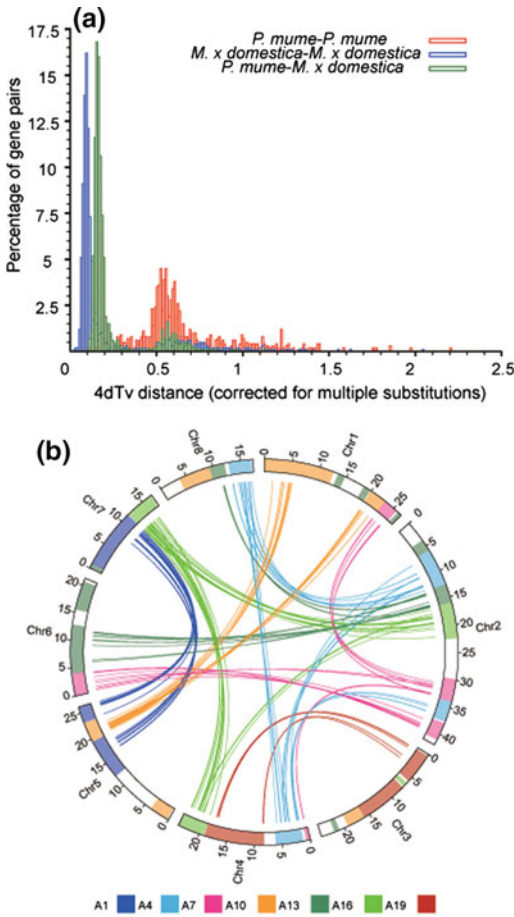
Fig. 5.6 Evolution of *P. mume*. **a** 4DTv distribution of duplicate gene pairs in *P. mume* and *M. × domestic*a, calculated based on alignment of codons with the HKY substitution model. **b** The duplication of *P. mume* by paralogous pairs in the *P. mume* genome (chromosomes Chr1 to Chr8). Each line represents a duplicated gene. The seven colours reflect the seven ancestral eudicot linkage groups (A1, A4, A7, A10, A13, A16 and A19)

2010). Similar to *EeDAM1*, conserved *CBF* sites were identified in the 1000-bp region upstream of the translation start codons in *DAM4* to *DAM6* in peach and Japanese apricot (Sergio Jiménez et al. 2009; Sasaki et al. 2011). In the *P. mume* genome, we identified 13 *CBF* orthologous genes

(Tables 5.10) and seven *CBF* regulons, *late embryogenesis-abundant (LEA) proteins/dehydrins* (Table 5.11). In the upstream regions of *DAM* genes in *P. mume*, we noted more putative CBF-binding sites on *DAM4*, *DAM5* and *DAM6* than in peach and found novel sites, one on *DAM1* and two on *DAM6* (Fig. 5.9b). We suggest that these additional sites render *P. mume* more sensitive to cold and result in early blooming in spring. In summary, our analyses have increased our understanding of the molecular control of dormancy, flowering regulation and acclimation to cold. The *CBF* and *DAM* genes that we identified in *P. mume* might allow molecular biology facilities to breed fruits and ornamental plants with disparate blooming times.

5.3.6 Floral Scent

Floral scent, determined by a complex mixture of low-molecular-weight volatile molecules, has a significant function in the reproductive processes of many plants and enhances the aesthetic properties of ornamental plants (Pichersky and Dudareva 2007). In earlier research, we determined that the dominant compound classes in *P. mume* flowers were benzenoid/phenylpropanoid from the cinnamic acid pathway, in addition to terpene compounds.

Genes related to the biosynthesis of volatile compounds, such as (*BEAT*) benzyl alcohol acetyltransferase, which catalyses the synthesis of benzyl acetate (Dudareva et al. 1998; Aranovich et al. 2007), have been identified in the *P. mume* genome. The *BEAT* gene family expanded notably in *P. mume* (34 members) compared with *M. × domestics* (16), *F. vesca* (14), *V. vinifera* (4), *Populus trichocarpa* (17) and *A. thaliana* (3) (Table 5.12). Twenty-six of 34 *P. mume BEAT* genes lay in clusters, the largest of which contained 12 members that were
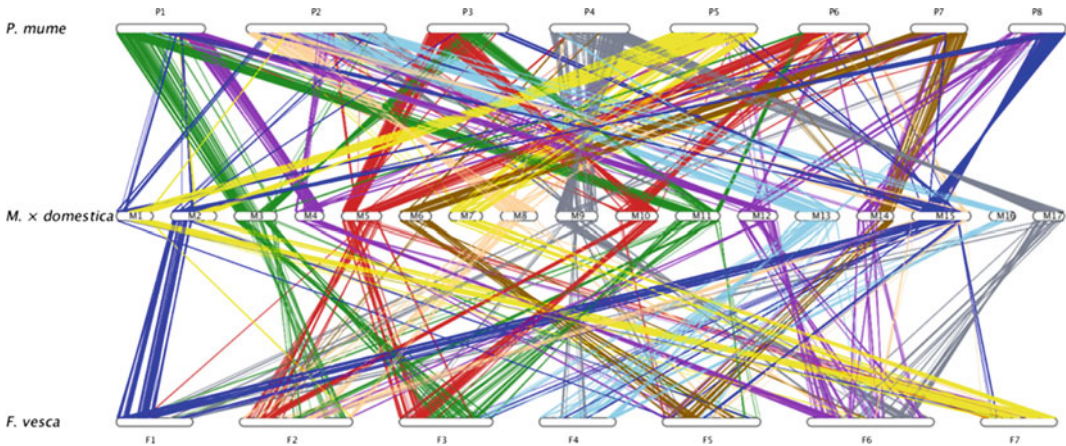
Fig. 5.7 Synteny between *P. mume*, *F. vesca* and *M.* × *domestica*. Schematic representation of the orthologs identified between *P. mume* (P1–P8), *F. vesca* (F1–F7) and *M.* × *domestica* (M1–M17). Each line represents an orthologous gene. The nine different colours represent the blocks reflecting the origin from the nine ancestral Rosaceae linkage groups

arranged in tandem (Fig. 5.11), suggesting that *BEAT* genes originated from serial duplication events, in contrast to the other sequenced plants.

In summary, the expansion of the *BEAT* gene family might increase the content of benzyl acetate and be related to the special fragrance of *P. mume*. Research on the *P. mume* genome should allow us to breed novel aromatic cultivars and other aromatic plants in the Rosaceae family.

5.4 Discussion

We are eager to establish an appropriate reference genome for Rosid species, which include one-third of all flowering plants (Hummer and Janick, 2009). Currently, draft genome sequences are available for three model Rosaceae species—*M.* × *domestica*, *F. vesca* and *P. mume*. The domesticated apple (*M.* × *domestica*), the main fruit crop of temperate regions throughout the world, is highly heterozygous and has a large

genome, which creates technical challenges in assembling its genome, resulting in 1629 metacontigs.

The strawberry (*F. vesca*) has a much smaller genome of ∼240 Mb, allowing functional gene studies within Rosaceae. Although the strawberry is useful for functional genomics research, most related high-value fruit plants in the Rosaceae species—peach (*P. persica*), pear (*Pyrus nivalis*) and cherry (*Prunus avium*)—are woody plants, not herbaceous ones. Unfortunately, the nearest relatives of woody fruit crops usually have a cumbersome polyploid genome.

After a detailed study of its origin and the current distribution in *P. mume*, we obtained a suitable sample from an isolated group in the origin area of *P. mume* for genomic sequencing and assembly. We report the genomic sequence of *P. mume* due to its small genome of ∼280 Mb and low heterozygosity. With whole-genome mapping, we increased the scaffold quality to 1.1 Mb in N50 and constructed *P. mume* pseudochromosomes using 779
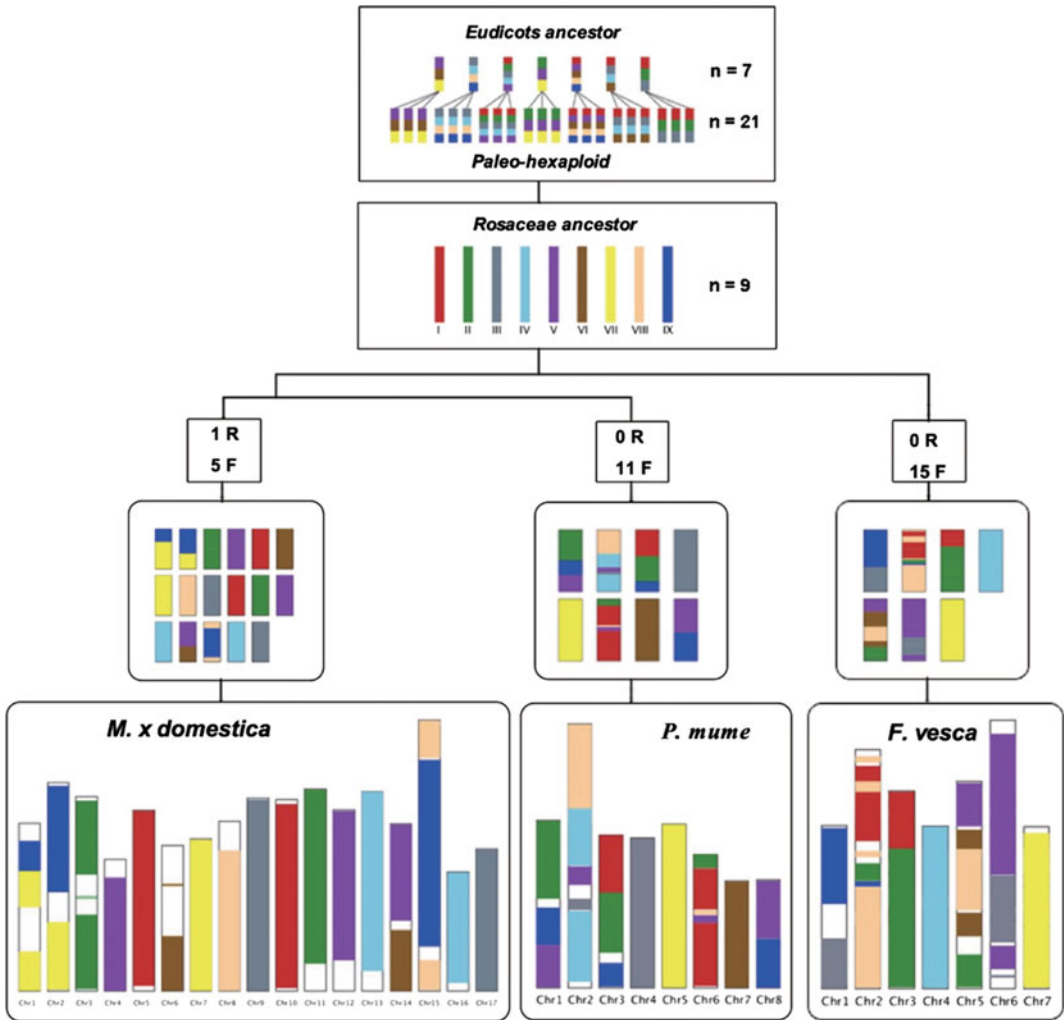
Fig. 5.8 Evolutionary model of the Rosaceae genome. The Rosaceae ancestor chromosomes are represented by nine colours. The various evolutionary processes from the common ancestor are indicated as R (whole-genome duplication (WGD)) and F (for fusions of chromosomes). In the second layer, different colours in each chromosome represent the origin of the common ancestral chromosomes. The current structure of the Rosaceae genome is shown at the bottom of the figure. In some regions, we were not able to determine which ancestral chromosome they came from, and those regions therefore represented as white spaces
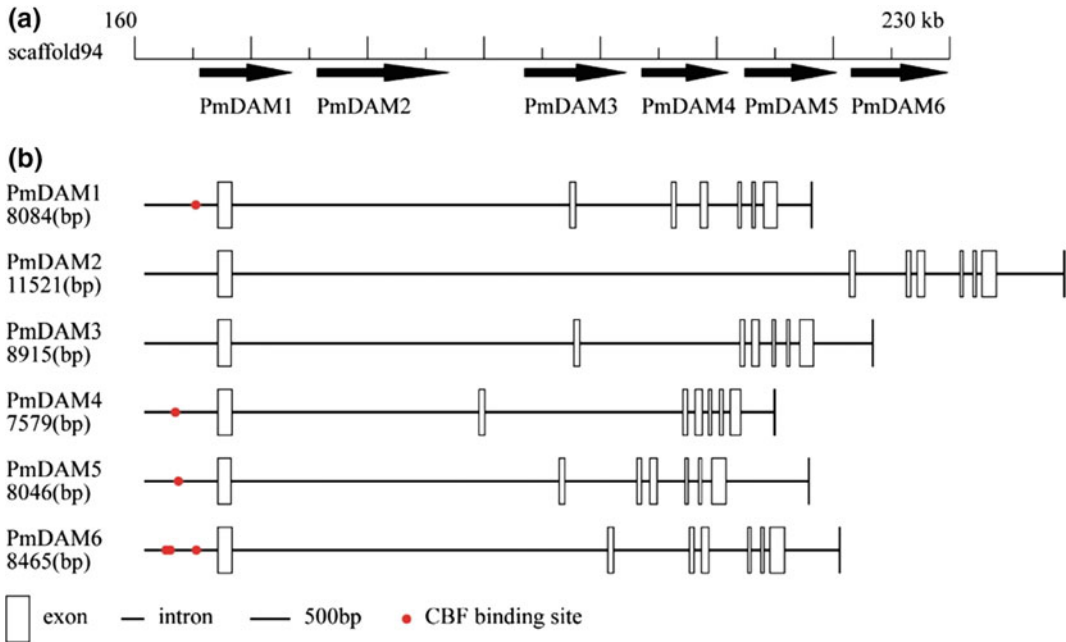
Fig. 5.9 Six tandemly arrayed *DAM* genes in *P. mume*. **a** Overview of *PmDAM* in the *P. mume* genome. Six *PmDAM* genes are located as tandem repeats. **b** Structures of *PmDAM* genes in *P. mume*. Boxes and lines represent exons and introns, whereas red points represent CBF-binding sites

Table 5.9 *DAM* gene orthologs of in *P. mume*

Gene name	Query species	ID	*P. mume* gene prediction	
			Scaffold	Genemark
PmDAM1	*Prunus persica*	gb|DQ863253.2	scaffold94	Pm004420
PmDAM2	*Prunus persica*	gb|DQ863255.1	scaffold94	Pm004419
PmDAM3	*Prunus persica*	gb|DQ863256.1	scaffold94	Pm004418
PmDAM4	*Prunus persica*	gb|DQ863250.1	scaffold94	Pm004417
PmDAM5	*Prunus persica*	gb|DQ863251.1	scaffold94	Pm004416
PmDAM6	*Prunus persica*	gb|AB437345.1	scaffold94	Pm004415

Fig. 5.10 Maximum likelihood rooted tree of six *P. mume* DAM genes; PtMADS27 was used as outgroup
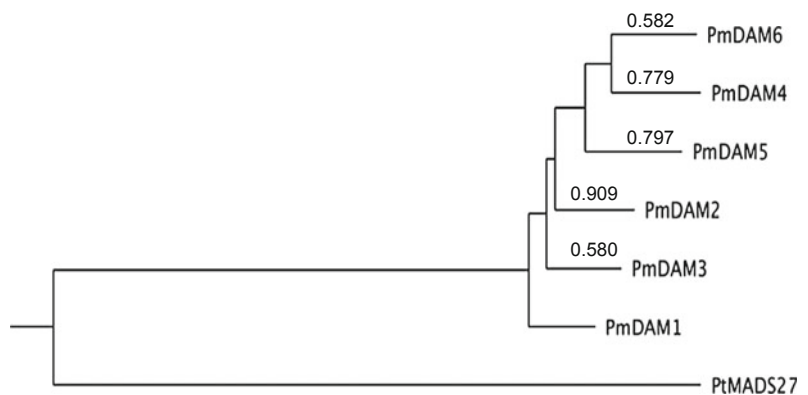
Table 5.10 CBF orthologs of in *P. mume*, *M.* × *domestica*, *F. vesca*, *P. trichocarpa*, *V. vinifera*, *O. sativa* and *A. thaliana*

Species	Number	Accession number
P. mume	13	Pm004870, Pm019385, Pm019386, Pm026227, Pm023766, Pm023767, Pm023768, Pm023769, Pm023770, Pm023772, Pm023775, Pm023777, Pm026221
M. × *domestica*	10	MDP0000154764, MDP0000155057, MDP0000189347, MDP0000195376, MDP0000198054, MDP0000262710, MDP0000400129, MDP0000451365, MDP0000652413, MDP0000833641
F. vesca	6	mrna13327.1, mrna13329.1, mrna30159.1, mrna30226.1, mrna32378.1, mrna32380.1
P. trichocarpa	14	POPTR_0001s08710.1, POPTR_0001s08720.1, POPTR_0001s08740.1, POPTR_0003s12120.1, POPTR_0004s19820.1, POPTR_0006s02180.1, POPTR_0009s14990.1, POPTR_0012s13870.1, POPTR_0012s13880.1, POPTR_0013s10330.1, POPTR_0015s13830.1, POPTR_0015s13840.1, POPTR_0016s02010.1, POPTR_0019s10420.1
O. sativa	11	Os01t0968800-00, Os02t0558700-00, Os02t0676800-01, Os02t0677300-01, Os03t0117900-01, Os04t0572400-00, Os08t0545500-00, Os09t0522000-01, Os09t0522100-00, Os09t0522200-02, Os11t0242300-00
V. vinifera	5	GSVIVT01019860001, GSVIVT01031387001, GSVIVT01031388001, GSVIVT01033793001, GSVIVT01033795001
A. thaliana	10	AT1G12610.1, AT1G12630.1, AT1G63030.1, AT2G35700.1, AT2G36450.1, AT4G25470.1, AT4G25480.1, AT4G25490.1, AT5G51990.1, AT5G52020.1

Table 5.11 Dehydrin orthologs in *P. mume*, *M.* × *domestica*, *F. vesca*, *P. trichocarpa*, *V. vinifera*, *O. sativa* and *A. thaliana*

Species	Number	Accession number
P. mume	7	Pm000687, Pm026682, Pm026683, Pm026684, Pm020945, Pm021811, Pm006114
M. × *domestica*	17	MDP0000126135, MDP0000129775, MDP0000178973, MDP0000196703, MDP0000265874, MDP0000269995, MDP0000360414, MDP0000529003, MDP0000595270, MDP0000595271, MDP0000629961, MDP0000689622, MDP0000698024, MDP0000770493, MDP0000862169, MDP0000868044, MDP0000868045
F. vesca	7	mrna14934.1, mrna14935.1, mrna14938.1, mrna14940.1, mrna17179.1, mrna21840.1, mrna27549.1
P. trichocarpa	8	POPTR_0002s01460.1, POPTR_0003s13850.1, POPTR_0004s16590.1, POPTR_0005s26930.1, POPTR_0009s12290.1, POPTR_0013s05870.1, POPTR_0013s05880.1, POPTR_0013s05890.1
O. sativa	7	Os01t0702500-01, Os02t0669100-01, Os11t0451700-00, Os11t0453900-01, Os11t0454000-01, Os11t0454200-01, Os11t0454300-01
V. vinifera	3	GSVIVT01018878001, GSVIVT01019440001, GSVIVT01023824001
A. thaliana	10	AT1G20440.1, AT1G20450.1, AT1G54410.1, AT1G76180.2, AT2G21490.1, AT3G50970.1, AT3G50980.1, AT4G38410.1, AT4G39130.1, AT5G66400.1

Table 5.12 Numbers of orthologous genes found in *P. mume* (Pm), *M.* × *domestics* (Md), *P. trichocarpa* (Pt), *A. thaliana* (At), *V. vinifera* (Vv), *F. vesca* (Fv) and *O. sativa* (Os) that synthesise volatile molecules

Type gene	Pm	Md	Pt	At	Vv	Fv	Os
PAL	2	6	5	4	5	2	9
ODO1	2	2	4	1	2	2	1
BPBT	13	25	27	11	12	12	29
CFAT	4	5	2	4	4	5	2

(continued)

Table 5.12 (continued)

Type gene	Pm	Md	Pt	At	Vv	Fv	Os
BSMT	21	32	25	23	25	15	13
CCMT	12	34	23	21	25	12	11
BEAT	34	16	17	3	4	14	–
OOMT	13	37	28	9	18	9	23
IEMT	2	44	30	15	12	14	13
EGS	9	13	18	8	18	10	7
IGS	2	10	17	8	17	10	7
POMT	6	43	32	15	12	14	11
SAMT	10	33	24	24	25	14	10
PAAS	6	4	5	2	5	6	7
r-terpinene-synthase	5	21	33	9	31	27	5
β-pinene-synthase	4	20	34	10	29	27	5
Germacrene	16	19	34	30	30	28	8
TPS10	8	13	13	10	7	9	8
Linalool synthase	1	2	2	1	–	1	2
CCD	6	12	16	7	7	6	5
Limonene-3-hydroxylase	68	111	108	93	50	56	97

SNP markers from eight linkages groups. We conclude that the combination of Illumina GA, whole-genome mapping technologies and the genetic map constructed by RAD can be used to perform de novo sequencing of plant genomes, allowing high-quality, rapid and low-cost sequencing of other plant species with similar conditions.

The sequence of the *P. mume* genome is a valuable resource for biological research and breeding. Based on the sequences of *P. mume*, *M. × domestica* and *F. vesca*, we reconstructed nine ancestral chromosomes of the Rosaceae family and inferred that they were shaped from an ancestor. Analysis of the *P. mume* genome and transcriptome can provide insights into the mechanisms of flowering scent, flowering dormancy and disease resistance. The genome also increases our knowledge of the evolution of the Rosaceae family and the function of the plant relative system.
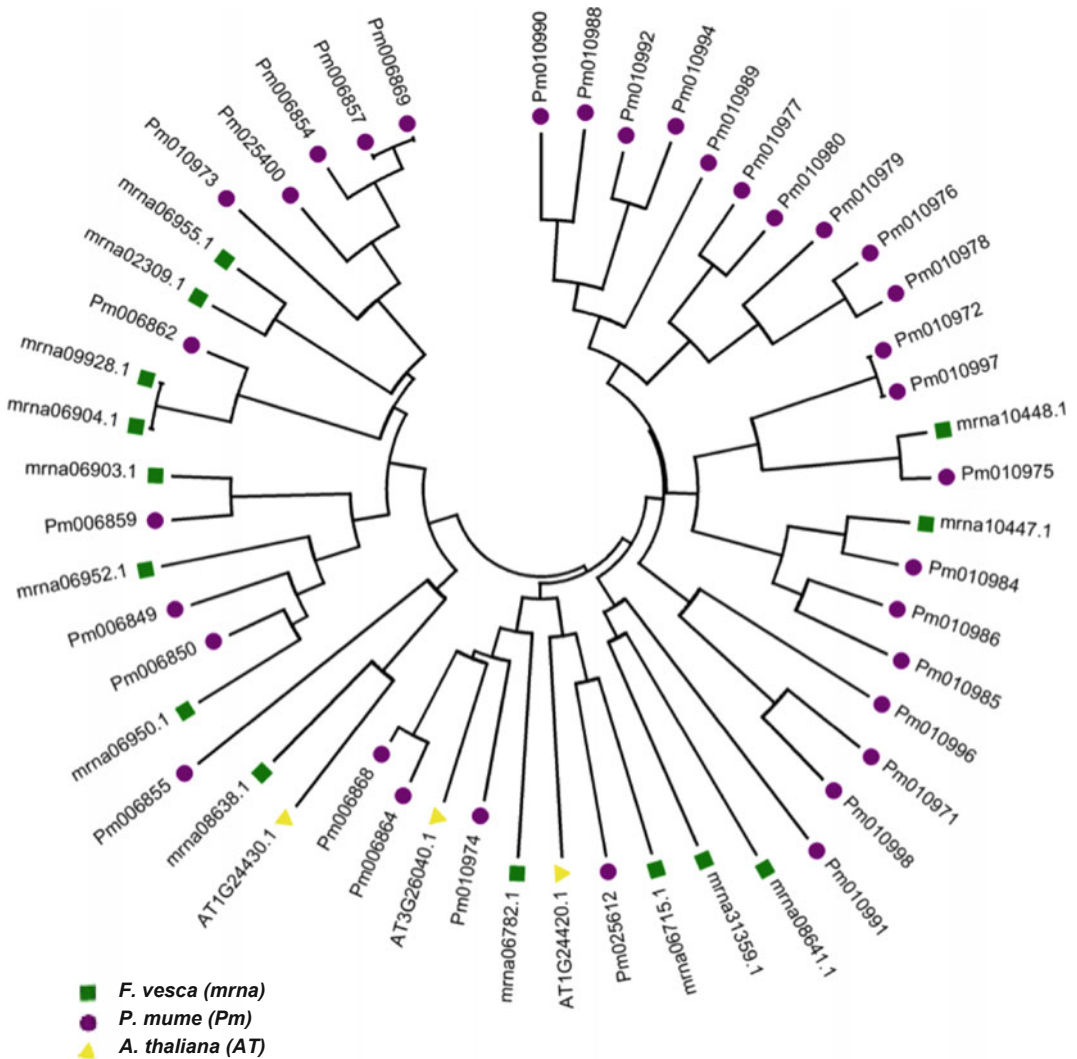
Fig. 5.11 Phylogenetic relationships of *BEAT* genes in *P. mume* (Pm), *F. vesca* (mrna) and *A. thaliana* (AT)

References

Aranovich D, Lewinsohn E, Zaccai M (2007) Post-harvest enhancement of aroma in transgenic lisianthus (*Eustoma grandiflorum*) using the *Clarkia breweri benzyl alcohol acetyltransferase* (*BEAT*) gene. Postharvest Biol Technol 43(2):255–260. https://doi.org/10.1016/j.postharvbio.2006.09.001

Bairoch A, Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 28(1):45–48. https://www.ncbi.nlm.nih.gov/pubmed/10592178

Baxter SW, Davey JW, Johnston JS, Shelton AM, Heckel DG, Jiggins CD et al (2011) Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. PLoS ONE 6(4): e19315. https://doi.org/10.1371/journal.pone.0019315

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27(2):573–580. https://www.ncbi.nlm.nih.gov/pubmed/9862982

Birney E, Clamp M, Durbin R (2004) GeneWise and genomewise. Genome Res 14(5):988–995. https://doi.org/10.1101/gr.1865504

Chen J (1995) Some aspects on Chinese Mei flower research. J Beijing For Univ 17(S1):1–7

Chutimanitsakun Y, Nipper RW, Cuesta-Marcos A, Cistue L, Corey A, Filichkina T et al (2011) Construction and application for QTL analysis of a restriction site associated DNA (RAD) linkage map in

barley. BMC Genom 12(1):4. https://doi.org/10.1186/1471-2164-12-4

Dudareva N, Raguso RA, Wang J, Ross JR, Pichersky E (1998) Floral scent production in *Clarkia breweri*. III. Enzymatic synthesis and emission of benzenoid esters. Plant Physiol 116(2):599–604. https://www.ncbi.nlm.nih.gov/pubmed/9489012

Elsik CG, Mackey AJ, Reese JT, Milshina NV, Roos DS, Weinstock GM (2007) Creating a honey bee consensus gene set. Genome Biol 8(1):R13. https://doi.org/10.1186/gb-2007-8-1-r13

Hampson SE, Gaut B, Baldi P (2004) Statistical detection of chromosomal homology using shared-gene density alone. Bioinformatics 21(8):1339–1348

Horvath DP, Sung S, Kim D, Chao W, Anderson J (2010) Characterization, expression and function of *DORMANCY ASSOCIATED MADS-BOX* genes from leafy spurge. Plant Mol Biol 73(1–2):169–179. https://doi.org/10.1007/s11103-009-9596-5

Hummer KE, Janick J (2009) Rosaceae: taxonomy, economic importance, genomics. In: Genetics and genomics of Rosaceae. Springer, pp 1–17

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449(7161):463–477. https://doi.org/10.1038/nature06148

Jiménez S, Lawton-Rauh AL, Reighard GL, Abbott AG, Bielenberg DG (2009) Phylogenetic analysis and molecular evolution of the dormancy associated MADS-box genes from peach. BMC Plant Biol 9(1):81

Jiménez S, Reighard G, Bielenberg D (2010) Gene expression of *DAM5* and *DAM6* is suppressed by chilling temperatures and inversely correlated with bud break rate. Plant Mol Biol 73(1–2):157–167

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110(1–4):462–467

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28(1):27–30

Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome Res 12(4):656–664

Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z et al (2010) *De novo* assembly of human genomes with massively parallel short read sequencing. Genome Res 20(2):265–272

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25(5):955

Majoros WH, Pertea M, Salzberg SL (2004) TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20(16):2878–2879

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res 17(2):240–248

Pichersky E, Dudareva N (2007) Scent engineering: toward the goal of controlling how flowers smell. Trends Biotechnol 25(3):105–110

Salamov AA, Solovyev VV (2000) Ab initio gene finding in Drosophila genomic DNA. Genome Res 10(4):516–522

Salse J, Abrouk M, Murat F, Quraishi UM, Feuillet C, (2009) Improved criteria and comparative genomics tool provide new insights into grass paleogenomics. Brief Bioinform 10(6):619–630

Sasaki R, Yamane H, Ooka T, Jotatsu H, Kitamura Y, Akagi T et al (2011) Functional and expressional analyses of *PmDAM* genes associated with endodormancy in Japanese apricot. Plant Physiol 157(1):485–497

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL et al (2011) The genome of woodland strawberry (*Fragaria vesca*). Nat Genet 43(2):109

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B (2006) AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic acids Res 34 (suppl_2):W435–W439

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24(8):1596–1599

Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Res 080978:108

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25(9):1105–1111

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ et al (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28(5):511

Van Ooijen JW, Voorrips R (2001) JoinMap® 3.0, software for the calculation of genetic linkage maps. Plant Research International, Wageningen, Netherlands, pp 1–51

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, et al (2010) The genome of the domesticated apple (*Malus × domestica Borkh*.). Nat Genet 42(10):833

Wisniewski M, Norelli J, Bassett C, Artlip T, Macarisin D (2011) Ectopic expression of a novel peach (*Prunus persica*) CBF transcription factor in apple (*Malus × domestica*) results in short-day induced dormancy and increased cold hardiness. Planta 233(5):971–983

Xing Z, Suo L, Liu H, Zhang Q-X (2009) The germplasm resources survey of *Prunus mume* at Tangmai. Northern Hort 10:46 (in Chinese)

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24(8):1586–1591

Zdobnov EM, Apweiler R (2001) InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17(9):847–848

Zhou S, Kile A, Bechner M, Place M, Kvikstad E, Deng W et al (2004) Single-molecule approach to bacterial genomic comparisons via optical mapping. J Bacteriol 186(22):7773–7782. https://doi.org/10.1128/JB.186.22.7773-7782.2004

Molecular Mapping and Gene Cloning of QTLs in *Prunus mume*

Zhihong Gao and Xiaopeng Ni

Abstract

Prunus is an economically important fruit tree genus and includes many stone fruit trees such as peach (*Prunus persica* Batsch), apricot, Japanese apricot (*Prunus mume*), sweet cherry (*Prunus avium* L.), European plum (*Prunus domestica* L.), Japanese plum (*Prunus salicina* Lindl.) and almond [*Prunus dulcis* (Mill.) D. A. Webb.]. The characterisation of genes associated with agriculturally important traits such as fruit ripening, dormancy, self-incompatibility, fruit quality and various other developmental processes is important to improve *Prunus* breeding programmes. Based on high-density molecular genetic maps, many genes are located on the genome, and subsequently, most of them were fine-mapped and further identified by positional cloning. Currently, with the availability of the *P. mume* genome sequence, the identification of new genes is significantly accelerated.

Z. Gao (✉) · X. Ni
College of Horticulture, Nanjing Agricultural University, No 1 Weigang, Nanjing 210095, People's Republic of China
e-mail: gaozhihong@njau.edu.cn

6.1 Genes Cloning Involved in Morphological and Fruit Characteristics

6.1.1 Introduction

Japanese apricot (*Prunus mume* Sieb. et Zucc.) is a climacteric fruit and produces large amounts of ethylene upon ripening. Ripening is accompanied by marked increases in the activities of two ethylene-biosynthetic enzymes, namely 1-amino cyclopropane-1-carboxylic acid (ACC) synthase and ACC oxidase. Fifteen complementary DNA clones corresponding to messenger RNAs differentially expressed in the pericarp of *P. mume* fruit in response to ripening, ethylene and wounding signals were isolated by differential display (Mita et al. 2006). Japanese apricot (*P. mume* Sieb. et Zucc.) exhibits *S*-RNase-based gametophytic self-incompatibility (GSI), as do other self-incompatible rosaceous fruit trees (Gao et al. 2012). In 2000, the first common *S*-RNase allele was PCR-amplified by oligonucleotide primers designed from conserved regions of Prunus *S*-RNase isolated from five self-incompatible and six self-compatible cultivars and used as a molecular marker for self-compatibility (Tao et al. 2000). Bud dormancy allows most deciduous fruit tree species to avoid injury in unsuitable environments, to synchronise their annual growth and to adapt to a temperate climatic zone (Yamane et al. 2008). To understand the molecular basis of the endodormancy

of buds of perennial plants, (Yamane et al. 2008) searched for the genes that are expressed preferentially in endodormant lateral buds of deciduous fruit trees (in this case, *P. mume*).

To construct genetic linkage maps, DNA markers, such as randomly amplified polymorphic DNAs (RAPDs) (Shimada et al. 1994), amplified fragment length polymorphisms (AFLPs) (Vos et al. 1995) or simple sequence repeats (SSRs) (Weber and May 1989) have been used for several fruit tree species. A genotyping-by-sequencing (GBS) technique, based on single nucleotide polymorphisms (SNPs) obtained from high-throughput next-generation sequencing techniques (e.g. the Illumina HiSeq platform), has recently been developed (Sun et al. 2013). This technique is a highly efficient and cost-effective way to construct a genome-wide high-density linkage map. The GBS approach is suitable even for highly heterozygous plant species, such as fruit trees, and has already been applied to develop DNA markers associated with disease resistance and fermentation characteristics in grapevine (*Vitis vinifera* L.) (Barba et al. 2014) and fruit skin colour in apple (Gardner et al. 2014). The first genome-wide mapping study focused on quantitative trait loci (QTLs) that affect stem growth and form, leaf morphology and leaf anatomy in an intraspecific cross derived from two different mei cultivars (Sun et al. 2014). The first high-density genetic map of mei (*P. mume*) was developed by specific locus amplified fragment sequencing (SLAF-seq) (Zhang et al. 2015). Bud dormancy is an important developmental stage affecting blooming date and leafing date (LD) in Japanese apricot (*P. mume*), but the genetic factors controlling the chilling requirement (CR) and heat requirement (HR) for dormancy release and bud burst time remain to be elucidated. Here, a quantitative trait locus (QTL) analysis using two F_1 segregating populations was conducted to identify loci affecting these traits (Kitamura et al. 2018).

6.1.2 Fruit Ripening in Japanese Apricot

Fruit species are typically defined in terms of two ripening types, climacteric and non-climacteric, where the earlier burst in respiration at the onset of ripening, in contrast to the latter. Non-climacteric fruits, including strawberries, grapes and citrus fruits, do not require climacteric respiration or increased ethylene levels at maturation. Climacteric fruit typically increases the biosynthesis of the gaseous hormone ethylene, required for as the ripening of fruit such as tomatoes, bananas, apples, pears and most stone fruit, including Japanese apricot (Vrebalov et al. 2002). Japanese apricot (*P. mume* Sieb. et Zucc.) is a climacteric fruit that produces large amounts of ethylene as it ripens. Ripening is accompanied by marked increases in the activities of two ethylene-biosynthetic enzymes, namely 1-aminocyclopropane-1-carboxylic acid (ACC) synthase and ACC oxidase. Mita et al. (1999) isolated cDNA clones for proteins that were involved in the biosynthesis and perception of ethylene during ripening, namely ACC synthase, ACC oxidase and the ethylene receptor. Northern blotting analysis revealed a markedly increased expression of ACC synthase prior to that of ACC oxidase and the increase in ethylene production during ripening. Overall, the levels of the mRNAs for the genes corresponded closely to the activity levels of ethylene-biosynthetic enzymes. Exposure to mature green *mume* fruit to ethylene for 12 h induced a strong expression of ACC synthase and ACC oxidase. Wounding of the pericarp of *mume* fruit induced the expression of ACC synthase, but not that of ACC oxidase. The rate of ethylene production increased slightly after wounding. These results suggest that the expression of genes for ACC synthase and ACC oxidase must be activated sequentially for a maximum production of ethylene during ripening of *mume* fruit, and several mechanisms are regulated by the expression of ethylene-biosynthetic genes during ripening.

Fifteen complementary DNA clones, corresponding to mRNAs differentially expressed in the pericarp of *P. mume* fruit in response to ripening, ethylene and wounding signals, were isolated by differential displays (Mita et al. 2006). Quantitative real-time PCR analysis distinctly showed that these genes are differentially regulated. Genes that were up-regulated during fruit ripening include *Pm15* (cinnamyl-alcohol dehydrogenase), *Pm21* (2-oxoacid-dependent dioxygenase), *Pm22* (1-acyl-sn-glycerol-3-phosphate acyltransferase), *Pm27* (unknown function), *Pm38* (alcohol dehydrogenase), *Pm41* (no homology), *Pm52* (no homology), *Pm65* (pectate lyase), *Pm68* (expansion), *Pm69* (serine carboxypeptidase) and *Pm94* (alcohol acyltransferase). The expression of most of these genes was also inducible by ethylene, and some of them were inducible by wounding. Genes *Pm3* (water channel protein, MIP) and *Pm8* (unknown function) were down-regulated during ripening. The expression of *Pm71* (no homology) and *Pm74* (NAC family protein) did not increase during ripening or in response to ethylene but was up-regulated in response to wounding. The possible physiological roles of these genes during ripening and in response to ethylene and wounding are discussed.

6.1.3 Gametophytic Self-incompatibility and S-RNase Genotypes

Japanese apricot (*P. mume* Sieb. et Zucc.) exhibits *S*-RNase-based gametophytic self-incompatibility (GSI), as do other self-incompatible rosaceous fruit trees. A single multi-allelic gene, called S-locus, controls the GSI. The *S*-locus in the pistils produces basic glycoproteins with ribonuclease (*S*-RNase) activity. The *S*-genotype of each cultivar needs to be determined in order to select a compatible pollinator and to promote breeding. The first common *S*-RNase allele was PCR-amplified by oligonucleotide primers designed from conserved regions of Prunus *S*-RNase isolated from five self-incompatible and six self-compatible cultivars in 2000 and used as a molecular marker for self-compatibility (Tao et al. 2000). Yaegaki et al. (2001) cloned three partial *S*-RNase genes, *MSRN-1, MSRN-2, and MSRN-3*, using the RT-PCR method in Japanese apricot and established *S*-genotypes of six representative cultivars (S_1S_7, S_2S_6, S_3S_4, S_3S_6, S_3S_6 and S_1S_5), which contribute to an efficient breeding programme through an extended elucidation of the *S*-genotype. Entani et al. (2003) cloned an S_9-*RNase gene* of a novel *S*-haplotype by RT-PCR (GenBank accession no. AB092646), located within the highly divergent genomic region of the *S*-locus and exhibiting *S*-haplotype-specific diversity among analysed *S*-haplotypes before and specifically expressed in pollen, but not in styles or leaves. Subsequently, Habu et al. (2008) identified a novel S_{11}-RNase, previously misidentified as S_6-RNase. In addition, the first and second intron sizes of S_1- to S_{11}- and S_f-RNases were presented. Since the differences among the second intron sizes of S_3-, S_9- and S_{10}-RNases, as well as those among S_5-, S_6-and S_{11}-RNases, appeared to be very small, primers were developed for allele-specific PCR amplification to distinguish *S*-RNase alleles; finally, three S-RNase genotypes were updated (S_6S_{10}, S_6S_{10} and $S_{11}S_f$). The first *S*-genotypes of native Chinese cultivars of *P. mume* have been reported by (Heng et al. 2008). Eleven Chinese cultivars were analysed, and the authors identified seven new *S*-alleles (PmS_{10}-*R*Nase to PmS_{16}-RNase), deposited in GenBank under the accession numbers DQ011150, DQ201191, DQ201192, DQ345781, DQ768219, DQ903312 and EF990750 and their corresponding cDNA sequences. Ten new *S*-RNase alleles (S_{17}-S_{26}) were amplified by allele-specific polymerase chain reaction (AS-PCR), and field-testing cross-pollination was used to confirm the *S*-genotypes of four Japanese apricot cultivars, which demonstrated high polymorphism of *S*-RNases in Japanese apricot (Xu et al. 2010). More *S*-genotypes of many cultivars were determined over time, eleven *S-RNase* alleles (S_1, S_2, S_7, S_{12}, S_{14}, S_{15}, S_{18}, S_{20}, S_{22}, S_{23} and S_{26}) and four new *S-RNase* alleles (S_{30}, S_{31}, S_{32} and S_{33}), with GenBank Accession Numbers JN232975,

JN232976, JN232977, JN232978, were identified (Wang et al. 2012). Based on the PCR method, (Gao et al. 2012) identified three new *S*-RNase genes and six new *SFB* genes from ten Japanese apricot cultivars native to China, using polyacrylamide gel electrophoresis (PAGE) and agarose gel electrophoresis, respectively. However, PAGE allowed a better separation and resolution than agarose gel electrophoresis; the identification of *S*-genotypes facilitated the production and breeding of Japanese apricot cultivars. Up to date, the S-RNase gene has been identified, and 36 genes including $S_{1\text{-}26}$, $S_{30\text{-}36}$, S_f-RNase, S_9-RNase (Entani) and S_{10}, have been registered (Heng et al. 2008).

The study of *SFB/SLF* genes in Japanese apricot pollen was performed later than that of the *S*-RNase gene. Yamane et al. (2003) identified the *S*-haplotype-specific F-box protein gene (*SFB*), a candidate gene for pollen S of Japanese apricot, $Pm\text{-}SFB^1$ and $Pm\text{-}SFB^7$ (GenBank accession numbers AB101440, AB101441) were cloned by 5′ and 3′ rapid amplification of cDNA ends (RACE), showing a high level of S-haplotype-specific sequence polymorphism. Their expression was specific to pollen, leading to the development of a molecular typing system for the S-haplotype in Japanese apricot. Entani et al. (2003) investigated the genomic structure of the *S*-locus region of S_1- and S_7-haplotypes of *P. mume* and identified 13 genes nearby the *S*-RNase gene. Among them, three *F-box* genes, named *S*-locus F-box (*SLF*), were confirmed, namely SLF^1, SLF^7 and SLF^9 (GenBank accession numbers AB092621, AB092622 and AB092645. In addition, (Ushijima et al. 2004) demonstrated that a deletion in $SFB^{4'}$ and an insertion in SFB^f lead to transcripts for $SFB^{4'}$ and SFB^f that lack the hypervariable regions HVa and HVb, which are suspected to be essential for the haplotype-specific interaction between *SFB* and *S*-RNase. As all *SFBs* identified in functional S-haplotypes of *Prunus* are intact, the fact that the partial loss-of-function mutations in $SFB^{4'}$ and SFB^f coincide with SC provides additional evidence that *SFB* is the pollen S gene in the

S-RNase-based GSI system of *Prunus*. The 1K0-26 is an intact SI S^7 haplotype, and a pollen-part mutant SC $S^{3'}$ haplotype through controlled pollination and segregation analyses. Cloning and DNA sequence analysis of the $S^{3'}$ locus revealed a 7.1-kb insertion in the pollen determinant $SFB^{3'}$, which makes a premature stop codon to produce transcripts for truncated dysfunctional *SFB*, just like the mode of mutation in SFB^f, another SC S-haplotype in Japanese apricot (Yamane et al. 2009). Heng et al. (2012) identified and characterised seven novel S-haplotype-specific F-box protein genes, $PmSFB^{10}$ to $PmSFB^{16}$, in *P. mume*, whose similarities among the deduced amino acid sequences of these *SFBs* ranged from 73.2 to 90.9%. Wang et al. (2013) isolated five new *SFB* alleles ($PmSFB^2$ (JQ356589), $PmSFB^{12}$ (JQ356586), $PmSFB^{15}$, $PmSFB^{40}$ (JQ356585), $PmSFB^{41}$ (JQ356593), $PmSFB^{42}$ (JQ356581) and $PmSFB^{43}$ (JQ356578)) from six Japanese apricot (*P. mume*) lines, using a specific *Prunus SFB* primer pair (SFB-C1F and Pm-Vb), which was designed from conserved regions of *Prunus SFB*. The five new *SFB* alleles share typical structural features with *SFB* alleles from other *Prunus* species and are polymorphic, with 67.08–96.91% amino acid identity. In addition, six new *SFB* genes ($PmSFB^{14}$, $PmSFB^{18}$, $PmSFB^{22}$, $PmSFB^{24}$, $PmSFB^{31}$ and $PmSFB^{34}$) were identified by polyacrylamide gel electrophoresis (PAGE) (Gao et al. 2012). So far, 15 *SFB* genes have been identified and registered, namely $PmSFB^1$, $PmSFB^2$, $PmSFB^3$, $PmSFB^7$, $PmSFB^{12}$, $PmSFB^{14}$, $PmSFB^{18}$, $PmSFB^{22}$, $PmSFB^{24}$, $PmSFB^{31}$, $PmSFB^{34}$, $PmSFB^{40}$, $PmSFB^{41}$, $PmSFB^{42}$, $PmSFB^{43}$, along with three *SLY* genes, $PmSLF^1$, $PmSLF^7$ and $PmSLF^9$, *and* two self-compatibility genes, $PmSFB^{3'}$ and $PmSFB^f$.

In studies on self-incompatibility gene identification, PCR technology is the most commonly used approach. So far, 36 *S*-RNase genes, 15 *SFB* genes, three *SLY* genes and two self-compatible genes have been identified and registered, which proves the feasibility of identifying genotypes based on the specificity of gene

sequences. Improving the genotypic identification method and developing new identification techniques can accelerate the development of Japanese apricot genotypes.

6.1.4 Bud Dormancy in Japanese Apricot

Bud dormancy allows most deciduous fruit tree species to avoid injury in an incompatible environment, to synchronise their annual growth and to adapt to a temperate climatic zone (Yamane et al. 2008). Dormancy states can be classified as follows: paradormancy, endodormancy and ecodormancy. Both endodormancy and paradormancy can be defined as a state induced by the perception of the promoting environmental or endogenous signalling cues, whether this originated solely within the meristem-containing tissue (endodormant) or in a structure distinct from the structure undergoing dormancy (paradormant) (Yamane 2014).

Genetic and molecular regulation of bud dormancy has been extensively studied in the model woody plant poplar (*Populus* spp.), with significant progress (Cooke et al. 2012; Rinne et al. 2009; Ruttink et al. 2007). Early biochemical studies on *Prunus* bud dormancy regulation have investigated seasonal carbohydrate concentration changes and absorption potentials (Marquat et al. 1999), while functional genomics could promote the discovery of gene function and identify gene networks associated with bud dormancy regulation at the transcriptional level on a genome-wide basis. In addition, the use of functional genomics can be useful in breeding, as functional genomics approaches can be used to generate robust molecular markers for bud dormancy traits.

To understand the molecular basis of endodormancy of buds of perennial plants, (Yamane et al. 2008) searched for the genes preferentially expressed in endodormant lateral buds of *P. mume*. This work identified a *MADS-box* gene with dormancy-associated expression. Seasonal expression analysis suggested that the gene was up-regulated during bud dormancy progression and down-regulated during dormancy release. Full-length cDNA cloning of the *MADS-box* gene and phylogenetic analysis revealed that the gene was similar to the *StMADS11* clade *MADS-box* genes of *Arabidopsis*, such as *SHORT VEGETATIVE PHASE* (*SVP*) and *AGAMOUS-LIKE24* (*AGL24*) (Yamane et al. 2008). Sequencing and expression analysis of the *evg* locus identified six *StMADS11* (*SVP/AGL24*)-clade MADS-box genes as candidate genes associated with terminal bud formation in peach (Bielenberg et al. 2008). These were named *DORMANCY-ASSOCIATED MADS-box 1–6* (*DAM1–6*) genes.

The *CBF* gene was isolated from *P. mume* by PCR and RT-qPCR, using primers designed according to the CBF transcription factors of peach and sweet cherry from GenBank (Zhang et al. 2012a). The *CBF* gene contains an AP2/EREB DNA-binding domain and two special short amino acid sequences; experiments indicated *PmCBF1* was induced under low-temperature stress.

Kitamura et al. (2016) performed a seasonal expression analysis of *PmDAM1–6* genes in flower buds of two Japanese apricot genotypes, namely high-chill and low-chill cultivars. The analysis revealed that *PmDAM3*, *PmDAM5* and *PmDAM6* expressions are closely associated with dormancy release in both flower and vegetative buds. In addition, a yeast two-hybrid screening demonstrated that *PmDAM6* could interact in yeast with the homolog of *Arabidopsis* SOC1 (*PmSOC1*). Synchronised expression patterns were detected in *PmDAM6* and *PmSOC1* during dormancy release in flower buds of the two genotypes. Taken together, these results suggest that the dimer of *PmDAM6* and *PmSOC1* may play a role in the regulation of dormancy transition and blooming time in Japanese apricot flower buds.

Wen et al. (2016) measured the concentrations of these two hormones in flower buds (collected during para-, endo- and ecodormancy and at the dormancy-release stage) and vegetative buds of *P. mume*, using liquid chromatography and tandem mass spectrometry (LC–MS/MS). Levels of expression of two zinc-finger family genes

CCCH and C2H2 and the GA20ox as candidates for dormancy-controlling genes were analysed by RT-qPCR; expression analyses indicated that both zinc-finger family genes exhibited high transcript levels in flower buds of both cultivars during paradormancy, concomitant with high ABA concentrations. In addition, high concentrations of GA3 in flower buds at the dormancy-release stage suggested that GA3 plays an important role in controlling the release of bud dormancy. High levels of expression of the GA20ox gene during ecodormancy and at the dormancy-release stage might have a prominent effect on dormancy release by regulating the gibberellin (GA) signal in Japanese apricot. These results will contribute to our understanding of ABA- and GA-mediated seasonal dormancy mechanisms in Japanese apricot.

In the GA signalling pathway, DELLA proteins are highly conserved and act as negative regulators. Lv et al. (2018) established a relationship between PmRGL2 in Japanese apricot and GA₄ levels during dormancy release of floral buds. Overexpression of PmRGL2 in poplar delayed the onset of bud dormancy and resulted in dwarf plants. Gene PmRGL2 exhibited a higher expression during ecodormancy and a relatively lower expression during endodormancy. The relative level of GA_4 showed an increasing trend at the transition from endodormancy to ecodormancy and displayed a similar expression pattern of genes related to GA metabolism, namely PmGA20ox2, PmGA3ox1 and PmGID1b, in both Japanese apricot and transgenic poplar. This study demonstrated that PmRGL2 plays a negative role in bud dormancy release by regulating the GA biosynthetic enzymes, GA20ox and GA3ox1 and the GA receptor, GID1b.

To provide suggestions for further explorations of protein-complex functions in association with bud growth and dormancy, six PmDAMs were cloned in P. mume and functionally analysed in yeast and tobacco to detect the roles of the genes paralogous to PmDAM6 (Zhao et al. 2018). The expression patterns, together with the sequence similarities, indicate that PmDAMs is divided into two sub-clades

within the SVP group. Moreover, PmDAMs takes part in the development of different plant organs, specifically the flower buds, in some intricate patterns. Furthermore, the PmDAM proteins have specific functions by forming the corresponding protein complex during the development of flower bud and induction of dormancy. In particular, when PmDAM1 dominates in flower buds in hot summer months, the protein complexes consist of PmDAM1 itself or with PmDAM2. With the decrease in temperatures in the following months, PmDAM6 is highly expressed, and the complex structure is gradually changed to the PmDAM6-protein complex due to strong binding tendencies with PmDAM1 and PmDAM3. Finally, homodimers of PmDAM6 prevailed to induce dormancy. The results obtained in the current study highlight the functions of PmDAMs in tissue development and dormancy, thereby providing suggestions for further explorations of protein-complex functions in association with bud growth and dormancy.

To evaluate their integral roles in bud development, PmDAM6 and six PmCBFs were cloned (Zhao et al. 2018). The consistency of expressions in either vegetative or reproductive buds provided a negative control from PmCBFs to PmDAM6 during the onset of dormancy. Besides, PmCBF5 could form heteromeric complexes with PmDAM1 and PmDAM6. Also, PmCBF1, PmCBF3 and PmDAM4 recognised the promoter of PmDAM6 by the alternative binding sites. Therefore, the interactions of these genes provide the base of an obvious model to respond to the coldness and engendered dormancy release. These findings will further facilitate the genetic control of bud dormancy and its increase in P. mume and may explain vernalisation.

6.2 Molecular Genetic Mapping and QTLs

6.2.1 RFLP-Based Genetic Maps

Yamane et al. (2003) developed SFB as a RFLP molecular marker for S-haplotypes and self-compatibility in Japanese apricot. Since

pistil and pollen *S*-alleles are tightly linked to each other as if they were the same single gene, the use of molecular markers for both pistil and pollen determinants would lead to a firm determination of S-haplotypes, especially when the S-RNase genes from different *S*-haplotypes provide the same PCR and RFLP bands (Tao et al.).

6.2.2 Genetic Maps Based on RAPD and AFLP Markers

The RAPD assay is applied in cultivar classification and identification and in the detection of mapping markers. The DNA polymorphisms and phylogenetic relationship by the RAPD assay, using genomic DNA of *P. mume* cultivars, classified those cultivars by fruit size and morphological traits, with a high detection rate of polymorphisms in their genome. In addition, this procedure is simpler than isozyme analysis (Shimada et al. 1994).

Over 14 years, 42 RAPD markers were successfully used to amplify genotypes into five groups due to the selection according to their characteristics for peach rootstock, instead of the limit morphologic characteristics, showed the genetic similarity is bigger between 05 Clone and 15 Clone (Mayer et al. 2008).

Amplified fragment length polymorphism (AFLP) is a polymerase chain reaction (PCR) based on markers that have been reliably used for genome mapping and genotyping (Vos et al. 1995). The AFLP markers have, for example, been used to study the genetic diversity in many fruit crops such as peach (*Prunus persica* L.) (Dirlewanger et al. 1998) and Japanese apricot (Fang et al. 2005). Fang et al. (2005) used amplified fragment length polymorphism (AFLP) markers to analyse 14 fruiting mei cultivars from China and Japan. All three dendrograms, generated by three sets of data, showed similar relationships among the fruiting mei cultivars. The corresponding main clusters contained the same cultivars, and the subgroups were closely related with the known geographic origins of the cultivars.

In studies using amplified fragment length polymorphism (AFLP) markers, 121 distinct genotypes of mei cultivars were discriminated from other species in the genetic stem, and the genetic diversity of mei accessions was highest in the Yunnan Province and decreased towards Eastern China and Japan (Yang et al. 2008).

6.2.3 Simple Sequence Repeat (SSR) or Microsatellite Markers

Recently, SSR markers have enabled a more reliable evaluation of genetic diversity and the construction of genetic maps because of the co-dominant inheritance and abundance of alleles per locus (Weber and May 1989). Gao et al. (2004) performed sequencing amplification fragments using simple sequence repeat (SSR) primer pairs and showed that the SSRs obtained included a microsatellite locus originally identified in peach. The microsatellite sequence of UDP96008 homogeneity between Japanese apricot and peach in GenBank was 98%. Twenty-four Japanese apricot genotypes, from diverse geographic areas, could be identified with 14 SSR primer pairs developed in different species of *Prunus*, resulting in a total of 129 alleles with an average of 2.5 alleles per primer pair.

Fourteen SSR markers, showing clear amplification and high polymorphisms, were used to assess the genetic diversity and relationships among 127 Japanese apricots from four distances in Asia. Fruiting and flower-ornamental germplasms derived from Japan and China represent one cluster, which supports the hypothesis that fruiting cultivars have been selected from flower-ornamentals.

Li et al. (2010) randomly tested 42 pairs of PCR primers flanking potential SSRs in the *P. mume* unigene dataset, and 14 pairs were identified as true-to-type SSR loci and could amplify polymorphic bands from 20 individual plants of *P. mume*. Subsequently, 14 EST-SSR primer pairs were used to test the transferability on peach and plum. The result showed that nearly 89% of the primer pairs produced target PCR bands in the two species. A high level of

marker polymorphism was observed in plum species (65%) and a low level in peach (46%), and clustering analysis of the three species indicated that these SSR markers were useful in the evaluation of genetic relationships and diversity between and within *Prunus* species.

6.2.4   REMAP and IRAP

The REMAP with IRAP techniques was performed to identify genetic relationships via genetic diversity analysis (Shen et al. 2011). The results showed that the observed number of alleles was 1.9955 in Japanese apricot populations, the effective number of alleles was 1.4887, Nei's gene diversity was 0.2910 and Shannon's information index was 0.4465. Eighty-four Japanese apricot varieties were divided into 18 groups at the similarity coefficient of 0.743 and into four groups at the similarity coefficient of 0.705 with the UPGMA cluster method. Fruiting mei and flowering mei clustered into different groups, showing that the complexity of inheritance and the extent of evolution were different. The PopGen32 analysis suggested that Japanese fruiting mei was introduced from Zhejiang Province in China, while Japanese flowering mei originated from Jiangsu Province in China.

6.2.5   Single Nucleotide Polymorphisms (SNPs) and Insertion-Deletions

With the advent of NGS technologies, entire genomes have been sequenced more efficiently and economically than ever before (Zhang et al. 2012b). The alignment of short reads obtained from different varieties of mei to the reference genome has provided the perfect opportunity to identify a large number of polymorphic DNA markers in parallel, including SNPs and InDels.

Numerous InDels have been generated using the NGS platform between two cultivars, using low-depth whole-genome sequencing (Sun et al. 2013). However, due to longer than expected insertions or deletions of three polymorphic InDels, a short sequence read including an InDel might be aligned with mismatched bases instead of gaps (Krawitz et al. 2010). A BWA short-read mapping tool was used to generate a high rate of variant bases at InDel positions. Thus, the mismatched InDels might be attributed to alignment with mismatched bases instead of gaps. Fifteen InDels were found useful for constructing high-resolution genetic maps, performing genome-wide association studies and designing genomic selection strategies in Japanese apricot.

Single nucleotide polymorphism (SNP) analysis represents the most effective approach because these markers are evenly distributed along the genome, highly numerous, highly polymorphic, co-dominant and have a potential functional effect. Genotyping-by-sequencing (GBS), an SNP identification method based on next-generation sequencing (NGS) technologies, is useful for the identification of a high number of SNP markers and the construction of high-density genetic linkage maps. In addition to a massive number of SNPs, InDels have also become valuable DNA markers and are applied in QTL mapping and marker-assisted selection because of their lower costs, relatively simple genotyping and easy transferability between populations (Hayashi et al. 2006; Pan et al. 2008). Using the Illumina Hiseq 2000 platform, 322 large-effect SNPs and 433 large-effect InDels were detected (Zhang et al. 2017), including 90 SNPs chosen randomly for validation using high-resolution melt analysis. The PFAM analysis was further conducted to better understand SNP effects on gene functions and two known QTL loci conferring the weeping trait, and their functional effects were analysed. This study highlights promising functional markers for molecular breeding and a whole-genome genetic basis of the weeping trait in Japanese apricot.

## 6.2.6   QTLs in *Prunus mume*

The first genome-wide mapping study of quantitative trait loci (QTLs) that affect stem growth and form leaf morphology and leaf anatomy in an

intraspecific cross was based on two different mei cultivars (Sun et al. 2014). Genetic mapping based on a high-density linkage map constructed from 120 SSRs and 1,484 SNPs led to the detection of multiple QTLs for each trait, some of which exert pleiotropic effects on correlative traits. Each QTL explains 3–12% of the phenotypic variance. Several leaf size traits share common QTLs, whereas growth-related traits and plant form traits might be controlled by a different set of QTLs. These findings provide unique insights into the genetic control of tree growth and architecture in mei and help to develop an efficient breeding programme for selecting superior mei cultivars.

A high-density genetic map is a valuable tool for fine-mapping loci controlling a specific trait, especially for perennial woody plants. The first high-density genetic map of mei (*P. mume*), developed by specific locus amplified fragment sequencing (SLAF-seq), has already been constructed (Zhang et al. 2015). The linkage map contains 8,007 markers, with a mean marker distance of 0.195 cM, making it the densest genetic map for the genus *Prunus*. Although weeping trees are widely used as landscape plants around

Fig. 6.1 Distribution of tightly linked markers and plausible candidate genes of weeping on pseudo-chromosome 7 of mei. **a** Candidate genomic region on pseudo-chromosome 7. **b** Schematic diagram of the order of the ten tightly linked markers and the 18 plausible candidate genes of weeping, based on gene annotation. This figure is available in black and white in print and in colour at DNA Research online
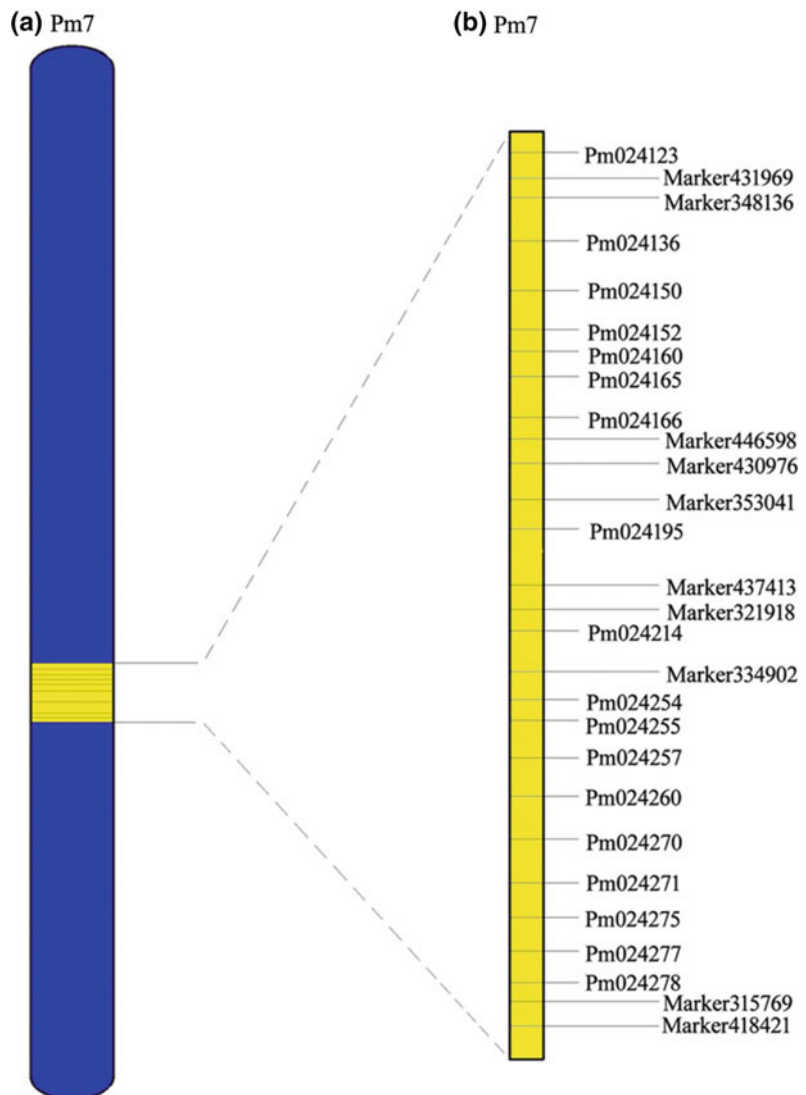
the world, little is known about the weeping controlling gene(s) (Pl). To test the utility of the high-density genetic map, Zhang et al. (2015) performed a fine-scale mapping of this important ornamental trait. In total, three statistic methods were performed progressively based on the result of inheritance analysis. Quantitative trait loci (QTL) analysis initially revealed that a locus on linkage group 7 was strongly responsible for the weeping trait. Mutmap-like strategy and extreme linkage analysis were then applied to fine-map this locus within 1.14 cM. Bioinformatics analysis of the locus identified some candidate genes. The successful localisation of weeping trait strongly indicates that the high-density map constructed using SLAF markers is a worthy reference for mapping important traits of woody plants (Fig. 6.1).

Bud dormancy is an important developmental stage affecting blooming date and leafing date (LD) in Japanese apricot (*P. mume*), but the

genetic factors controlling the chilling requirement (CR) and heat requirement (HR) for dormancy release and bud burst time remain to be elucidated. Here, a quantitative trait locus (QTL) analysis using two F_1 segregating populations was conducted to identify the loci affecting these traits (Kitamura et al. 2018). The genotyping-by-sequencing technique was used to construct two high-density genetic maps, one for NKSC, a population derived from high-chill 'Nanko' crossed with low-chill 'SC', covering 660.2 cM with 408 markers, and one for NINK, a population derived from low-chill 'Ellching' crossed with 'Nanko', covering 1,314.2 cM with 718 markers. We observed four traits: CR and HR for dormancy release, blooming date and LD over several years. To identify the QTL controlling, the down-regulation of DORMANCY-ASSOCIATED MADS-box6 (*PmDAM6*) in January leaf buds, in which PmDAM6 could act as a dose-dependent inhibitor of bud break, its
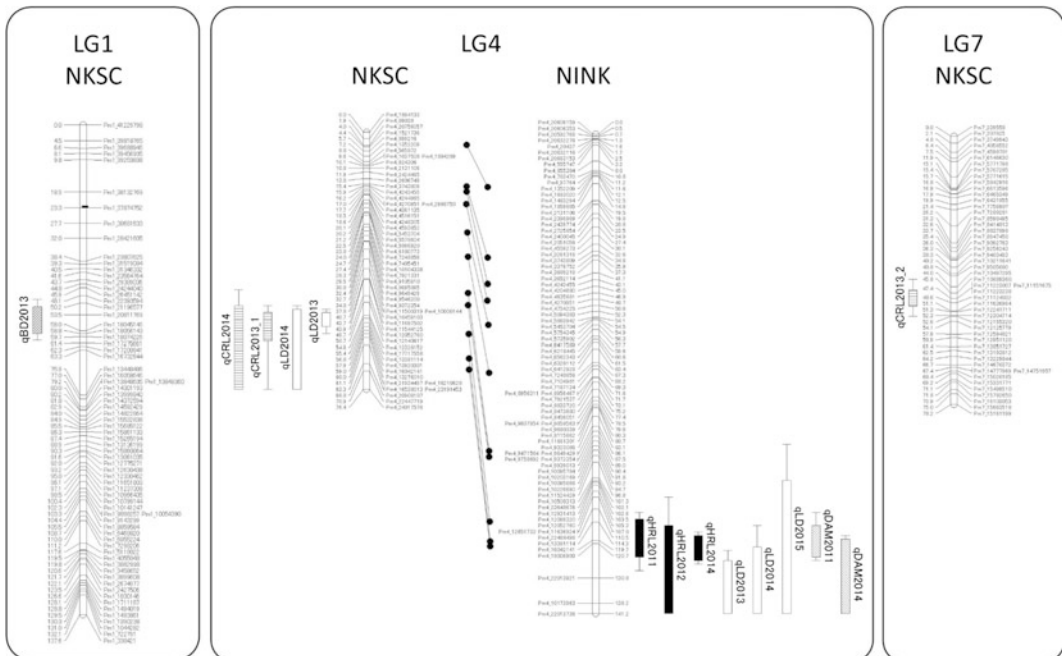

Fig. 6.2 Location of QTLs detected in the genetic linkage maps of NKSC and NINK. Boxes and bars indicate 1- and 1.5-LOD confidence intervals from peak LOD values, respectively. Black, white, striped, diagonal checkered and shaded boxes indicate QTLs controlling heat requirement for bud break of leaf buds (HRL), leafing date (LD), chilling requirement of leaf buds (CRL), blooming date (BD) and relative *PmDAM6* expression level (DAM), respectively

transcript levels in leaf buds were determined. All traits segregated in the analysed seasons in both populations. For leaf bud dormancy, CR and LD were highly correlated across years and traits in the NKSC population, while HR, LD and *PmDAM6* expression were highly correlated in the NINK population. The QTL analyses localised the significant QTLs controlling leaf bud CR and HR, LD and *PmDAM6* expression in leaf buds to a region in linkage group 4, which suggests that this locus controls dormancy release, bud break and PmDAM6 down-regulation in Japanese apricot leaf buds (Fig. 6.2).

References

Barba P, Cadle-Davidson L, Harriman J, Glaubitz JC, Brooks S, Hyma K et al (2014) Grapevine powdery mildew resistance and susceptibility loci identified on a high-resolution SNP map. Theor Appl Genet 127 (1):73–84

Bielenberg DG, Wang YE, Li Z, Zhebentyayeva T, Fan S, Reighard GL et al (2008) Sequencing and annotation of the evergrowing locus in peach [Prunus persica (L.) Batsch] reveals a cluster of six MADS-box transcription factors as candidate genes for regulation of terminal bud formation. Tree Genet Genomes 4 (3):495–507

Cooke JEK, Eriksson ME, Junttila O (2012) The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. Plant, Cell Environ 35(10):1707–1728

Dirlewanger E, Pronier V, Parvery C, Rothan C, Guye A, Monet R (1998) Genetic linkage map of peach [Prunus persica (L.) Batsch] using morphological and molecular markers. Theor Appl Genet 97(5–6):888–895

Entani T, Iwano M, Shiba H, Che FS, Isogai A, Takayama S (2003) Comparative analysis of the self-incompatibility (S-) locus region of Prunus mume: identification of a pollen-expressed F-box gene with allelic diversity. Genes Cells 8(3):203–213

Fang J, Qiao Y, Zhang Z, Chao CT (2005) Genotyping fruiting mei (Prunus mume Sieb. et Zucc.) cultivars using amplified fragment-length polymorphism markers. Amer Soc Hort Sci 40(2):325–328

Gao Z, Wang P, Zhuang W, Zhen Z (2012) Sequence analysis of new S-RNase and SFB alleles in Japanese Apricot (Prunus mume). Plant Mol Biol Rep 31 (3):751–762

Gao ZH, Shen ZJ, Han ZH, Fang JG, Zhang YM, Zhang Z (2004) Microsatellite markers and genetic diversity in Japanese apricot (Prunus mume). HortScience 39(7):1571–1574

Gardner KM, Brown P, Cooke TF, Cann S, Costa F, Bustamante C et al (2014) Fast and cost-effective genetic mapping in apple using next-generation sequencing. G3: Genes Genomes Genetics 4 (9):1681–1687

Habu T, Matsumoto D, Fukuta K, Esumi T, Tao R, Yaegaki H et al (2008) Cloning and characterization of twelve S-RNase alleles in Japanese apricot (Prunus mume Sieb. et Zucc.). J Jpn Soc Hort Sci 77(4):374–381

Hayashi K, Yoshida H, Ashikawa I (2006) Development of PCR-based allele-specific and InDel marker sets for nine rice blast resistance genes. Theor Appl Genet 113 (2):251–260

Heng W, Wu HQ, Chen QX, Wu J, Huang SX, Zhang SL (2008) Identification of S-genotypes and novel S-RNasealleles in Prunus mume. J Hort Sci Biotechnol 83(6):689–694

Heng W, Wu J, Wu HQ, Tao ST, Qi KJ, Gu C et al (2012) Identification and characterisation of SFBs in Prunus mume. Plant Mol Biol Rep 30(4):878–884

Kitamura Y, Habu T, Yamane H, Nishiyama S, Kajita K, Sobue T et al (2018) Identification of QTLs controlling chilling and heat requirements for dormancy release and bud break in Japanese apricot (Prunus mume). Tree Genet Genomes 14(2):101

Kitamura Y, Takeuchi T, Yamane H, Tao R (2016) Simultaneous down-regulation of DORMANCY-ASSOCIATED MADS-box6 and SOC1 during dormancy release in Japanese apricot (Prunus mume) flower buds. J Hort Sci Biotechnol 91(5):476–482

Krawitz P, Rödelsperger C, Jäger M, Jostins L, Bauer S, Robinson PN (2010) Microindel detection in short-read sequence data. Bioinformatics 26(6):722–729

Li X, Shangguan L, Song C, Wang C, Gao Z, Yu H et al (2010) Analysis of expressed sequence tags from Prunus mume flower and fruit and development of simple sequence repeat markers. BMC Genet 11(1):66

Lv L, Huo X, Wen L, Gao Z, Khalil-Ur-Rehman M (2018) Isolation and role of PmRGL2 in GA-mediated floral bud dormancy release in Japanese apricot (Prunus mume Siebold et Zucc.). Front Plant Sci 9:27

Marquat C, Vandamme M, Gendraud M, Pétel G (1999) Dormancy in vegetative buds of peach: relationbetween carbohydrate absorption potentials and carbohydrate concentration in the bud during dormancy and its release. Sci Hortic 79:151–162. https://doi.org/10.1016/S0304-4238(98)00203-9

Mayer NA, Lemos EGD, Pereira FM, Wickert E (2008) Characterization of three mume genotypes (Prunus mume Sieb. et Zucc.) by RAPD markers. Reva Brasil Fruticult 30(4):1045–1050

Mita S, Kirita C, Kato M, Hyodo H (1999) Expression of ACC synthase is enhanced earlier than that of ACC oxidase during fruit ripening of mume (Prunus mume). Physiol Plant 107(3):319–328

Mita S, Nagai Y, Asai T (2006) Isolation of cDNA clones corresponding to genes differentially expressed in

pericarp of mume (*Prunus mume*) in response to ripening, ethylene and wounding signals. Physiol Plant 128(3):531–545

Pan C-H, Li A-H, Dai Z-Y, Zhang H-X, Liu G-Q, Wang Z-B et al (2008) InDel and SNP markers and their applications in map-based cloning of rice genes. Rice Sci 15(4):251–258

Rinne PLH, Welling A, van der Schoot C (2009) Perennial life style of populus: dormancy cycling and overwintering. Genet Genomics Populus, Springer, New York, NY, USA, pp 171-200. http://link.springer.com/10.1007/978-1-4419-1541-2_9

Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, Fromm J et al (2007) A molecular timetable for apical bud formation and dormancy induction in poplar. Plant Cell 19(8):2370–2390

Shen Y, Ding X, Wang F, Cai B, Gao Z, Zhang Z (2011) Analysis of genetic diversity in Japanese apricot (Prunus mume Sieb. et Zucc.) based on REMAP and IRAP molecular markers. Sci Hort 132:50–58

Shimada T, Haji T, Yamaguchi M, Takeda T, Nomura K, Yoshida M (1994) Classification of mume (*Prunus mume* Sieb. et Zucc.) by RAPD assay. J JPN Soc Hortic Sci 63(3):543–551

Sun L, Zhang Q, Xu Z, Yang W, Guo Y, Lu J et al (2013) Genome-wide DNA polymorphisms in two cultivars of mei (Prunus mume sieb. et zucc.). BMC Genet 14 (1):98

Sun LD, Wang YQ, Yan XL, Cheng TR, Ma KF, Yang WR et al (2014) Genetic control of juvenile growth and botanical architecture in an ornamental woody plant, Prunus mume Sieb. et Zucc. as revealed by a high-density linkage map. BMC Genet 15(1):S1

Tao R, Habu T, Yamane H, Sugiura A, Iwamoto K (2000) Molecular markers for self-compatibility in Japanese apricot (Prunus mume). HortScience 35(6):1121–1123

Ushijima K, Yamane H, Watari A, Kakehi E, Ikeda K, Hauck NR et al (2004) The S haplotype-specific F-box protein gene, SFB, is defective in self-compatible haplotypes of Prunus avium and P-mume. Plant J 39 (4):573–586

Vos P, Hogers R, Bleeker M, Reijans M, Lee TVD, Hornes M et al (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23(21):4407–4414

Vrebalov J, Ruezinsky D, Padmanabhan V, White R, Medrano D, Drake R et al (2002) A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (Rin) locus. Science 296 (5566):343–346

Wang PP, Gao ZH, Ni ZJ, Zhuang WB, Zhang Z (2013) Isolation and identification of new pollen-specific SFB genes in Japanese apricot (Prunus mume). Genet Mol Res 12(3):3286–3295

Wang PP, Shi T, Zhuang WB, Zhang Z, Gao ZH (2012) Determination of S-RNasegenotypes and isolation of four novel S-RNasegenes in Japanese apricot (*Prunus mume* Sieb. et Zucc.) native to China. J Hort Sci Biotechnol 87(3):266–270

Weber JL, May PE (1989) Abundant class of human dna polymorphisms which can be typed using the polymerase chain-reaction. Am J Hum Genet 44(3):388–396

Wen LH, Zhong WJ, Huo XM, Zhuang WB, Ni ZJ, Gao ZH (2016) Expression analysis of ABA- and GA-related genes during four stages of bud dormancy in Japanese apricot (Prunus mume Sieb. et Zucc). J Hort Sci Biotechnol 91(4):362–369

Xu JX, Gao ZH, Zhang Z (2010) Identification of S-genotypes and novel S-RNase alleles in Japanese apricot cultivars native to China. Sci Hort 123(4):459–463

Yaegaki H, Shimada T, Moriguchi T, Haji T, Yamaguchi M, Hayama H (2001) Molecular characterization of S-RNase genes and S-genotypes in the Japanese apricot Prunus mume Sieb. et Zucc.). Sexual Plant Reprod 13(5):251–257

Yamane H (2014) Regulation of bud dormancy and bud break in Japanese apricot (*Prunus mume* Siebold. et Zucc.) and Peach [*Prunus persica* (L.) Batsch]: a summary of recent studies. J Jpn Soc Hort Sci 83 (3):187–202

Yamane H, Fukuta K, Matsumoto D, Hanada T, Mei G, Esumi T et al (2009) Characterization of a novel self-compatible S3′ haplotype leads to the development of a Universal PCR marker for two distinctly originated self-compatible S haplotypes in Japanese Apricot (Prunus mume Sieb. et Zucc.). J Jpn Soc Hort Sci 78(1):40–48

Yamane H, Kashiwa Y, Ooka T, Tao R, Yonemori K (2008) Suppression subtractive hybridization and differential screening reveals endodormancy-associated expression of an SVP/AGL24-type MADS-box gene in lateral vegetative buds of japanese apricot. J Amer Soc Hort Sci 133 (5):708–716

Yamane H, Ushijima K, Sassa H, Tao R (2003) The use of the S haplotype-specific F-box protein gene, SFB, as a molecular marker for S-haplotypes and self-compatibility in Japanese apricot (Prunus mume). Theor Appl Genet 107(8):1357–1361

Yang CD, Zhang JW, Yan XL, Bao MZ (2008) Genetic relatedness and genetic diversity of ornamental mei (Prunus mume Sieb. et Zucc.) as analysed by AFLP markers. Tree Genet Genomes 4(2):255–262

Zhang J, Zhang Q, Cheng T, Yang W, Pan H, Zhong J et al (2015) High-density genetic map construction and identification of a locus controlling weeping trait

in an ornamental woody plant (Prunus mume Sieb. et Zucc). DNA Res 22(3):183–191. https://doi.org/10.1093/dnares/dsv003

Zhang J, Zhang Q, Yang W (2012a) Cloning and expression of CBF transcription factor from *Prunus mume*. Acta Bot Bor-Occ Sin 32(8):1505–1510

Zhang J, Zhao K, Hou D, Cai J, Zhang Q, Cheng T et al (2017) Genome-wide discovery of DNA polymorphisms in Mei (*Prunus mume* Sieb. et Zucc.), an ornamental woody plant, with contrasting tree architecture and their functional relevance for weeping trait. Plant Mol Biol Rep 35(1):37–46

Zhang Q, Chen W, Sun L, Zhao F, Huang B, Yang W et al (2012b) The genome of Prunus mume. Nat Commun 3(1):1318

Zhao K, Zhou Y, Ahmad S, Xu Z, Li Y, Yang W et al (2018) Comprehensive cloning of *Prunus mume* dormancy associated *MADS-Box* genes and their response in flower bud development and dormancy. Front Plant Sci 9:457

Systems Biology in Japanese Apricot

7

Zhihong Gao and Xiaopeng Ni

Abstract

'Omics' research approaches have produced copious data for living systems, required for the development of systems biology to integrate multidimensional biological information into networks and models. The application of systems biology to plant science has occurred rapidly and increased our knowledge about circadian rhythms, multigenic traits, stress responses and plant defences, along with advances in the virtual plant project. Here, we define systems biology as the study of interactions among biological components using models and/or networks to integrate genes, metabolites, proteins, regulatory elements and biological components. Both component integration and dynamic interactions are the key features for systems biology. There is an immediate need for systems integration in plant biology, considering the large datasets generated from different omics technologies such as genomics, proteomics, transcriptomics, interatomic and metabolomics. In this chapter, we provide a comprehensive review on recent advances in the integration of multidimensional data into networks and the application of these networks in addressing questions in Japanese apricot.

7.1 Introduction

Biological research is the integration of multiple disciplines, including physiological, morphological, molecular, biochemical and genetic information (Trewavas 2006). These genome-scale studies assimilate large amounts of data into biologically meaningful interpretations, contributing to the development of modern systems biology. Due to the important role of dynamic modelling or multidimensional data analysis, the definition of systems biology is not unique.

The contention of systems biology has been accepted by numerous researchers who focus on the networks of combined component integration and dynamic interactions with key features, integration in plant biology, considering the large datasets generated from different omics technologies such as genomics, proteomics, transcriptomics, interatomic and metabolomics (Yuan et al. 2008). Contrasting with the immense application potential of disciplines with large datasets, there has been less development of platforms to integrate multidimensional data to derive models for describing biological interactions in plants (Trewavas 2006; Kirschner 2005; Yuan et al. 2008).

More prominent reviews concentrated on one or two aspects of plant systems biology or its

Z. Gao (✉) · X. Ni
College of Horticulture, Nanjing Agricultural University, No 1 Weigang, Nanjing 210095, People's Republic of China
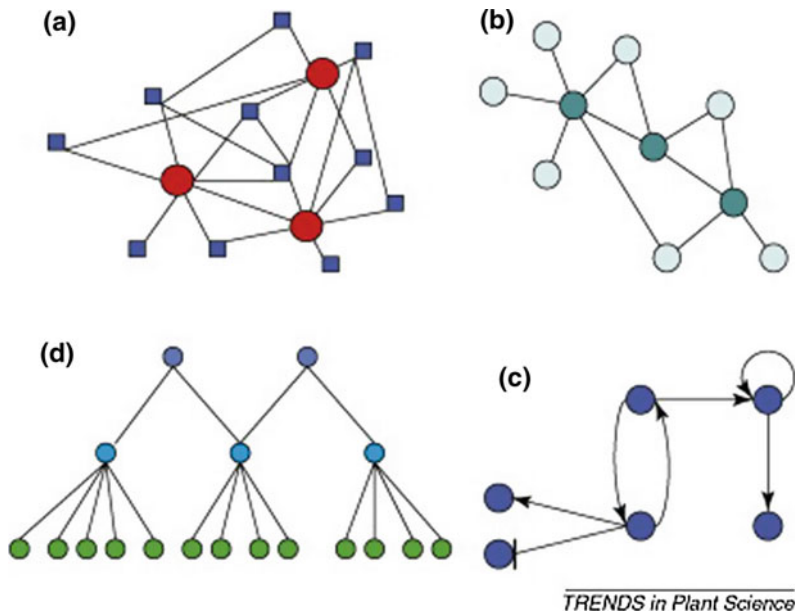e-mail: gaozhihong@njau.edu.cn

Fig. 7.1 Networks in plant systems biology. The four common networks used in plant systems biology studies are shown. **a** The gene-to-metabolite network often derives from correlation analysis of gene and metabolite profiling under multiple conditions. In this network, genes are typically symbolised differently to metabolites. Here, circles and the metabolites represent genes by squares, while lines symbolise the interactions. **b** Two types of interactomes have recently been recognised: genetic and physical interactomes. Interactomes can be derived from genetic or protein-binding assays such as yeast two-hybrid assay and co-immunoprecipitation. These can be visualised with circles representing genes or proteins and lines representing interactions. The genes in centralised hub locations with many interactions among multiple genes are often symbolised with a different colour. **c** The transcriptional regulatory network is highly diverse and can be presented as a hierarchical structure. The elements at the top are expected to be general regulatory genes. **d** The gene regulatory network can be derived from gene expression profiles, mutant analysis and other data. The gene regulatory network is dynamic, and system dynamics need to be visualised in the graph by the different symbols of the lines. The genes often represented by circles and lines frequently represent the interactions between genes. Different symbols at the end of the lines can describe different types of interactions, including gene activation and repression (Yuan et al. 2008)

history, discussing the techniques used in systems biology, such as microarray, next-generation sequencing, synthetic genetic array, pull-down assays, yeast two-hybrid system, systems biology markup language and other important tools (Ahmadian et al. 2006; Galbraith 2006; Kris et al. 2007; Boone et al. 2007). Currently, there is the need for a wide-ranging review covering different types of networks and broad application perspectives.

Constructing networks and analyses have become a common approach to describe biological systems. Networks can be either static or dynamic, and their components can include genes, proteins, cis-elements, metabolites and other molecules. Here, we focus on four network types, including gene-to-metabolite networks, protein–protein interaction networks, transcriptional regulatory networks and gene regulatory networks (Fig. 7.1). The first three types of networks are often static, whereas the gene regulatory network frequently emphasises the dynamic changes of processes.

7.2 Gene-to-Metabolite Networks

Gene-to-metabolite networks define the interactions among genes and metabolites and are typically constructed using multivariance analysis or data mining of gene profiling and metabolite profiling data under different conditions (Fig. 7.1a)

(Yuan et al. 2008). The distance calculated among genes and metabolites could be visualised in statistical analyses according to their profiling patterns. Early research used gene-to-metabolite networks to dissect the dynamic responses during sulphur and nitrogen starvation in *Arabidopsis* (Espinosa-Soto et al. 2004). The work integrated microarray-based gene profiling with liquid chromatography-mass spectrometry (MS) and Fourier-transform ion cyclotron MS-based metabolite profiling, using multivariate analysis methods including self-organising map and principal components analysis to derive gene-to-metabolite associations.

Microarray analysis is employed to provide expression profiles of single genes and new insights to elucidate the biological mechanisms responsible for fruit development. To evaluate the expression of genes mostly engaged in fruit development between *Prunus mume* and *Prunus armeniaca*, Li et al. (2012) identified differentially expressed transcripts along with the entire fruit life cycle by using microarrays spotted with 10,641 ESTs collected from *P. mume* and other *Prunus* EST sequences first. A total of 1418 ESTs were selected through quality control of microarray spots as the main method, including 707 up-regulated and 711 down-regulated genes showing more than two-fold differences in expression level were annotated by GO analysis. These differentially expressed genes were found to be involved in galactose metabolism, starch and sucrose metabolism and in the biosynthesis of other secondary metabolites. This experiment provided detailed information about fruit quality differences during development and ripening of these two species and laid a theoretical for the cloning of genes regulating a series of important rate-limiting enzymes involved in vital metabolic pathways during fruit development.

7.3 Plant Interactome and Protein Interaction Networks

Two types of interactomes have recently been recognised: genetic and physical interactomes (Fig. 7.1b). A genetic interactome is a network of

genes characterised on the basis of genetic interactions, serving to elucidate gene function within physiological processes (Boone et al. 2007).

The phenomenon of pistil abortion widely occurs in Japanese apricot and has seriously affected the yield (Shi et al. 2012b). Shi et al. (2012b) used a combination of two-dimensional gel electrophoresis (2-DE) and matrix-assisted laser MALDI-TOF/TOF approaches to identify the differentially expressed proteomes between perfect and imperfect flower buds in Japanese apricot. In total, 27 candidate protein spots were identified and classified into eight functional classifications and 10 process categories (Fig. 7.2). The majority of the identified proteins were mainly related to stress response and defence mechanisms (8), energy metabolism (4) and biosynthetic processes (4). Here, ACL, SAM, XTH and CCoAOMT could promote the formation of cell walls in perfect flower buds in Japanese apricot, which greatly contributes to pistil development. The SHT may be involved in the O-methylation of spermidine conjugates and could contribute to abnormal floral development. The identification of such differentially expressed proteins provides a new target for future studies to assess their physiological roles and significance in pistil abortion. The regulation and execution of Japanese apricot pistil abortion are a complex network of biochemical and cellular processes, and differentially abundant proteins are involved in multiple metabolic pathways.

Dormancy is of great significance in the growth and development of deciduous fruit trees. Zhuang et al. (2013) used the same methods as Shi et al. (2012a) to identify the differentially expressed proteomes of Japanese apricot flower buds at four critical stages. In their study, 32 identified proteins were classified into six functional categories: stress response and defence (11), energy metabolism (ten), protein metabolism (five), cell structure (three), transcription (one) and unclassified (two). The glyoxalase I homolog could help Japanese apricot survival under various biotic and abiotic stresses, greatly contributing to its dormancy. Enolase, thioredoxin family proteins and triose phosphate
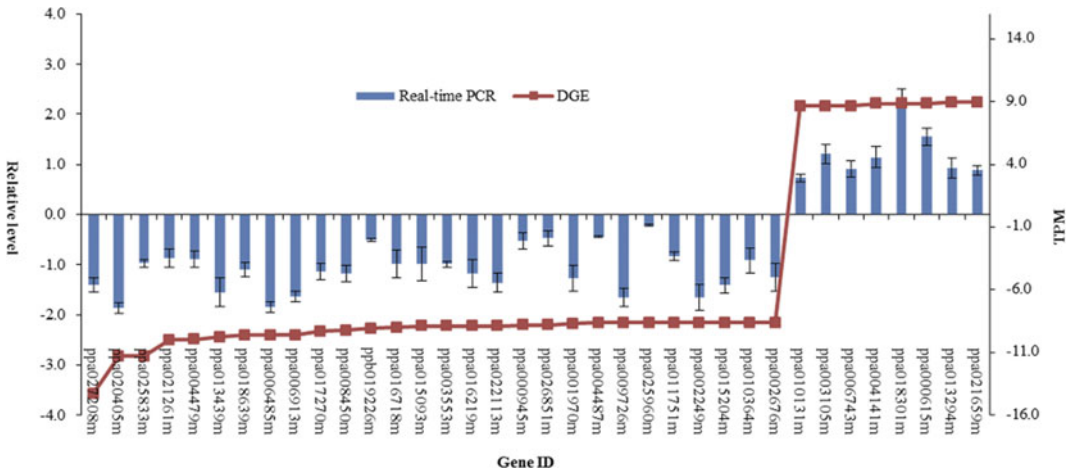
Fig. 7.2 Digital gene expression tag profiling and quantitative real-time PCR analysis of the expression of randomly selected genes. All real-time PCR reactions were repeated three times, and the data are presented as the mean ± SD. The x-axis indicates the different genes. The y-axis shows the expression levels: the left shows the relative expression level by qRT-PCR, and the right shows the tag number per million tags by DGE

isomerase provide adequate energy to complete consecutive dormancy release and bud break in Japanese apricot. Cinnamyl alcohol dehydrogenase 9 and arginase enhance the resilience of plants, enabling them to complete dormancy.

Overall, local-and gene family-based interaction networks have shown much potential for a plant interactome to provide a global view of changes in biological processes, identify the key regulatory proteins and enable in-depth understanding of signal transduction. The potential of interactomes in understanding the systemic regulation of biological processes is unquestionable, but without this work, protein interaction networks cannot be explored further.

7.4 Transcriptional Regulatory Networks

While the interactome describes protein–protein interactions, the transcription regulatory network describes the interaction between transcriptional regulatory genes and downstream genes (Fig. 7.1c) (Belostotsky and Rose 2005; Gilchrist

et al. 2006). The transcriptional regulatory network was one of the first types to be constructed in molecular and cellular biology (Belostotsky and Rose 2005).

The phenomenon of pistil abortion widely occurs in Japanese apricot, and imperfect flowers with pistil abortion seriously decrease the yield (Shi et al. 2012b). Shi et al. (2012b) investigated genes related to the pistil development of Japanese apricot, using high-throughput sequencing technology (Illumina); they first comparatively constructed the digital gene expression profile between the perfect and imperfect flowers in Japanese apricot. There were 689 significant differentially expressed genes between the two libraries; GO annotation revealed that highly ranked genes were those implicated in small molecule metabolism, cellular component organisation or biogenesis at the cellular level and in the fatty acid metabolism. Among them, late embryogenesis abundant protein (LEA), dicer-like 3 (DCL3) xyloglucan endotransglucosylase/hydrolase 2 (XTH2), pectin lyase-like superfamily protein (PPME1), lipid transfer protein 3 (LTP3), fatty acid biosynthesis 1

(FAB1) and fatty acid desaturase 5 (FAD5) might be related to pistil abortion in Japanese apricot by RT-PCR (Fig. 7.2).

Furthermore, proteomic and transcriptomic approaches were used to analyse the mechanisms of dormancy release following GA$_4$ treatment, based on two-dimensional gel electrophoresis (2-DE) and digital gene expression (DGE) profiling, respectively. Genes at the mRNA level associated with energy metabolism and oxidation–reduction played an important role in this

process. Analysis of the functions of identified proteins and genes and their related metabolic pathways would provide a comprehensive proteomic and transcriptomic view about the coordination of dormancy release after GA$_4$ treatment in Japanese apricot flower buds (Fig. 7.3) (Zhuang et al. 2013).

Zhang et al. (2018) analysed the transcriptome profiles at the three endodormancy stages and the natural flush stage, using RNA sequencing combined with phytohormone and sugar content

Fig. 7.3 Molecular model of dormancy release in Japanese apricot treated with GA$_4$. In this model, after GA$_4$ treatment, signal reception, including Ca^{2+} signalling, ROS signalling and hormone signalling modulate the expression of many kinds of genes and proteins, including SD (stress and defence), PM (protein metabolism), OR (oxidation–reduction), EM (energy metabolism) and CS (cell structure)
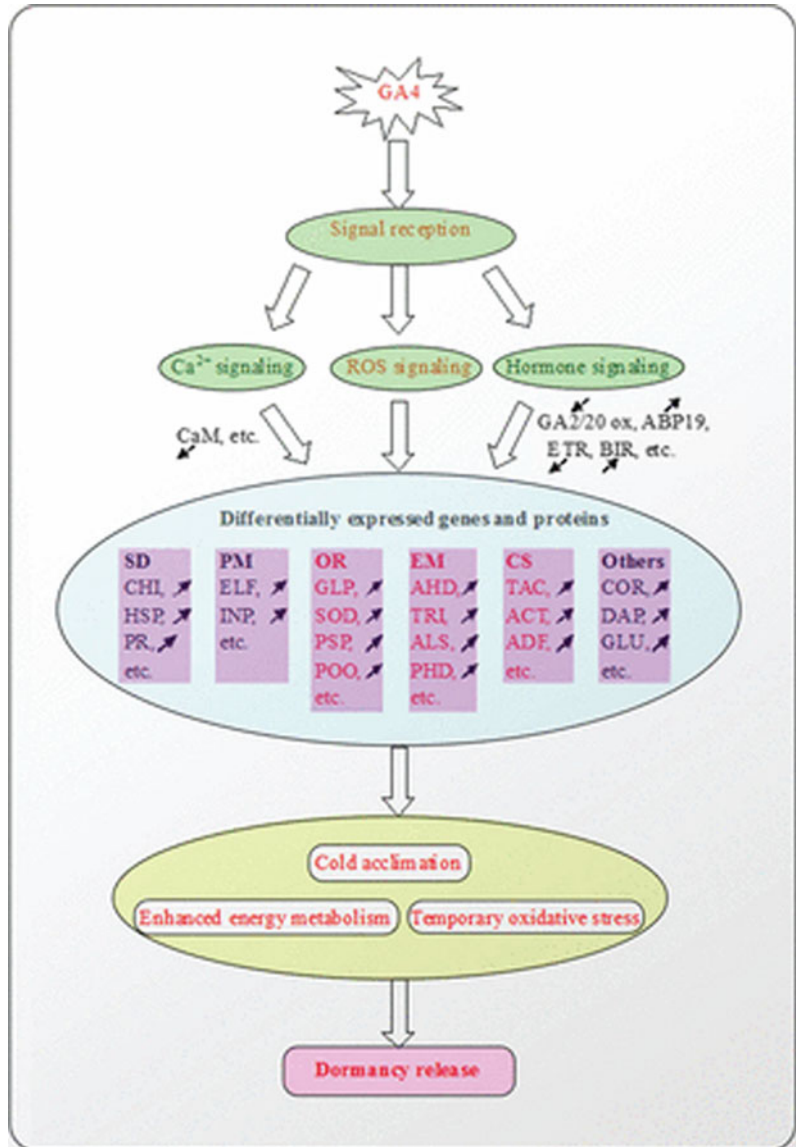
measurements. The authors observed significant alterations in hormone contents and carbohydrate metabolism, and α-amylases, glucan hydrolase family 17 and diphosphate-glycosyltransferase family might play a crucial role during interactions between hormones and sugars. The following hypothetical model for understanding the molecular mechanism of bud dormancy in *P. mume* is proposed: exposure to low temperatures induces the significant up-regulation of eight *C-repeat binding factor* genes, which directly promotes all six *dormancy-associated MADS-box* genes, resulting in dormancy establishment. The prolonged cold and/or subsequently increasing temperatures then decreases the expression levels of these two gene families, which alleviates the inhibition of *FLOWERING LOCUS T* and re-opens the growth-promoting pathway, resulting in dormancy release and initiation of the bud break process.

7.5 Gene Regulatory Networks

In recent years, genome-wide approaches have been applied to obtain a global view of gene regulatory networks underlying flower formation. Furthermore, mathematical models have been developed that can simulate certain aspects of this process and drive further experimentation.

A discrete model of the ABCE for floral organ identity specification was generated based on logical rules derived from existing experimental data (Fig. 7.4) (Espinosa-Soto et al. 2004). The model showed that all possible starting conditions of the network converge to one of five steady states (i.e. IM, sepals, petals, stamens or carpels). The recovery of steady states that correspond to the described expression patterns was also possible in mutant backgrounds, supporting the robustness of the model (Espinosa-Soto et al. 2004). It was noted that the steady states of reproductive organs appeared to be more stable than those of sepals or petals. This observation can be correlated with the fact that the presence and phenotypes of perianth organs are more variable than those of reproductive organs in angiosperms (Theissen and Melzer 2007). The

model also correctly predicted that AG is involved in a positive feedback loop to maintain its own expression (Gomez-Mena et al. 2005; Ó'Maoiléidigh et al. 2013).

Over the past three decades, extensive genetic and molecular analyses have led to the identification of a large number of key floral regulators and to detailed insights into how they control flower morphogenesis. In recent years, genome-wide approaches have been applied to obtain a global view of gene regulatory networks underlying flower formation. In Japanese apricot, a global view of genes (such as MADS-box gene, GRAS gene and *TCP* genes) has been obtained, but the study of gene regulatory networks continued (Xu et al. 2014; Lu et al. 2015; Zhou et al. 2016).

The MADS-box genes encode transcription factors that play crucial roles in plant development, especially in flower and fruit development. Xu et al. (2014) performed a genome-wide identification, characterisation and expression analysis of MADS-box genes in Rosaceae trees. In this study, 80 MADS-box genes were identified in *P. mume* and categorised into MIKC, Mα, Mβ, Mγ and Mδ groups based on gene structures and phylogenetic relationships. Expression analysis suggests that *P. mume* MADS-box genes have diverse functions in *P. mume* development, and the functions of duplicated genes diverged after the duplication events. In addition to its involvement in the development of female gametophytes, type I genes also play roles in male gametophyte development. In conclusion, MADS-box genes play a role in flower and fruit development and lay a foundation for selecting candidate genes for functional studies in *P. mume* and other species.

The GRAS gene family encodes transcriptional regulators that have diverse functions in plant growth and development, such as gibberellin and phytochrome, signal transduction, root radial patterning and axillary meristem formation and gametogenesis in the *P. mume* genome. Lu et al. (2015) investigated the phylogenetic relationships of *P. mume* GRAS genes and their locations and structures in the genome as well as the expression levels of different tissues. Out of 46 identified GRAS genes,
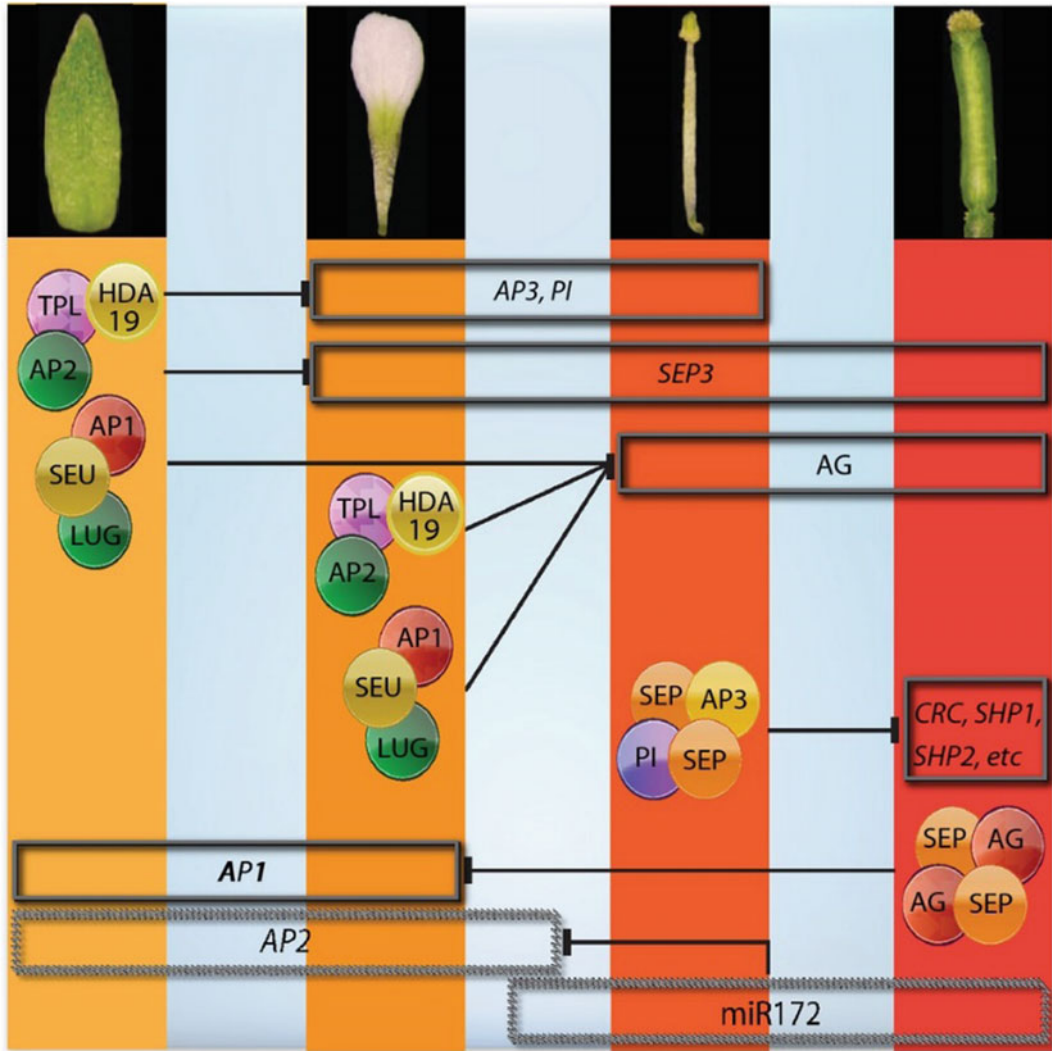
Fig. 7.4 Floral organ boundary maintenance by the ABCE proteins in *Arabidopsis thaliana* flowers

45 were located on the eight *P. mume* chromosomes. Phylogenetic results showed that these genes could be classified into 11 groups. Tissue expression analysis indicated that 13 genes showed high expression levels in roots, stems, leaves, flowers and fruits and were related to plant growth and development. Functional analysis of 24 GRAS genes and an orthologous relationship analysis indicated that many functioned during plant growth, flower and fruit development. In this sense, bioinformatics analysis provides valuable information to improve

the economic, agronomic and ecological benefits of *P. mume* and other Rosaceae fruit trees.

The TCP proteins, belonging to a plant-specific transcription factors family, have significant functions in plant development, especially flower and leaf development. A genome-wide analysis of *TCP* genes has been performed to explore their evolution in *P. mume* (Zhou et al. 2016). Nineteen *PmTCPs* were identified, and three of them contained putative miR319 target sites. Phylogenetic and comprehensive bioinformatics analyses of these genes

revealed that different types of *TCP* genes had undergone different evolutionary processes, and the genes in the same clade had similar chromosomal locations, gene structures and conserved domains. Genes in the *TCP-P* subfamily play major roles in both flower and gametophyte development. The *CIN* genes in double-petal cultivars might play key roles in the formation of the petal, while they are correlated with gametophyte development in the single-petal cultivar. The *CYC/TB1*-type genes have frequently been detected in petal and pistil formation processes. The less complex flower types of *P. mume* might result from the fact that there were only two *CYC-type* genes present in *P. mume*, associated with a lack of *CYC2* genes to control the identity of flower types. These results lay the foundation for further studies on the functions of *TCP* genes during flower development.

7.6 Conclusions

There are three domains that must be addressed to take full advantage of plant systems biology: (i) omics technology development; (ii) data integration into usable formats and (iii) data analysis within the domain of bioinformatics. Among these, bioinformatics probably deserves the greatest attention because it is essential that biological data be normalised, standardised and visualised to build integrated models. Due to the limitation of Japanese apricot systems biology, it is greatly tied to data modelling, in which analysis always involves generalisation, simplification and assumptions. Therefore, networks in systems biology might never completely represent the actual biological system. This interpretation of systems biology data, however, needs to be carefully considered, and systems biology approaches can be complemented by reductionist approaches.

Advances in plant systems biology need to take full advantage of the ever-increasing numbers of omics technologies. Advances are needed for both basic plant biology and the discovery of the key regulatory genes for agricultural purposes. By understanding systems integration, we should be able to accelerate molecular breeding.

References

Ahmadian A, Ehn M, Hober S (2006) Pyrosequencing: history, biochemistry and future. Clin Chim Acta 363 (1–2):83–94

Belostotsky DA, Rose AB (2005) Plant gene expression in the age of systems biology: integrating transcriptional and post-transcriptional events. Trends Plant Sci 10(7):347–353

Boone C, Bussey H, Andrews BJ (2007) Exploring genetic interactions and networks with yeast. Nat Rev Genet 8(6):437–449

Espinosa-Soto C, Padilla-Longoria P, Alvarez-Buylla ER (2004) A gene regulatory network model for cell-fate determination during Arabidopsis thaliana flower development that is robust and recovers experimental gene expression profiles. Plant Cell 16(11):2923–2939

Galbraith DW (2006) DNA microarray analyses in higher plants. OMICS 10(4):455–473

Gilchrist M, Thorsson V, Li B, Rust AG, Korb M, Kennedy K et al (2006) Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4. Nature 441(7090):173–178

Gomez-Mena C, de Folter S, Costa MM, Angenent GC, Sablowski R (2005) Transcriptional program controlled by the floral homeotic gene AGAMOUS during early organogenesis. Development 132(3):429–438

Kirschner MW (2005) The meaning of systems biology. Cell 121(4):503–504

Kris RM, Felder S, Deyholos M, Lambert GM, Hinton J, Botros I et al (2007) High-throughput, high-sensitivity analysis of gene expression in Arabidopsis. Plant Physiol 144(3):1256–1266

Li X, Korir NK, Liu L, Shangguan L, Wang Y, Han J et al (2012) Microarray analysis of differentially expressed genes engaged in fruit development between Prunus mume and Prunus armeniaca. J Plant Physiol 169 (17):1776–1788

Lu J, Wang T, Xu Z, Sun L, Zhang Q (2015) Genome-wide analysis of the GRAS gene family in Prunus mume. Mol Genet Genomics 290(1):303–317

Ó'Maoiléidigh DS, Graciet E, Wellmer F (2013) Gene networks controlling Arabidopsis thaliana flower development. New Phytol 201(1):16–30

Shi T, Gao Z, Wang L, Zhang Z, Zhuang W, Sun H et al (2012a) Identification of differentially-expressed genes associated with pistil abortion in Japanese apricot by genome-wide transcriptional analysis. PLoS ONE 7 (10):e47810

Shi T, Zhuang W, Zhang Z, Sun H, Wang L, Gao Z (2012b) Comparative proteomic analysis of pistil abortion in Japanese apricot (*Prunus mume Sieb*. et Zucc). J Plant Physiol 169(13):1301–1310

Theissen G, Melzer R (2007) Molecular mechanisms underlying origin and diversification of the angiosperm flower. Ann Bot 100(3):603–619

Trewavas A (2006) A brief history of systems biology. "Every object that biology studies is a system of systems." Francois Jacob (1974). Plant Cell 18 (10):2420–2430

Xu Z, Zhang Q, Sun L, Du D, Cheng T, Pan H et al (2014) Genome-wide identification, characterisation and expression analysis of the MADS-box gene family in Prunus mume. Mol Genet Genomics 289(5):903–920

Yuan JS, Galbraith DW, Dai SY, Griffin P, Stewart CN Jr (2008) Plant systems biology comes of age. Trends Plant Sci 13(4):165–171

Zhang Z, Zhuo X, Zhao K, Zheng T, Han Y, Yuan C et al (2018) Transcriptome profiles reveal the crucial roles of hormone and sugar in the bud dormancy of Prunus mume. Sci Rep 8(1):5090

Zhou Y, Xu Z, Zhao K, Yang W, Cheng T, Wang J et al (2016) Genome-Wide identification, characterization and expression analysis of the TCP gene family in Prunus mume. Front Plant Sci 7(120):1301

Zhuang W, Gao Z, Wang L, Zhong W, Ni Z, Zhang Z (2013) Comparative proteomic and transcriptomic approaches to address the active role of GA4 in Japanese apricot flower bud dormancy release. J Exp Bot 64(16):4953–4966

Small RNAs in *Prunus mume*

Zhihong Gao and Ting Shi

Abstract

Japanese apricot (*Prunus mume* Sieb. et Zucc) is an important economic fruit crop in China and Japan, with more than 200 cultivars in China. It is one of the most valuable materials used in the food and winemaking industries and is believed to contain many physio-chemicals beneficial for human health. Micro-RNAs (miRNAs) are a class of endogenous, small, noncoding RNAs that regulate gene expression by mediating gene silencing at transcriptional and post-transcriptional levels in higher plants. High-throughput sequencing is used to identify and quantitatively profile small RNAs in Japanese apricot. Previous studies have shown comparative identification of miRNAs between perfect and imperfect Japanese apricot flowers and some small RNAs related to the flower opening process.

8.1 Introduction

In the past, miRNA identification in flowers using sequencing approaches was commonly applied in studies of *Arabidopsis* (Yang et al. 2011), tomato (Wang et al. 2011b), orchid (An et al. 2011), *Boechera* (Amiteye et al. 2011), maize (Wang et al. 2011a), cotton (Pang et al. 2009), safflower and grape. Arabidopsis miRNAs (Rajagopalan et al. 2006) regulate multiple developmental events.

8.2 High-Throughput Sequencing of Small RNAs from Japanese Apricot Flower Bud Tissue

Taking a broader view of the high-throughput sequencing of small RNAs in Japanese apricot, it was observed that small RNAs of 24 nt dominated the library of unique species (Gao et al. 2012; Wang et al. 2014) (Fig. 8.1), as has been reported for many other plant species, such as *Arabidopsis thaliana* (Fahlgren et al. 2007), *Citrus trifoliata* (Song et al. 2010), *Medicago truncatula*, *Citrus sativus* and *Citrus Sinensis* (Xu et al. 2010). However, the most abundant small RNAs in the imperfect library were 21 nt long. Normally, the length of small RNAs is between 18 and 30 nt. Length distribution analysis is a helpful way to assess the composition of small RNA samples. For example, miRNA is normally 21 nt or 22 nt long, whereas siRNA is 24 nt long (Bartel 2004).

The overall distribution patterns of small RNAs (21-nt sRNAs = 30.33% and 24-nt sRNAs = 35.63%) in Japanese apricot are significantly different from that in *Pinus contorta*, a conifer species in which 21-nt RNAs are more abundant

Z. Gao (✉) · T. Shi
College of Horticulture, Nanjing Agricultural University, No. 1 Weigang, Nanjing 210095, People's Republic of China
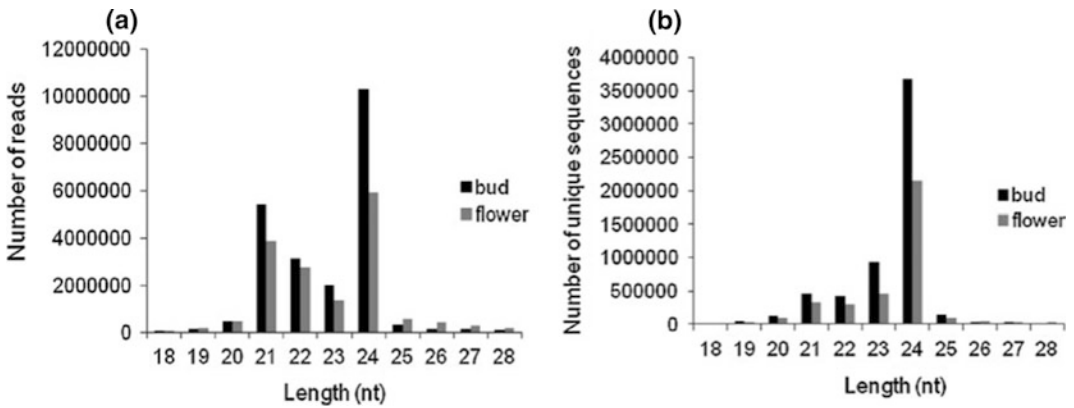e-mail: gaozhihong@njau.edu.cn

Fig. 8.1 Length distribution of sRNAs in bud and flower libraries of *P. mume*. length distribution of sRNA reads **a** and unique sequences, **b** in bud and flower libraries

(>50%) and 24-nt RNAs are less frequent (2.5%). A striking difference also exists when comparing Japanese apricot small RNAs with monocot species of rice. When compared with eudicotyledon species of sweet orange, the difference is not that evident, but still exists, such that 24-nt sRNAs are most frequent (>50%), while 21-nt sRNAs are less common (<20%). These analyses indicate that the small RNA transcriptome is complex across plant species and can be significantly different between phylogenetically distant plant families.

The majority of known miRNA reads are 21–22-nt in length for both libraries (Fig. 8.2a), and the 21-nt sRNAs are the most abundant ones, which is similar to the results reported for Phalaenopsis orchid (An et al. 2011). The majority of known unique miRNA sequences are 21–24 nt in length for both libraries (Fig. 8.2b), and the 24-nt miRNAs are more abundant than the 21-nt miRNAs, which may be due to the important biological functions of 21- and 24-nt miRNAs during flower opening in *P. mume*.
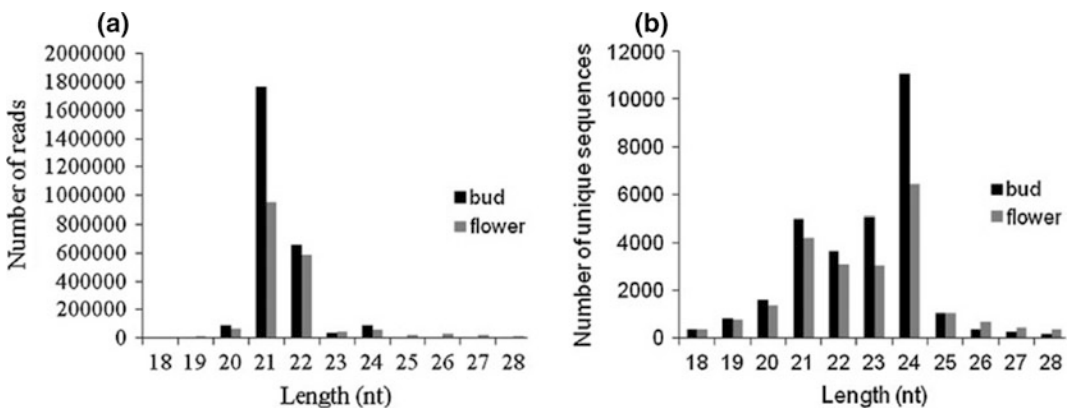

Fig. 8.2 Characterization of known miRNAs in bud and flower libraries of *P. mume*. Length distribution of known miRNA reads **a** and unique sequences, **b** in bud and flower libraries

8.3 miRNA Identification in Plant Flowers

In 2012, researchers identified 61 known miRNAs, belonging to 24 families (Fig. 8.3), and 61 potentially novel miRNAs/new members of known miRNA families (Gao et al. 2012). The analysis revealed that some miRNA families are expressed in the flowers of other plant species, but they are not flower specific. For example, miR156/miR157, miR166 and miR167 were represented most frequently in the libraries in this study. The involvement of three miRNA families (miR172, miR159/miR319 and miR156) in flowering-time regulation has recently been demonstrated in other investigations (Jones-Rhoades et al. 2006). Also, miR164, miR319, miR159 and miR167 specify particular cell types during the later stages of flower development (Nag and Jack 2010). Furthermore, the comparison of these species' miRNAs showed that miR156/miR157 and miR172 might be components of a regulatory pathway mediating the transition between the vegetative and the reproductive phases in plants. In addition, miR172 regulates stem cell fate and defines the inner boundary of APETALA3 and PISTILLATA expression domains in *Arabidopsis*

floral meristems (Zhao et al. 2007). By targeting APETALA2 and type III homeodomain-leucine zipper (HD-Zip) genes, miR166 regulates the temporal program of floral stem cells (Ji et al. 2011). It is believed that miR167, like miR160, targets mRNAs coding for ARF, which are DNA-binding proteins that are thought to control transcription in response to the phytohormone auxin. Transcriptional regulation is important for many of the diverse developmental responses to auxin signals, which include cell elongation, division and differentiation in both roots and shoots (Licausi et al. 2011).

8.4 Identification of Known miRNAs and Evolutionary Conservation

To investigate the evolutionary roles of these known miRNAs, extensive comparisons were performed against known miRNAs in other plant species, including *Physcomitrella, Selaginella, Picea, Pinus, Arabidopsis, Brassica, Ricinus, Glycine, Medicago, Gossypium, Solanum, Aquilegia, Citrus, Populus, Vitis, Oryza, Sorghum and Zea* (Fig. 8.4). Among the miRNA sequences obtained from Japanese apricot, members in five of the families (miR403, miR535, miR827,
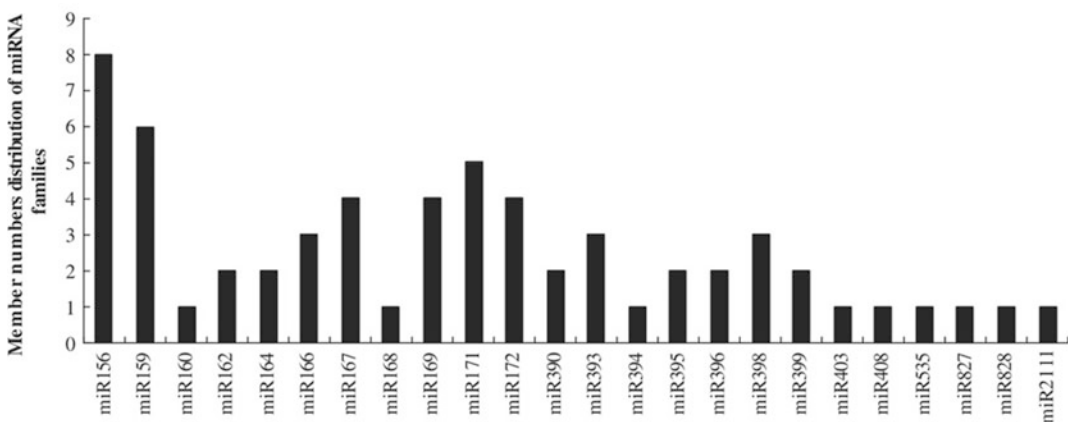

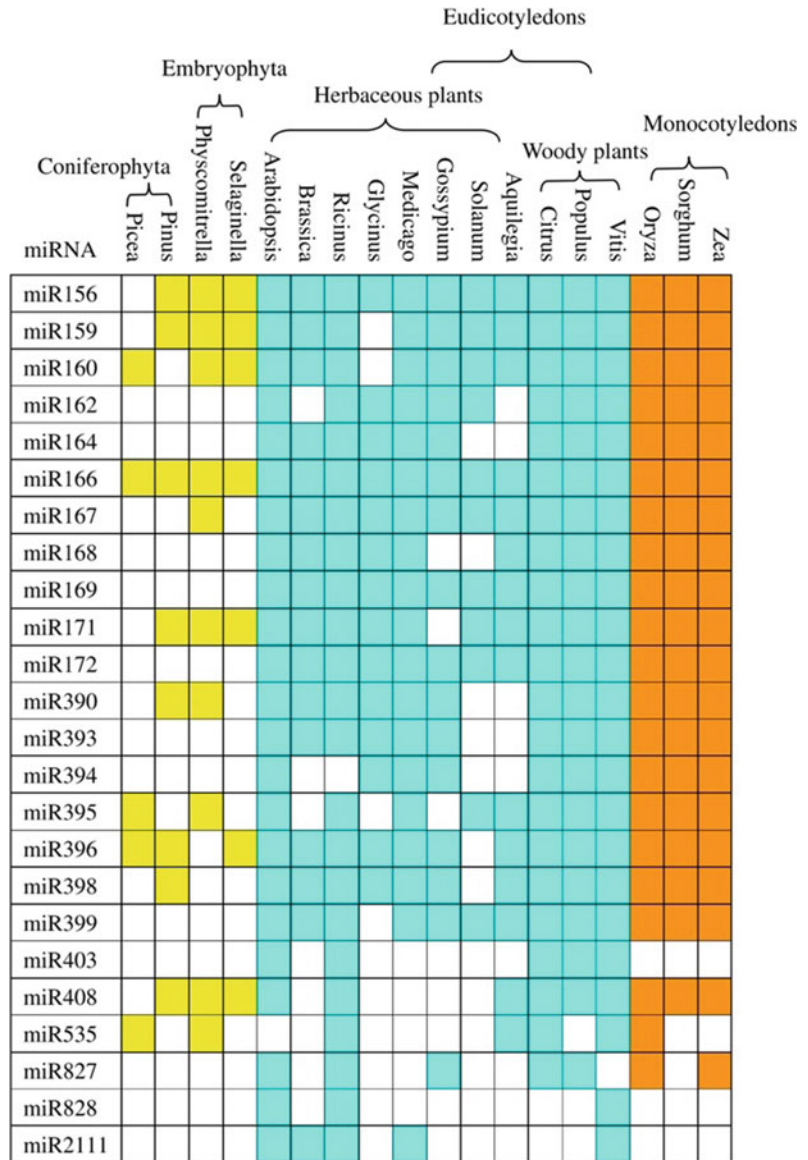**Fig. 8.3** Member numbers of identified miRNAs in each conserved miRNA family in Japanese apricot

Fig. 8.4 Known miRNAs from Japanese apricot, designated as pmu on the first column, and their homologs in other plant species

miR828 and miR2111) showed a lack of conservation of sequence identity compared to orthologues from 18 other plant species. Generally, Japanese apricot miRNAs had corresponding homologs in at least two plant species. Japanese apricot, *Arabidopsis* and *Vitis vinifera* shared 23 conserved miRNA families. Twelve (miR156, miR159, miR160, miR166, miR167, miR171, miR390, miR395, miR396, miR398, miR408 and miR535) out of 24 families had orthologues in Coniferophyta and Embryophyta, indicating that these 12 Japanese apricot miRNA families are ancient. Twelve (miR162, miR164, miR168, miR169, miR172, miR393, miR394, miR399, miR403, miR827, miR828 and miR2111) out of 24 families only had homologs
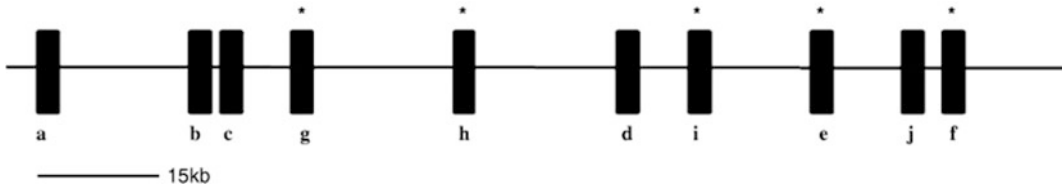
Fig. 8.5 Genomic organization of the miR395 family in *P. mume*. Thin black line represents genomic DNA fragments of scaffold28. Solid vertical bars represent miR395a-j genes. The locations of miR395 genes are roughly in proportion to their real physical locations.

Letters below the vertical bars indicate names of the genes. Asterisks denote miR395 genes whose miRNAs* have been detected in this study. Scale bar represents 15 kb

in angiosperms, indicating that these 12 Japanese apricot miRNA families are recent. In addition, miR403, miR828 and miR2111 only had homologs in eudicotyledons, suggesting that these miRNAs are probably involved in the regulation of specific development.

Most of the conserved miRNAs in *P. mume* were identical to their reference miRNAs sequences, indicating that they are highly conserved among plant species. Sometimes, miRNA family members cluster at a proximal distance on a chromosome in the genome to coordinately express and regulate multiple processes in plants (Li and Mao 2007). Members of the miR395 gene family are clustered with various cluster sizes and intergenic distances in *Arabidopsis* (Maher 2006), *Solanum demissum* (Li and Mao 2007), rice (Guddeti et al. 2005), *M. truncatula* (Guddeti et al. 2005), *Populus trichocarpa* (Lu et al. 2005) and *Brachypodium* (Lu et al. 2005) genomes. Wang et al. (2014) identified ten precursors of the miR395 family that clustered together on scaffold28 in the draft *P. mume* genome (Fig. 8.5). The maximum distance between two of these family members was approximately 15,000 nt, while the shortest distance was 142 nt. The miR395 family in *P. mume* was predicted to target mRNAs for ATP sulfurylase and sulfate transporters, both of which are involved in the sulfate metabolism pathway, which is in agreement with results from *Arabidopsis* (Liang et al. 2010).

8.5   Identification and Expression of miRNA-Guided Cleavage of Target Genes in *P. mume*

Wang et al. (2014) have detected the cleavage sites in seven predicted target genes in *P. mume* (Fig. 8.5). Four SPL genes (Pm030597, Pm002693, Pm002694 and Pm024528), one ARF gene (Pm010467), one SCL gene (Pm017821) and one AP2 gene (Pm018608) were confirmed to be targets of the miR156, miR167, miR171 and miR172 families, respectively. Moreover, the detected cleavage sites in seven predicted target genes were mapped onto the paired miRNA at the tenth or eleventh nucleotide from the 5′ end (Fig. 8.6). Four of these targets, which were predicted to be SPL transcription factors, were regulated by miR156 in *P. mume* (Wang et al. 2014).

Sequences analysis identified 92 conserved miRNAs and 33 novel miRNAs in *P. mume*. Differential expression analysis revealed 21 miRNA sequences during the process of *P. mume* flower opening. The results of qRT-PCR validated the expression of a series of conserved and novel miRNAs revealed by Solexa sequencing. Seven target genes were verified by 5′-RACE to be the targets of miR156, miR167, miR171 and miR172, respectively. These findings contribute to the identification of *P. mume* miRNAs and deepen our understanding of their roles in flower opening (Wang et al. 2014).

Fig. 8.6 Detection of miRNA-mediated mRNA cleavage using modified 5′-RACE and comparative relative expression of targets between two developmental stages by qRT-PCR. The arrows indicate the 5′ cleavage sites that were sequenced from 5′-end cleavage products, which were identified from cloned 5′-RACE products, with the frequency of clones shown
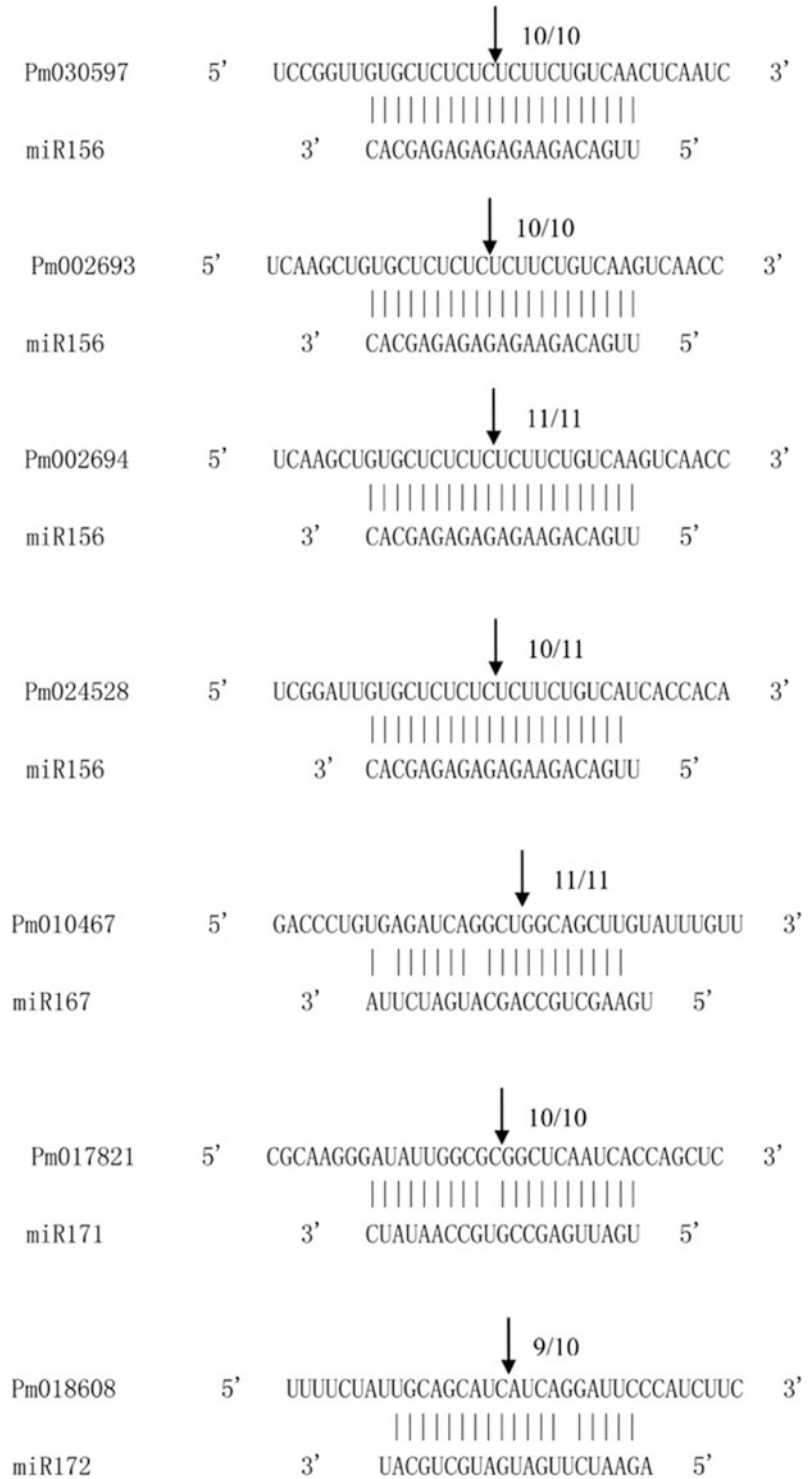
References

Amiteye S, Corral JM, Vogel H, Sharbel TF (2011) Analysis of conserved microRNAs in floral tissues of sexual and apomictic *Boechera* species. BMC Genom 12(1):500

An FM, Hsiao SR, Chan MT (2011). Sequencing-based approaches reveal low ambient temperature-responsive and tissue-specific microRNAs in *Phalaenopsis Orchid* (Research Support, Non-U.S. Govt). PLoS One 6(5):e18937

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116(2):281–297

Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS et al (2007) High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of *MIRNA* genes. PLoS ONE 2(2):e219

Gao Z, Shi T, Luo X, Zhang Z, Zhuang W, Wang L (2012) High-throughput sequencing of small RNAs and analysis of differentially expressed microRNAs associated with pistil development in Japanese apricot. BMC Genom 13(1):371

Guddeti S, Zhang DC, Li AL, Leseberg CH, Kang H, Li XG et al (2005) Molecular evolution of the rice miR395 gene family. Cell Res 15(8):631–638

Ji L, Liu X, Yan J, Wang W, Yumul RE, Kim YJ et al (2011) ARGONAUTE10 and ARGONAUTE1 regulate the termination of floral stem cells through two microRNAs in Arabidopsis. PLoS Genet 7(3):e1001358

Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAs and their regulatory roles in plants. Annu Rev Plant Biol 57:19–53

Li A, Mao L (2007) Evolution of plant microRNA gene families. Cell Res 17(3):212–218

Liang G, Yang F, Yu D (2010) MicroRNA395 mediates regulation of sulfate accumulation and allocation in *Arabidopsis thaliana*. Plant J 62(6):1046–1057

Licausi F, Weits DA, Pant BD, Scheible WR, Geigenberger P, van Dongen JT (2011) Hypoxia responsive gene expression is mediated by various subsets of transcription factors and miRNAs that are determined by the actual oxygen availability. New Phytol 190(2):442–456

Lu S, Sun Y-H, Shi R, Clark C, Li L, Chiang VL (2005) Novel and mechanical stress–responsive microRNAs in *Populus trichocarpa* that are absent from Arabidopsis. Plant Cell 17(8):2186–2203

Maher C (2006) Evolution of Arabidopsis microRNA families through duplication events. Genome Res 16(4):510–519

Nag A, Jack T (2010) Sculpting the flower; the role of microRNAs in flower development. Curr Top Dev Biol 91:349–378

Pang M, Woodward AW, Agarwal V, Guan X, Ha M, Ramachandran V et al (2009) Genome-wide analysis reveals rapid and dynamic changes in miRNA and siRNA sequence and expression during ovule and fiber development in allotetraploid cotton (*Gossypium hirsutum* L.). Genome. Biol 10(11):R122

Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. Genes Dev 20(24):3407–3425

Song C, Wang C, Zhang C, Korir NK, Yu H, Ma Z et al (2010) Deep sequencing discovery of novel and conserved microRNAs in trifoliate orange (*Citrus trifoliata*). BMC Genom 11(1):431

Wang L, Liu H, Li D, Chen H (2011a). Identification and characterization of maize microRNAs involved in the very early stage of seed germination (Research Support, Non-U.S. Gov't). BMC Genomics 12:154

Wang Y, Itaya A, Zhong X, Wu Y, Zhang J, van der Knaap E et al (2011b) Function and evolution of a microRNA that regulates a Ca2±ATPase and triggers the formation of phased small interfering RNAs in tomato reproductive growth. Plant Cell Online 23(9):3185–3203

Wang T, Pan H, Wang J, Yang W, Cheng T, Zhang Q (2014) Identification and profiling of novel and conserved microRNAs during the flower opening process in Prunus mume via deep sequencing. Mol Genet Genomics 289(2):169–183

Xu MJ, Liu Q, Nisbet AJ, Cai XQ, Yan C, Lin RQ et al (2010) Identification and characterization of microRNAs in *Clonorchis sinensis* of human health significance. BMC Genom 11(1):521

Yang XZ, Zhang HY, Li L (2011) Global analysis of gene-level microRNA expression in *Arabidopsis* using deep sequencing data. Genomics 98(1):40–46

Zhao L, Kim Y, Dinh TT, Chen X (2007) miR172 regulates stem cell fate and defines the inner boundary of *APETALA3* and *PISTILLATA* expression domain in *Arabidopsis* floral meristems. Plant J. 51(5):840–849

The Chloroplast Genome of *Prunus mume*

9

Zhihong Gao and Xiaopeng Ni

Abstract

The chloroplast genome of *Prunus mume* exhibits a circular DNA molecule of 157,712 bp, with a typical quadripartite structure. It consists of two inverted repeat regions (IRa and IRb) of 26,394 bp, separated by large (LSC) and small (SSC) single-copy regions of 85,861 and 19,063 bp, respectively. The chloroplast genomes of fruit *P. mume* exhibit a typical quadripartite structure with conserved genome arrangement, structure and divergence. Chloroplast genome length is 157,815 bp, including a pair of inverted repeat regions of 26,393 bp separated by a large single-copy region of 86,113 bp and a small single-copy region of 18,916 bp. In the organisation of the gene aspect, the chloroplast genome of *P. mume* encodes 112 unique genes, 19 of which are duplicated in the IR regions, resulting in 131 genes, 18 of which harbour one or two introns; the GC content is 38.9%. The chloroplast of *P. mume* genome encodes 133 genes (110 unique genes), including 94 protein-coding genes, 31 tRNA and 8 rRNA genes.

Z. Gao (✉) · X. Ni
College of Horticulture, Nanjing Agricultural University, No. 1 Weigang, Nanjing 210095, People's Republic of China
e-mail: gaozhihong@njau.edu.cn

9.1 The Chloroplast Genome

Prunus mume is an important fruit and ornamental crop and has been domesticated in China more than 3000 years ago. There are two varieties, namely fruit *P. mume* and ornamental *P. mume*. Recently, the complete chloroplast genomes of fruit *P. mume* (MH700953) and ornamental *P. mume* (NC_023798.1) have been sequenced. Chloroplasts play a crucial role in sustaining life, and the sequenced chloroplast genomes from a variety of land plants have enhanced our understanding of chloroplast biology, intracellular gene transfer, conservation, diversity and the genetic basis through which chloroplast transgenes can be engineered to enhance plant agronomic traits or to produce high-value agricultural or biomedical products (Daniell et al. 2016b).

The chloroplast development process initiates from multiple endosymbiosis of cyanobacteria and photosynthesis vectors in the plant. Chloroplasts are located in the cytoplasmic matrix and coated with bilayer membranes, showing a flat ellipsoidal or spherical shape. Chloroplasts are active metabolic centres that sustain life on earth by converting solar energy into carbohydrates via photosynthesis and oxygen release. Photosynthesis is often recognised as the key function of plastids, and play a vital role in other aspects of plant physiology and development, including the synthesis of amino acids, nucleotides, fatty acids, phytohormones, vitamins, a plethora of

metabolites and the assimilation of sulphur and nitrogen. Metabolites synthesised in chloroplasts are important for plant interactions with their environment and their defence against invading pathogens. Therefore, chloroplasts serve as metabolic centres in cellular reactions to signals and respond via retrograde signalling. The chloroplast genome encodes numerous key proteins involved in photosynthesis and other metabolic processes (Bobik and Burch-Smith 2015; Daniell et al. 2016a). Chloroplasts are also semi-autonomous genetic organelles within plant cells and contain independent chloroplast DNA (cpDNA). In the molecular machinery of the chloroplast division aspect, within the framework of the Molecular Machinery of Chloroplast Division, a study has found that the two chloroplast Ftsz types in the green lineage (FtsZ2 and FtsZ1) and the red lineage (FtsZA and FtsZB) copolymerise and have a conserved function (Chen et al. 2018). In unicellular algae, the cell cycle restricts the expression of nucleus-encoded chloroplast division proteins and Z-ring assembly to the S-phase, and the disrupted assembly of the chloroplast division machinery arrests cell-cycle progression (Chen et al. 2018).

The chloroplast genome of *P. mume* can be used to understand the structure of Rosaceae genomes and the rapid evolution of the Rosaceae genus. It also helps to measure and characterise the genetic diversity in domesticated and wild populations of fruit trees and to determine how this diversity relates to the tremendous phenotypic diversity in fruit trees. Furthermore, genome-based tools can be developed to improve breeding programs (Zhang et al. 2012).

9.2 Advances in Chloroplast Genome Sequencing Technology

One of the important factors in the rapid advancement of the chloroplast genomics field is the improvement in sequencing technologies. An early plant high-throughput method in chloroplast genome sequencing is the Sanger method,

also called 'first-generation sequencing technology'. However, this method is associated with numerous challenges, including the difficulty of building a quality library, the large number of PCR reactions and the possibility of DNA contamination from other organelles.

Entering the twenty-first century, The Roche's 454 Sequencing Platform, the ABI's Solid Sequencing platform and the Illumina's Solexa sequencing platform represent the second-generation sequencing technologies (NGS). The next-generation sequencing technology provides scientists with faster and cheaper methods for sequencing the chloroplast genome. The chloroplast genomes of *Vigna radiata* (Tangphatsornruang et al. 2010) and Lemnoideae (Wang and Messing 2011) were sequenced via this approach. Illumina is currently the major platform used for chloroplast genomes because it allows the use of rolling-circle amplification products. Investigators can then use bioinformatics platforms to perform de novo assembly without the need for reference genome sequences; from these assemblies, it is possible to identify consensus chloroplast genome sequences. This faster, cheaper and more efficient sequencing technology, heralding a new are of genomics research, promotes the study of chloroplast genomics development. At present, most of the chloroplast genomes included in the NCBI are sequenced through high-throughput sequencing technology. The chloroplast genome of *Mahonia bealei* was sequenced through the Illumina platform (Ma et al. 2013).

The third-generation sequencer, the PACBIO system, using single-molecule real-time (SMRT) sequencing, is currently widely used in chloroplast genome sequencing (38-43). Its advantage is the long read length (Eid et al. 2009), which facilitates de novo genome assembly, particularly in the four chloroplast junctions between the inverted repeat (IR) and single-copy regions (Daniell et al. 2016b). In addition, because of the single-molecule sequencing, the PC process is not required. This avoids systematic errors caused by PCR bias and results in higher fluxes, longer reading lengths and longer durations. The chloroplast genome of the mother plant was

sequenced by the Pacific Biosciences/SMRT technology, and the Single Nucleotide Polymorphisms (SNP) location was detected in 2014. In addition, *Potentilla micrantha* sequencing with the Illumina platform produced seven contigs covering only 90.59% of the chloroplast genome; by contrast, using the PacBio platform with error correction, the entire genome was successfully assembled in a single contig (Li et al. 2014; Ferrarini et al. 2013).

9.3 Characteristics and Structure of the Chloroplast Genome

The chloroplast genome has a highly conserved structure and content organisation; it comprises a double-stranded circular molecule with a typical quadripartite molecular structure. It contains a large single-copy region (LSC), a small single-copy region (SSC) and a pair of reverse complementary repeat regions (IR), where the IR region separates the LSC and SSC regions. The genome length is usually about 120–160 kb, and its differences are mostly due to IR expansion/contraction or loss. For example, the chloroplast genome of some algae plants does not have an IR region. Some leguminous plants lose one of the IR regions, while the chloroplast genome of *Pisum sativum* loses the IR segment and shortens its length. The chloroplast genome generally encodes 110–130 genes that are highly conserved in terms of composition and sequence. Furthermore, the conserved structure of cpDNA and its low nucleotide substitution rate play a vital role in phylogenetic studies. The advent of high-throughput sequencing technologies has facilitated rapid progress in the field of chloroplast genetics and genomics since the first chloroplast genome, from tobacco, was sequenced in 1986 (Shinozaki et al. 1986).

We found differences by comparing two chloroplast genomes with different genomic lengths. The cp genome of ornamental *P. mume* exhibited a circular DNA molecule of 157,712 bp, with a typical quadripartite structure. It consisted of two inverted repeat regions (IRa and IRb) of 26 394 bp, separated by large

(LSC) and small (SSC) single-copy regions of 85,861 and 19,063 bp, respectively. In contrast, the chloroplast genomes of fruit *P. mume* exhibit a typical quadripartite structure with conserved genome arrangement, structure and divergence. Chloroplast genome length is 157,815 bp, including a pair of inverted repeats of 26,393 bp, separated by a large single-copy region of 86,113 bp and a small single-copy region of 18,916 bp. In the organisation of the gene aspect, the chloroplast genome of ornamental *P. mume* encoded 112 unique genes, 19 of which were duplicated in the IR regions, giving 131 genes, 18 of which harboured one or two introns with 38.9% GC content. The chloroplast genome of fruit *P. mume* encoded 133 genes (110 unique genes), including 94 protein-coding genes, 31 tRNA and 8 rRNA genes. There were 18 duplicated genes. Furthermore, 12 intron-containing genes were found; Table 9.1 shows the genome distribution. The concentrations of GC contents of the chloroplast genome in LSC and SSC regions are significantly lower than that in the IR region. The main reason for this phenomenon is that all the eight rRNA genes with high GC content are distributed in the IR region. The IR region of the plant chloroplast genome is highly conserved, but the contraction and expansion of this region is a widespread phenomenon in plant evolution and used as an essential data for comparative analysis of plant chloroplast genomes (Ni et al. 2016). In general, the IR region is the most conserved region of the chloroplast genome. Expansion and contraction regions between IR, LSC and SSC regions are common during evolution and are the primary causes of differences in chloroplast genome length.

9.4 Analysis of the Phylogenetic Relationship According to the Chloroplast Genome Sequences

Phylogenetic relationships between *P. mume* and other Rosaceae trees are problematic because of frequent hybridisation, apomixes, presumed rapid radiation and complex historical diversity.

Table 9.1 List of annotated genes in the *Prunus salicina* Lindi chloroplast genome

| Category	Gene group	Gene name				
Photosynthetic	Subunits of photosystem I	*psaA*(x2)	*psaB*	*psaC*	*psaI*	*psaJ*
	Submits of photosystem II	*psbA*	*psbB*	*psbC*	*psbD*	*psbE*
		psbF	*psbH*	*psbI*	*psbJ*	*psbK*
		psbL	*psbM*	*psbN*	*psbT*	*psbZ*
	Subunits of NADH dehydrogenase	*ndhA*	*ndhB*(x3)	*ndhC*	*ndhD*	*ndhE*
		ndhF(x2)	*ndhG*	*ndhH*	*ndhI*	*ndhJ*
		ndhK(x2)				
	Subunits of cytochrome b/f complex	*petA*	*petB*	*petD*	*petG*	*petL*
		petN				
	Subunits of ATP synthase	*atpA*	*atpB*	*atpE*	*atpF*	*atpH*
		atpI				
	Large subunit of rubisco	*rbcL*				
Self-replication	Proteins of large ribosomal subunit	*rpl2*(x3)	*rpl14*	*rpl16*	*rpl20*	*rpl22*
		rpl23(x2)	*rpl32*	*rpl33*	*rpl36*	
	Proteins of small ribosomal subunit	*rps2*	*rps3*	*rps4*	*rps7*(x2)	*rps8*
		rps11	*rps12*	*rps14*	*rps15*	*rps16*
		rps18	*rps19*(x2)			
	Subunits of RNA polymerase	*rpoA*	*rpoB*	*rpoC1*	*rpoC2*	
	Ribosomal RNAs	*rrn23S*	*rrn16S*	*rrn5S*	*rrn4.5S*	
	Transfer RNAs	*trnR*-TCT	*trnI*-TAT	*trnC*-GCA	*trnT*-GGT	*trnI*-CAT(x2)
		trnS-GGA	*trnF*-GAA	*trnM*-CAT	*trnG*-GCC	*trnR*-ACG(x2)
		trnL-TAG	*trnH*-GTG	*trnY*-GTA	*trnP*-TGG	*trnV*-GAC(x2
		trnS-GCT	*trnS*-AGA	*trnQ*-TTG	*trnD*-GTC	*trnL*-CAA(x2)
		trnW-CCA	*trnT*-TGT	*trnfM*-CAT	*trnS*-TGA	*trnN*-GTT(x2)
		trnE-TTC				
Biosynthesis	Maturase	*matK*				
	Protease	*clpP*				
	Envelope membrane protein	*cemA*				
	Acetyl-CoA carboxylase	*accD*				
	c-type cytochrome synthesis gene	*ccsA(x2)*				
	Translation initiation factor	*infA*				
Unknown function	Conserved hypothetical chloroplast reading frames	*ycf1*(x2)	*ycf2*(x2)	*ycf3*(x2)	*ycf4*(x2)	*ycf15*(x4)

Genome sequences are frequently used to analyse phylogenetic relationships and genetic diversity and in evolution studies. Three independent genetic information genomes are chloroplasts, mitochondria and nuclear genomes. Compared with the nuclear and mitochondrial genome, the chloroplast genome has a small size, single-parental inheritance, a low nucleotide substitution rate, a haploid nature and a highly conserved genomic structure. It is therefore considered as the perfect model for diversity and evolution studies. Plastid phylogenomics offers new and deep insights into the phylogenetic relationships and diversification history of Rosaceae. The development of chloroplast phylogeny and time estimation facilitates future comparative evolutionary studies. In the chloroplast genome sequence, phylogenetic analysis is applied to evaluate the evolutionary relationships between species.

In a previous study, the authors constructed a phylogenetic tree based on the chloroplast genome of *P. mume* compared with nine other complete plastomes of the Rosaceae family, which is highly concordant with the results of other studies as shown in Fig. 9.1 (Wang et al. 2016).

The phylogenetic trees were reconstructed by researchers of the Nanjing Agricultural University

(unpublished data) with 26 published complete chloroplast genomes, including that of *P. mume*. Based on the results, *P. mume*, plum and apricot are closer to each other than to other Rosaceae species (Fig. 9.2). The tree branch length of plum is longer, while *P. mume* and apricot are similar, and the differences among *P. mume*, plum and apricot evolution are higher. *P. mume* is closer to the apricot than to the plum. The result of the phylogenetic analysis is consistent with the traditional classification system, indicating that the classification of Rosaceae is generally reasonable. The results of our phylogenetic analysis partially comply with the traditional classification system of the Chinese flora, e.g. the genera Rosa and Fragaria. With the emergence of more complete chloroplast genome sequences, the chloroplast genome is also expected to help in the identification of the deeper branches of phylogeny. This is consistent with the current classification system and provides a high support for the division of *P. mume* species into the apricot genus. Although there are differences in the phylogenetic tree structure and the molecular phylogeny of the Rosaceae family and its relationship with various genera, these chloroplast genome sequences will provide the genetic information necessary to understand the evolution of the plastid genome.

Fig. 9.1 Maximum likelihood (ML) phylogram of *Prunus mume* and nine Rosaceae species using whole chloroplast genome sequences. Numbers above each node indicate the ML bootstrap support values
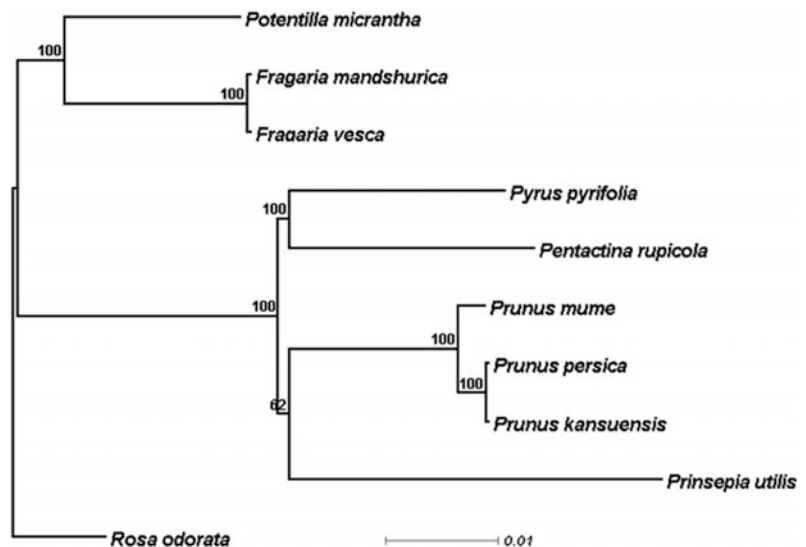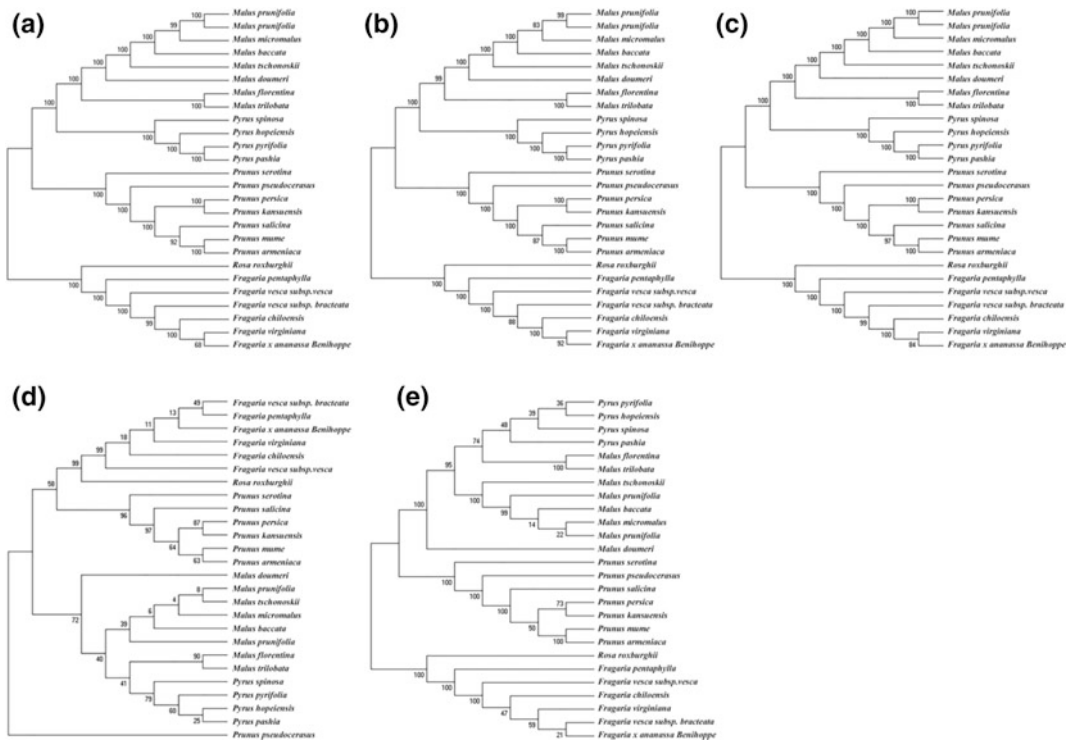
Fig. 9.2 Phylogenetic trees of the Rosaceae species chloroplast genome by MP. **a** Phylogenetic tree constructed using the complete chloroplast genome. **b** Phylogenetic tree constructed using the coding region. **c** Phylogenetic tree constructed by LSC. **d** Phylogenetic tree constructed using the intron. **e** IR constructs phylogenetic tree

References

Bobik K, Burch-Smith TM (2015) Chloroplast signaling within, between and beyond cells. Front Plant Sci 6:781

Chen C, MacCready JS, Ducat DC, Osteryoung KW (2018) The molecular machinery of chloroplast division. Plant Physiol 176(1):138–151

Daniell H, Chan HT, Pasoreck EK (2016a) Vaccination via chloroplast genetics: affordable protein drugs for the prevention and treatment of inherited or infectious human diseases. Annu Rev Genet 50:595–618

Daniell H, Lin CS, Yu M, Chang WJ (2016b) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol 17(1):134

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G et al (2009) Real-time DNA sequencing from single polymerase molecules. Science 323(5910):133–138

Ferrarini M, Moretto M, Ward JA, Surbanovski N, Stevanovic V, Giongo L et al (2013) An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome. BMC Genom 14 (1):670

Li Q, Li Y, Song J, Xu H, Xu J, Zhu Y et al (2014) High-accuracy de novo assembly and SNP detection of chloroplast genomes using a SMRT circular consensus sequencing strategy. New Phytol 204 (4):1041–1049

Ma J, Yang B, Zhu W, Sun L, Tian J, Wang X (2013) The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. Gene 528(2):120–131

Ni L, Zhao Z, Dorje G, Ma M (2016) The complete chloroplast genome of Ye-Xing-Ba (*Scrophularia dentata*; Scrophulariaceae), an Alpine Tibetan herb. PLoS One 11(7):e0158488

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T et al (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5(9):2043–2049

Tangphatsornruang S, Sangsrakru D, Chanprasert J, Uthaipaisanwong P, Yoocha T, Jomchai N et al (2010) The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput

pyrosequencing: structural organization and phylogenetic relationships. DNA Res 17(1):11–22

Wang W, Messing J (2011) High-throughput sequencing of three *Lemnoideae* (duckweeds) chloroplast genomes from total DNA. PLoS One 6(9):e24670

Wang S, Gao CW, Gao LZ (2016) Plastid genome sequence of an ornamental and editable fruit tree of Rosaceae, *Prunus mume*. Mitochondrial DNA A DNA Mapp Seq Anal 27(6):4407–4408

Zhang Q, Chen W, Sun L, Zhao F, Huang B, Yang W et al (2012) The genome of *Prunus mume*. Nat Commun 3:1318

Prunus mume Transcriptomics

10

Zhihong Gao and Ting Shi

Abstract

Large-scale transcriptome analysis using next-generation sequencing (NGS) technologies has proven extremely useful to obtain candidates for such unidentified genes. The NGS technologies are capable of generating high-throughput reads at a relatively low cost and have been used in various types of research, including genome sequencing, marker discovery and, in particular, transcriptome analysis. Genome-wide expression analyses are essential tools for elucidating molecular functions. The number of reads generated by the Roche/454 GS FLX is considerably lower than that generated by short-read platforms such as the Illumina Genome Analyzer. Transcriptome analysis using NGS technologies can capture nearly all the expressed sequences, including rare transcripts in a particular tissue at a specific developmental stage, due to the great depth of sequencing. Therefore, this approach is highly useful, especially for transcriptome analysis in non-model organisms such as *Prunus mume*.

Z. Gao (✉) · T. Shi
College of Horticulture, Nanjing Agricultural University, No. 1 Weigang, Nanjing 210095, People's Republic of China
e-mail: gaozhihong@njau.edu.cn

10.1 Pyrosequencing and Assembly

Habu et al. (2012) were the first to perform 454-pyrosequencing to examine gene expression patterns of the dormant leaf and flower buds at three different stages of dormancy. Out of 485,376 generated reads, they obtained 28,382 contigs and 85,247 singletons, of which 47,401 (41.7%) were annotated by BLAST searches against the non-redundant NCBI protein/nucleotide database (Fig. 10.1). Functional classification by the gene ontology (GO) term indicated that the genes obtained in this study have a relatively wide range of biological processes (Fig. 10.2). The authors searched for up-regulated genes in endodormant leaf and flower buds and found that 74 and 82 genes, respectively, were up-regulated at the endodormant stage as compared with the paradormant stage. From the EST obtained data, they constructed a 'Japanese apricot dormant bud EST database' (JADB) for future studies on dormancy in *Prunus* (Fig. 10.3).

This is the first report of a large-scale transcriptome analysis via 454-pyrosequencing in *P. mume*. This large-scale transcriptome analysis of dormant tissues revealed several candidate genes related to the control of endodormancy, but their functions have not yet been elucidated. However, the unique sequences obtained in this study include numerous novel sequences, which can be used to construct microarrays and as reference sequences for short-read
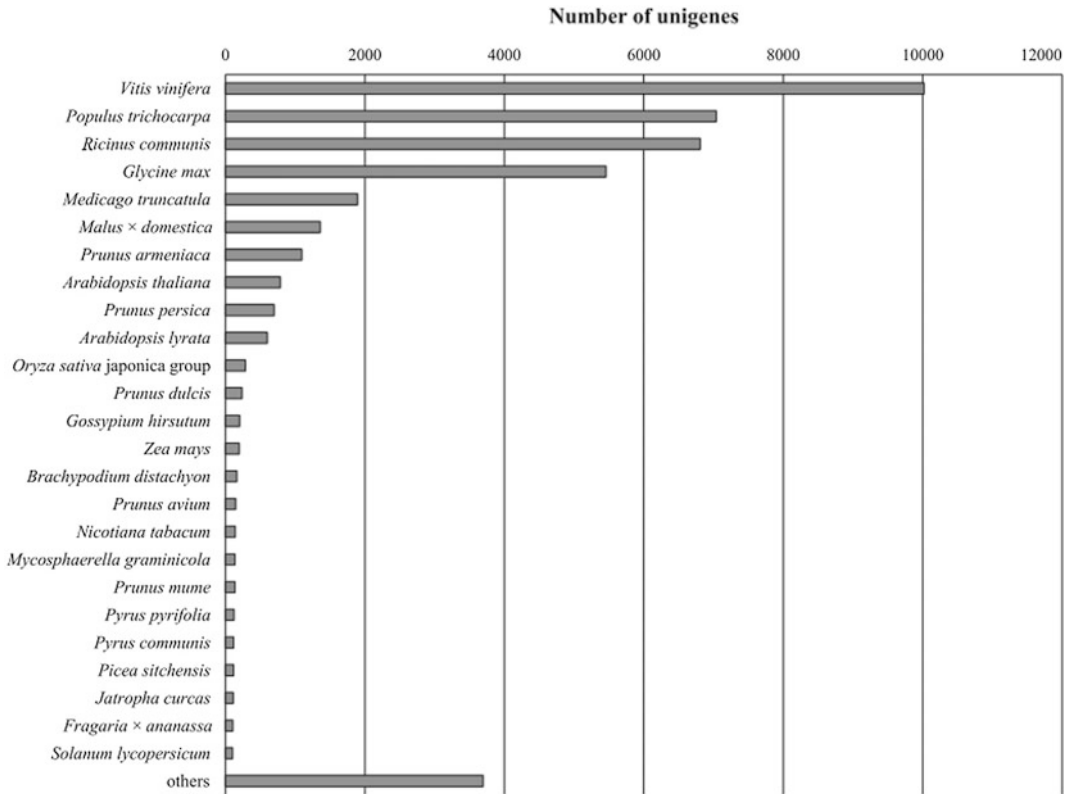
Fig. 10.1 Species distributions of top BLAST hits for each annotated unique sequence against the NCBI non-redundant databases. Only 2530 unique sequences showed closest hits to genes of *Prunus* species (*P. armeniaca*: 1095, *P. persica*: 698, *P. dulcis*: 238, *P. avium*: 146, *P. mume*: 136, and other *Prunus* species: 217), while most unique sequences showed closest hits to sequences from other species such as *Vitis vinifera* (10,019) and *Populus trichocarpa* (7039) (Habu et al. 2012)

sequencing; such an approach will likely help to elucidate the molecular mechanisms controlling endodormancy.

10.2 Comparing 454-Pyrosequencing with Illumina Genome Analyzer IIx (GAIIx) Single-End Sequencing

Habu et al. (2012) conducted a large-scale transcriptome analysis of unpollinated, self-pollinated and cross-pollinated pistils of *Prunus mume* to capture all of the molecular events induced by the GSI reaction. They obtained 40,061 unigenes from 77,521,310 reads from pollinated and unpollinated pistils to pollen grains. Among these unigenes, 29,985 and 27,898 unigene sequences showed at least one hit against the NCBI nr and TAIR10 protein databases, respectively, in BLASTX searches using an *E*-value cut-off of 1e-6. All obtained sequences from 454-pyrosequencing and GAIIx reads are shown in Table 10.1. The number of reads generated by the Roche/454 GS FLX is significantly lower than that generated by short-read platforms such as the Illumina Genome Analyzer (Habu and Tao 2014).
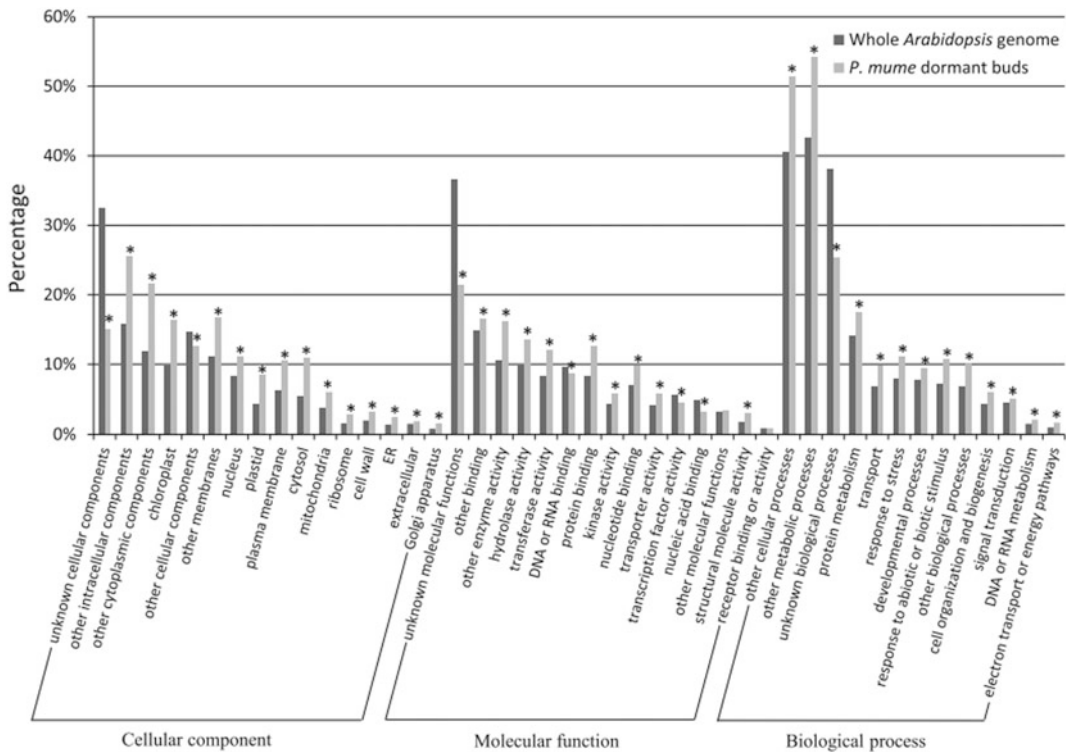
Fig. 10.2 Gene ontology (GO) assignments of unigenes. Proportion of the annotated unigenes and the *Arabidopsis* whole genes that matched various gene ontology categories. *Indicates a significant difference between the proportions of assigned GO terms in the annotated unigenes and the whole *Arabidopsis* genes (Fisher's exact test, $P < 0.001$) (Habu et al. 2012)

10.3 HiSeq 2000 Platform with Paired-End Sequencing

Jo et al. (2015) conducted a de novo transcriptome assembly for two selected *P. mume* cultivars, referred to as Takada and Wallyoung, by paired-end RNA sequencing. The authors obtained 9.14 and 9.48 GB sequence data from Takada and Wallyoung (NCBI accession numbers: SRX1187101 and SRX1187169), respectively. De novo transcriptome assembly identified 130,989 and 116,941 transcripts for Takada and Wallyoung, respectively (Table 10.2). In addition, Shi et al. (2012) used the digital gene expression tag profiling (DGE) method to perform a deep transcriptome analysis of *P. mume*. Details are shown in Chap. 14.

10.4 Transcriptome Sequencing (RNA-Seq) Analysis of the AP2/ERF Gene Family in *Prunus mume*

Du et al (2013) studied the expression of PmAP2/ERF genes in five sections of *P. mume*, young leaves, stems, young roots, flowers and immature fruits, by RNA-Seq. The expression patterns of 116 PmAP2/ERF genes are shown in Fig. 10.4 (ERF family) and Fig. 10.5 (AP2 and RAV families, Soloist). No expression was detected for five genes: PmAP2-14, PmAP2-15, PmERF2-4, PmERF4-2 and PmRAV-3. Seventy genes were detected in all five organs: bud, fruit, leaf, root and stem. However, only six genes (PmERF9-6, PmERF8-6, PmERF7-2,
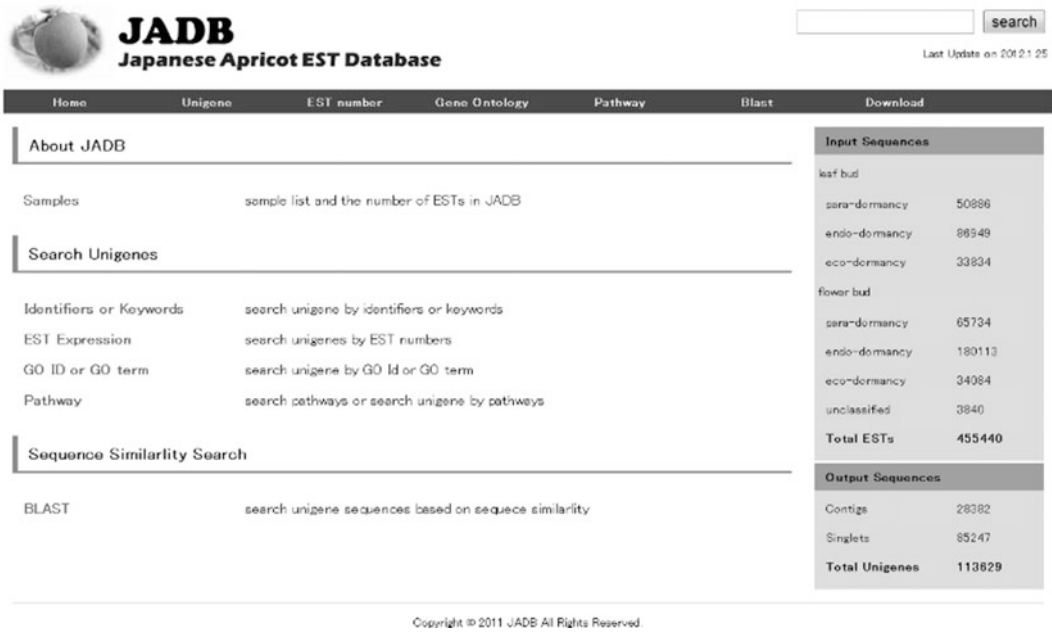
Fig. 10.3 Screenshot of the JADB web interface. Users can search unigenes in JADB using unigene identifiers, keywords, GO terms, pathway names and expression patterns (Habu et al. 2012)

Table 10.1 Samples and number of reads obtained from NGS platforms

Read set	Cultivar	Platform	No. of raw reads	No. of processed reads
CP_454	'Nanko'	GS FLX Titanium	231,107	212,240
SP_454	'Nanko'	GS FLX Titanium	280,823	252,608
UP_454	'Nanko'	GS FLX Titanium	205,649	186,404
NP_454	'Nanko'	GS FLX Titanium	157,615	137,856
NGP_454	'Nanko'	GS FLX Titanium	141,884	121,256
KP_454	'Kairyo-Uchida-Ume'	GS FLX Titanium	133,537	114,256
KGP_454	'Kairyo-Uchida-Ume'	GS FLX Titanium	163,964	140,487
CP_GAIIx	'Nanko'	GAIIx	30,987,870	27,302,728
SP_GAIIx	'Nanko'	GAIIx	27,505,143	23,990,492
UP_GAIIx	'Nanko'	GAIIx	27,634,681	25,062,983

Table 10.2 Summary of de novo assembled two *Prunus mume* transcriptomes (Jo et al. 2015)

Index	Takada	Wallyoung
Total trinity transcripts	130,989	116,941
Total trinity components	64,777	62,142
Percent GC	42.32	42.32
Contig N50	1864	2044
Median contig length	963	984
Average contig	1229.60	1320.65
Total assembled bases	161,063,714	154,438,141
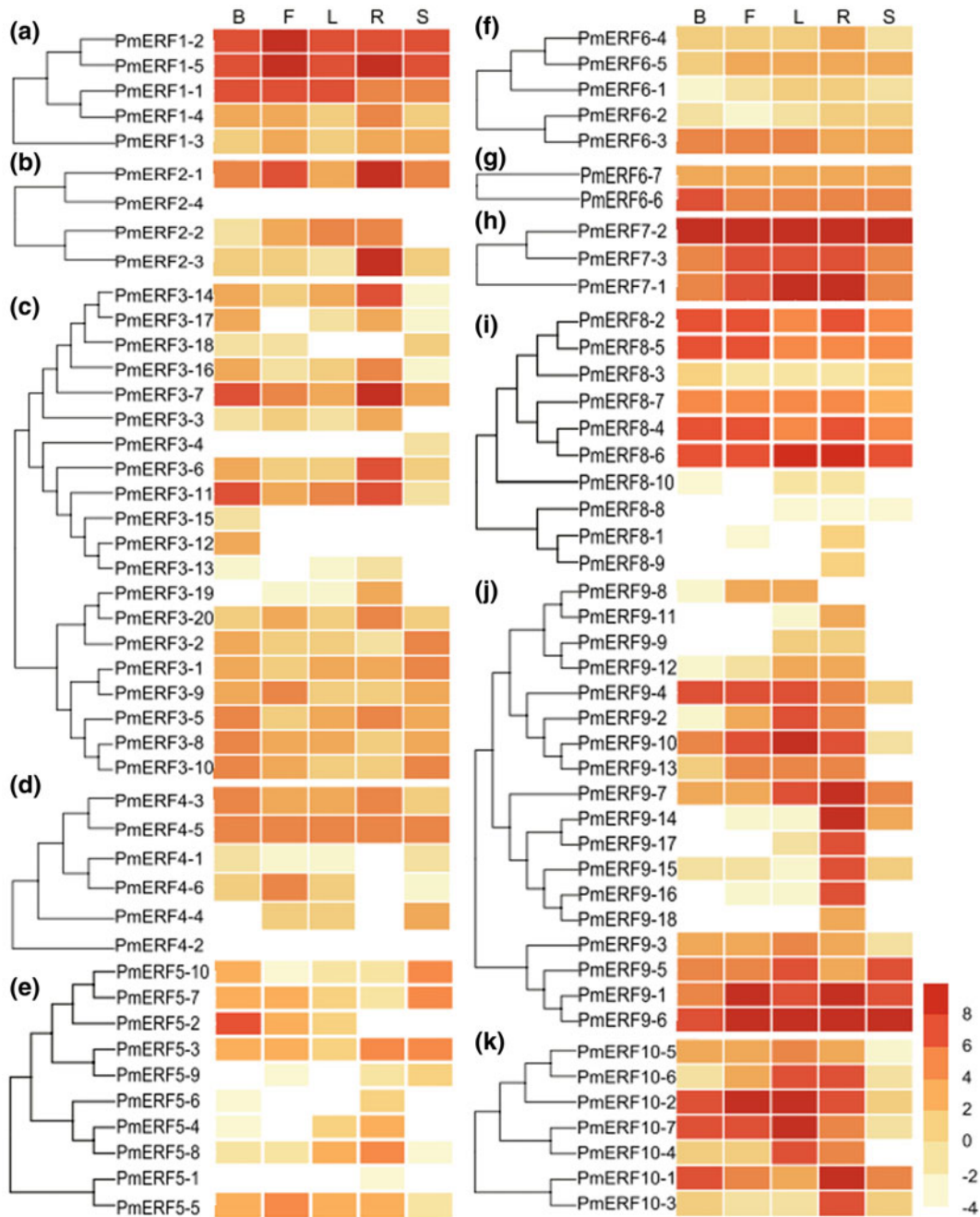
Fig. 10.4 Heat map showing expression levels of Chinese plum ERF family genes in five organs. The expression of all ERF genes identified in this study was measured by transcriptome sequencing (RNA-Seq) analysis using five Chinese plum organs: buds (B), fruits (F), leaves (L), roots (R) and stems (S). The ERF genes are divided into eleven groups: I (**a**), II (**b**), III (**c**), IV (**d**), V (**e**), VI (**f**), VI-L (**g**), VII (**h**), VII (**i**), IX (**j**) and X (**k**). The colour scale represents reads per kilo base per million normalised log$_2$-transformed counts, where light green indicates low levels and red indicates high levels (Du et al. 2013)

Fig. 10.5 Heat map showing expression levels of Chinese plum AP2, RAV and soloist genes in five organs. The expression of AP2 (**a**), RAV (**b**) and soloist (**c**) genes was measured by transcriptome sequencing (RNA-Seq) analysis using five Chinese plum organs: buds (B), fruits (F), leaves (L), roots (R) and stems (S). The colour scale represents reads per kilo base per million normalised $\log_2$-transformed counts, where light green indicates low levels and red indicates high levels (Du et al. 2013)
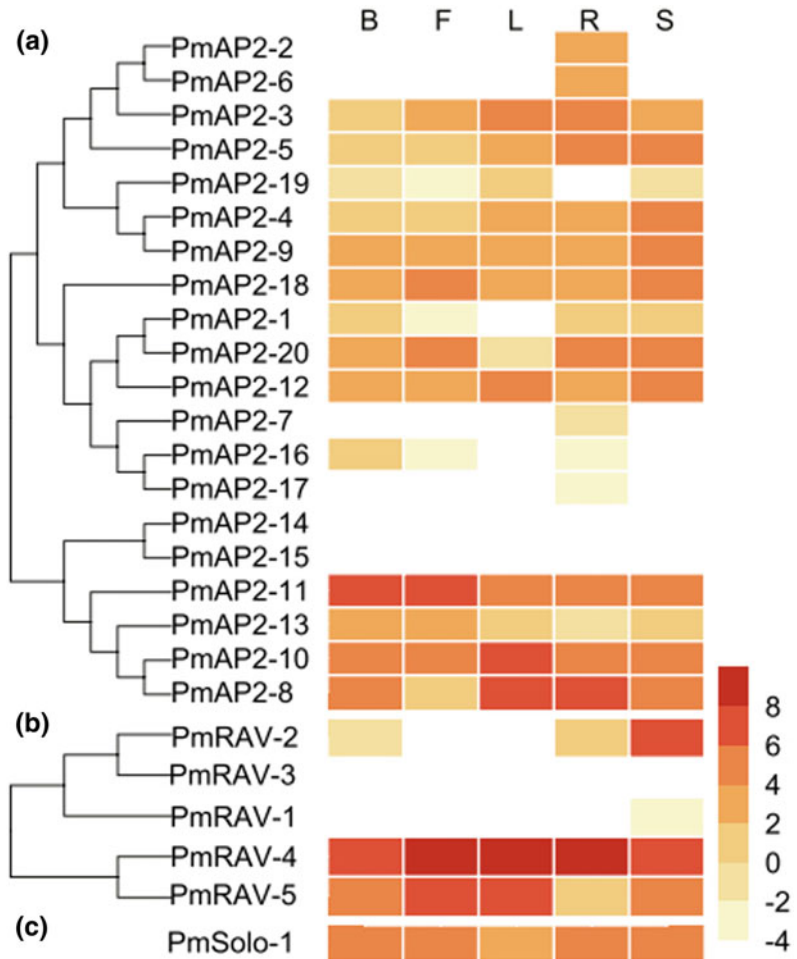
PmERF1-2, PmERF1-5 and PmRAV-4) showed high expression in all five organs. Most PmAP2/ERF genes showed some degree of tissue specificity, with 48 genes being the most abundant in roots, 23 genes in leaves, 18 genes in stems, 11 genes in buds and 11 genes in fruits (Du et al. 2013).

10.5 Conclusions

Large-scale transcriptome analysis using next-generation sequencing (NGS) technologies has proven extremely useful to obtain candidates for such unidentified genes (Zhang et al. 2012). The NGS technologies are capable of generating high-throughput reads at a relatively low cost and have been used in various types of research, including genome sequencing (Zhang et al. 2018a), marker discovery and, in particular, transcriptome analysis (Matsumura et al. 2010; Wang et al. 2010). Transcriptome analysis using NGS technologies can capture nearly all of the expressed sequences (Xu et al. 2014, 2015; Zhang et al. 2017, 2018b; Du et al. 2013), including rare transcripts in a particular tissue at a specific developmental stage, due to the great depth of sequencing (Kofler et al. 2011; Azim et al. 2014). Therefore, this approach is extremely useful, especially for transcriptome analysis in non-model organisms such as Japanese apricot.

References

Azim MK, Khan IA, Zhang Y (2014) Characterization of mango (*Mangifera indica* L.) transcriptome and chloroplast genome. Plant Mol Biol 85(1–2):193–208

Du DL, Hao RJ, Cheng TR, Pan HT, Yang WR, Wang J et al (2013) Genome-wide analysis of the *AP2/ERF* gene family in *Prunus mume*. Plant Mol Biol Rep 31 (3):741–750

Habu T, Tao R (2014) Transcriptome analysis of self- and cross-pollinated pistils of Japanese apricot (*Prunus mume* Sieb. et Zucc.). J Jpn Soc Hortic Sci 83(2):95–107

Habu T, Yamane H, Igarashi K, Hamada K, Yano K, Tao R (2012) 454-pyrosequencing of the transcriptome in leaf and flower buds of Japanese apricot (*Prunus mume* Sieb. et Zucc.) at different dormant stages. J Jpn Soc Hortic Sci 81(3):239–250

Jo Y, Lian S, Cho JK, Choi H, Chu H, Cho WK (2015) De novo transcriptome assembly of two different *Prunus mume* cultivars. Genomics Data 6:273–274

Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A et al (2011) PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS One 6(1):e15925

Matsumura H, Yoshida K, Luo SJ, Kimura E, Fujibe T, Albertyn Z et al (2010) High-throughput super SAGE for digital gene expression analysis of multiple samples using next generation sequencing. PLoS One 5(8):e12010

Shi T, Gao Z, Wang L, Zhang Z, Zhuang W, Sun H, Zhong W, Schönbach C (2012) Identification of differentially-expressed genes associated with pistil abortion in Japanese apricot by genome-wide transcriptional analysis. PLoS ONE 7(10):e47810

Wang XW, Luan JB, Li JM, Bao YY, Zhang CX, Liu SS (2010) De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. BMC Genomics 11:400

Xu Z, Zhang Q, Sun L, Du D, Cheng T, Pan H et al (2014) Genome-wide identification, characterisation and expression analysis of the MADS-box gene family in *Prunus mume*. Mol Genet Genomics 289(5):903–920. Available at: http://link.springer.com/10.1007/s00438-014-0863-z

Xu Z, Sun L, Zhou Y, Yang W, Cheng T, Wang J et al (2015) Identification and expression analysis of the SQUAMOSA promoter-binding protein (SBP)-box gene family in *Prunus mume*. Mol Genet Genomics 290(5):1701–1715

Zhang Q, Chen W, Sun L, Zhao F, Huang B, Yang W et al (2012) The genome of *Prunus mume*. Nat Commun 3:1318

Zhang J, Zhao K, Hou D, Cai J, Zhang Q, Cheng T et al (2017) Genome-wide discovery of DNA polymorphisms in mei (*Prunus mume* Sieb. et Zucc.), an ornamental woody plant, with contrasting tree architecture and their functional relevance for weeping trait. Plant Mol Biol Rep 35(1):37–46. Available at: http://link.springer.com/10.1007/s11105-016-1000-4

Zhang Q, Zhang H, Sun L, Fan G, Ye M, Jiang L et al (2018a) The genetic architecture of floral traits in the woody plant *Prunus mume*. Nat Commun 9(1):1702

Zhang Z, Zhuo X, Zhao K, Zheng T, Han Y, Yuan C, et al (2018b) Transcriptome profiles reveal the crucial roles of hormone and sugar in the bud dormancy of *Prunus mume*. Sci Rep 8(1)

Functional Genes in Bud Dormancy and Impacts on Plant Breeding

11

Yuto Kitamura, Wenxing Chen, Hisayo Yamane and Ryutaro Tao

Abstract

Bud dormancy is an important developmental stage that affects bud break and blooming in fruit trees. Because prolonged chilling exposure is necessary for dormancy release, climatic changes during dormancy season may affect the progress of bud dormancy and bud break. Although understanding the mechanisms of dormancy and genotype-dependent chilling requirement is important for stable fruit production in the future, the internal control mechanisms are largely unclear. For a better understanding of bud dormancy, numerous genetic and molecular analyses have been performed. For a molecular genomic approach, transcriptome analysis identified the MADS-box transcription factor gene called *PmDAM6* as a potential master regulator for dormancy release. Overexpression of *PmDAM6* in poplar (*Populus* spp.) could inhibit shoot growth and promote dormancy induction. The elucidation of the molecular network involved in PmDAMs is ongoing by taking advantage of *Prunus* *mume* whole-genome sequences. For the genetic approach, quantitative trait locus (QTL) analysis, using two F_1 segregating populations derived from the cross between high-chill and low-chill lines of *P. mume*, was conducted to identify loci affecting bud dormancy. By using the genotyping-by-sequencing technique, significant QTLs controlling chilling requirement and bud break date of leaf buds were identified at the terminal region of linkage group 4 (LG4) and co-localised with each other, suggesting that this locus controls dormancy release and bud break of leaf buds in *P. mume*. The use of genetic and molecular analyses to facilitate the breeding of *P. mume* cultivars that can adapt to various climatic conditions is considered an approach to mitigate global climate change.

11.1 Introduction

Prunus mume Sieb. et Zucc. is one of the most popular fruit tree species in East Asian countries, such as China, Taiwan, South Korea and Japan. Because the *P. mume* fruit is too sour to be eaten fresh, it is consumed as pickles, syrups or processed liqueurs. Salty pickles (Umeboshi in Japanese) are Japan's traditional preserved food products, and the consumption of processed liqueurs prepared from *P. mume* fruit (Umeshu in Japanese) has recently been increasing because they are rich in functional components such as

Y. Kitamura
Japanese Apricot Laboratory, Wakayama Fruit Experimental Station, Wakayama, Japan

W. Chen · H. Yamane · R. Tao (✉)
Graduate School of Agriculture, Kyoto University, Kyoto, Japan
e-mail: rtao@kais.kyoto-u.ac.jp

H. Yamane
e-mail: hyamane@kais.kyoto-u.ac.jp

polyphenols and organic acids, with considerable health effects. Moreover, the flowers of *P. mume* are used for ornamental purposes.

The whole *P. mume* genome has been sequenced, and the approximately 280-Mbp genome was assembled, containing 31,390 protein-coding genes (Zhang et al. 2012). Since this achievement, numerous studies about the genome-wide identification and characterisation of genetic factors controlling diverse traits in *P. mume* have been published, which include the genes associated with fundamental plant development and stress tolerance. The genes/gene family include *APETALA 2/ethylene-responsive element binding factor* (*AP2/ERF*), *late embryogenesis abundant* (*LEA*), *GRAS* gene family (Du et al. 2013a, b; Lu et al. 2015) and *MADS-box* genes, which are related to flowering (Xu et al. 2014), weeping trait, an ornamentally important characteristic (Zhang et al. 2015), and bud dormancy (Xu et al. 2014; Kitamura et al. 2018).

Recently, the impacts of global warming on the phenological development of several fruit tree species have been reported, and future shifts to suitable production areas are being predicted. In fact, global climate change also affects winter chilling accumulation and causes insufficient dormancy release to bloom in perennial crops (Atkinson et al. 2013; Sugiura et al. 2007, 2012). In Japan, *P. mume* trees usually bloom from early February until mid-March, when temperatures are typically relatively low. The trees bear fruits in June, after which they stop growing until September, which is when they enter an overwintering dormant stage (Fig. 11.1). During the blooming stage, pollination efficiency is considerably affected by weather conditions because most cultivars are self-incompatible, with pollination being dependent on pollinators such as honey bees (Miyake et al. 1995).

In perennial temperate plants, lateral buds enter endodormancy, a stage of inhibited growth triggered by endogenous factors regardless of conditions, after shoots stop growing and terminal buds are set (Lang 1987). Endodormancy is considered to be a defence mechanism that allows plants to protect themselves against low winter temperatures. Lateral buds under the control of paradormancy (e.g. apical dominance) in summer gradually shift into an endodormant state in autumn. After exposure to a certain amount of chilling, the endodormant buds shift into an ecodormant stage and are able to burst in response to growth-promoting factors such as warm temperatures (Faust et al. 1997; Horvath et al. 2003; Lang 1987) (Fig. 11.2). Responses of deciduous fruit trees such as *P. mume* to low and high temperatures during endodormancy and ecodormancy, respectively, are crucial factors for determining the genotype-dependent blooming time.

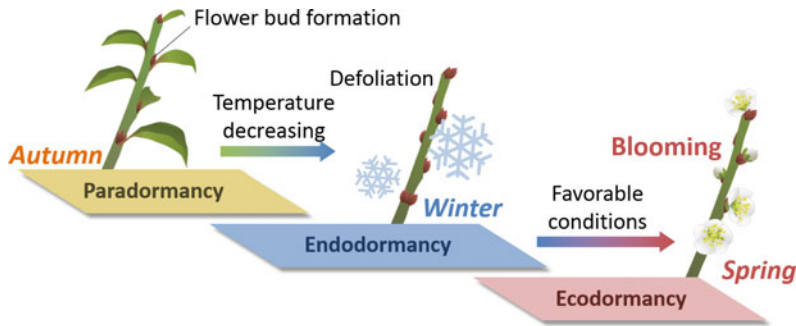**Fig. 11.1** Annual growth cycle of a *P. mume* tree
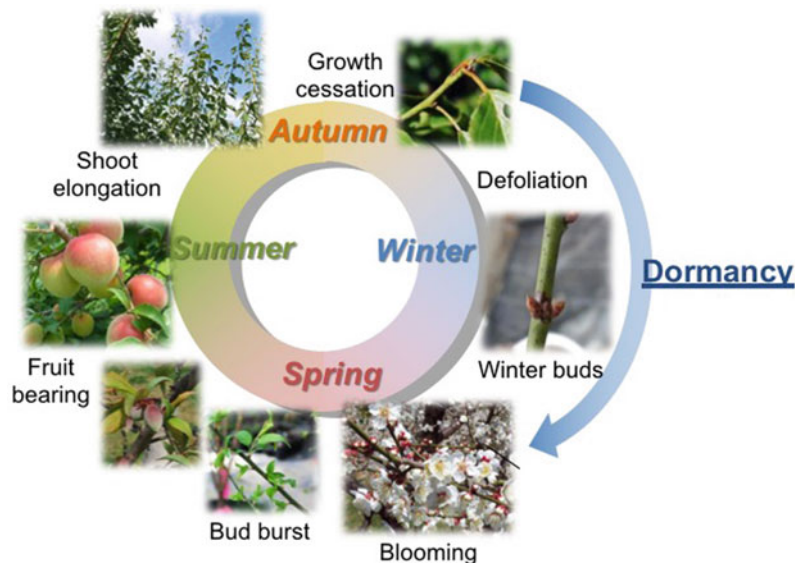
Fig. 11.2 Transition between dormancy phases in the lateral buds of temperate deciduous trees

To determine the genetic factors controlling both temperature requirements in *P. mume* dormant buds, continuous studies have been conducted. The chilling-responsive genes preferentially expressed in buds in the endodormancy-introducing stage have already been identified (reviewed by Yamane 2014). Additionally, quantitative trait locus (QTL) analyses using F_1 segregating populations derived from crosses between low- and high-chill genotypes have recently been performed. Here, we describe the overview of the genomic and genetic studies for bud dormancy of *P. mume*, focusing on dormancy-associated transcriptome analysis and QTL analyses for temperature requirements of dormant buds. Additionally, we discuss the possible candidate genes for dormancy regulators based on the results of synteny analyses among the detected loci located in relative *Prunus* genomes. Since these genetic insights can contribute to facilitate breeding, especially the development of low-chill cultivars, we discuss the applicability of the obtained genetic information to practical breeding of *P. mume*.

11.2 Transcriptome Studies to Unveil the *P. mume* Bud Dormancy Regulation Mechanism

11.2.1 Identification of Dormancy-Associated Genes by Transcriptome Analysis

The molecular basis of bud endodormancy has recently been investigated in several *Prunus* species. In peach [*Prunus persica* (L.) Batsch], a study of the *evergrowing*, in which shoot growth cessation and dormancy induction are inhibited, has revealed that six tandemly arrayed *MADS-box* genes, belonging to the *StMADS11* (*SVP/AGL24*) clade, are associated with terminal bud formation. These six genes have been designated *DORMANCY-ASSOCIATED MADS-box 1-6* (*DAM1-6*) (Bielenberg et al. 2004, 2008). Yamane et al. (2008) identified a *MADS-box* transcription factor that is up-regulated in dormant *P. mume* vegetative buds, using a method of suppression subtractive hybridisation with mirror orientation selection (SSH/MOS). They constructed two different SSH/MOS libraries containing cDNA pools that were expressed in endodormant buds minus those that were expressed in ecodormant or paradormant buds. In both of the two libraries, only one clone that showed higher similarity to *P. persica* dormancy-associated *MADS6* was obtained in common and named as *Prunus mume DAM6* (*PmDAM6*). The deduced amino acid sequence of the isolated gene contained the MADS-box domain at the N-terminal end and the putative I region and K box domain at the middle position, similar to those observed in other MIKC^c-type MADS-box genes. Another study by Sasaki et al. (2011) has revealed that six *DAM* homologs exist in tandem in the *P. mume* genome as well as in peach. Phylogenetic analysis showed that all six *PmDAMs* belong to the *StMADS11* (*SVP/AGL24*) clade of angiosperm *MADS-box* genes similar to *PpDAMs* of peach. A gene expression analysis has indicated that peach *PpDAM5* and *PpDAM6* are down-regulated during dormancy release and also during prolonged

low-temperature treatment, while the expression of *PmDAM4-6* is negatively correlated with chilling accumulation in vegetative buds (Jiménez et al. 2010; Sasaki et al. 2011; Yamane et al. 2011a, b). In *P. mume* flower buds, the down-regulation of *PmDAM3*, *5* and *6* expression synchronised with dormancy release and chilling accumulation was also demonstrated (Kitamura et al. 2016). Additionally, comprehensive expression analyses were performed to reveal the genes preferentially expressed in endodormant buds. The results of a 454-pyrosequencing analysis indicate that *PmDAM6* expression is considerably down-regulated in *P. mume* 'Nanko' flower buds during the transition from endodormancy to ecodormancy (Habu et al. 2012). The authors constructed a *P. mume* dormant bud EST database (JADB) from EST data obtained by this analysis and published it online (http://bioinf.mind.meiji.ac.jp/JADB). Custom microarray analysis could provide the transcript profile of *P. mume* dormant buds, including *PmDAM*s (Fig. 11.3), and provides a list of genes which are down-regulated by prolonged chilling exposure (Habu et al. 2014).

## 11.2.2 The Function of DAM Genes as Dormancy Regulators

Although *DAM* genes have been identified as biomarkers of dormancy phases and chilling responses, their physiological and molecular functions in controlling endodormancy release have not yet been elucidated. Sasaki et al. (2011) successfully produced transgenic poplar (*Populus tremula* × *Populus tremuloides*) plants constitutively expressing the *PmDAM6* gene under the control of the cauliflower mosaic virus 35S promoter (*35S:PmDAM6* poplar) to elucidate the physiological function of the gene. The *35S:PmDAM6* poplars showed growth cessation and terminal bud set under long-day conditions, which are environmental factors promoting shoot growth in poplar (Böhlenius et al. 2006; Ruttink et al. 2007), indicating the growth inhibitory functions of *PmDAM6* (Fig. 11.4; Sasaki et al. 2011; Yamane 2014). Considering the negative correlation between *DAM* gene expression levels

and endodormancy depth, PmDAM6 would inhibit the growth of dormant buds in a dose-dependent manner.

To predict the gene functions, clarification of molecular pathways, including the information of upstream and downstream genes or proteins, and cis-regulating motifs can be important approaches. Horvath et al. (2010) found that *EeDAM1* in leafy spurge (*Euphorbia esula*) contained putative C-repeat/DRE-Binding Factor (CBF) sites within the 2000-bp region upstream of the *EeDAM1* translation initiation codon. Because CBF sites are cis-regulating motifs targeted by the cold/drought stress CBF regulon, cold stress-responsive expression of *EeDAM1* was suggested to be mediated by the CBF protein. Similarly, conserved CBF sites were found within the 1000-bp region upstream of *PmDAM4* to *PmDAM6* translation initiation codons (Sasaki et al. 2011). These findings supported that *PmDAM4* to *6* have a closer relation to endodormancy release and chilling accumulation than *PmDAM1* to *3* in vegetative buds, suggesting that those genes are cold-responsive transcription factors mediated by CBF. However, out of six *CBF* homologs in *P. mume* (*PmCBF1–6*), which have recently been isolated, PmCBF1 showed interprotein interaction with PmDAM1 and 2 (Zhao et al. 2018a, b).

The candidate proteins which can interact with PmDAM6 have been isolated via the yeast two-hybrid technique (Kitamura et al. 2016). In this analysis, to avoid auto-activation in the yeast strain, a cDNA fragment corresponding to the partial amino acid sequence of PmDAM6 (amino acid positions +1 to +176) [PmDAM6 (M, I, K)] was used to construct the bait vector. Screening of more than 420,000 yeast colonies and sequencing of positive colonies indicated that PmDAM6 could interact with one of the MADS-box transcription factors, SUPPRESSOR OF OVER-EXPRESSION OF CONSTANS1 (SOC1) (PmMADS21) (Habu et al. 2012; Xu et al. 2014). In *Arabidopsis*, the SOC1 protein acts as an integrator of several floral inductive pathways, up-regulates downstream gene expression related to floral organ development and promotes flowering (Lee et al. 2000). Additionally, AGL24 and SOC1 heterodimers control floral meristem

A. Microarray
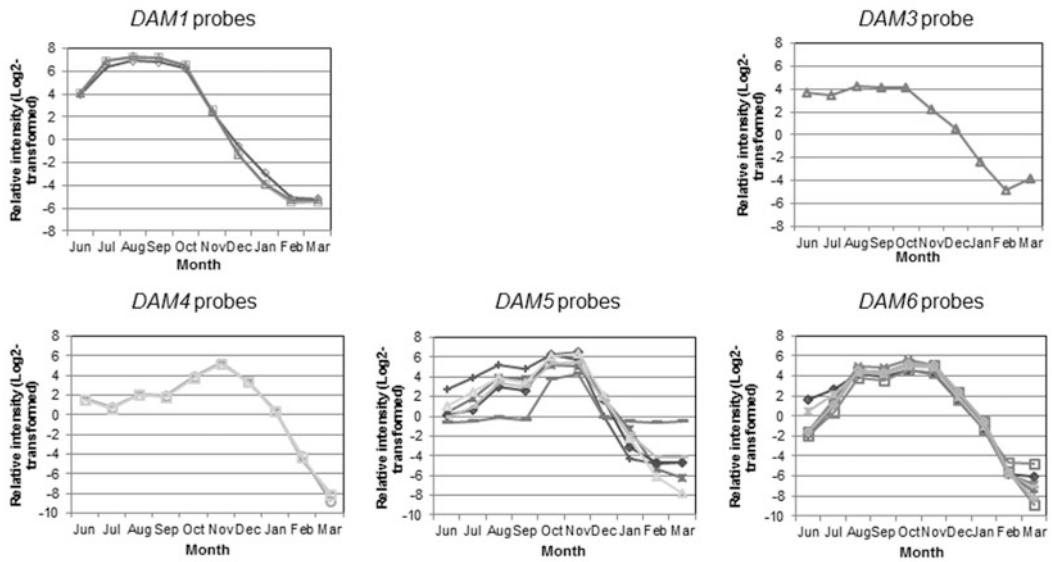
B. RT-qPCR
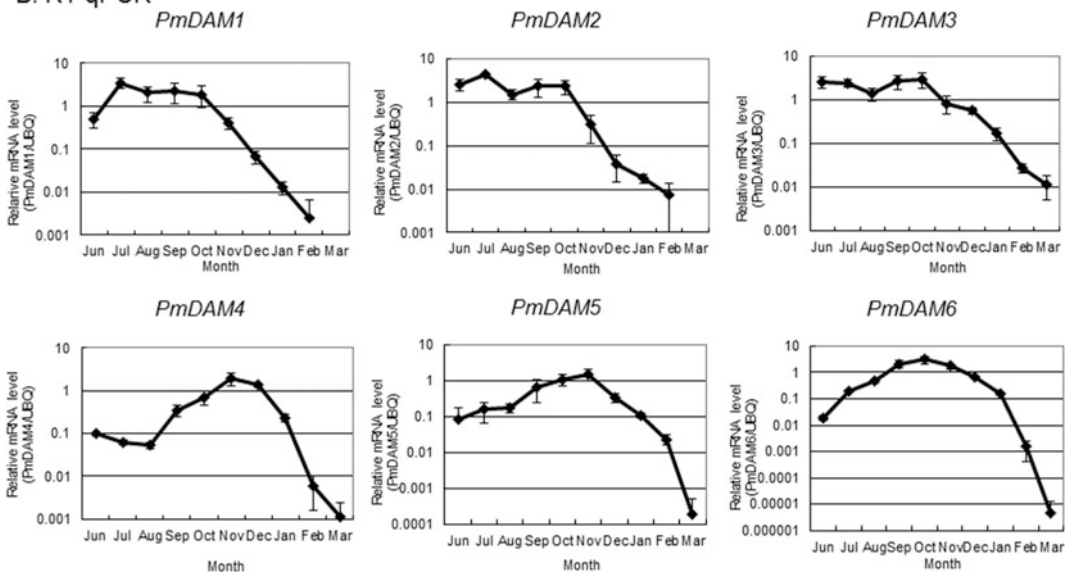
Fig. 11.3 Seasonal expression changes in *PmDAM1–PmDAM6* genes in Japanese apricot 'Nanko'. **a** Microarray results (Habu et al. 2014), **b** RT-qPCR results. In A, three *DAM1*-, one *DAM3*-, three *DAM4*-, six *DAM5*- and seven *DAM6*-annotated probes are shown in each graph. No *DAM2*-annotated probe was loaded on the 60K microarray (from Yamane 2014)

identity by regulating LEAFY (LFY) at the early stages of flower differentiation, and AGL24 and SOC1 expressions are repressed by LFY and AP1 in inflorescences (Lee et al. 2008; Yu et al. 2004). Considering that the seasonal expression pattern of *PmSOC1* is similar to that of *PmDAM6* in flower buds (Kitamura et al. 2016), a dimer of these proteins would be a key factor for regulating flower bud development, dormancy release and blooming time.
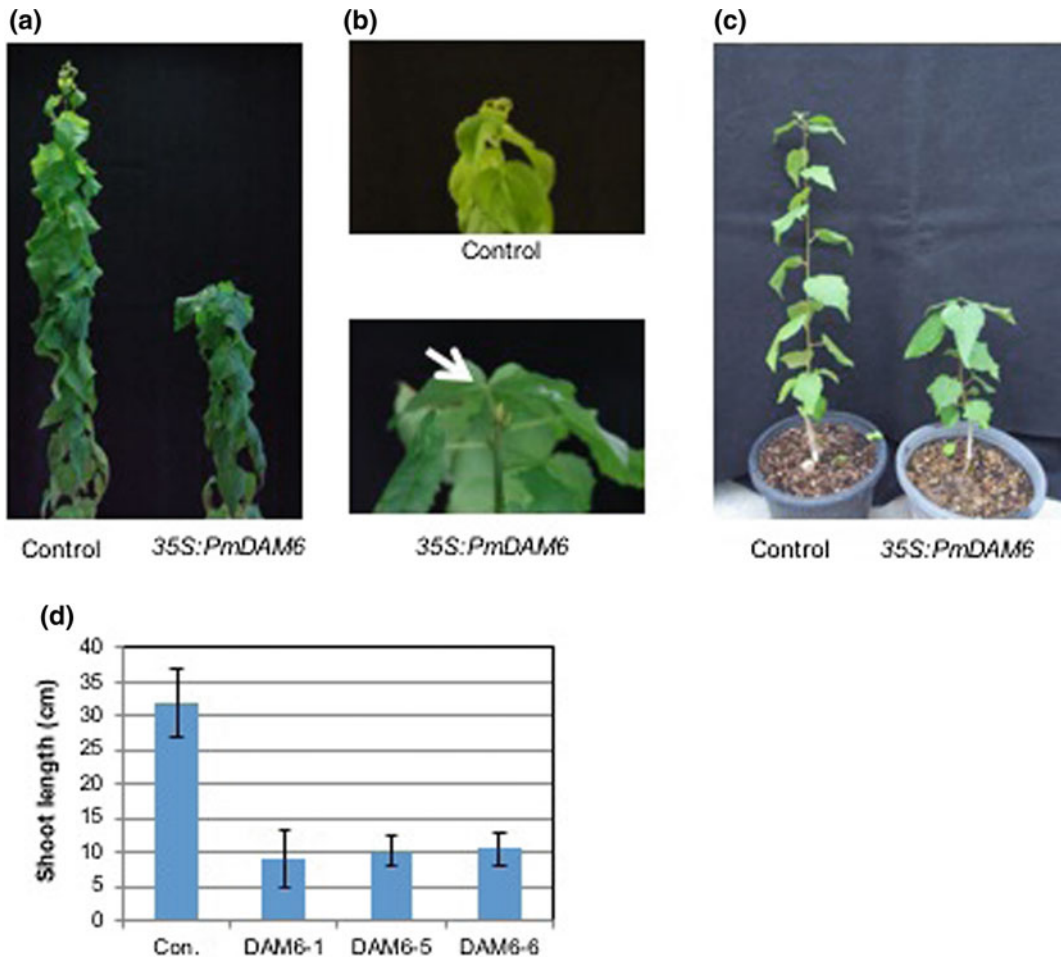
Fig. 11.4 Phenotypic observations of transgenic poplar plants overexpressing *PmDAM6*. **a** Shoot growth of *35S:PmDAM6* poplar was inhibited when grown under LD conditions (Sasaki et al. 2011). **b** Terminal bud set (indicated by arrow) was observed earlier in *35S:PmDAM6*. **c** Shoot growth of *35S:PmDAM6* was also inhibited when grown in isolated greenhouse conditions under natural daylengths. **d** Averaged shoot lengths of three lines of *35S:PmDAM6* poplars (DAM6-1, DAM6-5 and DAM6-6) and one control poplar (*emp*) with S.D. are shown (*n* = 4 − 5) (from Yamane 2014)

11.2.3 Relationship Between DAM and Flowering-Related Gene Family

FLOWERING LOCUS T (*FT*) and *TERMINAL FLOWER1* (*TFL1*) in *Arabidopsis* are key genes well-known as the transcription factors inducing and inhibiting flowering, respectively (Kardailsky et al. 1999; Kobayashi et al. 1999; Shannon and Meeks-Wagner 1993). The production of the FT protein is up-regulated by CONSTANS (CO) in response to long-day conditions in leaves, and the protein is transported to apical meristems to promote flower differentiation (Corbesier et al. 2007). In aspen trees, the CO/FT module not only controls flowering, but also mediates the cessation of growth under short-day conditions, suggesting its regulatory effects on dormancy (Böhlenius et al. 2006). Additionally, the overexpression of *Populus CENTRORADIALIS* (*CEN*), which is a *TFL1* homolog, results in delayed bud burst (Mohamed et al. 2010). These reports suggest that *FT* and *TFL1* promote and inhibit bud growth, respectively,

Fig. 11.5 Seasonal *PmTFL1* expression level changes in Japanese apricot 'Nanko' flower buds (FB) and vegetative buds (VB). Values are presented as the means of three replicates normalised against *ACTIN* expression levels. Error bars represent standard errors
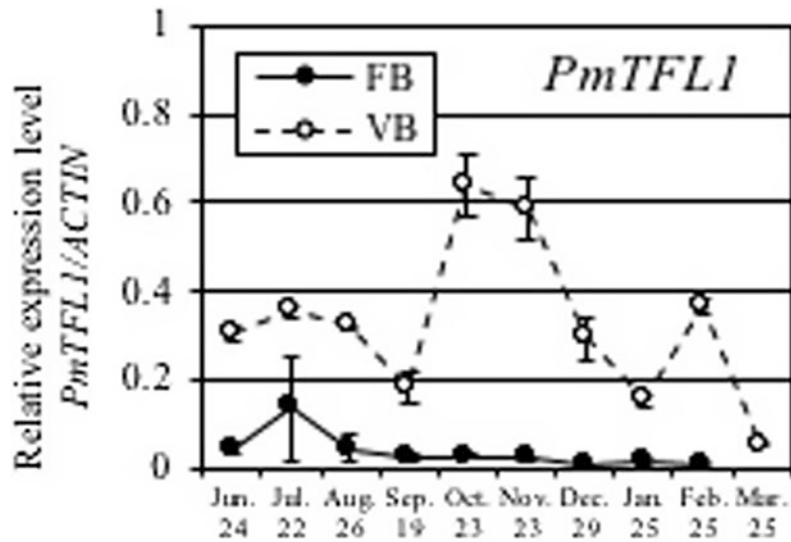
during dormancy. In *P. mume*, *FT* and *TFL1* homologs (*PmFT* and *PmTFL1*) have been identified, and it has been concluded these genes may be related to transitions in the reproductive/ vegetative and adult/juvenile phases (Esumi et al. 2009, 2010).

According to a previous study, SVP represses FT in *Arabidopsis* (Lee et al. 2007), and the DAM proteins of leafy spurge also negatively regulate *FT* expression (Horvath et al. 2008, 2010). Transgenic studies have confirmed that the ectopic expression of poplar *FT1* in plum and the overexpression of poplar *CEN/TFL1* in poplar decrease and increase the CR for endodormancy release, respectively (Mohamed et al. 2010; Srinivasan et al. 2012). Moreover, Ito et al. (2015) concluded that the expression patterns of pear *FT* and *TFL1* genes (i.e. *PpFT1a*, *PpFT2a*, *PpTFL1-1a* and *PpTFL1-2a*) in flower buds suggest that a balance between the expression of *FT* and *TFL1* influences the CR and flowering time. These studies collectively suggest the involvement of FT/TFL1 in the regulation of bud dormancy in perennial woody plants. However, the results of the expression analyses in our study revealed that *PmDAM* genes and *PmFT* were highly expressed simultaneously from October to November in *P. mume* flower buds (Esumi et al. 2009; Sasaki et al. 2011), which contradicts the idea that *PmFT* expression

is repressed by PmDAM proteins and is associated with dormancy regulation in *P. mume*. The decreasing *PmTFL1* expression levels in vegetative buds, observed from November to March (Fig. 11.5), seem to indicate that this gene is involved in increasing the CR, as reported for poplar (Mohamed et al. 2010). However, *PmTFL1* does not appear to be important for regulating flower bud dormancy because there was no correlation between its transcript abundance and flower bud dormancy depth. The PmDAM proteins might suppress organ enlargement throughout endodormancy and ecodormancy, independent of flowering or bud burst pathways mediated via PmFT/PmTFL1 in *P. mume*.

11.2.4 Genetic Analysis to Understand the Regulation of Bud Dormancy and Chilling Requirements in Prunus Mume

11.2.4.1 Quantitative Trait Locus Analysis Using the GBS Technique

Genetic approaches, including quantitative trait locus (QTL) analysis, are often used to reveal the

genetic factors controlling quantitative traits, such as chilling requirement (CR) and blooming date (BD). The first comprehensive QTL study of flower bud CR in *Prunus* was published by Fan et al. (2010). The authors detected major QTLs for CR and BD on the *Prunus* ($n = 8$) reference map. The QTLs for CR and BD have also been identified in almond (*Prunus dulcis*) (Sánchez-Pérez et al. 2012), apricot (*Prunus armeniaca*) (Dirlewanger et al. 2012; Olukolu et al. 2009; Socquet-Juglard et al. 2013) and sweet cherry (*Prunus avium*) (Castéde et al. 2014). Fine mapping and dissection of identified QTLs have been performed in peach (Zhebentyayeva et al. 2014). However, QTL analysis requires the generation of segregating populations, high-density genetic mapping and phenotype observation. In trees, the development of segregating populations takes time and is labour-intensive because of the large plant size and the long generation times, mainly because of a long juvenile phase. Consequently, a pseudo-testcross strategy is often used for linkage analyses of woody perennials, because F_1 populations derived from the cross of two heterologous parent genotypes show segregation for many traits (Grattapaglia and Sederoff 1994). Numerous QTL analyses based on this strategy have been conducted using full-sibling fruit trees (Ban et al. 2014; Kunihisa et al. 2014; Siviero et al. 2003; Weber et al. 2003; Yamamoto et al. 2014; reviewed by Iwata et al. 2016). This method is considered to be effective for genetic studies, especially in *P. mume*, because most available cultivars are self-incompatible, which makes generating an F_2 segregating population difficult. In fact, several genetic analyses of *P. mume* have been conducted using F_1 segregating populations (Sun et al. 2013; Zhang et al. 2015).

A genotyping-by-sequencing (GBS) technique, which is based on single nucleotide polymorphisms (SNPs) obtained from high-throughput next-generation sequencing techniques (e.g. the Illumina HiSeq platform), has recently been developed (Elshire et al. 2011). This technique is a highly efficient, cost-effective way to construct a genome-wide, high-density linkage map. The GBS approach is suitable even for highly heterozygous plant species, such as

fruit trees, and has already been applied to develop DNA markers associated with disease resistance and fermentation characteristics in grapevine (*Vitis vinifera* L.) (Barba et al. 2014; Yang et al. 2016) and fruit skin colour in apple (Gardner et al. 2014). Bielenberg et al. (2015) also used a GBS method to develop an SNP-based linkage map and to detect QTLs for CR and blooming time in peach.

11.2.4.2 QTLs for Temperature Requirements in *P. mume*

In *P. mume*, QTL analysis for dormancy-related traits has been conducted and reported by Kitamura et al. (2018). The authors collected phenotype data, including CR and heat requirement for bud break (bR), of both leaf and flower buds, blooming date and leafing date for several years. In addition, because down-regulation of *PmDAM* genes was well-correlated with both chilling and heat accumulation in leaf buds (Sasaki et al. 2011; Kitamura et al. 2016), the relative expression level of *PmDAM6* in leaf buds was also determined and used as one of the dormancy-related phenotypes for QTL analysis (Kitamura et al. 2018).

The genotyping-by-sequencing technique was used to construct two high-density genetic maps. The NKSC map, derived from a population from a cross between high-chill 'Nanko' and low-chill 'SC', consisted of eight linkage groups with 1233 SNP markers covering 1363.1 cM. The NINK map, derived from a population from a cross between low-chill 'Ellching' and 'Nanko', consisted of eight linkage groups with 995 SNP markers covering 1157 cM markers (Fig. 11.6; Yamane et al. in press). Significant correlations were observed between CRL and LD values in the NKSC population and among HRL, LD and PmDAM6 expression values in the NINK population. The QTL analyses conducted by Kitamura et al. (2018) successfully identified significant QTL loci for LD and CRL in NKSC and for HRL, LD and DAM6 expression levels in NINK. All these QTL regions overlapped with each other and were localised in the bottom region of linkage group (LD) 4 (Pm3
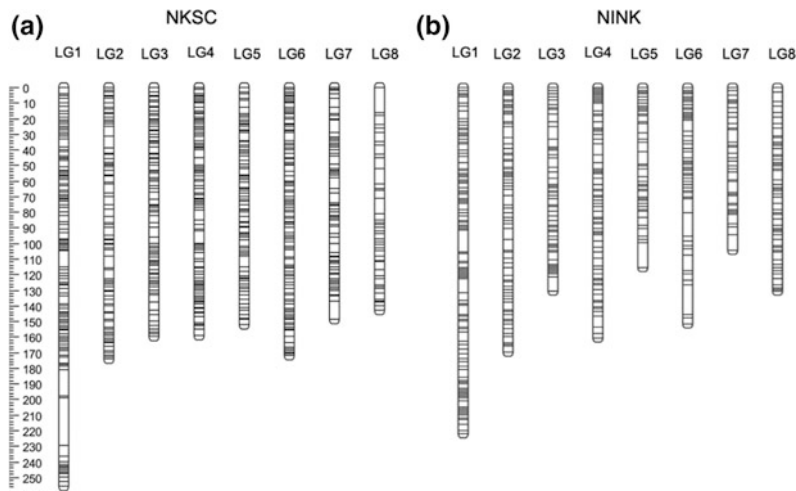
Fig. 11.6 Integrated genetic map constructed with SNP markers for NKSC (**a**) and NINK populations (**b**)

Table 11.1 QTLs detected in segregated populations of *P. mume* (Kitamura et al. 2018)

Population	Trait name	Year	LG	Peak LOD	Nearest marker[a]	PVE (%)[b]
NKSC ('Nanko' × 'SC')	Leafing date	2013	4	14.12	Pm4_13381114	49.2
		2014	4	9.82	Pm4_17717558	37.9
	Chilling requirement of leaf bud	2013	4	8.07	Pm4_13920301	32.1
			7	6.10	Pm7_11223007	25.4
		2014	4	4.18	Pm4_17717558	18.3
	Blooming date	2013	1	5.03	Pm1_20859384	21.6
NINK ('Ellching' × 'Nanko')	Heat requirement for bud break of leaf buds	2011	4	8.56	Pm4_16042141	74.3
		2012	4	7.48	Pm4_16042141	44.8
		2014	4	15.46	Pm4_18006903	74.6
	Leading date	2013	4	11.31	Pm4_22013736	58.6
		2014	4	9.82	Pm4_22013921	53.5
		2015	4	4.83	Pm4_13381114	31.4
	Relative expression level of PmDAM6	2011	4	8.10	Pm4_16042141	68.8
		2014	4	10.43	Pm4_10173963	61.0

[a]Pm1, Pm4, and Pm7 corresponds to Pm2, Pm3, and Pm8 chromosomes, respectively, in Zhang et al. 2012
[b]Phenotypic variation explained by each QTL

chromosome in Zhang et al. (2012)) throughout the experimental seasons, with high values of phenotypic variance explained (PVE) by the QTLs (Table 11.1; Fig. 11.7). The QTLs for CRL on LG7 and for BD on LG1 were also detected in the NKSC population. The finding that *PmDAM6* eQTL overlapped with QTLs controlling HRL and LD supported the idea that PmDAM6 acts as a dose-dependent growth inhibitor during both the dormancy release phase and the subsequent bud break phase. However, because the actual gene location of *PmDAM6*, at the bottom region of LG1, was not identified as QTL for dormancy traits, cis-regulatory and coding regions of *PmDAM6* may not explain the segregation of *PmDAM6* expression levels in this
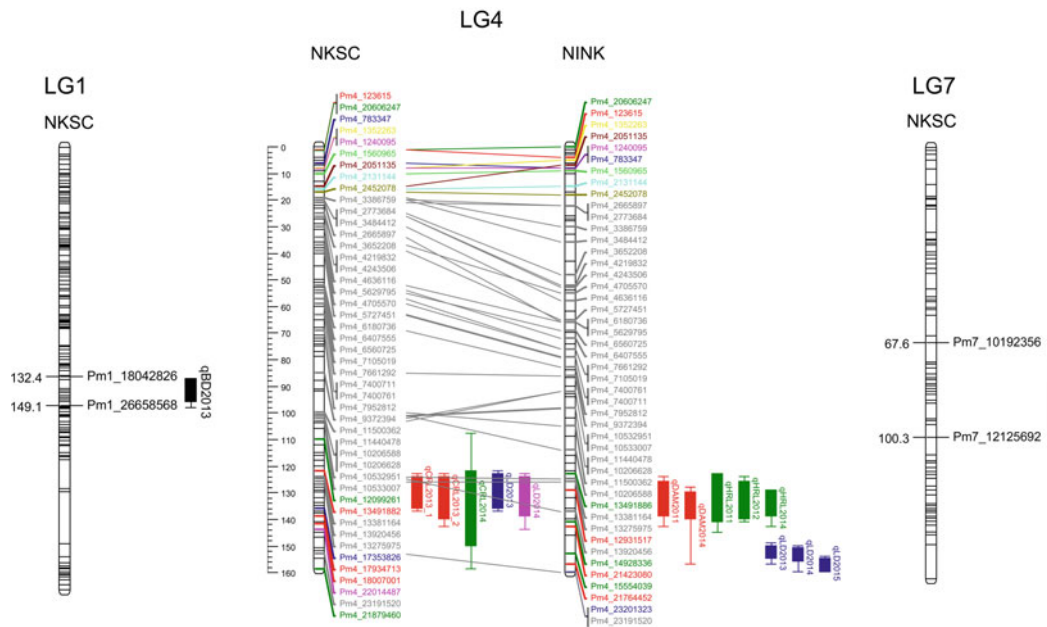
Fig. 11.7 Location of QTLs on LG1, LG4 and LG7 controlling dormancy traits detected in the genetic linkage maps of NKSC and NINK. Boxes and bars indicate 1- and 1.5-LOD confidence intervals from peak LOD values, respectively. QTLs for CRL in the 2013–14 season were detected on LG4 and LG7, and QTL for BD in the 2013–2014 season was detected on LG1, while QTLs for CRL in the 2014–2015 season, HRL, LD and DAM were detected only on LG4

population. In peach, on the other hand, co-localisation of the significant QTL region on LG1 for CR and BD with the DAM6 gene was reported (Bielenberg et al. 2015). Co-localisation of QTLs for CR and HR for leaf buds, LD and DAM6 expression in leaf buds at LG4 suggest that factor(s) on the LG4 QTL control endodormancy and ecodormancy release, the PmDAM6 transcripts levels in leaf buds and leaf bud break in *P. mume*.

For flower bud dormancy traits, no significant QTLs were identified in both populations except for BD in NKSC. The reasons were unclear, but possible explanations are that leaf bud break and flower bud break may be genetically differently regulated, and that the narrower segregation range for flower buds could not be effectively used to identify the same QTL as leaf buds. In the former case, flower bud break may be controlled by multiple minor factors undetectable in our QTL analysis or, alternatively, the alleles of major QTLs influencing blooming were homozygous in both parents of the analysed population. In the latter case, the phenotyping methodology for flower bud dormancy traits should be reconsidered. Further studies will be required to evaluate the relationship between leaf bud dormancy and flower dormancy in *P. mume*.

11.2.4.3 Syntenies to *Prunus* Relatives

Previously, QTLs for CR and BD have been identified on the LG4 of almond (Sánchez-Pérez et al. 2012) and sweet cherry (Castéde et al. 2014). The QTLs for apricot blooming initiation time are located on LG1 and LG4 (Dirlewanger et al. 2012; Olukolu et al. 2009; Socquet-Juglard et al. 2013). In peach, major QTLs for CR and BD have been detected on LG1, LG4 and LG7 (Bielenberg et al. 2015; Fan et al. 2010; Zhebentyayeva et al. 2014). In all these species, significant QTLs, which have relatively higher PVE values, were detected on LG4, and QTLs for CRL and LD on the LG4 of *P. mume* have

Table 11.2 List of dormancy characteristic-related QTLs located on LG4 of the *Prunus* genome

Species	Population	Trait name[a]	Year	References
Almond	'R1000' × 'Desmayo Largueta'	CU requirement	2009, 2010	Sánchez-Pérez et al. (2012)
		Flowering time	2009, 2010	
Apricot	'Goldrich' × 'Moniqui'	Flowering date	2006, 2007, 2009	Dirlewanger et al. (2012)
	'Harostar' × 'Rouge de Mauves'	Time to IRB[b]	2011	Socquet-Juglard et al. (2013)
Peach	Selfed F_2 of 'Contender' × 'Fla.92-2C'	Chilling requirement[c]	2008, 2009	Fan et al. (2010)
		Bloom date	2006, 2007, 2008, 2009	Zhebentysyevs et al. (2014)
	'Ferjalou Jalousia' × 'Fantasia'	Flowering date	2002, 2004, 2008	Dirlewanger et al. (2012)
	(('Summergrand' × *P. davidiana* P1908) × 'Summergrand') × 'Zephyr'	Flowering date	2006	
	Selfed F_2 of 'Hakuo' × 'UFGold'	Chilling requirement	2008, 2009	Bielenberg et al. (2015)
		Bloom date	2006, 2007, 2008, 2009, 2012	
Sweet cherry	'Regina' × 'Lapins'	Flowering date	2006, 2008, 2009, 2010	Dirlewanger et al. (2012)
	'Regina' × 'Lapins'	Flowering time	2006, 2008, 2009, 2010, 2011, 2012	Castéde et al. (2014)
	'Regina' × 'Garnet'	Chilling requirement	2010, 2011, 2012	
		Heat requirement	2011	
		Flowering time	2008, 2009, 2010, 2011, 2012	

[a]Trait names are represented in accordance with the reference articles
[b]Initial reproductive bud break
[c]The QTLs were detected with each CR calculation method based on the <7.2 °C model (Weinberger 1950), the 0–7.2 °C model (Eggert 1951), the Utah model (Richardson et al. 1974) and the Dynamic model (Fishman et al. 1987)

also been identified (Table 11.1). The findings of the dormancy-related QTLs located on LG4 in *Prunus* fruit tree species are summarised in Table 11.2.

To verify the interspecific co-localisation of QTLs, genome synteny analysis was performed (Kitamura et al. 2018). The sequences of the 1-kb region flanking the SNP site within QTL intervals retrieved from the *P. mume* genome were aligned to the sweet cherry genome, *P. avium* v1.0 pseudomolecule (Shirasawa et al. 2017) or to the peach genome, *P. persica* v2.0 scaffold (International Peach Genome Initiative et al. 2013) (Fig. 11.8). This analysis revealed that the position of the QTL on LG4 overlapped with QTLs of CR in peach (Bielenberg et al. 2015), but not with those in sweet cherry (Castéde et al. 2014), suggesting that endodormancy in peach and *P. mume* is regulated by factors which may be common to an extent, located on LG4 (Fig. 11.8). However, the QTL for BD on LG1 did not overlap with the peach QTL for CR and BD on LG1.
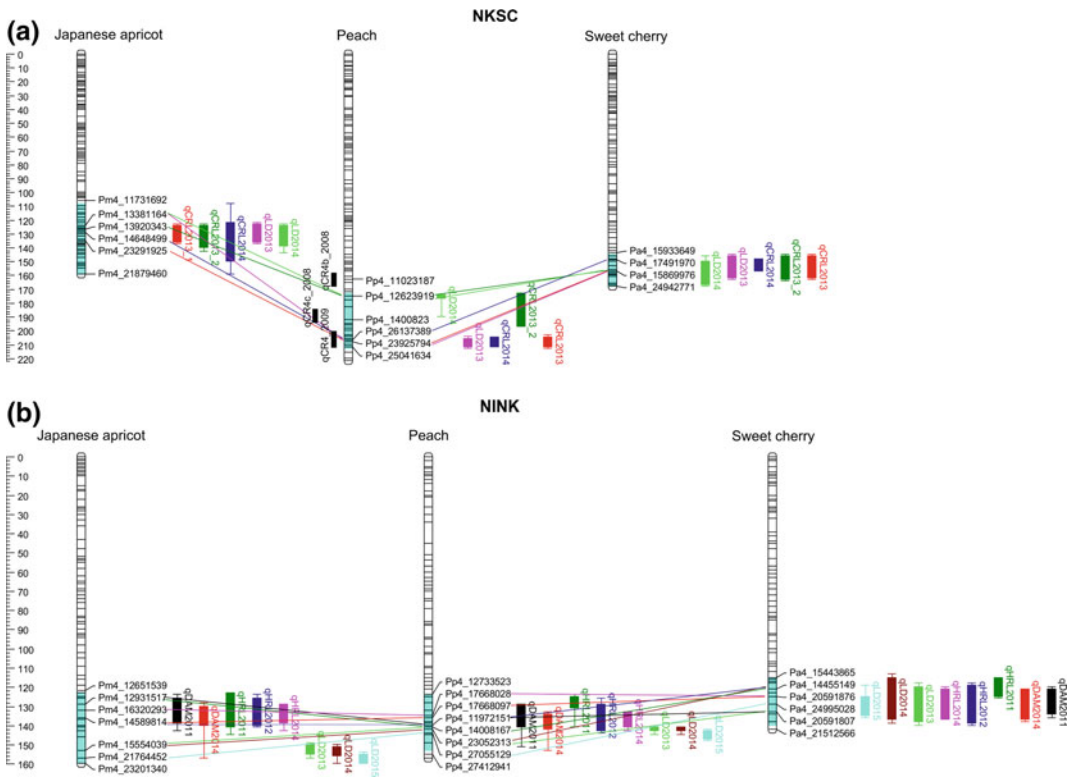
Fig. 11.8 Synteny of the QTL regions covering all detected QTLs on LG4 among Japanese apricot, peach and sweet cherry. All 1-LOD confidence intervals of the QTLs covering the widest regions in the linkage maps across all of the QTLs detected in LG4, namely qCRL2014 for the NKSC map and qDAM2014 and qLD2015 for the NINK map, were compared. The flanking markers of the QTL regions covering all detected QTLs and the nearest markers to peak positions of each QTL are indicated. In the NKSC map, the nearest markers to qCRL2013, qCRL2014 and pLD2013 (qLD2014) were Pm4 (Pm3 in Zhang et al. (2012))_13920343, Pm4_14648499 and Pm4_13381164, respectively. In the NINK map, the nearest markers to qDAM2011 (qHRL2011), qDAM2014 (qHRL2012), qHRL2014, qLD2013 (qLD2014) and qLD 2015 were Pm4_12931517, Pm4_14589814, Pm4_16320293, Pm3 (LG4)_15554039 and Pm4_21764452. The previously identified three LG4 QTL regions for CR in peach (Bielenberg et al. 2015) are shown in the left side of the NKSC peach map

11.3　Impacts on *P. mume* Breeding

11.3.1　Breeding Strategy for Global Climate Change Mitigation

As described in the introduction, global warming can affect fine bud burst, blooming and, consequently, stable fruit production in temperate fruit tree species. The fifth assessment report of the Intergovernmental Panel on Climate Change (IPCC) stated that the global average temperature might rise by 4.8 °C during the twenty-first century (RCP8.5 scenario) (IPCC 2013). Higher temperatures in winter may inhibit the fulfilment of the chilling requirement for bud endodormancy release, thereby causing delays in blooming or bud burst. To ensure stable fruit production in the future, the breeding of low-chill cultivars of temperate fruit tree species is currently becoming one of the important breeding objectives.

In Japan, new low-chill peach cultivars have been already developed, using cultivars originated from Florida and Brazil as cross parents

(Beppu et al. 2014; Sawamura et al. 2017). Although the blooming time of *P. mume* depends more heavily on HR than on CR (Kitamura et al. 2017), the delay of endodormancy release time due to a warmer climate might cause sporadic blooming or bud burst, resulting in unstable fruit development. Low-chill *P. mume* cultivars have not yet been bred strategically, but some native low-chill cultivars originating from Taiwan and China, such as 'Ellching' and 'SC', used in the QTL analysis described above, are known. In the Wakayama prefecture in Japan, a breeding program for low-chill cultivars using such foreign genetic resources is in progress (Japanese Apricot Laboratory, personal communications).

11.3.2 Use of Current Understanding of Genetic and Molecular Regulation Mechanism of Dormancy to Breeding in *P. mume*

Several genetic factors controlling or being highly related to dormancy have been identified in *P. mume*. Relative expression levels of *PmDAM* and *PmSOC1* genes are synchronically down-regulated during dormancy release, and these proteins are supposed to function in a dosage-dependent manner. In apricot, PmSOC1 genotypes correlate with the different amounts of CR (Trainin et al. 2013). In addition to *PmDAM* and *PmSOC1* genotypes, genotypes in the genomic region of LG4 QTL for dormancy may possibly have genetic relation with dormancy traits in *P. mume*, and respective studies are ongoing.

Regarding the cross-breeding of fruit trees, the seedling selection at a younger stage would be desired for efficient breeding because their large plant size and long juvenile phase cause higher labour costs. Molecular markers such as SNPs and PCR polymorphisms are useful for the selection of young seedlings. Moreover, since the down-regulation patterns of *PmDAMs* and *PmSOC1* expression levels in buds during dormancy release vary depending on the genotype,

the expression levels of these genes could be used as indices of the dormancy depth for each individual.

Because of the progress of massive sequencing and sequence-processing technologies, we are now in the position to carry out comprehensive genome analyses of many germplasm collections more easily and cost-effectively. The genome information is applicable for the identification of genetic loci regulating various agronomic traits by using the genome-wide association study (GWAS) approach, as well as for efficient breeding by using genomic selection (GS). In practical breeding, GWAS and GS have recently been used in several fruit tree species such as citrus and apple (Minamikawa et al. 2017; Moriya et al. 2017; Urrestarazu et al. 2017). In *P. mume*, the genome analysis of over 100 different genotypes is ongoing (Japanese Apricot Laboratory, personal communications), with the aim to develop new cultivars that adapt to abiotic stress and future climatic changes. These studies may contribute not only to sustainable fruit production and flower-ornamental cultures of *P. mume*, but also shed light on the genetic mysteries of various phenotypes in *P. mume*.

References

Atkinson CJ, Brennan RM, Jones HG (2013) Declining chilling and its impact on temperate perennial crops. Environ Exp Bot 91:48–62

Ban Y, Mitani N, Hayashi T, Sato A, Azuma A, Kono A, Kobayashi S (2014) Exploring quantitative trait loci for anthocyanin content in interspecific hybrid grape (*Vitis labruscana* × *Vitis vinifera*). Euphytica 198:101–114

Barba P, Cadle-Davidson L, Harriman J, Glaubitz JC, Brooks S, Hyma K, Reisch B (2014) Grapevine powdery mildew resistance and susceptibility loci identified on a high-resolution SNP map. Theor Appl Genet 127:73–84

Beppu K, Yagata M, Manabe T, Kataoka I (2014) New lower-chilling peach cultivar 'KU-PP1'. Hort Res (Japan) 13(Suppl 2):362 (in Japanese)

Bielenberg DG, Wang Y, Fan S, Reighard GL, Scorza R, Abbott AG (2004) A deletion affecting several gene candidates is present in the evergrowing peach mutant. J Hered 95:436–444

Bielenberg DG, Li Z, Zhebentyayeva T, Fan S, Reighard GL, Scorza R, Abbott AG (2008) Sequencing and annotation of the evergrowing locus in peach [*Prunus persica* (L.) Batsch] reveals a cluster of six MADS-box transcription factors as candidate genes for regulation of terminal bud formation. Tree Genet Genomes 4:495–507

Bielenberg DG, Rauh B, Fan S, Gasic K, Abbott AG, Reighard GL, Okie WR, Wells CE (2015) Genotyping by sequencing for SNP-based linkage map construction and QTL analysis of chilling requirement and bloom date in peach [*Prunus persica* (L.) Batsch]. PLoS One 10:e0139406. https://doi.org/10.1371/journal.pone.0139406

Böhlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O (2006) CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. Science 312:1040–1043

Castéde S, Campoy JA, Quero-Garcia J, Le Dantec L, Lafargue M, Barreneche T, Wenden B, Dirlewanger E (2014) Genetic determinism of phenological traits highly affected by climate change in *Prunus avium*: flowering date dissected into chilling and heat requirements. New Phytol 202:703–715

Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle I, Giakountis A, Farrona S, Gissot L, Turnbull C, Coupland G (2007) FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. Science 316:1030–1033

Dirlewanger E, Quero-García J, Le Dantec L, Lambert P, Ruiz D, Dondini L, Illa E, Quilot-Turion B, Audergon JM, Tartarini S, Letourmy P, Arùs P (2012) Comparison of the genetic determinism of two key phenological traits, flowering and maturity dates, in three *Prunus* species: peach, apricot and sweet cherry. Heredity 109:280–292

Du D, Hao R, Cheng T, Pan H, Yang W, Wang J, Zhang Q (2013a) Genome-wide analysis of the AP2/ERF gene family in *Prunus mume*. Plant Mol Biol Rep 31:741–750

Du D, Zhang Q, Cheng T, Pan H, Yang W, Sun L (2013b) Genome-wide identification and analysis of late embryogenesis abundant (LEA) genes in *Prunus mume*. Mol Biol Rep 40:1937–1946

Eggert FP (1951) A study of rest varieties of apple and in other fruit species grown in New York State. Proc Amer Soc Hort Sci 51:169–178

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379. https://doi.org/10.1371/journal.pone.0019379

Esumi T, Hagihara C, Kitamura Y, Yamane H, Tao R (2009) Identification of an FT ortholog in Japanese apricot (*Prunus mume* Sieb. et Zucc.). J Hort Sci Biotechnol 84:149–154

Esumi T, Kitamura Y, Hagihara C, Yamane H, Tao R (2010) Identification of a TFL1 ortholog in Japanese apricot (*Prunus mume* Sieb. et Zucc.). Sci Hort 25:608–616

Fan S, Bielenberg DG, Zhebentyayeva TN, Reighard GL, Okie WR, Holland D, Abbott AG (2010) Mapping quantitative trait loci associated with chilling requirement, heat requirement and bloom date in peach (*Prunus persica*). New Phytol 185:917–930

Faust M, Erez A, Rowland LJ, Wang SY, Norman HA (1997) Bud dormancy in perennial fruit trees: physiological basis for dormancy induction, maintenance, and release. HortScience 32:623–629

Fishman S, Erez A, Couvillon GA (1987) The temperature dependence of dormancy breaking in plants: mathematical analysis of a two-step model involving cooperative transition. J Theor Biol 124:473–483

Gardner KM, Brown P, Cooke TF, Cann S, Costa F, Bustamante C, Velasco R, Troggio M, Myles S (2014) Fast and cost-effective genetic mapping in apple using next-generation sequencing. G3-Genes Genomes Genet 4:1681–1687

Grattapaglia D, Sederoff R (1994) Genetic-linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137:1121–1137

Habu T, Yamane H, Igarashi K, Hamada K, Yano K, Tao R (2012) 454-pyrosequencing of the transcriptome in leaf and flower buds of Japanese apricot (*Prunus mume* Sieb. et Zucc.) at different dormant stages. J Jpn Soc Hortic Sci 81:239–250

Habu T, Yamane H, Sasaki R, Yano K, Fujii H, Shimizu T, Yamamoto T, Tao R (2014) Custom microarray analysis for transcript profiling of dormant vegetative buds of Japanese apricot during prolonged chilling exposure. J Jpn Soc Hort Sci 83:1–16

Horvath DP, Anderson JV, Chao WS, Foley ME (2003) Knowing when to grow: signals regulating bud dormancy. Trends Plant Sci 8:534–540

Horvath DP, Chao WS, Suttle JC, Thimmapuram J, Anderson JV (2008) Transcriptome analysis identifies novel responses and potential regulatory genes involved in seasonal dormancy transitions of leafy spurge (*Euphorbia esula* L.). BMC Genomics 9:536

Horvath DP, Sung S, Kim D, Chao W, Anderson J (2010) Characterization, expression and function of DORMANCY ASSOCIATED MADS-BOX genes from leafy spurge. Plant Mol Biol 73:169–179

International Peach Genome Initiative, Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F, Zuccolo A, Rossini L, Jenkins J, Vendramin E, Meisel LA, Decroocq V, Sosinski B, Prochnik S, Mitros T, Policriti A, Cipriani G, Dondini L, Ficklin S, Goodstein DM, Xuan P, Fabbro CD, Aramini V, Copetti D, Gonzalez S, Horner DS, Falchi R, Lucas S, Mica E, Maldonado J, Lazzari B,

Bielenberg D, Pirona R, Miculan M, Barakat A, Testolin R, Stella A, Tartarini S, Tonutti P, Arus P, Orellana A, Wells C, Main D, Vizzotto G, Silva H, Salamini F, Schmutz J, Morgante M, Rokhsar DS (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet 45:487–494

IPCC (2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp

Ito A, Saito T, Sakamoto D, Sugiura T, Bai S, Moriguchi T (2015) Physiological differences between bud breaking and flowering after dormancy completion revealed by DAM and FT/TFL1 expression in Japanese pear (*Pyrus pyrifolia*) Tree Physiol 36:109–120

Iwata H, Minamikawa MF, Kajiya-Kanegae H, Ishimori M, Hayashi T (2016) Genomics-assisted breeding in fruit trees. Breed Sci 66:100–115

Jiménez S, Reighard GL, Bielenberg DG (2010) Gene expression of DAM5 and DAM6 is suppressed by chilling temperatures and inversely correlated with bud break rate. Plant Mol Biol 73:157–167

Kardailsky I, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT, Chory J, Harrison MJ, Weigel D (1999) Activation tagging of the floral induced FT. Science 286:1962–1965

Kitamura Y, Habu T, Yamane H, Nishiyama S, Kajita K, Sobue T, Kawai T, Numaguchi K, Nakazaki T, Kitajima A, Tao R (2018) Identification of QTLs controlling chilling and heat requirements for dormancy release and bud break in Japanese apricot (*Prunus mume*). Tree Genet Gen 14:33

Kitamura Y, Takeuchi T, Yamane H, Tao R (2016) Simultaneous downregulation of DORMANCY-ASSOCIATED MADS-box6 and SOC1 during dormancy release in Japanese apricot (*Prunus mume*) flower buds. J Hort Sci Biotechnol 91:476–482

Kitamura Y, Yamane H, Yukimori A, Shimo H, Numaguchi K, Tao R (2017) Blooming date predictions based on Japanese apricot 'Nanko' flower bud responses to temperatures during dormancy. HortScience 52:366–370

Kobayashi Y, Kaya H, Goto K, Iwabuchi M, Araki T (1999) A pair of related genes with antagonistic roles in mediating flowering signals. Science 286:1960–1962

Kunihisa M, Moriya S, Abe K, Okada K, Haji T, Hayashi T, Kim H, Nishitani C, Terakami S, Yamamoto T (2014) Identification of QTLs for fruit quality traits in Japanese apples: QTL for early ripening are tightly linked to preharvest fruit drop. Breed Sci 64:240–251

Lang GA (1987) Dormancy–a new universal terminology. HortScience 22:817–820

Lee H, Suh S, Park E, Cho E, Ahn JH, Kim S, Lee I (2000) The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in *Arabidopsis*. Genes Dev 14:2366–2376

Lee J, Oh M, Park H, Lee I (2008) SOC1 translocated to the nucleus by interaction with AGL24 directly regulates LEAFY. Plant J 55:832–843

Lee JH, Yoo SJ, Park SH, Hwang I, Lee JS, Ahn JH (2007) Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. Genes Dev 21:397–402

Lu J, Wang T, Xu Z, Sun L, Zhang Q (2015) Genome-wide analysis of the GRAS gene family in *Prunus mume*. Mol Genet Gen 290:303–317

Minamikawa MF, Nonaka K, Kaminuma E, Kajiya-Kanegae H, Onogi A, Goto S, Yoshioka T, Imai A, Hamada H, Hayashi T, Matsumoto S, Katayose Y, Toyoda A, Fujiyama A, Nakamura Y, Shimizu T, Iwata H (2017) Genome-wide association study and genomic prediction in citrus: potential of genomics-assisted breeding for fruit quality traits. Sci Report 7:4721

Miyake M, Yamaguchi M, Haji T (1995) The self-compatibility in mume cultivars. J Jpn Soc Hort Sci 64(Suppl 2):116–117 (in Japanese)

Mohamed R, Wang CT, Ma C, Shevchenko O, Dye SJ, Puzey JR, Etherington E, Sheng X, Meilan R, Strauss SH, Brunner AM (2010) Populus CEN/TFL1 regulates first onset of flowering, axillary meristem identity and dormancy release in *Populus*. Plant J 62:674–688

Moriya S, Kunihisa M, Okada K, Iwanami H, Iwata H, Minamikawa M, Katayose Y, Matsumoto T, Mori S, Sasaki H, Matsumoto T, Nishitani C, Terakami S, Yamamoto T, Abe K (2017) Identification of QTLs for flesh mealiness in apple (*Malus × domestica* Borkh). Hort J 86:159–170

Olukolu BA, Trainin T, Fan S, Kole C, Bielenberg DG, Reighard GL, Abbott AG, Holland D (2009) Genetic linkage mapping for molecular dissection of chilling requirement and budbreak in apricot (*Prunus armeniaca* L.) Genome 52(10):819–828

Richardson EA, Seely SD, Walker DR (1974) A model for estimating the completion of rest for 'Redhaven' and 'Elberta' peach trees. HortScience 9:331–332

Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, Fromm J, Bhalerao RP, Boerjan W, Rohde A (2007) A molecular timetable for apical bud formation and dormancy induction in poplar. Plant Cell 19:2370–2390

Sánchez-Pérez R, Dicenta F, Martínez-Gómez P (2012) Inheritance of chilling and heat requirements for flowering in almond and QTL analysis. Tree Genet Genomes 8:379–389

Sasaki R, Yamane H, Ooka T, Jotatsu H, Kitamura Y, Akagi T, Tao R (2011) Functional and expressional analyses of PmDAM genes associated with endodormancy in Japanese apricot (*Prunus mume*). Plant Physiol 157:485–497

Sawamura Y, Suesada Y, Sugiura T, Yaegaki H (2017) Chilling requirements and blooming dates of leading peach cultivars and a promising early maturing peach selection, Momo Tsukuba 127. Hort J 86:426–436

Shannon S, Meeks-Wagner DR (1993) Genetic interactions that regulate inflorescence development in *Arabidopsis*. Plant Cell 5:639–655

Shirasawa K, Isuzugawa K, Ikenaga M, Saito Y, Yamamoto T, Hirakawa H, Isobe S (2017) The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding. DNA Res 24:499–508

Siviero A, Cristofani M, Machado MA (2003) QTL mapping associated with rooting stem cuttings from *Citrus sunki* vs. *Poncirus trifoliata* hybrids. Crop Breed Appl Biotechnol 3:83–88

Socquet-Juglard D, Christen D, Devenes G, Gessler C, Duffy B, Patocchi A (2013) Mapping architectural, phenological, and fruit quality QTLs in apricot. Plant Mol Biol Rep 31:387–397

Srinivasan C, Dardick C, Callahan A, Scorza R (2012) Plum (*Prunus domestica*) trees transformed with poplar FT1 result in altered architecture, dormancy requirement, and continuous flowering. PLoS One 7: e40715. https://doi.org/10.1371/journal.pone.0040715 . http://www.plosone.org/

Sugiura T, Kuroda H, Sugiura H (2007) Influence of the current state of global warming on fruit tree growth in Japan. Hort Res (Japan) 6:257–263 (in Japanese with English abstract)

Sugiura T, Sumida H, Yokoyama S, Ono H (2012) Overview of recent effects of global warming on agricultural production in Japan. JARQ 46:7–13

Sun L, Zhang Q, Xu Z, Yang W, Guo Y, Lu J, Pan H, Cheng T, Cai M (2013) Genome-wide DNA polymorphisms in two cultivars of mei (*Prunus mume* sieb. et zucc.) BMC Genet 14:98

Trainin T, Bar-Ya'akov I, Holland D (2013) ParSOC1, a MADS-box gene closely related to Arabidopsis AGL20/SOC1, is expressed in apricot leaves in a diurnal manner and is linked with chilling requirements for dormancy break. Tree Genet Genomes 9:753–766

Urrestarazu J, Muranty H, Denancé C, Leforestier D, Ravon E, Guyader A, Guisnel R, Feugey L, Aubourg S, Celton J-M, Daccord N, Dondini L, Gregori R, Lateur M, Houben P, Ordidge M, Paprstein F, Sedlak J, Nybom H, Garkava-Gustavsson L, Troggio M, Bianco L, Velasco R, Poncet C, Théron A, Moriya S, Bink MCAM, Laurens F, Tartarini S, Durel C-E (2017) Genome-wide association mapping of flowering and ripening periods in apple. Front Plant Sci 8:1923

Weber CA, Moore GA, Deng Z, Gmitter FG Jr (2003) Mapping freeze tolerance quantitative trait loci in a *Citrus grandis* × *Poncirus trifoliata* F1 pseudo-testcross using molecular markers. J Am Soc Hort Sci 128:508–514

Weinberger JH (1950) Chilling requirements of peach varieties. Proc Amer Soc Hort Sci 56:122–128

Xu Z, Zhang Q, Sun L, Du D, Cheng T, Pan H, Wang J (2014) Genome-wide identification, characterization and expression analysis of the MADS-box gene family in *Prunus mume*. Mol Genet Genom 289:903–920

Yamamoto T, Terakami S, Takada N, Nishio S, Onoue N, Nishitani C, Kunihisa M, Inoue E, Iwata H, Hayashi T, Itai A, Saito T (2014) Identification of QTLs controlling harvest time and fruit skin color in Japanese pear (*Pyrus pyrifolia* Nakai). Breed Sci 64:351–361

Yamane H (2014) Regulation of bud dormancy and bud break in Japanese apricot (*Prunus mume* Siebold & Zucc.) and peach [*Prunus persica* (L.) Batsch]: a summary of recent studies. J Jpn Soc Hort Sci 83:187–202

Yamane H, Kashiwa Y, Ooka T, Tao R, Yonemori K (2008) Suppression subtractive hybridization and differential screening reveals endodormancy-associated expression of an SVP/AGL24-type MADS-box gene in lateral vegetative buds of Japanese apricot. J Amer Soc Hort Sci 133:708–716

Yamane H, Ooka T, Jotatsu H, Hosaka Y, Sasaki R, Tao R (2011a) Expressional regulation of PpDAM5 and PpDAM6, peach (*Prunus persica*) dormancy-associated MADS-box genes, by low temperature and dormancy breaking reagent treatment. J Exp Bot 62:3481–3488

Yamane H, Ooka T, Jotatsu H, Sasaki R, Tao R (2011b) Expression analysis of PpDAM5 and PpDAM6 during flower bud development in peach (*Prunus persica*). Sci Hort 129:844–848

Yang S, Fresnedo-Ramírez J, Wang M, Cote L, Schweitzer P, Barba P, Takacs EM, Clark MD, Luby JJ, Manns DC, Sacks GL, Mansfield AK, Londo JP, Fennell AY, Gadoury D, Reisch BI, Cadle-Davidson LE, Sun Q (2016) A next-generation marker genotyping platform (AmpSeq) in heterozygous crops: a case study for marker assisted selection in grapevine. Hort Res 3:16002. https://doi.org/10.1038/hortres. 2016.2

Yu H, Ito T, Wellmer F, Meyerowitz EM (2004) Repression of AGAMOUS-LIKE 24 is a crucial step in promoting flower development. Nat Genet 36:157–161

Zhang Q, Chen W, Sun L, Zhao F, Huang B, Yang W, Tao Y, Wang J, Yuan Z, Fan G, Xing Z, Han C, Pan H, Zhong X, Shi W, Liang X, Du D, Sun F, Xu Z, Hao R, Lv T, Lv Y, Zheng Z, Sun M, Luo L, Cai M, Gao Y, Wang J, Yin Y, Xu X, Cheng T, Wang J (2012) The genome of *Prunus mume*. Nat Commun 3:1318

Zhang J, Zhang Q, Cheng T, Yang W, Pan H, Zhong J, Huang L, Liu E (2015) High-density genetic map construction and identification of a locus controlling weeping trait in an ornamental woody plant (*Prunus mume* Sieb. et Zucc). DNA Res 22:183–191

Zhao K, Zhou Y, Ahmad S, Xu Z, Li Y, Yang W, Cheng T, Wang J, Zhang Q (2018a) Comprehensive

cloning of *Prunus mume* dormancy-associated MADS-Box genes and their response in flower bud development and dormancy. Front Plant Sci 9:17

Zhao K, Zhou Y, Li Y, Zhuo X, Ahmad S, Han Y, Yong X, Zhang Q (2018b) Crosstalk of PmCBFs and PmDAMs Based on the changes of phytohormones under seasonal cold stress in the stem of *Prunus mume*. Int J Mol Sci 19:15

Zhebentyayeva TN, Fan S, Chandra A, Bielenberg DG, Reighard GL, Okie WR, Abbott AG (2014) Dissection of chilling requirement and bloom date QTLs in peach using a whole genome sequencing of sibling trees from an F2 mapping population. Tree Genet Genomes 10:35–51

Molecular and Developmental Biology: Self-incompatibility

12

Hisayo Yamane and Ryutaro Tao

Abstract

Although both self-compatible and self-incompatible cultivars exist in Japanese apricot (*Prunus mume* Sieb. et Zucc.), self-compatible ones have a horticultural advantage over self-incompatible ones in terms of fruit production. Therefore, self-compatibility is one of the important breeding objectives in Japan. Japanese apricot exhibits a homomorphic gametophytic self-incompatibility system in which self/nonself-recognition is controlled by a single multiallelic locus, the so-called *S* locus. During the last two decades, the ribonuclease gene, *S-RNase*, and the F-box gene, *SFB*, were identified as the pistil S and pollen S determinant genes, respectively, located within the *S* locus. Mutated versions of *SFB*, $S^{3\prime}$ and S^f, both of which contain a non-autonomous transposable element with the sequence resembling LTR of retrotransposons within the coding sequence, were reported to confer self-compatibility. Sequence similarity in the inserted sequence could be used to develop a universal PCR marker to detect $S^{3\prime}$ and S^f. Since $S^{3\prime}$ and S^f are the only self-compatible S haplotypes that have been found in Japanese apricot, the PCR marker would serve as a universal self-compatible marker for MAS in this species. Although the existence of a *Prunus*-specific gametophytic self-incompatibility recognition mechanism was supported by many reports, it has yet to be fully clarified. Recent genomic, molecular and evolutionary studies using whole-genome sequences, including *P. mume*, have provided new insights into the molecular network involved in the self-incompatibility recognition system in *Prunus*.

12.1 Introduction

Most *Prunus* fruit tree species, including Japanese apricot (*P. mume* Sieb. et Zucc.), exhibit a homomorphic gametophytic self-incompatibility (GSI) system in which self/nonself-recognition is controlled by a single multiallelic locus, the so-called *S* locus (Fig. 12.1). A self-incompatibility (SI) reaction is triggered when the same 'S allele' specificity is expressed in both the pollen and the pistil (Fig. 12.1). Thus, the growth of a pollen tube possessing either one of the two 'S allele' specificities carried by the recipient pistil is arrested in the style. The same type of GSI is found not only in other genera in the Rosaceae family, such as *Malus* and *Pyrus* in the tribe Maleae, but also in those of the families Solanaceae and the Plantaginaceae (McCubbin and Kao 2000; de Nettancourt 2001; Kao and Tsukamoto 2004; Takayama and Isogai 2005; Yamane and Tao 2009).

H. Yamane · R. Tao (✉)
Graduate School of Agriculture, Kyoto University, Kyoto, Japan
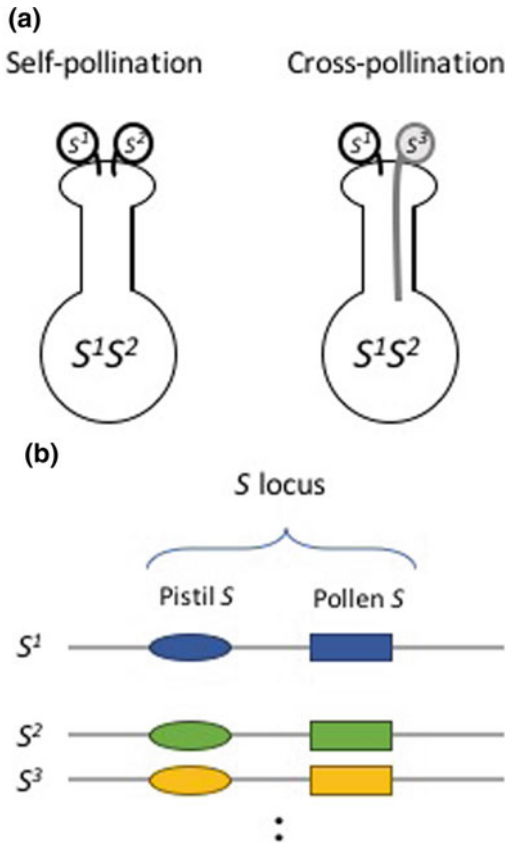e-mail: rtao@kais.kyoto-u.ac.jp

Fig. 12.1 Genetic control of gametophytic self-incompatibility (GSI). **a** In GSI, the pollen is rejected when its S haplotype matches either of the pistil S haplotypes. The cross is incompatible when the S haplotypes of the male and female parents match completely ($S^1S^2 \times S^1S^2$), while the cross is compatible when a portion of the S haplotypes is shared by the male and female parents ($S^1S^2 \times S^1S^3$). **b** The GSI is generally controlled by a single S locus which contains at least two genes, one for pollen specificity and another for pistil specificity (citation from Matsumoto and Tao 2016)

As Japanese apricot is unable to bear fruit parthenocarpically, fertilisation and seed formation are essential for good fruit production. Both self-compatible (SC) and SI cultivars exist in Japanese apricot. For the fruit production of SI cultivars in commercial orchards, cross-compatible and male fertile cultivars that flower simultaneously are inter-planted, and beehives are often placed in orchards to ensure fruit set (Westwood 1993). The SC cultivars of Japanese apricot have a horticultural advantage over SI

cultivars because Japanese apricot blooms very early in the spring, when pollinating insects are not very active. In cooler areas of Japan, spring temperatures and wind conditions are not suitable for bee activity. Therefore, the production of SC cultivars with good pomological characteristics is one of the major breeding objectives for Japanese apricot as well as other SI *Prunus* fruit tree species (Janick and Moore 1975; Yaegaki et al. 2002a). Yaegaki et al. (2002b) scored the self-unfruitfulness (SU) and self-fruitfulness (SF) of Japanese apricot cultivars and lines (Yaegaki et al. 2002a). The SU could be attributed to SI and/or male sterility. Interestingly, all ornamental cultivars are SU. In contrast to cultivars for fruit production, SU is preferable to SF in ornamental cultivars, as excessive fruit load may weaken the tree. Furthermore, the doubled flower characteristic, which is often preferable in ornamental Japanese apricot, could lead to sterility or low fruit set.

During the last two decades, genes for the two proteins in S locus controlling the allele specificity of GSI recognition in *Prunus* have been identified. It is now known that two separate genes at the S locus control male (pollen) and female (pistil) specificities (Fig. 12.2). Ribonuclease (RNase) and F-box genes were identified as the pistil S and pollen S determinant genes, respectively (see review by Kao and Tsukamoto 2004). Upon this finding, the term 'S haplotype' is used to describe the variants of the S locus, while the term 'allele' is used to describe the variants of the S locus genes, pistil S and pollen S. On the practical side, these findings led to the development of novel molecular techniques for S genotyping and SC screening (Tao et al. 1999; Yamane and Tao 2009; Yamane et al. 2003a). Molecular S genotyping and marker-assisted selection of SC offspring are now being successfully incorporated in *Prunus* breeding programmes worldwide.

In this chapter, we first summarise the characteristics of pistil S and pollen S in the GSI of Japanese apricot, along with the knowledge obtained with SC mutants. A molecular marker linked to the SC haplotype has been developed and applied in SC breeding in Japan. We also
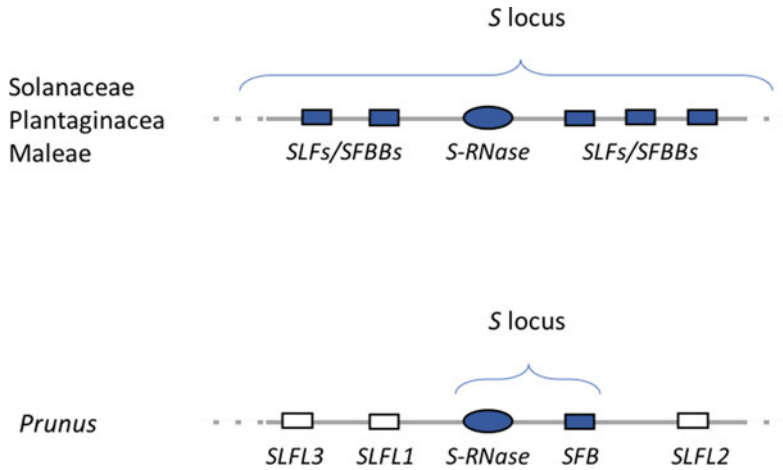
Fig. 12.2 Schematic diagram of *S* locus regions that genetically control GSI in *Prunus*. Although the specificity of the GSI reaction can be explained by assuming a single locus (*S* locus) with multiple co-dominant *S* alleles, recent molecular studies revealed that the pistil and pollen *S* specificities are determined by two different genes, *S-RNase* (pistil *S*) and *SFB* (pollen *S*), respectively. Other pollen-expressed F-box genes (e.g. *SLFL1–3*) flank the *S* locus. The *Prunus SLFL1–3* genes are not involved in determining pollen specificity (citation from Matsumoto and Tao 2016)

highlight the recently proposed distinct molecular recognition mechanism in the GSI of *Prunus*, which is different in terms of the SI/SC recognition mechanism from other plant species sharing the same pistil *S* and pollen *S* determinants, Solanaceae, Plantaginaceae and Maleae (Rosaceae). Some of those distinct *Prunus* GSI features were recently identified using transcriptome analysis and evolutionary analysis using whole-genome sequences of *Prunus*, including *P. mume*.

12.2 Identification of Self-incompatibility Determinants in *S* Locus Mutants

12.2.1 Identification and Characterisation of Pistil Determinant, *S-RNase* in Japanese Apricot

The physiology and mechanisms of GSI have been most extensively studied in solanaceous plant species. The cDNAs encoding glycoproteins co-segregated with *S* alleles were first cloned from *Nicotiana alata* (Anderson et al. 1986, 1989). The deduced amino acid sequence strongly implicated stylar RNase involvement in the recognition and rejection reaction in the style. Based on these and other studies, in the family Solanaceae, the *S* allele product in the pistil is a highly basic glycoprotein containing sequence motifs characteristics of the active site of the fungal RNase T2 (Kawata et al. 1988) and Rh (Horiuchi et al. 1988), called S-RNase (McClure et al. 1989). Previously, it has been reported that S-RNases are also associated with GSI of Maleae, *Pyrus* and *Malus*, in Rosaceae (Sassa et al. 1992, 1993, 1996; Broothaerts et al. 1995). The finding that these two families recruited the same molecule as the GSI pistil determinant was unexpected because Solanaceae (Asteridae) and Rosaceae (Rosidae) are phylogenetically distant (Chase et al. 1993; Igic and Kohn 2001). Later, Xue et al. (1996) cloned the S-RNase in *Antirrhinum* in Plantaginaceae, a family closely related to Solanaceae (Xue et al. 1996).

As *Prunus* belongs to the family Rosaceae, it was readily expected that *Prunus* also has an

S-RNase-based GSI system. However, S-RNases remained unidentified for several years after the cloning of the *Pyrus* and *Malus* S-RNases, mainly because polymerase chain reaction (PCR) cloning approaches for *Prunus* S-RNase were hindered by its relatively low DNA sequence similarity with Maleae *S-RNase*s and the presence of *Prunus* RNase genes that are not involved in GSI. The first clue to the cloning of *Prunus S-RNase* was obtained when N-terminal sequences of almond (*Prunus dulcis*) S-RNase were reported (Tao et al. 1997). On the basis of the N-terminal amino acid sequences, sweet cherry (*Prunus avium*) (Tao et al. 1999) and almond (Ushijima et al. 1998) *S-RNase*s were cloned.

Yaegaki et al. (2001) first determined *S*-genotypes of the main Japanese apricot cultivars by *S-RNase*-specific PCR band polymorphisms and assigned seven *S* alleles (*S1–S7*) (Yaegaki et al. 2001). Tao et al. (2002b) cloned the full-length cDNAs for *S-RNase* alleles of 'Nanko' and showed sequence similarity to other *Prunus S-RNase*s (Tao et al. 2002b). Tao et al. and Habu et al. reported 12 *S-RNase* alleles [*S1–S11* and *Sf* (a self-fertile allele)] (Table 12.1) (Tao et al. 2000, 2002a, b, 2003; Habu et al. 2008).

Generally, S-RNases are highly divergent, with allelic amino acid sequence identities ranging from approximately 30 to over 90% (Ushijima et al. 1998; McCubbin and Kao 2000). In spite of the high allelic sequence diversity, analysis of solanaceous S-RNase alleles revealed five conserved regions, C1–C5 (Ioerger et al. 1991). There is a single (RHV) hypervariable region in the rosaceous S-RNase, while two hypervariable regions (HVa and HVb) were found in solanaceous and plantaginaceous S-RNases (Ioerger et al. 1991; Xue et al. 1996). The different feature of the S-RNase structure between Maleae and *Prunus* is the number of introns, one and two in *S-RNase* of Maleae and *Prunus*, respectively.

12.2.2 Identification and Characterisation of Pollen Determinant, *SFB*, in Japanese Apricot

The pollen *S* determinant of the S-RNase-based GSI in Rosaceae, Solanaceae and Plantaginaceae was discovered many years after the stylar determinant, *S-RNase*. The subcentromeric location of the solanaceous and plantaginaceous *S* locus had long prevented chromosome walking (Entani et al. 2003). The first clue to the identification of the pollen *S* was obtained from the *S* locus of Plantaginaceae. Sequencing analysis of the *Antirrhinum hispanicum S* locus revealed the presence of a pollen-expressed F-box gene (*AhSLF* for *A. hispanicum S*-locus F-box), located 9 kb downstream of S^2-RNase (Lai et al. 2002). In Solanaceae, the recombination-suppressed region around the *S* locus was estimated to be 4.4 Mb in *Petunia* species (Wang et al. 2003). Map-based cloning and chromosome walking approaches identified multiple pollen-expressed F-box genes flanking the *S-RNase* gene, called *S-locus F-box* (*SLF*). Eventually, Kubo et al. (2015) concluded that the SLFs proteins collaboratively target all but self-S-RNases for degradation, a process known as the 'collaborative nonself-recognition model' (Kubo et al. 2010, 2015). Similarly, in the tribe Maleae, multiple pollen-expressed F-box genes flanking the S-RNase gene, called *S-locus F-box brothers* (*SFBB*), were identified as pollen determinants. Similar to Solanaceae, a collaborative nonself-recognition model could also explain the GSI system in Maleae.

The *S* locus of *Prunus*, which is located at the end of linkage group 6 (Dirlewanger et al. 2004), appeared to be considerably smaller than those of Solanaceae and Plantaginaceae. Two different research groups in Japan have successfully sequenced the *S* locus region of *Prunus* via genome walking, starting from S-RNase

Table 12.1 Main Japanese apricot cultivars cultivated in Japan (except for male-sterile cultivars and ornamental cultivars) and their S haplotypes (from Yamane and Tao 2009)

	S-RNase alleles[a]													S haplotypes
SI cultivars	S^1	S^2	S^3	$S^{3'}$	S^4	S^5	S^6	S^7	S^8	S^9	S^{10}	S^{11}	S^f	
'Baigo'							X				X			S^6S^{10}
'Nanko'	X							X						S^1S^7
'Gessekai'	X						X							S^1S^6
'Gyokuei'		X					X							S^2S^6
'Kagajizo'							X				X			S^6S^{10}
'Kairyouchidaume'			X		X									S^3S^4
'Ohshuku'	X					X								S^1S^5
SC cultivars														
'Kensaki'													XX	S^fS^f
'Koshinoume'				X									X	$S^{3'}S^f$
'Benisashi'								X					X	S^7S^f
'Hachiro'									X				X	S^8S^f
'Ryukyokoume'									X				X	S^8S^f
'Rinshu'										X			X	S^9S^f
'Jizoume'											X		X	$S^{10}S^f$
'Shinheidayu'											X		X	$S^{10}S^f$
'Orihime'												X	X	$S^{11}S^f$

[a] $S^{3'}$ is the mutant version of S^3

(Ushijima et al. 2001; Entani et al. 2003). Ushijima et al. (2003) conducted DNA sequencing and transcriptional analyses for the genomic regions that flank the almond (*P. dulcis*) S-RNase and identified polymorphic and non-polymorphic *S* locus F-box genes, called *SFB*, for the *S* haplotype-specific F-box protein gene and *SLF* for the *S-locus F-box*, respectively. Later, *SLF* has been replaced by *SLFL1* (Matsumoto et al. 2008). The features of *SFB*, such as the high level of allelic polymorphism, pollen-specific expression and the close physical distance to the *S-RNase*, all supported that *SFB* is the male determinant of GSI in almond. The same research group also found *SFB* at the cherry *S* locus during their attempt to compare the same S-RNase allele in SC and SI species of cherries, *P. cerasus* and *P. avium*, respectively (Yamane et al. 2003a). At around the same time, another research group in Japan analysed the *S* locus region of two different *S* haplotypes of Japanese apricot and found four F-box genes (Entani et al. 2003). Among them, SLF for S-locus F-box, which has a different name but is orthologous to *SFB* in almond and cherries, shows a high level of allelic sequence diversity and was supposed to be a candidate of pollen *S*. The other F-box genes found near *S-RNase*, *SLFL1*, *SLFL2* and *SLFL3* for *SLF-like genes 1*, *2* and *3*, respectively, showed a considerably lower allelic sequence diversity. The pollen S candidate of Japanese apricot was independently cloned by Yamane et al. (2003a, b) and named *PmSFB* (Fig. 12.2). As noted above, the *Prunus* pollen S was initially referred to by two different terms, '*SFB*' and '*SLF*'. We use '*SFB*' in this chapter to reflect the different features of the *Prunus SFB* compared to the *SLF* of Solanaceae and Plantaginaceae.

Yamane et al. (2003a, b) cloned SFB^1 and SFB^7 from the SI cultivar 'Nanko', using rapid amplification of the cDNA ends (RACE) technique with *SFB* gene-specific primers (Yamane et al. 2003b). Entani et al. (2003) demonstrated that SFB^1 of 'Nanko' was tightly linked to S^1-RNase. Genomic DNA blot analysis using *SFB*

as a probe yielded restriction fragment length polymorphism (RFLP) bands specific to each *S* allele of Japanese apricot (Yamane et al. 2003a, b).

The *SFB* contains only a single intron in the untranslated region, whereas no intron was found in solanaceous and plantaginaceous *SLF*s. A primary structural analysis of SFB revealed the presence of two variable (V1 and V2) and two hypervariable (HVa and HVb) regions (Ikeda et al. 2004). These regions appeared to be hydrophilic or at least not strongly hydrophobic, which suggests that they may be exposed on the surface and function in the allele specificity of the recognition response. The fact that positively selected sites appear to concentrate in the variable and hypervariable regions further supports the possibility that these regions could play an important role in SC/SI recognition.

12.3 SC Mutations and Development of Molecular Markers for SC

Among the genes in the *S* locus, mutations in the *S-RNase* that lead to dysfunction of the S-RNase are known to confer SC in Solanaceae and Rosaceae. In *Prunus*, SC is conferred by a low level of *S-RNase* transcription that leads to low levels of S-RNase accumulation in the style (Yamane et al. 2003b). It is possible to consider that the absence or low level of S-RNase in the style, leading to SC, is a common feature of S-RNase-based GSI. Mutations in the pollen *S*, however, lead to different outcomes depending on the family. So far, all known solanaceous and plantaginaceous PPMs are associated with competitive interaction, e.g. the presence of two different pollen S alleles in a pollen grain (Golz et al. 1999, 2001; Tsukamoto et al. 2006). Mutations that disrupt pollen *S* (*SLF*) function in Solanaceae and Plantaginaceae have not yet been found and therefore are thought to confer either SI or lethality, while mutations that disrupt

pollen S (*SFB*) function found in *Prunus* resulted in SC (Tsukamoto et al. 2006; Koichiro Ushijima et al. 2004; Yamane and Tao 2009).

Using *S-RNase* gene-specific PCR and genomic DNA blot analyses, Tao et al. (2000) found that empirically known SC cultivars of Japanese apricot had a common *S-RNase*, designated S^f *RNase*. Co-segregation of S^f *RNase* and SC was confirmed by segregation analysis (Tao et al. 2002a); however, there were apparently no substantial differences between the amino acid sequences derived from S^f *RNase* and from other SI *S-RNases*, suggesting that pollen S could be responsible for the SC in the S^f haplotype(Tao et al. 2002a). As predicted, DNA sequence analysis showed that the mutation in *SFB^f* is similar to that found in sweet cherry $SFB^{4'}$, conferring pollen-part mutated SC (Ushijima et al. 2004). The presence of a 6.8 kbp insertion in the middle of the SFB^f coding region makes transcripts for a defective SFB, lacking the C-terminal half that contains hypervariable regions that are putatively responsible for S specificity (Ushijima et al. 2004). The finding that the S^f haplotype confers SC in Japanese apricot led to the development of a molecular marker for SC in this species. A primer set designed from the unique sequence in the second intron of S^f-*RNase* was successfully amplified as S^f-*RNase*-specific PCR bands (Tao et al. 2003). Alternatively, primers derived from a 6.8 kbp inserted fragment sequence were used to identify the S^f haplotype by loop-mediated isothermal amplification (LAMP) method (Habu et al. 2006). The developed S^f-typing system has been used to detect S^f-carrying SC individuals at the juvenile stage during the SC breeding programme in the Wakayama prefecture, the main Japanese apricot production area in Japan. Finally, SC 'NK-14' was successfully developed as the first released 'MAS-applied' Japanese apricot cultivar in the world (Iwamoto, personal communication).

Very recently, another SC S haplotype, $S^{3'}$, was discovered in Japanese apricot as a pollen-part mutant S haplotype. In order to elucidate the molecular basis of the pollen-part mutation in the $S3'$ haplotype, Yamane et al.

(2009) have cloned and analysed the S locus region of the $S^{3'}$ haplotype. As expected, there is a structural alternation in *SFB* in the $S^{3'}$ haplotype, while there is no such mutation in *S-RNase* in the $S^{3'}$ haplotype. The mutated SFB^3, termed $SFB^{3'}$, contains a non-autonomous transposable element with the sequence resembling LTR of retrotransposons. Interestingly, most LTR regions and their adjacent regions in the inserted sequence to $SFB^{3'}$ showed an extremely high sequence identity to the inserted sequence to SFB^f of Japanese apricot (Ushijima et al. 2004). Not only the mode of mutation, but also the sequences inserted into *SFB* are similar in $S3^{3'}$ and S^f haplotypes. Because both segments of the LTR in the sequence inserted into $SFB^{3'}$ and into SFB^f are identical, it is supposed that the insertion event occurred quite recently. It is possible that the non-autonomous transposable elements inserted into $SFB^{3'}$ and SFB^f could be active retrotransposons in the same family. Sequence similarity in the inserted sequence could be used to develop a universal PCR marker to detect $S^{3'}$ and S^f. Since $S^{3'}$ and S^f are the only SC S haplotypes that have been found in Japanese apricot, the PCR marker would serve as a universal SC marker for MAS in Japanese apricot (Fig. 12.3). More than half of the diverse Japanese apricot cultivars and lines surveyed contained the S^f haplotype, while only 4 of 86 cultivars and lines had the $S^{3'}$ haplotype (Table 12.2). It is supposed that the mutated SC S haplotype could rapidly prevail over SI S haplotypes because SC has pomological advantage over SI. It is therefore possible that the limited number of $S^{3'}$ haplotypes could indicate that $S^{3'}$ was generated recently. An alternative explanation for the limited number of $S^{3'}$ could also be drawn from the limited number of S^3 in the cultivars and lines tested. The original SI S^3 was only present in 'Kairyou Uchidaume' and its progeny 1C1-10 (Habu et al. 2008). For this reason, it is possible that S^3 and its mutated version, $S^{3'}$, may be linked to undesirable characteristics that affect tree performance or fruit quality, and S^3 and $S^{3'}$ haplotypes may have been selected out under artificial or natural selection pressure.
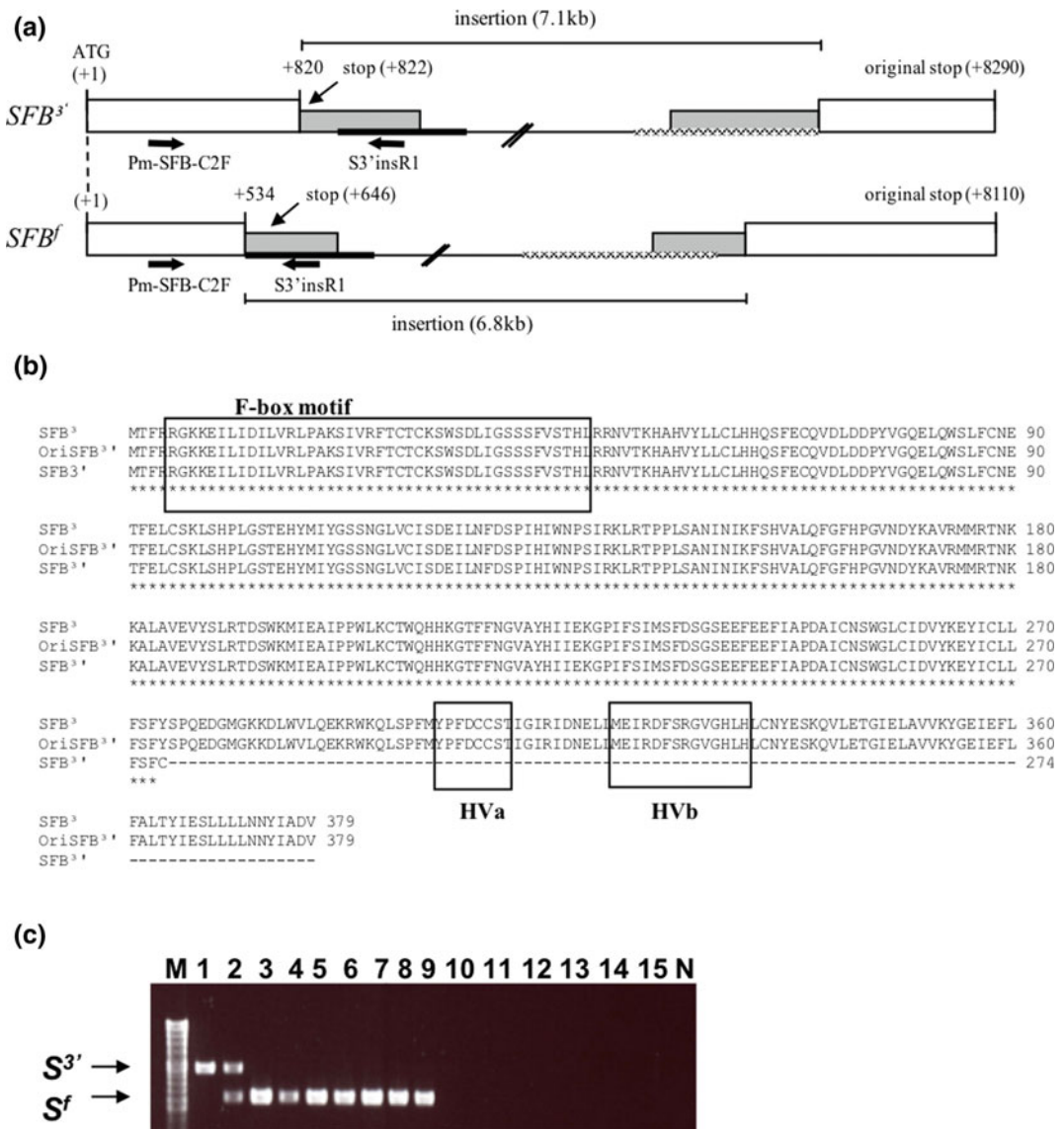
Fig. 12.3 Universal molecular marker for the selection of self-compatible (SC) cultivars in Japanese apricot. **a** Schematic representation of $SFB^{3'}$ (Yamane et al. 2009) as compared with SFB^f (Ushijima et al. 2004). The nucleotide sequences of the open box are similar to the coding region of a functional SFB. In the case of $SFB^{3'}$, nucleotide sequences of the open box are exactly the same as the coding region of the functional SFB^3. 'A' of the start codon, corresponding to that of the functional SFB, is positioned as +1. The sequences of both ends (grey box) of the inserted DNA fragments are identical in $SFB^{3'}$ as well as in SFB^f. Black and wavy thick lines in the inserted sequences indicate sequences that show high identities between $SFB^{3'}$ and SFB^f. The positions of the primers for the PCR amplification of $SFB^{3'}$ and SFB^f are indicated by arrows. **b** Amino acid sequence comparison between SFB^3, predicted intact $SFB^{3'}$ (OriSFB$^{3'}$) and SFB^3. The F-box motif and HVa and HVb regions are boxed. Universal primers distinguishing SC and SI in Japanese apricot. (C) The primers specifically amplify $S^{3'}$ and S^f haplotype. Lanes: M, 100-bp ladder; 1, 1K0-26 ($S^{3'}S^7$); 2, 'Koshinoume' ($S^{3'}S^f$); 3, 'Kensaki' (S^fS^f); 4, 'Benisashi' (S^7S^f); 5, 'Hachirou' (S^8S^f); 6, 'Ryukyou Koume' (S^8S^f); 7, 'Rinshuu' (S^9S^f); 8, 'Jizouume' ($S^{10}S^f$); 9, 'Orihime' (S^6S^f); 10, 'Nankou' (S^1S^7); 11, 'Oushuku' (S^1S^5); 12, 'Gessekai' (S^1S^6); 13, 'Gyokuei' (S^2S^6); 14, 'Kairyou Uchidaume' (S^3S^4); 15, 1C1-10 (S^3S^7); and N, negative control (PCR with no template DNA) (from Yamane et al. 2009)

Table 12.2 Characteristics of self-compatible Japanese apricot cultivars cultivated in Japan (from Yamane et al. 2009)

Cultivars and lines	Presence of $S^{3'}/S^f$	SU/SI/SC/SF[a]	Male sterility	Types	Reference
1C1-10	No	SI	No	Fruit production	Tao et al. (2002a)
1K0-26	$S^{3'}$	SF/SC	No	Fruit production	Tao et al. (2002a)
Akananiwa	No	SU	No	Ornamental	Yaegaki et al. (2002b)
Akebono	No	SU	MS	Fruit production	Yaegaki et al. (2002a, b)
Baigou	No	SU	No	Fruit production	Yaegaki et al. (2002a, b, 2003)
Benisashi	S^f	SF	No	Fruit production	Yaegaki et al. (2002b)
Bungo (Hiratsuka)	No	SU	MS	Fruit production	Yaegaki et al. (2002a, b)
Bungo (Kurume)	S^f	SF	No	Fruit production	Yaegaki et al. (2002b)
Bungo × Kousyuu Oujyuku	No	SU	MS	Fruit production	Yaegaki et al. (2002a, b)
Chouhanagata	No	SU	MS	Ornamental	Yaegaki et al. (2002a, b)
Fujibotan	S^f	SU	No	Ornamental	Yaegaki et al. (2002b)
Fujiedatankoubai	S^f	SU	MS	Fruit production	Yaegaki et al. (2002a, b)
Fujinoume	S^f	SF	No	Fruit production	Yaegaki et al. (2002b)
Futono	No	SU	No	Fruit production	Yaegaki et al. (2002b)
Gecchibai	S^f	SU	No	Fruit production	Yaegaki et al. (2002b, 2003)
Gessekai	No	SU	No	Fruit production	Yaegaki et al. (2002b, 2003)
Gojirou	No	SU	MS	Fruit production	Yaegaki et al. (2002a, b)
Gyokuei	No	SU	MS	Fruit production	Yaegaki et al. (2002a, b, 2003)
Hachirou	S^f	SF	No	Fruit production	Yaegaki et al. (2002b, 2003)
Hanakami	S^f	SU	No	Fruit production	Yaegaki et al. (2002b)
Ihara	No	SU	No	Fruit production	Yaegaki et al. (2002b)
Inazumi	S^f	SF	No	Fruit production	Yaegaki et al. (2002b, 2003)
Inkyo	No	SU	No	Fruit production	Yaegaki et al. (2002b)

(continued)

Table 12.2 (continued)

Cultivars and lines	Presence of $S^{3'}/S^f$	SU/SI/SC/SF[a]	Male sterility	Types	Reference
Ishikawa Oomiume	S^f	SU	No	Fruit production	Yaegaki et al. (2002b)
Issunnbai	S^f	SU	No	Fruit production	Yaegaki et al. (2002b)
Jizouume	S^f	SF	No	Fruit production	Yaegaki et al. (2002b, 2003)
Jyousyuushiro	No	SU	No	Fruit production	Yaegaki et al. (2002b)
Jyuurou	No	SU	No	Fruit production	Yaegaki et al. (2002b, 2003)
Kagajizou	No	SU	MS	Fruit production	Yaegaki et al. (2002a, b, 2003)
Kairyou Uchidaume	No	SU	No	Fruit production	Yaegaki et al. (2002b, 2003)
Kankoubai	S^f	SU	No	Ornamental	Yaegaki et al. (2002b)
Kasugano	No	SU	MS	Ornamental	Yaegaki et al. (2002a, b)
Kenkyou	No	SU	No	Ornamental	Yaegaki et al. (2002b)
Kensaki	S^f	SF	No	Fruit production	Yaegaki et al. (2002b)
Kichirobei	No	SU	No	Fruit production	Yaegaki et al. (2002b)
Kinsujiume	No	SU	MS	Ornamental	Yaegaki et al. (2002a, b)
Komukai	No	SU	MS	Fruit production	Yaegaki et al. (2002a, b)
Koshinoume	$S^{3'}$, S^f	SF/SC	No	Fruit production	Tao et al. (2002a), this study
Koubai	S^f	SU	No	Ornamental	Yaegaki et al. (2002b)
Koume	S^f	SU	No	Fruit production	Yaegaki et al. (2002b)
Koushuu Oujyuku	S^f	SU	No	Fruit production	Yaegaki et al. (2002b)
Koushuu Saishou	S^f	SF	No	Fruit production	Yaegaki et al. (2002b)
Koushuu Shinnko	S^f	SF	No	Fruit production	Yaegaki et al. (2002b)
Makitateyama	S^f	SU	MS	Ornamental	Yaegaki et al. (2002a, b)
Mangetushidare	No	SU	No	Ornamental	Yaegaki et al. (2002b)
Michishirube	S^f	SU	MS	Ornamental	Yaegaki et al. (2002a, b)
Muroya	S^f	SF	No	Fruit production	Yaegaki et al. (2002b)
Naniwa	No	SU	No	Fruit production	Yaegaki et al. (2002b)

(continued)

Table 12.2 (continued)

Cultivars and lines	Presence of $S^{3'}/S^f$	SU/SI/SC/SF[a]	Male sterility	Types	Reference
Nankou	No	SU	No	Fruit production	Yaegaki et al. (2002b, 2003)
Natsuka	S^f	SF	No	Fruit production	Yaegaki et al. (2002b)
Okituakabana	No	SU	MS	Ornamental	Yaegaki et al. (2002a, b)
Orihime	S^f	SF	No	Fruit production	Yaegaki et al. (2002b, 2003)
Oushuku	No	SU	No	Fruit production	Yaegaki et al. (2002b)
Oushukubai	No	SU	No	Ornamental	Yaegaki et al. (2002b)
Rinshuu	S^f	SF	No	Fruit production	Yaegaki et al. (2002b)
Ryuukyou Koume	S^f	SF	No	Fruit production	Yaegaki et al. (2002b, 2003)
Saju	No	SU	MS	Ornamental	Yaegaki et al. (2002a, b)
Sarasa	No	SU	No	Ornamental	Yaegaki et al. (2002b)
Seiyoubai	No	SU	MS	Fruit production	Yaegaki et al. (2002a, b)
Shimosukeume	S^f	SF	No	Fruit production	Yaegaki et al. (2002b)
Shiratamaume	S^f	SF	No	Fruit production	Yaegaki et al. (2002b)
Shirobotan	No	SU	MS	Ornamental	Yaegaki et al. (2002a, b)
Shirokaga	No	SU	MS	Fruit production	Yaegaki et al. (2002a, b, 2003)
Shukou	S^f	SF	No	Fruit production	Yaegaki et al. (2002b)
Sugita	No	SU	No	Fruit production	Yaegaki et al. (2002b)
Sumomoume	No	SU	MS	Fruit production	Yaegaki et al. (2002a, b)
Suzukishiro	S^f	SU	MS	Fruit production	Yaegaki et al. (2002a, b)
Taihei	S^f	SU	MS	Fruit production	Yaegaki et al. (2002a, b)
Tairinryokugaku	No	SU	No	Ornamental	Yaegaki et al. (2002b)
Taisyouume	S^f	SF	No	Fruit production	Yaegaki et al. (2002b)
Takadaume	S^f	SU	No	Fruit production	Yaegaki et al. (2002b)
Tamabotan	S^f	SU	No	Ornamental	Yaegaki et al. (2002b)
Tamagakishidare	No	SU	MS	Ornamental	Yaegaki et al. (2002a, b)

Table 12.2 (continued)

Cultivars and lines	Presence of $S^{3'}/S^f$	SU/SI/SC/SF[a]	Male sterility	Types	Reference
Tamaume	No	SU	No	Fruit production	Yaegaki et al. (2002b)
Tamaume × Koushuu Saisyou	No	SU	No	Fruit production	Yaegaki et al. (2002b)
Tobiume	No	SU	No	Ornamental	Yaegaki et al. (2002b)
Tougoro	$S^{3'}$, S^f	SF	No	Fruit production	Yaegaki et al. (2002b)
Touji	No	SU	No	Ornamental	Yaegaki et al. (2002b)
Tsukasashibori	No	SU	No	Ornamental	Yaegaki et al. (2002b)
Umetukuba No. 4	S^f	SF	No	Fruit production	Yaegaki et al. (2003)
Yaezakikankou	S^f	SU	MS	Ornamental	Yaegaki et al. (2002a, b)
Yakushiume	No	SU	No	Fruit production	Yaegaki et al. (2002b)
Yatsubusa (Akita strain)	S^f	SU	No	Fruit production	Yaegaki et al. (2002b)
Yatsubusa (Shimada strain)	$S^{3'}$, S^f	SF	No	Fruit production	Yaegaki et al. (2002b)
Yourou	No	SU	No	Fruit production	Yaegaki et al. (2002b)
Youseiume	S^f	SF	No	Fruit production	Yaegaki et al. (2002b)

[a]*SU* self-unfruitful (less than 10% of fruit set after self-pollination), *SI* self-incompatible, *SC* self-compatible, and *SF* self-fruitful (10% or more than 10% of fruit set after self-pollination)

12.4 Towards the Clarification of Molecular Recognition Mechanisms in the GSI of *Prunus*

12.4.1 Possible Distinct GSI Recognition Mechanism in *Prunus*

To date, many of the SC *Prunus* PPM S haplotypes are shown to encode a dysfunctional SFB (Tao and Iezzoni 2010). These findings indicate that the *Prunus* pollen S determinant acts to prohibit unknown mechanisms that work to inactivate the cytotoxic effects of the S-RNase. In

that case, it would be in contrast to the family Solanaceae, in which deletion or dysfunction of the pollen determinant are believed to lead either to SI or to lethality (Kubo et al. 2015) under the proposed 'collaborative non-S recognition system' model to explain Solanaceae and Maleae GSI systems.

Matsumoto and Tao (2016) hypothesised that the *Prunus* SFB helps to release the cytotoxic effects of self-S-RNases and to induce incompatible reactions. To explain the detoxification of *Prunus* S-RNases, many researchers predicted the existence of a hypothetical general inhibitor (GI) that detoxifies S-RNases in compatible pollen tubes (Tao and Iezzoni 2010; Luu et al. 2001; Matsumoto and Tao 2016). The SFB

protein is considered to be part of the SCF complex because the F-box motif is conserved at the N-terminal and was observed to be under purifying selection (Ikeda et al. 2004). This has been supported by the identification of the functional Skp1-like protein, which interacts with SFB (i.e. SSK1) (Matsumoto et al. 2012). In *Prunus*, as in other plants, multiple pollen-expressed F-box genes, *SFB*, and three *S-locus F-box like* genes (*SLFL1–3*), are located at the S locus and its flanking regions (Entani et al. 2003; Ushijima et al. 2003, 2004). The *SLFL1*, *SLFL2* and *SLFL3* exhibit a much lower level of allelic sequence polymorphism (i.e. approximately 92.5%) than *SFB* (i66–82.5%), and there are considerable differences in DNA sequences across the genes (Entani et al. 2003; Ushijima et al. 2003, 2004; Matsumoto et al. 2008). Based on the knowledge obtained from other taxa with S-RNase-based GSI, it is possible that E3 ligases that contain a protein orthologous to SLF or

SFBB may be involved in degrading S-RNases. Therefore, *Prunus SLFL* are the best GI candidates because phylogenetic studies have highlighted their close relationship with *SFBB* (Aguiar et al. 2015; De Franceschi et al. 2012; Matsumoto et al. 2008; Morimoto et al. 2015; Sassa et al. 2007, 2010). The SLFLs may help degrade S-RNases in a manner similar to the hypothesised activity of SFBBs. As expected, SLFLs can interact with SSK1, indicating they function as E3 ligases (Matsumoto et al. 2012). One possible explanation for how S-RNase cytotoxicity is controlled by SFBs and GIs to induce SI involves the degradation of the GI by SFB in a self-recognition-specific manner (i.e. GI degradation model; Fig. 12.4). All S-RNases are assumed to be recognised and inhibited by a GI unless they are affected by the cognate SFB. The SCF^{SFB} recognises the complex consisting of self-S-RNase and the GI and polyubiquitinates the GI for degradation to release cytotoxic
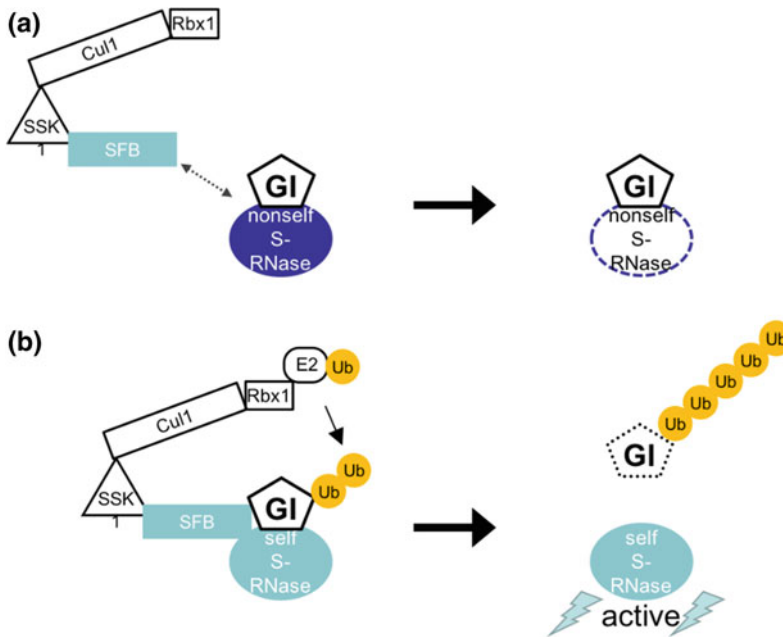

Fig. 12.4 The molecular basis of *Prunus* self-/nonself-discrimination involves the general inhibitor (GI) degradation model, which is based on *Prunus* pollen S biochemical functions. **a** Nonself-S-RNases are thought to be recognised and inhibited by the unidentified GI (proposed to be SLFLs). The inhibition of nonself-S-RNases is believed to be unaffected by SFB. **b** The SFB is assumed to recognise the self-S-RNase-GI complex and polyubiquitinated GI. Degradation of the polyubiquitinated GI by the ubiquitin-proteasome system leads to the release of active self-S-RNases, which enables the incompatibility reaction (from Matsumoto and Tao 2016)

self-S-RNases. In a compatible reaction, the SCF^{SFB} would not recognise complexes containing nonself-S-RNase, which would then be inhibited by a GI. However, this hypothesis needs to be carefully verified, which means the GI must be identified.

12.4.2 Use of *Prunus mume* Genome for Providing Insights into the GSI Recognition Mechanism in *Prunus*

Whole-genome sequences of *Prunus* genome are a powerful tool to uncover distinct molecular networks involved in *Prunus*-specific GSI recognition mechanisms. Comprehensive transcriptomic and evolutionary studies have been conducted using *Prunus* genome sequences, leading to important insights into the molecular network of GSI recognition mechanisms in *Prunus*.

Habu and Tao (2014) conducted a large-scale transcriptome analysis of pollinated and unpollinated pistils and pollen grains from Japanese apricot and found *SLFL1*, *SLFL2* and *SLFL3* in the pistil transcriptome. Among these, *SLFL3* showed transcriptional pattern changes appropriate for GI. Namely the expression of *SLFL3* was pollen-specific and highly up-regulated by pollination. It is interesting that *ASK2*, almost identical to *PavSSK1*, which is a gene for the SKP1-like1 protein interacting with SLFLs and SFB to form the SCF complex with Cullin1-likes (CUL1s) (Matsumoto et al. 2012), was also up-regulated upon pollination, especially after compatible pollination. The authors also found ubiquitin carboxyl-terminal hydrolase 16-like protein, which is a member of the deubiquitinating enzymes (DUBs), among the pollen-specific genes up-regulated by pollination. Furthermore, a ubiquitin-specific protease, UBP11, was also found as a self-pollinated pistil-specific up-regulated gene expressed in pollen. Although we have yet to clarify the SI/SC recognition mechanism of the S-RNase-based GSI in *Prunus*, it is likely that the ubiquitin/proteasome proteolytic pathway is involved in the system.

Deubiquitination processes may also be involved in the *S* haplotype-dependent protection and degradation of S-RNase.

Akagi et al. (2016) conducted a genome-wide analysis of the SFB/SLFL-like F-box genes using *Prunus* genomes and provided novel clues to the evolution of *SFB* in the *Prunus* genome (Akagi et al. 2016). They propose that the divergence between the ancestors of the *Prunus* pollen SFB and the other pollen S F-box genes occurred early in the establishment of eudicots and that *SFB* was generated from *Prunus*-specific duplication. These findings suggest that the *Prunus* species came to use the SFB gene as a pollen S factor after the *Prunus*-specific F-box duplications, which probably occurred around the time of the *Prunus* divergence from its common ancestor with *Malus* or *Fragaria*. They also hypothesised that *Prunus* SLFLs are still involved in the degradation of S-RNase proteins in a process similar to that by which SLF/SFBB genes are involved in the degradation of nonself-S-RNase (Kubo et al. 2010). In this case, it would be feasible to say that SLFL genes are strong candidates for the function of GIs that putatively target all S-RNase for degradation (Matsumoto and Tao 2016). This evolutionary scenario would suggest the possibility that SFB has taken over the function of discrimination between self- and nonself-S-RNases from the original pollen S factors. After the establishment of SFB, genes in the SLF/SFBB/SLFL clade (clade S) in *Prunus* might have lost the function to distinguish self- and nonself-S-RNases. This could have resulted in the acquisition of the GI function (Matsumoto and Tao 2016). Instead, the SFB gene, which was newly duplicated in the *Prunus* S locus, acts in the self-recognition or discrimination of self- and nonself-S-RNase. Morimoto et al. (2015) also conducted an evolutionary analysis of rosaceous S-RNases and reported that the S-locus genes have experienced several duplications during the evolution of the family Rosaceae.

Recently, sweet cherry pollen-part-mutation conferring SC was characterised through a comprehensive genomic analysis using the newly

published sweet cherry genome (Ono et al. 2018). Mutation was caused by the insertion inside of the gene coding glutathione S-transferase, putatively functioning in protein folding. Non-S-locus-derived mutation could confer SC, which may provide new insights into the molecular network involved in GSI reaction in *Prunus*. Genome sequences could be important tools to facilitate comprehensive genomic analysis, such as genome-wide association studies (GWAS), which would further contribute to uncover *Prunus*-specific GSI recognition mechanisms in the future.

References

Aguiar B, Vieira J, Cunha AE, Fonseca NA, Iezzoni A, van Nocker S, Vieira CP (2015) Convergent evolution at the gametophytic self-incompatibility system in *Malus* and *Prunus*. PLoS ONE 10(5):e0126138

Akagi T, Henry IM, Morimoto T, Tao R (2016) Insights into the *Prunus*-specific S-RNase-based self-incompatibility system from a genome-wide analysis of the evolutionary radiation of S locus-related F-box genes. Plant Cell Physiol 57(6):1281–1294

Anderson MA, Cornish E, Mau S-L, Williams E, Hoggart R, Atkinson A, Bonig I, Grego B, Simpson R, Roche P (1986) Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*. Nature 321 (6065):38

Anderson MA, McFadden GI, Bernatzky R, Atkinson A, Orpin T, Dedman H, Tregear G, Fernley R, Clarke AE (1989) Sequence variability of three alleles of the self-incompatibility gene of *Nicotiana alata*. Plant Cell 1(5):483–491

Broothaerts W, Janssens GA, Proost P, Broekaert WF (1995) cDNA cloning and molecular analysis of two self-incompatibility alleles from apple. Plant Mol Biol 27(3):499–511

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu Y-L (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. Ann MO Botl Gard 80(3):528–580

De Franceschi P, Dondini L, Sanzol J (2012) Molecular bases and evolutionary dynamics of self-incompatibility in the Pyrinae (Rosaceae). J Exp Bot 63(11):4015–4032

De Nettancourt D (2001) Incompatibility and incongruity in wild and cultivated plants, vol 3. Springer Science & Business Media

Dirlewanger E, Graziano E, Joobeur T, Garriga-Caldere F, Cosson P, Howad W, Arus P (2004) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. Proc Natl Acad Sci. USA 101:9891–9896

Entani T, Iwano M, Shiba H, Che FS, Isogai A, Takayama S (2003) Comparative analysis of the self-incompatibility (S-) locus region of *Prunus mume*: identification of a pollen-expressed F-box gene with allelic diversity. Genes Cells 8(3):203–213

Golz J, Su V, Clarke A, Newbigin E (1999) A molecular description of mutations affecting the pollen component of the *Nicotiana alata* S locus. Genetics 152 (3):1123–1135

Golz JF, Oh H-Y, Su V, Kusaba M, Newbigin E (2001) Genetic analysis of *Nicotiana* pollen-part mutants is consistent with the presence of an S-ribonuclease inhibitor at the S locus. Proc Natl Acad Sci USA 98 (26):15372–15376

Habu T, Tao R (2014) Transcriptome analysis of self- and cross-pollinated pistils of Japanese apricot (*Prunus mume* Sieb. et Zucc.). J. Japan. Soc. Hort. Sci 83:95–107

Habu T, Kishida F, Morikita M, Kitajima A, Yamada T, Tao R (2006) A simple and rapid procedure for the detection of self-compatible individuals in Japanese apricot (*Prunus mume* Sieb. et Zucc.) using the loop-mediated isothermal amplification (LAMP) method. Hort Sci 41(5):1156–1158

Habu T, Matsumoto D, Fukuta K, Esumi T, Tao R, Yaegaki H, Yamaguchi M, Matsuda M, Konishi T, Kitajima A (2008) Cloning and characterization of twelve S-RNase alleles in Japanese apricot (*Prunus mume* Sieb. et Zucc.). J Jpn Soc Hort Sci 77(4):374–381

Horiuchi H, Yanai K, Takagi M, Yano K, Wakabayashi E, Sanda A, Mine S, Ohgi K, Irie M (1988) Primary structure of a base non-specific ribonuclease from R*hizopus niveus*. J Biochem 103(3):408–418

Igic B, Kohn JR (2001) Evolutionary relationships among self-incompatibility RNases. Proc Natl Acad Sci USA 98(23):13167–13171

Ikeda K, Igic B, Ushijima K, Yamane H, Hauck NR, Nakano R, Sassa H, Iezzoni AF, Kohn JR, Tao R (2004) Primary structural features of the S haplotype-specific F-box protein, SFB, in *Prunus*. Sexual Plant Reprod 16(5):235–243

Ioerger T, Gohlke J, Xu B, T-h Kao (1991) Primary structural features of the self-incompatibility protein in Solanaceae. Sexual Plant Reprod 4(2):81–87

Janick J, Moore JN (1975) Advances in fruit breeding. Purdue University Press, West Lafayette

Kao T, Tsukamoto T (2004) The molecular and genetic bases of S-RNase-based self-incompatibility. Plant Cell 16(suppl. 1):S72–S83

Kawata Y, Sakiyama F, Tamaoki H (1988) Amino-acid sequence of ribonuclease T2 from Aspergillus oryzae. Eur J Biochem 176(3):683–697

Kubo K, Entani T, Takara A, Wang N, Fields AM, Hua Z, Toyoda M, S-i Kawashima, Ando T, Isogai A (2010) Collaborative non-self recognition system in S-RNase-based self-incompatibility. Science 330 (6005):796–799

Kubo K, Paape T, Hatakeyama M, Entani T, Takara A, Kajihara K, Tsukahara M, Shimizu-Inatsugi R, Shimizu KK, Takayama S (2015) Gene duplication and genetic exchange drive the evolution of S-RNase-based self-incompatibility in *Petunia*. Nat Plants 1(1):14005

Lai Z, Ma W, Han B, Liang L, Zhang Y, Hong G, Xue Y (2002) An F-box gene linked to the self-incompatibility (S) locus of *Antirrhinum* is expressed specifically in pollen and tapetum. Plant Mol Biol 50(1):29–41

Luu D-T, Qin X, Laublin G, Yang Q, Morse D, Cappadocia M (2001) Rejection of S-heteroallelic pollen by a dual-specific S-RNase in *Solanum chacoense* predicts a multimeric SI pollen component. Genetics 159(1):329–335

Matsumoto D, Tao R (2016) Distinct self-recognition in the *Prunus* S-RNase-based gametophytic self-incompatibility system. Hort J 85(4):289–305

Matsumoto D, Yamane H, Tao R (2008) Characterization of SLFL1, a pollen-expressed F-box gene located in the *Prunus* S locus. Sexual Plant Reprod 21(2): 112–121

Matsumoto D, Yamane H, Abe K, Tao R (2012) Identification of a Skp1-like protein interacting with SFB, the pollen S determinant of the gametophytic self-incompatibility in Prunus. Plant Physiol 159 (3):1252–1262

McClure BA, Haring V, Ebert PR, Anderson MA, Simpson RJ, Sakiyama F, Clarke AE (1989) Style self-incompatibility gene products of *Nicotlana alata* are ribonucleases. Nature 342(6252):955

McCubbin AG, T-h Kao (2000) Molecular recognition and response in pollen and pistil interactions. Annu Rev Cell Dev Biol 16(1):333–364

Morimoto T, Akagi T, Tao R (2015) Evolutionary analysis of genes for S-RNase-based self-incompatibility reveals S locus duplications in the ancestral Rosaceae. Hort J 84(3):233–242

Ono K, Akagi T, Morimoto T, Wünsch A, Tao R (2018) Genome re-sequencing of diverse sweet cherry (*Prunus avium*) individuals reveals a modifier gene mutation conferring pollen-part self-compatibility. Plant Cell Physiol 59(6):1265–1275

Sassa H, Hirano H, Ikehashi H (1992) Self-incompatibility-related RNases in styles of Japanese pear (*Pyrus serotina* Rehd.). Plant Cell Physiol 33(6):811–814

Sassa H, Hirano H, Ikehashi H (1993) Identification and characterization of stylar glycoproteins associated with self-incompatibility genes of Japanese pear, Pyrus serotina Rehd. Mol Gen Genet 241(1–2):17–25

Sassa H, Nishio T, Kowyama Y, Hirano H, Koba T, Ikehashi H (1996) Self-incompatibility (S) alleles of the Rosaceae encode members of a distinct class of the T 2/S ribonuclease superfamily. Mol Gen Genet MGG 250(5):547–557

Sassa H, Kakui H, Miyamoto M, Suzuki Y, Hanada T, Ushijima K, Kusaba M, Hirano H, Koba T (2007) S locus F-box brothers: multiple and pollen-specific F-box genes with S haplotype-specific polymorphisms in apple and Japanese pear. Genetics 175(4): 1869–1881

Sassa H, Kakui H, Minamikawa M (2010) Pollen-expressed F-box gene family and mechanism of S-RNase-based gametophytic self-incompatibility (GSI) in Rosaceae. Sexual Plant Reprod 23(1):39–43

Takayama S, Isogai A (2005) Self-incompatibility in plants. Annu Rev Plant Biol 56:467–489

Tao R, Iezzoni AF (2010) The S-RNase-based gametophytic self-incompatibility system in *Prunus* exhibits distinct genetic and molecular features. Sci Hort 124 (4):423–433

Tao R, Yamane H, Sassa H, Mori H, Gradziel TM, Dandekar AM, Sugiura A (1997) Identification of stylar RNases associated with gametophytic self-incompatibility in almond (*Prunus dulcis*). Plant Cell Physiol 38(3):304–311

Tao R, Yamane H, Sugiura A, Murayama H, Sassa H, Mori H (1999) Molecular typing of S-alleles through identification, characterization and cDNA cloning for S-RNases in sweet cherry. J Amer Soc Hort Sci 124 (3):224–233

Tao R, Habu T, Yamane H, Sugiura A, Iwamoto K (2000) Molecular markers for self-compatibility in Japanese apricot (*Prunus mume*). Hort Sci 35(6):1121–1123

Tao R, Habu T, Namba A, Yamane H, Fuyuhiro F, Iwamoto K, Sugiura A (2002a) Inheritance of S f-RNase in Japanese apricot (*Prunus mume*) and its relation to self-compatibility. Theor Appl Genet 105 (2–3):222–228

Tao R, Habu T, Yamane H, Sugiura A (2002b) Characterization and cDNA cloning for Sf-RNase, a molecular marker for self-compatibility, in Japanese apricot (*Prunus mume*). J Jpn Soc Hort Sci 71(5):595–600

Tao R, Namba A, Yamane H, Fuyuhiro Y, Watanabe T, Habu T, Sugiura A (2003) Development of the Sf-RNase gene-specific PCR primer set for Japanese apricot (*Prunus mume* Sieb. et Zucc.). Hortic Res (Japan)

Tsukamoto T, Hauck NR, Tao R, Jiang N, Iezzoni AF (2006) Molecular characterization of three non-functional S-haplotypes in sour cherry (*Prunus cerasus*). Plant Mol Biol 62(3):371

Ushijima K, Sassa H, Tao R, Yamane H, Dandekar A, Gradziel T, Hirano H (1998) Cloning and characterization of cDNAs encoding S-RNases from almond (*Prunus dulcis*): primary structural features and sequence diversity of the S-RNases in Rosaceae. Mol Gen Genet 260(2–3):261–268

Ushijima K, Sassa H, Tamura M, Kusaba M, Tao R, Gradziel TM, Dandekar AM, Hirano H (2001) Characterization of the S-locus region of almond (*Prunus dulcis*): analysis of a somaclonal mutant and a cosmid contig for an S haplotype. Genetics 158 (1):379–386

Ushijima K, Sassa H, Dandekar AM, Gradziel TM, Tao R, Hirano H (2003) Structural and transcriptional analysis of the self-incompatibility locus of almond: identification of a pollen-expressed F-box gene with

haplotype-specific polymorphism. Plant Cell 15 (3):771–781

Ushijima K, Yamane H, Watari A, Kakehi E, Ikeda K, Hauck NR, Iezzoni AF, Tao R (2004) The S haplotype-specific F-box protein gene, SFB, is defective in self-compatible haplotypes of *Prunus avium* and *P. mume*. Plant J 39(4):573–586

Wang Y, Wang X, Skirpan AL, Kao Th (2003) S-RNase-mediated self-incompatibility. J Exp Bot 54 (380):115–122

Westwood M (1993) Temperate-zone pomology physiology and culture. Timber Press, Portland, Oregon, p 523

Xue Y, Carpenter R, Dickinson HG, Coen ES (1996) Origin of allelic diversity in *Antirrhinum* S locus RNases. Plant Cell 8(5):805–814

Yaegaki H, Shimada T, Moriguchi T, Haji T, Yamaguchi M, Hayama H (2001) Molecular characterization of S-RNase genes and S-genotypes in the Japanese apricot Prunus mume Sieb. et Zucc.). Sexual Plant Reprod 13(5):251–257

Yaegaki H, Miyake M, Haji T, Yamaguchi M (2002a) Determination of self-fruitfulness in Japanese apricot (*Prunus mume* Sieb. et Zucc.) cultivars. Bull Natl Inst Fruit Tree Sci 1:55–60 (In Japanese with English summary)

Yaegaki H, Miyake M, Haji T, Yamaguchi M (2002b) Determination of self-fruitfulness in Japanese apricot Prunus mume Sieb. et Zucc.) cultivars. Bull. Natl. Inst.

Fruit Tree Sci 1: 55–60 (In Japanese with English summary)

Yaegaki H, Miyake M, Haji T, Yamaguchi M (2003) Inheritance of male sterility in Japanese apricot (*Prunus mume*). HortScience 38:1422–1423

Yamane H, Tao R (2009) Molecular basis of self-(in) compatibility and current status of S-genotyping in Rosaceous fruit trees. J Jpn Soc Hort Sci 78(2):137–157

Yamane H, Tao R, Mori H, Sugiura A (2003a) Identification of a non-S RNase, a possible ancestral form of S-RNases, in *Prunus*. Mol Genet Genom 269(1):90–100

Yamane H, Ushijima K, Sassa H, Tao R (2003b) The use of the S haplotype-specific F-box protein gene, *SFB*, as a molecular marker for *S*-haplotypes and self-compatibility in Japanese apricot (*Prunus mume*). Theor Appl Genet 107(8):1357–1361

Yamane H, Fukuta K, Matsumoto D, Hanada T, Mei G, Habu T, Fuyuhiro Y, Ogawa S, Yaegaki H, Yamaguchi M, Tao R (2009) Characterization of a novel selfcompatible $S^{3'}$ haplotype leads to the development of a universal PCR marker for two distinctly originated self-compatible S haplotypes in Japanese apricot (*Prunus mume* Sieb. et Zucc.). J. Japan. Soc. Hort. Sci 78:40–48

Molecular and Developmental Biology: Pistil Abortion

13

Zhihong Gao and Ting Shi

Abstract

To understand the role of miRNAs and genes in pistil development, high-throughput sequencing was used to identify pistil-development-related miRNAs and transcripts in Japanese apricot. A combination of two-dimensional gel electrophoresis (2-DE) and matrix-assisted laser desorption/ionisation time of flight/time of flight (MALDI-TOF/TOF) approaches was used to identify the differentially expressed proteomes between perfect and imperfect flower buds in Japanese apricot. The conclusions were as follows: ACL, SAM, XTH and CCoAOMT could promote the formation of cell walls in perfect flower buds in Japanese apricot, which greatly contributes to pistil development. Here, SHT may be involved in the O-methylation of spermidine conjugates and could contribute to abnormal floral development. The identification of such differentially expressed proteins provides a new target for future studies to assess their physiological roles and significance in pistil abortion. Comparative analysis showed that six potentially novel miRNAs were differentially expressed between perfect and imperfect flower buds. Target predictions of the 13 differentially expressed miRNAs resulted in 212 target genes. It is predicted that miR319/miR319a/miR319e target ARF2 genes and that miR160a targets ARF16/17. Auxin regulates a variety of physiological and developmental processes in plants. The ARF has been reported to regulate flower and leaf development. These findings provide valuable information for further molecular mechanisms associated with pistil development.

13.1 Introduction

Floral organs play an essential role in plant sexual reproduction. However, in most floral plants, only a few of the flowers and ovules actually give rise to mature seeds and fruits (Arathi et al. 1999). Several different mechanisms have been proposed to explain the phenomenon of female sterility, including pistil abortion. The most often discussed factors of female sterility are thought to be triggered by environmental and nutritional conditions (Zinn et al. 2010; Beppu and Kataoka 2011), low sink strength (Reale et al. 2009; Morio et al. 2004), influences of pathogens (Kocsis and Jakab 2008), occurrence of sporophytic or gametophytic mutations (Wang et al. 2008), ABCDE model and other related genes (Causier et al. 2003; Jofuku et al. 1994; Peng et al. 2008) and phytohormones (Ellis et al. 2005; Kumar et al. 2011; Lim et al. 2010; Olkamoto and Omori 1991).

Z. Gao (✉) · T. Shi
College of Horticulture, Nanjing Agricultural University, No. 1 Weigang, Nanjing 210095, People's Republic of China
e-mail: gaozhihong@njau.edu.cn

Morphological studies have shown that pistil development of staminate flowers in the olive is interrupted after differentiation of the megaspore mother cell. At that stage, no starch was observed in the pistils of staminate flowers; the plastids had few thylakoid membranes and grana, and the staminate flowers appeared very similar to proplastids (Reale et al. 2009). In *Arabidopsis*, heat stress reduced the total number of ovules and increased ovule abortion (Whittle et al. 2009). Early ovule degeneration was also caused by high temperatures in sweet cherry, and ovule development was regulated by gibberellin (GA) in sweet cherry flowers (Beppu and Kataoka 2011). In addition, GA suppressed the development of the embryo sac and shortened its longevity in grapes (Olkamoto and Omori 1991).

Japanese apricot (*Prunus mume* Sieb. et Zucc) has its origin in China and is an important economical fruit crop in China and Japan (Chu 1999). Because of its attractive flowers and highly valued fruits, it has been widely planted (Shi et al. 2009). The fruits are used in the food and vintage industry and contain various physiochemical substances beneficial for human health (Xia et al. 2010, 2011). However, imperfect flowers are common and seriously affect the yield (Gao et al. 2006) (Fig. 13.1). Imperfect flowers are characterised by either pistil below the stamens, withered pistils or absence of pistils, and hence fail to bear fruits (Hou et al. 2010).

Comparative proteomic analysis has been performed for perfect and imperfect flowers, and the different proteins have been analysed in both perfect and imperfect flowers for the young bud, mature bud and blossom stages; moreover, glucose metabolism, starch metabolism and photosynthesis were related to pistil abortion (Wang 2008). More recently, real-time quantitative reverse transcription polymerase chain reaction and in situ hybridisation have shown that *PmAG* mRNA was highly expressed in the sepals, carpel and stamens, and a weak signal was detected in the seed and the nutlet. No expression was detected in the leaves or petals, but no significant differential was expressed in perfect and imperfect flowers (Hou et al. 2010).

13.2 Proteomic Maps of Perfect and Imperfect Flower Buds of Japanese Apricot

Recently, global expression profiling approaches have been used to investigate the mechanisms of plant development (Hajduch et al. 2005; Soitamo et al. 2011; Xu et al. 2010). Similar to gene-expression profiling, proteomics based on 2D-PAGE followed by MALDI-TOF/TOF is able to simultaneously analyse changes and to classify temporal patterns of protein accumulation that occur in vivo (Katz et al. 2007).

Fig. 13.1 Photographs of perfect (left) and imperfect (right) flowers. Bar scale represents 1 cm (Shi et al. 2012b)

Proteomics has established itself as an increasingly used experimental tool for the investigation of complex cellular processes, including seed, ovule, embryo and endosperm development (Bai et al. 2010; Chen et al. 2006; Gallardo et al. 2003; Kwon et al. 2010; Liu et al. 2010; Martínez-García et al. 2011; Prassinos et al. 2011).

To identify a protein involved in the development of the flowers of Japanese apricot, the 2-DE technique was used on the proteome profiles of perfect and imperfect flower buds. The results showed a consistent pattern of protein expression levels on the gels, and image analysis revealed about 400 highly reproducible protein spots that were consistently observed in all replicates (Fig. 13.2).

Among the 27 protein spots, 16 up-regulated protein spots (spots 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 14, 15, 16, 17, 18, 19) and seven down-regulated protein spots (spots 20, 21, 22, 23, 24, 25, 27) showed a more than twofold difference, and four protein spots (spots 9, 12, 13, 26) were specifically expressed. Three protein spots (spots 9, 12, 13) showed specific expression in perfect flower buds, and one spot (spot 26) was specific to imperfect flower buds, which might be closely related to pistil development and female sterility.

Gene ontology categories were assigned to all 27 proteins according to their molecular functions, biological processes and cellular components (Fig. 13.3). Based on the molecular function, the genes were finally classified into eight categories: enzyme activity (14), unclassified (4), binding (3), molecular function (2), signalling (1), transcription factor activity (1), cell structure (1) and oxidising reaction (1), as shown in Fig. 13.3a. Additionally, ten biological processes were identified: stress response and defence (8), energy metabolism (4), biosynthetic processes (4), unclassified (2), signal transduction (2), protein metabolism (2), plant development (2), transcription (1), oxidation-reduction processes (1) and microtubule-based processes (1) (Fig. 13.3b).

Comparative analysis of the 2-DE maps of perfect and imperfect flower bud proteins in Japanese apricot was performed using the PDQuest software. More than 400 highly reproducible protein spots ($P < 0.05$) were detected, and 29 protein spots were showed more than twofold differences in expression values, of which 27 were confidently identified according to the databases.

Among these differentially expressed proteins, ACL is used mainly in fatty acid and sterol biogenesis (Suh et al. 2001). Several genes of the fatty acid biosynthesis pathway of *Brassica napus* are tightly regulated in a spatiotemporal manner by ACL activity (Ratledge et al. 1997). Fatty acids and sterols play important roles in many cellular and developmental processes, such as the generation of bio membranes, hormones and secondary messengers. In plants, the formation of pollen grains and seeds is closely
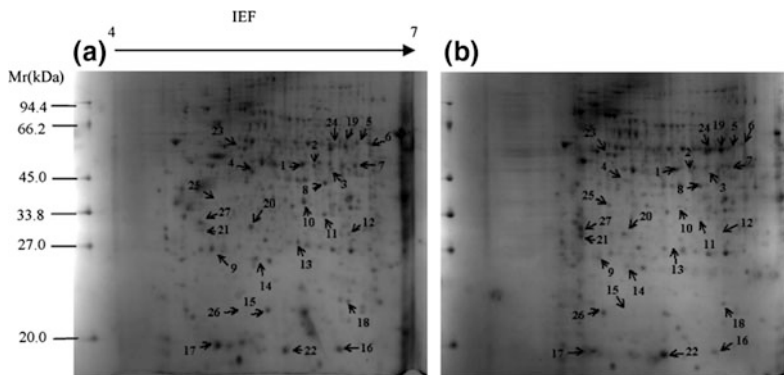

Fig. 13.2 Representative 2-DE patterns of perfect and imperfect flower buds in Japanese apricot. The proteins identified are marked with arrows and numbers, and the protein spot numbers correspond to those listed in Table 13.1 and the additional file 1. **a** Perfect flower buds; **b** imperfect flower buds (Shi et al. 2012b)
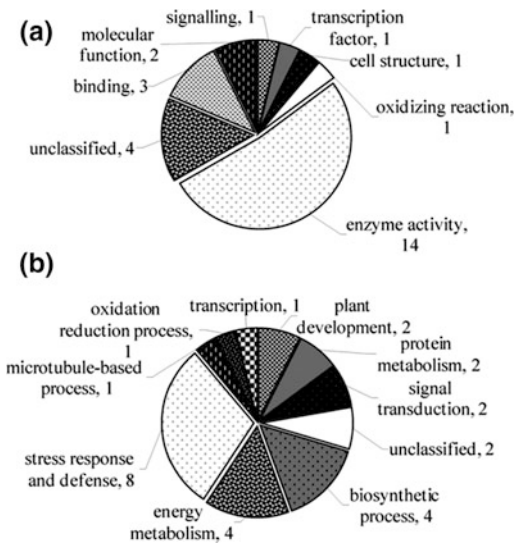
Fig. 13.3 Gene ontology of 27 differentially expressed proteins. Categorisation of proteins was performed according to the molecular function (**a**) and biological process (**b**). This categorisation was based on electronic annotation (http://www.geneontology.org/) and the literature (Shi et al. 2012b)

correlated with lipid production. An investigation of the time course of ACL expression in *Sordaria macrospora* suggests that ACL is specifically induced at the beginning of the sexual cycle and produces acetyl-CoA, which most probably is a prerequisite for fruiting body formation during later stages of sexual development, and ACL is essential for fruiting body maturation (Nowrousian et al. 1999). In our study, ACL family protein (spot 4) expression in perfect flower buds was higher than that in imperfect flower buds, indicating that the reduction of fatty acid and sterol biosynthesis increased the ratio of pistil abortion.

Our research found that the expression of SAM synthetase 1 (spots 1 and 3) and the SAM synthetase family protein (spot 2) in perfect flower buds was higher than that in imperfect flower buds. Here, SAM, a direct product of Met catabolism, is a substrate in numerous transmethylation reactions, including several reactions that occur in the biosynthesis of lignin. Lignin, a complex phenolic polymer, is important for mechanical support, water transport and defence in vascular plants (Campbell and Sederoff 1996);

p-Coumaroyl shikimate 3′-hydroxylase (C3′H) catalyses the ring *meta*-hydroxylation reaction, leading to the biosynthesis of lignin units, necessary for the biosynthesis of both G and S lignin units (Vanholme et al. 2010). Various alleles of the *C3′H*-deficient *reduced epidermal fluorescence8* (*ref8*) mutants exhibit severe female sterility in *Arabidopsis* (Weng et al. 2010). Highly lignified tissues, such as stem tissue, might be expected to have increased levels of SAMS in *Arabidopsis* (Shen et al. 2002). Also, SAM is the key compound for all transmethylation reactions such as methylation of pectin, DNA, RNA, histones and polyamine synthesis (Moffatt et al. 2002). High levels of SAM are also needed for pectin synthesis of cell walls. Pectin is transported as a highly methylated molecule into the cell wall and must be demethylated by pectin methyl esterase (PME) prior to insertion into the cell wall. Due to a decreased transmethylation capacity, the cell wall and especially pectin synthesis may have been affected. Our previous research found that the pistils of imperfect flowers stopped differentiation in early December and finally disintegrated, while the pistils of perfect flowers continued to differentiate and developed perfectly (Shi et al. 2011). This phenomenon might result from the strain formation of cell walls by SAM, which might be responsible for the lower expression of SAM and SAM synthetase family protein in imperfect compared to perfect flower buds.

During flower development, SHT is involved in the O-methylation of spermidine conjugates. Martin-Tanguy (1997) reported that elevated free polyamine and water-soluble polyamine levels (located in the ovaries) contribute to abnormal floral development, but amine conjugates (via transferases) have important functions in floral induction, floral evocation and reproduction in tobacco.

The XTHs (spot 11) are cell-wall enzymes that catalyse the cleavage and molecular grafting of xyloglucan chain functions in loosening and rearrangement of the cell wall. As they are involved in the modification of the load-bearing cell-wall components, they are believed to be extremely important in the regulation of growth

and development (Maris et al. 2011). Hyodo et al. (2003) showed that XTH9 tends to be expressed strongly in rapidly dividing and expanding tissues in *Arabidopsis.*

The CCoAOMT (spot 13) is an important enzyme and is involved in an alternative methylation pathway in lignin biosynthesis (Campbell and Sederoff 1996; Ye et al. 1994). Tissue print hybridisation showed that the expression of the CCoAOMT gene is temporally and spatially regulated and that it is associated with lignification in xylem and in phloem fibres in *Zinnia* organs (Ye et al. 1994). Lignin analysis showed that reduction in CCoAOMT alone resulted in a dramatic decrease in lignin content; the reduction in CCoAOMT also led to a dramatic alteration in lignin composition in tobacco (*Nicotiana tabacum* cv Xanthi) (Zhong et al. 1998). The levels of G lignin were most strongly reduced in line with the greatest decrease in CCoAOMT activity in alfalfa (*Medicago sativa* cv Regen SY) (Guo et al. 2001). These differentially expressed proteins might play an important role in pistil abortion in Japanese apricot.

Peroxidase was one of the main enzymes that eliminated active oxygen in the plant cell (Welinder 1992). Apparently, there is a close relationship between peroxidase (spot 7) and flower development (Sood et al. 2006; Sun et al. 2005). Genes encoding enzymes that detoxify reactive oxygen species (ROS), including ascorbate peroxidase and peroxidase, were down-regulated after ovules committed to abort (Sun et al. 2005). In our study, the expression of peroxidase (spot 7) in imperfect flower buds was lower than that in perfect flower buds. This result is consistent with previous reports. Although glutathione peroxidase (spot 26, GPX) belongs to the peroxidase family, there is no expression of glutathione peroxidase in perfect flower buds; in contrast, its expression is very high in imperfect flower buds. In general, the abundance of GPX increased upon treatment with various stresses. Therefore, we hypothesised that the forming of an imperfect flower is an actual adversity process, and the higher expression of GPX might lead to pistil abortion.

13.3 Genes Related to Pistil Abortion

To identify differentially expressed genes involved in the pistil development of flowers in Japanese apricot, we used Illumina sequencing on DGE from perfect (PF) and imperfect (IF) flower buds. A total of 3,476,249 and 3,580,677 tags were obtained from the PF and IF flower bud libraries, respectively. To increase the robustness of the approach, single-copy tags in the two libraries (141,270 in the PF and 142,287 in the IF library) were excluded from further analysis. After discarding the low-quality tags (tags containing 'N', adaptor sequences and copy number < 2), 3,331,468 and 3,434,800 tags (clean tags) remained in the PF and IF libraries, respectively, of which 129,933 (PF) and 126,485 (IF) distinct tags were obtained. There were 3448 more distinct tags in the PF than in the IF library, possibly representing genes related to pistil development. The percentage of distinct tags rapidly declined as the copy number increased, indicating that only a small portion of the transcripts was expressed at a high level under the conditions tested.

Using the *P*-value ≤ 0.05 as the threshold value, 333 differentially expressed genes, which could be categorised into 14 functional groups, were found (Fig. 13.4), which included four cellular components, four molecular functions and six biological processes. The four component categories were as follows: anchored to the membrane (9), ribosome (16), ATPase complex (10) and organelle part (96). Based on the molecular function, the genes were finally classified into four categories: ammonia ligase activity (3), acid-ammonia (or amide) ligase activity (3), Rho GTPase activator activity (2) and transferase activity (37). Additionally, six biological processes were identified: nitrogen cycle metabolism (3), proton transport (6), fatty acid metabolism (20), cellular component organisation or biogenesis at the cellular level (44), glutamine metabolism (3) and small molecule metabolism (81). The genes involved in small molecule metabolism [GO: 0044281] were the most significantly enriched in comparison

Fig. 13.4 Histogram showing the gene ontology functional enrichment of DEGs (Shi et al. 2012a)
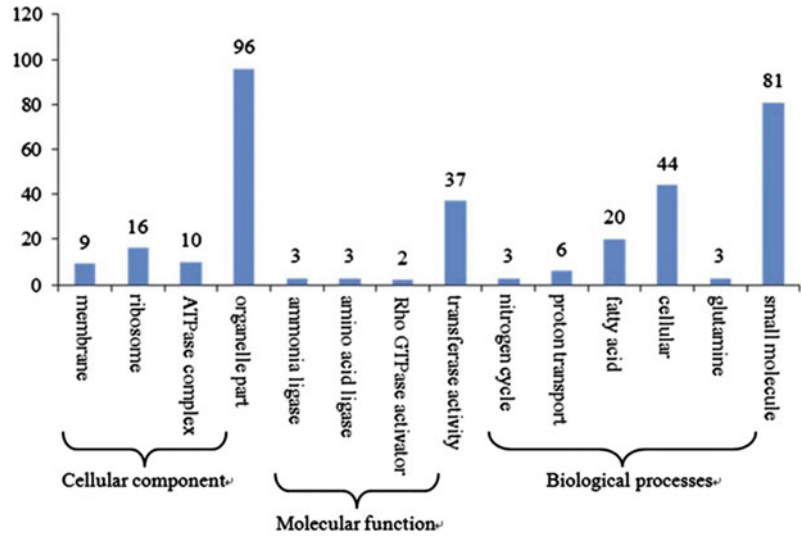
with those involved in the other five biological processes. Forty-four differentially expressed genes were involved in the cellular component organisation or biogenesis at the cellular level [GO: 0071841], which is carried out at the cellular level and results in the biosynthesis of constituent macromolecules, assembly, arrangement of constituent parts, or disassembly of a cellular component. Among the significantly enriched transcripts, 20 DEGs are involved in the regulation of fatty acid metabolism [GO: 0006631], including chemical reactions and pathways in which fatty acids and aliphatic monocarboxylic acids are hydrolyzed from natural oils.

A Q-value of ≤ 0.05 defined those genes that were significantly differentially expressed (enriched); 227 differentially expressed genes, associated with 12 metabolic and signal transduction pathways, were found (Fig. 13.5). The pathways with the unique genes were metabolic pathways (250 genes), cellular processes (66 genes) and genetic information processes (20 genes). We believe that these pathways are significant in the pistil abortion of Japanese apricot, in particular, metabolic pathways and cellular

Fig. 13.5 Histogram illustrating pathway enrichment analyses (Shi et al. 2012a)
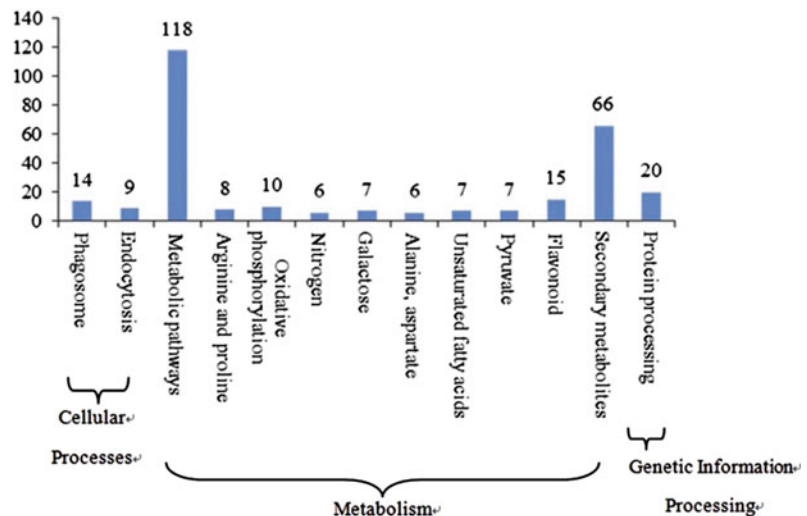
processes. In our study, metabolic pathways (ko01100) are large complexes comprising several metabolic patterns, including the biosynthesis of secondary metabolites (ko01110, 66 genes), flavonoid biosynthesis (ko00941, 15), oxidative phosphorylation (ko00190, 10 genes), arginine and proline metabolism (ko00330, eight genes), galactose metabolism (ko00052, seven genes), the biosynthesis of unsaturated fatty acids (ko01040, seven genes), pyruvate metabolism (ko00620, seven genes), nitrogen metabolism (ko00910, six genes) and alanine, aspartate and glutamate metabolism (ko00330, six genes). Cellular processes included phagosome (ko04145, 14 genes) and endocytosis (ko04144, nine genes). Genetic information processing only included protein processing in the endoplasmic reticulum (ko04141, 20 genes).

Of the DEGs with differences greater than fivefold (Table 13.1), 11 genes were present at higher levels in the IF library, associated with defence (2), metabolism (6), signal transduction (2) and transport (1) processes. The greatest differences between IF and PF DEGs were the *DCL3* (dicer-like 3) gene and the vacuolar ATPase subunit F family protein gene, both of which were present 8.95-fold higher in the IF library than in the PF library.

Forty-nine DEGs were less abundant in the IF library. Those present fivefold or more in the PF library are also listed in Table 13.1, in which 43 genes were classified as defence (5), development (2), metabolism (27), signal transduction (3), transcription (2) and transport (4) genes. The highest DEG was the fatty acid desaturase 5 (*FAD5*) gene, which was present at levels 14.82-fold higher than PF levels.

13.4 miRNAs Possibly Involved in the Regulation of Pistil Abortion in Imperfect Japanese Apricot Flower Buds

The characterisation and comparative profiling of entire sets of small RNAs (small RNA transcriptome), especially miRNAs, provide the foundation for unravelling the complex miRNA-mediated regulatory networks controlling pistil abortion in imperfect Japanese apricot flower buds. In this study, a number of miRNAs were differentially expressed between perfect and imperfect flower buds. Compared with the perfect library, four known miRNA genes and three potentially novel miRNA/new members of known miRNA family genes were expressed exclusively in imperfect flower buds. On the other hand, two potentially novel miRNAs/new members of known miRNA families were perfect-specific. Moreover, a total of seven known miRNAs and six potentially novel miRNAs exhibited significant expression changes between the perfect and imperfect libraries.

The relationship between the sRNA genes and the miRNA-target genes is one of the hot spots in the phenomenon of 'miRNA-associated transitivity'. Target prediction of these differential miRNAs could provide information on the biological processes regulated by miRNA. The annotations of these potential miRNA-target genes provided an alternate view of gene regulation of the pistil abortion trait formation in imperfect flower buds. It was discovered that these groups of predicted miRNA-target genes are possibly involved in the pistil abortion trait formation.

The majority of genes encoding transcription factors or F-box proteins play a significant role in plant development (Jones-Rhoades and Bartel 2004; Earley et al. 2010). In the present study, the predicted targets of miR319, miR319a, miR319e, miR160, miR393, miR394, miR6274, miR6295 and miR171d were either transcription factors or F-box proteins. In addition, it is predicted that miR319/miR319a/miR319e target *ARF2* genes and that miR160a targets *ARF16/17*. Auxin regulates a variety of physiological and developmental processes in plants, and ARF has been reported to regulate flower and leaf development. Also, ARF2 is a transcriptional suppressor that is involved in ethylene, auxin, ABA and brassinosteroid pathways, controlling the onset of leaf senescence, floral organ abscission and ovule development. The ARF2 promotes transitions between multiple stages of *Arabidopsis* development and positively regulates

Table 13.1 List of differentially expressed genes five or more times higher in the IF library (Shi et al. 2012a)

Gene	log2 ratio (IF/PF)	Description
Up-regulated genes		
Defence		
ppa004141m	8.86	Calcium-dependent protein kinase 13 (CPK13) (Kanchiswamy et al. 2010)
ppa018301m	8.86	CPR30 F-box and associated interaction domain-containing protein (Gou et al. 2009)
Metabolism		
ppa000615m	8.86	EMB3011 RNA helicase family protein (Meier 2012)
ppa010131m	8.67	Peptidyl-tRNA hydrolase family protein (Kroeger et al. 2009)
ppa006743m	8.67	AGD2 Pyridoxal phosphate (PLP)-dependent transferase superfamily protein
ppa006182m	8.86	Protein kinase superfamily protein
ppa021659m	8.95	Dicer-like 3 (DCL3) (Chen 2009; Bourc'his and Voinnet 2010)
ppa005799m	8.77	Pre-mRNA-splicing factor CWC26 (Laubinger et al. 2008; Chung et al. 2009)
Signal transduction		
ppa012743m	8.77	SLY2 F-box family protein
ppa003105m	8.67	Leucine-rich repeat protein kinase family protein
Transport		
ppa013294m	8.95	Vacuolar ATPase subunit F family protein
Down-regulated genes		
Defence		
ppa002249m	−8.61	Early-responsive to dehydration stress protein (ERD4)
ppa016718m	−8.99	ENODL5 early nodulin-like protein 5 (Borner et al. 2002)
ppa002203m	−9.37	FRO2 ferric reduction oxidase 2 (Sivitz et al. 2011; Chen et al. 2010)
ppa006485m	−9.61	Mitogen-activated protein kinase kinase kinase 15 (MAPKKK15) (Menges et al. 2008)
ppa021261m	−10.04	Late embryogenesis abundant protein (LEA) family protein (Thomashow 1999; Zinn et al. 2010; Goyal et al. 2005; Boavida et al. 2009)
Development		
ppa001970m	−8.71	Glutamine-rich protein 23 (GRP23)
ppa016920m	−8.99	BTB/POZ domain-containing protein
Metabolism		
ppa023612m	−5.96	Lipid transfer protein 3 (LTP3) (Liu et al. 2000)
ppa011478m	−7.11	Plant invertase/pectin methylesterase inhibitor superfamily protein
ppa023515m	−7.18	Cysteine proteinase superfamily protein
ppa009726m	−8.61	ABI-1-like 1 (ABIL1)
ppa025960m	−8.61	BEN1 NAD(P)-binding Rossmann-fold superfamily protein
ppa015204m	−8.61	SYD P-loop containing nucleoside triphosphate hydrolase superfamily protein
ppa002676m	−8.61	Pentatricopeptide repeat (PPR) superfamily protein
ppa005209m	−8.81	XF1 FAD/NAD(P)-binding oxidoreductase family protein
ppa026851m	−8.81	Subtilisin-like serine endopeptidase family protein
ppa007640m	−8.91	IRX9 Nucleotide-diphospho-sugar transferase superfamily protein

(continued)

Table 13.1 (continued)

Gene	log2 ratio (IF/PF)	Description
ppa003553m	−8.91	P-loop containing nucleoside triphosphate hydrolase superfamily protein
ppa022113m	−8.91	3-ketoacyl-CoA synthase 7 (KCS7)
ppb019226m	−9.08	Plant invertase/pectin methylesterase inhibitor superfamily protein
ppa007503m	−9.15	PLC-like phosphodiesterase superfamily protein
ppa017270m	−9.3	Alpha/beta-Hydrolases superfamily protein
ppa005976m	−9.55	Pectin lyase-like superfamily protein (PPME1) (Parre and Geitmann 2005; Bosch and Hepler 2005)
ppa018639m	−9.61	Cytochrome P450, family 735, subfamily A, polypeptide 1 (CYP735A1)
ppa013439m	−9.76	RING/U-box superfamily protein
ppa004479m	−9.95	Fatty acid biosynthesis 1 (FAB1) (Barkan et al. 2006)
ppa020149m	−10.16	Alpha dioxygenase
ppa005749m	−10.55	Purple acid phosphatase 22 (PAP22)
ppa004364m	−10.55	Non-specific phospholipase C3 (NPC3)
ppa019741m	−10.97	Xyloglucan endotransglucosylase/hydrolase 2 (XTH2) (Maris et al. 2011; Hyodo et al. 2003)
ppa025833m	−11.32	Alpha/beta-hydrolase superfamily protein
ppa020405m	−11.33	GDSL-like lipase/acylhydrolase superfamily protein
ppa027208m	−14.26	Fatty acid desaturase 5 (FAD5) (Heilmann et al. 2004)
ppa016543m	−14.82	Fatty acid desaturase 5 (FAD5) (Heilmann et al. 2004)
Signal transduction		
ppa015093m	−8.91	Cyclin D6
ppa016219m	−8.91	Stigma-specific Stig1 family protein
ppa012463m	−9.61	Pollen Ole e 1 allergen and extensin family protein
Transcription		
ppa011751m	−8.61	Myb domain protein 24 (MYB24)
ppa008450m	−9.23	Myb domain protein 73 (MYB73)
Transport		
ppa004487m	−8.61	MATE efflux family protein
ppa010364m	−8.61	TIP1;3 tonoplast intrinsic protein 1;3
ppa000945m	−8.81	HA9 H(+)-ATPase 9
ppa006913m	−9.61	AAC2 ADP/ATP carrier 2
Various functions		
ppa014641m	−8.71	Plant protein of unknown function (DUF868)
ppa020792m	−8.71	ARM repeat superfamily protein
ppa022037m	−8.81	Uncharacterised protein family (UPF0016)
ppa014616m	−8.81	Protein of unknown function (DUF1278)
ppa021534m	−10.16	Protein of unknown function (DUF679)
ppa008861m	−10.33	Family of unknown function (DUF716)

flower development. In this study, the expression of miR319/miR319a/miR319e was higher in imperfect than in perfect flower buds. Consequently, the expression of ARF2 was repressed by these miRNAs and thus regulated pistil development. Moreover, TCP2 (*TEOSINTE BRANCHED/CYCLOIDEA/PCF*) transcription factor genes and MYB33, which belong to a GAMYB-like family of transcription factors, are also targets of miR319/miR319a/miR319e in our prediction, which agrees with previous reports. Therefore, it is conceivable that the over-expression of miR319/miR319a/miR319e may contribute to an increase in imperfect flower ratios in pistil development.

References

Arathi H, Ganeshaiah K, Shaanker RU, Hegde S (1999) Seed abortion in *Pongamia pinnata* (Fabaceae). Amer J Bot 86(5):659–662

Bai S, Willard B, Chapin LJ, Kinter MT, Francis DM, Stead AD, Jones ML (2010) Proteomic analysis of pollination-induced corolla senescence in petunia. J Exp Bot 61(4):1089–1109. https://doi.org/10.1093/jxb/erp373

Barkan L, Vijayan P, Carlsson AS, Mekhedov S (2006) A Suppressor of fab1 challenges hypotheses on the role of thylakoid unsaturation in photosynthetic function. Plant Physiol 141(3):1012–1020

Beppu K, Kataoka I (2011) Studies on pistil doubling and fruit set of sweet cherry in warm climate. J Jpn Soc Hort Sci 80(1):1–13

Boavida LC, Shuai B, Yu HJ, Pagnussat GC, Sundaresan V, McCormick S (2009) A collection of Ds insertional mutants associated with defects in male gametophyte development and function in *Arabidopsis thaliana*. Genetics 181(4):1369–1385

Borner GHH, Sherrier DJ, Stevens TJ, Arkin IT, Dupree P (2002) Prediction of glycosylphosphatidylinositol-anchored proteins in *Arabidopsis*. A genomic analysis. Plant Physiol 129 (2):486–499

Bosch M, Hepler PK (2005) Pectin methylesterases and pectin dynamics in pollen tubes. Plant Cell Online 17 (12):3219–3226

Bourc'his D, Voinnet O (2010) A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. Science 330(6004):617–622

Campbell MM, Sederoff RR (1996) Variation in lignin content and composition. Plant Physiol 110:3–13

Causier B, Cook H, Davies B (2003) An *Antirrhinum* ternary complex factor specifically interacts with C-function and SEPALLATA-like MADS-box factors. Plant Mol Biol 52:1051–1062

Chen X (2009) Small RNAs and their roles in plant development. Annu Rev Cell Dev Biol 25:21–44

Chen Y, Chen T, Shen S, Zheng M, Guo Y, Lin J, Baluska F, Samaj J (2006) Differential display proteomic analysis of *Picea* meyeri pollen germination and pollen-tube growth after inhibition of actin polymerization by latrunculin B. Plant J 47(2):174–195. https://doi.org/10.1111/j.1365-313X.2006.02783.x

Chen WW, Yang JL, Qin C, Jin CW, Mo JH, Ye T, Zheng SJ (2010) Nitric oxide acts downstream of auxin to trigger root ferric-chelate reductase activity in response to iron deficiency in *Arabidopsis*. Plant Physiol 154(2):810–819

Chu MY (1999) China fruit records-Mei. China Forestry, Beijing (in Chinese)

Chung T, Wang D, Kim CS, Yadegari R, Larkins BA (2009) Plant SMU-1 and SMU-2 homologues regulate pre-mRNA splicing and multiple aspects of development. Plant Physiol 151(3):1498–1512

Earley K, Smith M, Weber R, Gregory B, Poethig R (2010) An endogenous F-box protein regulates ARGONAUTE1 in *Arabidopsis thaliana*. Silence 1 (1):1–10

Ellis CM, Nagpal P, Young JC, Hagen G, Guilfoyle TJ, Reed JW (2005) *AUXIN RESPONSE FACTOR1* and *AUXIN RESPONSE FACTOR2* regulate senescence and floral organ abscission in *Arabidopsis thaliana*. Development 132(20):4563–4574. https://doi.org/10.1242/dev.02012

Gallardo K, Le Signor C, Vandekerckhove J, Thompson RD, Burstin J (2003) Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiol 133(2):664–682

Gao Z, Wang S, Zhang Z (2006) Comparative study on flower and fruit characteristics of 29 varieties in Japanese apricot (*Prunus mume* Sieb. et Zucc.) Jiangsu Agri Sci 6:231–233

Gou M, Su N, Zheng J, Huai J, Wu G, Zhao J, He J, Tang D, Yang S, Wang G (2009) An F-box gene, CPR30, functions as a negative regulator of the defense response in Arabidopsis. Plant J 60(5):757–770

Goyal K, Walton LJ, Tunnacliffe A (2005) LEA proteins prevent protein aggregation due to water stress. Biochem J 388(Pt 1):151

Guo D, Chen F, Inoue K, Blount JW, Dixon RA (2001) Downregulation of caffeic acid 3-O-methyltransferase and caffeoyl coA 3-O-methyltransferase in transgenic alfalfa: impacts on lignin structure and implications for the biosynthesis of G and S lignin. Plant Cell 13:73–88

Hajduch M, Ganapathy A, Stein JW, Thelen JJ (2005) A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. Plant Physiol 137(4):1397–1419

Heilmann I, Mekhedov S, King B, Shanklin J (2004) Identification of the *Arabidopsis* palmitoyl-monogalactosyldiacylglycerol Δ7-desaturase gene FAD5, and effects of plastidial retargeting of *Arabidopsis* desaturases on the fad5 mutant phenotype. Plant Physiol 136(4):4237–4245

Hou J, Gao Z, Wang S, Zhang Z, Chen S, Ando T, Zhang J, Wang X (2010) Isolation and characterization of an *AGAMOUS* homologue *PmAG* from the Japanese Apricot (*Prunus mume* Sieb. et Zucc.). Plant Mol Biol Rep 29(2):473–480. https://doi.org/10.1007/s11105-010-0248-3

Hyodo H, Yamakawa S, Takeda Y, Tsuduki M, Yokota A, Nishitani K, Kohchi T (2003) Active gene expression of a xyloglucan endotransglucosylase/hydrolase gene, *XTH9*, in inflorescence apices is related to cell elongation in *Arabidopsis thaliana*. Plant Mol Biol 52:473–482

Jofuku KD, Boer BGW, Montagu MV, Okamuro JK (1994) Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*. Plant Cell 6(9):1211

Jones-Rhoades MW, Bartel DP (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell 14(6):787–799. https://doi.org/10.1016/j.molcel.2004.05.027

Kanchiswamy C, Takahashi H, Quadro S, Maffei M, Bossi S, Bertea C, Zebelo S, Muroi A, Ishihama N, Yoshioka H (2010) Regulation of *Arabidopsis* defense responses against *Spodoptera littoralis* by CPK-mediated calcium signaling. BMC Plant Biol 10(1):97

Katz E, Fon M, Lee YJ, Phinney BS, Sadka A, Blumwald E (2007) The citrus fruit proteome: insights into citrus fruit metabolism. Planta 226(4):989–1005. https://doi.org/10.1007/s00425-007-0545-8

Kocsis M, Jakab G (2008) Analysis of BABA (β-aminobutyric acid)-induced female sterility in *Arabidopsis* flowers. Acta Biol Szeged 52(1):247–249

Kroeger TS, Watkins KP, Friso G, Van Wijk KJ, Barkan A (2009) A plant-specific RNA-binding domain revealed through analysis of chloroplast group II intron splicing. Proc Natl Acad Sci USA 106(11):4537

Kumar R, Tyagi AK, Sharma AK (2011) Genome-wide analysis of auxin response factor (ARF) gene family from tomato and analysis of their role in flower and fruit development. Mol Genet Genom 285(3):245–260. https://doi.org/10.1007/s00438-011-0602-7

Kwon YS, Ryu CM, Lee S, Park HB, Han KS, Lee JH, Lee K, Chung WS, Jeong MJ, Kim HK (2010) Proteome analysis of *Arabidopsis* seedlings exposed to bacterial volatiles. Planta 232(5):1355–1370

Laubinger S, Sachsenberg T, Zeller G, Busch W, Lohmann JU, Rätsch G, Weigel D (2008) Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 105(25):8795

Lim PO, Lee IC, Kim J, Kim HJ, Ryu JS, Woo HR, Nam HG (2010) Auxin response factor 2 (ARF2) plays a major role in regulating auxin-mediated leaf longevity. J Exp Bot 61(5):1419–1430. https://doi.org/10.1093/jxb/erq010

Liu HC, Creech RG, Jenkins JN, Ma DP (2000) Cloning and promoter analysis of the cotton lipid transfer protein gene Ltp31. Biochim Biophys Acta-Mol Cell Biol Lipids 1487(1):106–111

Liu H, Liu YZ, Zheng SQ, Jiang JM, Wang P, Chen W (2010) Comparative proteomic analysis of longan (*Dimocarpus longan* Lour.) seed abortion. Planta 231(4):847–860. https://doi.org/10.1007/s00425-009-1093-1

Maris A, Kaewthai N, Eklöf JM, Miller JG, Brumer H, Fry SC, Verbelen JP, Vissenberg K (2011) Differences in enzymic properties of five recombinant xyloglucan endotransglucosylase/hydrolase (XTH) proteins of *Arabidopsis thaliana*. J Exp Bot 62(1):261

Martin-Tanguy J (1997) Conjugated polyamines and reproductive development: biochemical, molecular and physiological approaches. Physiol Plant 100(3):675–688

Martínez-García PJ, Dicenta F, Ortega E (2011) Anomalous embryo sac development and fruit abortion caused by inbreeding depression in almond (*Prunus dulcis*). Sci Hort 133:23–30

Meier I (2012) mRNA export and sumoylation—lessons from plants. Biochim Biophys Acta-Gene Regul Mechan 1819(6):531–537

Menges M, Dóczi R, Ökrész L, Morandini P, Mizzi L, Soloviev M, Murray JAH, Bögre L (2008) Comprehensive gene expression atlas for the *Arabidopsis* MAP kinase signalling pathways. New Phytol 179(3):643–662

Moffatt BA, Stevens YY, Allen MS, Snider JD, Pereira LA, Todorova MI, Summers PS, Weretilnyk EA, Martin-McCaffrey L, Wagner C (2002) Adenosine kinase deficiency is associated with developmental abnormalities and reduced transmethylation. Plant Physiol 128(3):812–821

Morio K, Kobayashi K, Ogiso E, Yokoo M (2004) Photosynthesis and dry-matter production during ripening stage in a female-sterile line of rice. Plant Prod Sci 7(2):184–188

Nowrousian M, Masloff S, Pöggeler S, Ulrich K (1999) Cell differentiation during sexual development of the fungus *Sordaria macrospora* requires ATP citrate lyase activity. Mol Cellular Biol 19(1):450

Olkamoto G, Omori N (1991) Effect of the levels of minerals, phytohormones and pistil extracts on in vitro ovule development and pollen tube growth in excised grape pistils. J Jpn Soc Hort Sci 60(3):521–529

Parre E, Geitmann A (2005) Pectin and the role of the physical properties of the cell wall in pollen tube growth of *Solanum chacoense*. Planta 220(4):582–592

Peng S, Luo T, Zhou J, Niu B, Lei N, Tang L, Chen F (2008) Cloning and quantification of expression levels of two MADS-box genes from *Momordica charantia*. Biol Plant 52(2):222–230

Prassinos C, Rigas S, Kizis D, Vlahou A, Hatzopoulos P (2011) Subtle proteome differences identified between post-dormant vegetative and floral peach buds. J Proteom 74(5):607–619. https://doi.org/10.1016/j.jprot.2011.01.018

Ratledge C, Bowater MDV, Taylor PN (1997) Correlation of ATP/citrate lyase activity with lipid accumulation in developing seeds of Brassica napus L. Lipids 32 (1):7–12

Reale L, Sgromo C, Ederli L, Pasqualini S, Orlandi F, Fornaciari M, Ferranti F, Romano B (2009) Morphological and cytological development and starch accumulation in hermaphrodite and staminate flowers of olive (Olea europaea L.). Sexual Plant Reprod 22 (3):109–119

Shen B, Li C, Tarczynski MC (2002) High free-methionine and decreased lignin content result from a mutation in the Arabidopsis S-adenosyl-L-methionine synthetase 3 gene. Plant J 29(3):371–380

Shi J, Gong J, Liu J, Wu X, Zhang Y (2009) Antioxidant capacity of extract from edible flowers of Prunus mume in China and its active components. LWT-Food Sci Technol 42(2):477–482

Shi T, Zhang QL, Gao ZH, Zhang Z, Zhuang WB (2011) Analyses on pistil differentiation process and related biochemical indexes of two cultivars of prunus mume. J Plant Resour Environ 20(4):35–41

Shi T, Gao Z, Wang L, Zhang Z, Zhuang W, Sun H, Zhong W (2012a) Identification of differentially-expressed genes associated with pistil abortion in Japanese apricot by genome-wide transcriptional analysis. PLoS One 7(10):e47810. https://doi.org/10.1371/journal.pone.0047810

Shi T, Zhuang W, Zhang Z, Sun H, Wang L, Gao Z (2012b) Comparative proteomic analysis of pistil abortion in Japanese apricot (Prunus mume Sieb. et Zucc). J Plant Physiol 169(13):1301–1310. https://doi.org/10.1016/j.jplph.2012.05.009

Sivitz A, Grinvalds C, Barberon M, Curie C, Vert G (2011) Proteasome-mediated turnover of the transcriptional activator FIT is required for plant iron-deficiency responses. Plant J

Soitamo AJ, Jada B, Lehto K (2011) HC-Pro silencing suppressor significantly alters the gene expression profile in tobacco leaves and flowers. BMC Plant Biol 11:68. https://doi.org/10.1186/1471-2229-11-68

Sood S, Vyas D, Nagar PK (2006) Physiological and biochemical studies during flower development in two rose species. Sci Hort 108(4):390–396

Suh MC, Yi SY, Lee S, Sim WS, Pai HS, Choi D (2001) Pathogen-induced expression of plant ATP: citrate lyase1. FEBS Lett 488(3):211–212

Sun K, Cui Y, Hauser B (2005) Environmental stress alters genes expression and induces ovule abortion: reactive oxygen species appear as ovules commit to abort. Planta 222:632–642. https://doi.org/10.1007/s00425-0050010-5

Thomashow MF (1999) Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. Annu Rev Plant Biol 50(1):571–599

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. Plant Physiol 153(3):895–905. https://doi.org/10.1104/pp.110.155119

Wang S (2008) Preliminary studies on differences of related characteristics between perfect flower and imperfect flower and protemics in Japanese apricot. Nanjing Agricultural University, Nanjing

Wang H, Liu Y, Bruffett K, Lee J, Hause G, Walker JC, Zhang S (2008) Haplo-insufficiency of MPK3 in MPK6 mutant background uncovers a novel function of these two MAPKs in Arabidopsis ovule development. Plant Cell 20(3):602–613. https://doi.org/10.1105/tpc.108.058032

Welinder KG (1992) Superfamily of plant, fungal and bacterial peroxidases. Curr Opin Struct Biol 2(3):388–393

Weng JK, Mo H, Chapple C (2010) Over expression of F5H in COMT deficient Arabidopsis leads to enrichment of an unusual lignin and disruption of pollen wall formation. Plant J 64(6):898–911

Whittle CA, Otto SP, Johnston MO, Krochko JE (2009) Adaptive epigenetic memory of ancestral temperature regime in Arabidopsis thaliana. Botany 87(6):650–657

Xia D, Shi J, Gong J, Wu X, Yang Q, Zhang Y (2010) Antioxidant activity of Chinese mei (Prunus mume) and its active phytochemicals. J Med Plants Res 4 (12):1156–1160

Xia D, Wu X, Shi J, Yang Q, Zhang Y (2011) Phenolic compounds from the edible seeds extract of Chinese Mei (Prunus mume Sieb. et Zucc) and their antimicrobial activity. LWT-Food Sci Technol 44(1):347–349

Xu Q, Liu Y, Zhu A, Wu X, Ye J, Yu K, Guo W, Deng X (2010) Discovery and comparative profiling of micro-RNAs in a sweet orange red-flesh mutant and its wild type. BMC Genom 11:246. https://doi.org/10.1186/1471-2164-11-246

Ye ZH, Kneusel RE, Matern U, Varner JE (1994) An alternative methylation pathway in lignin biosynthesis in Zinnia. Plant Cell 6:1427–1439

Zhong R, Morrison WH III, Negrel J, Ye ZH (1998) Dual methylation pathways in lignin biosynthesis. Plant Cell 10:2033–2045

Zinn KE, Tunc-Ozdemir M, Harper JF (2010) Temperature stress and plant sexual reproduction: uncovering the weakest links. J Exp Bot 61(7):1959–1968. https://doi.org/10.1093/jxb/erq053

Reconstruction of Ancestral Chromosomes of the Family Rosaceae

14

Zhihong Gao and Shahid Iqbal

Abstract

Genome research in Rosaceae plants is of great significance for improving fruit quality and production through various breeding and genetic engineering programmes. The reconstruction of the ancestral chromosomes of the Rosaceae family provides a new insight to study its evolution and phylogeny. Modern breeding programes, such as polyploidy, bacterial artificial chromosomes, whole-genome sequencing, physical and genetic maps, molecular markers and other omics studies, represent a fundamental ways to overcome the challenges and improve crops. This chapter introduces the latest progress in Rosaceae genomics in breeding and crop improvement.

14.1 Background and Significance

Rosaceae is a vast angiosperm family with about 3000 species, three subfamilies, 16 tribes and 8–100 genera, which are widely distributed in the temperate areas of the northern hemisphere (Hummer and Janick 2009; Phipps 2014), including numerous plants with attractive flowers (roses, meadow sweets, hawthorn, crabapple and rowans). Five Rosaceae genomes are available, including strawberry (*Fragaria vesca*), domesticated apple (*Malus domestica*), pear (*Pyrus bretschneideri*), peach (*Prunus persica*), and Mei (*Prunus mume*, related to apricot) genomes (Shulaev et al. 2011; Velasco et al. 2010; Verde et al. 2013; Wu et al. 2013; Zhang et al. 2012b) and thus providing essential resources for comparative analyses.

Phylogeny determination is the key for Rosaceae phylogenetic tree indicates that the active evolutionary radiation of the lineage inside the family corresponds to the new variance of the genera 3. This may be due to the rapid evolution of Rosaceae, indicating substantial phenotypic diversity and mutual morphological synapomorphies; for example, chromosome number, fruit type, and plant growth habits that are not easy to identify within the family (Potter et al. 2007, 2002; Morgan et al. 1994). *P. mume* (mei, $2n = 2x = 16$) is a woody-type perennial tree with an extensive penetration period of 3–5 years. As the most popular ornamental and fruit tree in East Asian countries, *P. mume* has been domesticated for more than 3000 years (Chen 1996). The *P. mume* genome can be used to understand the structure of the Rosaceae plant genome and its evolution. Furthermore, it helps to determine and identify the genetic diversity of the wild group of fruit plants and to determine how this diversity is related to the enormous

Z. Gao (✉) · S. Iqbal
College of Horticulture, Nanjing Agricultural University, No. 1 Weigang, Nanjing 210095, People's Republic of China
e-mail: gaozhihong@njau.edu.cn

phenotypic diversity of fruit trees. Moreover, different genome-based techniques can be developed for the breeding improvement programmes.

Zhang et al. (2012b) worked on DNA sequencing, linking the *P. mume* genome with existing data, and reconstructed nine ancestral chromosomes of the Rosaceae family; further, the chromosomal fusion, fission and replication histories of three subfamilies were described. The *P. mume* genome sequence is an important source for biological investigations and breeding. The genome has also increased our understanding of Rosaceae evolution. The analysis of the *P. mume* genome and transcriptome can provide insight into the mechanisms of flower fragrance, flowering dormancy and disease resistance (Zhang et al. 2012b). Therefore, studying the evolution of fruit trees is essential for understanding the evolution of angiosperms, also affecting other fields such as ecology and agriculture.

14.2 Description of the Evolutionary Model and Chromosomal Structure of Rosaceae and *Prunus mume*

Angiosperms produce seeds different from other plants, and protect seeds from water and other environmental effects (Ridley 1930; Seymour et al. 2013). In particular, the Rosaceae family produces several important fruits (apricots, cherries, apples, pears, peaches, plums, strawberries, etc.) with economic significance. Therefore, Rosaceae fruit trees provide an excellent system for relative and evolutionary research.

The first molecular phylogenetic study based on *rbcl* sequences proposed that Maleae (such as apples and pears), Rosoideae (such as strawberries, roses) and Amygdaleae (such as Tao Li) form a monophyletic branch, while Spiraeoideae is not monophyletic due to many different branches (Morgan et al. 1994). In recent years, phylogenetic studies with several genes have proved to be very useful for the phylogenetic reconstruction of fast radiated hard angiosperms, such

as the main lineages of the angiosperms Brassica and Caryophyllales (Huang et al. 2015; Yang et al. 2015; Zeng et al. 2014; Zhang et al. 2012a; Zimmer and Wen 2015). The Maleae ancestor is expected to be a hybrid of the Amygdaleae ancestors, partly because the number of basic chromosomes is 17; in addition to the early branch of the genus *Vauquelinia*, the number of basic chromosomes is 15, and the numbers of parental chromosomes are 9 and 8 for Spiraeoideae and Amygdaleae, respectively (Robertson et al. 1991) (Figs. 14.1 and 14.2).

14.3 Different Techniques for Rosaceae Crop Improvement

In Rosaceae crops, genome advancements are used to understand important agronomic traits, to reduce production costs and to improve product quality. They are also used for the genetic development of varieties through different breeding and genetic engineering programmes. The term "translational genomics" is used in genetic research to assess genomic data and critical findings. Several crop traits limit their specific genetic varieties, but breeding is improving to provide more genetic varieties. The improved new varieties have excellent properties and quality. Genomics enhances crop genetic research by identifying, quantifying and validating important genome sights to explain the genetic structure of traits. It can also effectively be used to manipulate the germplasm resources of descendants to obtain the best genetic combination and to ensure good growth and performance.

New genetic studies have investigated the most important physiological mechanisms for horticultural crop production, especially Rosaceae crops, such as dormancy, polyploidy, scion rootstock interaction and genetic diversity. According to these characteristics, Rosaceae species are ideal organisms to enhance our understanding of existing varieties and to design breeding programs for the next generation.
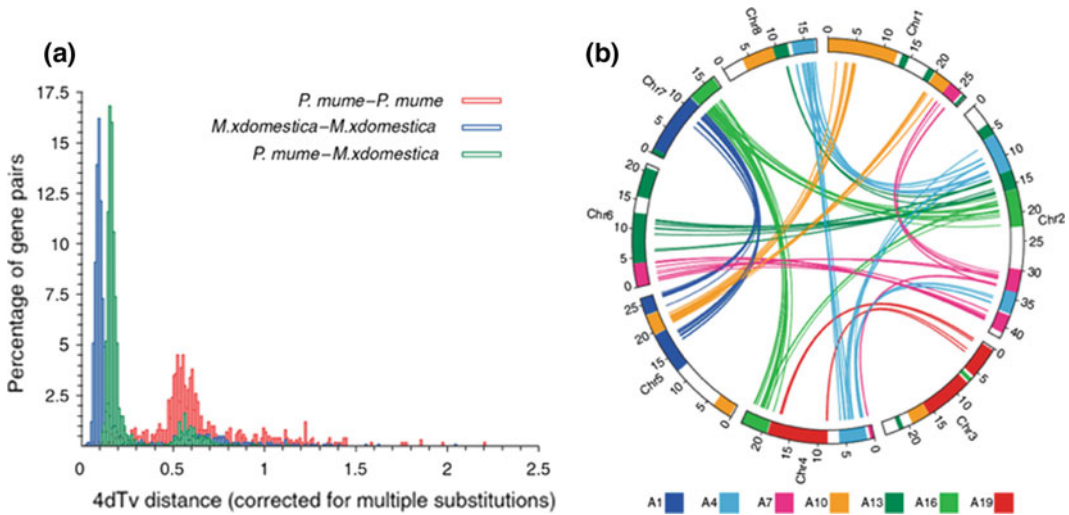
Fig. 14.1 *Prunus mume* evolution **a** 4DTv dispersal of duplicate gene combines in *P. mume* and *M. domestica*, intended through codon alignments using the HKY (Hasegawa, Kishino and Yano) substitution model. **b** The duplication of *P. mume* by paralogous pairs in the *P. mume* genome (chromosomes Chr1–Chr8). Each line shows a duplicated gene. The seven colours reflect the seven ancestral eudicot linkage groups (A1, A4, A7, A10, A13, A16 and A19) (Zhang et al. 2012b)

14.3.1   Whole-Genome Sequencing

Whole-genome sequencing (WGS) is the best technique for identifying genetic diversity and understanding the relationship between genetic inheritance and genetic traits. The rapid development of genome sequencing technology has led to whole-genome sequencing, which is a higher technology to discover genetic polymorphism. Whole-genome sequencing (WGS) has been performed for humans and many agronomic crops, especially horticultural crops such as members of the family Rosaceae. Many new tools can be used to study Rosaceae crops, with different specific areas such as expressed sequence tags (EST), single genomes, microarrays, proteomics and metabolomics. Gene function verification, with high-throughput plant transformation systems and cross-validation, is also being established.

The objective of plant genomics is to recognize, separate and regulate the functions of plant genes related to nutritional and generative phenotypes. Many phenotypes need similar activity and regulatory control and an accurate location of genes inside the plant. In recent years, this technique has been considered the best way to separate and identify genes of different phenotypes, and sometimes, complete genome sequencing is needed to obtain gene sequencing data. Some plant genome sequences have been achieved, including *Arabidopsis thaliana* (Goodman et al. 1995), *rice* (Jackson 2016) and many other crops. Furthermore, a large number of expressed sequence tags (EST) are currently being attained for numerous other plants as well as Rosaceae plants. The study of genomics and its phylogenetic information has progressed. Most genes have been recognized without whole-genome sequencing. Even so, the identification method of the EST gene does not provide relevant information about promoters and other non-coding regulations. In the past 10 years, *A. thaliana* has been the more extensively used model organisms in higher plant biology due to its highly compact genome size of about 125 Mb and small scattering and repeated DNA (Arabidopsis genome project 2000) (Goff et al. 2002).

Next-generation sequencing (NGS), combined with reference genome sequence (RGS) data, enables us to identify differences between
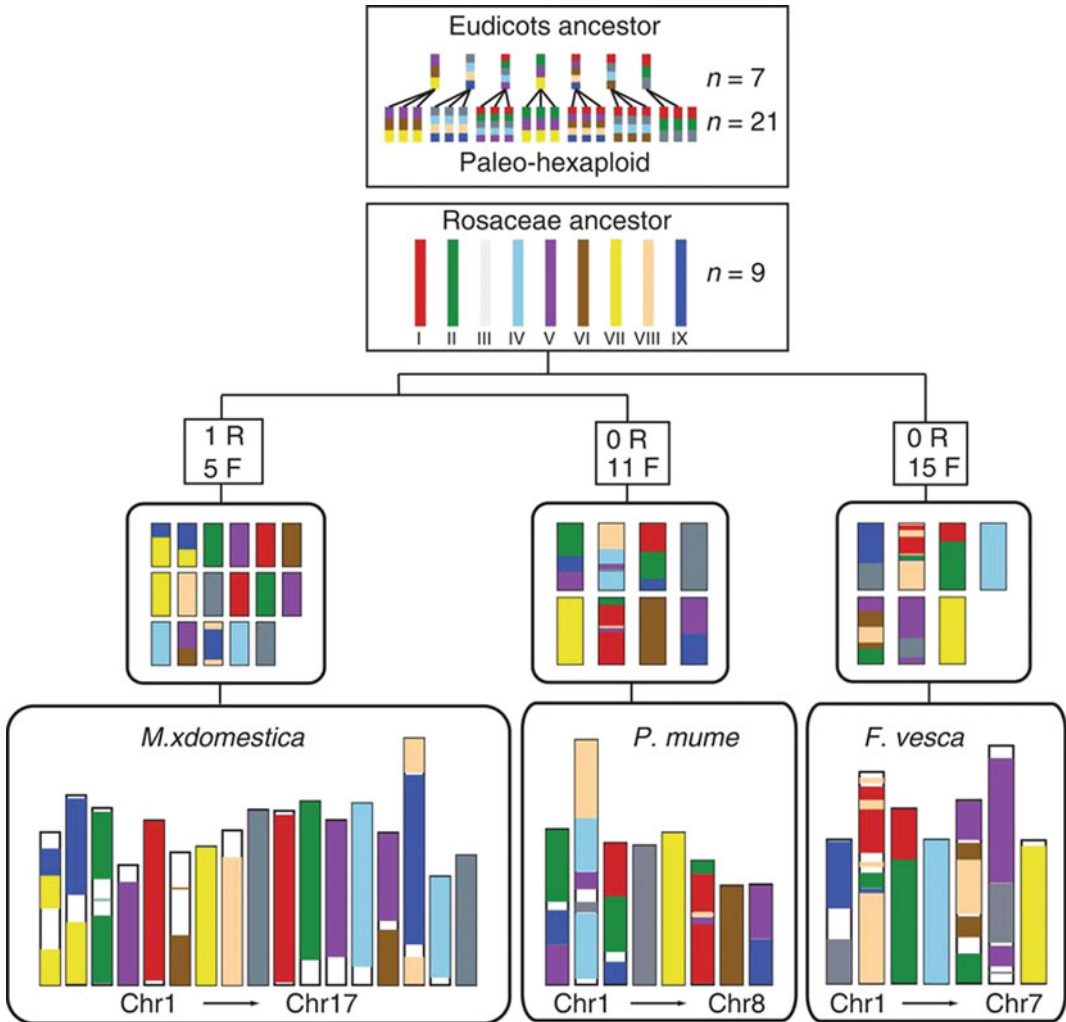
Fig. 14.2 Evolutionary model of the Rosaceae genome. The nine colours depict ancestor chromosomes of Rosaceae. Several evolutionary processes from the common ancestor represented as R (WGD) and F (chromosome fusion). On the second level, the different colours in each chromosome represent the origin of the chromosome from the common ancestor. The present Rosaceae genome structure is displayed at the bottom of the figure. White spaces showed undetermined ancestral chromosomes (unknown origin and their regions) (Zhang et al. 2012b)

individuals, strains and populations. Mapping sequence fragments to specific reference genome data sets effectively identify nucleotide polymorphisms. A whole-genome re-sequencing project, which finds 1001 strains of *Arabidopsis*, will lead to a data set that will promote future genetic research to determine the primary resources for alleles related through phenotypic diversity across whole genomes or species (http://1001genomes.org/) (Weigel and Mott

2009). Genotyping was carried out in rice using high-throughput technology based on genome-wide sequencing data generated through an Illumina genome analyser (Huang et al. 2009). Whole-genome de novo sequencing is a great novelty of next-generation sequencing. So far, this method has been implemented only in the bacterial genomes (Farrer et al. 2009; Moran et al. 2009), while some efforts are being made to achieve a similar progress in higher species.

14.3.2   Genetic Linkage Maps, Physical Maps and Bin Mapping

14.3.2.1   Genetic Linkage Maps

Genetic linkage maps have become a major tool for genetic and other breeding programmes in plant and animal species. They provide the opportunity to release the complex genetics of quantitative genetic traits through the identification and localisation of individual genes, the development of the whole-genome physical map, the assembly and annotation of the entire genome sequence; they also represent a useful marker library for marker-assisted breeding (MAB). Among these informative maps, the MAB map is constructed using the parental genotype directly involved in breeding programmes. The genetic linkage map of Rosaceae crops shows an extensive variety of germplasm resources and has important cultivated species of *Prunus*, including peach cultivars (Dirlewanger et al. 2006; Yamamoto et al. 2005), apricot (Dondini et al. 2007; Hurtado et al. 2002), apple (Hemmat et al. 1994; Kenis and Keulemans 2005; Naik et al. 2006), pear (Yamamoto et al. 2004), and raspberry (Graham et al. 2004). Specific markers are used for the construction of various genetic maps for a specific breeding programme.

Some applications of molecular markers in plant breeding require a short interval to mark the linkage map of the whole genome, the so-called saturation map (Tanksley et al. 1989), which is of great significance to detect the phenotypic variation of QTL by separating phylogenetic variation and other agronomic traits. To obtain more map advantages, we must select markers with high repeatability and transferability, such as RFLP, SSR or isozyme markers. The progress of molecular markers allows the improvement of taxonomy and classification, providing different tools for genetic linkage maps, fingerprinting and variety evaluation e.g. raspberry. Genetic linkage maps are the prerequisites for the study of the molecular information necessary for qualitative and quantitative inheritance and integration of molecular marker-assisted selection (MAS) and MAP-based gene cloning technology (Morgante and Salamini 2003). These maps are valuable for quantitative trait loci (QTL) analysis, fixing bacterial artificial chromosomes (BAC) and genome sequences. Mapping is also used to determine the ascertain location of identified markers in the genome. Mapping tools have stared to come up with the genetic linkage map of *Drosophila* and other model organisms at the beginning of the twentieth century (Sturtevant 1913), followed by the first basic physical map about 20 years later (Bridges 1935).

14.3.2.2   Physical Mapping

The physical map is an essential resource for location cloning, marker development and QTL localization. It can be used for various Rosaceae plants such as apple and peach. The physical map of peach is projected to cover 287.1 Mb of the total 290 Mb genome size (Baird et al. 1994) and is composed of 1899 overlapping groups, through which 239 are attached to eight linkage groups of the plum reference map. For peach, there is another generation map (Zhebentyayeva et al. 2006), amplified through the maximum information content fingerprint of the supplementary BAC library cloned from the existing BAC library. A total of 2511 tags are included in the physical framework. Because of sufficient hybridization data, the preliminary physical map of peach is biased towards the expressed genomic region. Permitting moderate estimates, the map covers the completely exact colour peach genome and is published in www.rosaceae.org. The apple genome physical map, containing 74,281 BAC clones based on BAC, has been constructed and represents about 10.53 haploid parallels (Han et al. 2007), while the *F. vesca* physical map is constructed with a large insert BAC Library (V. Shulaev and A.G. Abbott, personal communication).

For most important species of Rosaceae, numerous insertion libraries have been produced. A bacterial artificial chromosome (BAC) library

has been constructed for peach (Georgi et al. 2002), apricot (Vilanova et al. 2003), plum (Claverie et al. 2004) and apple (Vinatzer et al. 1998; Xu et al. 2001) as well as for cherry, strawberry and rose. Using the BAC library resources of the peach tree, the international Rosaceae mapping project (IRMP) is building the whole-genome map of the peach tree, based on the *Prunus* genetic map reference (Joobeur et al. 1998).

14.3.2.3 Bin Mapping

A previous study (Vision et al. 2000) proposed selective or "bin" mapping with the same concept as linkage information. Individuals with more recombinant divisions are usually more useful than those with fewer ones. The amounts and locations of these divisions of each given mapping population can be assumed from the marker genotype. Based on this information, a subdivision of an individual (bin set) with a broken point location can be selected, which allows mapping with minimal precision loss compared to the map attained via the complete group.

The concept of "bin" mapping was first applied to the TXE population of *Prunus* (Howad et al. 2005). The aim is to discover the smallest bin set that may allow any new marker to position in the unique region of the *Prunus* genome. Six plants were selected, dividing the TXE map into 67 fragments (bins) and allowing new markers to position in a map area with an average of 7.8 cM (the longest bin, 24.7 cM). This was verified in a group of 264 new SSR, using DNA from eight plants as analysis units: six selected plants with one parent ("Earlygold") and hybrid plants for obtaining TXE F2 populations. All these markers were effectively mapped to "bin", increasing the number of SSR in TXE to 449 and increasing the average marking density of the TXE graph (0.63 cM/marker). Considering that the leaves or DNA samples of these eight plants can be shifted to other laboratories for ease of use and that TXE is an extremely polymorphic and well-described population, "bin" mapping promotes further use of the *Prunus* reference map.

14.3.3 Polyploidy/Whole-Genome Duplication (WGD)

Polyploidy or whole-genome duplication (WGD) is an important mechanism and a critical component of the plant genome structure in many plants. About 30–80% of all angiosperms are polyploid. It is easy to detect polyploidy by comparing chromosome number, diploid process, chromosome structure and repeat events. In fact, all angiosperms, including seed plants, have to go through one or more rounds of polyploidy. In many eukaryotes, polyploidy is now considered as a fundamental process and plays a vital role in determining the genome structure and regulating gene mechanisms.

Whole-genome duplications are now a vibrant model organism for extensive evolutionary research. In addition, WGD occurs in a wide variety of spectral of phylogenetic lineages, containing many model systems that depend on cell, molecular and developmental biology. Budding yeast is the successor of ancient genome replication (Wolfe and Shields 1997) such as the zebrafish (Postlethwait et al. 2000). Because thousands of repetitive genes were added to the genome during genome replication, it is difficult to understand the evolutionary process of an individual repetitive gene (such as neo-functional and sub-functional grouping), and polyploidy may affect all duplicate genes, either generated by polyploidy or other replication events, representing a unique evolutionary process.

Whole-genome duplication (WGD) or polyploidy contains three or more complete sets of homologous chromosomes (Ramsey and Schemske 2002; Winge 1917), arising through various ways mostly from unreduced gametes of subsequent meiosis (Ahloowalia and Garber 1961; Bretagnolle and Thompson 1995; Harlan and DeWet 1975; Jorgensen 1928; Newton and Pellew 1929; Ramsey and Schemske 1998; Skalińska 1947). Determining the genetic mechanisms in evolutionary biology is the main challenge that helps to regulate diversity, new traits, functions and adaptations. Numerous studies have reported the transformative evolutionary potential of polyploidy (Levin 2002; Soltis et al. 2009; Soltis and Soltis 2000; Stebbins 1950).

Polyploidy is presently predictable as a critical evolutionary power in eukaryotes (Gregory and Mable 2005; Mable et al. 2011; Mable 2003), which provide biodiversity and genetic material for evolution (Levin 2002). Previous studies suggested that vertebrate ancestors had two polyploid events (Panopoulou and Poustka 2005), with successive polyploidy in amphibians and fish (Mable et al. 2011). Current studies of whole genomes have intensely changed the polyploidy paradigm. Genome analysis showed that flowering plants and eukaryotes were the results of genome-wide duplication with significant redundant genomes. For example, whole-genome sequence of *A. thaliana* (genome size (157 Mb) (Bennett et al. 2003) exposed a large number of repetitive genes for two or three rounds of genome replication (Bowers et al. 2003; Vision et al. 2000).

14.3.4   Candidate Gene Approach

The candidate gene method is the best technology for crop improvement and for studies of the genetic structure and designing the complex characteristics (Patocchi et al. 2009). The primary application of genetic crop improvement ranges from genome research to the reduction of a large amount of information of particular genes. The candidate gene method has been applied to plant genetics for a few decades now and is characterized through cloning and different quantitative trait loci (QTL). The candidate gene approach is used for the development of knowledge generated through various functions, structures and other comparative genomics and to determine the appropriate candidate genes that play a vital role in the phenotype of a particular trait. The study of the candidate gene is based on the presumption that the known gene (candidate gene) can correspond to the locus of interest (Pflieger et al. 2001). The most promising candidate is then selected from a large number of supposed candidate genes and gene sequences to verify the linkage with interesting traits. Fine-mapping based on the population allows the exact localization of candidate genes and loci.

After identifying candidate genes, different DNA markers, such as simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers, were developed. Subsequently, these gene-specific markers were mapped, and their positions were likened to the recognition sites of interesting traits. The candidate genes are identified through recognized qualitative or quantitative trait loci, which provide an excellent way to increase the limited sources. Functional analysis determines the induction of specific genes on the traits of interest and uses functional allele-specific markers to identify the changes in the related DNA sequences of functional differences. These markers are also known as "perfect markers" because they reduce the probability of recombination with related genetic markers. Gene-specific markers are typically the same and active and may be valuable in different genera of Rosaceae.

A previous study (Etienne et al. 2002) reported peach candidate genes, to isolate the major loci genes and QTLs for sugar content and acidity. A total of 18 candidate genes were selected, of which 12 were mapped, and a gene involved in solute accumulation was used to locate soluble solid concentrations along with QTL. In strawberries, the yellow fruit colour, associated with the flavanone 3-hydroxylase gene, is the best example of the significant improvement in Rosaceae through the candidate gene pathway (Deng and Davis 2001). The genes involved in cell wall modification and ethylene biosynthesis are Md-ACS1, Md-ACO1 and Md-Exp7, associated with firmness in apple (Costa et al. 2005, 2008; Oraguzie et al. 2004) and PpLDOX (leucoanthocyanidin dioxygenase) associated with browning during cold storage in peach (Ogundiwin et al. 2008). The recognition of particular candidate genes based on previous biological research findings is related to the natural progress of specific enzymes or proteins. The relationship among endogenious PG, fruit softening and the development of related markers are critical examples of biological processes, especially in genetic processes. In some other ways, a particular protein might be related to traits that can be used to identify candidate genes. For example, the

main resistance (R) genes frequently encode the nuclear binding site (NBS), leucine repeat sequence (LRR) protein kinase, and the NBS-LRR resistance gene analogues (RGA) can be used to identify candidate resistance genes (Baldi et al. 2004; Samuelian et al. 2008). The "Resistant gene map", reported by Lalli et al. (2005), describes the complexity of the genome site of this gene sequence in the disease resistance of *Prunus*.

Genome analysis can be used as a significant tool for identifying many genes related to complex traits. Similarly, the transcriptional profile of fire blight-challenged apple leaf tissue to identify 650 expression sequence tags (EST) (Malnoy et al. 2008; Norelli et al. 2009). Different bioinformatics tools were used to identify EST associated with blight. There was no pathogen of fire blight in EST isolated from apple tissues (Baldo et al. 2007). When BLAST was carried out, it showed a significant similarity with the known 2800 *Arabidopsis* genes, which are regulated by the bacterial challenge (Thilmony et al. 2006) or a universally developed resistance. It can also be identified by suppression of cDNA hybridisation (SSH) and cDNA AFLP transcriptional profiling. Until now, 28 candidate gene markers for resistance to fire blight have been screened, of which six are located near or known to be QTLs (Norelli et al. 2008). Since for most of the Rosaceae plants, their whole-genome sequences are available, many related methods can be used to scan the coding regions identified in the potential candidate genes and to further improve the proficiency of the candidate gene approach further. Co-localisation of candidate gene markers and specific loci does not verify the pathogenic role of the gene in specific phenotypes. This is the best genomics approaches for crop improvement of several Rosaceae species and accidental linkages. Therefore, more functional analyses are needed to establish a contributing role.

14.3.5 Proteomics and Metabolomics

Proteomics is the study of the proteome, using high-throughput identification and analysis of proteins and representing a developing field of research with various progressions in recent decades. This study has a high molecular significance for understanding the different protein supplements in the genome (Wilkins et al. 1996). Proteomics is the research of the multi-protein system, mainly focused on the interaction of multiple and different proteins and their role as a part of a more extensive system or network. Recent technologies in molecular biology have expanded the methods of detecting and analysing the whole protein and metabolite arrays in plants, especially in the developmental stage of the transcriptome. Proteomics (Pandey and Mann 2000) and metabolomics (Fiehn 2002) are two major technologies that can be used to better understand all the basic physiological and molecular processes of traits in positive or reverse genetic approaches (Fig. 14.3).

Proteins that play a vital role in plant growth and development are essential for improving crop biotechnology. These proteins retain cell homeostasis in a particular environment by controlling physiological and biochemical pathways. Previous studies have shown that genomics and proteomics are the two most significant technologies to maintain the discovery of new gene rolling, which can finally place in the pipeline of the crop improvement plan. Two-dimensional electrophoresis (2-DE) and mass spectrometry (MS) are two most extensively used proteomics approaches for the classification and identification of proteins in different proteomic states or environments. The progress of 2-DE is beneficial for biological technology, but associated with high labour intensity, insensitivity to low-copy number proteins and low repetition. To characterise the complete proteome, many gel-free proteomics technologies have also become valuable tools for researchers (Baggerman et al. 2005; Jayaraman et al. 2012; Lambert et al. 2005; Scherp et al. 2011).

Rosaceae plants may contain thousands of genes and are minor transcripts. They may also contain thousands of different proteins, containing various enzymes and structures and other metabolites, especially for secondary metabolism. Proteomics and metabolomics have an excessive
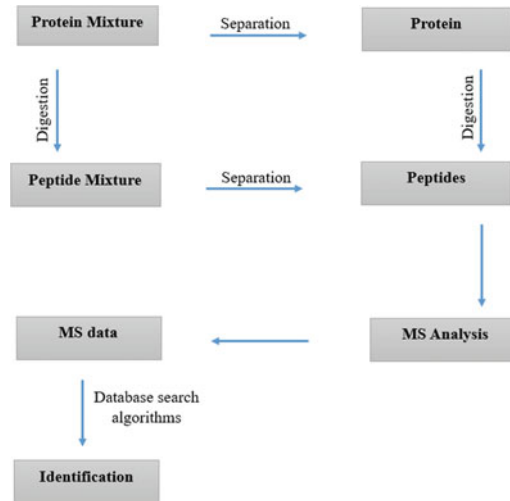
Fig. 14.3 General flowchart for proteomic analysis

potential for explaining biological processes, but new techniques in genomics advancement for crop improvement in the molecular biology of Rosaceae and related toolboxes are still being developed. Many challenges still require the extensive identification of proteins and metabolites (Fridman and Pichersky 2005), in addition to linking the network with specific proteins and metabolites to horticultural traits. Also, although different proteins can easily be associated with their encoding genes, it is problematic to connect a specific metabolite to its potential genetic source (Schauer and Fernie 2006). At present, these technologies are rarely used in Rosaceae. In apricot (Grimplet et al. 2003), proteomics is used to linked the expressed genes with their translated products, while a previous study (Alm et al. 2007) has investigated several proteins in strawberry to study the allergen content. In apple flesh, Guarino et al. (2007) noticed 303 distinctive proteins, of which 44 were identified and linked with 28 different genes. Furthermore, another study (Rudell et al. 2008) noticed more than 200 components in apple peel via metabolic profiling, of which 78 were identified.

Rosaceae plants are highly abundant in specific metabolites, and many of them are important for human health and nutrition. Target analysis has been used to study the interactions and biosynthesis of flavonoids, anthocyanins and phenols in several Rosaceae fruits. Modern improvements in metabolomics allow a comprehensive examination and inquiry of metabolic complexes. For example, a study on the four consecutive stages of strawberry fruit development, based on Fourier transform ion cyclotron mass spectrometry, was performed to investigate the unique information of the metabolic transformation of the immature fruit to the mature fruit (Aharoni and O'connell 2002). A study on the effects of UV irradiation and cold storage on apple fruit metabolism exposed changes in ethylene synthesis, acid metabolism, flavonoid pigment synthesis and fruit structure in primary and secondary pathways (Rudell et al. 2008). The extensive use of metabolomics in the Rosaceae family will increase the improvement of detection of new gene function in primary and secondary metabolism, and will provide a complete data set essential for the metabolic network model, which is related to metabolites promoting human health.

14.4   Conclusions

The perspective of genomics in Rosaceae crop improvement is important for different breeding programmes and for other research purposes to

improve the genetic value of various crops. The development at the genetic level is essential to address the sustainability of agriculture by reducing environmental impacts, minimising the use of land and water and reducing the use of chemicals to attain high-quality products with a good taste. Genomic technologies such as whole-genome sequencing, duplication, polyploidy, mapping techniques, omics studies and reconstruction of the ancestor of Rosaceae chromosomes and the genome are fundamental for Rosaceae crop evolution and improvement.

References

Aharoni A, O'connell AP (2002) Gene expression analysis of strawberry achene and receptacle maturation using DNA microarrays. J Exp Bot 53 (377):2073–2087

Ahloowalia B, Garber E (1961) The genus *Collinsia*. XIII. Cytogenetic studies of interspecific hybrids involving species with pedicled flowers. Bot Gaz 122(3):219–228

Alm R, Ekefjärd A, Krogh M, Häkkinen J, Emanuelsson C (2007) Proteomic variation is as large within as between strawberry varieties. J Proteome Res 6(8):3011–3020

Baggerman G, Vierstraete E, De Loof A, Schoofs L (2005) Gel-based versus gel-free proteomics: a review. Comb Chem High Throughput Screen 8(8):669–677

Baird WV, Estager AS, Wells JK (1994) Estimating nuclear DNA content in peach and related diploid species using laser flow cytometry and DNA hybridization. J Am Soc Hortic Sci 119(6):1312–1316

Baldi P, Patocchi A, Zini E, Toller C, Velasco R, Komjanc M (2004) Cloning and linkage mapping of resistance gene homologues in apple. Theor Appl Genet 109(1):231–239

Baldo A, Bassett C, Malnoy M, Korban S, Gasic K, Farrell R, Aldwinckle H, Norelli J (2007) Computational identification of candidate genes involved in response to fire blight in apples. In: Plant animal genomes XV conference, abstract. http://intl-pag.org/15/abstracts/PAG15P08c

Bennett MD, Leitch IJ, Price HJ, Johnston JS (2003) Comparisons with *Caenorhabditis* (∼100 Mb) and *Drosophila* (∼175 Mb) using flow cytometry show genome size in *Arabidopsis* to be ∼157 Mb and thus ∼25% larger than the *Arabidopsis* genome initiative estimate of ∼125 Mb. Ann Bot 91(5):547–557

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422(6930):433

Bretagnolle F, Thompson JD (1995) Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of autopolyploid plants. New Phytol 129(1):1–22

Bridges CB (1935) Salivary chromosome maps: with a key to the banding of the chromosomes of *Drosophila melanogaster*. J Hered 26(2):60–64

Chen J (1996) Chinese Mei flowers. Hainan Publishing House, Hankou, Hannan, China

Claverie M, Dirlewanger E, Cosson P, Bosselut N, Lecouls A, Voisin R, Kleinhentz M, Lafargue B, Caboche M, Chalhoub B (2004) High-resolution mapping and chromosome landing at the root-knot nematode resistance locus Ma from Myrobalan plum using a large-insert BAC DNA library. Theor Appl Genet 109(6):1318–1327

Costa F, Stella S, Van de Weg WE, Guerra W, Cecchinel M, Dallavia J, Koller B, Sansavini S (2005) Role of the genes Md-ACO1 and Md-ACS1 in ethylene production and shelf life of apple (*Malus domestica* Borkh). Euphytica 141(1–2):181–190

Costa F, Van de Weg W, Stella S, Dondini L, Pratesi D, Musacchi S, Sansavini S (2008) Map position and functional allelic diversity of Md-Exp7, a new putative expansion gene associated with fruit softening in apple (*Malus × domestica* Borkh.) and pear (*Pyrus communis*). Tree Genet Genomes 4(3):575–586

Deng C, Davis T (2001) Molecular identification of the yellow fruit color (c) locus in diploid strawberry: a candidate gene approach. Theor Appl Genet 103(2–3):316–322

Dirlewanger E, Cosson P, Boudehri K, Renaud C, Capdeville G, Tauzin Y, Laigret F, Moing A (2006) Development of a second-generation genetic linkage map for peach [*Prunus persica* (L.) Batsch] and characterization of morphological traits affecting flower and fruit. Tree Genet Genomes 3(1):1–13

Dondini L, Lain O, Geuna F, Banfi R, Gaiotti F, Tartarini S, Bassi D, Testolin R (2007) Development of a new SSR-based linkage map in apricot and analysis of synteny with existing *Prunus* maps. Tree Genet Genomes 3(3):239–249

Etienne C, Rothan C, Moing A, Plomion C, Bodenes C, Svanella-Dumas L, Cosson P, Pronier V, Monet R, Dirlewanger E (2002) Candidate genes and QTLs for sugar and organic acid content in peach [*Prunus persica* (L.) Batsch]. Theor Appl Genet 105(1):145–159

Farrer RA, Kemen E, Jones JD, Studholme DJ (2009) De novo assembly of the *Pseudomonas syringae* pv. syringae B728a genome using Illumina/Solexa short sequence reads. FEMS Microbiol Lett 291(1):103–111

Fiehn O (2002) Metabolomics—the link between genotypes and phenotypes. In: Functional genomics. Springer, pp 155–171

Fridman E, Pichersky E (2005) Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. Curr Opin Plant Biol 8 (3):242–248

Georgi L, Wang Y, Yvergniaux D, Ormsbee T, Inigo M, Reighard G, Abbott A (2002) Construction of a BAC library and its application to the identification of simple sequence repeats in peach [*Prunus persica* (L.) Batsch]. Theor Appl Genet 105(8):1151–1158

Goff SA, Ricke D, Lan T-H, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296 (5565):92–100

Goodman HM, Ecker JR, Dean C (1995) The genome of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 92 (24):10831–10835

Graham J, Smith K, MacKenzie K, Jorgenson L, Hackett C, Powell W (2004) The construction of a genetic linkage map of red raspberry (*Rubus idaeus* subsp. *idaeus*) based on AFLPs, genomic-SSR and EST-SSR markers. Theor Appl Genet 109(4):740–749

Gregory TR, Mable BK (2005) Polyploidy in animals. Gregory TR the evolution of the genome. Elsevier, London, pp 427–517

Grimplet J, Romieu C, Sauvage F, Lambert P, Audergon J, Terrier N (2003) Transcriptomics and proteomics tools towards ripening markers for assisted selection in apricot. In: XI Eucarpia symposium on fruit breeding and genetics 663, pp 291–296

Guarino C, Arena S, De Simone L, D'ambrosio C, Santoro S, Rocco M, Scaloni A, Marra M (2007) Proteomic analysis of the major soluble components in Annurca apple flesh. Mol Nutr Food Res 51(2):255–262

Han Y, Gasic K, Marron B, Beever JE, Korban SS (2007) A BAC-based physical map of the apple genome. Genomics 89(5):630–637

Harlan J, DeWet J (1975) On O. Winge and a prayer: the origins of polyploidy [in the higher plants]. Bot Rev 41(4):361–390

Hemmat M, Weedon N, Manganaris A, Lawson D (1994) Molecular marker linkage map for apple. J Hered 85 (1):4–11

Howad W, Yamamoto T, Dirlewanger E, Testolin R, Cosson P, Cipriani G, Monforte AJ, Georgi L, Abbott AG, Arus P (2005) Mapping with a few plants: using selective mapping for microsatellite saturation of the *Prunus* reference map. Genetics 171(3):1305–1309

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T (2009) High-throughput genotyping by whole-genome resequencing. Genome Res 19(6):1068–1076

Huang C-H, Sun R, Hu Y, Zeng L, Zhang N, Cai L, Zhang Q, Koch MA, Al-Shehbaz I, Edger PP (2015) Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. Mol Biol Evol 33 (2):394–412

Hummer KE, Janick J (2009) Rosaceae: taxonomy, economic importance, genomics. In: Folta KM, Gardiner SE (eds) Genetics and genomics of Rosaceae. Springer, New York, pp 1–17

Hurtado M, Romero C, Vilanova S, Abbott A, Llacer G, Badenes M (2002) Genetic linkage maps of two apricot cultivars (*Prunus armeniaca* L.), and mapping of PPV (sharka) resistance. Theor Appl Genet 105(2–3):182–191

Jackson SA (2016) Rice: the first crop genome. Rice 9 (1):14

Jayaraman D, Forshey KL, Grimsrud PA, Ané J-M (2012) Leveraging proteomics to understand plant–microbe interactions. Front Plant Sci 3:44

Joobeur T, Viruel M, de Vicente LM, Jauregui B, Ballester J, Dettori M, Verde I, Truco M, Messeguer R, Batlle I (1998) Construction of a saturated linkage map for *Prunus* using an almond × peach F2 progeny. Theor Appl Genet 97(7):1034–1041

Jorgensen G (1928) The experimental formation of heteroploid plants in the genus *Solanum*. J Genet 19 (2):133

Kenis K, Keulemans J (2005) Genetic linkage maps of two apple cultivars (*Malus* × *domestica* Borkh.) based on AFLP and microsatellite markers. Mol Breed 15(2):205–219

Lalli D, Decroocq V, Blenda A, Schurdi-Levraud V, Garay L, Le Gall O, Damsteegt V, Reighard G, Abbott A (2005) Identification and mapping of resistance gene analogs (RGAs) in *Prunus*: a resistance map for Prunus. Theor Appl Genet 111 (8):1504–1513

Lambert J-P, Ethier M, Smith JC, Figeys D (2005) Proteomics: from gel based to gel free. Anal Chem 77 (12):3771–3788

Levin DA (2002) The role of chromosomal change in plant evolution. Oxford University Press, London

Mable BK (2003) Breaking down taxonomic barriers in polyploidy research. Trends Plant Sci 8(12):582–590

Mable B, Alexandrou M, Taylor M (2011) Genome duplication in amphibians and fish: an extended synthesis. J Zool 284(3):151–182

Malnoy M, Xu M, Borejsza-Wysocka E, Korban SS, Aldwinckle HS (2008) Two receptor-like genes, Vfa1 and Vfa2, confer resistance to the fungal pathogen V*enturia inaequalis* inciting apple scab disease. Mol Plant Microbe Interact 21(4):448–458

Moran NA, McLaughlin HJ, Sorek R (2009) The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. Science 323(5912):379–382

Morgan DR, Soltis DE, Robertson KR (1994) Systematic and evolutionary implications of rbcL sequence variation in Rosaceae. Am J Bot 81(7):890–903

Morgante M, Salamini F (2003) From plant genomics to breeding practice. Curr Opin Biotechnol 14(2):214–219

Naik S, Hampson C, Gasic K, Bakkeren G, Korban SS (2006) Development and linkage mapping of E-STS and RGA markers for functional gene homologues in apple. Genome 49(8):959–968

Newton WCF, Pellew C (1929) *Primula kewensis* and its derivatives. J Genet 20:405–467

Norelli J, Gardiner S, Malnoy M, Aldwinckle H, Baldo A, Borejsza-Wysocka E, Farrell Jr R, Lalli D, Celton J,

Bassett C (2008) Using functional genomics to develop tools to breed fire blight resistant apple. In: Plant & animal genomes XVI conference, 2008

Norelli JL, Farrell RE, Bassett CL, Baldo AM, Lalli DA, Aldwinckle HS, Wisniewski ME (2009) Rapid transcriptional response of apple to fire blight disease revealed by cDNA suppression subtractive hybridization analysis. Tree Genet Genomes 5(1):27–40

Ogundiwin E, Peace C, Nicolet C, Rashbrook V, Gradziel T, Bliss F, Parfitt D, Crisosto C (2008) Leucoanthocyanidin dioxygenase gene (PpLDOX): a potential functional marker for cold storage browning in peach. Tree Genet Genomes 4(3):543–554

Oraguzie N, Iwanami H, Soejima J, Harada T, Hall A (2004) Inheritance of the Md-ACS1 gene and its relationship to fruit softening in apple (Malus × domestica Borkh.). Theor Appl Genet 108(8):1526–1533

Pandey A, Mann M (2000) Proteomics to study genes and genomes. Nature 405(6788):6837

Panopoulou G, Poustka AJ (2005) Timing and mechanism of ancient vertebrate genome duplications—the adventure of a hypothesis. Trends Genet 21(10):559–567

Patocchi A, Fernández-Fernández F, Evans K, Gobbin D, Rezzonico F, Boudichevskaia A, Dunemann F, Stankiewicz-Kosyl M, Mathis-Jeanneteau F, Durel C (2009) Development and test of 21 multiplex PCRs composed of SSRs spanning most of the apple genome. Tree Genet Genomes 5(1):211–223

Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. Mol Breeding 7(4):275–291

Phipps J (2014) Flora of North America North of Mexico, vol. 9, Magnoliophyta: Picramniaceae to Rosaceae. Oxford University Press, New York and Oxford

Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan Y-L, Kelly PD, Chu F, Huang H, Hill-Force A, Talbot WS (2000) Zebrafish comparative genomics and the origins of vertebrate chromosomes. Genome Res 10 (12):1890–1902

Potter D, Gao F, Bortiri PE, Oh S-H, Baggett S (2002) Phylogenetic relationships in Rosaceae inferred from chloroplast matK and trnL-trnF nucleotide sequence data. Plant Syst Evol 231(1–4):77–89

Potter D, Eriksson T, Evans RC, Oh S, Smedmark J, Morgan DR, Kerr M, Robertson KR, Arsenault M, Dickinson TA (2007) Phylogeny and classification of Rosaceae. Plant Syst Evol 266(1–2):5–43

Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu Rev Ecol Syst 29(1):467–501

Ramsey J, Schemske DW (2002) Neopolyploidy in flowering plants. Annu Rev Ecol Syst 33(1):589–639

Ridley HN (1930) The dispersal of plants throughout the world. L. Reeve & Company, Reeve, Ashford, England

Robertson KR, Phipps JB, Rohrer JR, Smith PG (1991) A synopsis of genera in Maloideae (Rosaceae). Syst Bot 16(2):376–394

Rudell DR, Mattheis JP, Curry EA (2008) Prestorage ultraviolet-white light irradiation alters apple peel metabolome. J Agric Food Chem 56(3):1138–1147

Samuelian SK, Baldo AM, Pattison JA, Weber CA (2008) Isolation and linkage mapping of NBS-LRR resistance gene analogs in red raspberry (Rubus idaeus L.) and classification among 270 Rosaceae NBS-LRR genes. Tree Genet Genomes 4(4):881–896

Schauer N, Fernie AR (2006) Plant metabolomics: towards biological function and mechanism. Trends Plant Sci 11(10):508–516

Scherp P, Ku G, Coleman L, Kheterpal I (2011) Gel-based and gel-free proteomic technologies. In: Gimble J, Bunnell B (eds) Adipose-derived stem cells. Springer, Totowa, NJ, pp 163–190

Seymour GB, Østergaard L, Chapman NH, Knapp S, Martin C (2013) Fruit development and ripening. Annu Rev Plant Biol 64:219–241

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP (2011) The genome of woodland strawberry (Fragaria vesca). Nat Genet 43(2):109

Skalińska M (1947) Polyploidy in Valeriana officinalis Linn. in relation to its ecology and distribution. J Linn Soc Lond Bot 53(350):159–186

Soltis PS, Soltis DE (2000) The role of genetic and genomic attributes in the success of polyploids. Proc Natl Acad Sci USA 97(13):7051–7057

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, de Pamphilis CW, Wall PK, Soltis PS (2009) Polyploidy and angiosperm diversification. Am J Bot 96(1):336–348

Stebbins G (1950) Variation and evolution in plants. Columbia University Press, New York

Sturtevant AH (1913) The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J Exp Zool 14(1):43–59

Tanksley S, Young N, Paterson A, Bonierbale M (1989) RFLP mapping in plant breeding: new tools for an old science. Nat Biotechnol 7(3):257

Thilmony R, Underwood W, He SY (2006) Genome-wide transcriptional analysis of the Arabidopsis thaliana interaction with the plant pathogen Pseudomonas syringae pv. tomato DC3000 and the human pathogen Escherichia coli O157: H7. Plant J 46(1):34–53

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D (2010) The genome of the domesticated apple (Malus × domestica Borkh.). Nat Genet 42(10):833

Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F (2013) The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet 45(5):487

Vilanova S, Romero C, Abernathy D, Abbott A, Burgos L, Llacer G, Badenes M (2003) Construction and application of a bacterial artificial chromosome (BAC)

library of *Prunus armeniaca* L. for the identification of clones linked to the self-incompatibility locus. Mol Genet Genome 269(5):685–691

Vinatzer B, Zhang H-B, Sansavini S (1998) Construction and characterization of a bacterial artificial chromosome library of apple. Theor Appl Genet 97(7):1183–1190

Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in *Arabidopsis*. Science 290 (5499):2114–2117

Weigel D, Mott R (2009) The 1001 genomes project for *Arabidopsis thaliana*. Genome Biol 10(5):107

Wilkins MR, Sanchez J-C, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF, Williams KL (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol Genet Eng Rev 13(1):19–50

Winge Ö (1917) The chromosome. Their numbers and general importance. CR Trav Lab Carlsberg 13:131–175

Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387(6634):6708

Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, Khan M, Tao S, Korban S, Wang H (2013) The genome of the pear (*Pyrus bretschneideri* Rehd.). Genome Res 23:396–408 (Crossref. Medline, Google Scholar)

Xu M, Song J, Cheng Z, Jiang J, Korban SS (2001) A bacterial artificial chromosome (BAC) library of *Malus floribunda* 821 and contig construction for positional cloning of the apple scab resistance gene Vf. Genome 44(6):1104–1113

Yamamoto T, Kimura T, Saito T, Kotobuki K, Matsuta N, Liebhard R, Gessler C, Van de Weg W, Hayashi T (2004) Genetic linkage maps of Japanese and European pears aligned to the apple consensus map. Acta Hortic 663:51–56

Yamamoto T, Yamaguchi M, Hayashi T (2005) An integrated genetic linkage map of peach by SSR, STS, AFLP and RAPD. J Jpn Soc Hortic Sci 74(3):204–213

Yang Y, Moore MJ, Brockington SF, Soltis DE, Wong GK-S, Carpenter EJ, Zhang Y, Chen L, Yan Z, Xie Y (2015) Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. Mol Biol Evol 32(8):2001–2014

Zeng L, Zhang Q, Sun R, Kong H, Zhang N, Ma H (2014) Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. Nat Commun 5:4956

Zhang N, Zeng L, Shan H, Ma H (2012a) Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. New Phytol 195(4):923–937

Zhang Q, Chen W, Sun L, Zhao F, Huang B, Yang W, Tao Y, Wang J, Yuan Z, Fan G (2012b) The genome of *Prunus mume*. Nat Commun 3:1318

Zhebentyayeva T, Horn R, Mook J, Lecouls A, Georgi L, Abbott A, Reighard G, Swire-Clark G, Baird W (2006) Physical framework for the peach genome. Acta Hortic 713:83–88

Zimmer EA, Wen J (2015) Using nuclear gene data for plant phylogenetics: progress and prospects II. Next-gen approaches. J Syst Evol 53(5):371–379