An investigation of genetic factors impacting sterol profiles in Australian olive oils

Submitted in fulfilment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

by

DEBASHREE LALA RAY

Masters (Biotechnology)

College of Engineering and Science

Victoria University

AUSTRALIA

March 2014

ABSTRACT

Analyses of phytosterol levels in Australian olive oils derived from the cultivar Barnea (*Olea europaea* L. cv. Barnea)reveal up to 4.8% campesterol, exceeding the IOOC standards that stipulate a level of <4%. It has been observed that sterol profiles appear to remain relatively consistent within individual cultivars, strongly implicating genetic factors as the cause of these different levels. The plant sterol biosynthetic pathway contains a bifurcation that leads to the formation of β-sitosterol or campesterol, with the flux controlled by the activity of two enzymes, SAM-24-methylene-lophenol-C-24-methyltransferase2 (SMT2) and C-4α-sterol-methyl-oxidase2 (SMO2). Thus, it is conceivable that the relative activity/expression of these enzymes could play a pivotal role in determining the relative amounts of β-sitosterol and campesterol in Australian olive oils. However, little is known about these enzymes in olives, nor the effect that many agronomic and/or processing practices have on the relative levels of these sterols in olive oil.

To fill this gap, this research describes the analysis of sterol levels in olive oils extracted from fruits grown, harvested and processed under various conditions. In addition, the coding sequences of *SMT2* and *SMO2* cDNAs and their expression levels during fruit development in olive cultivars Barnea, Frantoio and Picual were also investigated.

The analysis showed that no evaluated management or processing practices seem to have contributed in reducing the campesterol levels in commonly cultivated Australian olive oils. The *SMT2* gene in oliveappears to contain two loci; however no significant allelic and / or expression level differences were detected between the cultivars tested. The *SMO2* gene in olivealso appears to contain two loci, however results suggested the presence of allelic differences between the cultivars. qPCR analysis revealed that as *SMO2* expression in the fruit increases, the magnitude of this change does appear to vary between olive cultivars, in particular between Barnea and Picual.

DECLARATION

I, Debashree Lala Ray, declare that the PhD thesis entitled "An investigation of genetic factors impacting sterol profiles in Australian olive oils' is no more than 100,000 words in length, exclusive of tables, figures, appendices, references and footnotes. This thesis contains no material that hasbeen submitted previously, in whole or in part, for the award of any other academicdegree or diploma. Except where otherwise indicated, this thesis is my own work.

Debashree Lala Ray
March 2014

ACKNOWLEDGEMENTS

My sincere thanks and gratitude goes to my principal supervisor, Dr. Joshua Johnson, whose continuous advice and constant encouragement throughout the course of this thesis, as well as his great knowledge of molecular biology, have been invaluable to me. He has been instrumental in the completion of this work, by always being available for systematic guidance and providing insightful feedback whenever required, and was a great support throughout. Thank you for believing in me and being so understanding at all times.

I would like to sincerely thank my co-supervisor Dr. Sandra McKechnie, for proofreading my thesis and always being available for quick chats.

I would also like to thank Josh and Sandra for the wonderful teaching experience I had over the last four years alongside my PhD, which was not only a helpful diversion from my PhD work but had also provided me with the much needed financial support.

Special thanks go to Dr. Sarah Fraser and Dr. Thomas Yeager for their warm help and willingness to give their time for morale-boosting chats during my studies.

I would like to express a special “thank you” to Dr. Polydefkis Hatzopoulos, from Agricultural University of Athens, Greece, though I have never met him, he had always beenwilling to discuss my lab problems and providedme with lab protocols, relevant papers and valuable advice on few aspects of this project, which has been a critical turning point in my work.

My thanks to the Rural Industries Research and Development Corporation (RIRDC), for providing me with financial support throughout this project through an Industry Research Scholarship. I would like to thank Claudia Guillaume and Leandro Ravetti at Modern Olives, Lara, Victoria, for facilitating my research at theselaboratories for the first six months during my project and training me in sterol extraction and gas chromatography techniques.

Many thanks go to all the technical staff at Victoria University, especially Stacey, Dale, Joe, Ivan, Charmaine, Min and Mary for their assistance and friendship through this project.

I owe much to my two good old friends, Sudinna and Shruti, my very first friends in Australia, thank you for helping me get through the hardest times and for being by my side.

I must also thank my friends at Victoria University for being there and supporting me with friendly advice, cups of tea and random conversations which kept me sane throughout, in particular Zeinab, Warren, Lata, Rabia, Kasup, Asalangika and Nuwan.

Special thanks must go to my parents-in-law for their generosity and unconditional support throughout my studies. It is because of them, I never had to worry about family or home while I was deeply buried in my thesis work.

Much love and thanks must go to my parents, brother and sister-in-law, for not only supporting me on the good days but encouraging me to keep going on the dark days. To my Dad, the person responsible for my love of science and to my mum, who always believed in me and my ability and encouraged me to be the best that I could be, thank you.

To Goldie, my adorable cocker spaniel, I will now take you out on lots of new exciting walks, I promise!

Finally, I wish to thank my husband Ray and little son Arian, who have stood by me during this very long process, put up with my odd hours and offered me much needed emotional support. I could have never completed this work without the two of you by my side.

This work is dedicated to my parents.

LIST OF PUBLICATIONS AND PRESENTATIONS

Refereed Journal Articles

Guillaume, C., Ravetti, L., Ray, D. L. and Johnson, J. (2011). Technological factors affecting sterols in Australian olive oils. Journal of the American Oil Chemists' Society **89**(1): 29-39.

Ray, D. L. and Johnson, J. The role of sterol methyl transferase-2 (SMT2) and sterol methyl oxidase-2 (SMO2)on phytosterol biosynthesis and function in plants. Review paper (manuscript in preparation).

Ray, D. L. and Johnson, J. The molecular characterization and expression analysis of the *sterol methyl transferase-2* gene families in *Olea europaea*L.(manuscript in preparation).

Ray, D. L. and Johnson, J. The molecular characterization and expression analysis of the *sterol methyl oxidase-2* gene families in *Olea europaea*L.(manuscript in preparation).

Ray, D. L. and Johnson, J.Identification and evaluation of reference genes for gene expression normalization in *Olea europaea*L.mesocarp tissues by quantitative real-time RT-PCR(manuscript in preparation).

Conference Poster Presentations

Ray, D. L. and Johnson, J. (2009). An investigation of genetic factors impacting sterol profiles in Australian olive oils. Presented at the 2009 FoHES Postgraduate Research Conference, Victoria University, Melbourne, Australia.

Ray, D. L. and Johnson, J. (2011). The characterization of the *SMT2* gene family in *Olea europaea*L.Presented at the Joint Annual Conference for the Australian Society of Biochemistry and Molecular Biology (ASBMB), the Australian Society of Plant Scientists (ASPS) and the Australia and New Zealand Society for Cell and Developmental Biology (ANZSCDB), Cairns, Australia.

Ray, D. L. and Johnson, J.(2012). The characterization of the *SMT2* gene family in *Olea europaea*L.Presented at the 2012 FoHES Postgraduate Research Conference, Victoria University, Melbourne, Australia.

Ray, D. L. and Johnson, J. (2012). The characterization of the *SMO2* gene family in *Olea europaea*L.Presented at the 2012 FoHES Postgraduate Research Conference, Victoria University, Melbourne, Australia.

LIST OF ABBREVIATIONS

AOA	Australian Olive Association
AOCS	American Oil Chemists' Society
AORL	Australian Oils Research Laboratory
AS-PCR	allele-specific PCR
BL	brassinolide
bp	base pairs
BR	brassinosteroid
BSA	bovine serum albumin
cDNA	complementary DNA
cds	coding sequence
CIP	calf intestinal phosphatase
CTAB	cetyltrimethylammonium bromide
CV	coefficient of variance
DAF	days after flowering
dATP	adenosine deoxyribonucleoside triphosphate
dCTP	cytosine deoxyribonucleoside triphosphate
dGTP	guanosine deoxyribonucleoside triphosphate
DEPC	diethyl pyrocarbonate
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	equimolar mixture of dATP, dCTP, dGTP and dTTP
DTT	dithiothreitol
DPI	Department of Primary Industries
EDTA	ethylenediaminotetra-acetic acid
EST	expressed sequence tag
EVOO	extra virgin olive oil
E+U	Erythrodiol and Uvaol
EU	European Union
FAO	Food and Agriculture Organization of the United Nations

FFA	free fatty acids
g	gravitational force
GC	Gas chromatography
gDNA	genomic DNA
GSP	gene specific primers
HPLC	High-performance liquid chromatography
HR	hazard ratio
IOC/COI	International Olive Council
IOOC	International Olive Oil Council
IRC	inter-run calibrators
kb	kilobase pairs
LB	Luria-Bertani broth
LMW	low molecular weight
MOLS	Modern Olives Laboratory Services
mRNA	messenger RNA
MUFA	mono unsaturated fatty acids
NATA	National Association of Testing Authorities
NGS	next generation sequencing
NRQ	normalised relative quantities
NSW	New South Wales
OR	odds ratio
ORF	open reading frame
OS	2,3-oxidosqualene
PCR	polymerase chain reaction
PUFA	poly unsaturated fatty acids
qPCR	quantitative polymerase chain reaction
RACE	rapid amplification of cDNA ends
RIN	RNA integrity number
RIRDC	Rural Industries Research and Development Corporation
RNA	ribonucleic acid
RNase	ribonuclease
ROO	refined olive oil
RQ	relative quantities
RT	reverse transcription

SE	standard error
SD	standard deviation
SDS	sodium dodecyl sulfate
SMT1	cycloartenol-C-24-methyltransferase
SMT2	SAM-24-methylene-lophenol-C-24-methyltransferase
SMO2	C-4α-sterol-methyl oxidase
SNP	single nucleotide polymorphism
TAP	tobacco acid pyrophosphatase
TE	tris-EDTA buffer
TLC	thin layer chromatography
U	unit
USDA	United States Department of Agriculture
UTR	untranslated region
UV	ultraviolet
VIGS	virus induced gene silencing
V	volume

TABLE OF CONTENTS

CHAPTER 4 MOLECULAR CHARACTERISATION OF THE *STEROL METHYL TRANSFERASE-2 (SMT2)* GENE FAMILY OF OLIVES

CHAPTER 7 CONCLUSIONS AND FUTURE DIRECTIONS

LIST OF FIGURES

LIST OF TABLES

CHAPTER1

GENERAL INTRODUCTION AND LITERATURE REVIEW

Page

List of Figures and Tables in Chapter 1

Chapter 1General Introduction and Literature Review

ABSTRACT

Olive (*Olea europaea* L.) oil, a staple food for thousands of years in the Mediterranean region, has grown in popularity throughout the world, including Australia, at least partially due to its reported health benefits, some of which can be attributed to its phytosterol content. Due to its unique organoleptic properties, the production of extra virgin olive oil demands a premium price which has led to frequent adulteration of the oil with cheaper vegetable oils and refined or processed olive oils. To minimize such fraudulent cases of providing inferior quality of olive oil, the IOOC (International Olive Oil Council) has imposed strict standards that olive oils have to meet in order to be accepted as authentic on the international export market. Various analytical criteria, for quality, sensory and purity evaluation of each grade of olive oil, have been developed by the IOOC. Compositional analysis of the individual sterols in olive oil (apparent β-sitosterol, campesterol, stigmasterol, brassicasterol and Δ5-avenasterol) in conjunction with the total sterol content are considered very effective in assessing the degree of purity of the oil and the absence of other plant oils.These strict IOOC standards have led to a problem for Australian olive oil producers as analyses have shown that some Australian olive oils do not meet these international standards for the total content of sterols. It has been observed that Australian olive oil, specifically those derived from the cultivar Barnea (representing 40% of the olive crop in Australia), contain up to 4.8% campesterol, which exceeds the IOOC standards that stipulates a campesterol level of less than 4%. Analyses of the sterol content of theBarnea cultivar, on a year to year basis, have shown fluctuations in total and relative sterol levels, however campesterol levels have remained consistently high, strongly implicating genetic factors as the cause of these elevated levels. Characterization of the sterol biosynthetic pathway in plants has revealed that there is a key bifurcation in the pathway leading to the formation of β-sitosterol or, alternatively, campesterol, with the flux controlled by the activity of two branchpoint enzymes, SMT2(SAM-24-methylene-lophenol-C-24-methyltransferase) and SMO2 (Sterol-4α-methyl-oxidase). Functional characterisation of these two proteins and the genes encoding them have been undertaken in several plants, fungi and bacteria species. The role of these enzymes in determining the relative levels of sitosterol and campesterol has been supported by direct experimental evidence, showing that altering their levels leads to changes in the ratio of these sterols in

Arabidopsis thaliana and in *Nicotiana tabacum*. However, to date little is known about the genes encoding these enzymes in olives. Therefore, this study aimed to isolate and characterise the *SMT2* and *SMO2* gene families from olive, focussing on cultivars cultivated in Australia and known to exhibit a range of sterol content profiles, namely, Barnea, Frantoio and Picual. This was undertaken in orderto identify any differences in the alleles encoding these enzymes or the expressionin these genes between cultivars that may contribute to the observed differences in the sterol profiles of the olive oils derived from these cultivars.

This chapter describes the necessary background information on olive oil production and composition, adulteration of olive oil, IOOC standards for quality evaluation of olive oil, phytosterol content in olive oil and the non-compliance of Australian olive oils with international standards. Our current knowledge of the SMT2 and SMO2 enzymes, the genes that encode them and the evidence that suggests these enzymes play pivotal roles in determining campesterol and β-sitosterol levels in olive oils have also been reviewed.

1.1 OLIVE CULTIVATION AND OLIVE OIL PRODUCTION

1.1.1 Cultivation of olives

The olive family (Oleaceae), which represents one of the oldest agricultural tree crops, have been cultivated since ancient times, comprises of about 30 genera and 600 species(Foxhall 2007; Preedy 2010).According to the *Olea* database (http://www.oleadb.it/), there are about 1250 olive cultivars dispersed over 54 countries. Olive plants can propagate vegetatively and can survive for several centuries and retain their genetic characteristics for thousands of years (Preedy 2010).

According to classical taxonomy, olives have been classified as the following:

Kingdom: Green plantae

SubKingdom: Tracheobionata - vascular plants

Superdivision: Spermatophyta - seed plants

Division: Magnoliophyta - flowering plants

Class: Magnoliopsida - dicotyledons

Subclass: Asteridae

Order: Scrophulariales or Lamiales

Family: Oleaceae

Genus: *Olea*
Species: *europaea*

Apart from the genus*Olea*, the Oleacaea family contains other genera which are mainly valued as ornamentals, such as *Ligustrum* (Privet), *Syringa* (lilac) and*Fraxinus* (ash)(Therios 2008). The genus *Olea* L. itself comprises of more than 30 species, out of which only one species,*O. europaea*L. produces edible olives and oil commonly consumed. The *O. europaea* L. exist in two forms, one as wild olives or oleaster form (*Olea europaea*L.subsp. *europaea*var.*sylvestris*) and the other as cultivated olives (*Olea europaea*L.subsp. *europaea*var.*europaea*). Other subspecies of the *Olea*

*europaea*L.complex are *laperrinei*, *cuspidata*, *guanchica*, *maroccana* and *cerasiformis*.

Olive trees are tolerant to a broad range of environmental pressures including drought and high salinity (Preedy 2010). The cultivation of olives began in the Mediterranean basin approximately more than 3000 years ago (Preedy 2010) and have been consistently cultivated and used in that part of the world, however an increased awareness of the reported health benefits of olive oil over other fats and oils (Section 1.3) has made olive oil consumption increase (International Olive Oil Council 2010; Preedy 2010) globally and cultivation of olives has disseminated into new geographic areas such as Australia, North America and South America (Preedy 2010).

1.1.2 World olive oil production

Due to the growing appreciation of extra virgin olive oil (EVOO) as a healthier alternative to other vegetable oils and technological advances that have made the production of olives more cost effective, the olive industry has seen a rapid increase in the size of production (International Olive Oil Council 2010).

The European Union [EU] is the world's largest olive oil producers, accounting for 67% of global olive oil production where Spain (45%), Italy (31%) and Greece (22%) are the chief producers, while other countries with considerable production figures are Turkey, Egypt, Syria and Morocco (Figure 1.1) (Darnet *et al.* 2004; International Olive Oil Council 2010).

The vast majority (~90%) of olives produced globally are consumed within their native markets(Figure 1.2). It has been observed that over the last 10 crop years the consumption of olives has increased by 6%(International Olive Oil Council 2010).

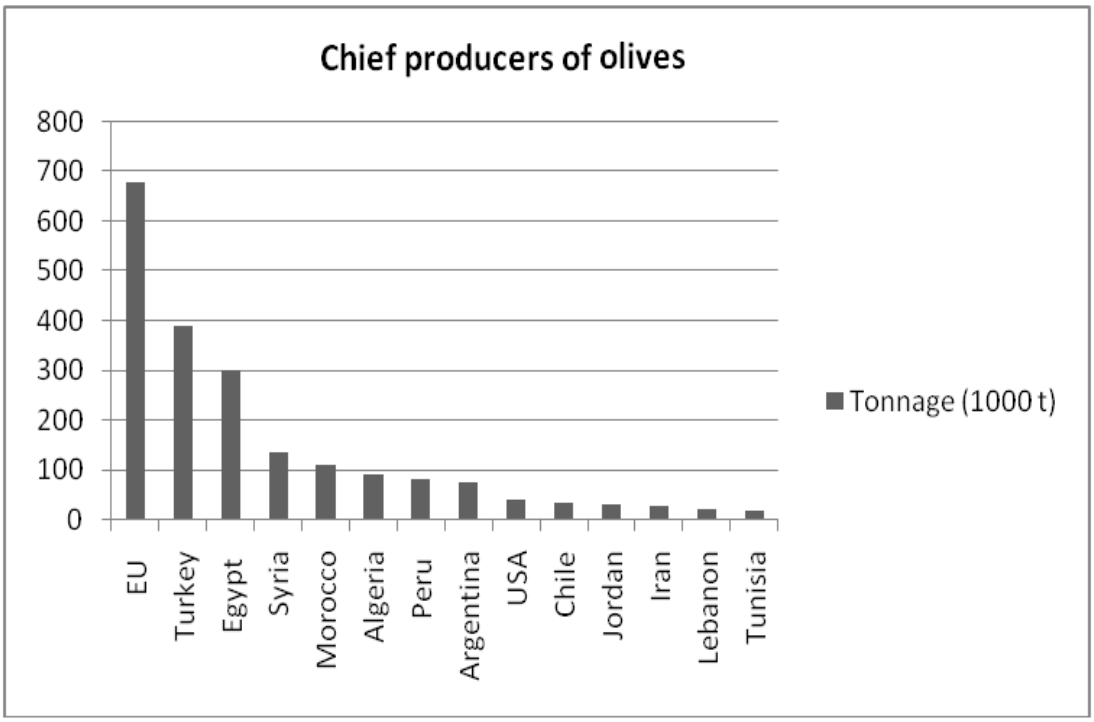


Figure 1.1 Chief olive producers in 2009/2010 crop year (International Olive Council 2010)

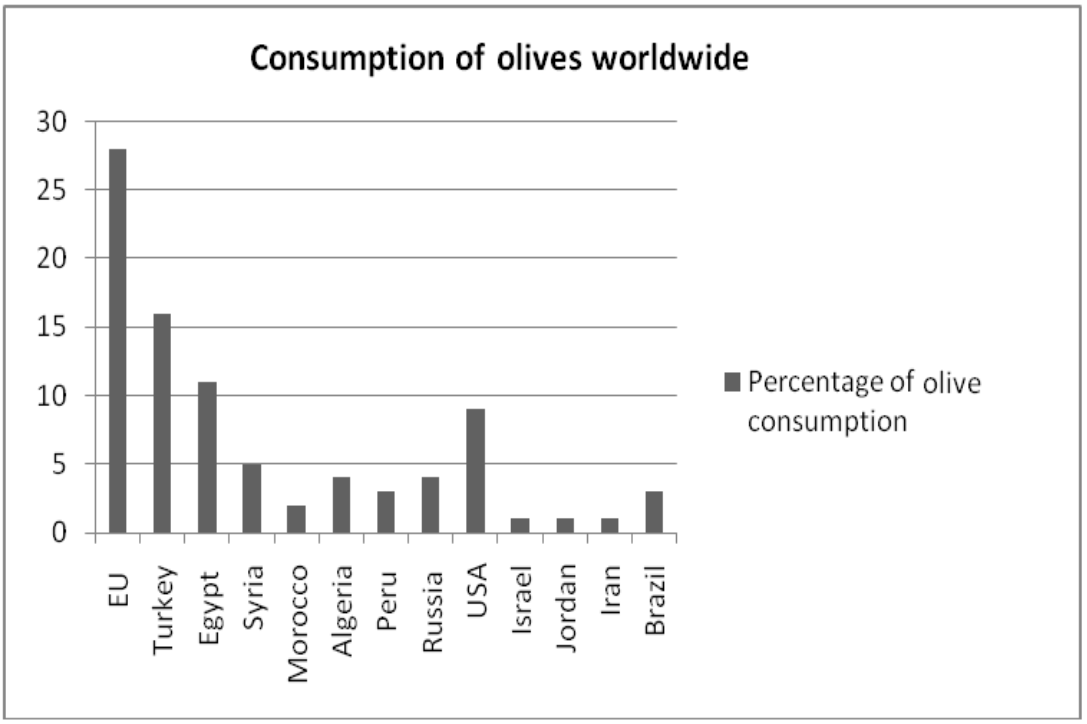


Figure 1.2 World trend of olive consumption (2000-2010) (International Olive Council 2010)

The consumption of olive oil worldwide has grown slowly from 2.59 million tonnes in 2000 to 2.84 million tonnes in 2010 (Rural Industries Reseach and Developmental Corporation 2010). There has been a decline in the per capita consumption of olive oil in the major olive producing and consuming nations over the last decade for reasons that are still unclear, however this decline has been partially compensated by increases in consumption by western nations such as USA, UK, Germany and Australia (Rural Industries Reseach and Developmental Corporation 2010).

The international trading of olive oil and olive pomace oil between the months of November 2009 and March 2010 in the six leading importing countries have been summarised by the IOOC (International Olive Oil Council 2010) (Table 1.1) which showed a total increase in its imports by 15, 283.2 tonnes (+3%) during that period.

Country	Increase in olive oil imports (%)
Australia	52
Japan	40
Canada	35
Brazil	24
USA	2
EU/27 (intra+extra EU)	0.5

Table 1.1 Increase in olive oil imports between November-2009 and March-2010

The crop season 2011/2012 has been a record season with all-time highs in production (3, 370, 000 tonnes), consumption (3, 112, 000 tonnes) and exports (802, 500 tonnes) (International olive oil council, 2013). The IOOC Member countries produced a total of 3, 296, 000 tonnes, of which 2, 444, 500 tonnes are produced by the EU countries, while the rest of the producing countries produced around 80, 000 tonnes. In terms of olive oil consumption during this crop season, the 27 Member States of the EU

consumed 1, 900, 000 tonnes, where the other IOOC member countries is provisionally assessed at 606 500 tonnes, andthat of the non-IOOC countries at 668, 000 tonnes. In terms on export, the IOOC Members accounted for 96% of total world exports (802, 500 tonnes), of which 67% (560, 000 tonnes) was by 27 Member States of the EU(International olive oil council, 2013).

1.1.3 Australian olive oil industry

In comparison to the world olive oil production, Australia is a minor and relatively new commercial sized producer of olive oil. The development of the Australian olive industry has been slow with the emergence of this industry around the mid-1990s which triggered its rapid expansion from a cottage industry to a commercial industry (Ravetti *et al.* 2010). Though small by world standards, the Australian olive industry today is a technically sophisticated industry that is increasingly export focused. The industry utilizes modern production methods and latest horticultural practices characterized by highly mechanized, irrigated olive groves and achieves extraction rates and olive yields equal to or better than its competitor countries (Niaounkis *et al.* 1992; Parker *et al.* 2004; Olson 2006; Sheppard 2008; Krichene *et al.* 2009; Frankel 2010).

Australia enjoys a classic Mediterranean climate and latitude suited to olive production, with dry hot summers and cold dry winters and large tracts of available land with adequate water supply (Mailer 2007). Australia also has relatively good public infrastructure, good storage facilities and few pests and diseases, all of which has contributed in the rapid and sustainable growth of the industry. About 95% of the olive oil produced in Australia is of the extra virgin grade, which does not involve the use of chemicals or refining during extraction and thus considered a high quality product (Section 1.6) (Hearn 2002; Mailer *et al.* 2008). The relevant olive industries located in Australia are the temperate South Eastern, South Western and Eastern Seaboard regions of the continent, of which north-central Victoria and north of Perth in Western Australia contributes over an estimated 70% of Australia's current olive oil production. There are no accurate statistics of the total number of olive groves and trees in Australia but according to current industry estimates there are around 10

million trees grown on over 800 groves covering more than 30,000 hectares (Preedy 2010). A wide range of olive cultivars are grown in Australia which includes Barnea, Frantoio, Picual, Manzanillo, Leccino, Coratina, and Arbequina, (Figure 1.3).

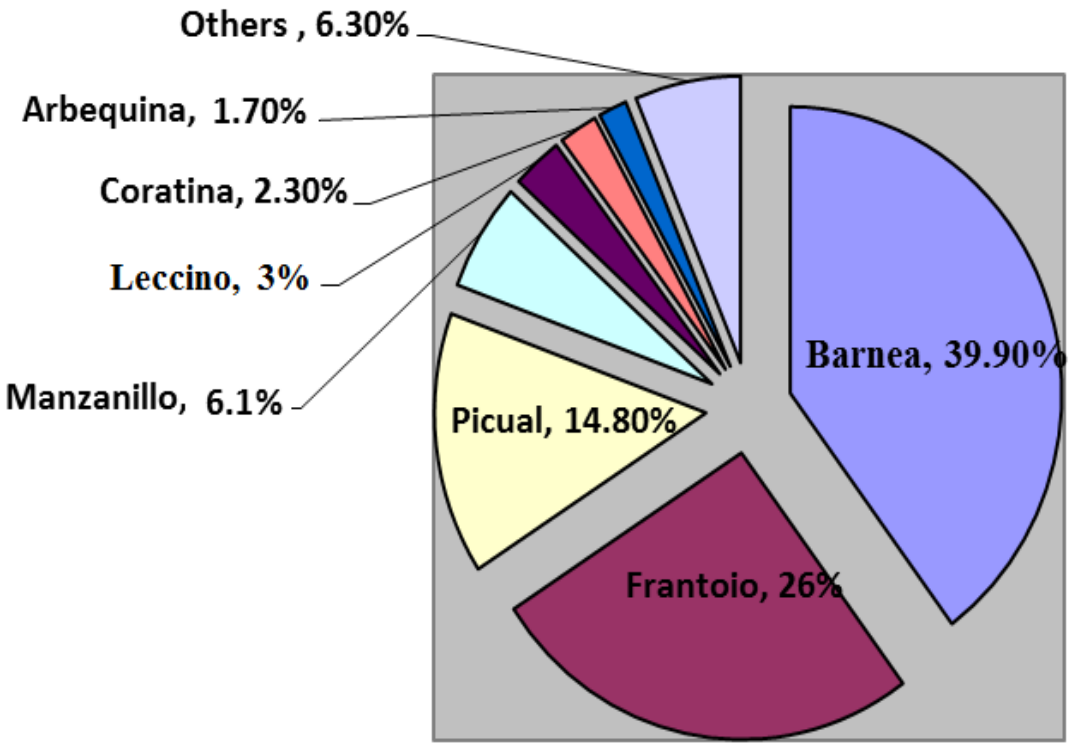


Figure 1.3 Australian olive cultivar distribution in the crop year 2008 (Ravetti *et al.* 2010)

There has been a rapid growth in the Australian olive industry from 2,500 tonnes in 2004 to around 15,000 tonnes in 2009 with an estimated retail value of over AU$185 million (Figure 1.4). From 2004-2009 the annual production of olive oil increased by 47% on average. This is likely due to increased awareness among the population about the health benefits of extra virgin olive oil (Section 1.3) and a greater access to information regarding how olive oil can be used and also the availability of tax incentives for Australian olive growers (Sheppard 2008). Due to an increase in the production rate of Australian olive oil, a significant level (~25%) of import replacement has been attained during the 2001-2009 period (Rural Industries Reseach and Developmental Corporation 2010).

Thanks to the excellence in freshness, taste and quality of Australian olive oils, the export volume of extra virgin olive oil has increased significantly from 501 tonnes in 2004 to 6,959 tonnes in 2009, showing an average annual increase of 85% (Rural Industries Reseach and Developmental Corporation 2010).

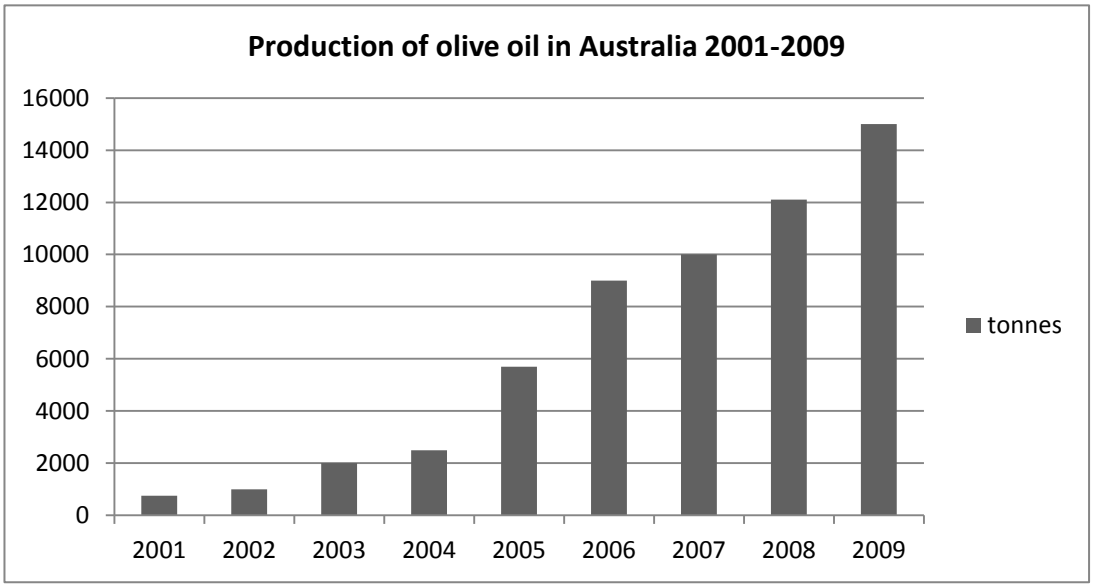


Figure 1.4 Production of olive oil in Australia: 2001-2009 (Ravetti, Guillaume et al. 2010)

In order to maintain and support the development of the industry on a national basis and to ensure the production of high quality Australian-produced olive products, the Australian Olive Association Ltd (AOA) was established in 1995 (http://www.australianolives.com.au/)(Mailer *et al.* 2008) and the growth of the Australian olive industry over the last decade has been supported by ongoing research and development (Hearn 2002; Mailer *et al.* 2008). With the help of research, the industry has established optimal harvest times for olives and identified olive cultivars that produce high quality olive oil suitable to the Australian environment (O' Brien 2008).

1.2 COMPOSITION OF OLIVE OIL

1.2.1 Structure of olive fruit

The olive fruit is a fleshy oval-shaped drupe which consists of a pericarp (skin or flesh) and endocarp (kernel, pit) (Figure 1.5).

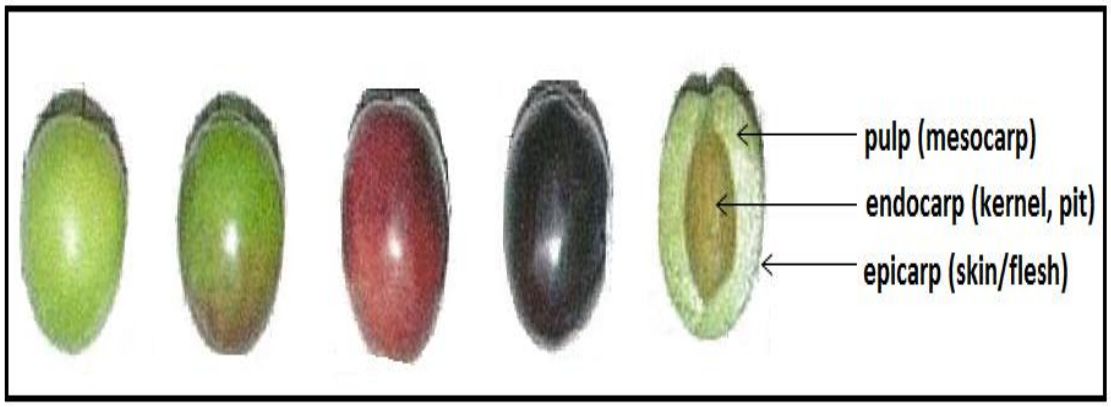


Figure 1.5 Structure of olive fruit at different stages of ripeness

The pericarp which accounts for about 65-83% of the total weight has two parts: the epicarp (skin) and the mesocarp (flesh, pulp). The epicarp is a protective tissue which is covered by a layer of wax which remains green throughout the growth phase due to the presence of chlorophyll, and later on changes to straw yellow, purple and black due to varying concentrations of chlorophyll, anthocyanins and carotenoids (Halvadakis *et al.* 2006; Therios 2008). The mesocarp which is the edible portion of the fruit, has low sugar content and high oil content which varies depending on fruit ripeness and cultivar(Boskou 2006; Therios 2008). The endocarp (stone or pit) encloses the olive seed kernel and is made up of fibrous lignin. It accounts for 3% of fruit weight and contains ~2-4% of total fruit oil (Boskou 2006). The average composition of the olive fruit is water (also called vegetation water) (50%), oil (22%), sugar (19.1%), cellulose (5.8%), proteins (1.6%) and minerals (1.5%) (Boskou 2006).

1.2.2 Olive oil extraction

Around 96-98% of the oil is concentrated in the pericarp region of the olive fruit, mainly in vacuoles within the mesocarp cells (Boskou 2006). Olive oil is separated from other phases in the olive paste (solid material and vegetative extract liquid) by purely mechanical means such as pressure, centrifugation and percolation.

Pressure: Initially, olive leaves are removed from the fruit by suction and then by circulating water to remove all impurities. When pressing is used, the olive paste is pressed by pushing the stacked load under the head of the press frame. This pressure between the synthetic fiber mats decreases the olive paste in volume and separates the solid phase from the liquid phase. The liquid phase contains the oily must which on further centrifugation separates into oil and vegetation water (Boskou 2006).

Centrifugation: In centrifugation, after washing and de-leafing the olives, the separation between the solid and liquid phases is performed by diluting the olive paste with water (Boskou 2006). The oil is then extracted from the paste using high speed rotating machines by direct continuous centrifugation which exploits the difference between specific weights of the immiscible liquids and solid matter.

Percolation: During percolation, rows of metal discs or plates are plunged into the olive paste, which when withdrawn has oil dripping from it and thus this process is called percolation or drippage. The steel plate is coated with oil due to different surface tensions of the liquid phases in the paste. This difference in interfacial tension of the oil and vegetation water through the steel plate facilitates the separation between the solid and the liquid phase.

After the extraction of olive oil from the olive paste, two residual products are left over (1) olive husk and (2) vegetation water. Additional oil can be obtained from the olive husk and is called pomace oil. Pomace oil is extracted with solvents such as hexane by continuous system of centrifugation. This raw pomace oil obtained from the husk (dark green in colour) has a medium-high free acidity depending upon the

extent to which fermentation has taken place and also contains impurities extracted by the solvent. Therefore refining of the pomace oil is done to make this oil edible. This involves the de-acidification, winterization and bleaching of the pomace oil. The final step is the deodorization of the pomace oil to eliminate volatile compounds that gives an unpleasant odour to the oil (Boskou 2006).

1.2.3 Chemical characteristics of olive oil

1.2.3.1 Fatty acids

Olive oil contains a high percentage of triglycerides and free fatty acids. Around 13 main fatty acids are present in olive oil: myristic acid (C14:0), palmitic acid (C16:0), palmitoleic acid (C16:1), heptadecanoic acid (C17:0), heptadecenoic acid (C17:1), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), linolenic acid (C18:3), archidic acid (C20:0), eicosenoic acid (C20:1), behenic acid (C22:0) and lignoceric acid (C24:0). Olive oil contains a small percentage of saturated fatty acids, such as palmitic acid (7.5-20%) and stearic acid (0.5-5%). Olive oil contains a very high percentage of monounsaturated fatty acid (MUFA), with predominance of oleic acid (55-83%) and also contains moderate quantities of linoleic acid (3.5-21%) and linolenic acid (0-1.5%), that are polyunsaturated (PUFA). The specific fatty acid composition in olive oil varies depending on various factors such as olive cultivar, climate, latitude, and stage of maturity (Boskou 2006; Preedy 2010).

1.2.3.2 Minor components of olive oil

In addition to the major triglyceride components, olive oil contains multiple minor components that are important for determining the quality, authenticity, aroma, flavor and shelf life of the oil. The minor components of olive oil can be separated into two classes based on their method of isolation: those in the unsaponifiable fraction and those in the soluble fraction.

I. Unsaponifiable fraction

The unsaponifiable fraction is defined as the fraction that has been extracted with solvents after the saponification of the oil. The components of unsaponifiable fraction in order of their increasing polarity are as follows: hydrocarbon, tocopherols, fatty alcohols, tri-terpenic alcohols, 4-methyl sterols, sterols, other terpenic compounds and polar pigments (chlorophyll, xanthophyll, and pheophytins) (Boskou 2006; Kailis *et al.* 2007).

The main hydrocarbon present in the unsaponifiable fraction of olive oil is squalene (around 400 mg / kg) (Covas *et al.* 2006). Squalene is formed by the condensation of six units of isoprene and is a precursor in the biosynthesis of cholesterol and other steroids.

Olive oil contains the tocopherols α-, β-, γ-, δ- (100-250 mg / kg), where α-tocopherol (vitamin E) makes up almost 95% of the total tocopherol content (Boskou 2006). The antioxidant properties of the tocopherols in olive oil play an important role in maintaining the stability of the oil (Boskou 2006). The tocopherol content of olive oils depends on various factors, involved in transportation, storage and method of olive oil processing. Tocopherol content is higher in olives that have been collected during the first period of harvesting than those which have been harvested towards the end of the period. The tocopherol content is much lower in olive oils that have been refined, bleached and deodorized (Boskou 2006).

Another important minor constituent of olive oil are fatty alcohols which can be further classified to aliphatic alcohols and terpenic alcohols. They are used as an important indicator used to differentiate various olive oil classes. Waxes are esters of fatty alcohols with fatty acids which are semi-soluble compounds primarily coming from the skin of the olive fruit. The content of wax in olive oil is very low, not exceeding 35mg/100g (Boskou 2006). Triterpenes are mainly concentrated in the skin of the olive fruit and the four most abundant triterpenes present in olive oil are oleanolic acid, maslinic acid, erythrodoil and uvaol alcohols (Covas *et al.* 2006).

Sterols are an important non-glyceride fraction of olive oil, which are broadly used for checking the authenticity and quality of the oil. Four classes of sterols are present in olive oil: 4α-desmethyl sterols (common sterols), 4α-methyl sterols, 4,4-dimethyl sterols (triterpene alcohols) and triterpene dialcohols. The saturated forms of plant sterols which lack double bonds in the steroid nucleus and alkyl side chains are referred to as plant stanols (Piironen *et al.* 2000). As these compounds and their presence in olive oil were the focus of the current study they are discussed at length below (Section 1.8).

The presence of pigments in the oil provides its colour ranging from green-yellow to golden depending on the degree of fruit ripeness and olive cultivar. The composition of pigments is used as a criterion for evaluating the quality of the oil. Two classes of pigments are present in olive oil (1) chlorophylls and pheophytins, (2) carotenoids. The total content of chlorophyll (a and b) and pheophytin (a and b) in olive oil ranges between 1-20 ppm. The total content of carotenoids in olive oil ranges between 1-20 ppm and the main carotenoids present in olive oil are lutein, β-carotene, laxanthin and neoxanthin.

II. Soluble fraction

The soluble fraction of olive oil mainly consists of phenolic compounds, which are usually obtained from the oil by extraction with methanol-water. Polyphenols in olive oil are recognized for their role as antioxidants which increases the oil's shelf life, as well as flavour determinants of the bitterness and pungency of olive oil (Covas *et al.* 2006). The concentration of polyphenols in olive oil vary from 0 to 1000 ppm or more, depending on altitude, harvesting time and processing conditions (Mailer *et al.* 2006). The major phenolic compounds present in olive oil are as follows: (1) simple phenols (eg. hydroxytyrosol, tyrosol, vanillic acid); (2) secoiridoids (eg. oleuropein glucoside) and ligstroside lacking a carboxymethyl group and the aglycone form of oleuropein glucoside and ligstroside; (3) polyphenols which are lignans and flavonols. Around 90% of the phenolic content of olive oil is made up of hydroxytyrosol, tyrosol, and secoiridoid derivatives. Oleuropein is the main glycoside present in olive oil which is responsible for the bitter taste of immature olives (Boskou 2006).

1.3 HEALTH BENEFITS OF OLIVE OIL

One of the factors that have contributed to the increase in global olive oil consumption has been the general perception of the health benefits of this oil over other dietary fats/oils. Olive oil has been widely studied for its effects on coronary heart diseases, particularly for its ability to reduce blood pressure and LDL cholesterol levels. Other health benefits of olive oil consumption reported in the medical literature include a reduced risk of developing a number of diseases including hypertension, atherosclerosis, Parkinson's disease, Alzheimer's disease, high blood pressure, diabetes and different types of cancers(Choudhury *et al.* 1995; Kiritsakis 1999; Harwood *et al.* 2000; Beardsell *et al.* 2001; Harwood *et al.* 2002; Shahtahmasebi *et al.* 2003; Parker *et al.* 2004; Xu *et al.* 2005; Quiles *et al.* 2006; Vossen 2007; Sheppard 2008; Preedy 2010).The sum of the research in this area is beyond the scope of this literature review and some of the relevant reviews have been discussed in this section.

The comparison of the levels of saturated and unsaturated fatty acids in different kinds of oils has shown that olive oil contains high levels of monounsaturated fatty acids (MUFA) (especially oleic acid) and low levels of saturated fats making it popular as a healthy oil (Alonso *et al.* 2006; Covas *et al.* 2006) (Section 1.2.3.1) (Table 1.2).

	Saturated fat (g/100g)	**MUFA (g/100g)**
Olive oil	13	73.9
Canola Oil	7	58.9
Peanut Oil	17	46.2
Lard	39	41.6
Sunflower oil	10	31.8
Corn Oil	13	29.3
Butter	62	26.8
Soybean oil	14	24.3

Table 1.2 Comparison of levels of saturated and unsaturated fatty acids in different kinds of oils (Alonso *et al.* 2006)

In Mediterranean countries, the main source of MUFA in the diet comes from olive oil (Alonso *et al.* 2006). The health benefits of MUFA-rich diets on plasma cholesterol levels, were the first to generate interest in olive oil (Stark *et al.* 2002). A large number of studies has provided evidenceon the cardiovascular benefits of a Mediterranean-diet rich in olive oil, by showing improvement in their lipid profile, through a decrease in total and LDL-cholesterol and an increase in the HDL-to-cholesterol ratio (Stark *et al.* 2002; Vincent-Baudry *et al.* 2005; López-Miranda *et al.* 2010).

A randomized, double-blind, five year-period crossover study was conducted on 22 healthy non-obese normocholesterolaemic volunteers receiving four cholesterol-loweringdiets: (1) an American Heart Association/National Cholesterol Education Program Step II diet (in these diets, saturated fat is replaced by carbohydrate resulting in low-fat and high-carbohydrate diet) and 3 high-MUFA diets: (2) olive oil (OO), (3) peanut oil (PO), and (4) peanuts and peanut butter (PPB) (Kris-Etherton *et al.* 1999). These results were compared with volunteers receiving the Average American diet (AAD). The results showed that the high-MUFA diets lowered total cholesterol by 10%and LDL cholesterol by 14% in comparison to the Step II diet. The OO, PO, and PPB diets decreased cardiovascular disease risk by an estimated 25%, 16%, and 21%, respectively, while the Step II diet lowered the risk only by 12%.

During 2003-2004,a randomized trial was conducted on 772 asymptomatic persons 55 to 80 years of age at high cardiovascular risk (Estruch *et al.* 2006). These participants were assigned to three different types of diets: (1) low-fat diet (sample size (n=257); (2) Mediterranean diet with virgin olive oil, 1 litre per week (n= 257), (3)Mediterranean diet with free nuts, 30 g/d (n= 258). The analysis of the results showed that the mean changes in the Mediterranean diet with olive oil group and the Mediterranean diet with nuts group were 0.39 mmol/L and 0.30 mmol/L, respectively, for plasma glucose levels; 5.9 mm Hg and 7.1 mm Hg, respectively, for systolic blood pressure; and 0.38 and 0.26, respectively, for the cholesterol-high-density lipoprotein cholesterol ratio, compared to the low fat diet. This trial concluded that Mediterranean diets supplemented with olive oil or nutsmay decrease the risk of coronary heart disease by improving the serum lipid profile.

A randomized, crossover controlled trial was conducted on twenty-eight stable coronary heart disease patients where a daily dose of 50 ml of virgin and refined olive oil (ROO) was sequentially administered over two periods of three weeks. It was observed that the intervention of olive oil led to the reduction of inflammatory markers such as interleukin-6 ($P<0.002$) and C-reactive protein ($P=0.024$) in the subjects, however, virgin olive oil (VOO) was found to be more effective that ROO (Fito *et al.* 2007).

In addition to having high levels of MUFA, olive oil also contains multiple minor components (Section 1.2.3.2) which are also thought to have a positive impact on cardiovascular health (Stark *et al.* 2002).

A study was conducted by Nicolaiew et al (1998) in healthy subjects to isolate the effects of oleic acid from other potentially biologically active compounds in olive oil. To determine if oleic acid alone was responsible for the ability to lower risk of cardiovascular disease extra virgin olive oil was compared with oleic acid-rich sunflower oil and an oral fat load was performed after 3 weeks. The analysis showed that fasting and postprandial plasma lipid levels were similar following consumption of either oil however when LDL oxidation susceptibility was evaluated by measuring the formation of conjugated dienes, the diene production decreased only after the olive oil diet though not significant ($P=0.085$) (Nicolaiew *et al.* 1998). This suggested that other of minor components of virgin olive oil have a mild effect on cardiovascular risk factors.

An important minor constituent of olive oil are phytosterols, which are widely accepted worldwide as one of the most important markers for determining the authenticity of olive oil (Section 1.2.3.2) (Bohacenko *et al.* 2001). In addition, phytosterols have been a focus of much research for the last half-century because of their potential health benefits (Pollak 1953; Pegel 1997; Normén *et al.* 2001; Berger *et al.* 2004; Rudzinska *et al.* 2005; Jiménez-Escrig *et al.* 2006). Phytosterols have long been known to reduce the level of plasma-cholesterol and LDL cholesterol in humans. The structural homology of phytosterols to cholesterol causes them to compete for absorption and uptake in the small intestine. This reduces the cholesterol content of

micelles and decreases its transport to the intestinal brush border membranes. As cholesterol is no longer soluble outside the micellar phase it forms co-crystals with phytosterols and is rapidly excreted by the liver, thereby reducing the level of plasma-cholesterol and LDL cholesterol (Trautwein *et al.* 2007).

A case control study was conducted by Pollak (1953) on 26 healthy subjects which showed that an average consumption of 8.1g of plant sterols (containing 75-80% sitosterol) daily for 2 weeks reduced the total plasma cholesterol levels by 27.7%. Pollak's findings were followed up by a number of studies where sitosterol preparations of different origin were used to study their effect in lowering blood cholesterol levels, and virtually all studies showed consumption of food supplemented with sitosterol significantly lowered blood cholesterol levels (Kritchevsky *et al.* 2005). The efficacy of phytosterol esters was also affirmed in a meta-analysis of 41 trials with various phytosterol enriched food products, and it was shown that an optimum daily dosage of 2 g/day sterols reduces LDL-cholesterol level by 10%, however higher doses provides only a small additional effect (Katan *et al.* 2005).

To study the effect of phytosterol enriched margarines on plasma total and LDL cholesterol levels, a randomised double-blind study was conducted on one hundred healthy non-obese normocholesterolaemic and mildly hypercholesterolaemic volunteers receiving a diet enriched with sterols from soybean, sheanut or ricebranoil or with sitostanol-ester (Weststratea *et al.* 1998). The analysis of results showed that in comparison to the controls, the plasma total- and LDL-cholesterol concentrations weresignificantly reduced by 8-13% (0.37-0.44 mmol/L) for margarines enriched in soybeanoil sterol-esters or sitostanol-ester. The LDL:HDL cholesterol ratio was also reduced by 0.37 and 0.33 units respectively for these margarines.

Phytostanols, hydrogenated forms of phytosterols, are more effective in lowering cholesterol levels in mammals. This is due to the fact that phytostanols are virtually unabsorbable and they remain in the intestinal lumen for longer periods and thus can interfere with the absorption of cholesterol continuously and in a much more efficient manner (Piironen *et al.* 2000; Sakouhi *et al.* 2009). A randomized one year double-blind study was conducted in 153 patients with mild hypercholesterolemia where 51 patients consumed margarine without sitostanol ester (control) and 102 patients

consumed margarine containing sitostanol ester (1.8 or 2.6 g of sitostanol per day) (Miettinen *et al.* 1995). In the sitostanol group the mean reduction in serum cholesterol level was 10.2% as compared with an increase of 0.1% in the control group.

Apolipoprotein E-deficient (apoE−/−) mice were used as animal models to study whether the antiatherogenic effects of phytosterols are associated with reductions in proinflammatory cytokine production (Ruiz-Canela *et al.* 2011). These models were fed a cholesterol-supplemented diet in the presence or absence of 2% dietary phytosterols for 14 weeks and then immunized with ovalbumin. The results showed that dietary phytosterols reduce plasma cholesterol concentrations and atherosclerotic lesion size by 20 and 60% respectively. In addition, phytosterol-treated mice relative to controls showed 10 times higher production of anti-inflammatory [interleukin (IL)-10] and reduced production of proinflammatory cytokines, IL-6 and tumor necrosis factor (TNF)-α (Nashed *et al.* 2005). The results concluded that phytosterol-rich diets were strongly associated with reduced concentrations of plasma cholesterol, proinflammatory cytokine production and incidence of atherogenesis in murine models with elevated susceptibility to atherosclerosis.

Thus these epidemiologic data, combined with clinical intervention studies clearly shows that consumption of olive oil, a rich source of MUFA, phytosterols and polyphenols can lower LDL-cholesterol levels and eventually reduce the risk of coronary heart diseases.

Apart from their cholesterol lowering property, there is epidemiological evidence that olive oil could exert a favorable effect on hypertensive patients, being much more stable and less susceptible to oxidation due to the presence of high levels of MUFA (Alonso *et al.* 2006).

A double blind randomized cross-over clinical trial was conducted on 23 mild to moderate hypertensive patients, where eleven patients were fed with a MUFA rich diet from EVOO and twelve patients were fed with a PUFA rich diet from sunflower oil for a period of six months. Thereafter the blood pressure and heart rate were measured for every patient every two months. The results showed that the systolic and

diastolic blood pressure were significantly lower for patients fed with MUFA rich diet compared with patients fed with PUFA rich diet. The daily drug dosage was reduced by 48% for patients with a MUFA diet and by 4% for patients with a PUFA diet. In particular, eight patients receiving the MUFA diet required no drug therapy while all patients receiving the PUFA diet required antihypertensive treatment (Ferrara *et al.* 2000). The analysis of results showed that consumption of MUFA rich diet is associated with lower systolic and diastolic blood pressure in addition to reduced need for antihypertensive medication.

A cohort study that has assessed the relationship between MUFA, olive oil and blood pressure within a Mediterranean population is the SUN (Seguimiento Universidad de Navarra) study. This study involved 6863 participants whose diet was assessed using a semiquantitative food-frequency questionnaire previously validated in Spain, with 136 items and with at least two year of follow-up. The analysis of the data after a median follow-up time of 28.5 months showed that the cumulative incidence of hypertension was 4.7% among men and 1.7% among women. Among men the odd ratios (OD) (95% confidence interval) of hypertension from the second to the fifth quintile of olive oil consumption were 0.55 (0.28-1.10), 0.75 (0.39-1.43), 0.32 (0.15-0.70), and 0.46 (0.23-0.94), respectively (P = 0.02 for linear trend) (Alonso *et al.* 2004). This clearly indicated that men with the highest baseline of olive oil consumption had a significantly lower risk of hypertension compared with those in the first quintile. In women no such association was found between olive oil consumption and reduced risk of hypertension.

There are several studies that have found an association with olive oil and a reduced risk of developing different types of cancers (Preedy 2010; Psaltopoulou *et al.* 2011). Olive oil contains at least 30 phenolic compounds recognized for their role as antioxidants, which quench the formation of reactive oxygen and nitrogen species, thus modulate oncogenes, tumour suppressor genes and signal transduction pathways leading to the inhibition of cell proliferation and induction of apoptosis.

Several case control studies have been conducted in Greece, Spain, Italy and France to study the association between olive oil consumption and reducing risk of breast cancer (Psaltopoulou *et al.* 2011).

Three case-control studies were carried out in Spain to study the effects of olive oil consumption on patients diagnosed with breast cancer incidents (Landa *et al.* 1994; Martin-Moreno *et al.* 1994; García-Segovia *et al.* 2006). The first trial was carried out in 1994, (762 cases- 988 controls), for olive oil intake, the odds ratios for the highest versus the lowest quartile of consumption was 0.66, with a significant dose-response trend. In the second trial conducted from 1988-1991 (100 breast cancer cases-100 controls), women in the highest tertile of monounsaturated fat consumption were at lower risk compared with women in the lowest tertile (RR (relative risk) = 0.30; 95% CI 0.1 to 1.08). The third trial was conducted from 1999-2001 in the Canary Islands on 755 women with 291 incident cases with confirmed breast cancer and 464 controls. The results showed that the odds ratio (OR) for women in the three upper quintiles of olive oil consumption (≥ 8.8 g/day) was 0.27 (95% CI 0.17-0.42) and OR for monounsaturated fat intake is 0.52 (95% CI 0.30-0.92), and thus supporting the protective role of olive oil consumption on breast cancer among Canaries women.

Animal studies involving immune deficient female mice lacking either/both B and T cells (SCID mice) were fed with defined diets supplemented with 0.2% cholic acid and 2% sterols (phytosterols or cholesterol mixture). MDA-MB-231estrogen receptor-negative human breast cancer cells were injected into their inguinal mammary fat pads of the mice after two weeks. It was observed those mice fed with enriched phytosterols showed a 40% reduction in serum cholesterol, 20-30 fold increase in serum β-sitosterol and campsterol concentration and 33% smaller breast tumour sizes than those animals fed with cholesterol diet (Awad *et al.* 2000).

A series of case-control studies were conducted at major hospitals in Uruguay where the relationship between dietary phytosterols and risk of specific cancers such as lung, breast, stomach and esophageal cancer were studied (Mendilaharsu *et al.* 1998; Ronco *et al.* 1999; De Stefani *et al.* 2000; Stefani *et al.* 2000; Bradford *et al.* 2007). These studies involved 100-500 patients diagnosed and histologically verified with specific cancers and frequency matched control patients of similar age, gender and residence. All patients were interviewed using food frequency questionnaires based on 64 food items considered representative of the local diet and dietary intake of total phytosterol, β-sitosterol, campesterol, and stigmasterol were assessed using published food

composition data. A strong inverse relationship was observed between total phytosterols with stomach cancer (Stefani *et al.* 2000) where OR for patients with the highest tertile of total phytosterols intake was 0.33 (95% confidence interval 0.17–0.65). In cases of lung, breast and esophageal cancers, the odds ratio for total phytosterol intake in the highest tertile was found to be 0.29 (95% CI = 0.14-0.63), 0.41 (95% CI = 0.26-0.65) and 0.21 (95% CI = 0.10-0.50) respectively (Mendilaharsu *et al.* 1998; Ronco *et al.* 1999; De Stefani *et al.* 2000). Thus these studies revealed that the intake of dietary phytosterols was associated with a potential protective effect in all four major cancers.

To study the potential immunomodulatory effect of phytosterols, hypercholesterolemic apolipoprotein E-deficient (apoE−/−) mice and normocholesterolemic C57BL/6J mice were fed with or without a 2% phytosterol supplement and treated with turpentine or saline and euthanized after 48 hours (Calpe-Berdiel *et al.* 2007). The cultured spleen lymphocytes of apoE−/− mice fed with phytosterols showed an increased secretion of Th1 lymphocyte cytokines (IL-2 and IFN-γ) compared to control fed animals. However no change in the levels of Th2 lymphocyte cytokines (IL-4 and IL-10) was observed. In C57BL/6J mice, the phytosterol rich diet inhibited intestinal cholesterol absorption without decreasing plasma cholesterol levels. The level of IL-2 production increased in C57BL/6J mice fed with phytosterols, however the production of IFN-γ, IL-4 and IL-10 remained unchanged. These studies showed that in apoE−/− and C57BL/6J mice, the phytosterol supplemented diet significantly increases the Th1/Th2 ratio, with the immune equilibrium towards a Th1 response.

After decades of epidemiological, clinical and experimental research, it is clear that consumption of olive oil have a positive influence on health outcomes, including coronary heart disease, hypertension, immune function and different types of cancers. These potential health benefits of olive oils can be partially attributed to the phytosterol content in olive oil. This has led to an increased interest in phytosterol research in the past decade, and incorporation of phytosterols in an increasing cultivars of foods to act as “nutriceuticals” (Engel *et al.* 2005; Fernandes *et al.* 2007). Commercially plants sterols are supplemented to spreads, margarines (Logicol-Australia, Benecol-UK), orange juices, mayonnaises, milk, yoghurt drinks (Logicol-

Australia), soy milk, meat and soups and green tea (Sierksma *et al.* 1999; Berger *et al.* 2004; Kritchevsky *et al.* 2005; Trautwein *et al.* 2007).

1.4 OLIVE OIL ADULTERATION AND INTERNATIONAL OLIVE OIL COUNCIL (IOOC) STANDARDS

1.4.1 Adulteration of olive oils

Due to its unique properties, olive oil demands a premium price (Boskou 2006; Therios 2008). This has led to frequent reports of adulteration of the oil with cheaper vegetable oils and refined or processed olive oils in order to improve profit margins (Firestone 2001; Ozen *et al.* 2002; Dourtoglou *et al.* 2003; Frankel 2010). The increase for demand for olive oil has seen an increase in cases of adulteration where olive oil producers has been found guilty of blending smaller amounts of olive oil with cheaper vegetable oils and re-labeling and pricing it as high quality extra virgin olive oil product (Tay *et al.* 2002). In addition to the fraudulent nature of this practice it has also caused issues for consumer health.

In 1981, a mass food poisoning broke out in the Madrid area of Spain which affected more than 20,000 people and resulted in more than 800 deaths. This incident was caused due to the consumption of aniline-laced industrial rapeseed oil which was sold as olive oil and cooking oil in the market(Firestone 2001; Ferragut 2007). In 1995, Turkish hazelnut oil, being three times cheaper than olive oil, was introduced illegally into the European Union in order to adulterate olive oil (Webster *et al.* 2001). These incidents have drawn the attention of the FDA in the United States and other government health organizations worldwide and as a result, in order to minimize such fraudulent cases of providing inferior or adulterated olive oil, the IOOC (International Olive Oil Council) has imposed strict standards that olive oils have to meet in order to be accepted as authentic on the international European export market(International Olive Oil Council 2003).

1.4.2 Classification of olive oils according to IOOC standards

The Codex Alimentarius Commission (http://www.codexalimentarius.org/), which was established in 1963 by FAO (Food and Agriculture Organization of the United Nations) and WHO (World Health Organization), is a collection of international standards and codes of practices related to food production and safety. Although the standards of Codex Alimentarius are universal standards which cover the regulations of all member countries, the codex standards for olive oil have mostly been adopted from the International Olive Oil Council (IOOC). The IOOC Trade Standard and the Codex Alimentarius Draft Standard has classified olive oil products into different classes: virgin olive oil, refined olive oil, olive oil and olive pomace oil (International Olive Oil Council 2006).

1.4.2.1 Virgin olive oil

These oils are obtained from the olive fruits solely by mechanical and other physical means under strict conditions that do not lead to alterations in the oil. The only treatments these oils undergo are washing, decanting, centrifuging and filtration.

Those virgin oils which are fit for consumption are as follows:

(1) Extra virgin olive oil (EVOO): This olive oil category has a free acidity, expressed as oleic acid, of not more than 0.8 grams per 100 grams.
(2) Virgin olive oil: This olive oil category has a free acidity, expressed as oleic acid, of not more than 2 grams per 100 grams.
(3) Ordinary virgin olive oil: This olive oil category has a free acidity, expressed as oleic acid, of not more than 3.3 grams per 100 grams.

Those virgin oils which are not fit for consumption, designated as lampante virgin olive oil, has free acidity, expressed as oleic acid, of more than 3.3 grams per 100 grams. These oils have an off flavor and/or off smell and are intended for refining.

1.4.2.2 Refined Olive oil

These oils are obtained by refining virgin olive oil with a high acidity level and/or organoleptic defects which are eliminated during refining but do not lead to alteration in their initial glyceride structure. This olive oil category has a free acidity, expressed as oleic acid, of not more than 0.3 grams per 100 grams.

1.4.2.3 Olive oil

These oils are obtained by blending refined olive oil and virgin olive oil. This olive oil category has a free acidity, expressed as oleic acid, of not more than 1%.

1.4.2.4 Olive pomace oil

These oils are obtained by treating olive pomace with solvents but not subjected to any re-esterification processes which can alter the initial glyceride structure. This olive oil category has a free acidity, expressed as oleic acid, of not more than 1.5 grams per 100 grams.

1.5 AUTHENTICITY AND QUALITY STANDARDS FOR OLIVE OIL

Various analytical criteria, for quality, sensory and purityevaluation of each grade of olive oil, have been developed by the IOOC in order to prevent unfair competition in the market and to ensure the genuineness and purity of the product(International Olive Oil Council 2011). The IOOC trade standard also lays down rules for hygiene, packing, labeling and also recommends methods for determining various analytical parameters (Harwood *et al*. 2000; Boskou 2006; Weisman 2009).

1.5.1 Basic quality parameters

The basic quality parametersare used for the evaluation of the primary quality of olive oil. The quality standards based on these parameters can be further divided into two types:

- Analytical quality controls for olive oil (Table 1.3)
- Sensory quality controls for olive oil

1.5.1.1 Analytical quality controls for olive oil

Free fatty acid content (FFA) (%): Free fatty acids in olive oil are formed due to the breakdown of triglycerols caused by the hydrolysis of the oil. Lower content of free fatty acids is associated with higher quality oil. The quantity of FFA is an important factor for classifying olive oil into different commercial grades (Weisman 2009).

Peroxide value (meq O_2/kg): The peroxide value (PV) is an indication of the amount of hydroperoxides generated in the oil from the oxidation of the polyunsaturated fatty acids in the oil. Peroxides are unstable leading to oil rancidity on accumulation. It is analysed by means of titration that liberates iodine from potassium iodide and is expressed in milequivalents of free oxygen per kilo of oil (meq O_2/kg). The IOOC standards permit EVOO to have PVs of up to 20 meq O_2/kg while PVs are greatly reduced for refined olive oils (Table 1.3).

UV Light Absorbency (nm): This is a more delicate indicator of oxidation to identify oils which are old or have been refined by heating the oil. A spectrophotometer is used to detect the presence of oxidised compounds in the ultraviolet spectrum at wavelengths of 232 and 270 nm. Absorbance is also measured at wavelengths of Δ K (difference in absorbance at 270nm and 266-274nm) to detect the presence of colour removing substances or refined or pomace oil.

Moisture and volatile matter (% m / m): The amount of moisture and volatile matter in olive oil formed as a result of the extraction method is measured. High moisture content impairs the quality of the oil by reducing flavor and level of

antioxidants. It is evaluated by measuring the loss of weight of the product by heating at103˚C ± 20˚C in a drying stove for 30 minutes and then measuring the difference in weight until no further difference is detected. It is expressed in terms of percentage of the total weight.

Insoluble impurities in light petroleum (% m/m): This is a measure of the presence of dirt, minerals, resins, oxidized fatty acids, alkaline soaps of palmitic and stearic acids, and proteins that are formed due to poor manufacturing practices during olive oil production. They are determined by dissolving the oil in n-hexane or petroleum ether and then removing the insoluble matter by filtration. It is expressed as a percentage of the total (Weisman 2009).

Flash point: This is the measure of the temperature at which the oil will burst into flames. Refined olive oil, pomace oil and seed oil have lower flash point than virgin olive oils.

Trace metals (mg/kg): This is a measure of the amount of copper and iron in mg / kg present in the oil which may originate from the soil and fertilizers or from contamination during processing and storage. They are determined by atomic absorption spectrometry.

1.5.1.2 Sensory quality controls for olive oil

One of the most important parameters in ensuring the acceptance of olive oil for consumption is sensory or organoleptic assessment. The primary sensory components of olive oil are aroma, flavour, pungency and bitterness. The positive attributes are fruity, green, spicy, citrus, fragrant, bitter, pungent and over-ripe. The negative attributes are muddy, winey, metallic, fusty, musty, rancid and burnt. Extra virgin olive oils should not have any sensory defects and must have some fruity attribute whereas most of these sensory compounds are lost in refined olive oil.

This assessment is carried out by 8-12 panel members who have been trained to IOOC standards. The rating of the olive oil is awarded on the basis of a scale of points ranging from 0 to 9, where 0 indicates that the oil has extreme defects and 9 indicates

that it has no defects. The rating of olive oil is accepted only if majority (atleast 5 out of 8) of the tasters agree about the same defect in a particular batch of olive oil. Depending on the overall rating of the panel, olive oil is classified as extra virgin olive oil, virgin olive oil, ordinary virgin oil or lampanate virgin olive oil.

1.5.2 Other quality parameters

Other quality parameters are mainly used for determining the purity of the oil rather than the quality itself (Table 1.6). These are mainly used to detect the presence of other vegetable oils or refined oils in virgin olive oils andolive-pomace oils (Harwood *et al.* 2000; Weisman 2009).

Fatty acid composition (%): This is a measure of the proportions of individual fatty acids in the olive oil (Table 1.4).The composition of fatty acids influences the stability, nutritional value and are characteristic of olive oils (Weisman 2009). These are analysed by assessing fatty acid methyl esters using gas chromatography.

Fatty acid	IOOC limit
Myristic acid (C14:0)	< 0.05
Palmitic acid (C16:0),	7.5-20.0
Palmitoleic acid (C16:1)	0.3-3.5
Heptadecanoic acid (C17:0)	≤ 0.3
Heptadecenoic acid (C17:1)	≤ 0.3
Stearic acid (C18:0)	0.5-5.0
Oleic acid (C18:1)	55.0-83.0
Linoleic acid (C18:2)	3.5-21.0
Linoleic acid (C18:3)	≤ 1.0
Archidic acid (C20:0)	≤ 0.6
Eicosenoic acid (C20:1)	≤ 0.4
Behenic acid (C22:0)	≤ 0.2*
Lignoceric acid (C24:0)	≤ 0.2

Table 1.4 IOOC standards for individual fatty acids in olive oil
*Limit raised to < 0.3 for olive-pomace oils.

Different classes of olive oil	Free fatty acid content (% m/m expressed in oleic acid)	Peroxide value (milleq. Peroxide oxygen per kg/oil)	Absorbency in ultra-violet (K1%1cm) (270nm)	Absorbency in ultra-violet (K1%1cm) Δ K	Absorbency in ultra-violet (K1%1cm) (232nm)	Moisture and volatile matter (% m/m)	Insoluble impurities in light petroleum (% m/m)	Flash point	Trace metals - Iron (mg / kg)	Trace metals - Copper (mg / kg)
Extra virgin olive oil	≤0.8	≤ 20	≤ 0.22	≤ 0.01	≤ 2.5	≤ 0.2	≤ 0.1		≤ 3.0	≤ 0.1
Virgin olive oil	≤2.0	≤ 20	≤ 0.25	≤ 0.01	≤ 2.6	≤ 0.2	≤ 0.1		≤ 3.0	≤ 0.1
Ordinary virgin olive oil	≤3.3	≤ 20	≤ 0.30	≤ 0.01		≤ 0.2	≤ 0.1		≤ 3.0	≤ 0.1
Lampanate virgin olive oil	≤3.3	no limit				≤ 0.3	≤ 0.2		≤ 3.0	≤ 0.1
Refined olive oil	≤0.3	≤ 5	≤ 1.10	≤ 0.16		≤ 0.1	≤ 0.05		≤ 3.0	≤ 0.1
Olive oil	≤1.0	≤ 15	≤ 0.90	≤ 0.15		≤ 0.1	≤ 0.05		≤ 3.0	≤ 0.1
Crude olive pomace oil	no limit	no limit				≤ 1.5		≥ 120˚C		
Refined olive pomace oil	≤0.3	≤ 5	≤ 2.00	≤ 0.20		≤ 0.1	≤ 0.05		≤ 3.0	≤ 0.1
Olive pomace oil	≤1.0	≤ 15	≤ 1.70	≤ 0.18		≤ 0.1	≤ 0.05		≤ 3.0	≤ 0.1

Table 1.3 Quality parameters of olive oil according to IOOC standards(International Olive Oil Council 2011)

Trans fatty acid isomers (%): Olive oil has very high level of unsaturated fatty acids (~80%) and only small amounts of saturated fatty acids where majority of the saturated fatty acids have the *cis* configuration (hydrogen atoms are present on the same side of the double bond of the carbon chain). When oils are exposed to high heat or pressure, the natural *cis* form of the fatty acids can get converted to *trans* form (hydrogen atoms are present on the opposite side of the double bond of the carbon chain). *Trans* fatty acids have adverse health effects as it tends to raise the levels of low density lipoprotein (LDL-bad cholesterol) and reduce the levels of high density lipoproteins (HDL-good cholesterol). Thus, the presence of high levels of *trans* fatty acids is indicative of refinement that precludes that categorisation of extra virgin olive oil.The IOOC limit of *trans* fatty acid isomers in olive oil is described in Table 1.6.

Sterol composition (total and individual) (mg / kg): Sterols are present in olive oil as minor components of the non-glycerine fraction. Compositional analysis of the sterol fraction is used to detect the presence of other plant oils in olive oil and to assess the degree of purity of the oil. The ratio of campesterol/stigmasterol has been reported as a quality index of olive oil, where a higher value of the ratio indicates a positive relation to the quality of the product as stigmasterol content is higher in adulterated oils (Ranalli *et al.* 1997; Koutsaftakis *et al.* 1999). The IOOC limit for total sterol content in different grades of olive oil is given in Table 1.6. The IOOC limit for individual sterols are described below (Table 1.5).

Type of Sterols	**IOOC limit**
Cholesterol	≤ 0.5
Brassicasterol	≤ 0.1 *
Campesterol	≤ 4.0
Stigmasterol	< campesterol in edible oils
Delta-7-stigmastenol	≤ 0.5
Apparent beta-sitosterol: beta-sitosterol + delta-5-avenasterol + delta-5-23-stigmastadienol + clerosterol + sitostanol + delta 5-24-stigmastadienol	≥ 93.0

Table 1.5 IOOC standards for sterol composition in olive oils
* Limit raised to < 0.2 for olive-pomace oils.

Erythrodiol and uvaol (%): These are two pentacyclic triterpenes which are concentrated in the skin and flesh of olive oil. The percentage oferythrodiol and uvaol, in the total sterol fraction, is a good indication between mechanically extracted oils and solvent extracted oils.

Waxes content (mg / kg): Waxes are esters of fatty acids with fatty alcohols. The main waxes which are present in olive oil are esters C36, C38, C40, C42, C44 and C46. Waxes mostly accumulate in the skin of olives therefore their content in pomace oils is higher as it contains a greater proportion of the fruit skin(Boskou 2006). The C40-C46 esters of the waxy proportions are least affected by the dewaxing process thus determination of the sum of C40-C46 aliphatic waxes is considered as an important purity parameter to detect the presence of olive-residue oil in olive oil.

Triacylglycerol analysis (ΔECN42): The presence of seed oils in olive oil and olive pomace oil is detected by determination of equivalent carbon number of the oil (ECN). ECN is calculated by measuring the difference between the real and theoretical content of triglycerides with an equivalent carbon number of 42. The real content of triglycerides is analysed by HPLC while the theoretical content of triglycerides is calculated by GC based on the content of C_{16} and C_{18} fatty acids in the oil.

Stigmastadiene content (mg / kg): During the process of oil refining, several unsaturated hydrocarbons with a steroideal structure, known as sterenes,are formed by dehydration of sterols. Stigmasta-3,5-diene originates from the dehydration of β-sitostero, which is considered as an effective tracer of olive oil adulteration with seed oils.

Content of 2-glyceryl monopalmitate (%): This is a measure of the percentage of palmitic acid at the 2-position of the triacylglycerols in olive oil. This test is used to determine if the oil has been re-esterified by synthetic means or by addition of animal fat.

Total unsaponifiablematter (g/kg): This is a measure of the total components of the oil that do not turn to soap under the process of saponification (Section 1.2.3.2 I).

According to IOOC standards, natural olive oils must contain less than 15g/kg of the unsaponifiable matter while olive pomace oil may contain upto 30g/kg of unsaponifiable matter.

1.6 ANALYTICAL METHODS FOR OLIVE OIL

In order to evaluate the authenticity of EVOO and detect the presence of adulterants that can devalue EVOO, a wide array of analytical methods have been published over the years (Firestone 2001; Ozen *et al.* 2002; Zou *et al.* 2009; Frankel 2010).

The most widely used analytical technique is UV Spectroscopy at 208-210 and 310-320 nm. For the quantification of seed oils in olive oils, gas chromatography is used for the analysis of fatty acid profile after methylation (Tay *et al.* 2002). Detection of adulteration with as little as 1% of vegetable oils have been claimed using a rapid reversed phase HPLC/differential refractive index system based on the equivalent carbon number (ECN) 42. Other current methodology for the analysis of olive oil products include infrared and near infrared (IR and NIR) spectroscopy, Fourier transform infrared spectroscopy, Raman and nuclear magnetic resonance (NMR) spectroscopy and a wide array of sophisticated statistical approaches (Tay *et al.* 2002; Zou *et al.* 2009; Frankel 2010).

1.7 THE COMPOSITION OF AUSTRALIAN OLIVE OILS AND IOOC STANDARDS

The natural and diverse environment and growing conditions in Australia have been shown to contribute to a wide range in olive oil quality, both biochemically and organoleptically (Mailer 2005). Australian growers primarily produce premium quality extra virgin olive oil where the oil is extracted from fresh olives within 24 hours of harvest (Mailer *et al.* 2008; Ravetti *et al.* 2010). Australia only exports extra virgin olive oil.

Different classes of olive oil	Trans fatty acid content (%) (C18:1 T)	Trans fatty acid content (%) (C18:2 T+C18:3T)	Total sterols (mg / kg)	Erythrodiol and uvaol content (% total sterols)	Waxes content (C40 + C42 + C44 + C46) (mg / kg)	Equivalent carbon number (ΔECN)	Stigmastadiene content (mg / kg)	Content of 2-glyceryl monopalmitate (2P) (%) C:16:0 ≤14.0%	Content of 2-glyceryl monopalmitate (2P) (%) C:16:0>14.0%
Edible virgin olive oils	≤ 0.05	≤ 0.05	≥ 1000	≤ 4.5	≤ 250	≤ 0.2	≤ 0.1	≤ 0.9	≤ 1.0
Lampanate virgin olive oil	≤ 0.1	≤ 0.1	≥ 1000	≤ 4.5*	≤ 300*	≤ 0.3	≤ 0.5	≤ 0.9	≤ 1.1
Refined olive oil	≤ 0.2	≤ 0.3	≥ 1000	≤ 4.5	≤ 350	≤ 0.3		≤ 0.9	≤ 1.1
Olive oil	≤ 0.2	≤ 0.3	≥ 1000	≤ 4.5	≤ 350	≤ 0.3		≤ 0.9	≤ 1.0
Crude olive pomace oil	≤ 0.2	≤ 0.1	≥ 2500	> 4.5**	> 350**	≤ 0.6		≤ 1.4	≤ 1.4
Refined olive pomace oil	≤ 0.4	≤ 0.35	≥ 1800	> 4.5	> 350	≤ 0.5		≤ 1.4	≤ 1.4
Olive pomace oil	≤ 0.4	≤ 0.35	≥ 1600	> 4.5	> 350	≤ 0.5		≤ 1.2	≤ 1.2

Table 1.6Purity parameters of olive oil according to IOOC standards(International Olive Oil Council 2011)

*When the oil has wax content between 300-350 mg/kg it is considered a lampante virgin olive oil if the total aliphatic alcohol content is ≤ 350 mg/kg or the erythrodiol + uvaol content is ≤ 3.5%

**When the oil has wax content between 300-350 mg / kg it is considered as crude olive-pomace oil if the total aliphatic alcohol content is > 350 mg/kg and the erythrodiol +uvaol content is > 3.5

The majority of the olive oil produced in Australia is analysed at IOOC accredited laboratories to keep accurate records of the olive oil quality produced in Australia on a year to year basis (Ayton *et al.* 2007; Mailer 2008; Mailer 2012).

Analysis of results in 2007 has shown that the average FFA content (%) in Australian olive oils was 0.33% which is well below the IOOC limit of 0.8% (Table 1.3). The average peroxide value in 2007 was 11meq O_2/kg which is within the limits of IOOC standards (Table 1.3) (Mailer 2012).

The average fatty acid profile for majority of the fatty acids from the olive oil samples harvested in 2007 is shown in Table 1.7. The results shows that the average fatty acid content in Australian olive oils fall well within the IOOC limit, however few samples (shown in bold) do not meet IOOC limit.

	Average (%)	**Range (%)**	**IOOC limits (%)**
C16:0	13.1	8.7-**21.9**	7.5-20.0
C18:1	71.1	**46.8**-80.7	55.0-83.0
C18:2	10.8	**3.4**-**28.3**	3.5-21.0
C18:3	0.7	0.5-**1.2**	≤1.0

Table 1.7 Fatty acid composition of Australian olive oils harvested in 2007 season
Ranges outside the IOOC limits are highlighted in bold

There is a broad range in the sterol content and profile in Australian olive oils (Table 1.8). In particular, Australian olive oils derived from the Barnea cultivar, contain up to 4.8% campesterol, as confirmed by the Australian Government Analytical Laboratories (Mailer 2007) which exceeds the IOOC standards that stipulate a campesterol level of less than 4%. It is important to note that this Israeli cultivar, Barnea, makes a major contribution to Australia's oil production (~40%) due to its excellent performance in Australian conditions (Figure 1.3). However these oils are excluded from export in the international market due to their high campesterol levels. High campesterol levels do not detract from the quality of the oils but their presence in high levels can devalue extra virgin olive oil in the international trading market as they don't adhere to the IOOC standards.

Sterols	Range (%)	Sterols	Range (%)
Cholesterol	0.03-0.16	Sitosterol	79.45-88.24
Brassicasterol	0-0.02	D-5- Avenasterol	5.21-13.66
24-Methylenecholesterol	0.02-0.48	D-5,23-Stigmastadienol	0-0.13
Campesterol	2.27-4.89	Clerosterol	0.2-0.93
Campestanol	0.1-0.25	Sitostanol	0.28-2.51
Stigmasterol	0.34-1.41	D-5,24-Stigmastadienol	0.21-1.27
D-7- Avenasterol	0.22-1.00	Apparent β sitosterol	93.83-96.38
D-7- Stigmastenol	0-0.52	Diols	0.64-3.09
D-7-Campesterol	0-0.59	Total sterols (mg / kg)	1131.7-2153.8

Table 1.8 Sterol content in Australian olive oils showing the range for each component (Mailer 2012)

In 2008, Mailer and Ayton conducted a survey on eleven different Australian olive cultivars (Arbequina, Barnea, Coratina, Corregiola, Frantoio, Koreneiki, Manzanillo, Leccino, Nevadillo Blanco, Pendolino and Picual) which represented majority of the Australian olive crop production. The olive fruit samples were collected at early and late maturity from four different geographical regions including Northern NSW/Southern Qld, Central Victoria, Western Australia and Southern Victoria/Tasmania over two years (2005 and 2006) (Mailer *et al.* 2008). This survey was designed to study the influence of natural variations (cultivars, site, harvest timing and growing season) on the quality of olive oil. This survey reports that out of 143 samples of olive oil that were tested over 2 years, 87 of them did not comply with one or more of the tests that are required for the authentication of EVOO in the international market.

The analysis of the total sterol content for eleven cultivars revealed that there is a strong relationship between the olive cultivar and the amount of phytosterol (Figure 1.6). According to international standards EVOO must have a minimum of 1000 mg/kg of total sterol content. In this study of eleven cultivars, it was revealed that the total sterol levels in few cultivars are sometimes below the IOOC limit such as Koroneiki (789mg/kg), Coratina (918mg/kg) and Pendolino (883mg/kg) while others have higher levels (>1000 mg/kg) (Mailer *et al.* 2008).

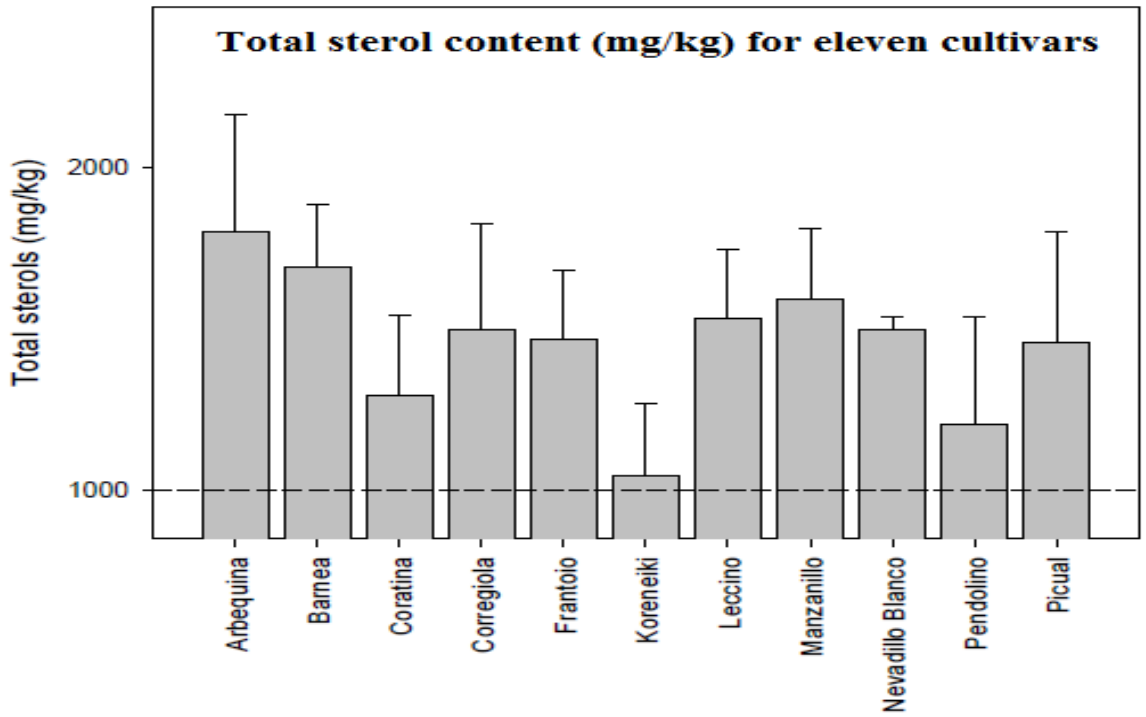


Figure 1.6 Total sterol content (mg / kg) for eleven cultivars (Mailer *et al.* 2008).
Each bar indicates the mean for the cultivar ± standard deviation.
IOOC requirement (>1000 mg / kg) is indicated by the dotted line.

Analysis of the campesterol content between eleven cultivars has showed that cultivars Barnea (4.8%) and Koroneiki (approximately 4.0%) contain higher campesterol levels in comparison to others, where Barnea clearly exceeds the IOOC limit of 4.0% (Figure 1.7).

As the area/site where olives are cultivated can have a strong influence on olive oil quality due to different temperature, rainfall and soil type, the olive cultivars were selected from four extreme sites in this study (Mailer *et al.* 2008). It was observed that the mean campesterol content regardless of the growing season, cultivar or time of harvest, was the highest in cultivars grown in Central Victoria (Figure 1.8). It is important to note that majority of the Barnea trees are grown in this region of Australia (Mailer *et al.* 2008).

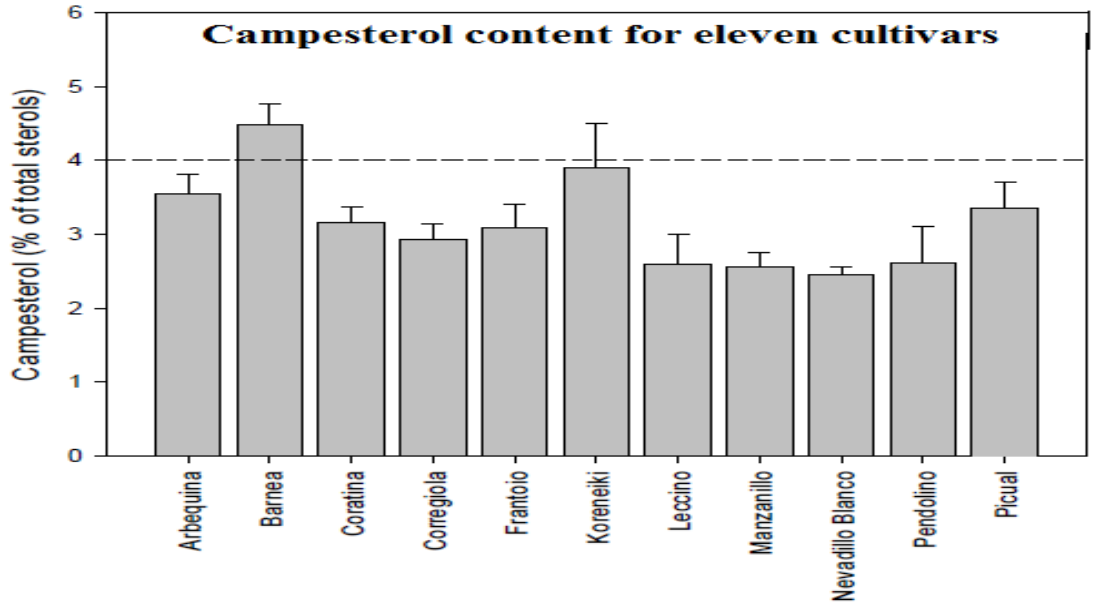


Figure 1.7 Mean campesterol content (% of total sterols) for eleven cultivars (Mailer and Ayton 2008)
Each bar indicates the mean for the cultivar ± standard deviation.
IOOC limit (<4%) is indicated by the dotted line.

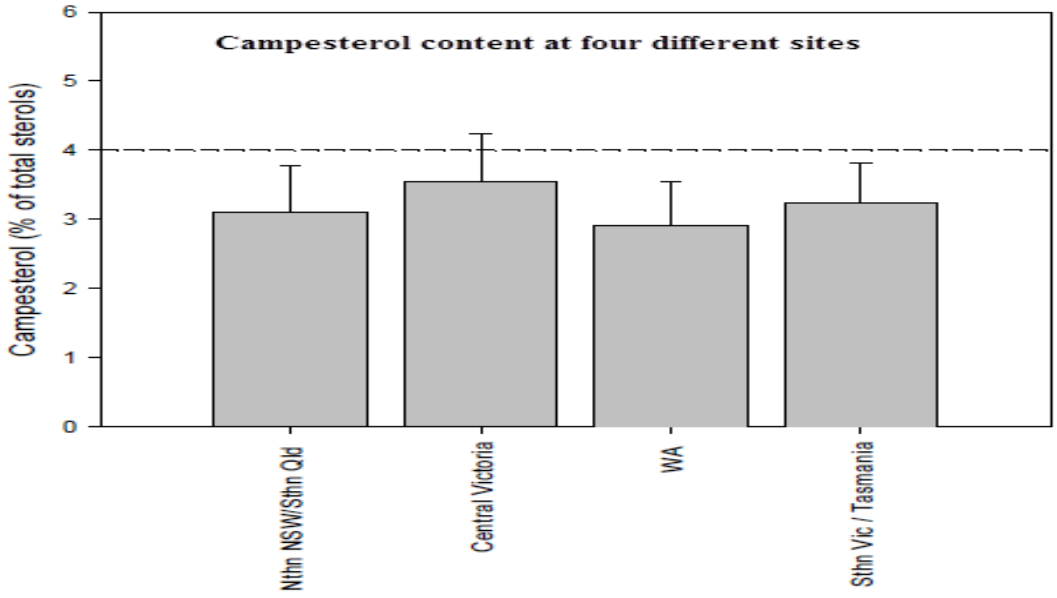


Figure 1.8 Campesterol content (% of total sterols) for four different sites in Australia (Mailer *et al.* 2008).
(Northern New South Wales/Southern Queensland, Central Victoria, Western Australia and Southern Victoria/Tasmania). This data represents the mean campesterol content in each of the four sites regardless of growing season, cultivar or time of harvest. Each bar indicates the mean for the cultivar ± standard deviation. IOOC limit (<4%) is indicated by the dotted line.

Mailer and Ayton also studied the comparison between early and late harvest times to see if the variability between oil qualities is sufficient to cause the oil to be outside the international limits for EVOO. The results showed that there was very little variation in the campesterol content in harvest timing for all cultivars, with six cultivars having slightly low levels in late harvest, while Barnea and Koroneiki continue to have high campesterol levels with no difference between early and late harvested fruit (Figure 1.9). In order to overcome seasonal differences, samples were collected at early and late harvest from four sites over two years (2005 and 2006). There was no differences for the two seasons for campesterol content in all cultivars, except that there was an average increase in campesterol content from 3.5 to 4.14% in Koroneiki and 2.02 to 2.6% in Pendolino from 2005 to 2006 (Figure 1.10).

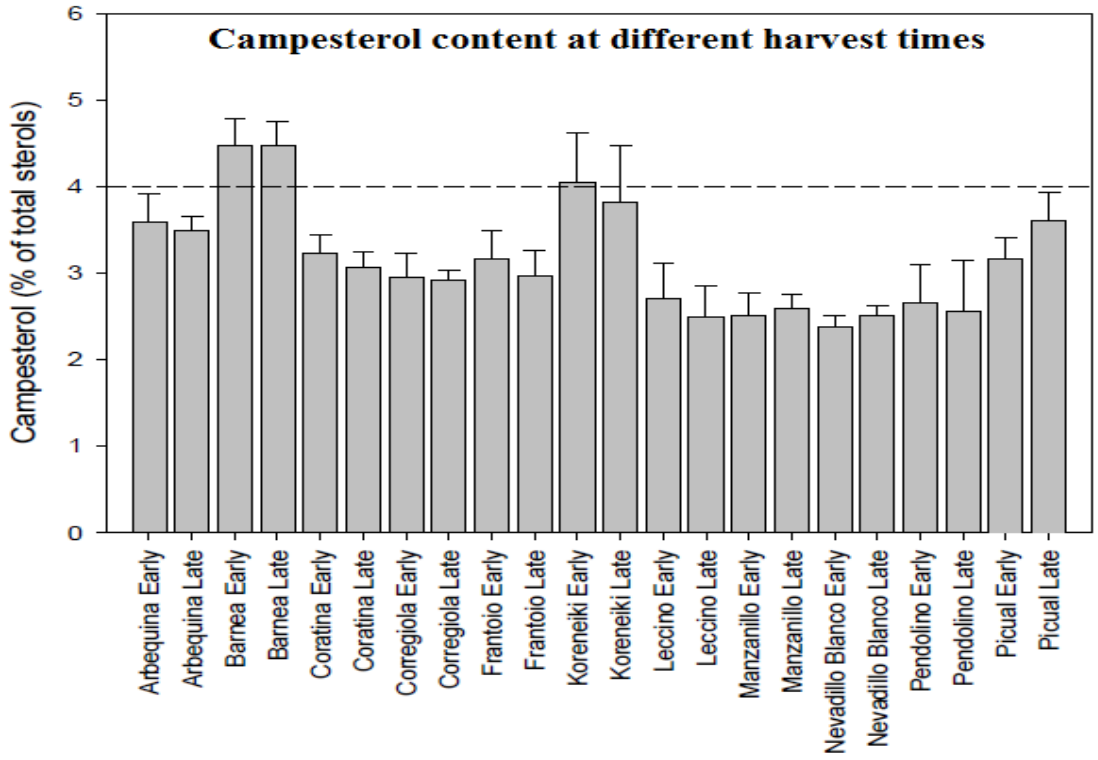


Figure 1.9 Campesterol content (% of total sterols) for eleven cultivars at different harvest times (Mailer *et al.* 2008)
Each bar indicates the mean for the cultivar ± standard deviation.
IOOC limit (<4%) is indicated by the dotted line.

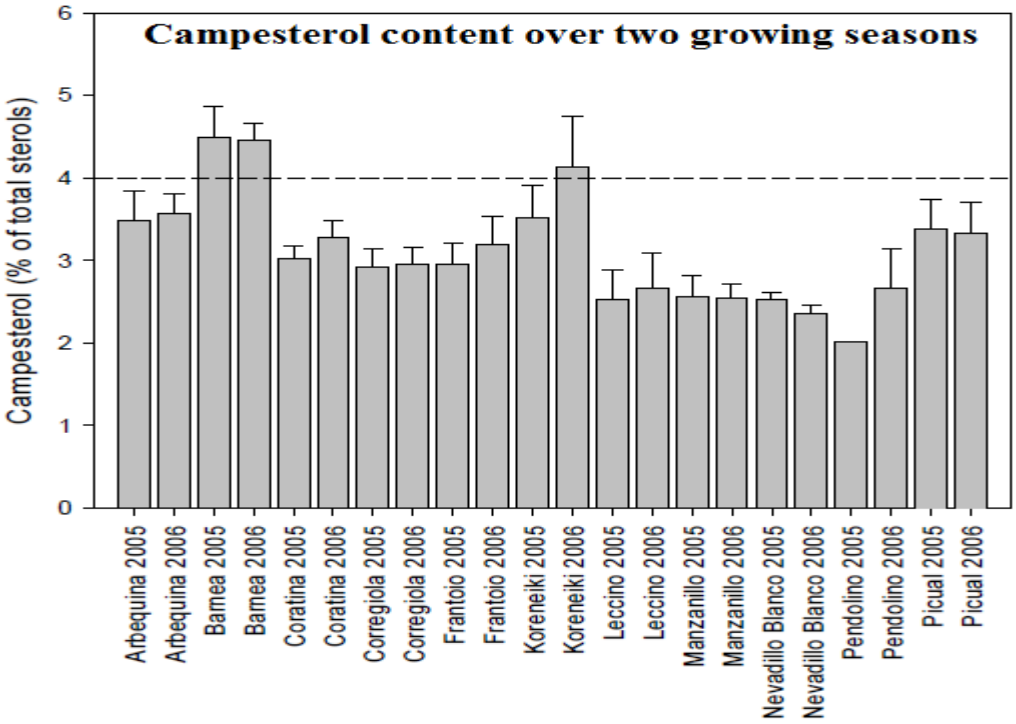


Figure 1.10 Campesterol content (% of total sterols) over two growing seasons (2005 and 2006) (Mailer *et al.* 2008)
Each bar indicates the mean for the cultivar ± standard deviation.
IOOC limit (<4%) is indicated by the dotted line.

These results imply that a large quantity of high quality and authentic extra virgin olive oil (40-45%) being generated in Australia could not be classified as extra virgin olive oil, mainly due to naturally high levels of campesterol in many Australian olive oils. In order to prevent the rejection of these oils on the international market due to non-compliance with the IOOC standards, high quality oil with exceptional characteristics such as organoleptic quality and oxidative stability are now being blended with inferior quality oil to achieve compliance with these trade standards.

However this situation is not limited to Australia as other olive producing countries such as Chile, Argentina, New Zealand, Spain, France and Italy are also encountering similar issues (Mailer 2007; Mailer *et al.* 2008).It has been observed that the olive cultivar Cornicabra, which accounts for more than 14% of total Spanish production, routinely exceeds 4% campesterol levels (Salvador *et al.* 2001; Salvador *et al.* 2003). Argentinean olive oils especially those derived from the cultivars Barnea and Arbequina have consistently shown to have high campesterol levels (4-5.5%) (Ceci *et al.* 2007; Pardo *et al.* 2011). Comparative extraction trials were carried out on 72 olive

oil samples from the cultivar Koroneiki in Greece which showed an average campesterol content of 4.2% (Koutsaftakis *et al.* 1999).

To ensure the genuineness and authenticity of the product, it is important to impose strict standards in order to eliminate cases of adulteration in olive oil however the IOOC standards are merely a generalized standard for the characterization of olive oil by analytical methods that were primarily developed on and therefore suitable for cultivars produced in the traditional olive oil producing countries. This has led to the creation of trade barriers for producers who are producing genuine olive oil, however their products get rejected on the basis of stringent regulations (Mailer *et al.* 2008). This claim has been backed by research that the composition of the minor components of the olive oil would be affected by various geographical and technological factors like cultivar, climate, crop year, degree of fruit ripeness, temperature during fruit maturation, storage time of fruits prior to oil extraction, processing and extraction methods (Aparicio *et al.* 2002; Ranalli *et al.* 2002; Boskou 2006). Thus natural factors play an important role in determining the quality and authenticity characteristics of olive oil. The survey conducted by Mailer and Ayton (2008), clearly showed that campesterol levels in Australian cultivar Barnea remains consistently high across seasons, sites and harvest times, but the impact of horticultural and processing practices on sterol levels of Australian olive oils such as fruit size, irrigation, fruit maturity, malaxing time, malaxing temperature, delays between harvest and process and storage time has not been undertaken. It would be interesting to see if these processing practices affect the campesterol levels in this cultivar.

In order to prevent unfair trading practices, many organizations around the world including FOSFA (Federation of Oils, Seeds and Fats Associations), the European Community, USDA (United States Department of Agriculture) and others within individual countries have set regulations which can restrict the trade of genuine high quality product if natural product variability is not considered (Mailer *et al.* 2008). Recently, in July 2010, the members of Standards Australia arranged a meeting with the members of AOA and other relevant stakeholders (retailers, growers, consumer associations) in order to develop Australian standards for olive oil which will prevent unscrupulous producers to make profits at the expense of local producers who are struggling to compete and also provide a level playing field for growers (Bustos

2010). The Australian standards for olive oil are similar to IOOC standards except it allows a higher level of linolenic acid and campesterol reflecting the actual properties of Australian olive oil (www.standards.org.au). New tests have also been included that will allow the traders to identify fresh oils from old/refined oils.

The Codex Australia (http://www.daff.gov.au/agriculture-food/codex) has been made aware of these concerns and since then Codex Australia has mounted a progressively stronger case against the adoption of IOOC Standards by Codex Alimentarius. It is important to note that Australia accounts for only 0.07% of the world's production and no more than 0.05% of the world's exports (Field 2009). Thus, due to Australia's insignificance by volume in the world's export market, it would be a more difficult task for Codex Australia to negotiate changes to the trading standards. Therefore in order to solve this legislation problem, more research must be undertaken to persuade the Codex Alimentarius about the high quality of Australian olive oils and convince them that such high campesterol levels in our Australian cultivars are merely due to strong influence by genetic factors and growing seasons and they bear no relationship with adulteration or oil quality.

1.8 THE PHYTOSTEROL CONTENT IN OLIVE OIL

Phytosterols are an essential component of plant membranes and theyplay an important role in regulating plant membrane fluidity and permeability(Benveniste 2002).In addition, sterols are an essential part of cellular signalling systems as well asbeing general precursors in hormone biosynthesis.

Phytosterols resemble cholesterol both in structure and biological function (Piironen *et al.* 2000) (Figure 1.11).In vertebrates, cholesterol is the unique sterol and ergosterol is the major sterol in most fungi while most higher plants contains a complex mixture, in which cholesterol is a minor component and 24-ethyl sterols (sitosterol and stigmasterol) account for more than 60% and 24-methyl sterols (campesterol) accounts for less than 40% (Benveniste 2002; Benveniste 2004; Trautwein *et al.* 2007).

All phytosterols are made up of a steroid nucleus, 5,6-double bond and a hydroxyl group at carbon 3 in the β-position (Figure 1.11). The major differences in structure between different sterols are mainly found in their alkyl side chains, for example, cholesterol resembles the structure of phytosterols but it comprises of 8 carbon atoms in its side chain, while phytosterol side chains contain 9-10 carbon atoms (Piironen *et al.* 2000). Some phytosterols can vary by the presence or absence of a methyl or ethyl group at carbon-24, saturation and position of double bond and geometry of the substitution at carbon-24 (Trautwein *et al.* 2007).

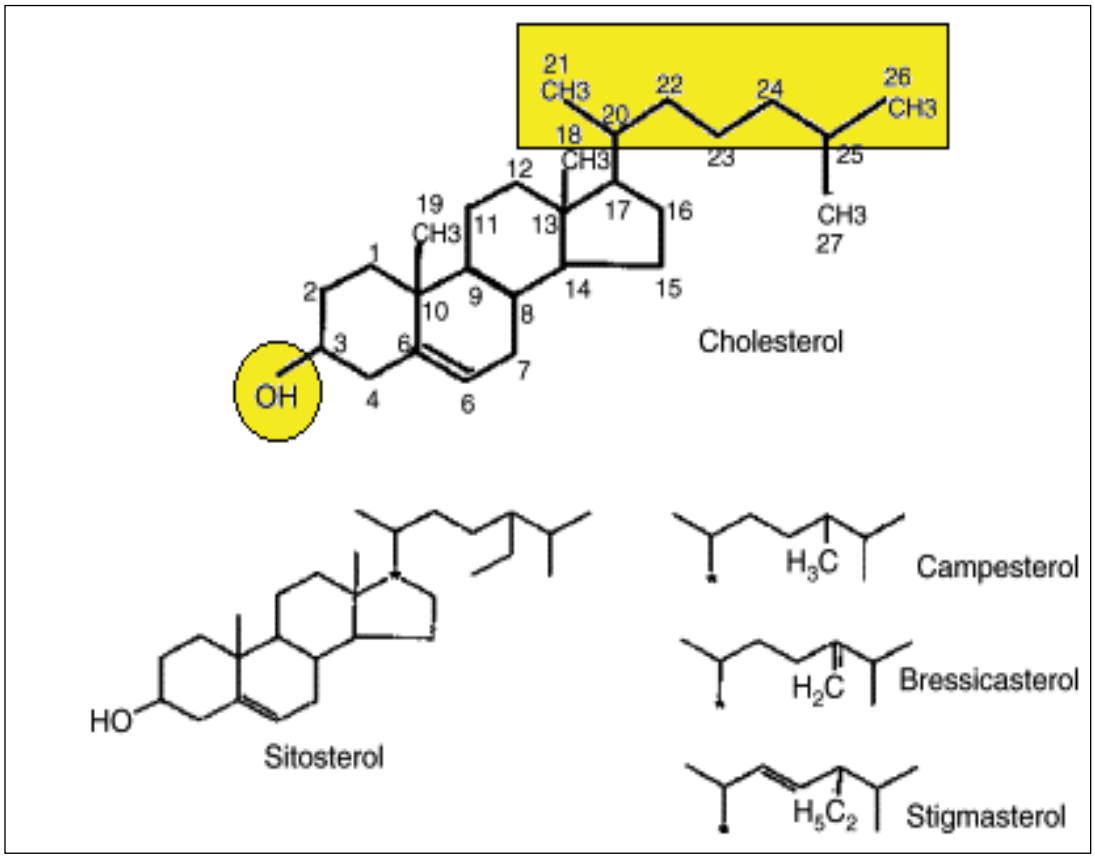


Figure 1.11 Structure of cholesterol and phytosterols (Depicted based on Patel *et al.* 2006).

Based on their structure, plant sterols can be divided into 4-desmethyl sterols, 4α-monomethyl sterols and 4,4-dimethyl sterols (Benveniste 2002). The 4α-monomethyl sterols and 4,4-dimethyl sterols are considered to be intermediates in the formation of 4-desmethyl sterols. The predominant plant sterols are β-sitosterol, campesterol and stigmasterol, which are classified as 4-desmethyl sterols, all of which have double bonds in the side chain (Figure 1.11) (Rudzinska *et al.* 2005). In addition to their role

in cell membranes and regulating membrane fluidity and permeability, campesterol is also the precursor of compounds such as brassinosteroids, which are an important class of hormones involved in higher plant growth and development (Benveniste 2004).

Plant sterols are present in three forms in olive oils: free sterols, sterol esters and sterol glycosides (Piironen *et al.* 2000). Free sterols and to some extent sterol glycosides are integral components of the membrane lipid bilayer and also play an important role in signal transduction. Sterol esters mainly act as a storage form of sterols and are located intracellularly. About 2-3 mg of total (free + esterified) sterols per gram of dry weight is frequently found throughout all plant species (Benveniste 2004).

The principal sterols present in olive oil are β-sitosterol (75-90%), Δ-5-avenasterol (5-20%) and campesterol (2-4.8%), while other sterols present in smaller quantities are stigmasterol, cholesterol, brassicasterol, chlerosterol, ergosterol, sitostanol, campestanol, Δ-7-avenasterol, Δ7-cholestenol, Δ-7-campestenol,Δ-7-stigmastenol,Δ-5,23-stigmastadienol, Δ-5,24-stigmastadienol, Δ-7,22-ergostadienol, Δ-7,24-ergostadienol, 24-methylene cholesterol and 22,23-dihydrobrassicasterol (Boskou 2006; Preedy 2010). According to the European Union Commission (http://ec.europa.eu/index_en.htm), the total content of sterols in virgin olive oils varies between 1000-2000 mg / kg. The sterol content of refined olive oils are lower as the refining process gives rise to significant loses of sterols while sterol content of solvent extracted olive oils is about three times higher than that of virgin olive oils (Boskou 2006). The differences in sterol composition and total sterol content can be attributed to various factors like cultivar, degree of fruit ripeness, crop year, storage time of fruits prior to oil extraction, processing and geographical factors (Koutsaftakis *et al.* 1999; Boskou 2006). As mentioned before (Section 1.5.2), phytosterols in olive oil are used to determine the authenticity of the oil.

1.9 INFLUENCE OF GENETIC FACTORS ON THE STEROL CONTENT OF OLIVE OILS

The sterol content in olive oils seems to be primarily influenced by genetic factors (Section 1.7). In order to study the influence of genetic factors on the composition of sterols in olive oil, it is important to understand the plant sterol biosynthetic pathway in order to identify factors that may influence the relative amounts of these compounds in the olive oil.

1.9.1 Plant sterol biosynthetic pathway

The general sterol biosynthetic pathway in plants has been well characterized by biochemical approaches (Benveniste 2004). As demonstrated by biochemical fractionation studies and by subcellular *in situ* and *in vivo* localization studies of sterol biosynthesis enzymes, plant sterol biosynthesis is known to take place in the endoplasmic reticulum (ER) of the plant cell(Boutte´ *et al.* 2009). However at steady state, sterols get transported from the ER to the plasma membrane via the Golgi apparatus where they mostly accumulate and help inregulating the fluidity and thepermeability of the phospholipid bilayers (Figure 1.12). The isolation of *Arabidopsisthaliana* sterol mutant lines that present altered sterol profiles from the wildtype has provided insight into the respective impact of sterol profiles on the phenotypes of adult plants and has also provided valuable tools for furthercharacterization of sterol biosynthesis and function at the cellular andsubcellular level(Clouse 2002; Benveniste 2004; Boutte´ *et al.* 2009).Molecular studies on these mutants have also demonstrated that correct sterol composition is required for embryonicpattern formation, cell division, cell elongation, cellpolarity, cellulose accumulation and ethylenesignalling pathways(Schrick *et al.* 2002; Souter M *et al.* 2002; Luo *et al.* 2007).

As mentioned before in Section 1.8, higher plants contains a complex mixture of sterols where cholesterol is a minor component and 24-ethyl sterols (sitosterol and stigmasterol) account for more than 60% and 24-methyl sterols (campesterol) account for less than 40%.

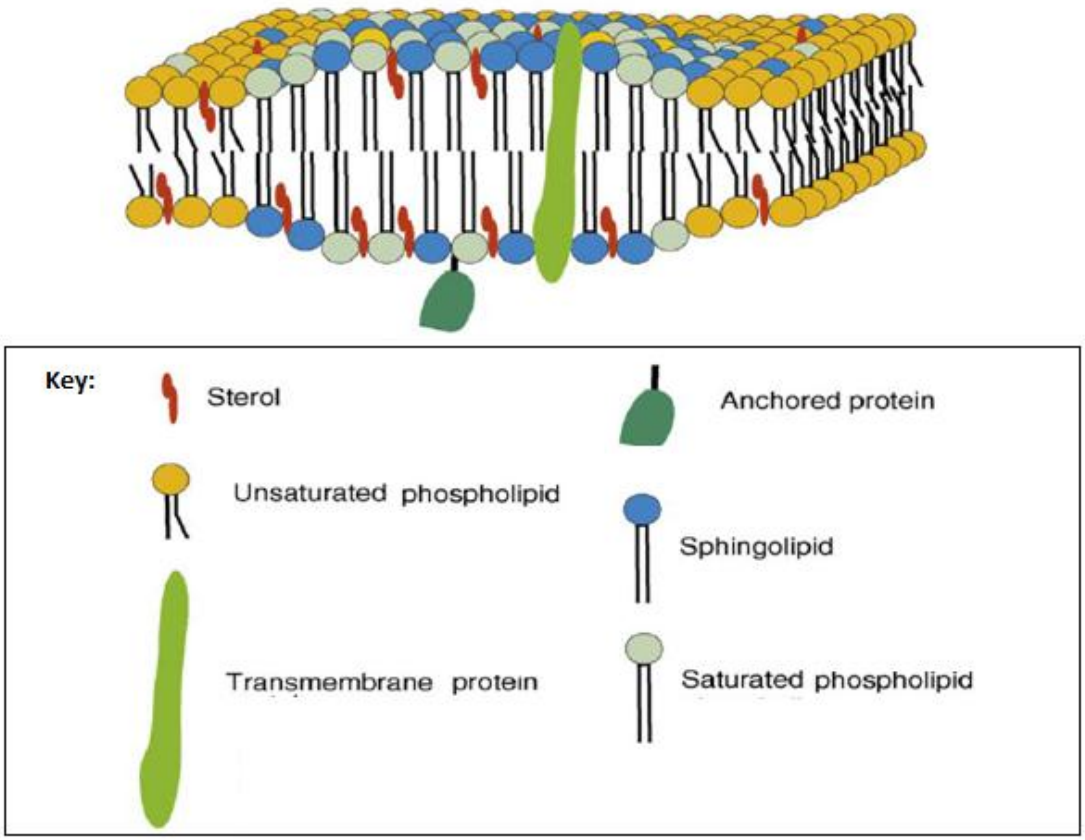


Figure 1.12 Hypothetical sterol enriched membrane raft model in a plasma membrane showing the subcellular localization of sterol binding proteins(Depicted based on Boutte´ *et al.* 2009)

The sterol biosynthetic pathway involves more than 30-enzyme catalyzed reactionsto convert the starting substrate, hydroxyl-3-methyl glutaryl coenzyme A reductase (HMG-CoA) to mevalonate and then to isopentyl pyrophosphate (IPP), which serves as the fundamental building block for the biosynthesis of all the terpenoids (Benveniste 2002). IPP is then converted to squalene which undergoes oxygenation catalyzed by squalene epoxidase to produce 2,3-oxidosqualene (OS). The pathway from IPP to OS is the same in all eukaryotes, however profound differences exist in the biosynthetic segment downstream of OS. In non-photosynthetic eukaryotes, OS is cyclized to form lanosterol, precursor of cholesterol in vertebrates and of ergosterol in most fungi. However in photosynthetic eukaryotes OS cyclase catalyses the conversion of OS to cycloartenol, which is a precursor of phytosterols (Benveniste 2002).

Cycloartenol undergoes methylation to produce 24-methylene cycloartenol which is catalyzed by SMT1 (sterol methyl transferase 1). The passage of 24-methylene cycloartenol to end-product sterols involves the removal of two methyl groups at position 4 and one methyl group at position 14 catalyzed by a distinct SMO (sterol-4α-methyl-oxidase) enzymatic complex (Figure 1.13). This enzymatic complex contains (i)the sterol-4α-methyl-oxidase (SMO) which produces a 4α-carboxylicsterol derivative; (ii)4α-carboxysterol-C3-dehydrogenase/C4-decarboxylase which oxidatively decarboxylases the 4α-carboxylic acid to produce an 3-oxosteroid; and (iii)a 3-keto reductase, which reduces the keto group of the 3-oxosteroidto yield the corresponding monodemethylatedsterol. There are two distinct SMOs, SMO1 and SMO2, which are involved in the removal of the first and the second C4-methyl of the phytosterol precursors respectively. The SMO1 enzymatic complex catalyzes the first oxidative demethylation converting 24-methylene cycloartenol to cycloeucalenol (Figure 1.13). The cyclopropane ring of cycloeucalenol opens catalyzed by cyclopropyl sterol isomerase to produce obtusifoliol. The sterol C14-demethylase is a member of the cytochrome P450 gene family, which catalyzes the oxidative removal of the 14*α*-methyl group of obtusifoliol to give 4α-methyl ergostatrienol. This product is then reduced to form 4α-methyl ergostadienol by the enzyme Δ14-reductase. After the removal of the 14α methyl group and reduction of the 14 double bond, 4α-methyl ergostadienol is isomerized to form 24-methylene-lophenol.

In plants, the sterol pathway is essentially linear from cycloartenol until it reaches 24-methylene-lophenol, with a bifurcation downstream of this compound resulting in two pathways leading to the formation of sitosterol and campesterol, with the sitosterol pathway predominating in activity (Figure 1.13) (Benveniste 2004). This key intermediate, 24-methylene-lophenol, can undergo methylation as catalyzed by the enzyme SMT2 (SAM-24-methylene-lophenol-C-24-methyltransferase), to produce 24-ethylidiene lophenol. At this step the process of second oxidative demethylation takes place but the reaction is catalyzed by the SMO2 enzymatic complex which converts 24-ethylidiene lophenol to Δ7-avenasterol. Alternatively, the substrate 24-methylene-lophenol can undergo oxidation catalysed by the SMO2 enzymic complex to produce episterol. Δ7-avenasterol or episterol is then converted to 5-derhydroavenasterol or 5-dehydroepisterol, respectively, catalyzed by C-5-desaturase (STE1). These products are then converted to isofucosterol or 24-methylene

cholesterol catalyzed by Δ-7 reductase (DWF-5) and eventually to sitosterol or campesterol catalyzed by Δ-24 sterol reductase (DIM)(Figure 1.13) (Piironen *et al.* 2000; Benveniste 2002). Campesterol is also the precursor of compounds such as brassinosteroids, which are an important class of steroidal plant hormones involved in higher plant growth and development. The brassinosteroid pathway has been reviewed by others and will not be covered in this chapter (Yokota *et al.* 1997; Jang *et al.* 2000).

Thus, it has been observed that the orientation of the sterol biosynthetic flux towards β-sitosterol or campesterol is mainly controlled by the two branchpoint enzymes, SMT2 and SMO2 in plants (Schaeffer *et al.* 2001).

1.9.2 Sterol-adenosyl-L-methionine-(Adomet)-sterol-C-24-methyltransferase (SMT)

1.9.2.1 The structure of SMT

SMTs are members of the methyltransferase superfamily which are characterized by the presence of a methyl transferase domain and an S-adenosyl-L-Met binding region(Carland *et al.* 2010). SMT is a tetramer which is composed of four identical subunits. The three dimensional structure of AdoMet-dependent methyl transferase enzymes shows a folding pattern with the central parallel β-sheet surrounded by α-helices (Figure 1.14) (Nes 2000). SMTs catalyse the crucial C-methylation reactions where the enzymes bind to a Δ24-sterol acceptor molecule and a methyl donor AdoMet cofactor to produce a 24-methylated (ethylated) sterol product and Ado Hcy(Song *et al.* 2007).

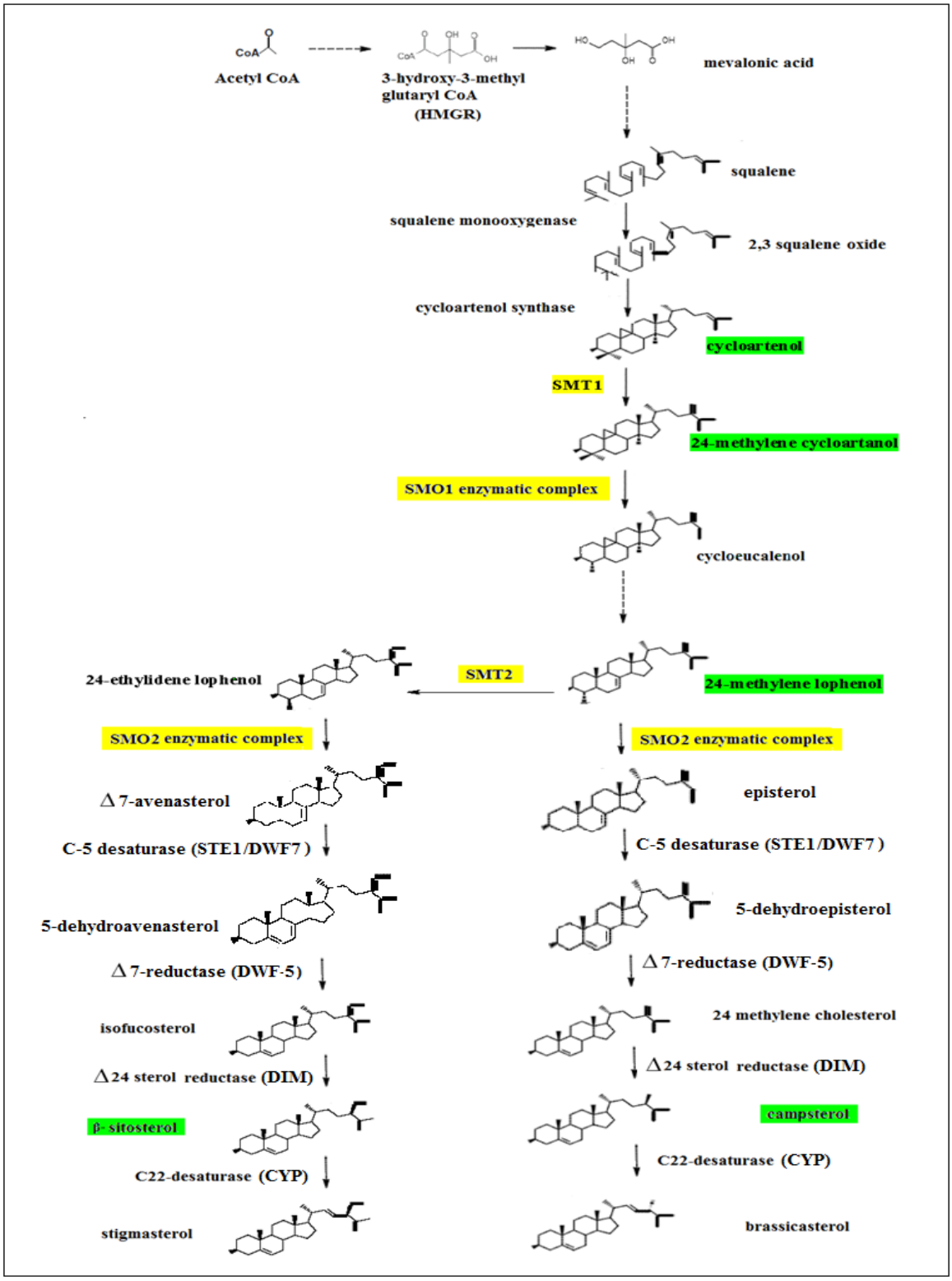


Figure 1.13 Biosynthetic pathway for plant sterols (Schaller 2004)
Schematic diagram of the biosynthetic pathway of acetyl CoA to 24-methyl and 24-ethyl sterols. The key enzymes of the pathway are highlighted in yellow and their respective substrates and end products are highlighted in green. SMT1 (cycloartenol-C-24-methyltransferase), SMT2 (SAM-24-methylene-lophenol-C-24-methyltransferase), SMO (sterol 4α methyl oxidase).

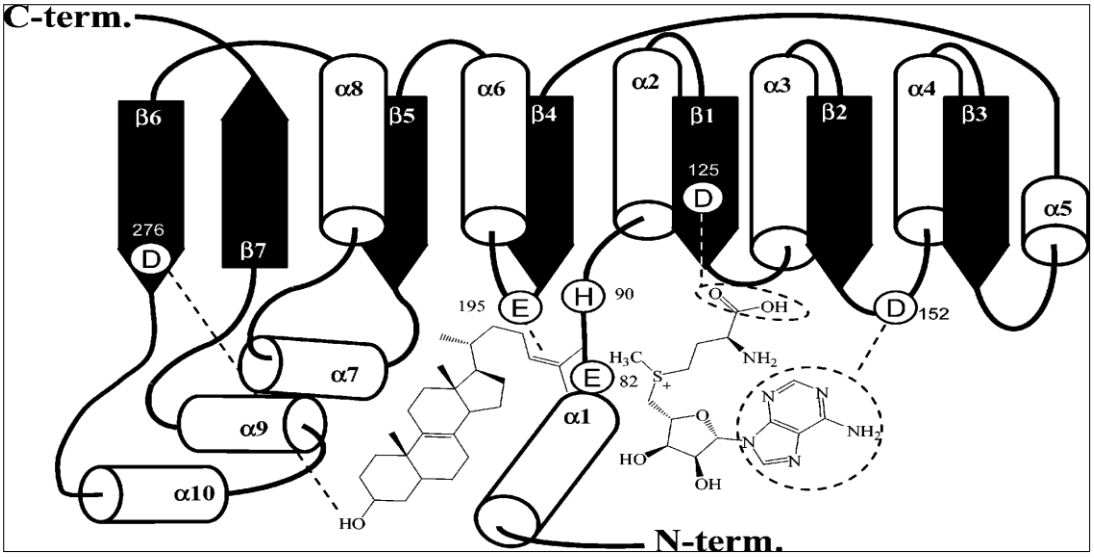


Figure 1.14 SchematicrepresentationoftheSMT structure basedonknowncrystallographicstructures detectedbythefold-recognition analysis (Depicted based on Nes *et al.* 2004); spatialarrangementofthesecondarystructureelements inrelationtosterolandAdoMetsubstrates.

Previous research on SMTs cloned from fungi, protozoa and plants has shown that these enzymes are active at pH that ranges from 6 to 8 with an optimum pH of 7.5, theoretical pI values that range from 5.4 to 7.5and predicted molecular masses of the monomeric SMTs that range from 38 to 43 kDa (Song *et al.* 2007).

1.9.2.2 The catalytic activity of SMTs

In higher plants, S-adenosylmethionine (Adomet)-sterol-C-methyltransferases (SMTs) are believed to be ER-localised enzymes whichare responsible for the addition of an extra methyl or ethyl group to the carbon-24 of the sterol side chain(Bouvier-Navé *et al.* 1997). This enzyme catalyses the conversion of cycloartenol, theubiquitous C30 intermediate of the phytosterol pathway, to produce a mixture of 24-methyl(ene) sterols and 24-ethyl(idene) sterols. In *S.cerevisiae*, the SMT convertszymosterol into fecosterol (Moore *et al.* 1969), but SMT is involved in two distinct transmethylation reactions in plants. According to the substrate specificity studies, cycloartenol is the substrate of the first methylation reaction, resulting in 24-methylene-cycloartanol catalysed by SMT1 enzyme, whereas 24-methylene-lophenol is the preferred substratefor the second methylation, yielding 24-ethylidene-

lophenol catalysed by SMT2 enzyme, finally resulting in the production of β-sitosterol and stigmasterol and away from brassinosteroid biosynthesis (Benveniste 2002).

1.9.2.3 SMTs in plants and microorganisms

The *SMT* gene *(ERG6)* was first cloned from *S. cerevisiae*and this *ERG6* expression system was used for the functional characterisation of the *SMT* gene. The results demonstrated that two different SMT enzymes, SMT1 and SMT2, are involved in catalysing the two distinct methyltransfer reactions in the sterol biosynthetic pathway (Bouvier-Navé *et al.* 1997; Bouvier-Navé *et al.* 1998; Schaeffer *et al.* 2000; Benveniste 2002).

Molecular cloning of *SMT*geneshas been achieved in a number of plant species including *Arabidopsis thaliana, Glycine max, Gossypium hirsuturm, Nicotiana tabacum, Oryza sativa, Ricinuscommunis* and *Zea mays*(Husselstein *et al.* 1996; Shi *et al.* 1996; Grebenok *et al.* 1997; Schaeffer *et al.* 2000; Luo *et al.* 2008; Neelakandan *et al.* 2010).

The *Arabidopsis* sequencing project confirmed that the *Arabidopsis* genome contains three distinct SMT genes (*At*5g13710, *At*1g20330, *At*1g76090) encoding *SMT1*, *SMT2* and *SMT3* (Diener *et al.* 2000; Benveniste 2004). SMT1 enzyme, located in the vacuole of higher plants, is the most homologous with yeast *ERG6* and it catalyses the initial methyl transfer reaction and also serves as a branch point enzyme between cholesterol and the more abundant sterols (Diener *et al.* 2000). SMT2 and SMT3, located in the endoplasmic reticulum of the higher plants, catalyse the second methyl transfer reaction which determines the flux between the β-sitosterol and campesterol segment of the pathway (Bouvier-Navé *et al.* 1997). Sequence alignment of *SMT1* and *SMT2* proteins has revealed four highly conserved domains of the SMT enzyme. The first domain is Region II (LDXGCGXGGPXRXI) which corresponds to the G-rich consensus motif described for all AdoMet-dependent methyltransferases. SMT2 contains three more domains namely, Region I (YEW/F/YGWGXSFHF), Region III (IEATCHAP) and Region IV (KPGXXF/YXXYEW) domains which are believed to be typical of methyltransferases acting on a sterol substrate (Benveniste 2002; Benveniste 2004). Also SMT2 possess a hydrophobic domain at the N-terminal position which has approximately 25 amino acids, while SMT1 are devoid of such a hydrophobic domain (Bouvier-Navé *et al.* 1997). Two full length *SMT2* cDNAs of 1421bp and 1249bp [*AtSMT2*-

1 (Accession: X89867 and *AtSMT2-2* (Accession: U71400respectively], encoding polypeptides of 361 and 351 amino acids respectivelyhave been cloned and isolated from *Arabidopsis.* The two *At*SMT2gene families have been found to have 78.38 % similarity at nucleotide level and 82% sequence identity at amino acid level.

To study the effects of β-sitosterol and campesterol on the development of cotton fibres (*Gossypium hirsuturm*), two key *SMT2* genes (*GhSMT2-1* and *GhSMT2-2)* were isolated from developing fibres of upland cotton (Luo *et al.* 2008). The full length cDNA of *GhSMT2-1* was 1151bp, which consisted of a 8bp 5' UTR, a 1086bp ORF and a 57bp 3'UTR. The full length cDNA of *GhSMT2-2* was 1166bp, which consisted of a 18bp 5' UTR, a 1086bp ORF and a 62bp 3'UTR. Both the *GhSMT2* genes encoded a polypeptide of 361 amino acids and with a predicted molecular mass of 40kDa. The two *Gh*SMT2 genes shared 86.2% similarity to each other at nucleotide level. Southern blotting analysis was conducted to investigate the copy number of *SMT2* genes which has revealed that *GhSMT2-1* or *GhSMT2-2* is a single copy gene in upland cotton. To study the role of *SMT2* genes in the growth and development of cotton plants, expression profiles of the two *GhSMT2* genes in different organs and at various stages of fibre development were investigated. For both *GhSMT2-1* and *GhSMT2-2* genes, highest expression levels were detected in 10 DPA (day post anthesis) fibers, and the lowest expression levels were observed in leaves and cotyledons. The analysis revealed that the expression level of *GhSMT2-1* was10 times higher than *GhSMT2-2* in all organs and tissues detected. While both *GhSMT2* genes play a crucial role in fibre elongation in upland cotton, the effects of *GhSMT2-2* onfibre growth appeared weaker. This could be due to loss of activity of duplicated genes during the polyploidization of this species (Luo *et al.* 2008).

Two full length SMT2 cDNAs (*GmSMT2-1* and *GmSMT2-2)* have been cloned and isolated from soybean (*Glycine max*), containing ORFs of 1086bp and 1092bp and encoding polypeptides of 361 and 363 amino acids, respectively (Neelakandan *et al.* 2009). The two *Gm*SMT2 genes shared 89% similarity to each other at nucleotide level. Comparison of their amino acid sequences revealed 96% sequence identity at amino acid level with only five differences at amino acid positions 175, 183, 261, 262 and 263 which lie outside the active site regions. These SMT2 proteins were functionally expressed in *E.coli* and their enzyme activity was determined. Both soybean SMT1 and SMT2 had distinct substrate specificities and product outcomes, where *Gm*SMT2-1 and *Gm*SMT2-2 had substrate specificity for 24-methylene lophenol, while *Gm*SMT1 had substrate specificity for cycloartenol similar to that

observed in *At*SMT enzymes (Diener *et al.* 2000). Steady state kinetic analysis were performed using Michaelis-Menten equation which revealed *Gm*SMT2-1 and *Gm*SMT2-2 generated k_{cat} values of 0.8 min^{-1} and 1.34 min^{-1} respectively while *Gm*SMT1 generated a K_{cat} value of 0.6 min^{-1}. This indicated that the catalytic efficiency of *Gm*SMT2-2 for the natural substrate is slightly faster than *Gm*SMT2-1 whereas the catalytic efficiency of *Gm*SMT2-1 is more similar to the catalytic efficiency of *Gm*SMT1. Previous research has suggested that the first methyl transfer step catalysed by SMT1 is rate-limiting, thus offering venues for "fine" control of sterol composition in phytosterol biosynthesis (Schaeffer *et al.* 2000; Neelakandan *et al.* 2010). Further, expression analysis of SMT2-1,SMT2-2 andSMT1in leaf, roots and seeds and at various stages of growth of soybean was investigated. It was shown that SMT1 and SMT2 were constitutively expressed in all tissues during the life history of soybean, however expression levels declined to very low levels in the latter stages of seed maturation. Unlike *Arabidopsis*, where the expression of *At*SMT2-2 was generally lower, in soybean, SMT2-2 was expressed generally at higher levels than SMT2-1 orSMT1 in all tissues, suggesting different requirements for the SMT2 isoforms in soybean physiology (Neelakandan *et al.* 2009).

In another study the transcriptional regulation of the *GmSMT2-1* and *GmSMT2-2* genes in soybean seedlings were monitored under different environmental conditions such as cold, dehydration, salt and abscisic acid (Neelakandan *et al.* 2010). The results suggested that environmental conditions induce the differential changes in transcript levels of *GmSMT2* genes to some extent and the two *GmSMT2* genes work in parallel to overcome these environmental stresses.

To shed additional light on the transcriptional regulation of soybean *SMT2* genes, the upstream promoter region of the *GmSMT2* geneswere cloned and characterised (Neelakandan, Nguyen et al. 2010). A number of important *cis*-elements and transcription factor binding sites were identified for both *Gm*SMT2-1 and *Gm*SMT2-2 enzymes based on homology to previously characterized regulatory motifs in plants, however, the upstream sequence of *Gm*SMT2-2 had few additional *cis*-elements such as TGA-element (AACGAC) for auxin responsiveness,TATC-box (TATCCCA) for gibberellin responsiveness,T/G box (AACGTG) for induction by Jasmonic acid and RY repeat or legumin box (GTGCATG). The RY repeat or legumin boxcommonly found in legume storage protein gene promoters are responsible for high levels of expression in seeds, which may explain the high level of

expression of *SMT2-2* gene insoybean seed tissues as observed in previous studies (Neelakandan *et al.* 2009; Neelakandan *et al.* 2010).

1.9.2.4 Evidence supporting the functional role of the *SMT* gene families in determining the sterol content in plants

Research has provided direct evidence that *SMT1* and *SMT2* activities in plants play an important role in the accumulation of cholesterol, campesterol and sitosterol and also helps in balancing the morphogenetic effects upon growth and development in these plants.

Modulation of the expression of the *SMT1* gene has been achieved in *Arabidopsis*(Diener *et al.* 2000) and in tobacco (Schaeffer *et al.* 2000). The *SMT1* cDNA was isolated from a tobacco callus library and transformation in tobacco was done via *Agrobacterium* using a transfer DNA (T-DNA) vector. Mutants lacking this enzyme were used to identify the function of *SMT1* gene. Results showed that overexpression or co-suppression of *SMT1* led to concomitant depletion or accumulation of cycloartenol, the substrate of the enzyme. In contrast, in *Arabidopsis* a null-*SMT1* mutant accumulated cholesterol and not cycloartenol, at the expense of sitosterol, which indicated that side-chain methylation of cycloartenol is an essential step prior to further metabolization. Also the modulation of *SMT1* gene expression was reported in tobacco seeds using a seed specific *SMT1* expression cassette (Holmberg *et al.* 2002). Co-suppression of the *SMT1* gene led to the decrease in cycloartenol level, while overexpression of the gene resulted in increase in the level of sitosterol by 50%, while levels of cycloartenol and cholesterol decreased by 53% and 34% respectively. These results suggested that SMT1 controls the level of cholesterol in plants (Diener *et al.* 2000) and also supports the sub-functions of these enzymes where SMT1 catalyses the earlier methyl transfer step (in a different compartment) while SMT2 catalyses the later methyl transfer step.

In order to understand the physiological function of SMT2 in *Arabidopsis*, the expression of this gene was modulated in *Arabidopsis* transformed with the cDNA encoding *AtSMT2-1*(Schaeffer *et al.* 2001). Northern blotting analysis revealed that the plants from this set of transgenic lines could be classified into two groups: high sitosterol (hs) plants and high campesterol (hc) plants (Figure 1.15). Plants overexpressing the transgene showed a concomitant increase of the amount of sitosterol (65-70%) and a parallel drop in the amount

of campesterol and brassicasterol (from 11 to 4% and from 2 to 0.5% respectively). Such plants were designated as hs plants and their phenotypic characterization displayed a reduced stature and growth rate which could be restored by brassinosteroid treatment. Plants showing co-suppression of *SMT2-1* revealed a dramatic accumulation of campesterol at the expense of sitosterol. These plants were designated as hc plants and showed phenotypic traits such as reduced growth, increased branching and drastically reduced fertility. In case of hc plants, these abnormalities in their phenotype could not be restored by exogenous brassinosteroids (Schaeffer et al., 2001). Therefore it was shown that *SMT2-1* plays a very crucial role in balancing the ratio of campesterol to sitosterol in plants. It is also worth noting that in mutants where *SMT2* was co-suppressed, showed an accumulation of 24-methylene-lophenol, the substrate of *SMT2*(Bouvier-Navé *et al.* 1997). This would indicate that in sterol biosynthesis, the subsequent metabolisation of 24-methylene-lophenol by *SMO2* is probably a slow step. Thus, in high campesterol plants, it is plausible that an increase in the expression of *SMO2* would enhance the biosynthetic flux in the 24-methyl sterol segment of the pathway.

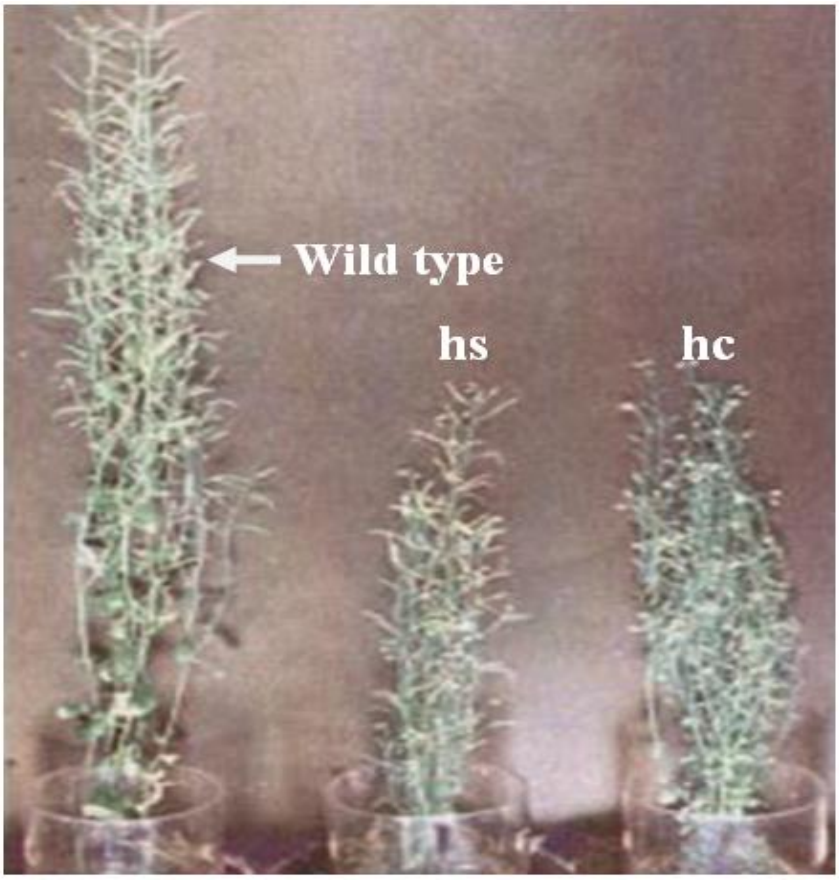


Figure 1.15 Arabidopsis phenotypes transformed with *SMT2-1*(Schaeffer *et al.* 2001) Developmental modifications of greenhouse grown *Arabidopsis* plants in case of overexpression (hs plants) and co-suppression (hc plants) of *SMT2-1* compared with the wild type.

Similarly, in case of tobacco plants transformed with *SMT2-1* construct, an overexpression of *SMT2-1* resulted in a dramaticincrease of β-sitosterol due to elevated 24-methylenelophenol-C24-methyltransferase activity at the expense of 24 methyl sterols (campesterol) (Schaeffer *et al.* 2000).

As mentioned earlier in the previous section (Section 1.9.2.3), two full length SMT2 cDNAs (*GmSMT2-1* and *GmSMT2-2)* were cloned and isolated from soybean (*Glycine max*) and kinetic studies revealed *GmSMT2-2* had a higher catalytic efficiency among the two, exhibiting regulated transcript abundance in developing soybean seeds (Neelakandan *et al.* 2009). To further understand the physiological function of SMT2 in soybean sterol biosynthesis in regulating the metabolic flux, a set of transgenic *Arabidopsis* plants harbouring seed targeted overexpression of *GmSMT2-2* (abbreviated as SM2), *GmSMT1* (SM1)and in combination with *AtHMGR1* genes (SH) were prepared. A fourth construct was also prepared where all the three genes, *Gm*SMT2-2, *Gm*SMT1 and HMGR1 were overexpressed (SHS)(Neelakandan *et al.* 2010). The results revealed a higher level of SMT1 and SMT2 expression in genetically modified SHS seeds, leading to a maximum increase in the total sterol content as compared to the other constructs.The alteration in sterol content was the least in SM2 seeds. Overall in the transgenic seed constructs, sitosterol and stigmasterol levels increased, cycloartenol levels decreased and campesterol levels were more or less same, except in the SHS seeds, where there was a reduction. The campesterol levels seemed to be tightly regulated in soybean seed tissues since the sitosterol:campesterol ratio were not drastically altered. This is in contrast with the *Arabidopsis*and tobaccomutants, which showed that overexpression of SMT2-1 enzyme leads to a significant drop in campesterol levels. The interpretation of this data implied that SMT2 genes play a key role in fine tuning the end-product C-24 ethylsterol (sitosterol and stigmasterol) levels in soybean sterol biosynthesis(Neelakandan *et al.* 2010).

Carland and Nelson (2010) conducted another study recently to investigate the effect of deficiency of *SMT2* and *SMT3* in *Arabidopsis* mutants. It was observed that in the *SMT2* mutants, there were moderate developmental defects, including serrated floral organs, reduced stature and aberrant cotyledon vein patterning, which suggested that the loss of SMT2 can be at least partially complemented by SMT3. A transcript null *SMT3* mutant was identified in order to test the developmental roles of SMT2 and SMT3. Although *SMT3* null mutants appeared as wild type, *SMT3 SMT2* double mutants showed enhanced defects in

comparison to *SMT2* null mutants, such as defective root growth, loss of apical dominance, sterility, homeotic floral transformation and discontinuous cotyledon vein pattern. It was also noticed that although the sterol content in these mutants was dramatically affected the BR levels remained undisrupted, which suggests that the vascular patterning defects were due to deficiency in particular sterols, rather than in BRs (Carland *et al.* 2010).

Recently a study was conducted to investigate the downregulating effects of the three BR-biosynthetic enzymes DET2, DWF4 and SMT2 from *Arabidopsis thaliana,* by simultaneous suppression of the three genes with an RNA interference construct (Chung *et al.* 2010). DET2 and DWF4 are key enzymes of brassinosteroid synthesis in *A. thaliana* in the downstream segment of the pathway (Benveniste 2004). Results showed that triple knock-down transgenic plants displayed a dwarf phenotype with altered development, which partially resembled BR-deficient mutants. The transcript levels of the *DET2, DWF4* and *SMT2* genes were quantified using real-time quantitative PCR (qRT-PCR) and the results showed that the mRNA levels of the *DET2, DWF4* and *SMT2* genes from the transgenic plants were down-regulated up to 20, 56 and 52%, respectively. The campesterol content of the triple knock-down transgenic plants was 4.2 fold higher than that of the wild type plants, while the sitosterol and stigmasterol contents were reduced to 35 and 31% respectively (Table 1.11). This result clearly indicates that due to the knocking down of the *SMT2* gene, the substrate 24-ethylidene-lophenol was used up by the *SMO2* enzyme in the 24-ethyl segment of the pathway, leading to an elevated campesterol level in these dwarf mutants. Simultaneous suppression of these three genes has also resulted in reduced formation of BR-specific biosynthesis intermediates in these transgenic plants, where the levels of 6-deoxocathasterone, 6-deoxoteasterone,3-dehydro-6-deoxoteasterone, 6-deoxotyphasterol,6-deoxocathasterone,typhasterol, and castasterone weredecreased to 19, 50, 15, 43, 19, 38, and 54%, respectively, in comparison to their levels in wild- type plants. Thus, it was observed in this study that activities of the *SMT2* gene and the downstream genes in the BR pathway play an important role maintaining campesterol levels in plants.

In view of these coarse and fine control points in phytosterol biosynthesis, past research has firmly established that SMT2 biosynthesis step is a crucial regulatory point that can drastically impact the levels of sitosterol and campesterol in plants.

1.9.3 Sterol 4*α*-methyl oxidase (SMO)

1.9.3.1 The structure of SMO

SMO belongs to a small family of membrane-bound non-haem iron oxygenases located in the endoplasmic reticulum of the cells that are involved in lipid oxidation (Darnet *et al.* 2004). Extensive work has been reported on the SMOcomponent of the enzyme complex (Section 1.9.3.3) (Bard *et al.* 1996; Darnet *et al.* 2003). Hydropathy plots derived from the SMO enzymes isolated from different species (Section 1.9.3.3) revealed at least three and possibly four large hydrophobic regions and a C-terminal hydrophilic region are present in SMO. The SMO amino acid sequences of all species studied possess atleast three conserved histidine-rich motifs (HX3H, HX2HH and HX2HH) (Bard *et al.* 1996). The function of these tripartite motifs may be to provide the ligands for a presumed catalytic di-iron centre asproposed previously for other enzymes possessing similar motifs (Darnet *et al.* 2004).

1.9.3.2 The catalytic activity of SMO

In higher plants, the conversion of cycloartenol into functional phytosterols involves the removal of two methyl groups at C-4 and one methyl group at C-14 catalysed by a SMO enzymatic complex (Section 1.9.1). This enzymatic complex consists of sterol 4α-methyl oxidase (SMO), 4α-carboxysterol-C3-dehydrogenase/C4-decarboxylase and a 3-keto reductase. The oxidation step is performed by SMO which stepwise converts the methyl group to an alcohol, an aldehyde and finally to a 4*α*-carboxylic sterol derivative. Subsequently, this acid is oxidatively decarboxylated by 4*α*-CD (4*α*-carboxysterol C3dehydrogenase/C-4-decarboxylase) to produce a 3-oxosteroid, which is finally stereospecifically reduced by an NADPH-dependent SR (sterone reductase) to give the corresponding monodemethylated sterol (Figure 1.13) (Darnet *et al.* 2004).

In animals and fungi, the 2 methyl groups are removed sequentially, while in higher plants, two distinct oxidation systems, SMO1 and SMO2 are involved in the removal of the first and second C4-methyls of phytosterol precursors (Benveniste 2004). The preferred substrates of the first demethylation step is 24-methylene-24-dihydro-cycloartenol which is catalysed by

SMO1 enzymatic complex whereas the substrate for the second demethylation step is 24-ethylidene-lophenol which is catalysed by SMO2 enzymatic complex (Section 1.9.1). In both animal and plants, membrane-bound cytochrome b5 was shown to be an obligatory electron carrier from NAD(P)H to the SMO.

1.9.3.3 SMO in plants and microorganisms

The *Saccharomyces cerevisiae* sterol-C4-methyl-oxidase (*Sc*SMO) gene, *ERG25* has been cloned and isolated revealing a 309 amino acid polypeptide with a molecular mass of 36.48kDa (Bard *et al.* 1996). Two putative TATA motifs are present at 182bp and 84bp along withthree histidine motifs H^{59}RLFH, H^{172}KQHH, andH^{256}HDLHHH(Bard *et al.* 1996).A typical dilysine motif (KKXX) for endoplasmic reticulum retention is present at the COOH terminal of the enzyme.

Cloning and characterisation of the genes encoding C-4 sterol methyl oxidase has been achieved in a number of other species including *Glomus intraradices, Neurospora crassa, Candida albicans, Penicillium chrysogenum* and *Aspergillus* species (Wang *et al.* 2008; Oger *et al.* 2009). SMO protein sequences from all species revealed a characteristic hydrophobicity profile (Figure 1.16), a typical dilysine motif (KKXX) and a pattern of histidine-rich motifs which are typical of C-4 methyl sterol oxidases from *Sc*SMO.

In plants, cloning and functional characterisation of sterol 4α-methyl oxidases have been achieved infew species such as *Arabidopsis*, maize, and tobacco(Darnet *et al.* 2001; Chan *et al.* 2010).In *Arabidopsis thaliana*, multiple *SMO* genes have been identified, where five different *SMO* sequences belong to two distinct gene families SMO1 and SMO2 based on substrate specificity. SMO1 contains three isoforms [*At*SMO1-1 (*At4g12110)*, *At*SMO1-2 (*At4g22756)*, *At*SMO1-3 (*At4g22753)*] while SMO2 contains two isoforms[*At*SMO2-1 (*At1g07420)* and *At*SMO2-2 (*At2g29390*)](Darnet *et al.* 2001). The protein sequences of SMO1 and SMO2 in *Arabidopsis thaliana* showed 29-32% and 33% identity with the yeast and human SMO respectively, however SMO1 and SMO2 are 89% identical to each other (Darnet *et al.* 2001).Typical of the ERG3-ERG25 family from animal, fungi or plants, both *At*SMOs are characterized by the presence of three histidine-rich motifs, HX^3H, HX^2HH, HX^2HH, which exhibit a topology and spacing of amino acids within the histidine

motifs(Figure 1.16)(Darnet *et al.* 2001).This high sequence identity including the presence and spacing of the three histidine-rich motifs strongly suggests that *SMO1* and *SMO2* in *Arabidopsis thaliana* code for two plant SMO isoenzymes. The gene structure of *At*SMO1s and *At*SMO2s are different where

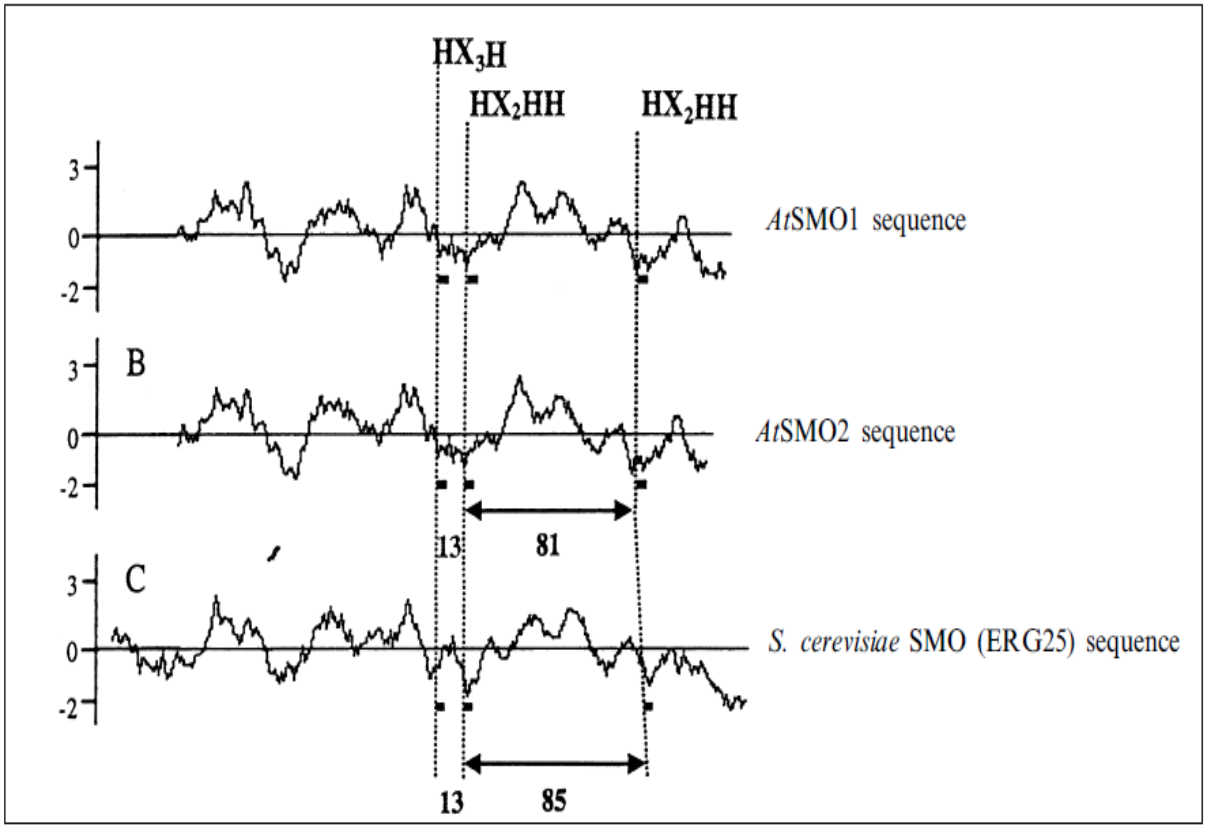


Figure 1.16Hydropathy plots of the *At*SMO1, *At*SMO2 and *S. cerevisiae* SMO
Vertical lines indicate regions corresponding to the three histidine-rich motifs that show homology to the membrane-bound enzymes of the ERG3-ERG25 family. *At: A. thaliana.*(Depicted based on Darnet *et al.* 2001)

*At*SMO1scontains 5 exons and *At*SMO2scontains 7 exons and the size and the conserved number of exons rarely changes within each family. The organization of the coding regions of the two families (*At*SMO1s and *At*SMO2s)are conserved and differs between the two families (Darnet *et al.* 2004). The two full length*At*SMO2cDNAs, *At*SMO2-1 (AF346734)and *At*SMO2-2 (AF327853), which have been cloned and isolated from *Arabidopsis thaliana* cDNA librarycontains an ORF of 801 and 783bp, encoding polypeptides of 267 and 261 amino acids respectively (Darnet *et al.* 2001).

The function of the proteins encoded by the *SMO* cDNAs from *Arabidopsis thaliana* was investigated by complementation of a yeast *ERG25* mutant strain lacking SMO activity in *Saccharomyces cerevisiae*. The *ERG25* product (*Sc*SMO) in *Saccharomyces cerevisiae*, is an

essential enzyme since disruption of *ERG25* is lethal as *ERG25* strain requires ergosterol supplementation in order to maintain viability (Bard *et al.* 1996). This study showed that even in the absence of ergosterol in the medium, *At*SMO1, *At*SMO2 and *Sc*SMO could restore growth of this strain. Also in the yeast erg25 mutant background, following transformation with *At*SMO1 and *At*SMO2, there was absence in the accumulation of 4α-monomethylated sterols which suggests that they catalyze oxidation of both 4,4-dimethyl and 4α-methyl-sterol intermediates. This study showed that *At*SMO1 and *At*SMO2 genes can complement the null, lethal phenotype, thereby confirming the activity of SMOs coded by these genes. This was the first functional identification of *SMO* genes from plants (Darnet *et al.* 2001).

In addition to the functional identification of *SMO* genes from *Arabidopsis*, this study has revealed that the complementation level of *At*SMO2 appeared to be slightly higher than *At*SMO1. This reason could be attributed to their differences in substrate specificityand catalytic efficiency towards 4,4-dimethyl-zymosterol (the physiological substrate of *Sc*SMO). The apparent lower demethylation effciency obtained with the *At*SMO isoenzymes would suggest that *in vivo*, the initial oxidation steps catalyzed by the SMO are partially rate-limiting relative to the overall demethylation process in the *ERG25* mutant background (Darnet *et al.* 2001).

Previous research on the microsomal preparation isolated from *Zea mays* has also suggested that two distinct oxidative systems are involved in the oxidative C4-monodemethylationof 4,4-dimethyl- and 4α-methylsterols(Pascal *et al.* 1993). In contrast to the observed substrate specificityand catalytic efficiency towards 4,4-dimethyl-zymosterol of *At*SMOs, C4-methylated-Δ8-sterolssuch as 4,4-dimethyl-zymosterol were shown to be poor substratesof SMOs isolated from maize. These observations indicate that substrate specificity of SMO is different between *Arabidopsis thaliana* and *Zea mays*,which could possibly reflectthat SMOs from different plants are quite diverse.

A query of the Genbank/EMBL databases for *SMO2*sequences reveals sequences from a few others plant species such as *Gossypium arboretum*(Accession: AF352575.1),*Vitis vinifera* (Accession: XM_002282617.1),*Glycine max* (Accession: NM_001253115.1), *Solanum lycopersicum* (Accession:NM_001246951.1)and *Ricinus communis*(Accession: XM_002520459.1), however their functional characterisation has not been conducted to date (Chan *et al.* 2010).

1.9.3.4 Evidence supporting the functional role of the *SMO* gene families in determining the sterol content in plants

A study was conducted by Darnet and Raheir (2004) to elucidate the precise functions of the *SMO1* and *SMO2* gene families. The expression of these genes were reduced by using a VIGS (virus-induced gene silencing) approach in *Nicotiana benthamiana*, which should result in the accumulation of the substrates of the targeted enzymes (Darnet *et al.* 2004; Purkayastha *et al.* 2009). Two cDNA fragments corresponding to *Nt*SMO1 and *Nt*SMO2 were cloned from *Nicotiana tabacum* and were inserted into the viral TTO1 vector, and were used for silencing experiments.

The sterol levels of the tobacco plants infected either with TTO1-SMO1 or TTO1-SMO2 were analysed by comparing the sterol levels obtained from control plants. The results indicated that following the silencing with *SMO1*, a significant 3-fold relative increase in the total level of 4,4-dimethyl sterols was observed when compared with the different series of control and *Nt*SMO2 plants. Importantly, a substantial accumulation of 24-methylene cycloartenol (4-6 fold higher) and cycloartenol (2-4 fold) were obtained, while qualitative and quantitative levels of 4α-methyl sterols were not affected.

In case of silencing of *SMO2*, an accumulation of 4α-methyl sterols was found, with no change in the levels of 4,4-dimethylsterols. The most noteworthy difference observed was the subsequent rapid methylation of transiently accumulated 24-methylene-lophenol by branch point enzyme SMT2, leading to excessive accumulation of 24-ethylidene-lophenol and its reduced derivative 24-ethyllophenol (10-15 fold increase). This subsequently led to the increased accumulation of sitosterol and stigmasterol and decreased accumulation of campesterol. As the substrates at this branch point are not limiting, these distinct biochemical phenotypes suggested that overexpression or downregulation of *SMO2* could play an important role in determining the ratio of sitosterol to campesterol in plants (Darnet *et al.* 2004; Schaller 2004).

1.10 GENETIC RESEARCH IN OLIVES

Cultivated olive species bear hermaphroditic flowers which exhibit various degrees of self-incompatibility and therefore this allogamous mode of reproduction leads to a high level of genetic heterozygosity (Preedy 2010). Olives are diploid in nature containing 2n=2x=46 chromosomes. The nuclear DNA content of various olive cultivars have been determined using flow cytometry which reveals that the olive genome size is estimated to contain between 2.90 and 3.07 billion basepairs (where 1 pg corresponds to approximately 10^9 base pairs) (Bennett *et al.* 1976; Loureiro *et al.* 2007).

Unlike the large agricultural crops of rice (*Oryza sativa*) and maize (*Zea mays*), relatively small amounts of data are available on olive genetic sequences in the GenBank database (Preedy 2010). As of 2008, only 1037 core nucleotide, 24 ESTs (expressed sequence tags) and two GSS (genome survey sequence) were submitted in NCBI (National Center for Biotechnology Information). *Olea europaea ssp.*contains 1029 core nucleotide sequences which have been divided into sequences from genomic untranslated regions and translated regions which approximately contains 608 sequences (59% of total) and 421 sequences (41% of total), respectively (Preedy 2010).

In the past few years high throughput next generation sequencing (NGS) technologies havebecome popular due to their ability to produce complete genome sequences and large sets of expressed sequence tags (ESTs), which simultaneously allows discovery of novel genes as well as the quantitative analysis of the expression of these genes in several tissues and organs (Alagna *et al.* 2009). NGS technology has been used to study the transciptome of olive drupes from two *Olea europaea* cultivars, *Coratina* characterized by a very high phenolic content and *Tendellone* characterized by an oleuropein-lacking natural variant (Alagna *et al.* 2009). Four different cDNA libraries were sequenced from the beginning to the end of the drupe developmental stage (45 DAF and 135 DAF) byusing 454 pyrosequencing technologies (Roche). A total of 261,485 reads were obtained which has provided a large amount of sequence data on expressed genes in these cultivars that play an important role in olive fruit metabolism (Alagna *et al.* 2009). This sequencing data is a valuable resource for gene discovery and characterisation in olives.

A comparative proteomic approach based on two-dimensional electrophoresis and mass spectrometry has been used to study the proteomic profile of olive drupes during development, to reveal modulation in the biosynthesis of compounds related to major quality traits of olives and olive oil (Bianco *et al.* 2013). For this, the total protein content extracted from drupe mesocarp of olive cultivar Coratina at three different developmental stages (45, 110 and 150 DAF) were analysed. The image analysis revealed 247 protein spots as differentially accumulated during olive fruit development. 68 unique proteins identified out of 121 protein spots, were differentially accumulated in relation to biochemical processes controlling major fruit development and ripening traits. Several differentially accumulated protein spots associated to fatty acids biosynthesis and aroma compounds were also detected.

More recently,another comprehensive study where the regulation of olive miRNA under different developmental phases and tissues related to alternate bearing was performed(Yanik *et al.* 2013). For this, six olive small RNA libraries were constructed from fruits and leaves and sequenced by high-throughput Illumina sequencing.Bioinformatics analyses of 93,526,915 reads identified 135 conserved miRNA, belonging to 22 miRNA families in the olive. Among the six libraries, expression of olive tree miRNAs varied greatly, indicating the contribution of diverse miRNA in balancing between reproductive and vegetative phases.

A recent paper (Dündar, 2013) describedcDNAs that are associated with alternate bearing in olive were isolated and analysed from the leaves of trees in “on year” and in “off year” in July (when fruits start to appear) and in November (harvest time). qRT-PCR analyses of the cDNA libraries was conducted to confirm cDNA molecules that are associated with different developmental stages and fruits and a number of candidate cDNAs associated with “on year” and “off year” were isolated. Screening of ~500 cDNAs revealed 37 putative transcripts corresponding to known gene functions which were annotated with gene names and Gene Ontology (GO) terms. In this study, seven commonly used reference genes [Olest34, alpha-tubulin, beta-tubulin, beta-actin, 26S rRNA, 18S rRNA and Glyceraldehyde phosphate dehydrogenase] were chosen to identify the reference gene which is least spatially and temporally variable and GAPDH was decided to be used as an appropriate reference gene for olives to normalize the copy numbers of the cDNAstested(Dündar, 2013).

1.11 GENE EXPRESSION ANALYSIS AND REFERENCE GENE VALIDATION USING REAL-TIME PCR (qPCR)

As the current research utilises qPCR to characterise the expression of genes during olive fruit development, a brief overview of this technique and its application are covered.

The qPCR technique relies upon the detection and quantitation of the amount of PCR product formed at every cycle of the PCR (characterised by the quantification cycle Cq) using a DNA-binding fluorescent reporter molecule.SYBR Green I dye is a double-stranded DNA binding dye which intercalates to DNA double helix and upon excitation emits light leading to an increase in fluorescence. In a solution the SYBR Green I dye exhibits very weak fluorescence, however when bound to a DNA molecule fluorescence increases dramatically. The amplification of the PCR product bound to SYBR Green I dye is represented in the form of a plot where measured fluorescence is plotted against the PCR cycle number (Figure 1.18). The amplification plot is divided into three phases (Karlen *et al.* 2007). **Phase I**: It represents the lag phase where no amplification can be detected. This phase is used to detect the baseline-subtracted fluorescence.**Phase II**: It represents the exponential phase where the amount of PCR product approximately doubles in each cycle (100% amplification efficiency). On a log scale graph [Figure 1.17 (right)] this phase corresponds to the linear phase which calculates two values. Firstly, a threshold line is calculated which is no more than half way up the linear part of the reaction. It is the level of detection at which a reaction reaches a fluorescent intensity above background. Secondly, at this phase the Cq value (quantification cycle) is determined which is defined as the cycle number at which the fluorescence crosses the threshold line.**Phase III**: It represents the non-exponential plateau phase where the DNA concentration no longer increases exponentially due to the limitation of one or more reactants.

In RT-qPCR experiments using a relative quantification approach, the expression level of the target genes are normalised using internal control genes known as reference genes to derive fold change gene expression. This normalisation strategy improves the fidelity of the quantification process by controlling any variation in the expression level of the biological samples that might have been introduced due to various factors such as RNA integrity, initial sample amount, reverse transcription efficiency etc. Some of the most common and best

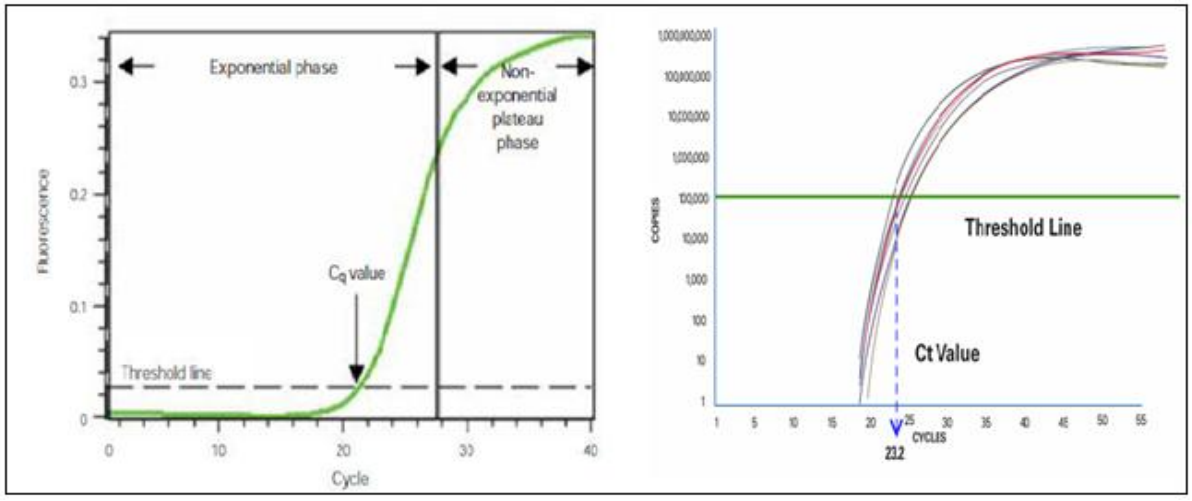


Figure 1.17 Representation of real time PCR amplification curve. Left: Regular view, Right: Logarithmic view (Bio-Rad manual)

known reference genes that have been used over the last few decades in plants and animals include glyceraldehyde 3-phosphate dehydrogenase, 18S or 26S RNA, elongation factor 1 alpha, ubiquitin carrier protein, actin, α-tubulin, β-tubulin and TATA-Box binding protein (Czechowski *et al.* 2005; Jarošová *et al.* 2010; Gamm *et al.* 2011; Long *et al.* 2011). These genes have been chosen as reference genes for their housekeeping roles in basic cellular and metabolic processes. It is assumed that these reference genes have constant level of expression in different tissues and under different treatments and has no inter-individual variability. Unfortunately, it is impossible to find an ideal reference gene for normalisation in various samples as it is unreasonable to assume that the transcription of any gene will be absolutely resistant to any fluctuations in the cell cycle or nutrition status (Jarošová *et al.* 2010; Long *et al.* 2011). It has been shown that in many experiments the use of a single reference is not acceptable as it would likely produce erroneous conclusions in expression patterns (Vandesompele *et al.* 2002; Gutierrez L *et al.* 2008). Recent reports have also shown that the most commonly used traditional reference genes may be inappropriate for normalization in qPCR experiments due to their expression variability under different experimental conditions (Schmittgen *et al.* 2000; Czechowski *et al.* 2005). The importance of expression stability in the choice of reference genes has prompted the development of various software packages such as GeNorm (Hellemans *et al.* 2007), BestKeeper (Pfaffl *et al.* 2004) and NormFinder (Andersen *et al.* 2004) to identify them. Therefore for accurate analysis of RNA transcription it is crucial to choose reference genes that are minimally regulated in a particular experiment in a given species and organs/tissues.

In plants a number of reference gene validation attempts have been reported covering both model and crop species such as *Arabidopsis* (Czechowski *et al.* 2005), rice (Jain M *et al.* 2006), wheat (Long *et al.* 2011) grapevine (Gamm *et al.* 2011), barley (Jarošová *et al.* 2010), soybean(Hu *et al.* 2009) and cotton (Artico *et al.* 2010). The analysis of the *Arabidopsis thaliana* Affymetrix ATH1 microarray data has revealed that there are hundreds of potential reference genes that outperform traditional reference genes in terms of expression stability, where most of these genes are expressed at much lower levels than the traditional reference genes. This list of new *Arabidopsis* reference genes were successfully employed to search for reference genes in unrelated species such as *Vitis vinifera*(Reid *et al.* 2006) and *Coffea arabica*(Cruz *et al.* 2009).

Although some reference genes were recently suggested for the major tissues (leaves, fruits and pedicels) of olive (Dündar, 2013), no firmly established information about the identification of valid reference genes in olive is available to date. Therefore, in order to normalise the expression data of target genes in olives it is very important to identify stable reference genes which show consistent expression with the olive samples. The Czechowski et al. (2005) data may also be used to select a panel of reference genes which may be appropriate for a specific set of chosen experimental conditions and tissue types in olives.

1.12 AIMS OF THE PROJECT

As discussed earlier the sterol composition andtotal sterol content of olive oil has been found to be affected by various geographical and technological factors like cultivar, climate, crop year, degree of fruit ripeness, temperature during fruit maturation, storage time of fruits prior to oil extraction, processing and extraction methods. However, the impact of other agronomical practices on sterol levelsof Australian olive oils, specifically, fruit size, irrigation, fruitmaturity, malaxing time, malaxing temperature and delays between harvest and process is not known. The current study aimed to investigate the impact of these practices on the sterol profiles of Barnea, Frantoio and Picual.

There is clear evidence that the gene families *SMT2*and *SMO2*play crucial roles in determining the ratio of campesterol to sitosterol in plants. It is plausible that allelic differences in these genes and/or differences in the expression patterns of these genes may

impact on the sterol profiles observed in the oils derived from these cultivars. However, little is known about the genes encoding these enzymes in olives. As a first step in investigating these enzymes in olives the current study aimed to isolate and sequence *SMT2* and *SMO2* cDNAs from Barnea, Frantoio and Picual to characterize the gene families in olives and identify any allelic differences between these cultivars. This study also aimed to investigate the expression levels of the *SMT2* and *SMO2* genes throughout fruit development to identify any inter-cultivar differences that may also impact on the sterol profiles in these Australian olive oils.

1.12.1 Major aims

To characterise the sterol profiles of Australian olive oils by studying the impact of various horticultural and processing practices.

To isolate and characterise *SMT2* and *SMO2* genes from cultivars of olivesgrown in Australiathat produce oils varying significantly in their sterol profiles and identify any allelic differences or differences in the expression of these genes that may contribute to differences in these observed sterol profiles.

1.12.2 Specific aims

- To study the impact of horticultural and processing practices such as fruit size, irrigation, fruitmaturity, malaxing time, malaxing temperature, delays between harvest and process and storagetime on sterol levels of olive oils produced from cultivars Barnea, Frantoio and Picual.
- To isolate and characterise the *SMT2* and *SMO2* gene families from olivecultivars Barnea, Frantoio and Picual and identify any allelic differences between cultivars.
- To select and validate stable reference genes to study the expression of *SMT2* and *SMO2* gene familiesin oliveusing a qPCR approach.
- To characterise the expression of *SMT2* and *SMO2* gene familiesin the developing olive fruit to identify any differences that may contribute to the specific sterol profiles in the Australian olive oils under investigation.

CHAPTER 2

MATERIALS AND METHODS

List of Figures and Tables in Chapter 2

Chapter 2 Materials and Methods

MATERIALS

2.1 CHEMICALS, REAGENTS AND KITS

The materials used during the course of this research are listed below. The methods for preparation and sterilisation of the reagents are included in Appendix I. The information on plasmid cloning vectors used in this study is included in Appendix II.

2.1.1 General buffers and solutions

To prepare all buffers and solutions, analytical grade chemicals were used unless otherwise stated. Distilled and de-ionised milli-Q water (Milli-Q Plus Ultra-Pure Water System, Millipore, Billerica, MA, USA) was used to prepare all buffers and solutions and were sterilized by autoclaving at 121˚Cfor 15 minutes or, whereindicated, filter sterilized using a 0.22 μM or 0.45 μM millipore membrane filter.

For RNA solutions, all chemicals, water, buffers, plasticware and glassware used were RNase free. All glassware, pestle and mortars, forceps and spatulas were covered with aluminium foil and baked at 200˚C for at least 12 hours prior to use. DEPC-treated water (0.1%) was prepared with distilled and de-ionized milli-Q water and was used to prepare all buffers for RNA work in RNase-free glassware. All electrophoresis tanks, tray, combs and plastic containers were sprayed with RNase Erase (MP Biomedicals) and rinsed with DEPC treated water. Disposable plastic ware (pipette tips and eppendorf tubes) were purchased RNase-free, and RNase-free barrier tips (LightLabs: AvantGuard barrier tips) were used for pipetting. Micropipettes, centrifuges and benchtops were sprayed with RNase Erase prior to each extraction process.

A number of buffers/enzymes were supplied with kits for use in PCR reactions, DNA purifications, restriction digestions, ligation, cloning, sequencing and radiolabelling reactions

as outlined below. In addition, the following buffers and solutions were prepared in the laboratory for various experimental works (Appendix I):

- 2X CTAB
- 5% CTAB solution
- 1% CTAB solution
- REB solution
- 3M Sodium Acetate (pH 5.2)
- TE Buffer
- Ethidium bromide (10mg/mL) (Sigma)
- TBE electrophoresis buffer
- TAE electrophoresis buffer
- SSC Buffer
- Denaturation solution
- Neutralisation solution
- Washing buffer
- Denhardt's prehybridisation buffer
- 10% SDS (electrophoresis grade, BioRad)
- 0.5M EDTA (Ajax Chemicals)
- Ampicillin (50mg/mL) (Sigma)
- BigDye sequencing dilution buffer
- DEPC water (0.1%)
- Choloroform: isoamyl alcohol (24:1)
- phenol: choloroform: isoamyl alcohol (25:24:1)
- Buffered phenol (for DNA isolation)

2.1.2 Enzymes

A number of enzymes were used in this project especially for PCR, cloning and sequencing reactions, many of which were supplied with the kits. They include:

- RNase A (20mg/mL) (Invitrogen)
- Platinum *Taq* polymerase (Invitrogen)
- Platinum *Taq* polymerase high fidelity (Invitrogen)
- Various restriction endonucleases (Promega, MBI Fermentas and New England Biolabs)
- The Klenow fragment of *E. coli* DNA polymerase I (Promega)
- Alkaline Protease solution (Promega)
- Calf Intestinal Phosphatase (CIP)
- Tobacco Acid Pyrophosphatase (TAP)
- T4 RNA ligase

2.1.3 Microbiological media

- LB broth (Luria-Bertani broth): 1% w/v bacto-tryptone, 0.5% bacto-yeast extract, 1% w/v NaCl
- LB-ampicillin broth (100 μg/ml): 1% w/v bacto-tryptone, 0.5% bacto-yeast extract, 1% w/v NaCl, 100 μg/ml ampicillin
- SOC media: 2% Tryptone, 0.5% Yeast Extract, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, 20mM glucose

2.1.4 Commercial kits and reagents

Various commercially available kits and reagents were used in this research study which are summarised below in Table 2.1 and 2.2 respectively.

Name of kits	Supplier	Experiment
Thermoscript RT-PCR system	Invitrogen	cDNA synthesis (Section 2.7.2)
ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit v.3	Applied Biosystems	DNA Sequencing (Section 2.6.6.4)
Prime-a-Gene Labeling System	Promega	Radiolabeling of probes (Section 2.9.3.3)
TOPO TA cloning and sequencing kit	Invitrogen	Cloning and DNA Sequencing (Section 2.6.6.1)
Wizard *Plus* SV Minipreps DNA Purification System	Promega	Plasmid purification (Section 2.6.6.2)
Wizard® SV Gel and PCR Clean-Up System	Promega	Purification of PCR products (Section 2.6.5)
Gene Racer kit	Invitrogen	5' and 3' RACE PCR (Section 2.6.3)
RNeasy Plant Mini Kit	Qiagen	Total RNA extraction (Section 2.4.2)
DNase I, RNase free kit	Qiagen	DNA-free RNA isolation (Section 2.4.2)
GeneRacer SuperScript™ III RT Module	Invitrogen	Reverse transcription of total RNA to create RACE-ready first-strand cDNA (Section 2.6.2)
LightCycler® Master^{PLUS}DNA MasterSYBR Green I	Roche	qPCR Expression analysis (Section 2.10)

Table 2.1 List of commercial kits used in this research study

Name of reagents	Supplier	Experiment
Platinum Taq Polymerase	Invitrogen	"Hot-Start" PCR (Section 2.7.3)
Platinum Taq Polymerase high fidelity	Invitrogen	"Hot-Start" PCR (Section 2.6.3)
Radiolabeled (α_{32}P dCTP) 50μCi, 3,000 Ci/mmol	Perkin Elmer	Radiolabelling of the probes (Section 2.9.3.3)
Illustra ProbeQuantG-50 Micro Columns	GE Healthcare	Purification of radiolabelled probes (Section 2.9.3.3)
RNase ERASE Spray	MP Biomedicals	Removal of RNase contamination from benchtops, glassware and plasticware (Section 2.4.2)
100bp molecular weight marker	Roche	Determination of size of PCR products (Section 2.6.3 and 2.7.3)
1kb molecular weight marker	Roche	Determination of size of genomic DNA samples (Section 2.9.1)
DNA grade agarose	Applichem	Agarose gel electrophoresis Section 2.6.3, 2.7.3, 2.9.1)

Table 2.2 List of commercial reagents used in this research study

2.2 PLANT SAMPLES

Samples were collected from individual olive trees of the cultivars Barnea, Frantoio and Picual. These trees were labelled such that the olive fruit samples collected over different time points came from the same tree in order to avoid any biological variation.

2.2.1 Plant material collection for sterol analysis

Olive fruits that were used for olive oil extraction for sterol analysis were collected and processed from the selected Victorian groves: Boort Estate (Boort, Victoria) and BoundaryBend Estate (Boundary Bend, Victoria) over two consecutive years, 2007 and 2008.

2.2.2 Plant material collection for DNA/RNA extraction

All plant materials (olive fruits and leaves) which were used for the investigation of *SMT2* and *SMO2* gene families were collected from the Boort site in Northern Victoria during the 2009 crop season.

Olive fruit was collected and preserved at a number of developmental stages during the growth period of the olive. Specifically, olive fruitwas collected at 96, 109, 116, 122, 136, 150, 167 and 170 days after flowering (DAF) in the 2009 crop season. Olive fruit thatwas collected for RNA extraction was harvested in a manner to establish an RNase-free environment (Section 2.1.1). Briefly the procedure involved:

- Olive fruit from the different cultivars was first collected in RNase-free 50 mL tubes placed on ice.
- The harvested olive fruits were chopped into pieces < 0.5 cm across using sterile scalpel blades from the mesocarp region and then placed in sterile 50 mL tubes containing 50 mL of RNA later Tissue Collection: RNA stabilization solution (Applied Biosystems). The tubes were sealed and gently mixed by inverting.
- The tubes were stored at 4˚C overnight and then moved to -20˚Cuntil the end of the harvest period. Samples were placed at -80˚Cfor long term storage.

For DNA extraction, leaves from the olivecultivars mentioned above were collected from the trees and stored at -20˚Cand then snap frozen and stored at -80˚Cfor further use.

2.3 STEROL ANALYSIS

2.3.1 Horticultural and processing factors evaluated

Cultivation, harvest and processing practices were evaluated to assess their impacts on the sterol content and profile of the Australian olive cultivars, Barnea, Frantoio and Picual.The horticultural and processing practices which were evaluated werecultivar, crop year, fruit size, irrigation, fruit maturity, malaxing time, malaxing temperature and delays between

harvest and processing (Table 2.3). Fruit maturity was measured using the maturity index (MI) developed by the CIFA Alameda del Obispo, Spain (Boskou 2006). For this, fruit from the threecultivars was harvested at three different times, early harvest, middle harvest and late harvest with 2–3 weeks between harvests (Table 2.3). Three replicates of each treatment were processed. Each replicate typically consisted of two units of 700 g of olive paste each. Replicates of each treatment were processed during the 2007 and the 2008 seasons. All samples were evaluated in duplicate.

2.3.2 Method of extraction of sterols from olive oil

Olive oil was extracted using the experimental olive oil mill (Abencor^{R}) and sterol analysis was conducted according to the official method IOOC/T.20/No10/Rev and analysed by an Agilent Technology 6890N gas chromatography system (International Olive Oil Council 2009). Sterols were quantified using 5α-cholestan-3β-ol as the internal standard. The data subjected to a statistical analysis was assessed through an analysis of variance using the SAS version 8.02 (SAS Institute Inc., Cary, NC, USA). Separation of the means was obtained using the least square means test and significant differences were defined at $P \leq 0.05$. Every aspect was analysed separately. No interactions were evaluated in this project. Briefly the procedure involved:

- Olive oil was saponified in a 2N ethanolic solution of potassium hydroxide.
- After several rinses with ethyl ether and distilled water the unsaponifiable fraction was separated from the saponifiable fraction.
- The sterol fraction was extracted from the unsaponifiable fraction by thin layer chromatography (TLC). For this, basic silica gel plates (20 cm x 20 cm) were immersed in 0.2N ethanolic solution of potassium hydroxide and dried for two hours in a fume cupboard. A 5% solution of the unsaponifiables with chloroform was prepared and applied on the silica gel plates and the sterol compounds were separated using a toluene-acetone mixture (95:5 V/V) as an eluent. After separation the plates were sprayed uniformly with 2,7-dichlorofluorescein solution (0.2% ethanolic solution) and observed under ultraviolet light (wavelength: 366 or 254 nm) to identify the sterol bands.

Olive cultivar
Barnea
Frantoio
Picual
Crop year
2007
2008
Fruit size
Barnea: small (<2.00g), medium (2.00–3.00g), large (>3.00g)
Frantoio: small (<1.40g), medium (1.40–2.00g), large (>2.00g)
Picual: small (<2.20g), medium (2.20–3.20g), large (>3.20g).
Irrigation (Kc-crop factor)*
Kc-0.74 (Normal treatment)
Kc-0.32 (1/2X)
Kc-1.48 (2X)
Fruit maturity
Early harvest (MI : between 1.00 and 2.00)
Middle harvest (MI : between 2.5 and 3.5)
Late harvest (MI : between 4 and 5)
Malaxing time
15minutes (1/2 X)
30minutes (Standard-X)
60minutes (2 X)
Malaxing temperature
28˚C (Standard)
18˚C(Cold)
38˚C(Hot)
Delays between harvest and process
Immediate processing (<12 hours)
Medium processing (48 hours)
Delayed processing (120 hours)

Table 2.3 Details of horticultural and olive oil processing factors for evaluation of sterol content and composition

*****Crop factors (Kc) are crop specific evapotranspiration values that must be replaced with irrigation for a crop to produce commercial yield (Lazzara *et al.* 2010)

MI: maturity index of harvested fruits(Boskou 2006)

- The sterol bands were scraped off the silica gel plates, filtered with hot chloroform, washed several times with ethyl ether and evaporated to dryness.
- The sterols were transformed into trimethyl-silyl ethers by adding a silylation reagent consisting of a 9:3:1 (V/V/V) mixture of pyridine/hexamethyl disilazane/trimethyl chlorosilane in the ratio of 50 µL for every milligram of sterols.
- The silylated sterols were analysed by capillary column gas chromatography. The analytical column was a DB-5 5% phenyl-methyl-siloxane stationary phase (30 m x 0.25 mm x 0.25 µm). The gas chromatographic conditions were as follows: inlet temperature: 280˚C; oven temperature 267˚C; detector temperature: 290˚C; split ratio: 30:1; amount injected 1:l. Hydrogen was used as the gas carrier at a flow rate of 1.2 ml/min.

GENERAL MOLECULAR METHODS

2.4 PREPARATION OF GENOMIC DNA AND RNA

2.4.1 Extraction of genomic DNA from olive leaves

Genomic DNA (gDNA) was extracted from olive leaves using the CTAB method (Murray *et al.* 1980) with some modifications. This involved the grinding of frozen leaf tissue under liquid nitrogen to obtain a fine powder. The cells were then lysed by adding 2X CTAB solution (preheated to 60˚C) and incubated at 60^{o}C for 15minutes. Equal volumes of chloroform: isoamyl alcohol were added to the mixture and the aqueous phase was separated from the organic phase by centrifugation at 4,000g for five minutes. The aqueous phase was collected and 5% CTAB was added in a ratio 1/5 of the sample's volume, followed by extraction with chloroform: isoamyl alcohol and centrifugation at 8,000g for 12 minutes. The aqueous phase was collected again and 1%CTAB solution was added in a ratio 1:1.3 in regard to the sample's volume and incubated at room temperature for 1 hour, followed by centrifugation at 8,000g for 12 minutesto collect the aqueous phase. The pellet was resuspended in REB solution and the DNA was precipitated in 2 volumes of cold 95% ethanol. The pellet was resuspended in ddH_2O and RNA was hydrolysed by adding RNase A (final concentration 10µg/mL) and incubated at 37˚C for 30 minutes. After RNase treatment,

equal volume of equilibrated phenol was added to the sample, mixed thoroughly, followed by centrifugation at 10,000g for 12 minutes to obtain the aqueous phase. This was repeated using phenol: chloroform: isoamyl alcohol (25:24:1) and then with chloroform: isoamyl alcohol (24:1). Finally, the supernatant was collected and 3M sodium acetate (pH 5.2) (0.1 times supernatant volume) and 100% ethanol were added (2 times of supernatant volume) for an overnight DNA precipitation at -20°C. The precipitated DNA was collected by centrifugation (10,000g) for 30 min at 4°C, washed with 70% ethanol twice, air dried at room temperature for 10 minutes and re-suspended in 30-50 µL of ddH_2O. DNA was stored at -20˚C until further use.The integrity of the purified gDNA was visually analysed using agarose gel electrophoresis (section 2.5.2) and the concentration was determined using a fluorometer (section 2.5.1).

2.4.2 Extraction of total RNA from olive fruits and DNase treatment of RNA samples

Total RNA was extracted from the mesocarp of olive fruitsusing the RNeasy plant mini kit (Qiagen) according to the supplied protocol. Frozen olive fruit pieces (section 2.2.1) were thawed on ice and were retrieved from the RNA *Later* solution with sterile forceps and approximately 100mg of olive fruit tissuewasplaced in an RNAse-free mortar and ground to a fine powder under liquid nitrogen. The cells were then lysed using 450µL of buffer RLT which contains guanidine thiocyanate which aids in effective denaturation and disruption of the cells. The samples were then centrifuged through a QIAshredder homogenizer and the supernatant of the flowthrough was transferred to a new eppendorf tube, 0.5 volumes of ethanol was added and the sample was transferred to the RNeasy Mini spin column for selective binding of the RNA to the membrane. The column was washed with buffer RW1 and the RNA was treated with DNase I to remove any contaminating genomic DNA using the RNase-free-DNase set (Qiagen). Briefly, 10µL of DNase I and 70 µL of buffer RDD was added to the RNeasy Mini spin column containing the bound RNA and incubated at room temperature for 15 minutes. The DNase I was then removed by a second wash with buffer RW1 and other contaminants were washed away during two subsequent washes with buffer RPE. DNase treated RNA was then eluted in RNase free ddH_2O and stored at -80˚C until required.

Total RNA was extracted from each olive fruit sampling date namely, 96, 109, 116, 122, 136, 150, 167 and 170 DAF harvested in the 2009 crop season. A pooled sample of RNA was also prepared by pooling equal amount of RNA from each timepoint harvested in the 2009 crop season and was used for cDNA synthesis (Section 2.7.2) and RACE cDNA synthesis (Section 2.6.2). RNA isolated from olive fruit tissues harvested at individual timepoints in the 2009 (96, 109, 116 and 136 DAF) crop season for each cultivar were used for gene expression analysis using qPCR (Section 2.10).

2.5 QUANTIFICATION AND ELECTROPHORESIS OF DNA/RNA

2.5.1 Estimation of DNA/RNA concentration using fluorometery

Genomic DNA, PCR products and total RNAsampleswere quantified using the Qubit fluorometer (Invitrogen). To quantify samples at very low/high concentrations, specific dsDNA/RNA HS (high sensitivity) dyes and dsDNA/RNA BR (broad range) dyes were used. To calibrate the fluorometer two standards were prepared with a fixed amount of nucleic acid (provided with the kit) and combined with the buffer and dye. For quantifying each sample, a 1:200 working solution was prepared by diluting the dye with their respective buffer. Approximately 1-5 µL of DNA/RNA samples were added to an individual assay tube containing the working solution to make up a final volume of 200 µL. The samples were then mixed by vortexing for 2-3 seconds, incubated at room temperature for 2 minutes and the concentration was measured by the fluorometer.

2.5.2 Agarose gel electrophoresis

To visualise and assess the purity and integrity of gDNA, PCR products and total RNAsamples, all samples were resolved electrophoretically in a 1% agarose gel. For agarose gel electrophoresis, all gels (1-2%) were prepared with DNA grade agarose (Applichem) with 1X TBE and ethidium bromide (0.6 µg/mL) for DNA samples. For RNA samples, 1X TAE buffer and ethidium bromide (0.6 µg/mL) were used. Appropriate molecular weight markers (MBI Fermentas) (Table 2.2) were used to estimate the size of the DNA fragments. Aliquots to be separated by agarose gel electrophoresis were mixed with 6X Loading Dye Solution (MBI Fermentas) to a final loading concentration of 1X and electrophoresed at 35-100V for

times ranging from 45 minutes to 14 hours (details specified in relevant sections below). The gels were visualised using a UV/White Darkroom (UVP, Upland CA, USA) and analysed with Labworks Analysis Software (UVP).

Agarose gel electrophoresis was used for various purposes throughout this research study:

- Visual assessment of the quality of genomic DNA (Section 2.9.1)
- Visual assessment of the quality of RNA (Section 2.5.2)
- Visual assessment of the results of PCR reactions and gel-purification of the PCR products (Section 2.6.3, 2.7.3)
- Separation of restriction digested gDNA for Southern blotting (Section 2.9.1)
- Visual assessment of PCR reactions/restriction digestions to confirm the success of cloning reactions (Section 2.6.6.3, 2.7.4).

2.5.3 Quantitation of RNA samples using a Bioanalyzer

The integrity of all RNA samples wasassessed by gel electrophoresis by visualising two distinctbands of the two rRNA subunits (28S rRNA and18S rRNA). In addition to this, the quality of the RNA samples were judged by their RNA integrity number (RIN) calculated by the Agilent 2100 Bioanalyzer (Schroeder *et al.* 2006). The Agilent Bioanalyzer uses the lab-on-chip technologies for the standardization of RNA quality control where the RNA samples are electropheretically separated on a micro-fabricated chip and then detected via laser induced fluorescence. The software generated a gel like image and estimated the concentration of each sample. It automatically generates the ratio of 18S to 28S ribosomal units and displays a RIN value after taking the entire electrophoretic trace into account. A RIN value of 1 denoted completely degraded RNA and a RIN value of 10 denoted high quality RNA. In this study only RNA samples with RIN values ≥8 were used for subsequent analysis.

2.6AMPLIFICATION OF PARTIAL CODING SEQUENCES OF *SMO2* AND *SMT2* GENES FROM BARNEA

A BLASTn screen[http://www.biomedcentral.com/content/supplementary/1471-2164-10-399-S1.xls]of the 454 sequencing data generated from developing olives(Alagna *et al.* 2009)with the *SMO2/SMT2* sequences of other plant species revealedexpressed sequence tags

(ESTs) of putative *SMT2*(365bp)(Cluster Id: OLEEUCl064665)genes (Section 4.2.1) and *SMO2* (238bp*)* (Cluster Id: OLEEUCl011741) (Section 5.2.1) genes. In order to isolate the full length coding sequences of these gene families in olives, thesepartial sequences were used for the amplification and cloning of the uncharacterized 5' and 3' regions of the *SMT2* and *SMO2* genes from the olive cultivar Barnea using a RACE **(**Rapid amplification of cDNA ends) PCR approach. The strategy for the amplification of full length *SMT2*/*SMO2* coding sequences has been shown in Figure 2.1.

2.6.1Design of oligonucleotide primers for RACE PCR

Forward and reverse gene specificprimers (GSPs) were designed using the ESTs of *SMO2* (Section 5.2.1) and *SMT2* genes (Section 4.2.1)obtained from the Alagna *et al* data (2009) (Figure 2.1 and Table 2.4). For the amplification of the 5' ends, 5' gene Racer primer and the reverse GSP was used, and for the amplification of the 3' ends, 3' gene Racer primer and the forward GSP was used (Table 2.4). The 5' gene Racer primer and the 3' gene Racer primer were provided with the kit (Section 2.1.4). In order to avoid the formation of primer dimers and primer hairpin formation the quality of each primers were assessed using the quality scores as rated by Net Primer software (http://www.premierbiosoft.com/netprimer/index.html).

All the primers were commercially synthesized (GeneWorks Pty Ltd, South Australia, Australia) with cartridge purification. The primers were supplied as a dried precipitate and stock solutions of the primers with an initial concentration of 100µM were prepared with ddH_2O and stored at -20˚C until required.

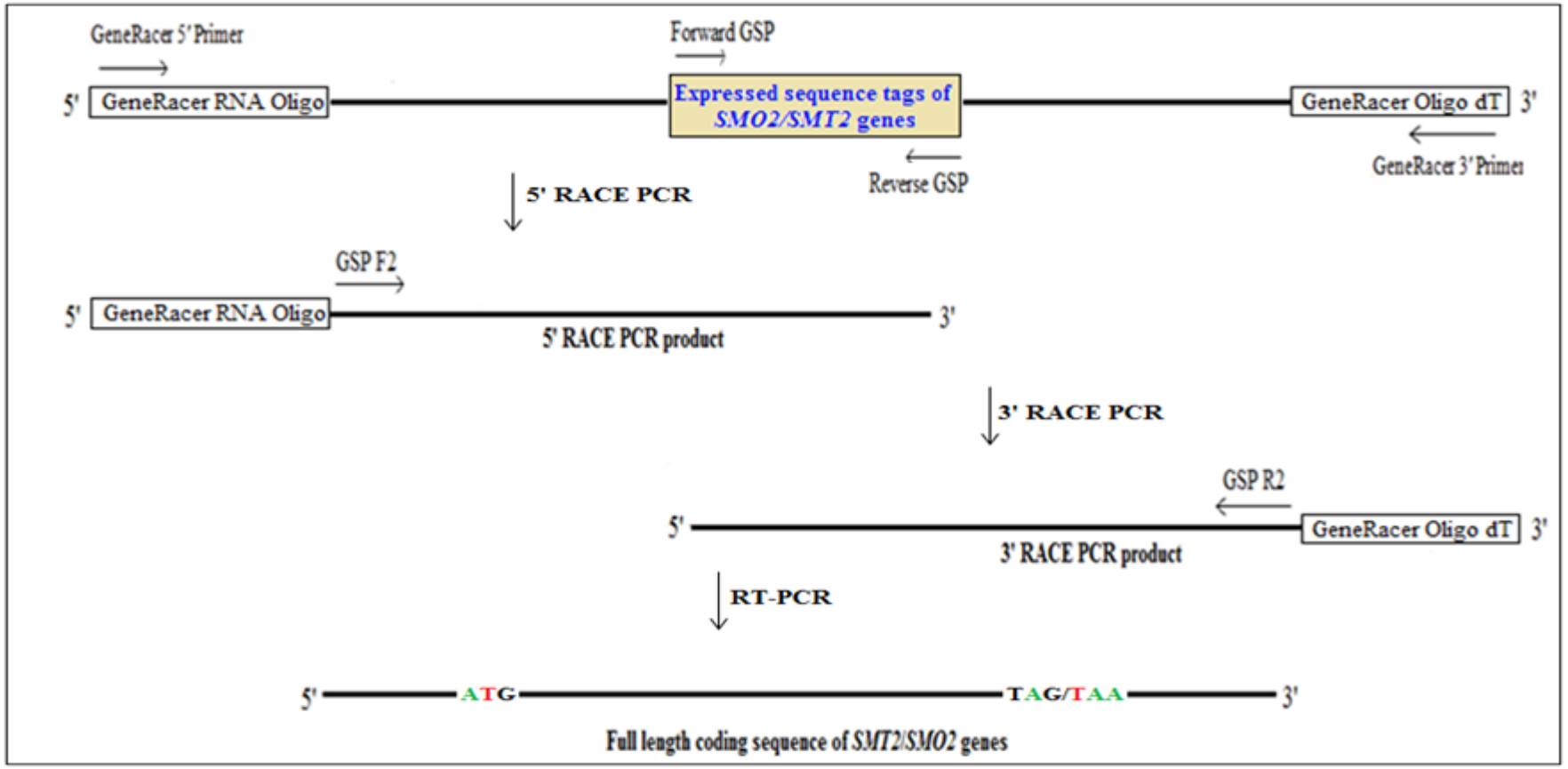


Figure 2.1Strategy for the amplification of full length *SMT2/SMO2* genes

The partial *SMT2/SMO2*ESTsidentified from the Alagna *etal* data (2009) are depicted as the brown box. The GeneRacer 5' and 3' primer and the forward and reverse gene specific primers (GSPs) used for the amplification of 5' and 3' RACE ends have beenshown with arrows. The forward and reverse primers designed for the amplification of full length coding sequences of the *SMT2/SMO2*genes (F2 and R2) have also been shown. The START codon and STOP codon of the ORFs have been shown highlighting that the designed GSPs sit outside the putative ORFs.

Primer pairs and sequences (5'-3')	Annealing temperature used for PCR (°C)	Purpose
GeneRacer-F: GGACACTGACATGGACTGAAGGAGTA **SMT2-R**: AGTGAAAATTCCCGACTCGCCACCTCTA	68	5' RACE PCR (SMT2)
GeneRacer-R: GCTGTCAACGATACGCTACGTAACG **SMT2-F**:GATAAACAATGCGAAGTCGTGTGCG	68	3' RACE PCR (SMT2)
GeneRacer-F: GGACACTGACATGGACTGAAGGAGTA **SMO2-R**: ATAACCCTTGTCAGTACCAAATATCC	65	5' RACE PCR (SMO2)
GeneRacer-R: GCTGTCAACGATACGCTACGTAACG **SMO2-F**: CCTGCCATYACTGGTCCCCATCTG	68	3' RACE PCR (SMO2)

Table 2.4 Primer pairs for 5' and 3' RACE PCR for*SMT2* and *SMO2* genes

2.6.2 Synthesis of RACE cDNA

Total RNA extracted from fruit pooled over development (Section 2.4.3) from the cultivar Barnea was used for the synthesis of RACE cDNA using the GeneRacer kit (Invitrogen), according to the supplied protocol. Briefly, this technique involved the following steps:

- The total RNA (1 μg) isolated from Barnea was treated with calf intestinal phosphatase (CIP) enzyme to remove 5′ phosphate groups. The dephosphorylation reaction was carried out in a total volume of 10 μL containing 1-2 μg RNA, 1X CIP Buffer (see Appendix I), 40 U RNaseOUT and 10 U CIP enzyme.The reaction mixture was incubated at 50˚C for 1 hour and the RNA was precipitated overnight with 0.1V 3 M sodium acetate (pH 5.2), 2V 95% ethanol and muscle glycogen (final concentration 0.2 mg / ml) and the pellet was resuspended in 7 μL DEPC water.

- The dephosphorylated RNA was then treated with tobacco acid pyrophosphatase (TAP) toremove the 5′ cap structure from intact, full-length olive mRNA. The decapping reaction was carried out in a total volume of 10μL containing 7μL dephosphorylated RNA (1X TAP Buffer (See Appendix I), 40U RNaseOUT and 0.5 U TAP enzyme. This reaction mixture was incubated at 37˚C for 1 hour and the RNA was precipitated overnight with 0.1V 3M sodium acetate (pH 5.2), 2V 95% ethanol and mussle glycogen (final concentration 0.2 mg/ml) and the pellet was resuspended in 7μL DEPC water.
- The gene Racer RNA Oligo was then ligated to the 5′ end on the olive mRNA using the T4 RNA ligase enzyme. At first the 7μL decapped RNA was added to a tube containing the lyophilized, pre-aliquoted gene Racer RNA Oligo and mixed well by pipetting up and down several times. This mixture was incubated at 65˚C for 5 minutes to relax the secondary structure and then placed on ice for 2 minutes. To this reaction mixture, the following reagents were added to carry out the ligation reaction: 1X ligase buffer, 1mM ATP, 40U RNaseOUT and 5U T4 RNA ligase enzyme. This reaction mixture was incubated at 37˚C for 1 hour and the RNA was precipitated overnight with 0.1V 3M sodium acetate (pH 5.2), 2V 95% ethanol and mussel glycogen (final concentration 0.2mg/ml) and the pellet was resuspended in 10μL DEPC water.
- The RNA oligo ligated RNA was then reverse transcribed to form the first strand cDNA using the SuperScript III RT enzyme. At first, to the 10 μL RNA Oligo ligated RNA, 50μM oligo dT primer and 0.5mM dNTP mix was added and incubated at 65˚C for 5 minutes to relax any RNA secondary structure. To this reaction mixture, the following reagents were added to a final volume of 20μL to carry out the synthesis of cDNA: 1X First Strand buffer, 0.1M DTT, 40U RNaseOUT and 200U SuperScript III RT enzyme. The reaction mixture was incubated at 55ºC for 60 minutes and then terminated by incubating at 70ºC for 15 minutes. *E. coli* RNase H (2U) was added to the mixture to remove the RNA template and incubated at 37ºC for 20 minutes. The cDNA synthesis reactions were then stored at -20ºC until further use or were used immediately for RACE PCR reactions to amplify the 5’ and 3’ ends of the genes of interest using forward and reverse GSPs (Section 2.6.2).

2.6.3RACE PCR

RACE PCR was conducted using the RACE cDNA synthesized (section 2.6.2) and Platinum *Taq* Polymerase High Fidelity. As mentioned in Section 2.6.1, for the amplification of the 5' ends, 5' gene Racer primer (complementary to the gene Racer RNA Oligo) and the reverse GSP was used and for the amplification of the 3' ends, 3' gene Racer primer (homologous to the 3' gene Racer oligo dT primer) and the forward GSP was used.

The RACE PCR reaction was carried out in a total volume of 50μL containing 1μL RACE cDNA (Section 2.6.2), 1X High Fidelity PCR Buffer, 0.2mM dNTPs, 2mM $MgSO_4$, 0.2μM gene specific primer (forward/reverse) (Section 2.6.1), 5'/3' gene Racer primer (0.6μM) and 2.5U Platinum *Taq* Polymerase High Fidelity. For RACE work, touchdown PCR (Gene Racer Kit, Invitrogen) was carried out to increase the specificity and reduce amplification background (Table 2.5).

The following cycling conditions were used for RACE PCR:

Temperature (˚C)	Time	Number of cycles
94	2 minutes	1
94	30 seconds	5
72	2 minutes	
94	30 seconds	5
70	2 minutes	
94	30 seconds	20-25
60-68	30 seconds	
72	2 minutes	
72	10 minutes	1

Table 2.5 Cycling conditions for touchdown RACE PCR

2.6.4 Controls for RACE PCR

Both positive and negative controls were used at all times during RACE work. The RACE PCR kit from Invitrogen provided a HeLa total RNA to be used as a positive control for each enzymatic step and PCR. The RACE-ready cDNA which was synthesized from the control HeLa total RNA was then used to amplify β-actin gene (750bp) using control primers provided with the kit.

The following negative PCR controls were used for RACE work:

- No template: To ensure that the PCR cocktail or primers were not contaminated with DNA.
- Gene Racer primer and template and no GSP: To ensure that single primers did not amplify anything due to non-specific binding.
- One GSP and template and no gene Racer primer: To ensure that single primers did not amplify anything due to non-specific binding.

The RACE PCR reaction for the controls were carried out in a total volume of 50 µL containing 1µL HeLa RT template, 1X High Fidelity PCR Buffer, 0.2mM dNTPs, 2mM $MgSO_4$, 0.2µM gene specific primer (control primers A orcontrol primer B.1), gene Racer primer (5' or 3') (0.6µM) and 2.5U Platinum *Taq* Polymerase High Fidelity. To set up the negative control reactions, the HeLa RT template or primers were replaced with sterile ddH_2O.

2.6.5Purification of PCR products

PCR products were purified using the Wizard SV Gel and PCR Clean-Up System (Promega) as per the supplier's protocol. This step was done to remove excessnucleotides and primers from the PCR reactions. This technique is based on the ability of DNA to bind to silica membranes containing chaotropic salts (Promega manual). At first, DNA fragments were separated using gel electrophoresis and the band(s) of interest were excised from the gel and dissolved in 10µL of membrane binding solution (per 10mg of agarose gel) containing guanidine isothiocyanate, followed by incubation at 65˚C for 10 minutes to melt the agarose. Briefly this involved:

- The dissolved gel mixture was transferred and bound to the SV Minicolumn assembly by centrifugation at 10,000g for 1minute.
- The column was washed with 700μL of membrane wash solution, previously diluted with 95% ethanol and centrifuged at 10,000g for 1minute. This was followed by a second wash with 500μL of membrane wash solution and centrifuged at 10,000g for 5 minutes.
- The SV Minicolumn was transferred to a clean 1.5ml microcentrifuge tube and the purified DNA was eluted in 50μL nuclease free water.
- The purified PCR products were stored at -20˚C and were used for cloning and sequencing (Section 2.8).

2.6.6Cloning and sequencing of PCR products

2.6.6.1 Cloning of PCR products

The purified PCR products were cloned into the pCR II-TOPO vector (Invitrogen) followed by the transformation of the recombinant vector into competent *E.coli* cells using the TOPO TA cloning and sequencing kit (Invitrogen).

The cloning reaction was carried out in a total volume of 6μL containing the following reagents: 4μL of fresh purified PCR product, 1μL salt solution (1.2M NaCl, 0.06M $MgCl_2$), 1μL pCR II-TOPO vector and ddH_2O to a final volume of 6μL. The reaction mixture was incubated for 20-30 minutes at room temperature (~22˚C) and then stored on ice prior to transformation.

Immediately after the cloning reaction transformation of the ligated TOPO vector construct into chemically competent *E. coli* cells were performed. One Shot TOP10 chemically competent cells (Invitrogen) [F- *mcr*A Δ (*mrr-hsd*RMS-*mcr*BC) Φ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *deo*R *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G] were used in this study for all transformation experiments. Briefly, this involved:

- 2μL of the TOPO® Cloning reaction were added to a vial containing 50μL One Shot TOP10 chemically competent cells and mixed gently.
- The reaction mixture was the incubated on ice for 30 minutes followed by a heat shock treatment of the cells at 42˚C for 30 seconds.
- The tubes were then transferred to ice immediately and 250μL of SOC medium (Appendix I) was added, followed by shaking the tubes horizontally at 37˚C for one hour at 200rpm.
- Different volumes (50-100) μL of the transformation reaction were then spread on pre-warmed LB-kanamycin/ampicillin plates and incubated at 37˚C overnight.
- White or light blue colonies were picked from the plates for the purification of plasmid DNA as described below (Section 2.6.6.2).

2.6.6.2 Purification of plasmid DNA

Around 10-12 white or light blue colonies were cultured overnight at 37˚C in a shaking incubator (~200 rpm) in LB medium containing 50μg/mL ampicillin. The plasmid DNA was then isolated using the Wizard® Plus MiniprepsDNA Purification System (Promega).
This involved the following steps:

- About 1-10mL of the overnight culture was centrifuged for 5 minutes and the pellet was resuspended in 250μL of cell resuspension solution
- To lyse the cells 250μL of cell lysis solution was added and incubated for 5 minutes.
- To the solution, 10μL of alkaline protease solution was added and incubated for 5 minutes.
- Neutralisation solution (350μL) was added to the tube, mixed well and centrifuged at 10,000g for 10 minutes.
- The clear lysate was then transferred to the SV spin column assembly and centrifuged at 10,000g for 1 minute.
- The column was washed with 750μL of column wash solution, previously diluted with 95% ethanol and centrifuged at 10,000g for 1minute. This was followed by a second wash with 250μL of column wash solution and centrifuged at 10,000g for two minutes.

- The SV spin column was transferred to a clean 1.5mL microcentrifuge tube and the purified DNA was eluted in 100µL nuclease free water.
- The purified plasmid DNA solutions were stored at -20˚C and used for sequencing reactions.

2.6.6.3 Analyzing transformants

Inorder to confirm the presence and sizeof the inserts, the plasmids were analysedbyrestriction enzyme analysis. The restriction digestion was carried out in a total volume of 20uL containing 50-100ng purified plasmid DNA, 0.2µL Acetylated Bovine serum albumin (BSA) (10µg/µL), 2µL 10X Restriction enzyme buffer and 2µL E*coR1* restriction enzyme (10U/µL). The digestion mixture was incubated at 37˚C for ~2-3 hours followed by incubation at 70˚C for 5 minutes to inactivate the enzyme. The E*coR1*digests were electrophoresed in a standard 1% TBE gel (Section 2.5.2) at 80V for 1 hour and the bands were visualised to confirm the presence and the size of the inserts.

2.6.6.4 Sequencing reactions

All plasmid clones with confirmed inserts were sequenced using the ABI Prism Bigdye Terminator V3.1 Cycle Sequencing Kit (Applied Biosystems) which is based on the chain termination method (Sanger *et al.* 1977). The sequencing reaction was carried out in a final volume of 20µL according to a modified version of the standard protocol, containing 1µL BigDye Premix, 3.5µL 5X reaction buffer [250mM Tris HCl (pH 9.0), 10mM $MgCl_2$], 1µL of forward or reverse primer (3-5pM) and plasmid DNA (300-400ng) as the template.The primers which were used for the initial sequencing of cDNA clones were designed on vector based sequences (Table 2.6).

Primer	Sequence 5'→ 3'	Length (bp)
M13 Forward	5´-GTAAAACGACGGCCAG-3´	16
M13 Reverse	5´-CAGGAAACAGCTATGAC-3´	17

Table 2.6 Vector-based primer sequences for sequencing
*Full vectors maps are given in Appendix II

Both forward and reverse primers were used in separate reactions to sequence both strands of each clone in order to ensure accuracy of the sequences in regions of low quality or ambiguous sequences through replicate sequencing reactions. The following cycling conditions were used for the sequencing reaction:

Temperature (˚C)	Time	Number of cycles
96	1 minute	1
96	10 seconds	30
50	5 seconds	
60	4 minutes	
5	2 minutes	1

Table 2.7 Cycling conditions for sequencing reactions

2.6.6.5 Sequencing precipitation of DNA

The sequencing reactions were ethanol precipitated according to the supplied protocol (Applied Biosystems). Briefly this involved:

- 3µL 3M sodium acetate (pH 5.2), 62.5µLethanol (commercial 96% (v/v) analytical grade stock)and 14.5µLddH_2O were added to each of the 20µL sequencing reactions (Section 2.6.6.4).

- The samples were vortexed (Deluxe bench vortex- MT19DL, Chiltern Scientific) briefly and incubated at room temperature for 15 minutes.
- The samples were precipitated by centrifugation at 10,000g for 30 minutes.
- The pellet was washed twice with 75% ethanol and the tubes were allowed to stand at 70-90^{o}C for 1 minute.
- The tubes were wrapped in aluminium foil or stored in dark containers and sent to Micromon DNA sequencing facility,Department of Microbiology, Monash University, Clayton 3168, Australiaand separated on an ABI 373A automated sequencer.

2.6.7 Analysis of sequencing results

All sequencing results were analysed using the BioEdit software package 7.0.9, [http://www.mbio.ncsu.edu/bioedit/bioedit.html; (Hall 1999)].

2.6.7.1 Quality assessment of DNA sequences

The software package Chromas v2.01 (http://www.technelysium.com.au/chromas.html) was used for the quality assessment of the sequenced DNA samples by visual analysis of their trace files. The raw sequence data were imported into this software and ambiguous regions, where the forward and reverse sequences disagreed, were manually corrected by visual checking of the chromatogram, in those cases where this was not possible, replicate sequencing reactions were performed to resolve the ambiguity.

2.6.7.2 Comparison of two sequences by pairwise alignment

In order to compare two sequences pairwise alignments were conducted with the BioEdit software. This programme also providedthe % identity of the aligned regions (and % similarity in the case of amino acid alignments).

2.6.7.3 Alignment of multiple sequences

Multiple sequence alignment was performed withthe “clustalw” program in BioEdit. The “Sequence Identity Matrix” option was been used to create identity matrices which displayed the percentage of identical residues between sequences.

2.6.7.4Identification of open reading frames and translation of sequences

Open reading frames (ORF) within the *SMT2* and *SMO2* cDNA sequences were identified using the ExPASy translate tool (http://web.expasy.org/translate/). Translation of ORFs were conducted using the “translate or reverse translate” command in BioEdit. The start of an ORF was assumed to be an ATG codon ending with a stop codon (TAA, TAG or TGA) and the nucleotide sequences were trimmed to reflect the coding sequences of the cDNAs. The longest ORFs generated this way were confirmed by comparing to non - redundant (nr) database of BLASTx (Altschul *et al*, 1990).

2.6.7.5 BLAST analyses of sequences

The Basic Local Alignment Search Tool (BLAST)was used to identify similar nucleotide (BLASTn) or protein sequences (BLASTx) in the non-redundant (nr) databases (http://blast.ncbi.nlm.nih.gov/Blast/).

2.6.7.6 Amino acid sequence analysis

The predicted hydropathy profile of the putative protein products of the isolated gene sequences were determined using a number of bioinformatic analysis programs including, TMpred(http://www.ch.embnet.org/software/TMPRED_form.html), ProtScale (http://web.expasy.org/cgi-bin/protscale/protscale.pl), SMART (http://smart.embl-heidelberg.de/) and InterProScan (http://www.ebi.ac.uk/Tools/pfa/iprscan/) which makes a prediction of membrane-spanning regions and their orientation. To determine to which protein families the isolated proteins belong to, the Pfam Protein Families Database (http://pfam.janelia.org/) was used.

2.7 AMPLIFICATION OF FULL LENGTH CODING SEQUENCES OF *SMO2* AND *SMT2* FROM BARNEA, FRANTOIO AND PICUAL

2.7.1 Design of oligonucleotide primers

To amplify the full length coding sequences of the *SMO2/SMT2* gene families from Barnea, Frantoio and Picual,primers were designed based on sequences identified in the putative 5' and 3' UTRs of the *SMO2/SMT2* RACE PCR products characterized from Barnea (Section 2.6) (Table 2.8).

Primer pairs and sequences (5'-3')	Annealing temperature used for PCR (°C)	Expected size of PCR product (bp)	Purpose
SMO2-F2:CWCAASCATYCGCGGAWTT **SMO2-R2**: AACRWGCTTCAGAWTGATTCT	50	943	To amplify full length *SMO2* cDNA from three olive cultivars
SMT2-F2 :TAGCCCCTTTCACTCCGC **SMT2-R2** : CAAATAACCAACAAAATCTAACATA	48	1201	To amplify full length *SMT2* cDNA from three olive cultivars

Table 2.8 Primer pairs for the amplification of full length coding sequenceof*SMT2* and *SMO2* genes from Barnea, Frantoio and Picual
Mixed base code: W=AT, S= GC, Y=CT, R= AG

2.7.2 Synthesis of cDNA from total RNA

Full length cDNA transcripts were synthesised from total RNA isolated from Barnea, Frantoio and Picual (Section 2.4.2) using the ThermoScript RT-PCR kit (Invitrogen). First strand cDNA synthesis was performed withThermoScript RT.This involved the following steps: (i) 1µg total olive RNA was primed with oligo dT primer (50µM) at 65ºC for 5 minutes; (ii) the denatured RNA was then added to a reaction mixture containing 1X cDNA synthesis buffer, 0.1M DTT, 40U RNaseOUT and 15U ThermoScript RT enzyme to a total volume of 20 µL; (iii) the reaction mixture was incubated at 55ºC for 45 minutes and (iv) the reaction was terminated by incubating at 85ºC for 5 minutes. *E. coli*RNase H (2U) was added to the mixture to remove the RNA template and incubated at 37ºC for 20 minutes. The cDNA synthesis reactions were stored at -20ºC until further use or were used immediately for PCR reactions to amplify full length cDNAs of the *SMO2*/*SMT2* gene families using gene specific primers (Section 2.7.3).

2.7.3 PCR conditions

Polymerase chain reaction (PCR) was carried out in a total volume of 50 µL containing 50ng template cDNA, 1X PCR Buffer, 0.2mM dNTPs, 1.5mM $MgCl_2$, 0.5µM primer (forward and reverse) and 2U Platinum *Taq* Polymerase. In order to prevent the formation of primer dimers and non-specific primer binding "Hot start" PCR was used through the use of the Platinum Taq Polymerase (Invitrogen). The thermal cycling was conducted in a PTC-100 or PTC-200 Peltier Thermal Cycler (MJ Research). The cycling conditions for the amplification of PCR products are outlined in Table 2.9. All PCR products were analysed by agarose gel electrophoresis to estimate their size(s) (Section 2.5.2). A no-template control (NTC) was used for each PCR reaction that contained all the components of the reaction except the cDNA template. It helps to detect if there is any contamination in the reaction mixes leading to non-specific amplification of products.

Temperature (˚C)	Time	Number of cycles
95	5 minute	1
95	30 seconds	35
Annealing temperature (primer dependent)	30 seconds	
72	2 minutes	
72	10 minutes	1

Table 2.9 Cycling conditions for PCR

2.7.4 Cloning, sequencing and analysis of full length *SMO2* and *SMT2* genes

The full length cDNAs of the *SMO2*/*SMT2* gene families amplified from Barnea, Frantoio and Picual cultivars were gel purified, cloned, sequenced and analysed as described for the RACE PCR products (Sections 2.6.6).

2.8 IDENTIFICATION OF ALLELIC DIFFERENCES OF THE *SMT2* AND *SMO2* GENE FAMILIES BETWEEN BARNEA, FRANTOIO AND PICUAL

Multiple sequence alignments of the full length *SMT2*and *SMO2* sequences isolated and sequenced from the three olive cultivars (Section 2.10) identified transcripts that were present in all three cultivars as well as transcripts with putative cultivar-specific SNPs. To gauge whether the apparently cultivar specific transcripts were present in other cultivars, primers were designed to specifically target those putative alleles to determine whether they were actually unique to that cultivar, or they had just been missed in the initial screen.Alelle-specific primers were designed such that the 3’ nucleotide in the forward and the reverse primer were unique to each of the alleles to be amplified. Allele specific PCR was conducted by starting with an initial annealing temperature determined using the NetPrimer software (www.premierbiosoft.com). To confirm that the PCR products were real and not false positives due to non-specificity of AS-primer annealing,the allele specificity was checked by ensuring unambiguous sequence data was generated from all AS-primer pairs during the

initial amplification and then the annealing temperature was increased in 1˚Cincrements for each allele under investigation. The allele specificity was achieved by checking if the product disappeared simultaneously in all three cultivars at higher annealing temperature; in that case it was assumed that the allele was present in all cultivars. However, if the product disappeared in two cultivars but still appears in the third cultivars, it was interpreted as the putative allele is cultivar specific.The allele specific primers which were designed for *SMT2* and *SMO2* alleles are shown in Table 2.10 and Table 2.11 respectively.

Primer pairs and sequences (5'-3')	Annealing temperature calculated[c]/Actual* (°C)	Expected size of PCR product (bp)
OeSMT2-1d F: CTTAAATTTGAAGCCGGGT R: TTTAAAACCCGGTAGATCT	53/55	303

Table 2.10 Allele specific primers designed for a putatively unique *SMT2* allele
Nucleotides unique to the *SMT2* allele are highlighted in yellow.
[c] Melting temperature determined by NetPrimer software
Actual*: the highest annealing temperature after which the products disappear in the gel

2.9 SOUTHERN BLOTTING

In order to investigate the copy number of *SMT2* and*SMO2* genes inolives, Southern blotting was conducted using theupward capillary transfer technique (Sambrook *et al.* 1989).

2.9.1 Restriction digestion and electrophoresis of gDNA

10µg of genomic DNA isolated from the leaves of olive cultivars Barnea, Frantoio and Picual(Section 2.4.1) were each digested separately with three different restriction enzymes *Eco*RI, *Hind*III and *Bam*H1.

Primer pairs and sequences (5'-3')	Annealing temperature Calculated[c]/Actual (°C)	Expected size of PCR product (bp)
OeSMO2-1e F: CCAGACTAAGAATAATACCCCT R: GAACCATCCACAACCACAA	53/64	401
OeSMO2-1f F: TTATCTTATCCTGTCTTCAAATACG R: TGTGTCCCCAATAGAAAATGAG	57/62	127
OeSMO2-1g F: GACTAAGAATAACACCCCGGCAT R: GGCAAGGGAAGACTACTTCGCAT	61/66	141
OeSMO2-1h F: CCAGTTATGATTTTATCTTATCCTGTT R: CCAAATATTCAGTCCATGTAGACG	53/58	494
OeSMO2-2a F: GCTCAGGGGAAATGTGTCAA R: ATAGTCGTGGAAATCAGCCCCT	58/62	477
OeSMO2-2b F: GCTCAGGGGAAATGTGTCAA R: TTGTGGCAAATCCAAGGAG	58/63	319
OeSMO2-2c F: GCTCAGGGGAAATGTGTCAA R: ATGGAAATCGGCACCC	50/54	473

Table 2.11 Allele specific primers designed for putatively unique *SMO2* alleles
Nucleotides unique to the each of the *SMO2* alleles are highlighted in yellow.
[c] Melting temperature were predicted by NetPrimer software

The following reaction setup was used for restriction digestion:

Component	Volume
genomic DNA	10µg
Acetylated Bovine serum albumin (BSA) (10µg/µL)	0.2µL
10X Restriction enzyme buffer	2µL
Restriction enzyme (10U/µL)	2µL
Total volume	20µl

Table 2.12 Reaction setup for restriction digestion

The restriction digests were incubated at 37˚C for 2-4 hours followed by incubation at 70˚C for 5 minutes to inactivate the enzyme. The digested gDNA was separated in a 1% TAE gel by agarose gel electorophoresis (Section 2.5.2) at 35V overnight. The distance migrated by the molecular weight markers (100bp and 1kbp) (Table 2.2) in the gel were recorded to determine the size of the hybridised probe in the blot.

2.9.2 Southern blotting of gels via upward capillary transfer technique

The agarose gels containing the digested gDNA were prepared for blotting according to the neutral transfer gel treatment protocol outlined by Hybond N+ membrane supplier (Amersham).

This involved the following steps:

- The gel was soaked in three gel volumes of depurination buffer containing 0.125N HCl for 10 minutes with gentle agitation in an orbital shaker and rinsed with ddH_2O. This pre-treatment is done to cleave the immobilized DNA molecules into smaller fragments by removing random purines. This step facilitates the transfer of DNA molecules as smaller fragments transfer more quickly than the larger ones.
- The gel was soaked in three gel volumes of denaturation buffer (0.5M NaOH, 1.5M NaCl) for 30 minutes with gentle agitation in an orbital shaker and rinsed with ddH_2O.
- The gel was then equilibrated with three gel volumes of neutralisation buffer (0.5M Tris-HCl pH 7.2, 1.5M NaCl) for 30 minutes with gentle agitation in an orbital shaker

and rinsed with ddH_2O. The step equilibrates the pH of the solution from alkaline to neutral, resulting in the formation of neutral ssDNA.

- After pre-treatment of the agarose gels, the Southern blotting was set up according to the method of upward capillary transfer (Sambrook *et al.* 1989) as illustrated in Figure 2.6. This involved the transfer of the DNA fragments from the gel to a positively charged Hybond N+ nylon membrane by the capillary action of a high salt buffer.
- The pre-treated agarose gel containing the fractionated restriction fragments was placed on a filter paper wick which was submerged in a reservoir containing the transfer buffer (10X SSC).
- The Hybond N+ membrane was placed on top of the agarose gel and covered with three pieces of Whatman 3MM paper and a stack of paper towel that were held in place with a weight ~0.4-0.5 kg. The transfer was allowed to proceed for 16-20 hours after which the blotting assembly was disassembled.
- The membrane containing the bound DNA fragments were allowed to air dry.
- The crosslinking of the DNA fragments to the membrane was done by exposure to UV radiation on a UV/visible Darkroom (UVP, Upland CA, USA) at 320nm for exactly 4 minutes.

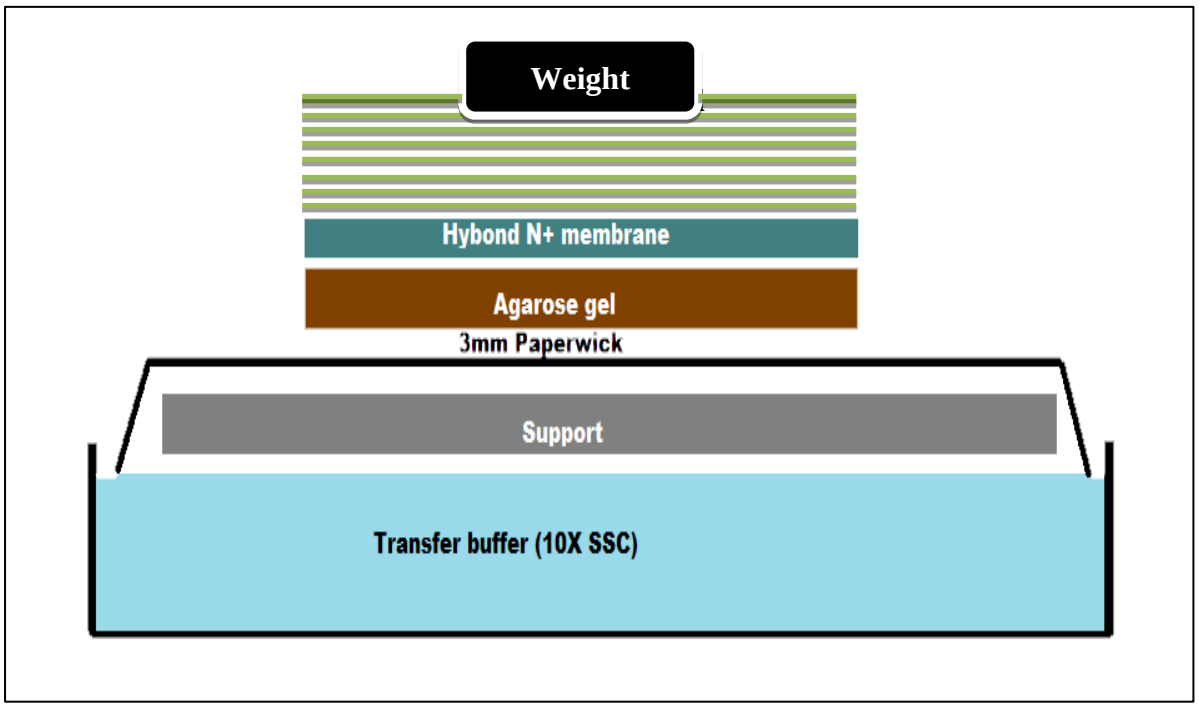


Figure 2.2The upward capillary Southern blotting assembly

2.9.3 Southern Hybridisation using radiolabelled probes

2.9.3.1 Probe design

The Southern blots were probed with either*SMT2* or *SMO2* probes which were amplifiedby RT-PCR using standard PCR reaction conditions (Section 2.7.3). The template RNA was isolated from the olive cultivar Barneaas described in Section 2.4.2. The primer sequences used for the amplification of the probes were based on the RACE cDNA data (Sections 2.7.4) (Table 2.13). Following amplification, the probes were visually assessed for the size by agarose gel electorphoresis (Section 2.5.2), gel purified (Section 2.6.5), sequenced (Section 2.6.6.4) and quantified with the Qubit fluorometer (Section 2.5.1) prior to labelling. These probes were also amplified and sequenced by PCR using the genomic DNA isolated from the olive cultivar Barneaas described in Section 2.4.1 and 2.6.6.4 to ensure the presence/absence of introns within the probes.

Probe	Primer pairs and sequences (5'-3')	Annealing temperature (°C)	Expected size of PCR products (bp)
SMO2	F: TCACGAAAGTGTCTTCTTCTTGTCGG R: ATAACCCTTGTCAGTACCAAATATCC	65	671
SMT2	F: GATAAACAATGCGAAGTCGTGTGCG R:AGTGAAAATTCCCGACTCGCCACCTCTA	65	518

Table 2.13 Primer pairs for designing *SMO2* and *SMT2* probes for Southern blotting

2.9.3.2 Identification of probe specific restriction enzyme recognition sites

The target DNA sequences of the *SMT2* and *SMO2* genes which were used as probes in the Southern blotting experiments were analysed for any restriction enzyme recognition sites for Eco*RI*, *Bam*HI and *Hind*III enzymes within the probe, using the 'Restriction Map' programme in BioEdit.

2.9.3.3 Radiolabeling of probes

The purified *SMT2* and *SMO2* PCR products were radiolabelled with α-^{32}P using the Prime-a-Gene Labelling System (Promega). Briefly, this involved an initial denaturation of the amplified probe (~25ng) at 96-100˚C for 2 minutes followed by addition of 1X of labelling buffer, unlabelled dNTP mixture containing dATP, dGTP, dTTP to a final concentration of 25mM each, nuclease-free BSA to a final concentration of 400ng/µL, [α^{32}P]-dCTP 50µCi, 3,000Ci/mmol to a final concentration of 333nM and 5 units of the Klenow fragment (100U/mL). This reaction mixture was incubated at 37˚C for one hour before terminating the reaction by the addition of 20mM of EDTA. After labelling, any unincorporated labelled nucleotides were removed from the reaction by size-exclusion chromatography using ProbeQuant G-50 micro columns (GE Healthcare) using the suppliers protocol. Briefly, this process involved the column preparation by centrifugation at 735g for one minute, applying 50µL of labelled probe to the column and elution of the probe by centrifugation at 735g for two minutes. The labelled probe was denatured at 96˚C for 5 minutes, briefly chilled on ice and applied to the Southern blots for hybridisation.

2.9.3.4 Hybridisation and washing of the blots

The Southern blots were placed in the Denhardt's pre-hybridisation buffer (Appendix I) in glass hybridisation tubes at 15 rpm for atleast one hour at 65˚Cto minimise any non-specific binding of the probe to the membranes. The denatured radiolabelled probes (Section 2.9.3.3) were added to the prehybridisation buffer andhybridisation was carried out overnight at 65˚C. After hybridisation, the blots were washed twice in a low stringency wash solution (2XSSC, 0.1% SDS) at 65˚Cfor 10 minutes, followed by washing twice, 10 minutes each in a medium stringency wash solution (1X SSC, 0.1% SDS) at 65˚Cand finally four times, 5 minutes each, in a high stringency wash solution (0.1X SSC, 0.1% SDS) at 65˚C.

2.9.3.5 Imaging and detection of hybridised probe

To detect the hybridised probe, the blot was exposed to a BAS-MS 2340 imaging plate (FujiFilm) for 6–12 hours. The scanning of the image was conducted in a FLA-3000

phosphorimager (FujiFilm) and the results were analysed using the Image Gauge v3.12 (FujiFilm) software.

2.10 EXPRESSION ANALYSIS OF *SMT2* AND *SMO2* GENESUSING REAL TIME QUANTITATIVE PCR (qPCR)

The expression of *SMT2* and *SMO2* genes were quantitatively measured across three different olive cultivars Barnea, Frantoio and Picual during the 2009 crop season using real time quantitative PCR approach to identify any expression differences between the cultivars that may contribute to the differences observed in the relative levels of campesterol and sitosterol in olive oils.

2.10.1 RNA isolation and cDNA synthesis

Total RNA isolated from olive fruit tissues harvested at individual timepoints in the 2009 (96, 109, 116 and 136 DAF) crop season from olive cultivars Barnea, Frantoio and Picual (Section 2.4.2) were used for gene expression analysis using qPCR For all samples, the cDNA was synthesized from 100ng of total RNA in a 20µL reaction volume using Thermoscript RT-PCR system as described in Section 2.7.2. The cDNA samples were diluted 5-fold and a final volume of 10µL cDNA was used for all real time PCR reactions.

2.10.2 Preparation of standards and experimental setup for qPCR

Standards for quantitative PCR were used to generate calibration curves to calculate the PCR efficiency of the samples. PCR amplification efficiency is the rate at which a PCR product is generated and is expressed in terms of percentage. In a perfectly efficient reaction the amount of PCR product doubles in quantity with every cycle. Standard curves of real time PCR reactions are represented graphically as a semi-log regression line plot of Cqvalues plotted against log of input nucleic acid (Hellemans *et al.* 2007; Karlen *et al.* 2007; Bustin *et al.* 2009).The slope of the log-linear portion of the standard curve is used to calculate the PCR efficiency by the following equations:

Exponential Amplification = $10^{(-1/slope)}$

Efficiency = $[10^{(-1/slope)}] - 1$

For preparation of the standards, cDNA reactions synthesized from RNA isolated from the individual timepoints in the 2009 (96, 109, 116 and 136 DAF) crop season from Barnea, Frantoio and Picual were pooled and aliquoted in single-use tubes. Five series of 4-fold dilutions were prepared using the undiluted cDNA pool [1:1 (STD1), 1:4 (STD4), 1:16 (STD16), 1:64 (STD64), 1:256 (STD256)]. These dilution series were diluted further in a 1:5 dilution with sterile ddH_2O and a final volume of 10µL cDNA was used in each real time PCR reaction. These five sets of serial dilutions were used as a template for real-time PCR to generate standard curves to determine efficiency of each primer set.

To ensure methodological reproducibility, a total of 12 samples were measured in duplicate with each gene, where all duplicates were independent measurements coming from independent RNA extractions and cDNA synthesis reactions. As the LightCycler Carousel (Roche) has a maximum of 32 wells, all 12 samples in duplicate were measured in one run.
All samples used in the qPCR study, with their cultivar name, date and year of harvest, negative controls, standards and their abbreviations has been shown in Table 2.14. The experimental setup used in the 32-well LightCycler Carousel to study the expression of all genes under investigation between different samples has been shown in Table 2.15.

Sample Name	Cultivar	Developmental stage (DAF)	Year of Harvest
B-1-09	Barnea	96	2009
B-2-09	Barnea	109	2009
B-3-09	Barnea	116	2009
B-4-09	Barnea	136	2009
F-1-09	Frantoio	96	2009
F-2-09	Frantoio	109	2009
F-3-09	Frantoio	116	2009
F-4-09	Frantoio	136	2009
P-1-09	Picual	96	2009
P-2-09	Picual	109	2009
P-3-09	Picual	116	2009
P-4-09	Picual	136	2009

Sample Name	Abbreviation
STD1	Standard 1
STD4	Standard 4
STD16	Standard 16
STD64	Standard 64
STD256	Standard 256
NTC	No template control
NRT	No reverse transcriptase control

Table 2.14 Information for all olive cDNA samples, standards and negative controls used in the qPCR study
Details on preparation of cDNA for each sample/standard/negative control have been summarised in Sections 2.10.1 and 2.10.2

Samples	B-1-09	B-2-09	B-3-09	B-4-09	F-1-09	F-2-09	F-3-09	F-4-09	P-1-09	P-2-09	P-3-09	P-4-09	NTC	NRT	Total number of reactions
Run															28

Table 2.15 Experimental setup used to study the expression of all genes under investigation between different samples
For each gene, expression levels were analysed for 12 samples in duplicate in a single run. Details of all samples with their cultivar name, date and year of harvest, negative controls, standards has been shown in Table 2.14. The green boxes represent individual reactions in duplicate.

2.10.3 Selection of reference genes for normalisation in olives

As of writing this, there was no available published information about the validation of reference genes in the developing fruit of olives(Section 1.95). Therefore, in order to normalise the expression data of *SMT2* and *SMO2* genes in olives it was important to identify stable reference genes that show consistent expression within the olive samples. As information about the stability of reference genes in olives is lacking, reference genes were chosen from previous research studies in plants such as grapevine, *Arabidopsis*, wheat, soybean, and barley (Czechowski *et al.* 2005; Hu *et al.* 2009; Jarošová *et al.* 2010; Gamm *et al.* 2011; Long *et al.* 2011).The list of new *Arabidopsis* reference genes identified using the Affymetrix ATH1 microarray data (Section 1.95)was also used for this purpose. The publication of olive EST sequences has facilitated the genome-wide mining of reference genes in olives (Alagna *et al.* 2009). A total of eight reference genes were chosen based on their stability in other plants as well as the availabilty of putative ESTs for these genes in a publicly available database of olive ESTs (http://www.oleadb.it/) (Georgio 2007) where consensus sequences derived from atleast 10 ESTs were chosen for designing primers for the amplification of each of the selected reference genes (Table 2.17). Four traditional housekeeping genes (Nicot *et al.* 2005; Expósito-Rodríguez *et al.* 2008; Cruz *et al.* 2009; Artico *et al.* 2010) were selected which were GAPDH, TUBA, EF1-alpha and UBQ (Table 2.17). Three new novel reference genes identified in the *Arabidopsis thaliana* Affymetrix ATH1 microarray data (Czechowski *et al.* 2005) were chosen which were PP2A, PTB and TIP2 (Table 2.16). In addition 60S RBP was also included as Gamm, Héloir et al. (2011) have demonstrated that it is one of the suitable reference genes to normalize gene expression data in two different grapevine organs (leaves and berries).

2.10.4 Primer design for reference genes

All primers which were used for qPCR assay (Table 2.16 and 2.17) were designed using the Beacon designer^{TM} software (http://www.premierbiosoft.com). This software helps to design SYBR Green PCR primers that are specific to the target gene by avoiding significant cross homologies to non-specific products. It automatically avoids regions of secondary structures that can prevent the extension of SYBR Green primers. It also gives a rating of the primer pair by a ranking algorithm by determining how closely the primer pairs and primer search parameters match with the target values.

Gene name	Gene Annotation	Suggested function	Accession number/Cluster ID	Primer sequence (5' to 3')	Amplicon length (bp)	Melting temperature (˚C)*	Reference**
60S RBP	60S ribosomal protein L18	component of the large 60S ribosomal subunit	Olea EST database: Cluster ID-OLEEUCI011221	F:GTAAGAGCAAGAAGACCAAG	101	55	(Gamm, Héloir et al. 2011)
				R:GCTTCCAGTTCTCCTCAC			
PP2A	Protein phosphatase 2A	cell growth and signalling	Olea EST database: Cluster ID-OLEEUCI010038	F:AGATCGGTGAAATACTTCCACACG	189	56	(Czechowski, Stitt et al. 2005)
				R:TCGTGGATACTACTCAGTGGAGACTG			
PTB	Polypyrimidine tract-binding protein	alternative splicing	Olea EST database: Cluster ID-OLEEUCI031691	F: CTTCTCCGAAATAAACCAGAT	156	56	(Czechowski, Stitt et al. 2005)
				R: GGTGTCAGCTCCAGTTGTAA			
TUBA	Tubulin alpha	cytoskeletal structural protein	Olea EST database: Cluster ID-OLEEUCI051890	F: AGAACACCTCAGCAACAC	100	51	(Jarošová and Kundu 2010)
				R: AACTACCAGCCACCAACT			
TIP2	Aquaporin tonoplast intrinsic protein	water transport regulation	Olea EST database: Cluster ID-OLEEUCI011159	F: ACTTGTTGTAAGCAATGG	104	51	(Czechowski, Stitt et al. 2005)
				R: TGATTCATTAAGCGTTGG			
UBQ	Ubiquitin carrier protein	protein degradation	AF429430.1	F: AATGAAGTCTGTCTCTCCTTTGG	150	51	(Padilla, Hernandez et al. 2009)
				R: AAGGGAAATCCCATCAACG			
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase	glycolysis	EF506530.1	F: ACAGCTCCTGGTAAGGGTGA	210	56	(Long, Liu et al. 2011)
				R: GGCTTGCGTCAAGAAGTCTC			
EF1-alpha	Elongation factor 1 alpha	protein synthesis	XM_002527974.1	F: GAATGGTGATGCTGGTCTTG	191	56	(Long, Liu et al. 2011)
				R: CACTTGGCACCCTTCTTAGC			

Table 2.16 Details of eight candidate reference genes used for qPCR

*Melting temperature determined by Beacon designer software

**References where each of these genes has been used as reference genes for normalisation in qPCR analysis in other plants have been shown.

The forward and reverse primer pairs designed for the eight reference genes, along with their abbreviated names, function, accession numbers and amplicon length have been shown in Table 2.16.

2.10.5 Primer design for target genes

In this study the expression levels of all of the *SMT2* and *SMO2* alleles which were identified from the sequence analysis of their full length coding sequences (Section 2.7) were assessed throughout olive fruit development. For this, all *SMT2* and *SMO2* alleles were aligned together using the multiple sequence alignment feature of Bioedit (2.6.7.3) and primers based on conserved sequences common to all of the identified alleles were designed using the Beacon Designer software (http://www.premierbiosoft.com) for qPCR analysis (Table 2.10).

Primer pairs and sequences (5'-3')	Annealing temperature used for qPCR (°C)*	Expected size of PCR product (bp)	Purpose
F: CTAAGGAGTTACAAGGAT R: ACTCGCCACCTCTAGT	46	246	To amplify *SMT2* cDNAs for qPCR
F: GAAAGTGTCTTCTTCTTGTCGG R: GACGGCAAGGGAAGACTACT	56	221	To amplify *SMO2* cDNAs for qPCR

Table 2.17 Primer sequences for *SMT2* and *SMO2* probes for qPCR analysis
*Melting temperature determined by Beacon designer software

2.10.6 Amplification and sequencing of the reference and target genes

Primers designed for the eight reference and two target (*SMO2* and *SMT2*) genes were used to amplify these genes from the cDNA (Section 2.10.1) synthesized from pooled Barnea RNA (Section 2.4.2) using typical PCR conditions (Section 2.7.3) to ensure each primer set amplified products of the expected size and sequence. For this, the gel purified PCR products (~10ng) were directly used in the sequencing reactions (Section 2.6.6.4) and the analysis of the sequencing results was conducted using BioEdit software (Section 2.6.7) to confirm the PCR specificity of the primer pairs under the experimental conditions.

2.10.7 qRT-PCR methodology

All reactions were performed in 20µL volumes containing 2µL of primer mix (5µM of each forward and reverse primer) (Table 2.16 and 2.17), 10µL of 5-fold diluted cDNA (10ng) (Section 2.10.1) and 4µL SYBR Green I Master mix reagent. The thermal cycling conditions used for all qPCR reactions are shown in Table 2.18. In order to confirm that the desired product had been amplified a melting curve analysis was performed at the end of PCR. If a single product was amplified, the analysis produced only one melting peak while non-specific products or primer-dimers led to the production of multiple peaks. The sequence of the expected *SMT2/SMO2* amplicons generated in the qPCR analysis were also analysed for their melting temperatures using the Beacon designer^{TM} software (http://www.premierbiosoft.com) to determine any expected melting temperature differences between the alleles.

2.10.8 qPCR analysis

To evaluate the stability of eight reference genes under study across different timepoint samples of three olive cultivars, the expression stabilities of the tested genes were validated with two software programs, qBase Plus software version 2.4 (Biogazelle) and BestKeeper (Pfaffl *et al.* 2004).

LightCycler Programme	Analysis Mode	Cycles	Segment	Target Temperature (˚C)	Time	Acquisition mode**
Pre-incubation	None	1		95	10 minutes	None
Amplification	Quantification	45	Denaturation	95	10 seconds	None
			Annealing	primer-dependent*	0-10 seconds	None
			Extension	72	amplicon-size dependent [amplicon size/25]	Single
Melting curve	Melting curve	1	Denaturation	95	0 seconds	None
			Annealing	65	60 seconds	None
			Melting	95 Ramp Rate= 0.1˚C/second)	0 seconds	Continuous
Cooling	None	1		40	30 seconds	None

Table 2.18 Thermal cycling conditions programmed specifically for the LightCycler Carousel-Based system PCR run with the LightCycler FastStart DNA Master^{PLUS} SYBR Green I dye

*Annealing temperature used for each primer pair are shown in Table 2.16 and 2.17

****** Acquisition mode indicates the frequency with which data is acquired

2.10.8.1 qBase Plus

The quantification of reference genes and target genes were compared by exporting the Cq values from the LightCycler carousel to the qBase Plus software. The raw Cq values were then converted to normalised relative quantities (NRQs) by using the classic delta-delta Ct method with multiple reference genes (Hellemans *et al.* 2007)to derive fold change gene expression using the following equation:

$$NRQ = \frac{E_{goi}^{\Delta Ct,goi}}{\sqrt[f]{\prod_{o}^{f} E_{ref_o}^{\Delta Ct,ref_o}}}$$

where, NRQ= normalised relative quantities

GOI= genes of interest

Ref= reference gene

Ct= Cq (quantification cycle)

E= efficiency

The average expression stability of the reference genes were analysed using the Genorm module integrated in qBase^{plus}. GeNorm is a statistical algorithm which relies on the principle that two ideal reference genes will be equally expressed in all samples irrespective of any experimental condition or tissue type and are minimally regulated (Jarošová *et al.* 2010; Gamm *et al.* 2011; Uddin *et al.* 2011). GeNorm M determines expression stability measure (M value) of all the reference genes under investigation based on the geometric averaging of multiple reference genes and mean pairwise variation of a gene from all other reference genes in a set of samples. Lower M values reflect greater stability of the reference genes (Figure 2.3). GeNorm M ranks the candidate genes from the most unstable gene to a single most stable gene. GeNorm V calculates the normalisation factor (NF_n) by calculating the geometric mean of the expression levels of the stable most reference genes by step-wise inclusion of a less stable gene (Czechowski *et al.* 2005; Uddin *et al.* 2011). The programme calculates the pairwise variation V_n/V_{n+1} between two sequential normalisation factors NF_n and NF_{n+1}. A large variation indicates that the added gene has significant contribution on the normalisation and thus should be included for calculation. If the variation is low (V_n/V_{n+1} <0.15) then that indicates that the added reference gene is not required for calculation of the normalisation factor and thus can be excluded (Figure 2.3).

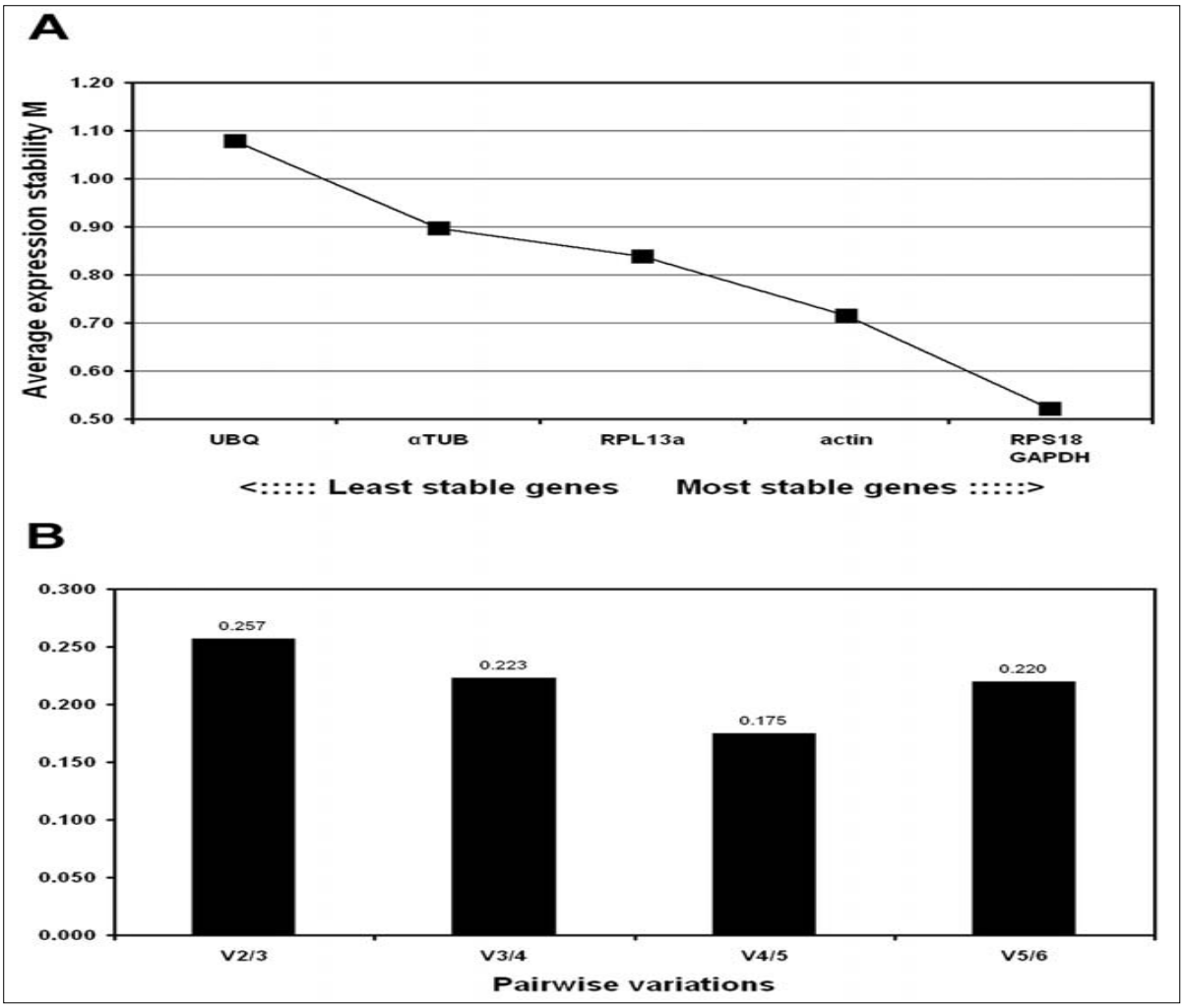


Figure 2.3 Selection of most suitable reference genes in the insect species *Apis mellifera L.* by genorm software (A) Stepwise exclusion of the least stable reference gene by calculating the average expression stability values (M value) of six reference genes plotted from least stable (left) to most stable (right) genes. (B) Determination of optimal number of reference genes for normalisation by pairwise variation analysisbetween two sequential normalisation factors NF_n and NF_{n+1} (Scharlaken *et al.* 2008).

2.10.8.2 BestKeeper analysis

The stability of the eight reference genes was also evaluated using the Excel based tool Bestkeeper (Pfaffl *et al.* 2004). BestKeeper ranks the stability of candidate reference genes by performing a statistical analysis of the Cq values based on three variables: Pearson correlation coefficient (r), standard deviation (SD) and percentage covariance (CV). It performs numerous pair-wise correlation analysis of all pairs of candidate reference genes by combining all highly correlated (and putatively stably expressed) reference genes into an

index value (BestKeeper index) by calculating the geometric mean. If the reference genes are stably expressed, their expression levels will be highly correlated (Pfaffl *et al.* 2004).The correlation between each candidate reference gene and the index is calculated to determine the relationship between the index and the contributing reference genes byPearson correlation coefficient (r), coefficient of determination (r^2) and the probability p values.

2.10.8.3 Statistical analysis

To determine the significant differences between experimental samples, unpaired-t tests, Mann-Whitney 1 tests or one-way ANOVA tests were performed using the built-in statistical tools in the qBase Plus software by comparing the fold change scores. All statistical tests were two-sided and statistical significance was set at a p-value less than or equal to 0.05 with accompanying 95% confidence intervals (low and high).

CHAPTER 3

TECHNOLOGICAL FACTORS AFFECTING STEROLS IN AUSTRALIAN OLIVE OILS

List of Figures and Tables in Chapter 3

ABSTRACT

Phytosterols are important lipids related to the quality of olive oil and broadly used to determine the authenticity of the oil. Recent analyses have identified that some Australian olive oils would not meet international standards for total content of sterols or for certain individual components. Australian olive oil, specifically those derived from thecultivar Barnea (representing 41% of the olive crop in Australia), contain up to 4.8% campesterol, which exceeds the IOOC standards that stipulates a campesterol level of less than 4%. Previous research has shown that the composition of the minor components of the olive oil can be affected by various geographical and technological factors like cultivar, climate, crop year, degree of fruit ripeness, temperature during fruit maturation, storage time of fruits prior to oil extraction, processing and extraction methods. In this chapter the horticultural and processing practices that may have an impact on the sterol content and profile of three cultivars commonly cultivated in Australia were analysed. The evaluation was undertaken using three different cultivars, Barnea, Frantoio and Picual and the processing practices evaluated were: irrigation, fruit size, maturity, malaxing time, malaxing temperature and delays between harvest and process. Processing practices particularly affect triterpene dialcohols and stigmasterol while horticultural practices and fruit characteristics tend to affect more significantly other sterols such as β-sitosterol, sitostanol, Δ5-avenasterol and Δ7-avenasterol. Interestingly, no evaluated management or processing practice seems to have contributed in reducing the campesterol levels in Australian olive oils, particularly those derived from theolive cultivar Barnea, strongly implicating genetic factors as the cause of these elevated levels.

3.1 INTRODUCTION

Phytosterols have been a focus of much research for the last half-century because of their potential health benefits (Section 1.3). A growing source of phytosterols in the Western diet is from olive oil, which is considered a functional food, not only due to the presence of sterols but also because of its high levels of monounsaturated fatty acids (especially oleic acid), antioxidants and other phytochemicals (Alonso, 2006; Covas, 2006).

It has been found that a significant number of olive oil samples of largely cultivated cultivars in Australia do not meet international olive oil standards for the total content of sterols or for certain minor components (Mailer,2006). In particular,Australian olive oils derived from the cultivar Barnea, contain up to 4.8% campesterol, as confirmed by the Australian Government Analytical Laboratories (Mailer 2007) which exceeds the IOOC standards that stipulates a campesterol level of less than 4%. Also some olive oil samples derived from the Frantoio cultivar, showed extremely low total sterol levels, barely above or even under the minimum 1,000 ppmestablished as international limit. It is important to note that Barnea oil represented 41% and Frantoio oil 26% of the olive oils produced in Australia in 2006 (Ravetti 2006). The IOOC standards are merely a generalized standard for the characterization of olive oil by analytical methods that were primarily developed on and therefore suitable for cultivars produced in the traditional olive oil producing countries. These current standards fail to recognise the natural variation of olive oils, which has been observed not only in Australian olive oils but also in many new growingregions such as Chile, Argentina, New Zealand, Spain, France and Italy (Mailer 2007; Mailer *et al.* 2008).This has led to the creation of trade barriers for producers who are producing genuine olive oil, however their products get rejected on the basis of overly stringent regulations (Mailer *et al.* 2008).

Previous studies have reported that sterol composition and total sterol content of olive oils can be affected by cultivar, crop year, degree of fruit ripeness, storage time of fruits prior to oil extraction, processing and by geographical factors (Koutsaftakis *et al.* 1999; Aparicio *et al.* 2002; Ranalli *et al.* 2002). In 2008, Mailer and Ayton conducted a survey on eleven different Australian olive cultivars to study the influence of natural variation (cultivars, site, harvest timing and growing season) on the sterol composition of olive oils. The survey showed that there are some significant correlations between culturaland processing practices

and sterol content and composition, however, interestingly thecampesterol levels in Australian cultivar Barnea remained consistently high regardless of growing season, site or time of harvest(Mailer 2007).

In this chapter, the effect of olive cultivars and season on the phytosterol profile of the olive oils was studied. The influence of major horticultural and olive oil processing practices such as fruit size, irrigation, fruit maturity, malaxing time, malaxing temperature, delays between harvest and process on total sterols and their composition in different Australian olive cultivars has not yet been undertaken. These technological factors were also evaluated to study their impact on sterol levels of Australian olive oils and to further identify whether they are responsible for the highercampesterol levels, as observed in the Australian olive cultivar, Barnea(Guillaume *et al.* 2011).

3.2 EFFECT OF THE CULTIVAR AND YEAR ON STEROL COMPOSITION

3.2.1 Effect of the cultivar on sterol composition

In order to study the effect of olive cultivars on the phytosterol profile of the olive oils, the composition of sterol and triterpene dialcohols were studied in three different olive cultivars, commonly cultivated in Australia, namely Barnea, Frantoio and Picual (Table 3.1). The analysis revealed that with the exception of cholesterol, Δ-campesterol and erythrodiol + uvaol, all other compounds were significantly affected by thecultivar($P<0.001$). Campesterol, β-sitosterol and Δ7-avenasterol were the most affected with F values of 3125, 368 and 451 respectively. The cultivar Barnea showed exceptionally higher levels of campesterol (exceeding the IOOC limit of 4%) in comparison to Frantoio and Picual.

	Frantoio	Picual	Barnea	Std. Err.	F[a]	Significance
Cholesterol	0.16 a	0.14 b	0.13 b	0.006	1.980	0.140
24-Methylene cholesterol	0.27 b	0.29 a	0.18 c	0.005	67.540	0.000
Campesterol	3.39 c	3.53 b	4.88 a	0.028	3125.000	0.000
Campestanol	0.23 a	0.21 b	0.17 c	0.004	16.760	0.000
Stigmasterol	0.92 b	1.07 a	0.73 c	0.013	76.050	0.000
Δ7-Campesterol	0.12 b	0.16 a	0.16 a	0.008	2.978	0.052
Δ7-Stigmastenol	0.40 a	0.39 a	0.34 b	0.006	9.235	0.000
Apparent β-sitosterol[b]	93.87 a	93.80 a	92.96 b	0.028	157.700	0.000
Δ5,23-Stigmastadienol	0.21 a	0.15 b	0.05 c	0.010	26.870	0.000
Clerosterol	1.08 a	1.11 a	0.92 b	0.014	18.050	0.000
β-Sitosterol	82.57 b	86.09 a	85.94 a	0.088	368.000	0.000
Sitostanol	0.69 a	0.62 b	0.70 a	0.007	16.120	0.000
Δ5-Avenasterol	8.61 a	5.47 b	4.93 c	0.083	451.000	0.000
Δ5,24-Stigmastadienol	0.75 a	0.39 b	0.42 b	0.011	174.600	0.000
Δ7-Avenasterol	0.66 a	0.44 c	0.48 b	0.007	127.100	0.000
Erythrodiol + Uvaol	1.08 a	1.01 b	1.05 a	0.013	2.037	0.130
Total Sterols (in ppm)	1,855.44 b	1,731.69 c	1,968.92 a	10.629	47.510	0.000

Table 3.1Sterol and triterpene dialcohol concentrations (values as % total sterols) of oils processed from fruit of three different cultivars: Frantoio, Picual and Barnea
Mean sample size = 216. Means followed by the same Roman letter within each row do not present significant differences (Duncan's multiple range test α = 0.05)
Std. Err.: Standard error
[a] F tests the effect of the cultivar
[b] Apparent β-sitosterol = Δ5,23-stigmastadienol + clerosterol +β-sitosterol + sitostanol +Δ5-avenasterol +Δ5,24-stigmastadienol

3.2.2 Effect of the year on sterol composition

The variations of sterols and triterpene dialcohols between two consecutive years, 2007 and 2008 were studied to identify any seasonal variations that might have an impact on the phytosterol profile of the olive oils (Table 3.2). Most of the compounds were significantly affected by the year, particularly cholesterol, campestanol, stigmasterol, Δ7-stigmasterol, apparent β-sitosterol, Δ5,23-stigmastadienol, Δ5,24-stigmastadienol and erythrodiol + uvaol ($P<0.001$). Campesterol levels were not affected by the season.

	2007	2008	Std. Err.	F[a]	Significance
Cholesterol	0.21 a	0.07 b	0.006	126.500	0.000
24-Methylene cholesterol	0.23 b	0.25 a	0.004	2.974	0.085
Campesterol	3.98 a	3.90 a	0.027	1.617	0.200
Campestanol	0.22 a	0.16 b	0.003	57.680	0.000
Stigmasterol	1.09 a	0.84 b	0.013	72.520	0.000
Δ7-Campesterol	0.16 a	0.08 b	0.005	43.330	0.000
Δ7-Stigmastenol	0.29 b	0.41 a	0.006	91.050	0.000
Apparent β-sitosterol[b]	93.37 a	93.76 b	0.025	45.300	0.000
Δ5,23-Stigmastadienol	0.33 a	0.01 b	0.011	223.800	0.000
Clerosterol	0.93 b	0.95 a	0.007	0.961	0.330
β-Sitosterol	85.16 a	84.44 a	0.087	11.460	0.001
Sitostanol	0.64 b	0.67 a	0.006	2.838	0.093
Δ5-Avenasterol	5.94 b	7.06 a	0.085	30.970	0.000
Δ5,24-Stigmastadienol	0.38 b	0.63 a	0.010	120.100	0.000
Δ7-Avenasterol	0.49 b	0.54 a	0.007	8.120	0.005
Erythrodiol + Uvaol	0.95 a	1.18 a	0.013	62.640	0.000
Total Sterols (in ppm)	1,767.92 b	1,834.06 a	9.854	7.625	0.006

Table 3.2Sterol and triterpene dialcohol concentrations (values as % total sterols) of oils processed from fruit in two different years

Mean sample size = 216. Means followed by the same Roman letter within each row do not present significant differences (Duncan's multiple range test α = 0.05)

Std. Err.: Standard error

[a] F tests the effect of the year

[b] Apparent β-sitosterol = Δ5,23-stigmastadienol + clerosterol +β-sitosterol + sitostanol +Δ5-avenasterol +Δ5,24-stigmastadienol

3.3 EFFECT OF HORTICULTURAL PRACTICES ON STEROL COMPOSITION

The horticultural practices that were evaluated to study their impact on sterol composition of the extracted olive oils were fruit maturity, fruit size and irrigation.

3.3.1 Effect of fruit maturity on sterol composition

The total content of sterols and triterpene dialcohols during different stages of fruit maturation is summarised in Table 3.3. β-sitosterol, sitostanol, Δ5-avenasterol and Δ7-

	<2	2–4	>4	Std. Err.	F^a	Significance
Cholesterol	0.18 a	0.13 b	0.12 b	0.014	2.055	0.130
24-Methylene cholesterol	0.17 b	0.22 a	0.24 a	0.012	3.375	0.038
Campesterol	3.91 b	3.92 b	4.03 a	0.071	0.300	0.740
Campestanol	0.17 b	0.17 b	0.20 a	0.007	1.930	0.150
Stigmasterol	0.75 b	0.77 b	0.83 a	0.020	1.478	0.230
Δ7-Campesterol	0.22 a	0.08 b	0.11 b	0.016	7.445	0.001
Δ7-Stigmastenol	0.37 a	0.31 c	0.34 b	0.017	1.116	0.330
Apparent β-sitosterol[b]	93.81 a	93.89 a	93.61 b	0.067	1.613	0.200
Δ5,23-Stigmastadienol	0.10 a	0.05 b	0.11 a	0.021	0.796	0.450
Clerosterol	0.91 b	0.97 a	1.00 a	0.019	2.146	0.120
β-Sitosterol	87.00 a	84.99 b	84.58 b	0.217	14.880	0.000
Sitostanol	0.95 a	0.58 b	0.56 b	0.023	65.190	0.000
Δ5-Avenasterol	4.43 b	6.83 a	6.78 a	0.221	16.560	0.000
Δ5,24-Stigmastadienol	0.42 c	0.47 b	0.57 a	0.024	3.701	0.028
Δ7-Avenasterol	0.40 b	0.55 a	0.53 a	0.016	10.700	0.000
Erythrodiol + Uvaol	1.16 a	1.02 b	0.92 c	0.028	6.424	0.002
Total Sterols (in ppm)	1,728.99 c	1,915.09 a	1,853.32 b	27.783	4.105	0.019

Table 3.3Sterol and triterpene dialcohol concentrations (values as % total sterols) of oils processed from fruit with maturity index of <2.00, 2.00–4.00 and >4.00

Mean sample size = 36. Means followed by the same Roman letter within each row do not present significant differences (Duncan's multiple range test α = 0.05)

Std. Err.: Standard error

[a] F tests the effect of the maturity index

[b] Apparent β-sitosterol = Δ5,23-stigmastadienol + clerosterol + β-sitosterol + sitostanol +Δ5-avenasterol +Δ5,24-stigmastadienol

avenasterol are significantly ($P < 0.001$) affected by maturity index. Among them sitostanol is the one most affected (F value of 65.2). During ripening, the amount of β-sitosterol decreases, while the amount of Δ5-avenasterol and Δ7-avenasterol significantly increases. Nonetheless, apparent β-sitosterol and campesterol did not change significantly between ripening stages.

3.3.2 Effect of fruit size on sterol composition

The total content of sterols and triterpene dialcohols of oils processed from fruit of small, medium and large size within each cultivar is summarised in Table 3.4. The composition of

	Small	Medium	Large	Std. Err.	F^a	Significance
Cholesterol	0.12 a	0.13 a	0.14 a	0.012	0.130	0.880
24-Methylene cholesterol	0.17 c	0.22 b	0.28 a	0.013	6.519	0.002
Campesterol	4.12 a	3.85 b	3.85 b	0.081	1.228	0.300
Campestanol	0.25 a	0.14 b	0.18 b	0.013	7.878	0.001
Stigmasterol	0.62 b	0.86 a	0.78 a	0.026	7.764	0.001
Δ7-Campesterol	0.23 a	0.26 a	0.19 a	0.031	0.343	0.710
Δ7-Stigmastenol	0.34 b	0.36 b	0.39 a	0.011	2.147	0.120
Apparent *β*-sitosterol[b]	93.66 a	93.83 a	93.48 a	0.087	1.326	0.270
Δ5,23-Stigmastadienol	0.03 a	0.02 a	0.03 a	0.004	1.454	0.240
Clerosterol	1.14 b	1.25 a	1.22 a	0.033	0.935	0.400
β-Sitosterol	86.88 a	85.66 b	83.52 c	0.226	28.680	0.000
Sitostanol	0.86 a	0.65 b	0.51 c	0.018	68.650	0.000
Δ5-Avenasterol	4.25 c	5.83 b	7.74 a	0.205	43.360	0.000
Δ5,24-Stigmastadienol	0.60 a	0.57 a	0.55 a	0.031	0.241	0.790
Δ7-Avenasterol	0.51 b	0.47 b	0.73 a	0.023	14.480	0.000
Erythrodiol + Uvaol	1.19 a	1.12 b	0.89 c	0.031	10.330	0.000
Total Sterols (in ppm)	1,998.16 a	2,002.29 a	1,947.13 b	23.947	0.544	0.580

Table 3.4Sterol and triterpene dialcohol concentrations (values as % total sterols) of oils processed from fruit of small, medium and large size within each cultivar
Mean sample size = 36. Means followed by the same Roman letter within each row do not present significant differences (Duncan's multiple range test α = 0.05)
Std. Err.: Standard error
[a] F tests the effect of fruit size
[b] Apparent β-sitosterol = Δ5,23-stigmastadienol + clerosterol + β-sitosterol + sitostanol +Δ5-avenasterol +Δ5,24-stigmastadienol

campestanol, stigmasterol, β-sitosterol, sitostanol, Δ5-avenasterol, Δ7-avenasterol and erythrodiol and uvaolare significantly affected by fruit size. While the concentration of β-sitosterol, sitostanol and erythrodiol+uvaol significantly decreased with fruit size, concentration of Δ5-avenasterol and Δ7-avenasterol increased (Table 3.4).

3.3.3Effect of irrigation on sterol composition

The analysis of the effect of irrigation on sterol and triterpenedialcohols concentrationsfrom olive fruits receiving three different irrigation regimes: 1/2X, X and 2 X is summarised in Table 3.5. 24-methylene cholesterol, stigmasterol, Δ7-stigmastenol, apparent β-sitosterol and

	½ X	X	2 X	Std. Err.	F^a	Significance
Cholesterol	0.22 a	0.12 c	0.17 b	0.018	2.636	0.076
24-Methylene cholesterol	0.40 a	0.27 b	0.26 b	0.015	9.700	0.000
Campesterol	4.03 a	3.95 a	3.83 b	0.066	0.844	0.430
Campestanol	0.31 a	0.25 b	0.23 b	0.012	4.364	0.019
Stigmasterol	0.92 a	0.80 b	0.75 b	0.020	7.102	0.001
Δ7-Campesterol	0.15 b	0.24 a	0.16 b	0.018	2.810	0.065
Δ7-Stigmastenol	0.54 a	0.51 a	0.41 b	0.009	24.350	0.000
Apparent *β*-sitosterol[b]	92.85 c	93.42 b	93.73 a	0.079	13.200	0.000
Δ5,23-Stigmastadienol	0.16 a	0.08 b	0.09 b	0.018	2.084	0.130
Clerosterol	1.58 a	1.12 b	1.10 b	0.062	7.361	0.001
β-Sitosterol	83.96 c	85.29 a	84.69 b	0.203	3.737	0.027
Sitostanol	0.73 a	0.66 b	0.72 a	0.018	1.543	0.220
Δ5-Avenasterol	5.81 b	5.79 b	6.59 a	0.177	2.284	0.110
Δ5,24-Stigmastadienol	0.59 a	0.49 b	0.55 a	0.023	1.927	0.150
Δ7-Avenasterol	0.60 a	0.48 b	0.49 b	0.013	10.530	0.000
Erythrodiol + Uvaol	0.93 a	0.99 a	0.93 a	0.033	0.358	0.700
Total Sterols (in ppm)	1,933.71 a	1,851.10 b	1,992.07 a	29.511	1.954	0.150

Table 3.5Sterol and triterpene dialcohol concentrations (values as % total sterols) of oils processed from fruits receiving three different irrigation regimes: 1/2X, X and 2X
Mean sample size = 36. Means followed by the same Roman letter within each row do not present significant differences (Duncan's multiple range test α = 0.05)
Std. Err.: Standard error
[a] F tests the effect of irrigation regime during oil accumulation
[b] Apparent β-sitosterol = Δ5,23-stigmastadienol + clerosterol + β-sitosterol + sitostanol +Δ5-avenasterol +Δ5,24-stigmastadienol

Δ7-avenasterol are amongst the significantly affected compounds. It is noteworthy that while stigmasterol and Δ7-stigmastanol decrease with higher levels of irrigation, apparent β-sitosterol significantly increases. There was no significant change in the campesterol content of olive fruits receiving three different irrigation regimes.

3.4 EFFECT OF PROCESSING PRACTICES ON STEROL COMPOSITION

The processing practices that were evaluated to study their impact on sterol composition of the extracted olive oils were malaxing time, malaxing temperature and delays between harvest and processing.

3.4.1 Effect of malaxing time on sterol composition

The effect of the malaxing time of the olive paste on sterol and triterpenedialcohols concentrationsofoils processed at malaxing times of 15, 30 and 60 minutes is summarised in Table 3.6. The composition of erythrodiol+uvaol were the only components to be significantly affected ($P < 0.001$) by malaxing time while stigmasterol and Δ7-stigmasterol were affected but to lesser extent ($P < 0.01$) where they tend to increase with more malaxing time.There was no significant change in the campesterol content of oils processed at different malaxing times.

3.4.2 Effect of malaxing temperature on sterol composition

The composition of erythrodiol+uvaol and the total level of sterols were significantly affected ($P<0.001$) by malaxing temperature, where these components tend to increase with higher malaxing temperature (Table 3.7). The stigmasterol concentration was affected however to a lesser extent ($P= 0.014$). There was no significant change in the campesterol content of oils processed at different malaxing temperatures.

3.4.3 Effect of delay between harvest and process on sterol composition

The delay between harvest and processing significantly affected the percentage of erythrodiol+ uvaol and stigmasterol ($P<0.001$) (Table 3.8). Both erythrodiol+uvaol and stigmasterol levels increased with longer days between harvesting and processing. Campestanol was also significantly affected ($P<0.001$), however it decreased withlarger delays between harvesting and processing.

	15 min	30 min	60 min	Std. Err.	F^{a}	Significance
Cholesterol	0.15 a	0.14 a	0.13 a	0.015	0.121	0.890
24-Methylene cholesterol	0.26 a	0.25 a	0.22 b	0.007	2.471	0.089
Campesterol	4.00 a	3.99 a	3.85 b	0.062	0.583	0.560
Campestanol	0.23 a	0.20 b	0.19 b	0.011	1.658	0.200
Stigmasterol	0.89 c	0.97 b	1.07 a	0.028	3.490	0.034
Δ7-Campesterol	0.13 a	0.13 a	0.09 b	0.012	1.082	0.340
Δ7-Stigmastenol	0.30 b	0.32 b	0.41 a	0.017	4.198	0.018
Apparent β-sitosterol[b]	93.58 a	93.55 a	93.60 a	0.057	0.008	0.920
Δ5,23-Stigmastadienol	0.17 a	0.14 b	0.13 b	0.023	0.239	0.820
Clerosterol	0.82 b	0.85 a	0.86 a	0.020	0.317	0.730
β-Sitosterol	85.23 a	84.84 b	84.67 b	0.206	0.633	0.530
Sitostanol	0.62 a	0.63 a	0.63 a	0.011	0.004	0.970
Δ5-Avenasterol	6.29 b	6.62 a	6.79 a	0.203	0.522	0.590
Δ5,24-Stigmastadienol	0.45 b	0.47 b	0.52 a	0.029	0.406	0.670
Δ7-Avenasterol	0.49 a	0.46 b	0.48 a	0.014	0.506	0.600
Erythrodiol + Uvaol	0.81 c	1.05 b	1.17 a	0.031	14.810	0.000
Total Sterols (in ppm)	1,710.33 b	1,813.10 a	1,794.17 a	24.160	1.732	0.180

Table 3.6Sterol and triterpene dialcohol concentrations (values as % total sterols) of oils processed at malaxing times of 15, 30 and 60 min

Mean sample size = 36. Means followed by the same Roman letter within each row do not present significant differences (Duncan's multiple range test α = 0.05)

Std. Err.: Standard error

[a] F tests the effect of malaxing time

[b] Apparent β-sitosterol = Δ5,23-stigmastadienol + clerosterol +β-sitosterol + sitostanol +Δ5-avenasterol +Δ5,24-stigmastadien

	18 °C	28 °C	38 °C	Std. Err.	F^{a}	Significance
Cholesterol	0.17 a	0.21 a	0.21 a	0.015	0.764	0.470
24-Methylene cholesterol	0.26 a	0.26 a	0.25 a	0.008	0.208	0.810
Campesterol	3.95 a	3.98 a	3.93 a	0.064	0.004	0.960
Campestanol	0.21 a	0.22 b	0.21 a	0.008	0.213	0.810
Stigmasterol	0.89 b	0.95 b	1.12 a	0.035	4.439	0.014
Δ7-Campesterol	0.13 b	0.09 b	0.18 a	0.013	3.798	0.026
Δ7-Stigmastenol	0.34 b	0.39 b	0.38 b	0.012	1.526	0.220
Apparent *β*-sitosterol[b]	93.59 a	93.35 a	93.22 a	0.066	2.706	0.071
Δ5,23-Stigmastadienol	0.23 a	0.24 a	0.28 a	0.034	0.174	0.840
Clerosterol	0.98 a	0.99 a	0.99 a	0.019	0.003	0.970
β-Sitosterol	84.21 a	84.05 a	84.33 a	0.191	0.179	0.840
Sitostanol	0.66 a	0.69 a	0.68 a	0.012	0.658	0.520
Δ5-Avenasterol	6.96 a	6.85 a	6.45 a	0.190	0.657	0.520
Δ5,24-Stigmastadienol	0.57 a	0.53 a	0.49 a	0.023	1.047	0.350
Δ7-Avenasterol	0.53 b	0.62 a	0.55 b	0.019	2.044	0.130
Erythrodiol + Uvaol	0.86 c	1.01 b	1.25 a	0.034	13.980	0.000
Total Sterols (in ppm)	1,669.97 c	1,806.86 b	1,924.26a	25.487	9.656	0.000

Table 3.7Sterol and triterpene dialcohol concentrations (values as % total sterols) of oils processed at temperatures of 18, 28 and 38˚C
Mean sample size = 36. Means followed by the same Roman letter within each row do not present significant differences (Duncan's multiple range test α = 0.05)
Std. Err.: Standard error
[a] F tests the effect of malaxing temperature
[b] Apparent β-sitosterol = Δ5,23-stigmastadienol + clerosterol +β-sitosterol + sitostanol +Δ5-avenasterol +Δ5,24-stigmastadienol

	<12 h	48 h	120 h	Std. Err.	F[a]	Significance
Cholesterol	0.08 a	0.08 a	0.06 b	0.009	0.815	0.450
24-Methylene cholesterol	0.25 a	0.25 a	0.23 a	0.007	0.435	0.650
Campesterol	3.90 a	3.88 a	3.90 a	0.063	0.001	0.990
Campestanol	0.21 a	0.17 b	0.15 b	0.006	9.638	0.000
Stigmasterol	0.98 b	1.07 b	1.31 a	0.037	8.241	0.001
Δ7-Campesterol	0.07 b	0.07 b	0.14 a	0.011	5.816	0.004
Δ7-Stigmastenol	0.37 a	0.33 b	0.31 b	0.011	2.568	0.082
Apparent β-sitosterol[b]	93.56 a	93.63 a	93.42 b	0.045	1.934	0.150
Δ5,23-Stigmastadienol	0.21 a	0.25 a	0.12 b	0.025	2.517	0.086
Clerosterol	0.98 a	0.95 b	0.94 b	0.011	1.084	0.340
β-Sitosterol	84.25 a	84.66 a	84.80 a	0.220	0.552	0.580
Sitostanol	0.65 a	0.62 b	0.61 b	0.011	1.018	0.360
Δ5-Avenasterol	6.90 a	6.62 b	6.50 b	0.215	0.303	0.740
Δ5,24-Stigmastadienol	0.58 a	0.52 b	0.47 b	0.026	1.218	0.300
Δ7-Avenasterol	0.58 a	0.54 a	0.48 b	0.017	2.630	0.077
Erythrodiol + Uvaol	1.00 a	1.11 b	1.39 c	0.032	15.670	0.000
Total Sterols (in ppm)	1,817.20 a	1,802.15 a	1,776.41 a	18.0228	0.432	0.650

Table 3.8Sterol and triterpene dialcohol concentrations (values as % total sterols) of oils extracted from fruit within 12 hour of harvesting, 48 hour from harvesting and 120 hour from harvesting

Mean sample size = 36. Means followed by the same Roman letter within each row do not present significant differences (Duncan's multiple range test α = 0.05)

Std. Err.: Standard error

[a] F tests the effect of the delay between harvesting and processing

[b] Apparent β-sitosterol = Δ5,23-stigmastadienol + clerosterol +β-sitosterol + sitostanol +Δ5-avenasterol +Δ5,24-stigmastadienol

3.5 DISCUSSION

Analysis of the sterol content of olive oils derived from the commonly cultivated Australian olive cultivars have shown that significant number of olive oil samples do not meet international olive oil standards for the total content of sterols or for certain minor components (Mailer *et al.* 2006; Mailer 2007; Mailer *et al.* 2008). Previous research hasreported that sterol composition and total sterol content of olive oils can be affected by cultivar, crop year, degree of fruit ripeness, harvest time, site, storage time of fruits prior to oil extraction, processing and by geographical factors (Koutsaftakis *et al.* 1999; Aparicio *et al.* 2002; Ranalli *et al.* 2002; Mailer *et al.* 2008).

In this chapter, the sterol composition of olive oils was analysed during two consecutive seasons and in three different olive cultivars, commonly cultivated in Australia and known to have contrasting phytosterol profiles (Mailer *et al.* 2008), namely Barnea, Frantoio and Picual, to confirm and further investigate the effect of cultivar and harvest season on the phytosterol profile of the olive oils. Furthermore, the horticultural and processing practices that may have an impact on the sterol content and profile of the olive oils extracted from these olive cultivars, were also analysed. The horticultural practices that were evaluated were irrigation, fruit size and maturity. The processing practices that were evaluated were malaxing time, malaxing temperature and delays between harvest and processing.

The analysis of the sterol composition of olive oils extracted from three different cultivars and two consecutive years have shown that though most of the sterol compounds were significantly affected by the season, the cultivar has shown the most significant level of effect on the composition of different sterols (Figure 3.1). Only cholesterol, erythrodiol + uvaol and Δ7-campesterol had levels of significance higher than 0.001. All other compounds investigated were significantly affected by the cultivar. The cultivar Barnea showed exceptionally higher levels of campesterol which exceeded the IOOC limit of 4%.

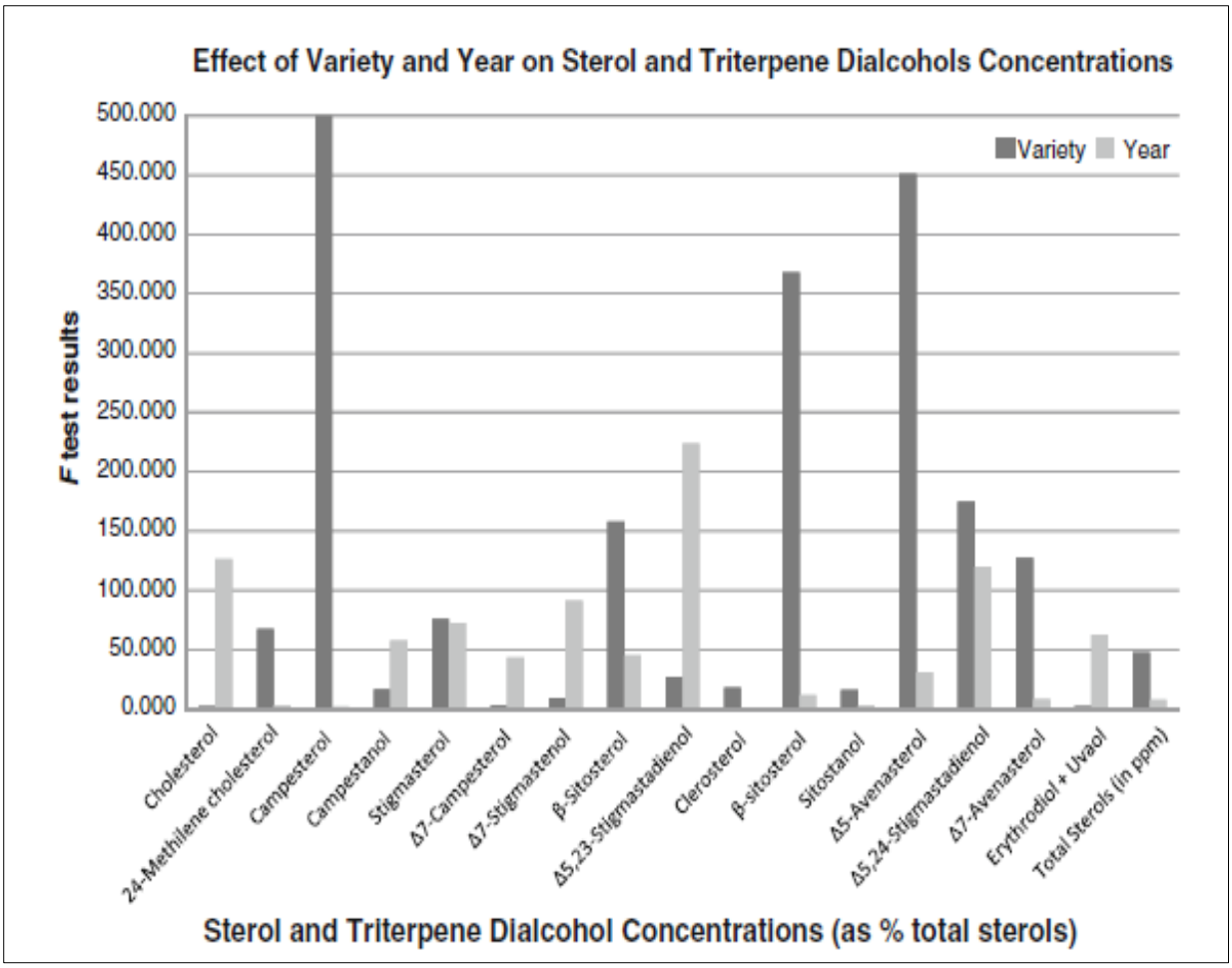


Figure 3.1Effect of cultivar and year on sterol and triterpene dialcohols concentrations

Horticultural practices such as irrigation and fruit characteristics such as maturity and size showed that most of these compounds such as β-sitosterol, sitostanol, Δ5-avenasterol and Δ7-avenasterol are significantly affected which is in broad agreement with previous studies (Figure 3.2) (Koutsaftakis *et al.* 1999). However, the campesterol levels remained consistently high regardless of fruit size/maturity/irrigation regimes.

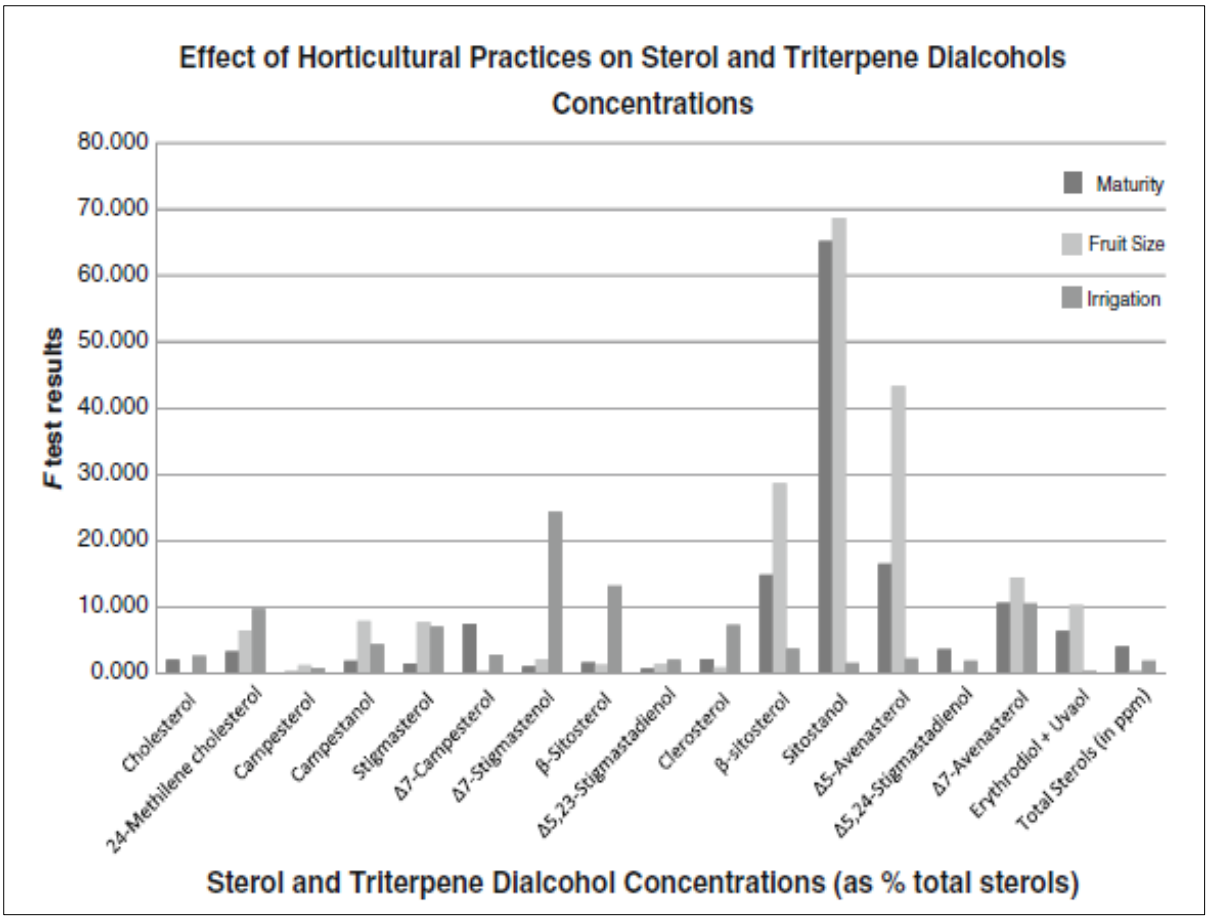


Figure 3.2Effect of horticultural practices on sterol and triterpene dialcohols concentrations

The malaxing time at the paste preparation stage is a very important parameter of good manufacturingpractice. Differences in malaxing time of the olive paste significantly affected only the erythrodiol+uvaol content and to a lesser extent, the content of stigmasterol and Δ7-stigmasterol.

Processing temperature isanother important parameter during the olive oil manufacturing process. Correspondingly, the composition of erythrodiol+uvaol was significantly increased with increase in malaxing temperature which is in agreement with previous research studies (Koutsaftakis *et al.* 1999). In addition, the total sterol levels were also significantly affected by malaxing temperature.

Similar to the other processing parameters (such as malaxing time/temperature) evaluated, the delay between harvest and processing significantly affected the percentage of erythrodiol+ uvaol and stigmasterol.

Thus all processing practices had asignificant impact on the concentrations of triterpene dialcohols and stigmasterol, howeverthere was no significant change in the campesterol content of these oils (Figure 3.3).

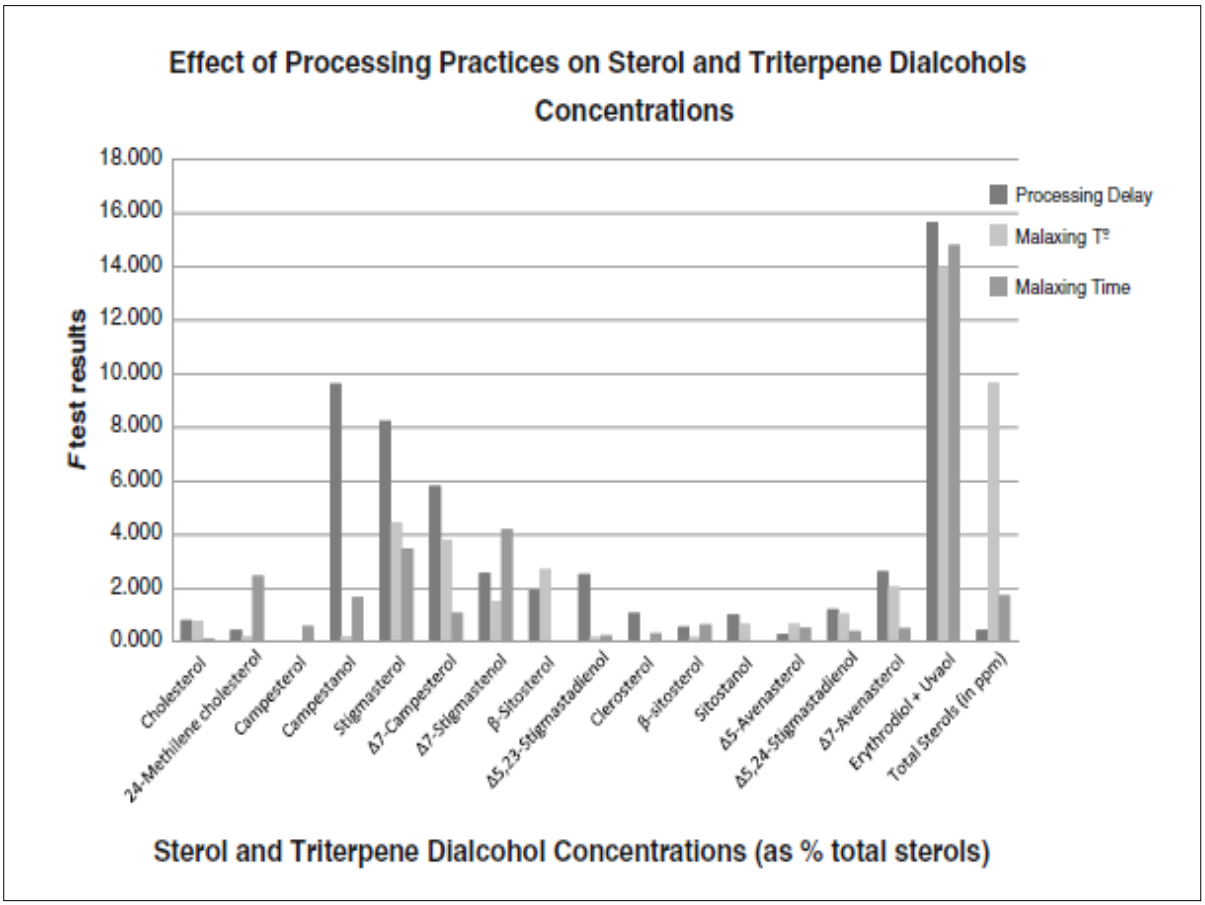


Figure 3.3Effect of processing practices on sterol and triterpene dialcohols concentrations

In conclusion, different horticultural and processing practices seem to have a significant impact on some aspects of the sterol content and profile of the most commonly cultivated cultivars of olive in Australia, however no evaluated management or processing practices seems to have contributed in affecting the relative campesterol levels, in these olive oils. It is important to point out that though most of the compounds were significantly affected by the season, the cultivar has shown the most significant level of effect on the composition of different sterols, particularly the campesterol level in the cultivarBarnea which exceeds the IOOC limit of 4%. Previous research studies has shown that oils extracted from olive cultivars grown in other parts of the world such as Argentina, Spain and Greece also contain higher campesterol levels (Aparicio *et al.* 1997; Salvador *et al.* 2001; Ceci *et al.* 2007;

Pardo*et al.* 2011) however all these olive oil samples are not accepted as extra virgin olive oils in the international market on the basis of stringent IOOC regulations (Section 1.6).

These consistent levels of campesterol observed within these Australian olive cultivars across more than one season and despite various horticultural and processing practices (Section 3.2, 3.3 and 3.4)(Salvador *et al.* 2001) strongly suggest that the genetic makeup of the cultivars is the most important determinant of relative sterol levels. Previous research has shown that the orientation of the sterol biosynthetic flux towards sitosterol or campesterol is mainly controlled by the two branchpoint enzymes, SMT2 and SMO2 in plants (Section 1.9.1). Therefore these results warranted an investigation into these key enzymes of the sterol pathway in olives which may conceivably play a role in determining their individual phytosterol profiles in these cultivars.

CHAPTER 4

MOLECULAR CHARACTERISATION OF THE*STEROL METHYL TRANSFERASE-2 (SMT2)* GENE FAMILY OF OLIVES

List of Figures and Tables in Chapter 4

ABSTRACT

Analyses of phytosterol levels in Australian olive (*Olea europaea* L.) oils have identified that the relative levels of these compounds vary between cultivars. Notably, extra virgin oils derived from the cultivarBarnea(representing 41% of the olive crop in Australia), have campesterol levels up to 4.8% of total phytosterol content, which exceeds the international standards for extra virgin olive oil that stipulate a campesterol level of less than 4%.The sterol content of the Barnea cultivar, on a year to year basis, has shown fluctuations in total and relative sterol levels, however campesterol levels have remained consistently high, strongly suggesting that the sterol levels are influenced by genetic factors. The characterized sterol biosynthetic pathway in plants contains a bifurcation that leads to the formation of β-sitosterol or campesterol, with the flux controlled by the activity of two branch-point enzymes, SAM-24-methylene-lophenol-C-24-methyltransferase2 (SMT2) and C-4α-sterol-methy-oxidase2 (SMO2). Experimental evidence has demonstrated that SMT2 can influence campesterol and β-sitosterol levels in tobacco.Thus, it is conceivable that the relative activity or expression of this enzyme could play a pivotal role in determining the relative amounts of β-sitosterol and campesterol in Australian olive oils. In this chapter, characterization of full length *SMT2* cDNAs were studied from olive using 5'-3' RACE approach to identify any allelic differences between the olive cultivars Barnea, Frantoio and Picual.The RACE data of the *SMT2* cDNA generated a 1105bp 5' RACE product and two different sized 3' RACE products (768 and 747bp). This RACE data was used to design primers flanking the putative coding sequence and led to the amplification of a 1201bp product from RNA isolated from the olive cultivarsBarnea, Frantoio and Picual.Sequencing and analysis of the full length cDNA identified an ORF of 1083bp encoding a putative protein of 360 amino acids. Further analysis identified four putative *SMT2* coding sequences in olive, namely *OeSMT2-1a, OeSMT2-1b, OeSMT2-1c* and *OeSMT2-1d.* Comparison of these *SMT2* sequences between the olive cultivarsBarnea, Frantoio and Picual revealed high conservation between cultivars (99.4-100% sequence identity). Southern blotting approach was adopted to investigate the copy number of *SMT2* genes and the results suggested at least two copies of the *SMT2* gene in the olive genome. The presence of four distinct *SMT2* sequences expressed in the developing fruit of the diploid olive may represent four alleles of two *SMT2* genes. Results from the allele specific PCR suggested that no *SMT2* allelic differences seem to be present between the olive cultivars which was further supported by the Southern blotting results which failed to identify any distinct RFLPs between the olive cultivars.

4.1 INTRODUCTION

As discussed in Section 1.9.1 investigations into the gene families encoding SAM-24-methylene-lophenol-C-24-methyltransferase2 (SMT2) protein are of interest due to its potential role in determining the ratio of campesterol to β-sitosterol in plants.

Analyses of phytosterol levels in Australian olive oils have identified that Australian olive oils, specifically those derived from the cultivar*Barnea* (representing 41% of the olive crop in Australia), contain up to 4.8% campesterol, which exceeds the IOOC standards for extra virgin olive oil that stipulate a campesterol level of less than 4% (Section 1.7)(Mailer 2007). It has been observed that although relative sterol percentages in Australian olive oils fluctuate seasonally, general trends in sterol profiles appear to remain consistent within individual cultivars, strongly implicating genetic factors as the cause of these different levels. The sterol biosynthetic pathway (Section 1.9.1) is essentially linear from cycloartenol until it reaches 24-methylene-lophenol, with a bifurcation downstream of this compound resulting in two pathways leading to the formation of β-sitosterol or, alternatively, campesterol, with the flux controlled by the activity of two branch-point enzymes, SAM-24-methylene-lophenol-C-24-methyltransferase2 (SMT2) and C-4α-sterol-methyl-oxidase2 (SMO2)(Schaeffer *et al.* 2001; Darnet *et al.* 2004) (Figure 1.13). Research has demonstrated that experimentally modulating SMT2 levels can influence campesterol and β-sitosterol levels in tobacco, *Arabidopsis* and soybean (Schaeffer *et al.* 2000; Schaeffer *et al.* 2001; Neelakandan *et al.* 2009). Thus, it is conceivable that the relative activity or expression of these enzymes could play a pivotal role in determining the relative amounts of β-sitosterol and campesterol in Australian olive oils.

Our current knowledge of the *SMT2* gene families in other plants have been summarized in Section 1.9.2. Briefly, past research work on the SMT2 gene families has resulted in the identification of two isoforms of SMT2 in *Arabidopsis thaliana, Glycine max*(soybean),*Nicotiana benthamiana*(tobacco) and *Gossipum hirsutum*(cotton)(Section 1.9.2.3). Closer inspection of these SMT2 amino acid sequences revealed four conserveddomains in the protein sequences corresponding to the substrate binding segment for the sterol (Region I, Region IIIand Region IV) and AdoMet (Region II). However, to date nothing is known about the genes encoding these enzymes in olives. Therefore a better understanding of the sequence information of this gene family in olives was required to study

whether they play a role in determining the relative amounts of β-sitosterol and campesterol in Australian olive oils.

In a recent study, the transciptome of olive drupes from two olivecultivars, Coratina and Tendellonewas studied using454 pyrosequencing technologies (Alagna *et al.* 2009). This chapter describes the use of this data as a starting point in the isolation and characterization of the full length coding sequences of the *SMT2* genes from three different olive cultivars, Barnea, Frantoio and Picual and an analysis of the putative allelic differences between these olive cultivars that could conceivably play a role in impacting the individual phytosterol profiles of the oils derived from them.

4.2 CHARACTERIZATION OF5' AND 3' RACE PRODUCTS OF *SMT2* GENES IN BARNEA

4.2.1Screening and identification of partial *SMT2* sequences

A BLASTn screen of the 454 sequencing data generated from developing olives with the *SMT2* sequences of *Arabidopsis* revealed a 365bp sequence (Cluster Id: OLEEUCl064665)(Alagna *et al.* 2009) with 72.8% identity to *SMT2-1*(Accession number:NM101884) and 70.2% identity to *SMT2-2*(Accession number:U71400.1) from *A. thaliana*. Subsequent translation revealed that this partial sequence encoded a 122 amino acid sequence that was 83.6% and 77.8% identical to the homologous section of the *A. thaliana* SMT2-1 and SMT2-2 enzymes. This partial sequence was used for the amplification and cloning of the uncharacterized 5' and 3' regions of the *SMT2* genes in the olive cultivar Barnea.

4.2.2RNA extraction from Barnea fruits

Total RNA was extracted from the mesocarp of Barnea fruits harvested in the 2009 season for eight sample timepoints between 96 and 170 DAF revealing all samples had intact, high quality RNA (Figure 4.1).

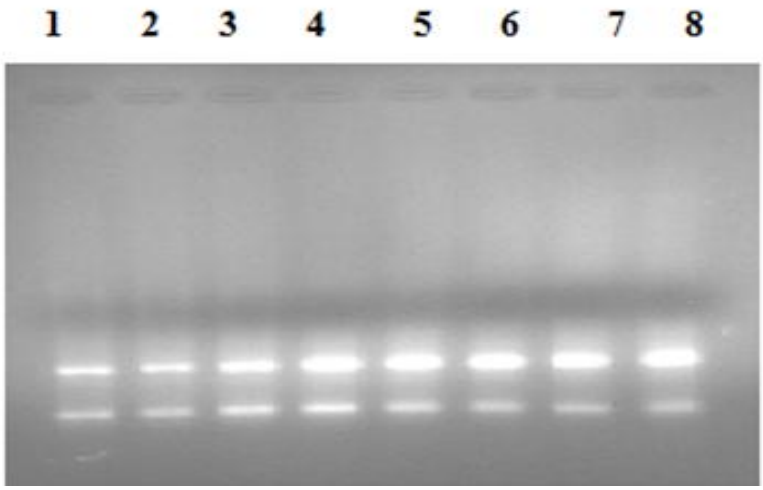


Figure 4.1 Total RNA extracted from eight timepoint samples of cv Barnea fruits(Lane1: 96 DAF, Lane2: 109 DAF, Lane3: 116 DAF, Lane4: 122 DAF, Lane5: 136 DAF, Lane6: 150 DAF, Lane7: 167 DAF, Lane8: 170 DAF)

4.2.3 RACE PCR products of *SMT2* genes from cv. Barnea

To ensure the RACE technique was functional, the RACE-ready cDNA synthesized from the control HeLa total RNA supplied with the GeneRacer kit (Invitrogen), was used to conduct RACE PCR (Section 2.6.4) using theGene Racer 5′ Primer and Control Primer B.1 to amplify the 5′ end of the β-actin cDNA (~900bp) and Gene Racer 3′ Primer and Control Primer A to amplify the3′ end of the β-actin cDNA (~1800bp) which resulted in the amplification of products of expected size (data not shown). Control Primer A and Control Primer B.1 was also used to amplify a partial section (~750bp) of the β-actin cDNA which also resulted in the amplification of expected size product (data not shown).

3’ RACE PCR was conducted using the 3’ GeneRacer primer and the forward GSP (Section 2.6.3 and Table 2.4) which resulted in the amplification of a ~750bp fragment (Figure 4.2).

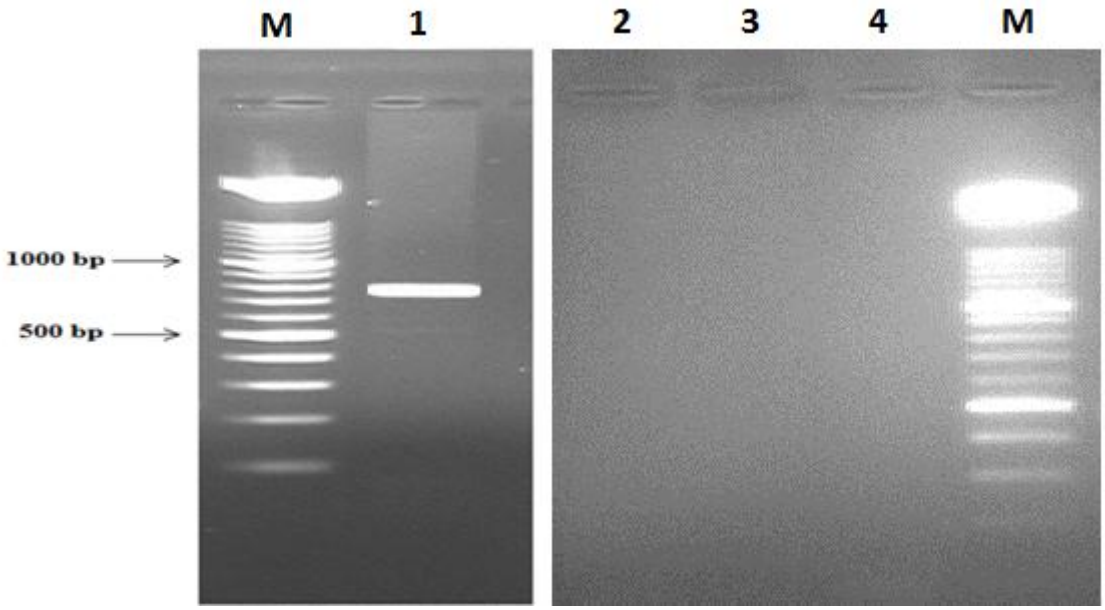


Figure 4.2 Agarose gel electrophoresis of 3' RACE PCR product of *SMT2* gene amplified from cv. Barnea RNA

M: molecular weight marker; Lane1: Barnea *SMT2*3' RACE product; Lane 2:water only negative control; Lane 3:negative control (Gene Racer primer and template and no GSP);Lane 4:negative control (GSP and template and no gene Racer primer)

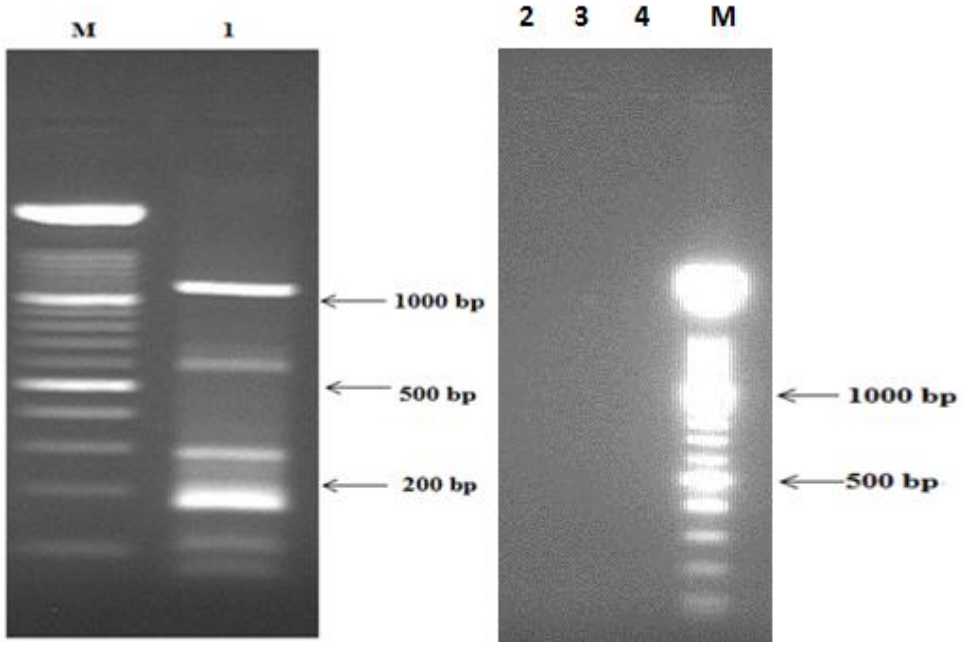


Figure 4.3Agarose gel electrophoresis of 5' RACE PCR product of *SMT2* gene amplified from cv. Barnea RACE cDNA

M: molecular weight marker; Lane1: Barnea *SMT2*5' RACE; Lane 2:water only negative control; Lane 3:negative control (Gene Racer primer and template and no GSP);Lane 4:negative control (GSP and template and no gene Racer primer)

5’ RACE PCR was conducted using the 5’ GeneRacer primer and the reverse GSP (Section 2.6.3 and Table 2.4) resulting in the amplification of multiple products of different sizes: ~1100bp, ~650bp, ~275bp and ~175bp (Figure 4.3).

The four products amplified in the 5’ RACE reaction (175bp, 275bp, 650bp and 1100bp) and the ~750bp 3’ RACE product from Barnea were gel purified, cloned and sequenced as described in Section 2.6.5 and 2.6.6. The digestion of ten plasmids containing the cloned 3' RACE products revealed eight inserts in the expected size range although small variations in insert size appear to be present (Figure4.4). The digestion of ten plasmids containing the ~1100bp cloned 5' RACE products revealed nine inserts in the expected size range (Figure 4.5). The digestion of 5 plasmids containing each of the smaller sized 5’ RACE products (650bp and 275bp)revealed expected size products. The digestion of 5 plasmids containing the 175bp 5’ RACE product revealed only three of the 5 clones with expected size inserts.

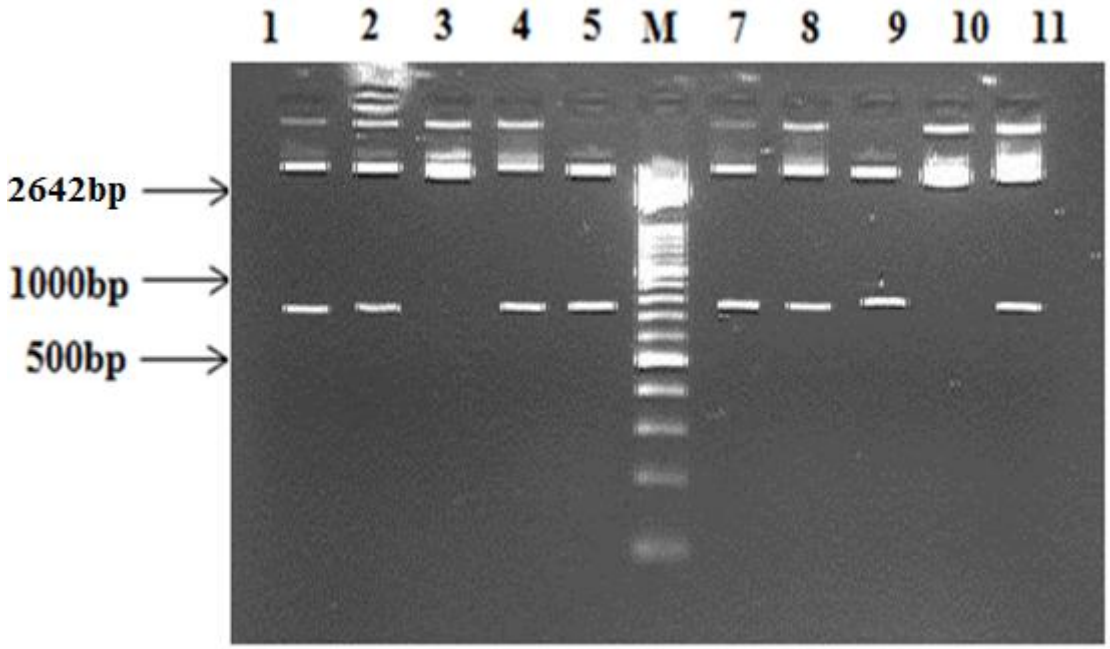


Figure 4.43’ RACE PCR results for cv. Barnea *SMT2* clones
*Eco*RI digested 5’ RACE clones:SMT2-1 (3’), SMT2-2 (3’), SMT2-3 (3’), SMT2-4 (3’), SMT2-5 (3’)(Lanes 1-5 respectively) and SMT2-6 (3’), SMT2-7 (3’), SMT2-8 (3’), SMT2-9 (3’), SMT2-10 (3’) (Lanes 7-11 respectively), M: 100bp molecular weight marker.

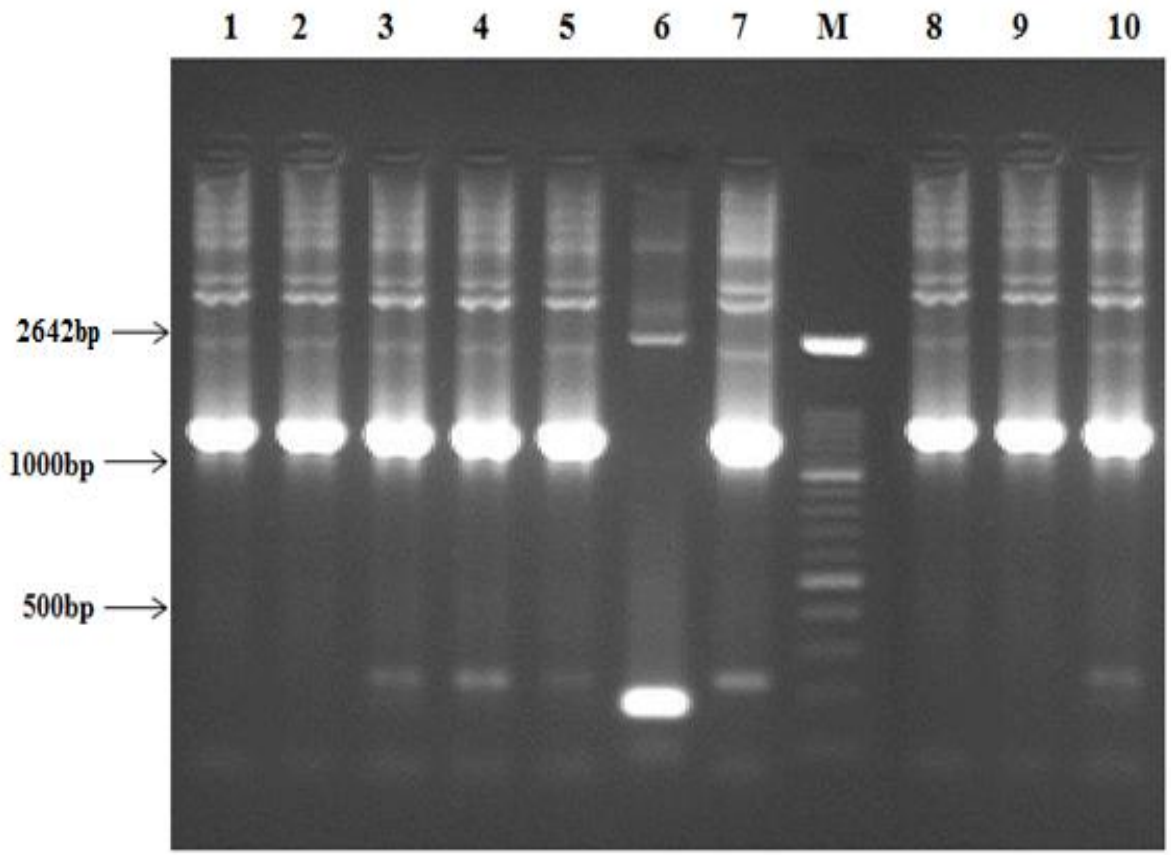
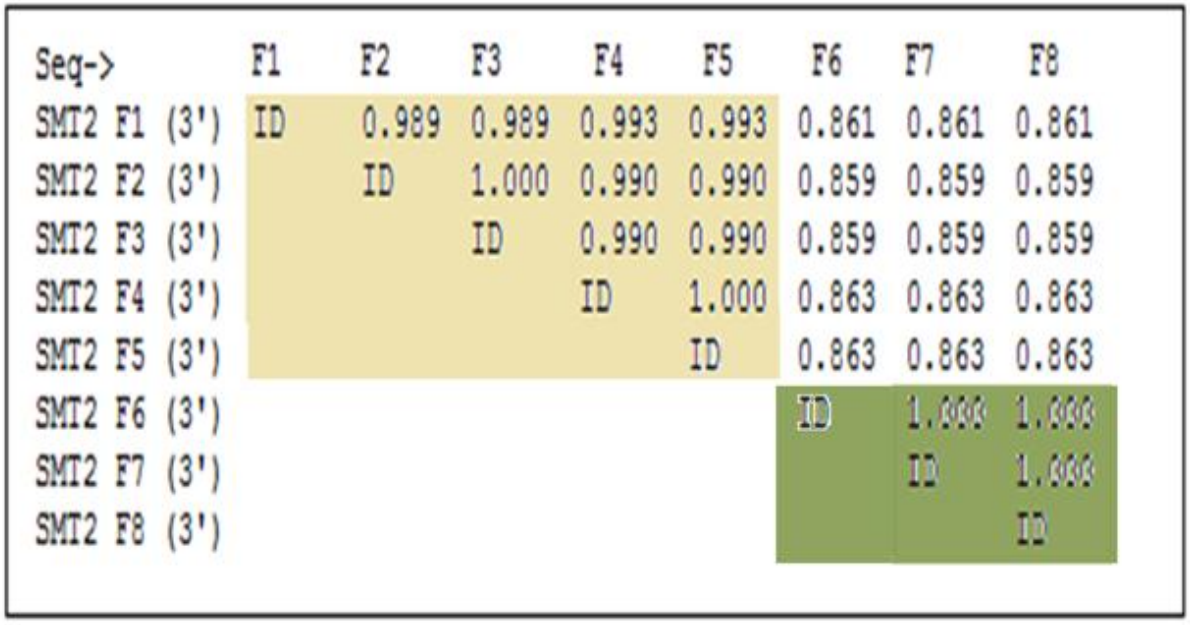


Figure 4.5 5' RACE PCR results for cv. Barnea *SMT2* clones
PCR amplified 5' RACE products:SMT2-1 (5'), SMT2-2 (5'), SMT2-3 (5'), SMT2-4 (5'), SMT2-5 (5'), SMT2-6 (5') and SMT2-7 (5')(Lanes 1-7 respectively) and SMT2-8 (5'), SMT2-9 (5'), SMT2-10 (5') (Lanes 8-10 respectively), M: 100bp molecular weight marker.

4.2.4 Sequencing and analysis of 3' RACE PCR products of *SMT2* genes from Barnea

The sequencing of the eight 3' RACE products revealed two distinct-sized products. Clones F1-F5 contained a 747bp product with 98.9-100% sequence identity between each other (Figures 4.6). While clones F6, F7 and F8 were 718bp (excluding the poly-A tail) with 100% identity between each other (Figures 4.6).



Seq->	F1	F2	F3	F4	F5	F6	F7	F8
SMT2 F1 (3')	ID	0.989	0.989	0.993	0.993	0.861	0.861	0.861
SMT2 F2 (3')		ID	1.000	0.990	0.990	0.859	0.859	0.859
SMT2 F3 (3')			ID	0.990	0.990	0.859	0.859	0.859
SMT2 F4 (3')				ID	1.000	0.863	0.863	0.863
SMT2 F5 (3')					ID	0.863	0.863	0.863
SMT2 F6 (3')						ID	1.000	1.000
SMT2 F7 (3')							ID	1.000
SMT2 F8 (3')								ID

Figure 4.6 Nucleotide sequence similarity of eight *SMT2* 3' RACE clones deduced from olive cultivar Barnea[F1- SMT2-1, F2- SMT2-2, F3-SMT2-3, F4-SMT2-4, F5-SMT2-5, F6-SMT2-6, F7-SMT2-7, F8-SMT2-8, F9-SMT2-9. ID-100% identical, seq-sequences].

4.2.5 Sequencing and analysis of 5' RACE PCR products of *SMT2* genes from Barnea

The sequencing of the nine ~1100bp 5' RACE products revealed a 1105bp product from each. These clones appear to be highly conserved with nucleotide sequence similarity ranging from 98.9- 99.9% (Figure 4.7). The three smaller size cloned RACE products (175bp, 275bp and 650bp)were all 5' truncated sequences that were 100% identical to clone F4 over the corresponding sequences (Figure IIIA in Appendix III).

Seq->	F1	F2	F3	F4	F5	F6	F7	F8	F9
SMT2 F1 (5')	ID	0.994	0.993	0.998	0.996	0.995	0.993	0.994	0.996
SMT2 F2 (5')		ID	0.999	0.992	0.990	0.993	0.990	0.990	0.992
SMT2 F3 (5')			ID	0.991	0.990	0.992	0.989	0.990	0.991
SMT2 F4 (5')				ID	0.998	0.997	0.995	0.996	0.998
SMT2 F5 (5')					ID	0.997	0.995	0.994	0.996
SMT2 F6 (5')						ID	0.994	0.993	0.995
SMT2 F7 (5')							ID	0.993	0.995
SMT2 F8 (5')								ID	0.998
SMT2 F9 (5')									ID

Figure 4.7 Nucleotide sequence similarity of nine*SMT2* 5' RACE clones deduced from olive cultivar Barnea[F1- SMT2-1, F2- SMT2-2, F3-SMT2-3, F4-SMT2-4, F5-SMT2-5, F6-SMT2-6, F7-SMT2-7, F8-SMT2-8, F9-SMT2-9. ID-100% identical, seq-sequences].

4.2.6 Analysis of the overlapping sections of the 5' and 3' RACE PCR products of *SMT2* genes fromcv. Barnea

Comparison of the *SMT2* clones from both 5' and 3' (both sequence classes) RACE sequencing data of Barnea revealed that the 519bp overlapping section of the clones match significantly with sequence similarity ranging from 98.4-100% (Figure IIIC in Appendix III).

4.2.7 Comparison of the deducedBarnea *SMT2* RACEclones with *SMT2* coding sequences from other plants

The multiple sequence alignments of the 5' and 3' RACE *SMT2* clones with previously published SMT2 coding sequences from other plants such as *Riccinus communis* SMT (Accession: XM_002510554.1),*Glycine max* SMT2-1(Accession:FJ483973.1), *Glycine max* SMT2-2(Accession:FJ483974.1), *Gossipum hirsutum* SMT2-1 (Accession:EU308589.1), *Gossipum hirsutum* SMT2-2 (Accession:EU308590.1), *Nicotiana tabacum* SMT2 (Accession: U71108.1); *Arabidopsis thaliana* SMT2-1 (Accession: NM101884.3)

and*Arabidopsis thaliana* SMT2-2 (Accession: U71400.1) revealed a 1031bp partial coding sequence from the 5’ RACE clones and a 570bp partial coding sequence from the 3’ RACE clones. The overlapping of the 5’ and 3’ RACE *SMT2* clones revealed a 1083bp open reading frame encoding a putative SMT2 protein of 360 amino acids (Figure IIIA in Appendix III).

4.3CHARACTERIZATION OFFULL LENGTH *SMT2* CODING SEQUENCES FROM cv. BARNEA

Primers based on sequences in the putative 5' and 3' UTRs of the *SMT2* sequences characterized from Barnea (Table 2.8, Figure IIIB in Appendix III), were used to amplify full-length coding sequences from Barnea cDNA by RT-PCR resulting in the amplification of an expected sized product of ~1200bp (Figure 4.8). Cloning of this PCR product and subsequent *Eco*RI digestion to remove the inserts revealed nine clones containing inserts of the expected size (Figure 4.9).

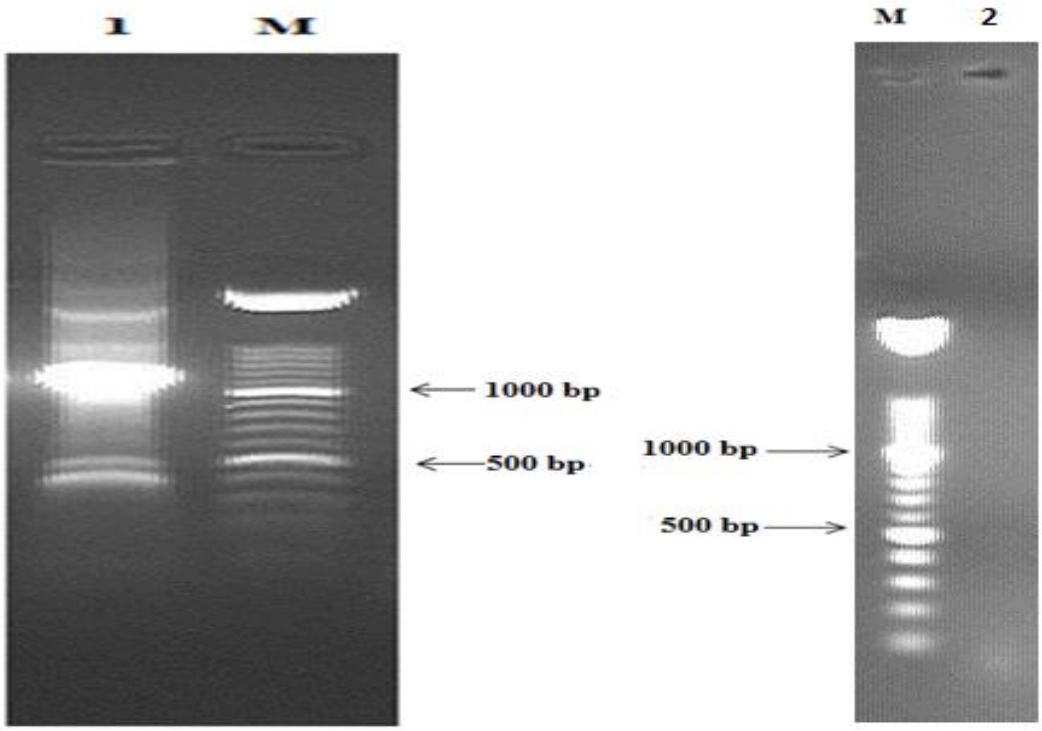


Figure 4.8Amplified product of full length *SMT2* genes from cv. Barnea
Lanes: M: 100bp molecular weight marker; Lane1: Barnea *SMT2*; Lane 2: water only negative control

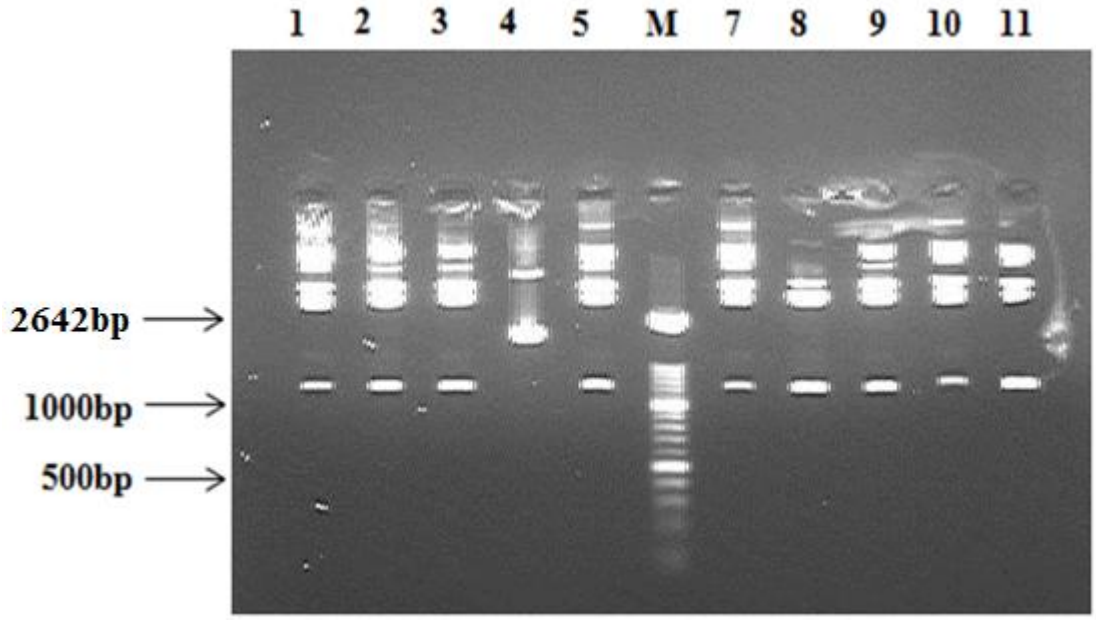


Figure 4.9 Full length coding sequence *SMT2* clones from cv. Barnea
*Eco*RI digested clones:SMT2-1, SMT2-2, SMT2-3, SMT2-4, SMT2-5 (Lanes 1-5 respectively) and SMT2-6, SMT2-7, SMT2-8, SMT2-9 and SMT2-10 (Lanes 6-10 respectively), M: 100bp molecular weight marker.

4.3.1 Sequencing and analysis of full length *SMT2*ORFs from cv. Barnea

Sequencing of the nine full length *SMT2*clones from Barnea revealed all clones contained a 1201bp insert with 99.1-100% sequence identity between each other (Figure 4.10 and Figure IIIB in Appendix III). Clone F1 was selected as representative of all Barnea clones and henceforth referred to as *OeSMT2-1a*. The other *SMT2* clones and the SNPs identified in these are discussed later in the chapter (Section 4.6).

The full length cDNA sequence of *OeSMT2-1a* consisted of1201bp, which included 61bp of 5'UTR, 56bp of 3' UTR and an open reading frame of 1083bp encoding a putative protein of 360 amino acid residues (Accession: KC862262)(Figure 4.11). The calculated molecular mass and pI of *OeSMT2-1a* were 40.1 kDa and 6.28, respectively.

Seq->	F1	F2	F3	F4	F5	F6	F7	F8	F9
Bar-F1-OeSMT2	ID	1.000	1.000	1.000	1.000	0.995	0.994	0.994	0.997
Bar-F2-OeSMT2		ID	1.000	1.000	1.000	0.995	0.994	0.994	0.997
Bar-F3-OeSMT2			ID	1.000	1.000	0.995	0.994	0.994	0.997
Bar-F4-OeSMT2				ID	1.000	0.995	0.994	0.994	0.997
Bar-F5-OeSMT2					ID	0.995	0.994	0.994	0.997
Bar-F6-OeSMT2						ID	0.991	0.991	0.994
Bar-F7-OeSMT2							ID	1.000	0.995
Bar-F8-OeSMT2								ID	0.995
Bar-F9-OeSMT2									ID

Figure 4.10 Nucleotide sequence similarity between Barnea *SMT2*clones
The nine Barnea clones are designated as F1, F2, F3, F4, F5, F6, F7, F8 and F9. Cultivars: Bar: Barnea. ID-100% identical, seq-sequences.

The putative OeSMT2-1a amino acid sequence revealed four signature motifs characteristic of SMT2 proteins. The conserved regions in the protein sequence correspond to the substrate binding segment for sterol [Region I $^{86\text{-}96}$(YEWGWGQSFHF), Region III$^{196\text{-}205}$ (YSIEATCHAP) and Region IV$^{219\text{-}229}$ (KPGSMYVSYEW) and AdoMet[Region II$^{128\text{-}142}$(LDAGCGVGGPMRAI)]. The hydropathy analysis of the putative OeSMT2-1a protein (Section 2.6.7.6) predicted that the encoded polypeptide contained two hydrophobic domains, one within the N-terminal region at amino acid residues 16-33 and the other within the C-terminal region at amino acid residues 312-331. Pfam analysis (Section 2.6.7.6) suggests that OeSMT2belongs to the family of SAM dependent methyltransferases which includes a group of sequence related proteins that catalyse the distinct patterns of 24-alkyl sterols.

Figure 4.11 The cDNA and deduced amino acid sequences of *OeSMT2-1a*(Accession: KC862262)

Underlined letters represent the DNA sequence of PCR primers used for RACE PCR (F1 and R1) and amplification of full length *SMT2* genes (F2 and R2). The four conserved domains of SMT2 proteins (Region I, II, III and IV) are shown in grey boxes.

4.3.2 Homology analysis of the deducedOeSMT2 protein of Barnea with other plant species

To further characterize the homology of OeSMT2-1a protein with SMT2 proteins characterised in other plant species, a comparison of these amino acid sequences were conducted (Figure 4.12 and 4.13). OeSMT2-1a showed high degree of sequence identity with SMT2 sequences isolated from various other plant species (Figure 4.12-A) such as *Arabidopsis thaliana SMT2-1*(81.2%), *Arabidopsis thaliana SMT2-2*(77.2%), *Ricinus communis* (84.2%), *Glycine max* SMT2-1(82.8%),*Glycine max* SMT2-2(82.4%), *Gossipum hirsutum* SMT2-1 (82.8%), *Gossipum hirsutum* SMT2-2 (82.5%) and *Nicotiana tabacum* (85.8%).

OeSMT2-1a showed limited sequence identity to *Arabidopsis thaliana*SMT1 (29.0%), (Figure 4.12-B). The four conserved domains identified in the OeSMT2-1a sequence(as described above) align perfectly with these motifs in other plants SMT2s at positions 86-96, 128-142, 196-205, 219-229 (Figure 4.13) except in domain IV where an aminoacid substitution from leucine (L) to methionine (M) is observed at position 224 in OeSMT2-1a protein, however this is similar to the domain IV of *Arabidopsis* SMT2-1.

	Oe-1a	At2-1	At2-2	Gm-1	Gm-2	Gh-1	Gh-2	Rc	Nt
Oe-1a	ID	0.812	0.772	0.828	0.824	0.828	0.825	0.842	0.858
At2-1		ID	0.825	0.853	0.851	0.856	0.850	0.883	0.825
At2-2			ID	0.801	0.802	0.814	0.798	0.823	0.788
Gm-1				ID	0.958	0.889	0.895	0.900	0.837
Gm-2					ID	0.881	0.879	0.879	0.829
Gh-1						ID	0.911	0.900	0.839
Gh-2							ID	0.900	0.831
Rc								ID	0.850
Nt									ID

A

Seq->	Oe-1a	At1
Oe-1a	ID	0.290
At1		ID

B

Figure 4.12 Comparison ofthe OeSMT2-1aproteins with cloned SMT2 (A) and SMT1 (B) from other plants at amino acid level based on identity percentage

Oe1: OeSMT2-1a(Accession: KC862262),Rc: *Riccinus communis* SMT (Accession: XM_002510554.1),Gm1: *Glycine max* SMT2-1(Accession:FJ483973.1), Gm2: *Glycine max* SMT2-2(Accession:FJ483974.1), Gh1: *Gossipum hirsutum* SMT2-1 (Accession:EU308589.1), Gh2: *Gossipum hirsutum* SMT2-2 (Accession:EU308590.1), Nt: *Nicotiana tabacum* SMT2 (Accession: U71108.1); At2-1: *Arabidopsis thaliana* SMT2-1 (Accession: NM101884.3), At2-2: *Arabidopsis thaliana* SMT2-2 (Accession: U71400.1),At1: *Arabidopsis thaliana* SMT1 (Accession: AF195648.1), ID-100% identical, Seq: Sequences).

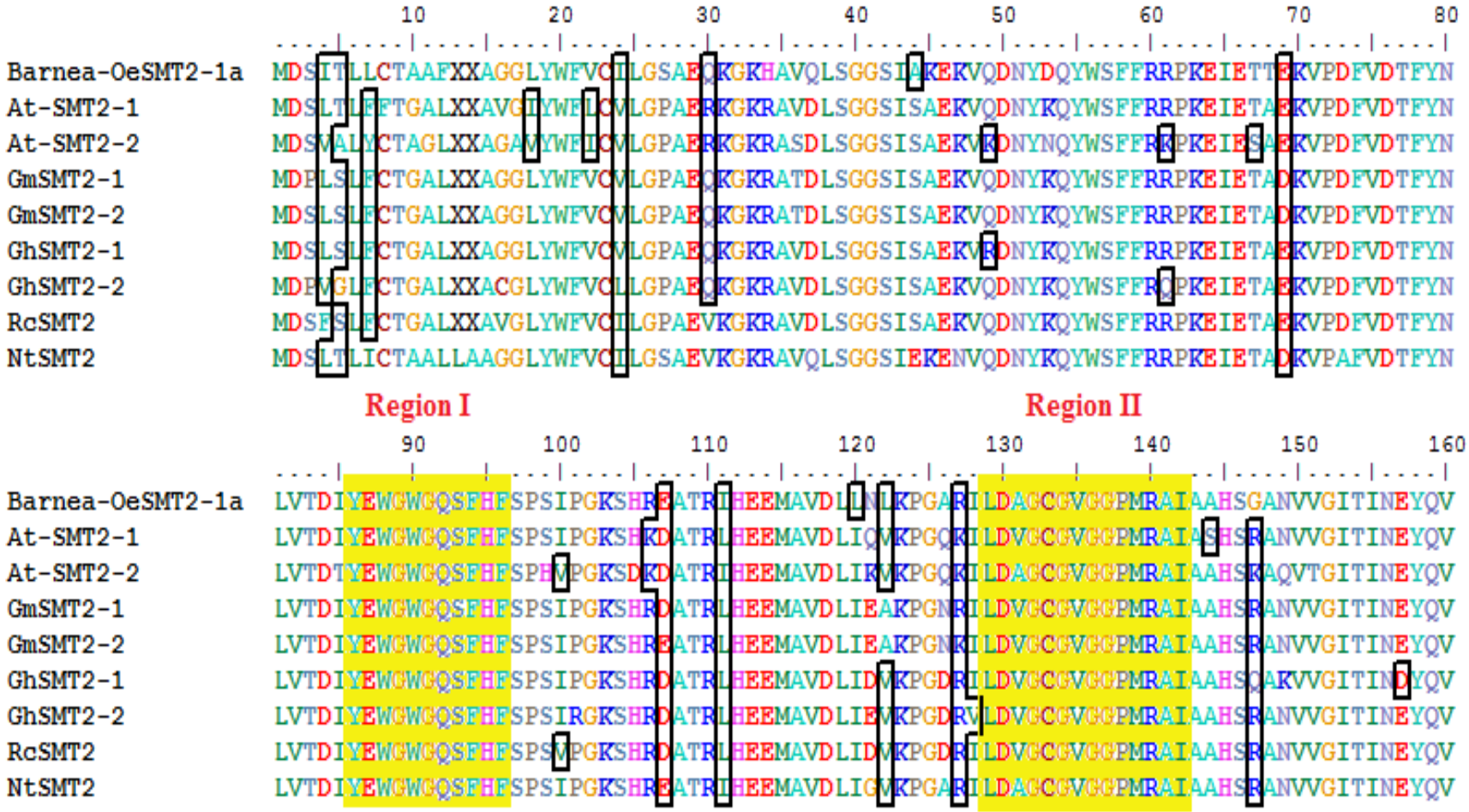


Figure4.13 Allignment of putative protein products of the OeSMT2-1a cDNA isolated from Barnea with other plant SMTs (Continued)

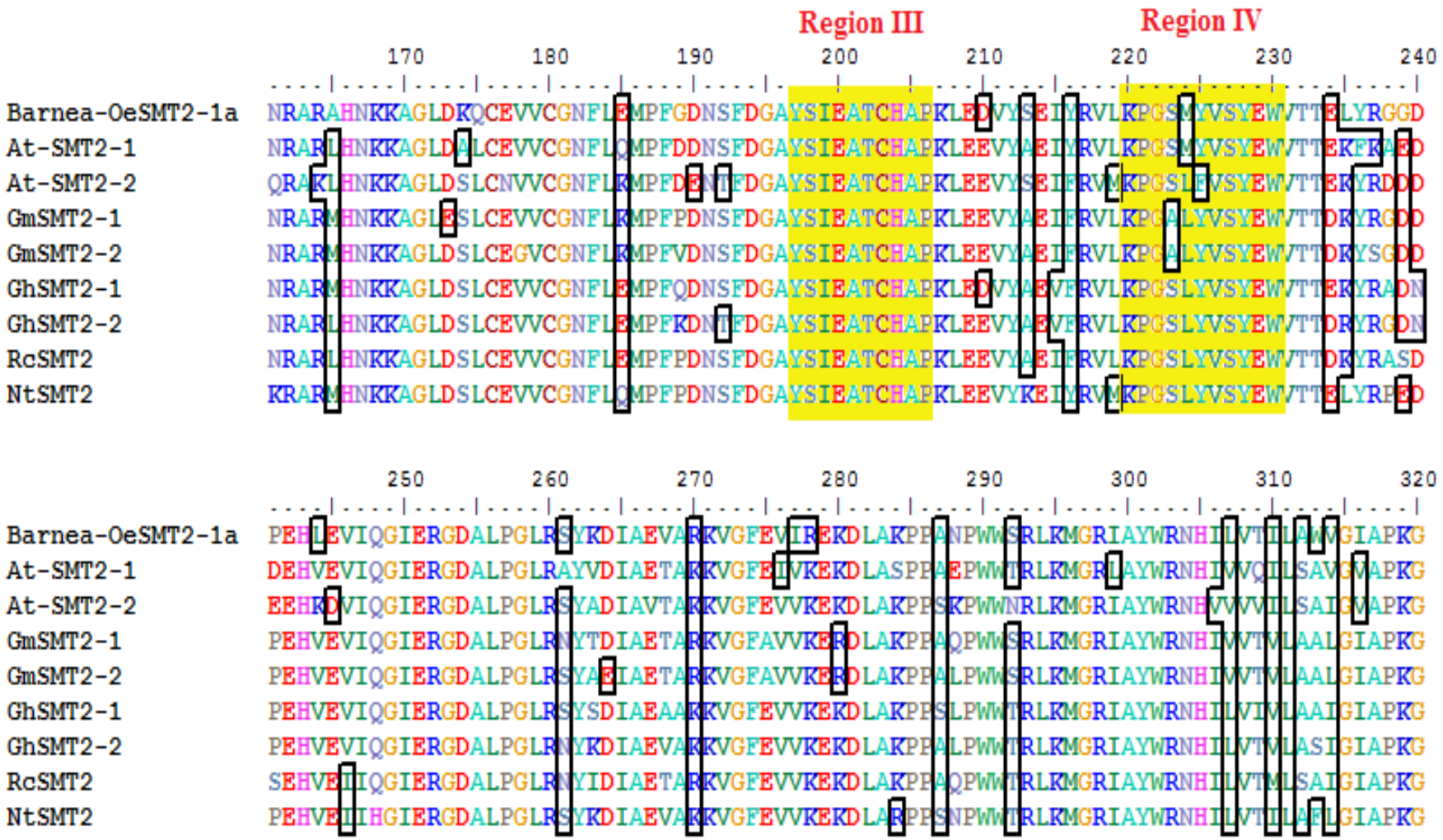


Figure4.13 Allignment of putative protein products of the OeSMT2-1a cDNA isolated from Barnea with other plant SMTs (Continued)

Figure 4.13 Alignment of putative protein products of the OeSMT2-1a cDNA isolated from Barnea with other plant SMTs (Continued) Oe1: OeSMT2-1a(Accession: KC862262),Rc: *Riccinus communis* SMT (Accession: XM_002510554.1), Gm1: *Glycine max* SMT2-1(Accession: FJ483973.1), Gm2: *Glycine max* SMT2-2 (Accession: FJ483974.1), Gh1: *Gossipum hirsutum* SMT2-1 (Accession: EU308589.1), Gh2: *Gossipum hirsutum* SMT2-2 (Accession: EU308590.1), Nt: *Nicotiana tabacum* SMT2 (Accession: U71108.1); At2-1: *Arabidopsis thaliana* SMT2-1 (Accession: NM101884.3), At2-2: *Arabidopsis thaliana* SMT2-2 (Accession: U71400.1). The four conserved domains of SMT2 proteins (Region I, II, III and IV) of all species are highlighted in yellow.

4.4 *SMT2* SOUTHERN BLOTTING OF BARNEA, FRANTOIO AND PICUAL

In order estimate the copy number of the *SMT2* gene family and identify any inter-cultivar RFLPs, a Southern blotting experiment was performed with a 518bp *SMT2* probe amplified from Barnea cDNA(Table 2.13, Figure 4.14) as described in Section 2.9. The probe was also amplified from genomic DNA and sequenced which confirmed their identity and the absence of introns within the probe (data not shown). To get a preliminary idea of any expected differences in copy number between the threeolive cultivars Barnea, Frantoio and Picual, the Southern Blotting experiment was performed using genomic DNA isolated from each cultivar (Section 2.4.1) and digested separately with three different restriction enzymes *Eco*RI, *Hind*III and *Bam*H1 (Section 2.9.1).

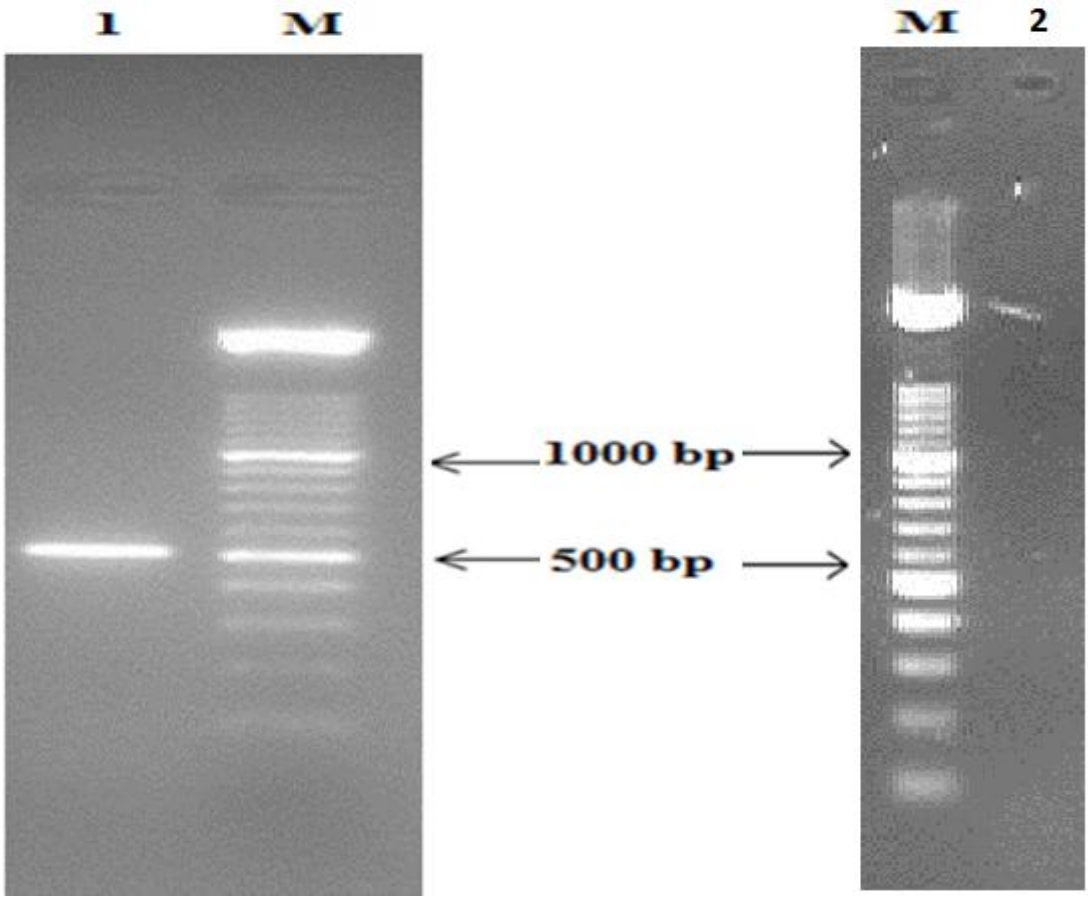


Figure 4.14 Amplification of *SMT2* probe (518bp) for Southern blotting
Lane 1: SMT2 probe; Lane 2: water only negative control; M: 100bp molecular weight marker.

The Southern blotting of cv. Barnea gDNA in the *Eco*RI digest revealed one weak hybridizing fragment approximately 5.5kb and one strong hybridizing fragment approximately7.5 kb (Figure 4.15).The *Eco*RI digest in Frantoio and Picual revealed the same hybridizing fragments of 5.5kb and 7.5kb as cv. Barnea.In all three olive cultivars, the *Hind*III digestproduced a strong hybridizing fragment ranging from 10 to 13kb, however the intensity of the band was weaker in cv. Barnea in comparison to cv. Frantoio and cv.Picual (Figure 4.15).The Southern blotting of Barnea gDNA in the *Bam*HI digest revealed one weak hybridizing fragment approximately 0.7kb and one strong hybridizing fragment approximately 1kb. The *Bam*HI digest in Frantoio and Picual revealed the same hybridizing fragments of 0.7kb and 1kb as Barnea (Figure 4.15).

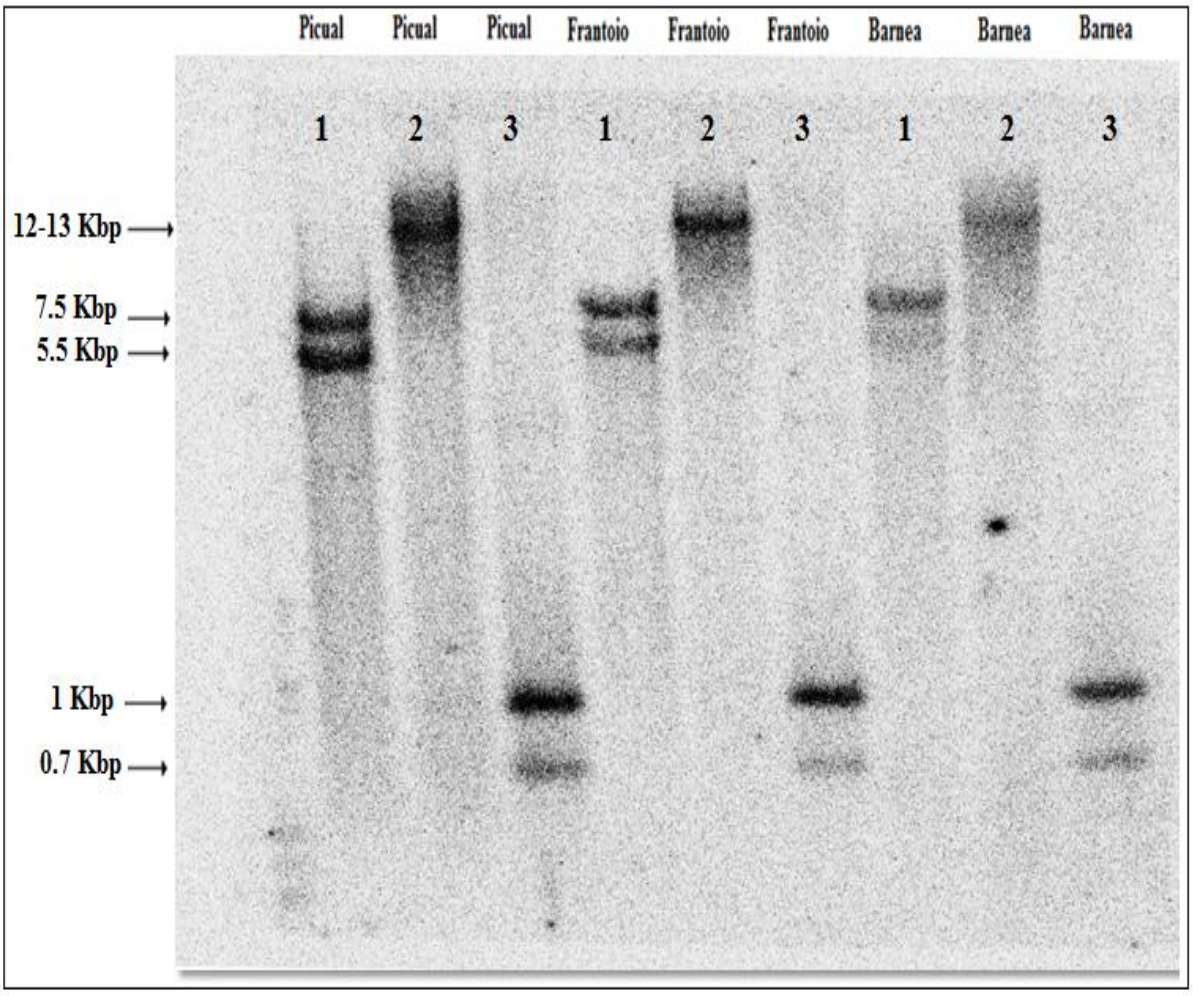


Figure 4.15 Southern blotting of gDNA isolated from Barnea, Frantoio and Picual hybridised with a radiolabeled 518bp *SMT2* cDNAprobe
Genomic DNA was digested with *Eco*RI (Lane 1), *Hind*III (Lane 2) and *Bam*HI (Lane 3) as indicated. Approximate molecular weights of the hybridising fragments (Kbp) are indicated on the left.

4.5CHARACTERIZATION OFFULL LENGTH *SMT2* GENES IN OLIVE CULTIVARS, FRANTOIO AND PICUAL

4.5.1RNA extraction from Frantoio and Picual drupes

Total RNA was extracted from the mesocarp of Frantoio and Picual fruits harvested from the 2009 crop for eight sample timepoints between 96 and 170 DAF revealing all samples had intact high quality RNA (Figure 4.16).

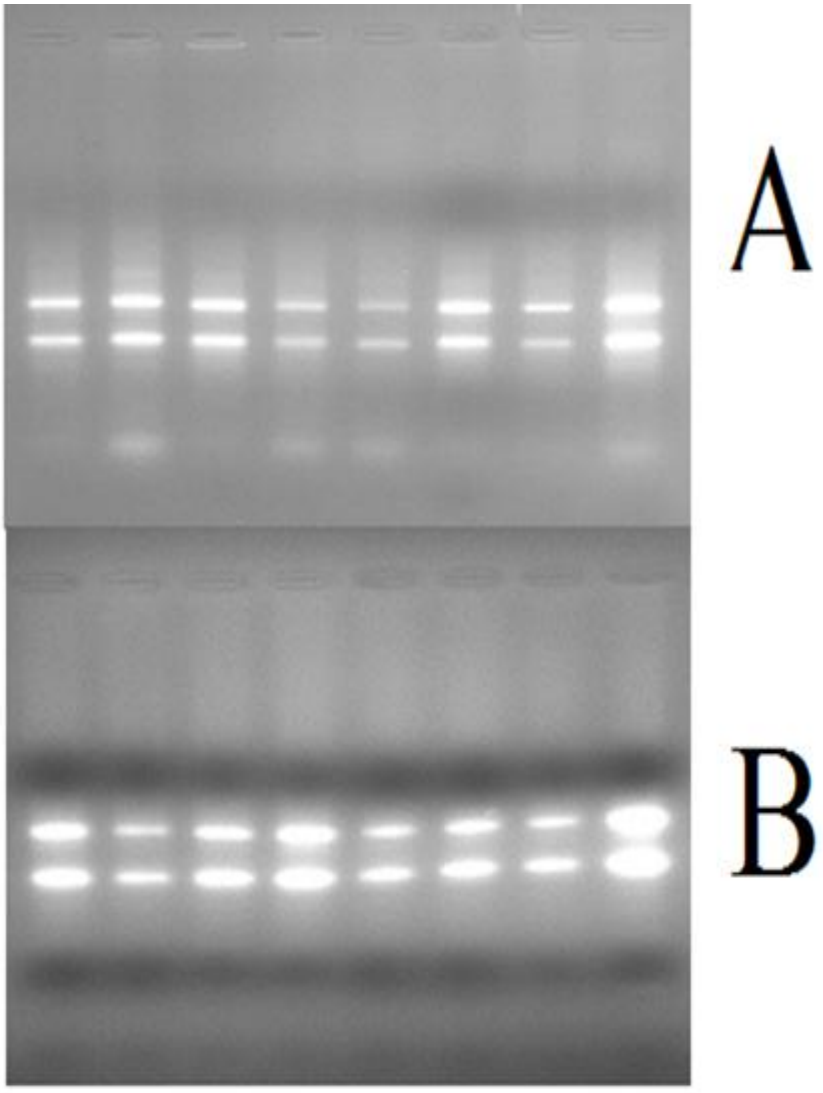


Figure 4.16 Total RNA extracted from eight time point samples of Frantoio (A) and Picual (B) fruits
(Lane1: 96 DAF, Lane2: 109 DAF, Lane3: 116 DAF, Lane4: 122 DAF, Lane5: 136 DAF, Lane6: 150 DAF, Lane7: 167 DAF, Lane8: 170 DAF)

4.5.2 Amplification of full length genes of *SMT2* from Frantoio and Picual

Primers based on sequences in the 5' and 3' UTRs of the *SMT2* sequences characterized from Barnea were used to amplify *SMT2* coding sequences from Frantoio and Picual cDNA resulting in the amplification of the products of the same size as those amplified from Barnea(~1200bp) (Table 2.8, Figure IIIB in Appendix III, Figure 4.17). Cloning of these PCR products and subsequent *Eco*RI digestion to remove the inserts revealed ten and nine clones from Frantoio and Picual, respectively, containing inserts of the expected size (Figure 4.18).

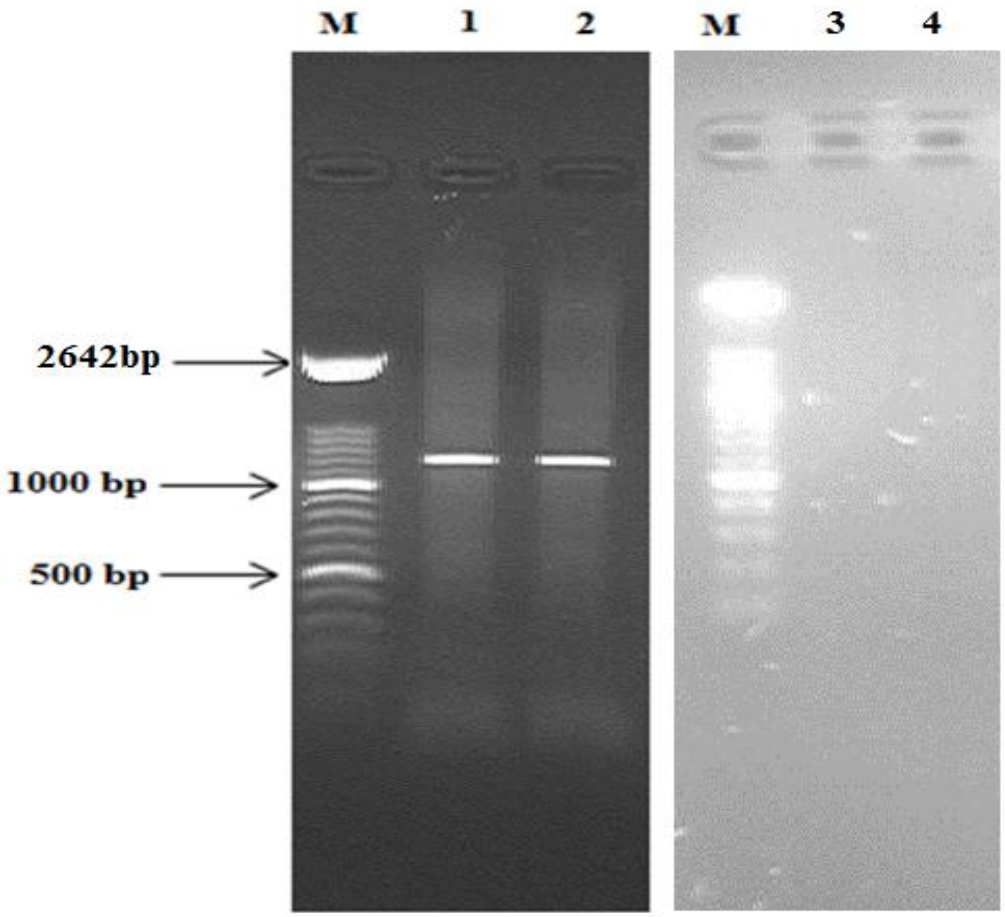


Figure 4.17Amplified product of full length *SMT2* genes from Frantoio and Picual
Lanes: M: 100bp molecular weight marker; Lane1: Frantoio *SMT2*, Lane2: Picual *SMT2*, Lane3: water only negative control-Frantoio *SMT2*, Lane4: water only negative control-Picual *SMT2*

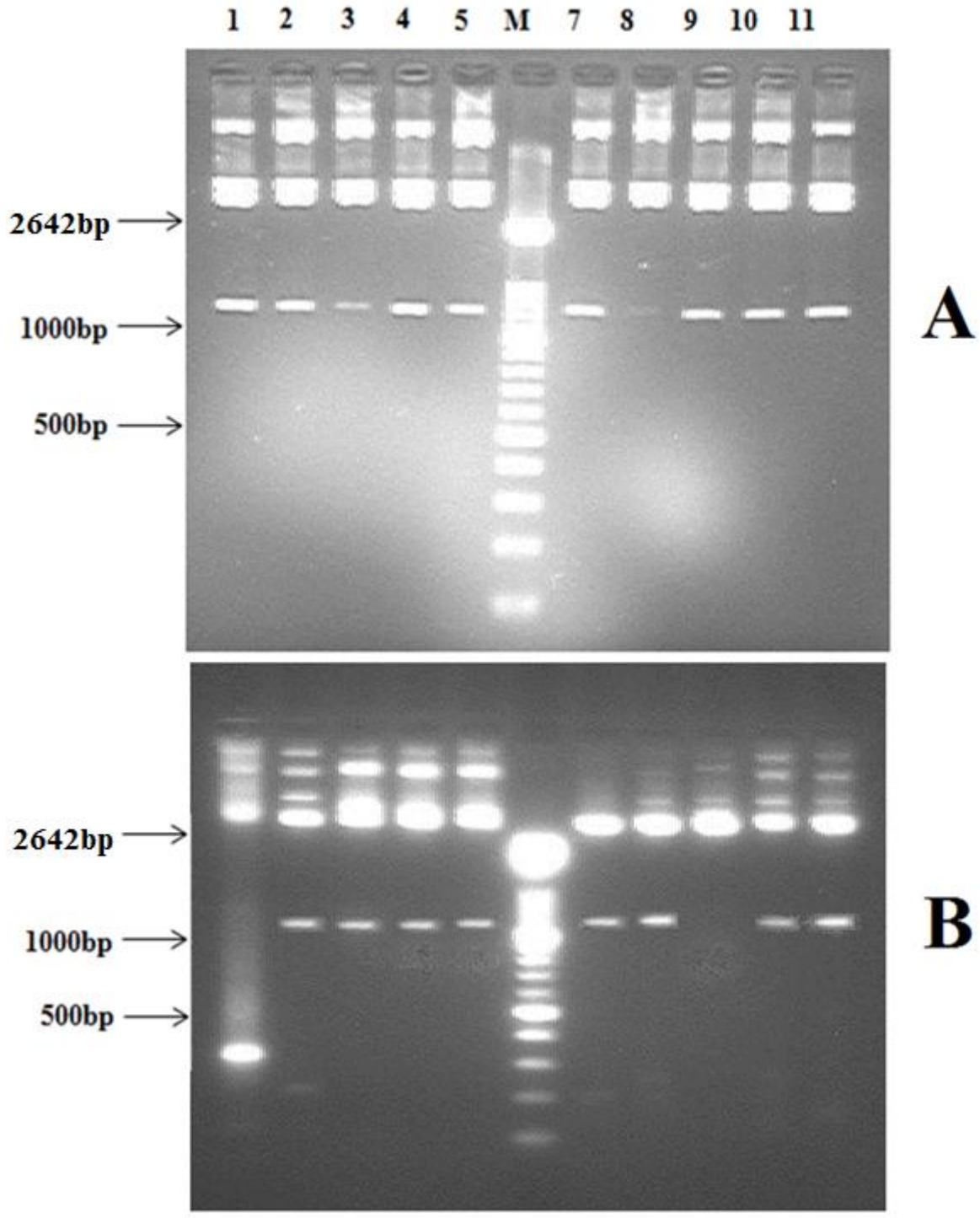


Figure 4.18 Full length *SMT2* clones from Frantoio (A) and Picual (B)
*Eco*RI digested clones:SMT2-1, SMT2-2, SMT2-3, SMT2-4, SMT2-5 (Lanes 1-5 respectively) and SMT2-6, SMT2-7, SMT2-8, SMT2-9 and SMT2-10 (Lanes 6-10 respectively), M: 100bp molecular weight marker.

4.5.3 Sequencing and analysis of full length *SMT2* genes from Frantoio

Ten full length *SMT2* clones were sequenced from the cultivar Frantoio (Figure 4.19 and Figure IIIB in Appendix III).These clones showed 99.1-100% DNA sequence identity to Barnea *OeSMT2-1a* allele.

Seq->	B1	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Bar-OeSMT2-1a	ID	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.995	0.994
Fran-F1-OeSMT2		ID	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.995	0.994
Fran-F2-OeSMT2			ID	1.000	1.000	1.000	1.000	1.000	0.995	0.995	0.994
Fran-F3-OeSMT2				ID	1.000	1.000	1.000	1.000	0.995	0.995	0.994
Fran-F4-OeSMT2					ID	1.000	1.000	1.000	0.995	0.995	0.994
Fran-F5-OeSMT2						ID	1.000	1.000	0.995	0.995	0.994
Fran-F6-OeSMT2							ID	1.000	0.995	0.995	0.994
Fran-F7-OeSMT2								ID	0.995	0.995	0.994
Fran-F8-OeSMT2									ID	1.000	0.991
Fran-F9-OeSMT2										ID	0.991
Fran-F10-OeSMT2											ID

Figure 4.19 Nucleotide sequence similarity between Barnea *OeSMT2-1a* and Frantoio *SMT2* clones
Frantoio clones showing 99.1-100% DNA sequence identity to Barnea *OeSMT2-1a* allele has been highlighted in yellow box. The ten Frantoio clones are designated as F1, F2, F3, F4, F5, F6, F7, F8, F9 and F10. Cultivars: Bar: Barnea; Fran: Frantoio. ID-100% identical, seq-sequences.

4.5.4 Sequencing and analysis of full length *SMT2* genes from Picual

Nine full length *SMT2* clones were sequenced from the cultivar Picual (Figure 4.20 and Figure IIIB in Appendix III). These clones showed 99.1-100% DNA sequence identity to Barnea *OeSMT2-1a* allele.

Seq->	B1	F1	F2	F3	F4	F5	F6	F7	F8	F9
Bar-OeSMT2-1a	ID	1.000	1.000	1.000	0.995	0.995	0.995	0.994	0.994	0.994
Pic-F1-OeSMT2		ID	1.000	1.000	0.995	0.995	0.995	0.994	0.994	0.994
Pic-F2-OeSMT2			ID	1.000	0.995	0.995	0.995	0.994	0.994	0.994
Pic-F3-OeSMT2				ID	0.995	0.995	0.995	0.994	0.994	0.994
Pic-F4-OeSMT2					ID	1.000	1.000	0.991	0.991	0.991
Pic-F5-OeSMT2						ID	1.000	0.991	0.991	0.991
Pic-F6-OeSMT2							ID	0.991	0.991	0.991
Pic-F7-OeSMT2								ID	1.000	1.000
Pic-F8-OeSMT2									ID	1.000
Pic-F9-OeSMT2										ID

Figure 4.20 Nucleotide sequence similarity between Barnea *OeSMT2-1a* and Picual*SMT2* clones
Picual clones showing 99.1-100% DNA sequence identity to Barnea *OeSMT2-1a* allele has been highlighted in yellow box. The nine Picual clones are designated as F1, F2, F3, F4, F5, F6, F7, F8 and F9. Cultivars: Bar: Barnea; Pic: Picual. ID-100% identical, seq-sequences.

4.6 IDENTIFICATION OF SNPS IN *SMT2* ALLELES IN OLIVE CULTIVARS BARNEA, FRANTOIO AND PICUAL

All *SMT2* clones isolated from the three cultivars, Barnea, Frantoio and Picual were aligned together to identify cultivar-specific alleles and/or alleles having key SNPs that are present in all cultivars (Figure 4.21, 4.22 and Figure IIIB in Appendix III).

The nine*SMT2* clones from Barnea revealed 99.0-100% sequence identity between each other (Figure 4.10 and Figure IIIB in Appendix III). Clones F1-F5 showed 100% identity between each other and as mentioned before in Section 4.3.1, clone F1 was selected as representative of all Barnea clones and referred to as *OeSMT2-1a*. Frantoio clones F1- F7 and Picual clones F1-F3 showed 100% sequence identity to the Barnea *OeSMT2-1a* allele.

Barnea clone F6 was 99.5% similar to *OeSMO2-1a* however unique SNPs were observed at a fewpositions. This clone had a C→T nucleotide substitution at position 97 that would lead to an amino acid substitution from histidine (H) to tyrosine (Y) at amino acid 33. A nonsense

mutation (A→T) at position 835 led to an introduction of a premature stop codon at amino acid 279 replacing amino acid lysine (K). This clone is henceforth referred to as *OeSMT2-1b*.Frantoio clones (F8 and F9) and Picual clones (F4, F5 and F6) showed the same characteristic SNPs as observed in Barnea *OeSMT2-1b* allele.

Barnea clones F7 and F8 were 100% identical to each other andwere 99.4% similar to *OeSMO2-1a.* These clones had two silent mutations at positions 372 (C→T) and 402 (A→G). At position 638, an A→G nucleotide substitution led to the substitution from glutamic acid (E) to non-polarglycine (G) at amino acid 213. The nucleotide substitution (T→C) which was seen at position 865 in *OeSMT2-1b* allele leading to an amino acid substitution from tryptophan (W) to arginine (R) was also observed in these clones. Another nucleotide substitution (A→G) was seen at position 998 which led to an amino acid substitution from tyrosine (Y) to cysteine (C) at amino acid 333. Clone F7 was chosen as the representative clone henceforth referred to as *OeSMT2-1c.* A Frantoio clone (F10) and Picual clones (F7, F8 and F9) showed the same characteristic SNPs as observed in Barnea *OeSMT2-1c* allele.

Barnea clone F8was 99.4% similar to *OeSMO2-1a* however unique SNPs were observed at several places. This clone had two silent mutations at positions 372 (C→T) and 402 (A→G) as in *OeSMT2-1c* allele however did not show the SNPs at positions 638, 865 and 998 as observed in this allele. In addition this clone had the T→A nucleotide substitution at position 1075 leading to the substitution of serine (S) to threonine (T) as observed in *OeSMT2-1b* allele. This clone appeared to be unique and is henceforth referred to as *OeSMT2-1d.*This allele was not identified in the characterised clones isolated from Frantoio and Picual.

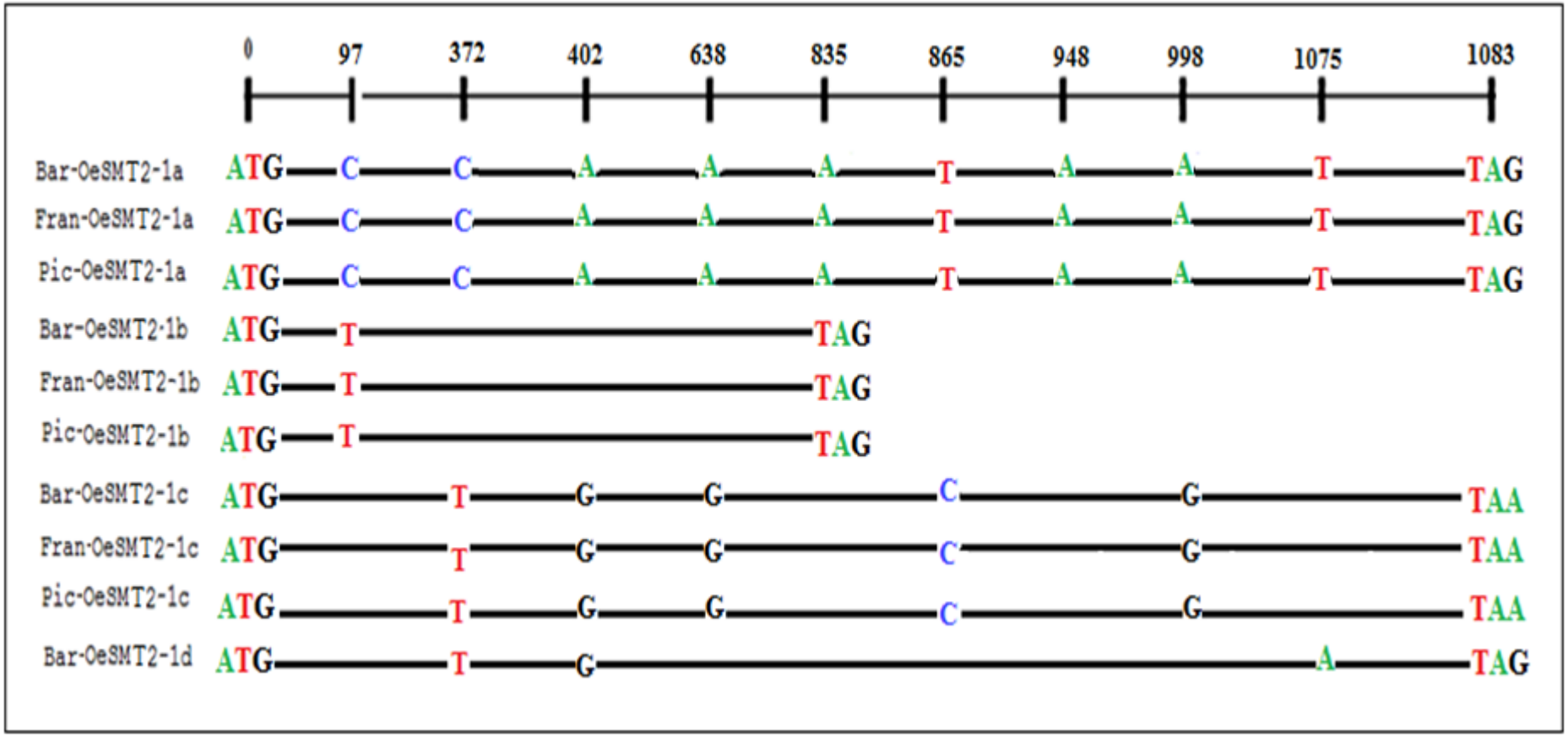


Figure 4.21 Line diagram of Barnea, Frantoio and PicualOe*SMT2-1* alleles highlighting identified SNPs with their respective positions in the coding sequence

Barnea Oe*SMT2-1a* has been used as a reference sequence to align all alleles. Continuous lines indicate 100% identical to reference sequence. Start codon ATG and STOP codon TAG/TAA has also been shown. Cultivars: Bar: Barnea; Fran: Frantoio, Pic: Picual

.

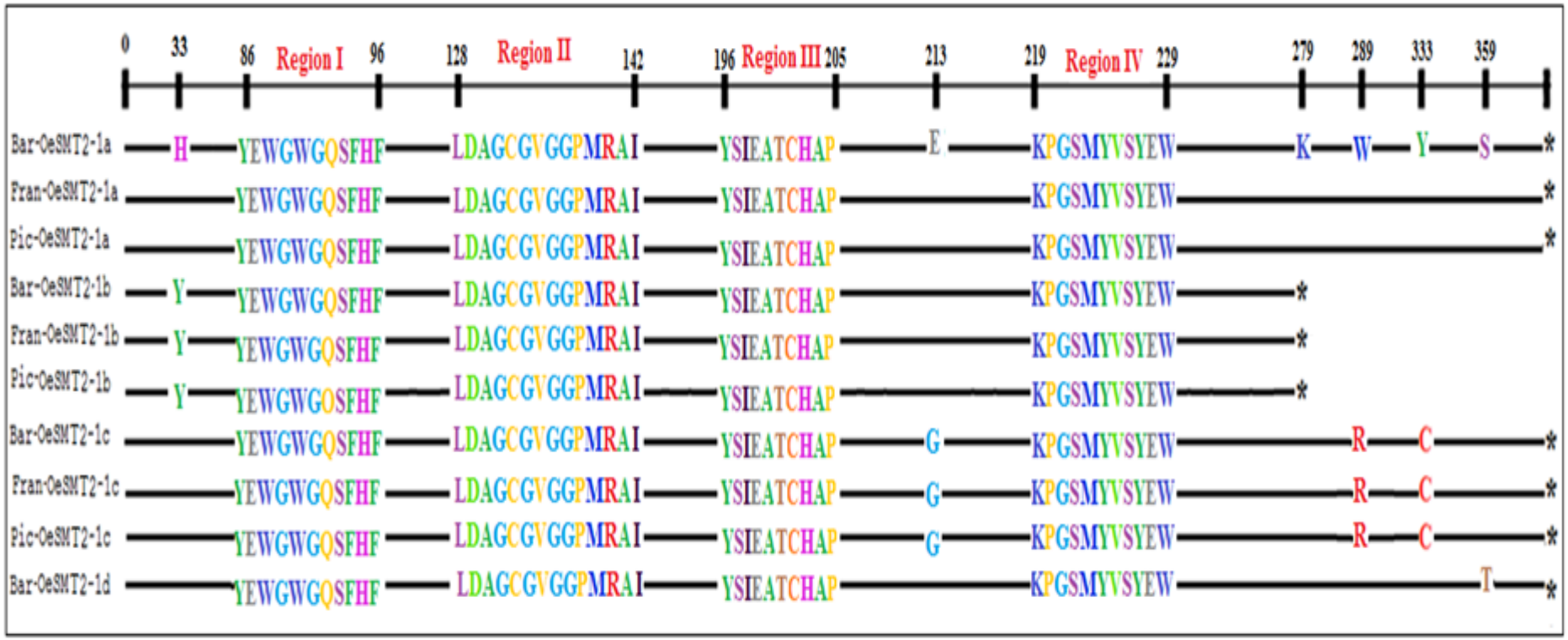


Figure 4.22 Line diagram of Barnea, Frantoio and PicualOeSMT2-1proteins

SNPs that lead to change in their amino acid sequence are shown at their respective amino acid position.Barnea Oe*SMT2-1a* has been used as a reference sequence to align all alleles. Continuous lines indicate 100% identical to reference sequence. The fourconserved domainscommonly observed in SMT2 proteins are also shown. Cultivars: Bar: Barnea; Fran: Frantoio, Pic: Picual; *: STOP codon.

4.7 TESTING FOR THE PRESENCE OF THE *OeSMT2-1d* ALLELE IN BARNEA, FRANTOIO AND PICUAL

The sequencing of the full length *SMT2* genes from Barnea, Frantoio and Picual revealed one cultivar specific allele, *OeSMO2-1d*, which was found only in Barnea, but not in Frantoio and Picual. To test for the presence of this allele in all cultivars, allele specific PCR was conducted (Figure 4.23) (Section 2.10) with a primer pair based on SNPs of *OeSMO2-1d* cDNA sequences, revealing amplification of the predicted 303bp product when an optimized annealing temperature of 55˚C was used in all cultivars (Figure 4.23).

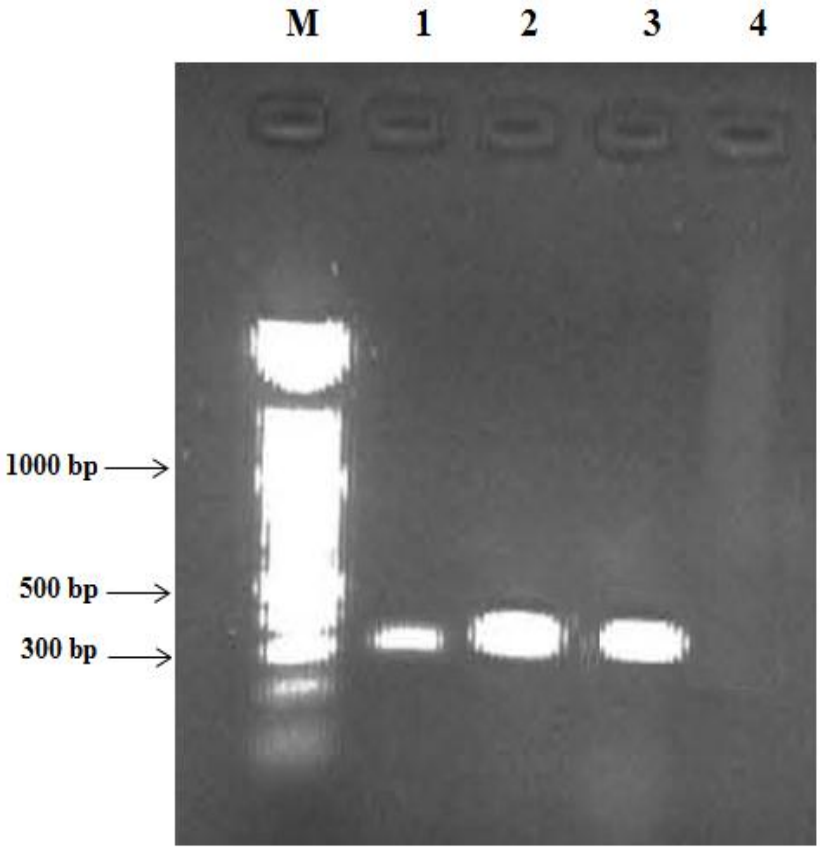


Figure 4.23 Allele-specific PCR productsof *OeSMT2-1d* amplified from Barnea (Lane 1), Frantoio (Lane 2) and Picual (Lane: 3); Lane 4: water only negative control. M: molecular weight marker (100bp).

4.8 DISCUSSION

In higher plants, S-adenosylmethionine (Adomet)-sterol-C-methyltransferases (SMTs) are involved in two distinct transmethylation reactions where cycloartenol is the substrate of the first methylation reaction, resulting in 24-methylene-cycloartanol catalysed by SMT1 enzyme, and 24-methylene-lophenol is the preferred substrate for the second methylation, yielding 24-ethylidene-lophenol catalysed by SMT2 enzyme, ultimately resulting in the production of β-sitosterol and stigmasterol separate from the campesterol biosynthetic pathway.

As described in Section 1.8.2.4 there is evidence that overexpression of *SMT2* results in increase of β-sitosterol levels due to elevated 24-methylenelophenol-C24-methyltransferase activity at the expense of 24 methyl sterols (campesterol) as observed in various plant species such as *Arabidopsis*, soybean and tobacco.However, no information is available about the *SMT2* genes in olives. A BLASTn screen of the 454 sequencing data generated from developing olives(Alagna *et al.* 2009) revealed a 365bp sequence for the *SMT2* gene, however the structure and organization of the *SMT2* gene family and any allelic variations between the olive cultivars, Barnea, Frantoio and Picual that may impact the total and relative sterol content in these olive oils have not been studied; some of these shortcomings have been addressed in the present chapter.

Based on this partial 365bp olive *SMT2*sequence, *SMT2*cDNAs were isolated from the olive cultivar Barnea using a RACE-PCR based approach. To study the structure of *SMT2* genes in olives, the full length *SMT2* genes were amplified from Barnea using primers based on sequences in the 5' and 3' UTRs of the *SMT2* sequences obtained from the RACE sequencing data. Further, to get an estimate of the expected copy number of the *SMT2* gene family in olives, genomic DNA was isolated from the leaves of olive cultivarsBarnea, Frantoio and Picualandwere digested with three different restriction enzymes *Eco*RI, *Hind*III and *Bam*H1 and Southern blotting experiment was conducted with a 518bp *SMT2* probe. Finally the full length *SMT2* genes were amplified from the other two olive cultivars commonly cultivated in Australia, Frantoio and Picual to identify any allelic differences between the cultivars.

4.8.1Olive SMT2 is encoded bytwo gene families

The 5’ RACE sequencing data of Barnea showed that all clones had very high sequence conservation with 98.9-99.9% sequence identity between each other. The analysis of 3’RACE products suggested that alternative usage of polyadenylation sites generates two *SMT2*transcripts in Barnea, where one sequence class contained 747bp with a poly-A tail while the second sequence class contained a poly-A tail 29bp upstream from the first sequence class; however the sequences are 100% identical besides this.The overlapping sections of the 5’ and 3’ RACE PCR products of the *SMT2*clones and subsequent sequencing of full length *SMT2*clones from Barnea revealed high sequence conservation (>98.9%) between the clones. Clone F1 was selected as a reference sequence for Barnea and was named as *OeSMT2-1a*(Accession: KC862262). Further analysis of the Barnea clones revealed the presence of three additional alleles(referred to as *OeSMT2-1b, OeSMT2-1c* and *OeSMT2-1d*)based on the identification of key SNPs within the coding sequences.

The Southern blotting data suggested that there are 2 loci within the Barnea genome that encode SMT2.Though a single band was observed in the *Hind*IIIdigest, two *SMT2*hybridizing fragments were detected in the*Bam*HIand *Eco*RI digest.Since within the target DNA sequence there is asingle *Bam*HI restriction site but not a*Hind*III or*Eco*RI site (data not shown), one could suggest the presence of a single *SMT2*gene in *O. europaea.* However, the presence of two strong hybridizing bands observed in the*Eco*RI digest suggests the presence of two *SMT2* genes in *O. europaea*.

Thus, the sequencing results and Southern analysis data, taken together suggested the presence of two *SMT2* gene families in olives, where the presence of the four *SMT2* sequences may represent four alleles of two *SMT2* genes, a conclusion that is supported by the Southern blotting data which clearly shows the presence of two hybridising fragments. This would be similar to the *SMT2* gene families of *A. thaliana*(Bouvier-Navé *et al.* 1997), *N. tabacum*(Schaeffer *et al.* 2000),*G. max*(Neelakandan *et al.* 2009) and *G. hirsutum*(Luo *et al.* 2008), which have all been found to contain two copies of the *SMT2* gene. However it is important to note that the two isolated *SMT2* gene families share 78.38%, 89.0% and 86.2% sequence similarity between each other at nucleotide level in *Arabidopsis*, soybean and cotton respectively, however the Barnea *SMT2* alleles share >99% similarity at nucleotide

level. This may indicate either *SMT2* alleles in olives are highly conserved unlike other plant species, or an additional *SMT2* allele may be present in olives which might have escaped detection in our screeningprocedures due to more divergentDNA sequence composition and thus this possibility should not be excluded and would require further investigation. The presence of two strong hybridizing bands observed in theSouthern data also supports the presence of two *SMT2* genes in *O. europaea*.

The full length cDNA sequence of *OeSMT2-1a*had an open reading frame of 1083bp and encoded a protein of 360amino acids with a predicted molecular mass of 40.1kDa. This is similar to the two *SMT2* cDNAs cloned and isolated from cottonboth containing an ORF of 1086bp and encoding polypeptides of 361 amino acids (Luo *et al.* 2008). However, *OeSMT2-1a* shares higher sequence similarity with the *SMT2*-1 cDNAs cloned and isolated from *Arabidopsis*(Bouvier-Navé *et al.* 1997) andsoybean (Neelakandan *et al.* 2009) whereboth of them have an ORF of 1086bp encoding polypeptides of 361 amino acids. While the *SMT2*-2 cDNAs cloned and isolated from *Arabidopsis*(Bouvier-Navé *et al.* 1997) andsoybean (Neelakandan *et al.* 2009)contained an ORF of 1080bp and 1092bp encoding polypeptides of 351 and 363 amino acids, respectively.Furthermore, the typical conserved structures characterized by the sterol C-24-methyltransferase (Section 1.9.2.3), such as region I(YEWGWGQSFHF), region II (LDAGCGVGGPMRAI), and region III (YSIEATCHAP) and region IV (KPGSMYVSYEW) were present in olive SMT2-1a deduced protein. The sequence alignment of OeSMT2-1a and other SMT2 homologuesfrom few plant species revealed that the spacing of these signature motifs in OeSMT2-1a align with the corresponding SMT2 sequence of other plant SMT2s (Figure 4.13).*Arabidopsis*SMT2 possess a hydrophobic domain at the N-terminal position which has approximately 25 amino acids, while SMT1 are devoid of such a hydrophobic domain (Bouvier-Navé *et al.* 1997). The hydropathy analysis of the putative OeSMT2-1a proteinpredicted that the encoded polypeptide contained two hydrophobic domains, one within the N-terminal region at amino acid residues 16-33 and the other within the C-terminal region at amino acid residues 312-331.

Pfam analysis suggested that *OeSMT2*belongs to the family of SAM dependent methyltransferases which includes a group of sequence related proteins that catalyse the distinct patterns of 24-alkyl sterols.

BLAST searches revealed that Oe*SMT2* sequences are similar in the ORF size, in amino acid sequence and in the presence of key motifswith *SMT2* sequences isolated from various other plant species such as *Arabidopsis*, cotton, tobacco, castor oil and soybean (77.2-85.8%), whereas the same can't be said for SMT1. Thus, it is clear that these clones encode SMT2 enzymes.

As mentioned before, SMTs are involved in two distinct transmethylation reactions, where the first and second methylation reactions are catalysed by SMT1 and SMT2 respectively (Section 1.9.2.3) (Benveniste 2004). As our study focuses on isolating genes encoding the SMT2 enzyme in olives, the OeSMT2-1a protein was aligned with *Arabidopsis thaliana*SMT1 which revealed limited sequence identity (29.0%) between them strongly suggesting that these genes encode SMT2 enzymes, and not SMT1 (Figure 4.12-B).

4.8.2No *SMT2* allelic differences appear to be present between Barnea, Frantoio and Picual

This chapter focused on characterising the *SMT2* gene family in olives and a comparison of the families in different olive cultivars. To this end, full length *SMT2* cDNAs were isolated from two additional olive cultivars, Frantoio and Picual that have contrasting phytosterol profiles(Mailer *et al.* 2008), and their sequences were compared against the Barnea *SMT2* clones to identify any allelic differences between the olive cultivars.

Comparison of the *SMT2* clones between olive cultivars Barnea, Frantoio and Picual revealed high sequence conservation between the cultivars (99.4-100%) (Figure IIIB in Appendix). Apart from theOe*SMT2-1a*alleleidentified in Barnea, Frantoio and Picual, SNPs were identified at several places along the nucleotide sequences of the Barnea, Frantoio and Picual clones suggesting that there may be additional *OeSMT2*alleles in these cultivars.

Based on their key SNPs, three *OeSMT2*-1 alleles namely *OeSMT2-1a,OeSMT2-1b* and *OeSMT2-1c* were identified from the sequencing data of all three olive cultivars indicating these alleles are most likely present in all the cultivars. In particular, *OeSMT2-1b* alleleswhich encoded a smaller SMT2 protein were identified in Barnea, Frantoio and Picualstrongly suggesting that this is a true allele and not the result of a mutation introduced

through the experimental methodology in the amplification, cloning and sequencing of this allele.

Only one allele (*OeSMT2-1d*) identified from the sequencing data of Barnea (Figure IIIB and IIIB in Appendix III), was absent in the sequencing data of Frantoio and Picual. As the screen of the *SMT2* PCR products was not exhaustive it was possible that this allele had been missed in the sequencing of clones from Frantoio and Picual. To test for the presence of this allele in these other olive cultivars, allele specific PCR was conducted under stringent conditions using primers that was specific for *OeSMT2-1d.*

AS-PCR results suggested that *OeSMO2-1d* is likely to be present in Frantoio and Picual. Sequencing of the AS-PCR products from the three cultivars were not conducted as the four SMT2 sequences identified have 100% sequence identify in the amplified region that would confirm their presence, with the only difference at the 3' end of the AS-primers. Further confirmation of this allele in Frantoio and Picual would require further screening of clones in future studies.

Thus, results from the allele specific PCR suggested that no *SMT2* allelic differences seem to be present between the olive cultivars. This was further supported by the Southern blotting results which failed to identify any distinct RFLPs between the olive cultivars.

In conclusion, the results from the isolation of full length *SMT2* genes from olive cultivars Barnea, Frantoio and Picual and the subsequent Southern blotting data revealed that there are 2 *SMT2* heterozygous loci with the same alleles in all 3 cultivars. Therefore it would appear that allelic differences in the *SMT2* gene family are unlikely to be responsible for the contrasting phytosterol profiles observed in the olive oils derived from these cultivars(Section 1.7 and 3.2.1).

CHAPTER 5

MOLECULAR CHARACTERISATION OF THE*STEROL METHYL OXIDASE-2* GENE FAMILYOF OLIVES

List of figures and tables in Chapter 5

ABSTRACT

The olive cultivarBarnea which represents 41% of the olive crop grown in Australia, contain up to 4.8% campesterol, which exceeds the IOOC standards for extra virgin olive oil that stipulate a campesterol level of less than 4%.It has been observed that although relative sterol percentages in Australian olive oils fluctuate seasonally, general trends in sterol profiles appear to remain consistent within individual cultivars, strongly implicating genetic factors as the cause of these different levels. The characterized sterol biosynthetic pathway in plants contains a bifurcation that leads to the formation of β-sitosterol or campesterol, with the flux controlled by the activity of two branch-point enzymes, SAM-24-methylene-lophenol-C-24-methyltransferase2 (SMT2) and C-4α-sterol-methy-oxidase2 (SMO2). Experimental evidence has demonstrated that SMO2 can influence campesterol and β-sitosterol levels in tobacco. Thus, it is conceivable that the relative activity or expression of these enzymes could play a pivotal role in determining the relative amounts of β-sitosterol and campesterol in Australian olive oils. In this chapter, characterization of full length *SMO2* cDNAs were conducted from olives using 5'-3' RACE approach to identify any allelic differences between the olive cultivars Barnea, Frantoio and Picual.

The RACE data of the *SMO2* cDNA generated a 885bp5' RACE product and two different size 3' RACE products (471bp and 459bp).This RACE data was used to design primers flanking the putative coding sequence and led to the amplification of a ~950bp product from RNA isolated from the olive cultivarsBarnea, Frantoio and Picual.Sequencing and analysis of the full length cDNA revealed the presence of atleast two distinct gene families namely *OeSMO2-1a* and *OeSMO2-2a* with an ORF of 822bp and 783bp respectively, and encode proteins of 274 and 261 residues.Southern blotting approach was adopted to investigate the copy number of *SMO2* genes and the results suggest that *SMO2* gene is encoded by a small multigene family in the olive genome.AS-PCR results suggestedthe presence of additional *SMO2* alleles in olives which was further supported by the Southern blotting results which identified distinct RFLPs between the olive cultivars.

5.1 INTRODUCTION

As discussed in Section 1.9.1, investigation into sterol-methyl transferase 2 (*SMT2*) and C4α-sterol-methyl oxidase 2 (*SMO2*) gene families are of interest due to their potential role in determining the ratio of β-sitosterol to campesterol in plants.

In Chapter 4, the investigation of the *SMT2* gene family in olives have been described which suggested that there is little allelic variation in that gene family. Therefore it is plausible that the increased levels of campesterol relative to the other phytosterols in the olive oil from Barnea is due to differences in the activity orincreased expression of the *SMO2* gene(s)in this cultivar, so characterization of the *SMO2* gene family may prove interesting.

The*SMO2* gene family has been characterised in a number of plants species including,*Arabidopsis thaliana, Solanum lycopersicum*(tomato),*Ricinus communis*(castor oil),*Nicotiana benthamiana*(tobacco),*Glycine max*(soybean),*Vitis vinifera*(grapevine),*Gossypium arboretum* (cotton) (Section 1.9.3.3). Briefly, past research on the *SMO2* gene family has resulted in the identification of two isoforms of SMO2 in *Arabidopsis thaliana,* namely *AtSMO2-1* and *AtSMO2-2* encoding predicted polypeptides of 267 and 261 amino acids, respectively(Darnet *et al.* 2004). Closer inspection of these SMO2 amino acid sequences revealed three characteristic histidine rich motifs (HX3H, HX2HH and HX2HH). The function of these tripartite motifs may be to provide the ligands for a presumed catalytic di-iron center as proposed previously for other enzymes possessing similar motifs (Darnet *et al.* 2004).

However to date no information is available on the full length *SMO2* genes in olives. Therefore a better understanding of this gene family sequence in olives was required to further study their role in determining the relative amounts of β-sitosterol and campesterol in Australian olive oils.A partial sequence of theputative *SMO2*gene was identified from the Alagna *et al*. data (2009) which was used in this study as a starting point in the isolation and characterization of the full length coding sequences of the *SMO2* genes from three different olive cultivars, Barnea, Frantoio and Picual and an analysis of the putative allelic differences between these olive cultivars that could conceivably play a role in determining their individual phytosterol profiles.

5.2 CHARACTERIZATION OF5’ AND 3’ RACE PRODUCTS OF *SMO2* GENES IN cv. BARNEA

5.2.1 Screening and identification of partial *SMO2* sequences

A BLASTn screen of the 454 sequencing data generated from developing olives with the *SMO2* sequences of *Arabidopsis* revealed a 238bp sequence(Cluster Id: OLEEUCl011741)(Alagna *et al.* 2009)with 79.9% identity to *SMO2-2* from *A. thaliana*(Accession number:AT2G29390.2). Subsequent translation revealed that this partial sequence encoded a 77 amino acid sequence that was 94.8% identical to the homologous section of the *A. thaliana* SMO2 enzyme. This partial sequence was used for the amplification and cloning of the uncharacterized 5’ and 3’ regions of the *SMO2* genes in the olive cultivar Barnea.

5.2.2RNA extraction from Barnea fruits

As described in Chapter 4 (Section 4.2.2), total RNA was extracted from the mesocarp of Barnea fruits harvested in the 2009 season for eight sample timepoints between 96 and 170 DAF revealing all samples had intact high quality RNA (Figure 4.1).

5.2.3 RACE PCR products of *SMO2* genes from cv. Barnea

3’ RACE PCR was conducted using the 3’ GeneRacer primer and the forward GSP (Section 2.6.3 and Table 2.4) which resulted in the amplification of a ~500bp fragment (Figure 5.1).

5’ RACE PCR was conducted using the 5’ GeneRacer primer and the reverse GSP (Section 2.6.3 and Table 2.4) resulting in the amplification of a smear with a prominent ~950bp product (Figure 5.2).

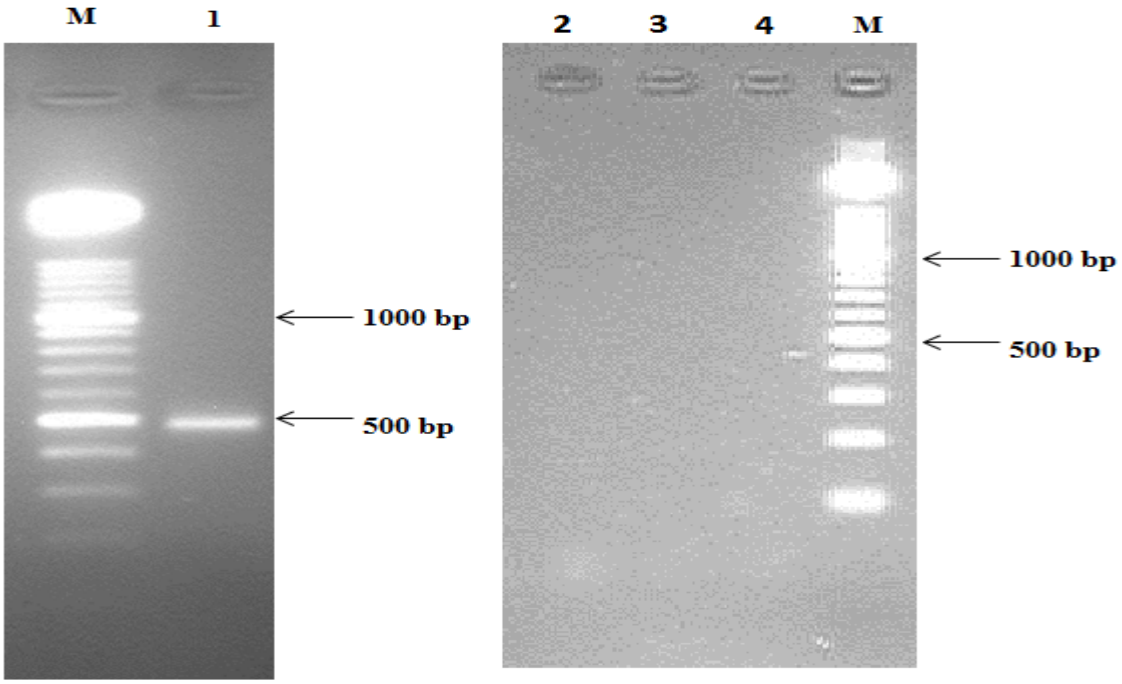


Figure 5.1 Agarose gel electrophoresis of 3' RACE PCR product of *SMO2* gene amplified from Barnea RNA

Lanes: M: molecular weight marker; Lane1: Barnea *SMO2*3' RACE; Lane2: water-only negative control; Lane 3:negative control (Gene Racer primer and template and no GSP);Lane 4:negative control (GSP and template and no gene Racer primer)

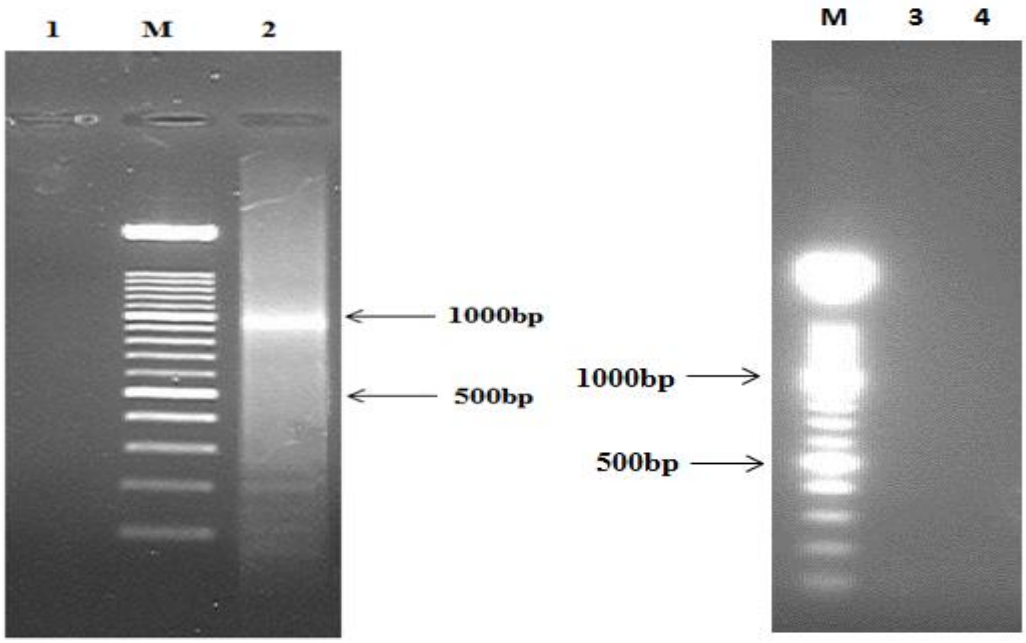


Figure 5.2 Agarose gel electrophoresis of 5' RACE PCR product of *SMO2* gene amplified from Barnea RNA

Lanes: M: molecular weight marker; Lane1: water-only negative control, Lane2: Barnea *SMO2* 5' RACE; Lane 3:negative control (Gene Racer primer and template and no GSP);Lane 4:negative control (GSP and template and no gene Racer primer)

These 5’ and 3’ RACE PCR products of *SMO2* gene from Barnea were gel purified, cloned and sequenced as described in Section 2.6.5 and 2.6.6. The digestion of ten plasmids containing the cloned 3' RACE products revealed inserts in the expected size range ~500bp (Figure 5.3), although variation existed in the size of the inserts.The digestion of ten plasmids containing the cloned 5' RACE products revealed inserts in the expected size range ~950bp (Figure 5.4).

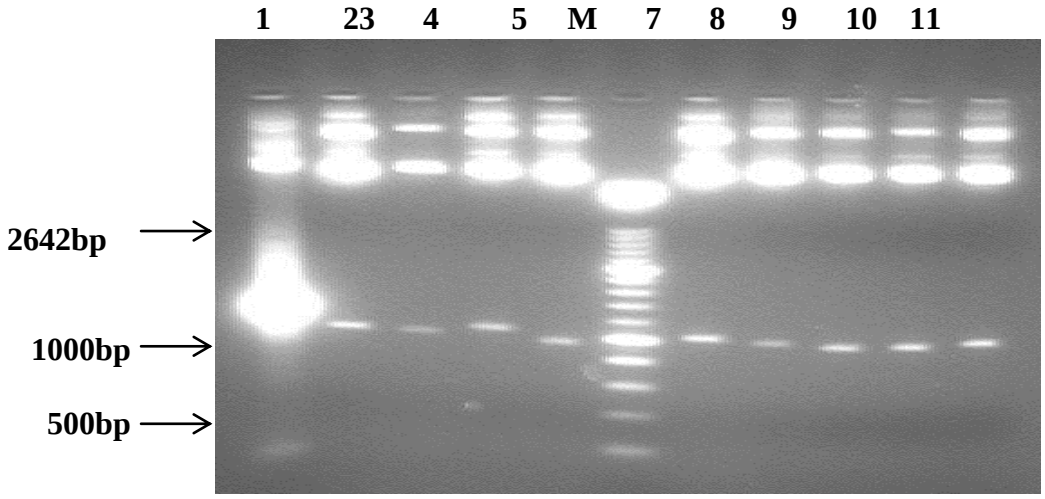


Figure 5.33’ RACE Barnea *SMO2* clones
PCR amplified product: SMO2-1 (3’) (Lane 1), *Eco*RI digested clones: SMO2-2 (3’), SMO2-3 (3’), SMO2-4 (3’), SMO2-5 (3’) (Lanes 2-5 respectively) and SMO2-6 (3’), SMO2-7 (3’), SMO2-8 (3’), SMO2-9 (3’), SMO2-10 (3’) (Lanes 6-10 respectively), M: 100 bp molecular weight marker.

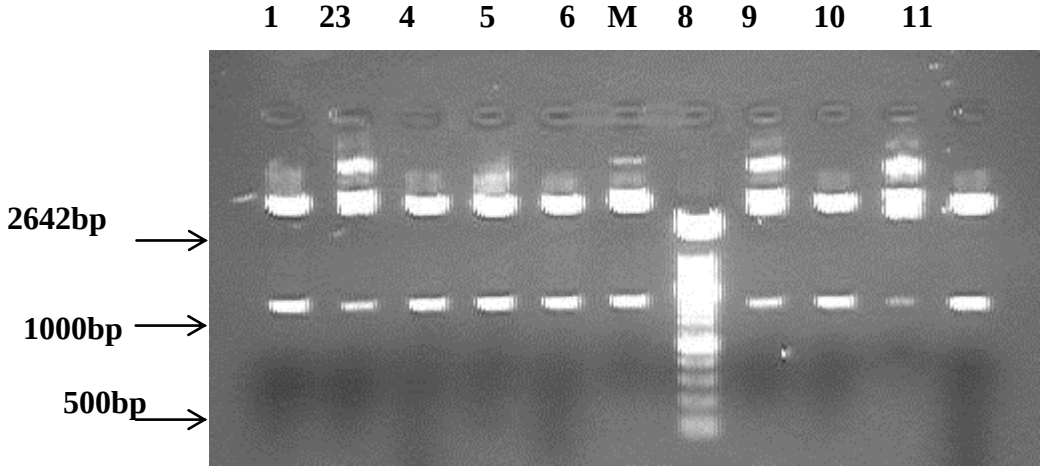


Figure 5.4 5’ RACE Barnea *SMO2* clones
*Eco*RI digested clones:SMO2-1 (5’), SMO2-2 (5’), SMO2-3 (5’), SMO2-4 (5’), SMO2-5 (5’), SMO2-6 (5’), (Lanes 1-6 respectively) and SMO2-7 (5’), SMO2-8 (5’), SMO2-9 (5’), SMO2-10 (5’) (Lanes 8-11 respectively), M: 100bp molecular weight marker

5.2.4 Sequencing and analysis of 3' RACE PCR products of *SMO2* genes from cv. Barnea

The sequencing of the ten 3' RACE products produced two different size products, a 471bp product (clone 1-5) and a 459bp product (clones 6-10). Clones 1, 2, 4 and 5 are highly conserved with nucleotide sequence similarity ranging from 97.2- 99.1%. Clones 6, 8, 9 and 10 are highly conserved with nucleotide sequence similarity ranging from 95.4-99.1%.Clone 3 is 88.1-89.8% similar to clones 1-5, while 73.3-75.1% similar to clones 6, 8, 9, and 10. Clone 7 is 73.3-80.4% similar to clones 1-5, while 94.3-98.4% similar to clones 6, 8, 9, and 10.The similarity between clones 1-5 and clones 6, 8, 9, and 10 ranges from 73.3-81.2% (Figure 5.5).

Seq->	1	2	3	4	5	6	7	8	9	10
SMO2-1 (3')	ID	0.980	0.898	0.985	0.991	0.812	0.804	0.785	0.785	0.808
SMO2-2 (3')		ID	0.881	0.978	0.972	0.797	0.789	0.770	0.770	0.793
SMO2-3 (3')			ID	0.890	0.894	0.738	0.733	0.747	0.751	0.738
SMO2-4 (3')				ID	0.980	0.800	0.791	0.773	0.773	0.795
SMO2-5 (3')					ID	0.812	0.803	0.784	0.784	0.807
SMO2-6 (3')						ID	0.980	0.954	0.958	0.991
SMO2-7 (3')							ID	0.943	0.947	0.984
SMO2-8 (3')								ID	0.986	0.954
SMO2-9 (3')									ID	0.958
SMO2-10 (3')										ID

Figure 5.5 Nucleotide sequence similarity of ten *SMO2* 3' RACE clones deduced from olive cultivar Barnea
[1- SMO2-1, 2- SMO2-2, 3-SMO2-3, 4-SMO2-4, 5-SMO2-5, 6-SMO2-6, 7-SMO2-7, 8-SMO2-8, 9-SMO2-9, 10-SMO2-10, ID-100% identical, seq-sequences]

5.2.5 Sequencing and analysis of 5' RACE PCR products of *SMO2* genes fromcv. Barnea

The sequencing of the ten *SMO2* 5' RACE products revealed an 885bp product from each. Clones 1-5 are highly conserved with nucleotide sequence similarity ranging from 98.3-

99.4% (Figure 5.6). Again, clones 6-10 are highly conserved with nucleotide sequence similarity ranging from 97.1-99.6%. The similarity between clones 1-5 and clones 6-10 ranged from 94.3-97.2%.

Seq->	1	2	3	4	5	6	7	8	9	10
SMO2-1 (5')	ID	0.993	0.985	0.989	0.983	0.968	0.966	0.948	0.946	0.950
SMO2-2 (5')		ID	0.989	0.994	0.987	0.972	0.970	0.952	0.951	0.954
SMO2-3 (5')			ID	0.986	0.984	0.967	0.964	0.945	0.945	0.949
SMO2-4 (5')				ID	0.984	0.969	0.967	0.949	0.948	0.951
SMO2-5 (5')					ID	0.964	0.962	0.943	0.943	0.946
SMO2-6 (5')						ID	0.990	0.971	0.971	0.975
SMO2-7 (5')							ID	0.971	0.971	0.975
SMO2-8 (5')								ID	0.990	0.994
SMO2-9 (5')									ID	0.996
SMO2-10 (5')										ID

Figure 5.6 Nucleotide sequence similarity of ten *SMO2* 5' RACE clones deduced from olive cultivar Barnea
[1- SMO2-1, 2- SMO2-2, 3-SMO2-3, 4-SMO2-4, 5-SMO2-5, 6-SMO2-6, 7-SMO2-7, 8-SMO2-8, 9-SMO2-9, 10-SMO2-10, ID-100% identical, seq-sequences].

5.2.6 Analysis of the overlapping sections of the 5' and 3' RACE PCR products of *SMO2* genes fromcv. Barnea

Comparison of the *SMO2* clones from both 5' and 3' RACE sequencing data of Barnea revealed that the 238bp overlapping section of the clones match significantly with sequence similarity ranging from 91.4-100% (Figure 5.7 and Figure IVA in Appendix IV). The clones 1-5 from the 5' RACE and 6-10 from the 3' RACE showed >96% identity over the overlapping section (Figure 5.7). And clones 6-10 from the 5' RACE and 1-5 from the 3' RACE showed >98% identity over the overlapping section (Figure 5.7).

5.2.7 Comparison of the deducedBarnea *SMO2* RACEclones with *SMO2* coding sequences from other plants

The multiple sequence alignments of the 5' and 3' RACE *SMO2* clones with previously published SMO/SMO2 coding sequences from other plants such as *Riccinus communis SMO*(Accession: XM_002520459.1),*Solanum lycopersicumSMO* (Accession: NM_001246951.1), *Gossipum arboreum SMO* (Accession: AF352575.1), *Vitis vinifera* partial *SMO2-2* (Accession: XM_002282617.1),*Glycine maxSMO* (Accession: NM_001253115.1), *Nicotiana benthamiana* partial *SMO* cds (Accession: AY321104.1), *Arabidopsis thaliana SMO2-1* (Accession: AF327853) and *Arabidopsis thaliana SMO2-2* (Accession: AF346734) (Figure IVA in Appendix IV) revealed a 752bp partial coding sequence from the 5' RACE clones and a 319bp partial coding sequence from the 3' RACE clones. The overlapping of the 5' and 3' RACE *SMO2* clones revealed two distinct classes of clones with either 783bp or 822bp open reading frames encoding proteins of 261 or 274 amino acids respectively (Figure IVA in Appendix IV).

5.3CHARACTERIZATION OFFULL LENGTH *SMO2* CODING SEQUENCES FROM cv. BARNEA

Primers based on sequences in the 5' and 3' UTRs of the *SMO2* sequences characterized from Barnea (Table 2.8, Figure IVB in Appendix IV), were used on Barnea cDNA resulting in the RT-PCR amplification of a product of predicted size (~950bp) (Figure 5.8). Cloning of this PCR product and subsequent *Eco*RI digestion to remove the inserts revealed ten clones from Barnea containing inserts of the expected size (Figure 5.9).

Seq->	1(5')	2(5')	3(5')	4(5')	5(5')	6(5')	7(5')	8(5')	9(5')	10(5')	1(3')	2(3')	3(3')	4(3')	5(3')	6(3')	7(3')	8(3')	9(3')	10(3')
1(5')	ID	0.991	0.995	0.995	0.974	0.931	0.927	0.931	0.931	0.931	0.936	0.931	0.931	0.927	0.927	0.982	0.991	0.982	0.982	0.982
2(5')		ID	0.995	0.995	0.974	0.931	0.927	0.931	0.931	0.931	0.936	0.931	0.931	0.927	0.927	0.982	0.991	0.982	0.982	0.982
3(5')			ID	1.000	0.978	0.936	0.931	0.936	0.936	0.936	0.940	0.936	0.936	0.931	0.931	0.987	0.995	0.987	0.987	0.987
4(5')				ID	0.978	0.936	0.931	0.936	0.936	0.936	0.940	0.936	0.936	0.931	0.931	0.987	0.995	0.987	0.987	0.987
5(5')					ID	0.914	0.910	0.914	0.914	0.914	0.919	0.914	0.914	0.919	0.910	0.965	0.974	0.965	0.965	0.965
6(5')						ID	0.987	0.991	0.991	0.991	0.987	0.982	0.991	0.987	0.987	0.940	0.931	0.931	0.931	0.931
7(5')							ID	0.995	0.995	0.995	0.991	0.987	0.995	0.991	0.991	0.936	0.927	0.927	0.927	0.927
8(5')								ID	1.000	1.000	0.995	0.991	1.000	0.995	0.995	0.940	0.931	0.931	0.931	0.931
9(5')									ID	1.000	0.995	0.991	1.000	0.995	0.995	0.940	0.931	0.931	0.931	0.931
10(5')										ID	0.995	0.991	1.000	0.995	0.995	0.940	0.931	0.931	0.931	0.931
1(3')											ID	0.995	0.995	0.991	0.991	0.944	0.936	0.936	0.936	0.936
2(3')												ID	0.991	0.987	0.987	0.940	0.931	0.931	0.931	0.931
3(3')													ID	0.995	0.995	0.940	0.931	0.931	0.931	0.931
4(3')														ID	0.991	0.936	0.927	0.927	0.927	0.927
5(3')															ID	0.936	0.927	0.927	0.927	0.927
6(3')																ID	0.982	0.991	0.991	0.991
7(3')																	ID	0.982	0.982	0.982
8(3')																		ID	0.991	0.991
9(3')																			ID	0.991
10(3')																				ID

Figure 5.7 Nucleotide sequence similarity of the overlapping sections of *SMO2* 5'and3' RACE clones deduced from olive cultivar Barnea
[1- SMO2-1, 2- SMO2-2, 3-SMO2-3, 4-SMO2-4, 5-SMO2-5, 6-SMO2-6, 7-SMO2-7, 8-SMO2-8, 9-SMO2-9, 10-SMO2-10, ID-100% identical, seq-sequences] The clones 1-5 from the 5' RACE and 6-10 from the 3' RACE showing >96% identity have been highlighted in the brown box. And clones 6-10 from the 5' RACE and 1-5 from the 3' RACE showing >98% identity have been highlighted in the green box.

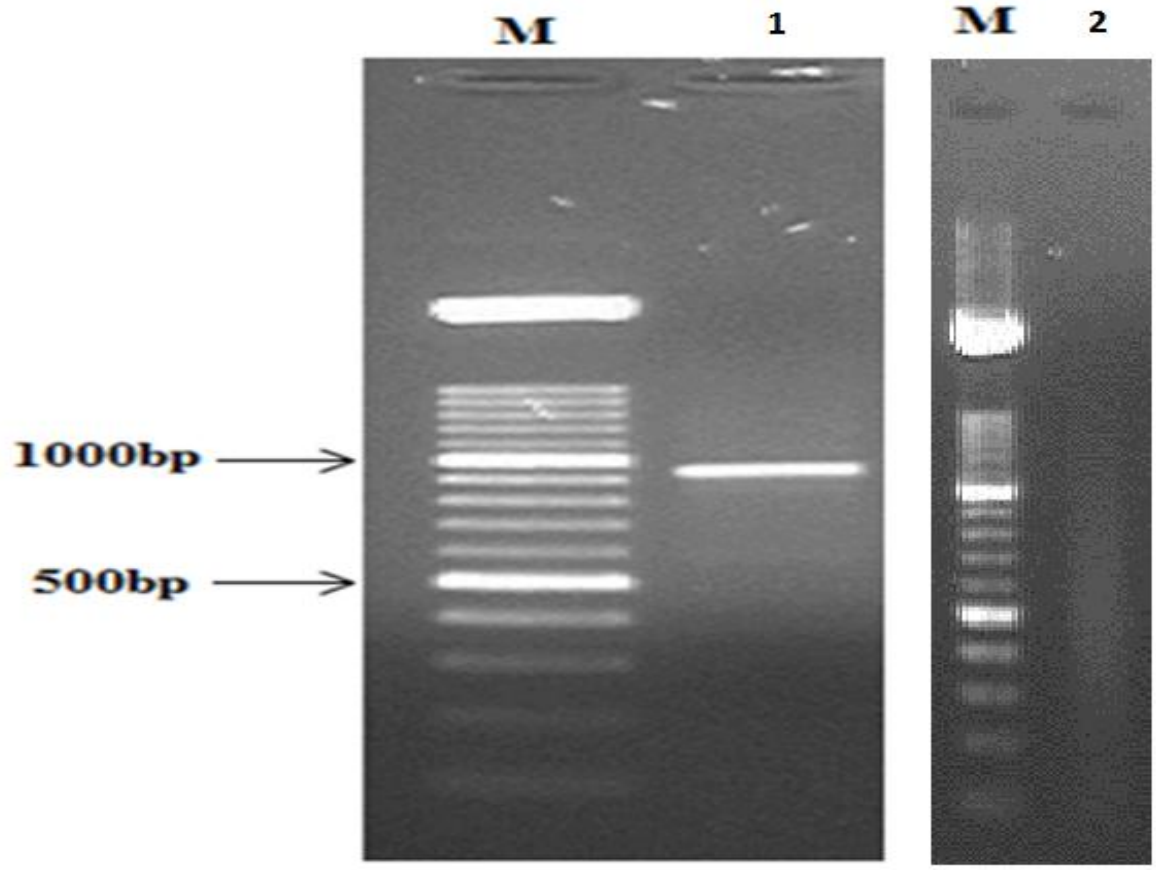


Figure 5.8Amplified product of full length *SMO2* genes from cv. Barnea
Lanes: M: 100bp molecular weight marker; Lane1: Barnea *SMO2*; Lane 2: water only negative control

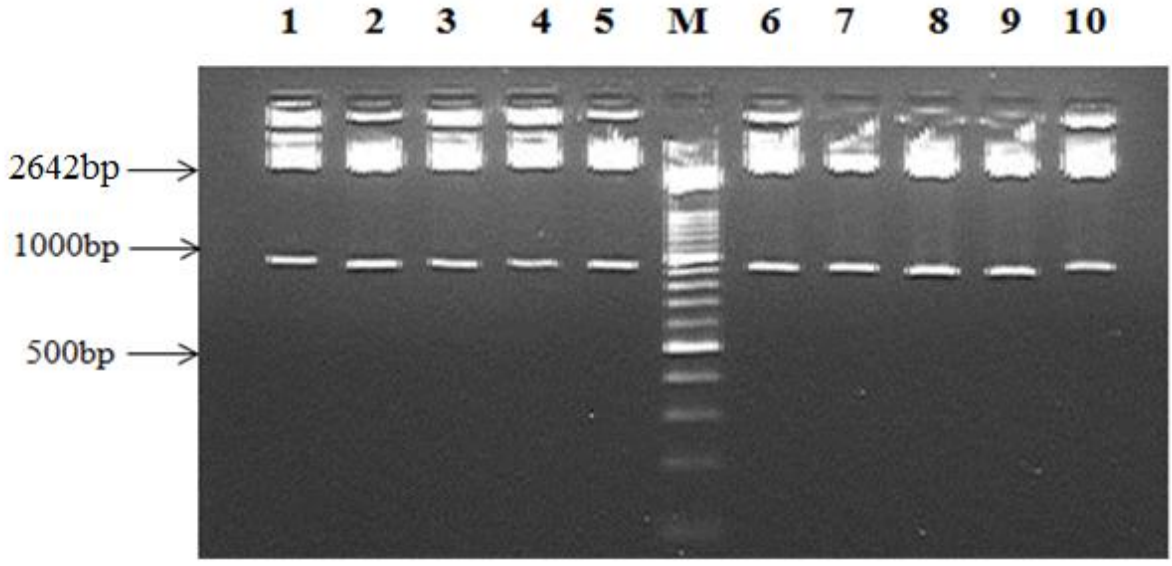


Figure 5.9 Full length *SMO2* clones from cv. Barnea
*Eco*RI digested clones:SMO2-1, SMO2-2, SMO2-3, SMO2-4, SMO2-5 (Lanes 1-5 respectively) and SMO2-6, SMO2-7, SMO2-8, SMO2-9 and SMO2-10 (Lanes 6-10 respectively), M: 100bp molecular weight marker.

5.3.1 Sequencing and analysis of full length *SMO2* genes from cv. Barnea

Sequencing of ten full length *SMO2*clones from Barnea revealed six clones (F1, F2, F3, F4, F5 and F6) containing 944bp sequences with 99.1-99.8% sequence identity andfour clones (F7, F8, F9 and F10) containing 950bp sequences with 97.3-100% sequence identity (Figure 5.10 and Figure IVB in Appendix IV).The DNA sequence identity between these two sequence classes ranged from 89.2-91.6%. Clone F1 and clone F7 were selected as representatives of each sequence class and henceforth referred to as *OeSMO2-1a* and *OeSMO2-2a*. The other *SMO2* clones and the SNPs identified in these are discussed later in the chapter (Section 5.6).

The cDNA sequence of *OeSMO2-1a* consisted of944bp nucleotides, which included 25bp of 5'UTR, 96bp of 3' UTR and an open reading frame of 822bp encoding a putative protein of 274 amino acid residues (Accession: KC862263)(Figure 5.11). The calculated molecular mass and pI of *OeSMO2-1a* were 31.8 kDa and 8.55, respectively.

Seq->	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Bar-F1-OeSMO2	ID	0.998	0.997	0.995	0.993	0.993	0.897	0.897	0.892	0.916
Bar-F2-OeSMO2		ID	0.998	0.994	0.992	0.994	0.896	0.896	0.891	0.915
Bar-F3-OeSMO2			ID	0.993	0.991	0.993	0.895	0.895	0.890	0.914
Bar-F4-OeSMO2				ID	0.997	0.997	0.897	0.897	0.892	0.916
Bar-F5-OeSMO2					ID	0.995	0.895	0.895	0.890	0.914
Bar-F6-OeSMO2						ID	0.895	0.895	0.890	0.914
Bar-F7-OeSMO2							ID	1.000	0.991	0.975
Bar-F8-OeSMO2								ID	0.991	0.975
Bar-F9-OeSMO2									ID	0.973
Bar-F10-OeSMO2										ID

Figure 5.10 Nucleotide sequence similarity between Barnea *SMO2*clones
The ten Barnea clones are designated as F1, F2, F3, F4, F5, F6, F7, F8, F9 and F10.
Cultivars: Bar: Barnea. ID-100% identical, seq-sequences.

The cDNA sequence of *OeSMO2-2a* consisted of950bp nucleotides, which included a 25bp 5'UTR, 141bp 3' UTR and an open reading frame of 783bp encoding a protein of 261 amino acid residues (Accession: KC862264)(Figure 5.12). The calculated molecular mass and pI of *OeSMO2-2* was 30.2 kDa and 8.92, respectively.

The coding sequences of *OeSMO2-1a* and *OeSMO2-2a* share 91.24% similarity to each other at amino acid level and are 89.65% identical at the nucleotide level. The hydropathy analysis of the two putative SMO2 proteins (Section 2.6.7.6) predicted that the encoded polypeptides contained a C-terminal hydrophilic region and at least four putative hydrophobic transmembrane domains within the N-terminal region at amino acid residues 31-53, 77-96, 113-131 and 167-194 for OeSMO2-1aand residues 31-52, 77-96, 108-129 and 167-194 for OeSMO2-2a. Pfam analysis suggests that OeSMO2belongs to the fatty acid hydroxylase superfamily which includes integral membrane enzymes such as C-5 sterol desaturases and C-4 sterol methyl oxidases which are involved in catalysing desaturations and hydroxylations.Further, the OeSMO2 protein sequences revealed three distinct histidine-rich motifs H^{127}RILH, H^{140}SVHH and H^{220}DYHH in both isoforms (Figure 5.11 and Figure 5.12).

5.3.2 Homology analysis of the deducedOeSMO2 proteins of Barnea with other plant species

To further investigate the homology of proteins OeSMO2-1a and OeSMO2-2a with SMO proteins characterized in other plant species, a comparison of these amino acid sequences were conducted (Figure 5.13 and Figure 5.14). OeSMO2-1a and OeSMO2-2a proteins showed high degree of sequence identity with SMO/SMO2 sequences isolated from various other plant species (Figure 5.13-A) such as *Arabidopsis thaliana SMO2-1*(80.2-82.6%), *Arabidopsis thaliana SMO2-2*(82.7-84.2%), *Ricinus communis* (85.3-85.5%), *Solanum lycopersicum* (84.6-85.1%), *Gossypium arboretum* (82-82.1%), *Nicotiana benthamiana*(75-76.9%), *Vitis vinifera* (83.5-83.7%) and *Glycine max* (83.8-84.1%). OeSMO2-1a and OeSMO2-2a showed limited sequence identity to *Arabidopsis thaliana*SMO1-1(35-35.6%), SMO1-2 (34.2-35.8%) and SMO1-3 (35.1-36.1%) (Figure 5.13-B). Three histidine rich motifs (described above) have been identified in olive SMO2s which align perfectly with these motifs in other plants SMOs at positions 127-131, 138-144, 220-224 (Figure 5.14).

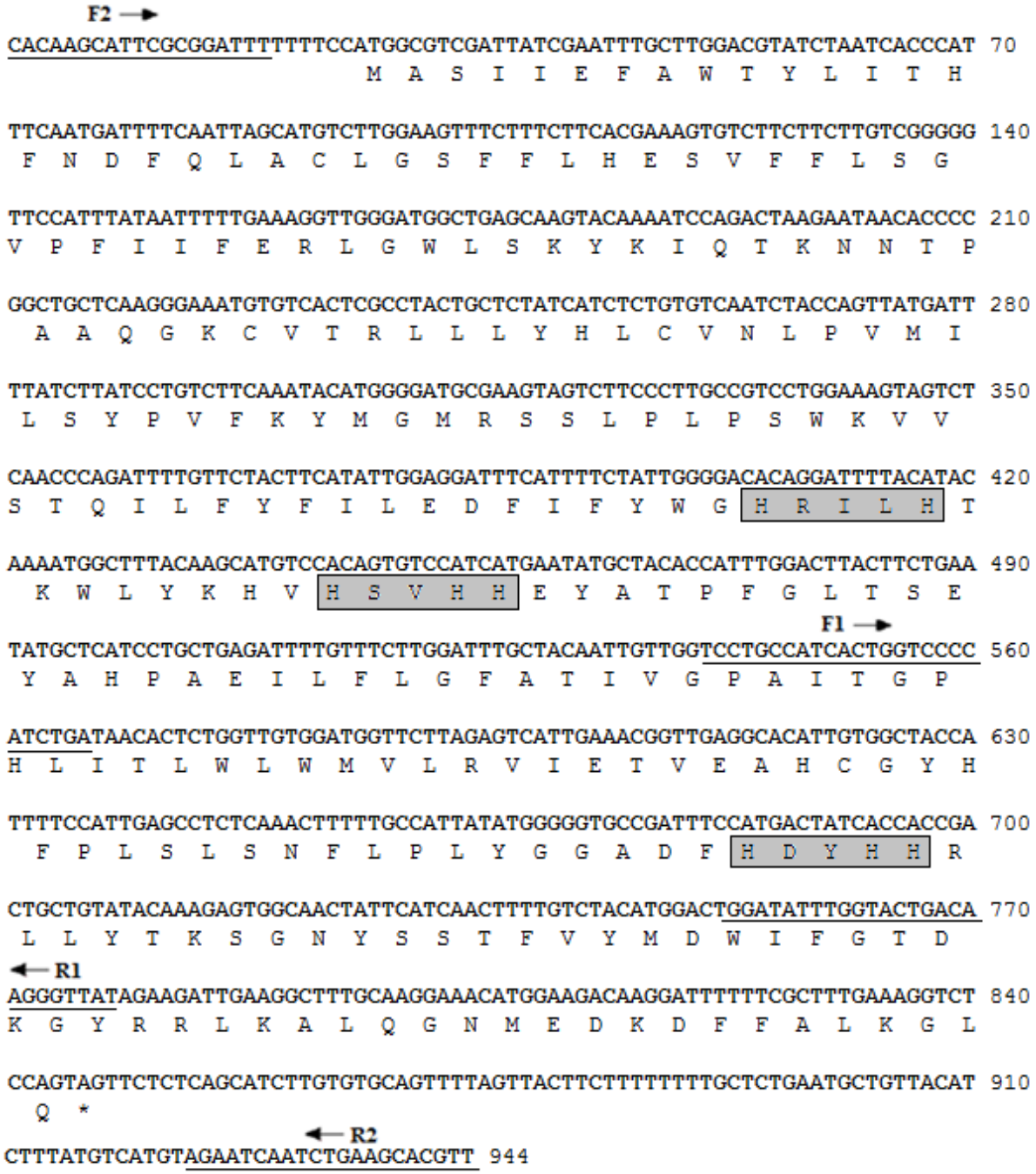


Figure 5.11The full length cDNA and deduced amino acid sequences of *OeSMO2-1a*(Accession: KC862263)

Underlined letters represent the DNA sequence of PCR primers used for RACE PCR (F1 and R1) and amplification of full length *SMO2* genes (F2 and R2). The three histidine rich motifs of SMO2 proteins are shown in grey boxes.

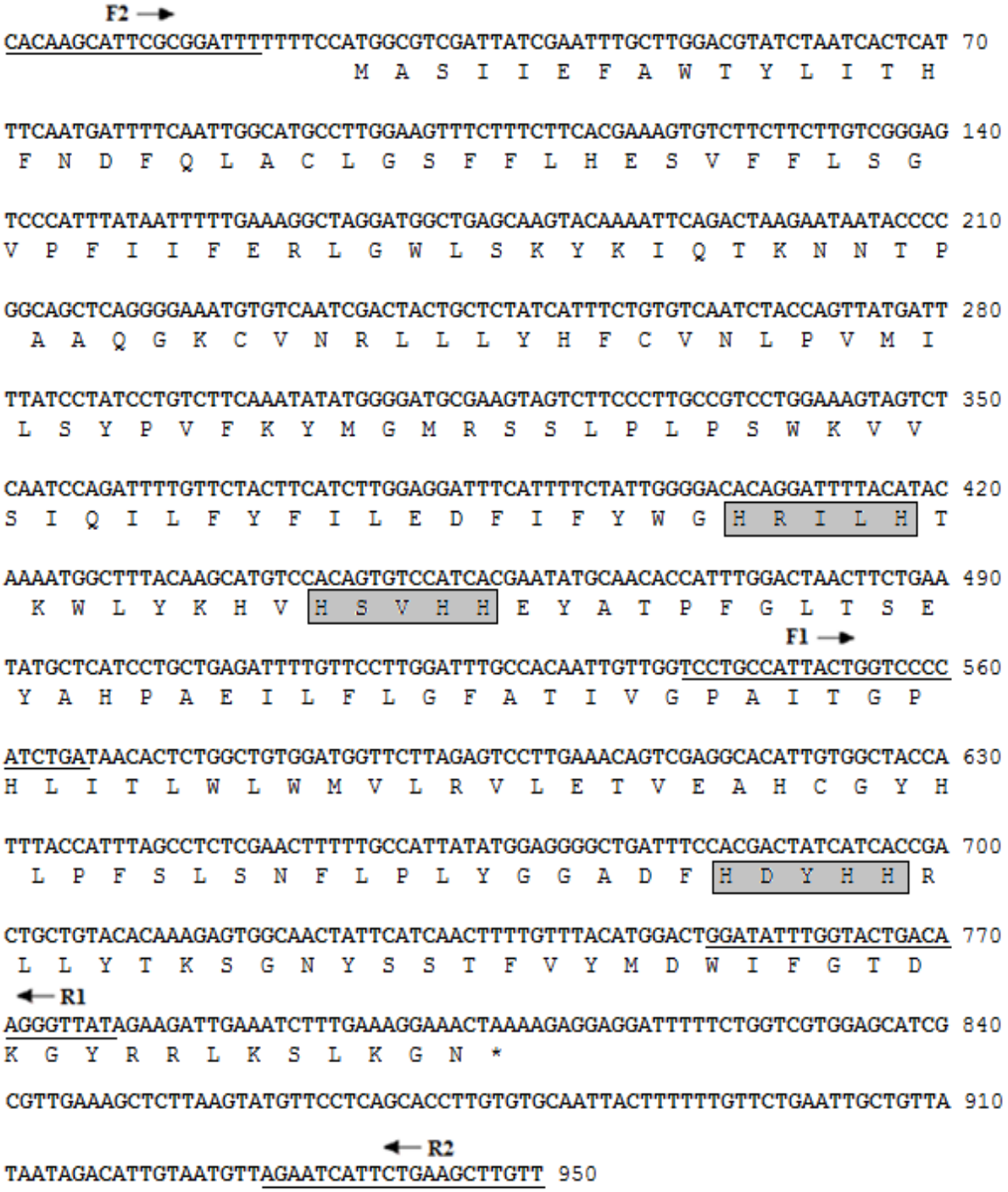


Figure 5.12 The full length cDNA and deduced amino acid sequences of *OeSMO2-2a*(Accession: KC862264)

Underlined letters represent the DNA sequence of PCR primers used for RACE PCR (F1 and R1) and amplification of full length *SMO2* genes (F2 and R2). The three histidine rich motifs of SMO2 proteins are shown in grey boxes.

Seq->	Oe-1	Oe-2	Rc	Sl	At2-1	At2-2	Ga	Gm	Nb	Vv
Oe-1	ID	0.915	0.853	0.846	0.802	0.827	0.820	0.838	0.750	0.835
Oe-2		ID	0.855	0.851	0.826	0.842	0.821	0.841	0.769	0.837
Rc			ID	0.866	0.814	0.840	0.907	0.911	0.780	0.877
Sl				ID	0.810	0.836	0.851	0.837	0.780	0.840
At2-1					ID	0.890	0.821	0.800	0.746	0.796
At2-2						ID	0.836	0.826	0.736	0.822
Ga							ID	0.867	0.754	0.859
Gm								ID	0.749	0.856
Nb									ID	0.729
Vv										ID

A

Seq->	Oe-1	Oe-2	At1-1	At1-2	At1-3
Oe-1	ID	0.905	0.331	0.324	0.346
Oe-2		ID	0.319	0.308	0.324
At1-1			ID	0.832	0.704
At1-2				ID	0.728
At1-3					ID

B

Figure 5.13 Comparison ofthe OeSMO2-1a and OeSMO2-2aproteins with cloned SMO/SMO2 (A) and SMO1 (B) from other plants at amino acid level based on identity percentage

Oe1: OeSMO2-1a(Accession: KC862263),Oe2: OeSMO2-2a(Accession: KC862264),Rc: *Riccinus communis* SMO (Accession: XM_002520459.1), Sl: *Solanum lycopersicum* SMO (Accession:NM_001246951.1), Ga: *Gossypium arboreum* SMO (Accession: AF352575.1), Vv: *Vitis vinifera* SMO2-2 (Accession: XM_002282617.1), Gm: *Glycine max* SMO (Accession: NM_001253115.1), Nb: *Nicotiana benthamiana* partial SMO cds (Accession:AY321104.1, At2-1: *Arabidopsis thaliana* SMO2-1 (Accession: AF327853), At2-2: *Arabidopsis thaliana* SMO2-2 (Accession: AF346734), At1-1: *Arabidopsis thaliana* SMO1-1 (Accession: NM_117281.2), At1-2: *Arabidopsis thaliana* SMO1-2 (Accession: NM_118404.4), At1-3: *Arabidopsis thaliana* SMO1-3 (Accession: NM_118403.1), ID-100% identical, Seq: Sequences).

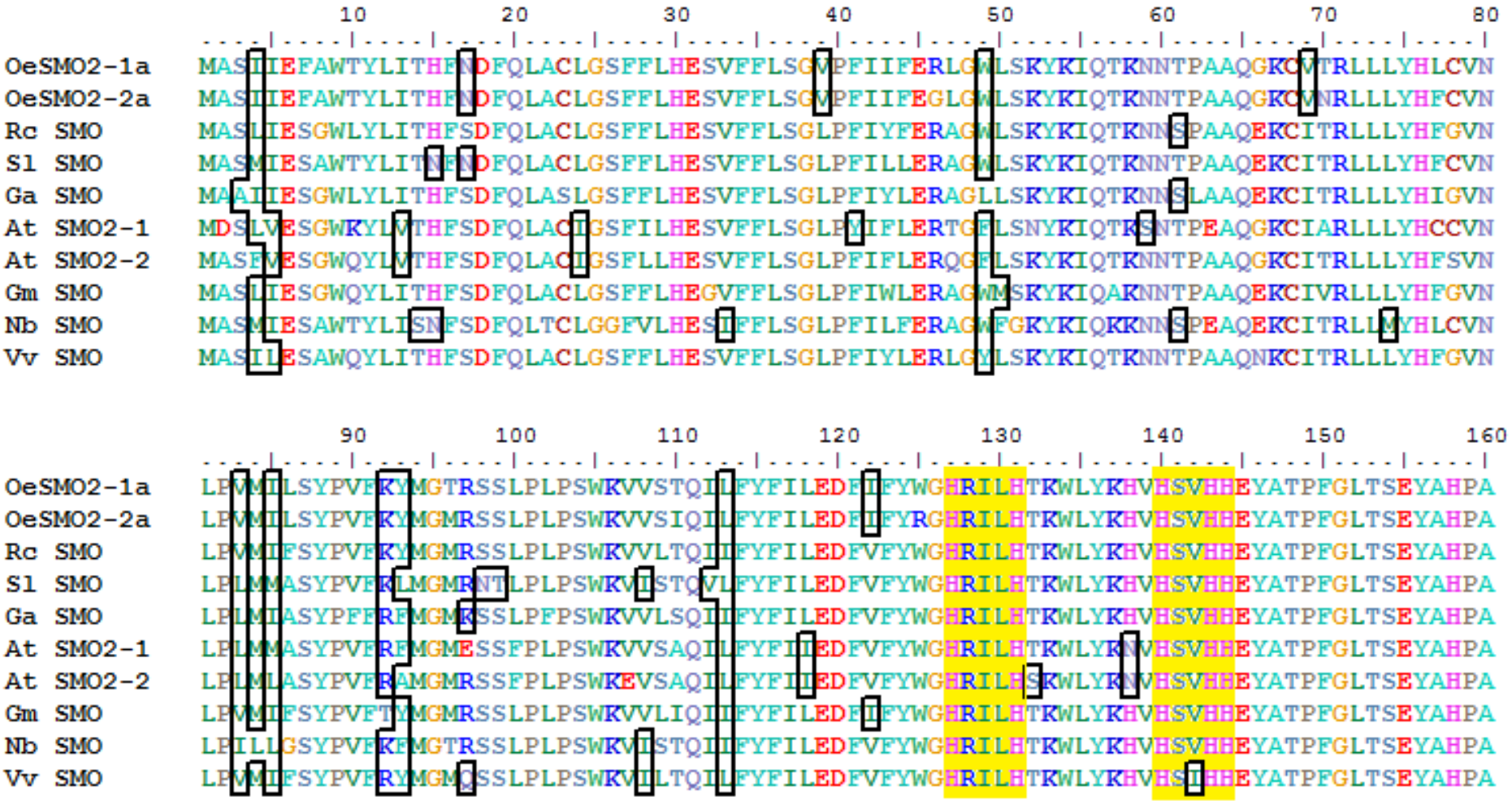


Figure 5.14 Allignment of putative protein products of the OeSMO2-1a and OeSMO2-2a cDNAs isolated from Barnea with other plant SMOs (continued)

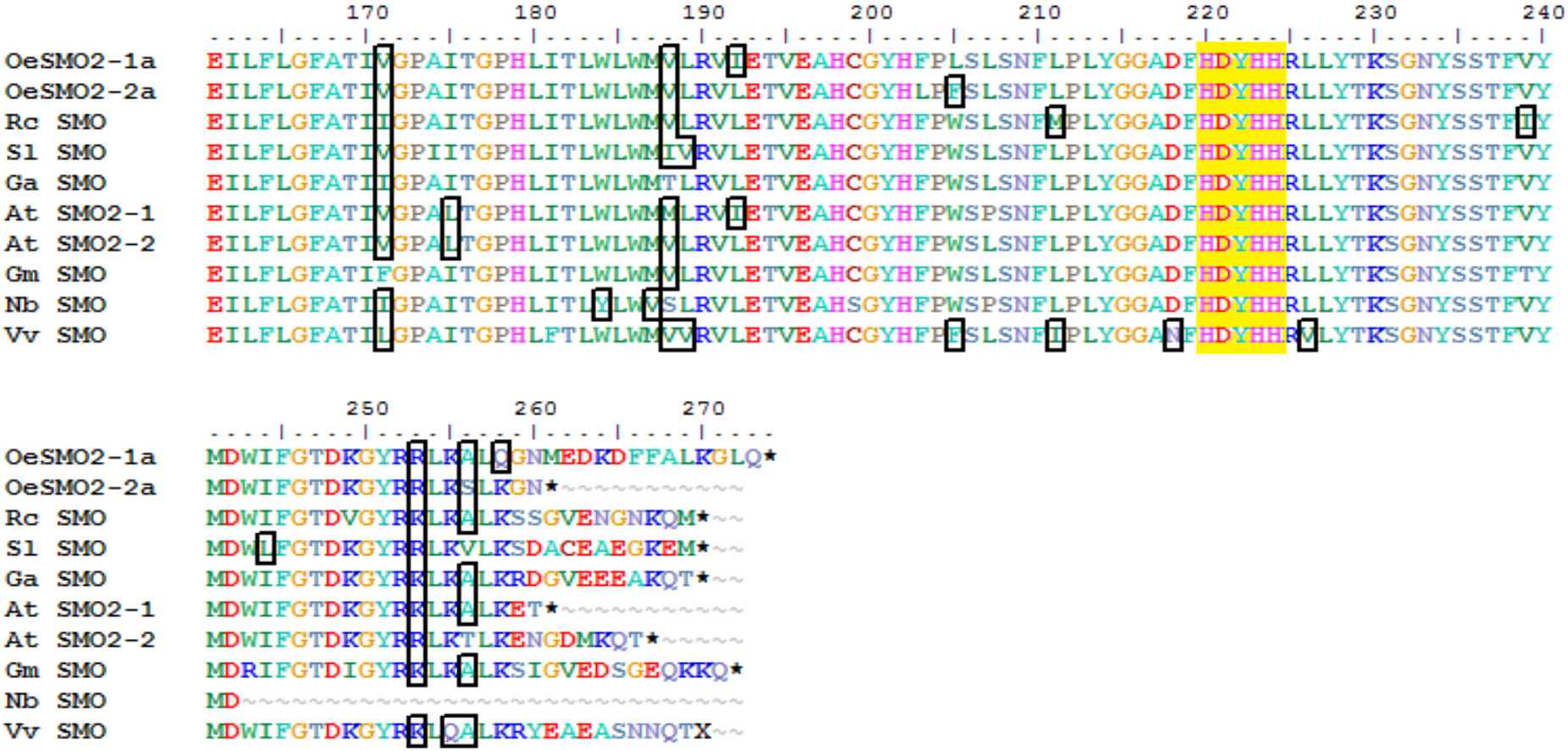


Figure 5.14 Allignment of putative protein products of the OeSMO2-1a and OeSMO2-2a cDNAs isolated from Barnea with other plant SMOs

OeSMO2-1a (Accession: KC862263),OeSMO2-2a (Accession: KC862264),Rc: *Riccinus communis SMO* (Accession: XM_002520459.1), Sl: *Solanum lycopersicumSMO* (Accession: NM_001246951.1), Ga: *Gossipum arboreum SMO* (Accession: AF352575.1), Vv: *Vitis vinifera* partial *SMO2-2* (Accession: XM_002282617.1), Gm: *Glycine maxSMO* (Accession: NM_001253115.1), Nb: *Nicotiana benthamiana* partial *SMO* cds (Accession: AY321104.1, At SMO2-1: *Arabidopsis thaliana SMO2-1* (Accession: AF327853), At SMO2-2: *Arabidopsis thaliana SMO2-2* (Accession: AF346734). The three characteristic histidine boxes encountered in SMO2 proteins of all species are highlighted in yellow.

5.4 *SMO2* SOUTHERN BLOTTING OF BARNEA, FRANTOIO AND PICUAL

In order to estimate the copy number of *SMO2* genes inolives, a Southern blotting experiment was performed with a 671bp *SMO2* probe amplified from Barnea cDNA(Table 2.13, Figure 5.15) as described in Section 2.9. The probe was also amplified from genomic DNA and sequenced which confirmed their identity and the absence of introns within the probe (data not shown). To get a preliminary idea of any expected differences in copy number, or the presence of RFLPs between the threeolive cultivars Barnea, Frantoio and Picual, the Southern blotting experiment was performed using genomic DNA isolated from each cultivar (Section 2.4.1) and digested separately with three restriction enzymes *Eco*RI, *Hind*III and *Bam*H1 (Section 2.9.1).

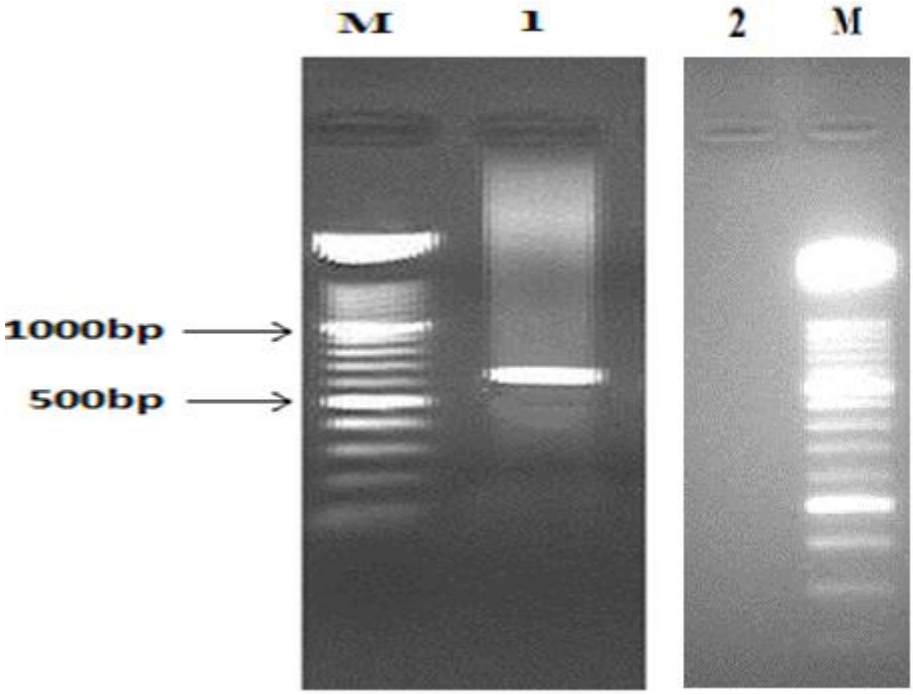


Figure 5.15 Amplification of *SMO2* probe (671bp) for Southern blotting
Lane 1: SMO2 probe; Lane 2: water only negative control; M: 100bp molecular weight marker.

The Southern blotting of Barnea gDNA in the *Eco*RI digest revealed three weak hybridizing fragments approximately 10kb, 7kb and 5.5kb and a large intensely hybridizing fragment between the range of 3kb and 4kb that may likely represent 3 or 4 individual fragments in that size range based on the band intensity (Figure 5.16, Figure 5.17).The *Eco*RI digest in

Frantoio revealed four strong hybridizing fragments of approximately 10kb, 7kb, 4.2kb and 4.0kb. It appears as if Picual has the same weak hybridizing fragments of 10kb and 7kb and the intensely hybridizing group of fragments as Barnea, plus an additional, slightly smaller fragment in place of the missing 5.5kb fragment.

The Southern blotting of Barnea gDNA in the *Hind*III digest appears to have four hybridising fragments approximately 7kb, 4.2kb, 1.7kb and 1kb respectively, plus an intensely hybridizing large molecular weight fragment (10-20kb) (Figure 5.16, Figure 5.17).Its pattern is otherwise identical to the Picual pattern. The Frantoio pattern contains a unique smaller fragment (2.5kb) in place of the largest clear band (~10kb) in Picual and Barnea.In all three olive cultivars, the *Bam*HI digest produced an intensely strong hybridizing fragment ranging from 10-18kb (Figure 5.16).

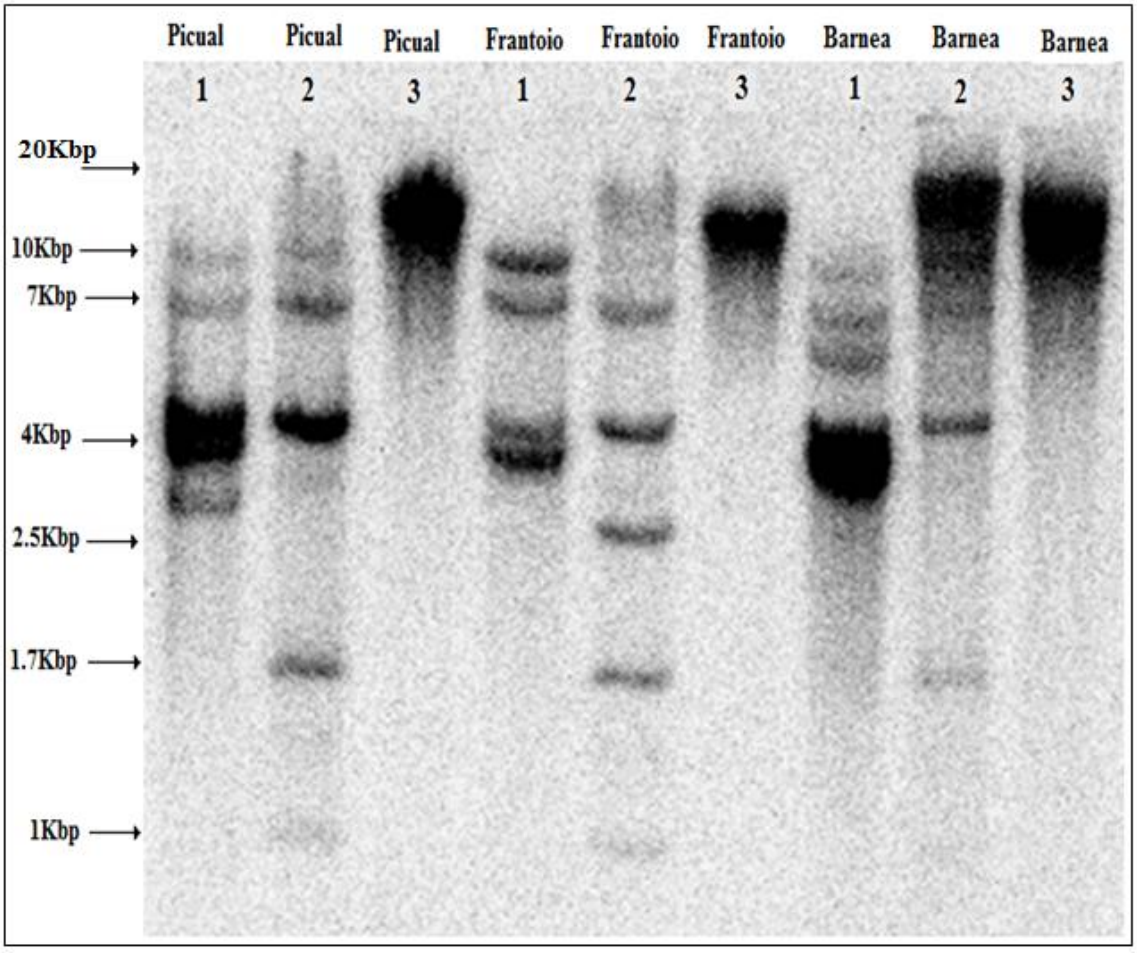


Figure 5.16 Southern blotting of gDNA isolated from Barnea, Frantoio and Picual hybridised with a radiolabeled 671bp *SMO2* cDNAprobe

Genomic DNA was digested with *Eco*RI (Lane 1), *Hind*III (Lane 2) and *Bam*HI (Lane 3) as indicated. Approximate molecular weights of the hybridising fragments (Kbp) are indicated on the left.

	*Hind*III digest		
Size of fragments (Kbp)	Picual	Frantoio	Barnea
18	ND	ND	strong hybridizing fragment ranging from 10-20 Kbp
10	+	ND	
7	+	+	+
4.2	+	+	+
2.5	ND	+	ND
1.7	+	+	+
1	+	+	+

	*Eco*RI digest		
Size of fragments (Kbp)	Picual	Frantoio	Barnea
10	+	+	+
7	+	+	+
5.5	ND	ND	+
4.2	strong hybridizing fragment ranging from 3.5-4.2 Kbp	+	strong hybridizing fragment ranging from 3-4.2 Kbp
4.0		+	
3		ND	
2.8	+	ND	

Figure 5.17 Hybridising pattern of *SMO2* probe in Barnea, Frantoio and Picual
Distinct RFLPs observed between cultivars are highlighted in yellow. As no RFLPs were identified for *Bam*HI digest therefore details were omitted. ND: not detected; +: presence of hybridising fragment.

5.5CHARACTERIZATION OFFULL LENGTH *SMO2* GENES IN OLIVE CULTIVARS, FRANTOIO AND PICUAL

5.5.1RNA extraction from Frantoio and Picual drupes

As described in Section 4.5.1, total RNA was extracted from the mesocarp of Frantoio and Picual fruits harvested from the 2009 crop for eight sample timepoints between 96 and 170 DAF revealing all samples had intact high quality RNA.

5.5.2 Amplification of full length genes of *SMO2* from Frantoio and Picual

Primers based on sequences in the 5' and 3' UTRs of the *SMO2* sequences characterized from Barnea were used on Frantoio and Picual cDNA resulting in the amplification of the expected sized products of ~950bp (Table 2.8, Figure IVB in Appendix IV, Figure 5.18).Cloning of these PCR products and subsequent *Eco*RI digestion to remove the inserts revealed nine and ten clones from Frantoio and Picual respectively containing expected size inserts (Figure 5.19).

5.5.3 Sequencing and analysis of full length *SMO2* genes from Frantoio

Nine full length *SMO2* clones were sequenced from the cultivar Frantoio (Figure 5.20 and Figure IVB in Appendix IV).These clones showed 98.9-100% DNA sequence identity to Barnea *OeSMO2-1a* allele and 88.9-89.6% DNA sequence identity to *OeSMO2-2a* allele.

5.5.4 Sequencing and analysis of full length *SMO2* genes from Picual

Ten full length *SMO2* clones were sequenced from the cultivarPicual (Figure 5.21 andFigure IVB in Appendix IV). These clones have 99.3-100% DNA sequence identity to Barnea *OeSMO2-1a* allele and 89.1-89.6% DNA sequence identity to Barnea *OeSMO2-2a* allele.

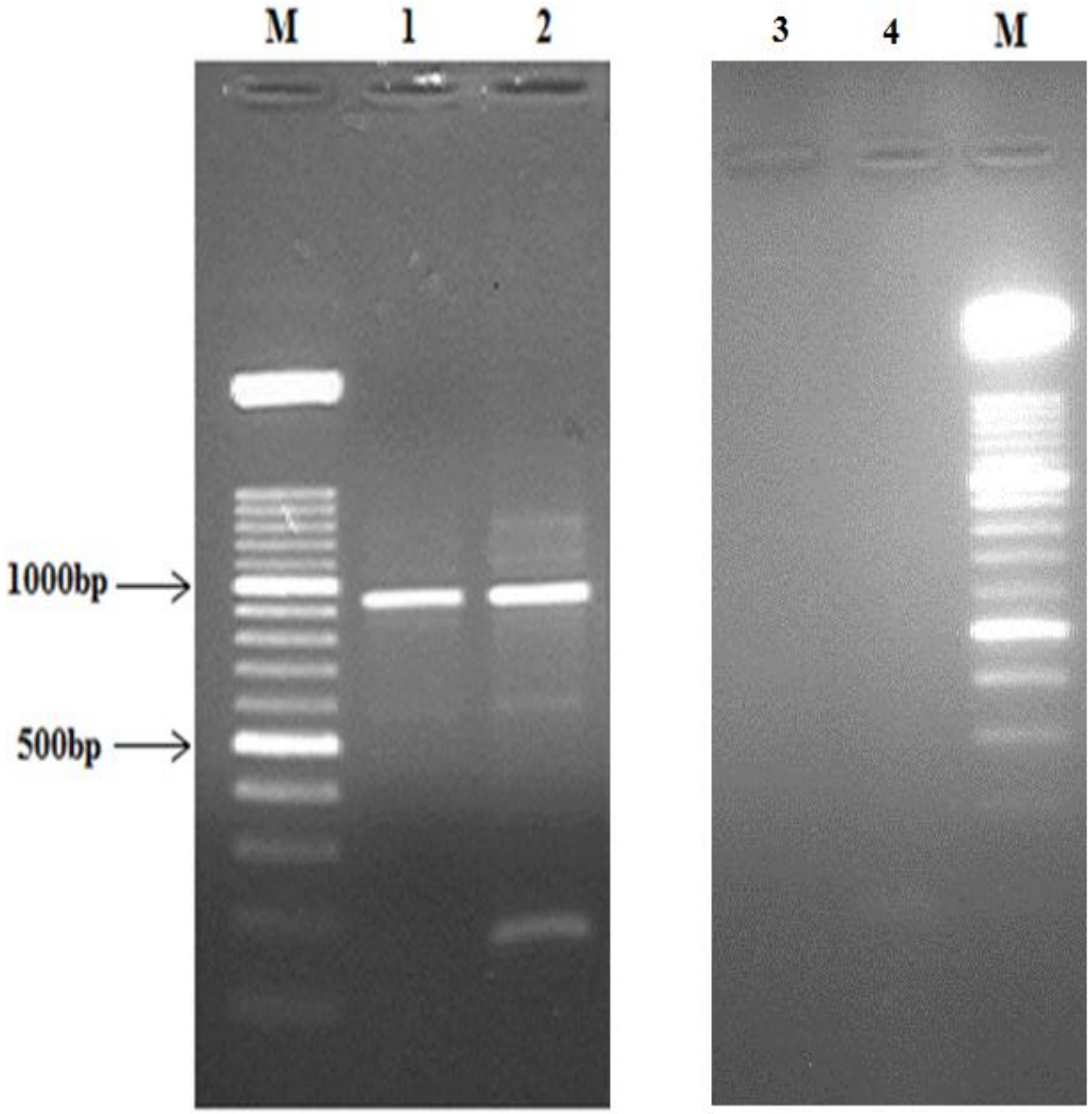


Figure 5.18 Amplified products of full length *SMO2* genes from Frantoio and Picual Lanes: M: molecular weight marker; Lane1: Frantoio *SMO2*; Lane2: Picual *SMO2*; Lane 3: water only negative control (Frantoio); Lane 4: water only negative control (Picual)

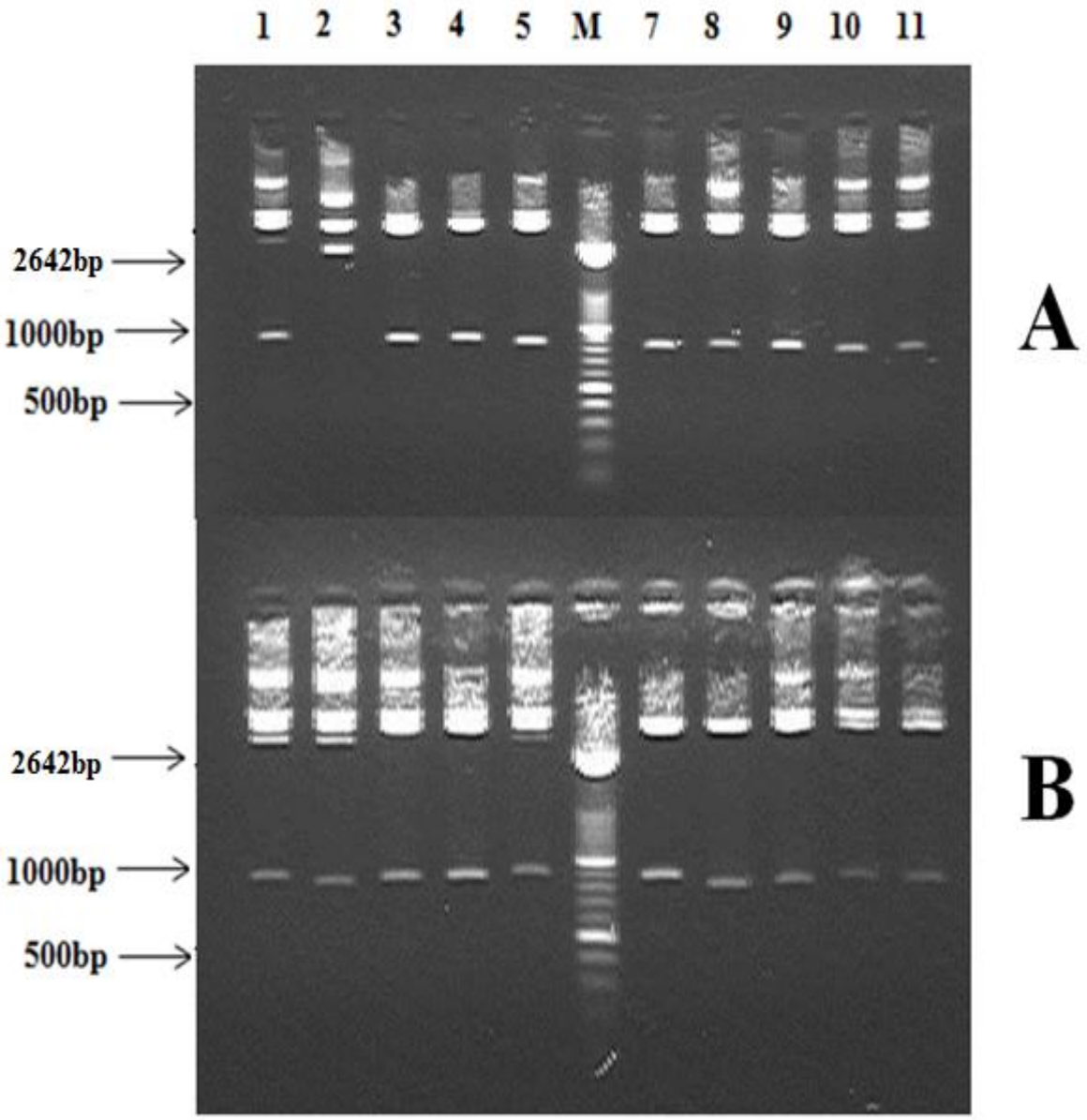


Figure 5.19 Full length *SMO2* clones from Frantoio (A) and Picual (B)
*Eco*RI digested clones:SMO2-1, SMO2-2, SMO2-3, SMO2-4, SMO2-5 (Lanes 1-5 respectively) and SMO2-6, SMO2-7, SMO2-8, SMO2-9 and SMO2-10 (Lanes 6-10 respectively), M: 100bp molecular weight marker.

Seq->	B1	B2	F1	F2	F3	F4	F5	F6	F7	F8	F9
Bar-OeSMO2-1a	ID	0.896	1.000	0.998	0.997	0.996	0.996	0.997	0.989	0.987	0.992
Bar-OeSMO2-2a		ID	0.896	0.895	0.894	0.892	0.892	0.894	0.891	0.890	0.889
Fran-SMO2-F1			ID	0.998	0.997	0.996	0.996	0.997	0.989	0.987	0.992
Fran-SMO2-F2				ID	0.996	0.995	0.995	0.996	0.987	0.986	0.991
Fran-SMO2-F3					ID	0.996	0.993	0.995	0.986	0.985	0.992
Fran-SMO2-F4						ID	0.992	0.993	0.985	0.984	0.991
Fran-SMO2-F5							ID	0.996	0.985	0.984	0.991
Fran-SMO2-F6								ID	0.986	0.985	0.992
Fran-SMO2-F7									ID	0.998	0.981
Fran-SMO2-F8										ID	0.980
Fran-SMO2-F9											ID

Figure 5.20 Nucleotide sequence similarity between Barnea *SMO2* (*OeSMO2-1a* and *OeSMO2-2a*) and Frantoio *SMO2* clones

Frantoio clones showing 98.9-100% DNA sequence identity to Barnea *OeSMO2-1a* allele has been highlighted in yellow box. Frantoio clones showing 88.9-89.6% DNA sequence identity to *OeSMO2-2a* allele has been highlighted in grey box. The nine Frantoio clones are designated as F1, F2, F3, F4, F5, F6, F7, F8 and F9. Cultivars: Bar: Barnea; Fran: Frantoio. ID-100% identical, seq-sequences.

Seq->	B1	B2	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Bar OeSMO2-1a	ID	0.896	1.000	0.996	0.995	0.995	0.997	0.993	0.997	0.996	0.995	0.993
Bar OeSMO2-2a		ID	0.896	0.894	0.892	0.891	0.896	0.892	0.896	0.895	0.891	0.890
Pic-SMO-F1			ID	0.996	0.995	0.995	0.997	0.993	0.997	0.996	0.995	0.993
Pic-SMO-F2				ID	0.993	0.991	0.993	0.990	0.993	0.992	0.991	0.990
Pic-SMO-F3					ID	0.990	0.992	0.989	0.992	0.991	0.990	0.989
Pic-SMO-F4						ID	0.992	0.989	0.992	0.991	0.992	0.991
Pic-SMO-F5							ID	0.996	1.000	0.998	0.992	0.991
Pic-SMO-F6								ID	0.996	0.995	0.989	0.987
Pic-SMO-F7									ID	0.998	0.992	0.991
Pic-SMO-F8										ID	0.991	0.990
Pic-SMO-F9											ID	0.998
Pic-SMO-F1												ID

Figure 5.21 Nucleotide sequence similarity between Barnea *SMO2* (*OeSMO2-1a* and *OeSMO2-2a*) and Picual*SMO2* clones

Picual clones showing 99.3-100% DNA sequence identity to Barnea *OeSMO2-1a* allele has been highlighted in yellow box. Picual clones showing 89.1-89.6% DNA sequence identity to *OeSMO2-2a* allele has been highlighted in grey box. The ten Picual clones are designated as F1, F2, F3, F4, F5, F6, F7, F8, F9 and F10. Cultivars: Bar: Barnea; Pic: Picual. ID-100% identical, seq-sequences.

5.6 IDENTIFICATION OF SNPS IN *SMO2* ALLELES IN OLIVE CULTIVARS BARNEA, FRANTOIO AND PICUAL

SMO2 clones isolated from the three cultivars, Barnea, Frantoio and Picual were aligned together to identify cultivar specific alleles and/or alleles having key SNPs that are present in all cultivars(Figure 5.22, Figure 5.23, Figure 5.24).

All *SMO2* clones sequenced from the cultivars Frantoio and Picual (Section 5.5.3, Section 5.5.4 and Figure IVB in Appendix IV) appear to be far more similar to *OeSMO2-1a* allele than to *OeSMO2-2a* allele as shown in Figures 5.22, 5.23 and5.24, however key SNPs that were identified in some of the Barnea *SMO2* alleles were also seen in Frantoio and Picual. Alleles common to all cultivars and alleles specific to each cultivar have been described in details below.

5.6.1 Alleles common to all cultivars

Frantoio clones F1 and F2 and Picual clone F1 showed 100% sequence identity to the Barnea *OeSMO2-1a* allele.

Barnea clones F2, F3, F4, F5 and F6 showed high sequence identity (99.1-99.8%) to *OeSMO2-1a,* howeverkey SNPs were identified at several places along the nucleotide sequences of these clones (Figure 5.22 and 5.23) which were also seen in Frantoio and Picual clones.

Barnea clone F2 had a C→A nucleotide substitution at position 183 that would not lead to any amino acid substitution. Clone F2 had significant sequence identity (99.7%) to *OeSMO2-1a* and henceforth referred to as *OeSMO2-1b.* Frantoio clone F3 and Picual clone F2 showed same key SNPs as observed in Barnea *OeSMO2-1b* allele.

Barnea clone F3 also showed the silent C→A nucleotide substitution at position 183 as that observed in *OeSMO2-1b*. In addition, this clone also had a A→G nucleotide substitution at position 719 that would result in the substitution of tyrosine (Y) →cysteine (C) at amino acid 240. This clone had 99.8% sequence identity to *OeSMO2-1a* and henceforth has been

referred to as *OeSMO2-1c*. Frantoio clone F4 and Picual clone F3 revealed the same SNPs as observed in Barnea *OeSMO2-1c* allele.

Barnea clones F4, F5 and F6 showed a G→A nucleotide substitution at position 729 which introduced a stop codon at amino acid position 243. These three clones had 99.3-99.5% sequence identity to *OeSMO2-1a*. These clones henceforth will be refered to as *OeSMO2-1d*. Frantoio clones F5 and F6 and Picual clone F4 also contained the same SNPs as observed in Barnea *OeSMO2-1d.*

5.6.2 Alleles unique to Barnea

Barnea clones F7, F8, F9 and F10 showed higher sequence identity (97.5-100%) to *OeSMO2-2a* than to *OeSMO2-1a* (89.2-92.0%).Howeverkey SNPs were identified at several places along the nucleotide sequences of these clones (Figure 5.22 and 5.23).

Clone F7 and F8 from Barnea are identical and henceforth will be refered to as *OeSMO2-2a*. Clone F9 in Barnea was 99.1% similar to *OeSMO2-2a* however unique SNPs (T→C) were observedat positions 79 (F→L), 225 (silent) and 490 (F→L) which substituted phenylalanine(F) to hydrophobic leucine (L) at position 79 and 490, however the SNP at position 225 was a silent mutation. At position 136, a G→A nucleotide substitution would lead to the non-conservative change glycine (G) to arginine (R). At position 373, A→T nucleotide substitution would lead to the substitution of amino acid arginine (R) to hydrophobic tryptophan (W). At position 695, G→A nucleotide substitution has led to the substitution of non-polar glycine (G) with aspartic acid (D). At position 754, A→G nucleotide substitution has led to the substitution of positively charged arginine (R) with non-polar glycine (G).This clone is henceforth referred to as *OeSMO2-2b.* This allele was absent in the sequencing data of Frantoio and Picual.

Clone F10 in Barnea was 97.5% similar to *OeSMO2-2a.* Silent mutations were observed at nucleotide positions 432, 441, 456, 492, 504, 525, 553, 582, 585, 624, 645, 648, 651, 660, 669 and 684. This clone also contained the G→A nucleotide substitution (136) and A→T nucleotide substitution (373) as observed in *OeSMO2-2b.* At amino acid 190, there was an amino acid substitution from positively charged arginine (R) to non-polar glycine (G). At

amino acid 192, there was an amino acid substitution from hydrophobic leucine (L) to isoleucine (I). Two more amino acid substitutions were seen in this clone, one leading to the substitution of hydrophobic leucine to aromatic hydrophobic phenylalanine at amino acid position 203 and another from phenylalanine (F) to leucine (L) at amino acid position 205. SNPs identified at several positions (432, 441, 456, 490, 492, 504, 525, 553, 574, 582, 585, 609, 615, 624, 645, 648, 651, 660, 669, 684 and 695) in this clone in comparison to *OeSMO2-2a*, however aligns perfectly with nucleotides at those positions in the *OeSMO2-1a* allele (Figure 5.21 and 5.22). Again, with respect to sequence similarity, this clone is only 92.0% similar to *OeSMO2-1a*. This clone has been henceforth referred to as *OeSMO2-2c*. This allele was absent in the sequencing data of Frantoio and Picual.

5.6.3 Alleles unique to Frantoio

Two additional cultivar specific alleles were identified in Frantoio which were absent in the sequencing data of Barnea and Picual (Figure 5.22 and 5.23).

Frantoio clones F7 and F8 had 98.9% sequence identity to Barnea *OeSMO2-1a*.Four silent mutations were seen at position 180, 186, 381, 547, 549 and 717. These clones had a A→G nucleotide substitution at position 385 within the conserved histidine motif that lead to an amino acid substitution from isoleucine (I) to valine (V) at position 129. Another A→G nucleotide substitution was seen at position 790 that leads to an amino acid substitution from lysine (K) to glutamic acid (E). At position 817, a C→A nucleotide substitution leads to amino acid substitution from glutamine (Q) to lysine (K) at position 273. These two SNPs at position 790 and 817 align perfectly at their respective positions in the Barnea *OeSMO2-2* alleles. These clones henceforth will be referred to as *OeSMO2-1e*.

Clone F9 in Frantoio was 99.2% similar to Barnea *OeSMO2-1a*. This clone also showed the silent C→A nucleotide substitution at position 183 as observed in *OeSMO2-1b* and *1c*. A A→G point mutation in nucleotide position 280 of the clone substituted methionine (M) with valine (V). At position 361, T→C nucleotide substitution led to an amino acid substitution from phenylalanine (F) to leucine (L). At position 606, T→C nucleotide substitution was seen within the conserved histidine domain, however this mutation is silent. Another T→C nucleotide substitution was seen at position 716 which substituted amino acid valine to

alanine leading to a side chain alteration at this position. A G→A nucleotide substitution at position 729 introduced a stop codon at amino acid position 243 as observed in *OeSMO2-1d* allele. This clone is henceforth referred to as *OeSMO2-1f.*

5.6.4 Alleles unique to Picual

Two additional cultivar specific alleles were identified in Picual which were absent in the sequencing data of Barnea and Frantoio (Figure 5.22 and 5.23).

Picual clones F5, F6, F7 and F8 showed 99.3-99.7% DNA sequence identity to Barnea *OeSMO2-1a* allele however revealed two subsequent mutations at nucleotide positions 189 and 190, of which the former mutation is silent and the latter one substitutes amino acid alanine (A) to hydrophilic neutral serine (S). Clone F5 had been chosen as the representative clone and henceforth referred to as *OeSMO2-1g* (Figure 5.22 and 5.23)

Picual clones F9 and F10 showed 99.3-99.5% DNA sequence identity to Barnea *OeSMO2-1a* allele however key SNPs are identified at few positions. Three silent mutationsT→C, C→T and T→C are seen at positions 240, 270 and 714 respectively. At position 729 G→A nucleotide substitutions introduces a stop codon at amino acid position 243 as observed in Barnea *OeSMO2-1d* allele. Clone F9 had been chosen as the representative clone and henceforth referred to as *OeSMO2-1h.*

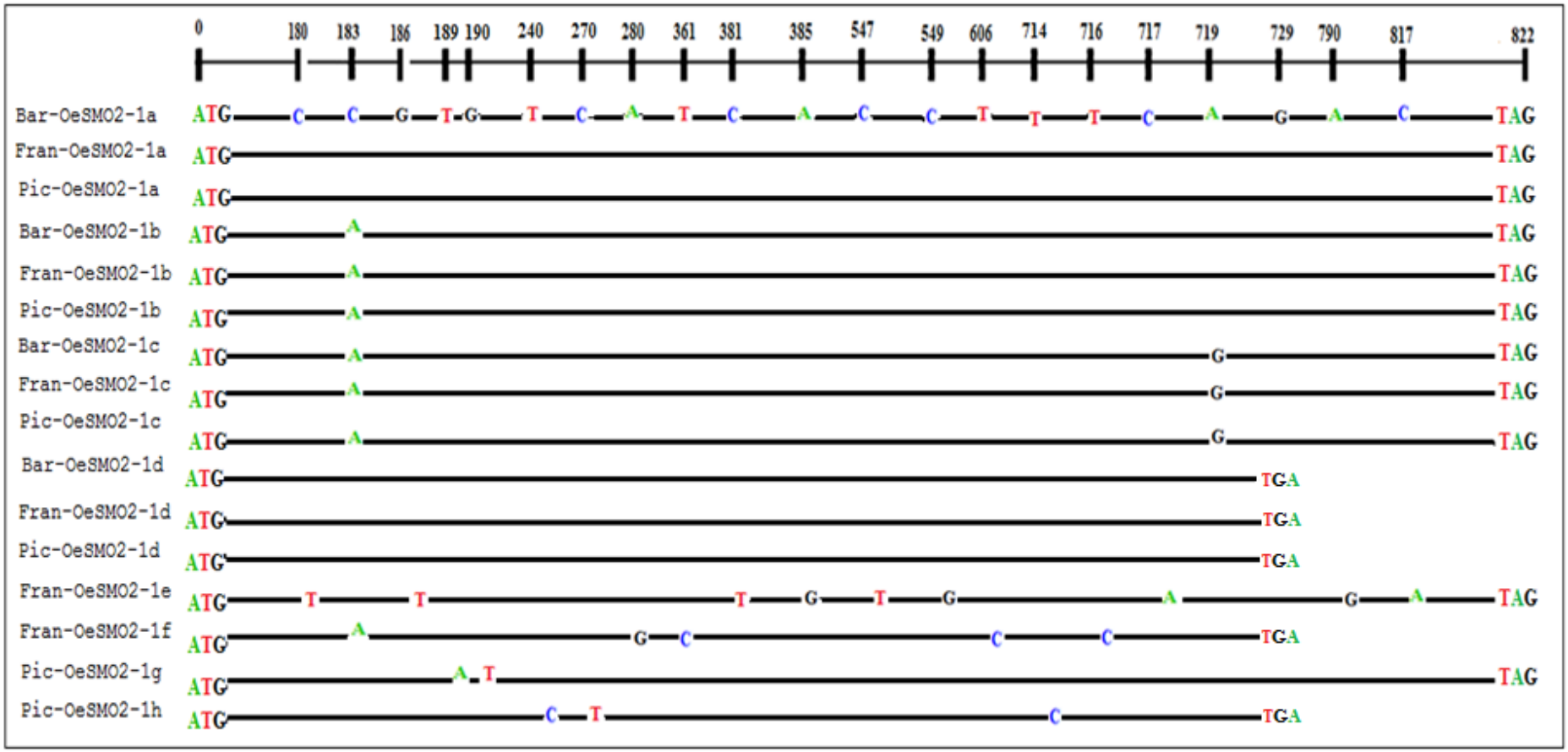


Figure 5.22 Line diagram of putative Barnea, Frantoio and PicualOe*SMO2-1* alleles highlighting identified SNPs with their respective positions in the coding sequence

Barnea Oe*SMO2-1a* has been used as a reference sequence to align all alleles. Continuous lines indicate 100% identical to reference sequence. Start codon ATG and STOP codon TAG has also been shown. Cultivars: Bar: Barnea; Fran: Frantoio, Pic: Picual.

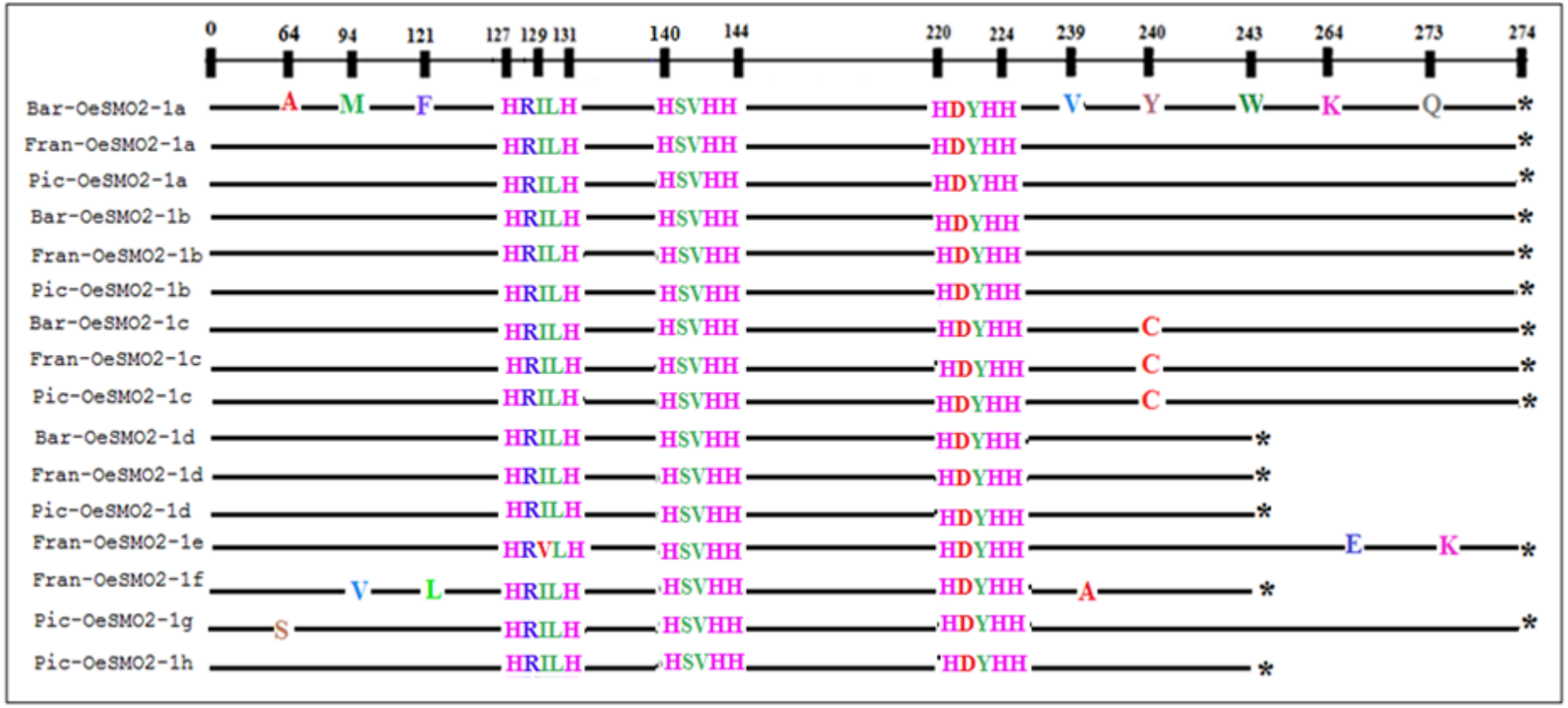


Figure 5.23 Line diagram of Barnea, Frantoio and PicualOeSMO2-1proteins

SNPs that lead to change in their amino acid sequence are shown at their respective amino acid position.Barnea Oe*SMO2-1a* has been used as a reference sequence to align all alleles. Continuous lines indicate 100% identical to reference sequence. The three rich histidine boxes encountered in SMO2 proteins are also shown. Cultivars: Bar: Barnea; Fran: Frantoio, Pic: Picual; *: STOP codon.

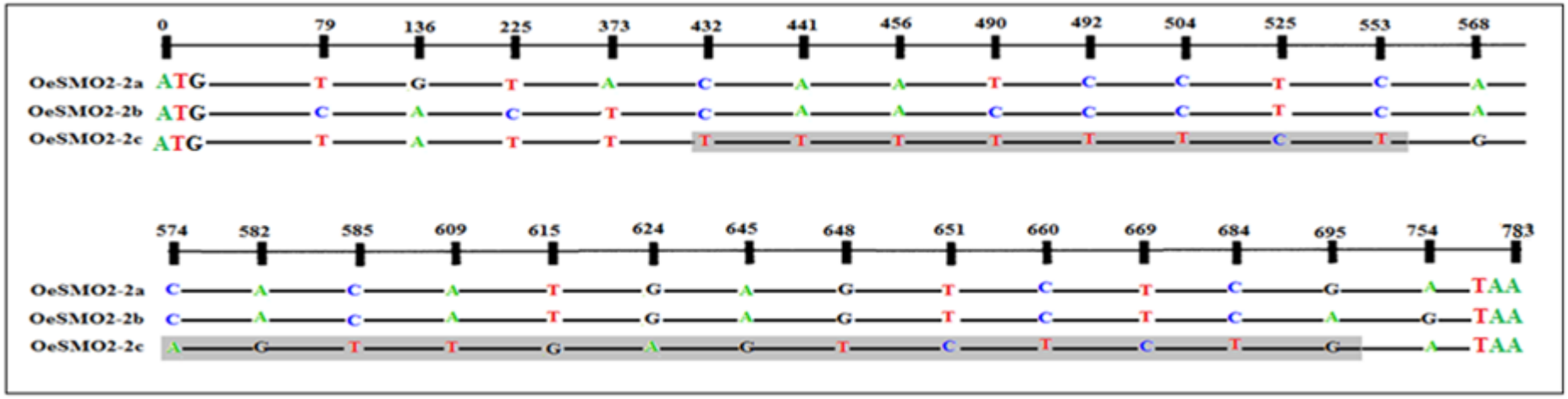


Figure 5.24 Line diagram of Barnea *SMO2-2* alleles highlighting identified SNPs with their respective position in the coding sequence
SNPs which have been identified at several positions in *OeSMO2-2c* in comparison to *OeSMO2-2a*, however aligning perfectly with nucleotides at those positions in the *OeSMO2-1a* allele are highlighted in grey.

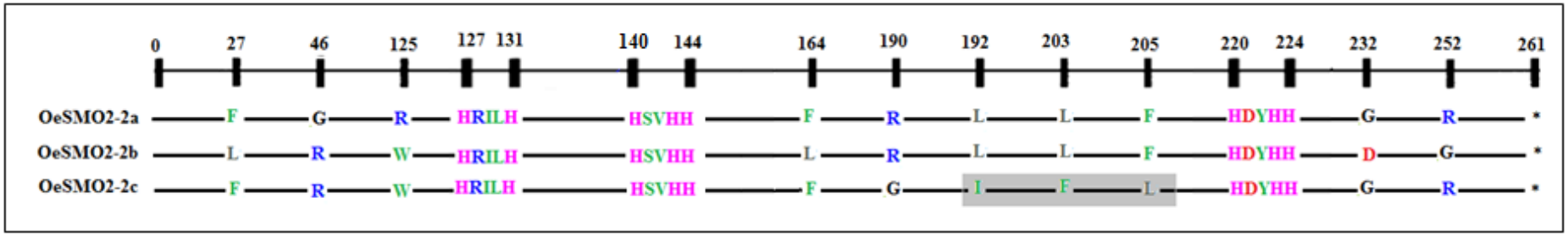


Figure 5.25 Line diagram of Barnea OeSMO2-2proteins
SNPs that lead to change in their amino acid sequence are shown at their respective amino acid position. The three rich histidine boxes encountered in SMO2 proteins are also shown. The protein region where the SNPs which have been identified at several positions in *OeSMO2-2c* in comparison to *OeSMO2-2a*, however aligning perfectly with nucleotides at those positions in the *OeSMO2-1a* allele are highlighted in grey.

5.7 TESTING FOR THE PRESENCE OF THE CULTIVAR SPECIFIC *OeSMO2* ALLELES IN BARNEA, FRANTOIO AND PICUAL

5.7.1 Allele specific PCR of the cultivar specific *OeSMO2* alleles

The sequencing of full length *SMO2* genes from Barnea, Frantoio and Picual revealed seven cultivar specific alleles namely*OeSMO2-1e, OeSMO2-1f,OeSMO2-1g, OeSMO2-1h, OeSMO2-2, OeSMO2-2a* and *OeSMO2-2b*. As the screening of clones was not exhaustive, a preliminary test for the presence of these seven alleles in all cultivars was conducted using allele specific PCR (Figure 5.26) (Section 2.8). This was conducted using primer pairs based on unique SNPs of each of the apparently cultivar-specific alleles. This revealed amplification of only the appropriate cDNA clones from the respective cultivars when optimized annealing temperature was used (data not shown).

Amplification of sections of the *SMO2* gene of Barnea, Frantoio and Picual corresponding to the *OeSMO2-1g* allele with its respective allele specific primer pair produced products similar in size (141bp) to those predicted from the *OeSMO2-1g* allele (Figure 5.26).

Allele specific primer pair corresponding to the *OeSMO2-1h* allele amplified PCR products of similar in size (494bp) to those predicted from the *OeSMO2-1h* allele only in Frantoio and Picual. At a higher annealing temperature the 494bp band disappeared in Barnea.

Amplifications of sections of the *SMO2* gene corresponding to the *OeSMO2-1e* allele with its respective allele specific primer pair produced products similar in size (401bp) to those predicted from the *OeSMO2-1e* allele in all three cultivarsBarnea, Frantoio and Picual.

Amplifications of sections of the *SMO2* gene of Barnea, Frantoio and Picual corresponding to the *OeSMO2-1f* allele with its respective allele specific primer pair produced products similar in size (127bp) to those predicted from the *OeSMO2-1f* allele.

Amplifications of *OeSMO2-2a* allele, *OeSMO2-2b* and *OeSMO2-2c* alleles with its respective allele specific primer pair produced PCR products of expected size (477bp, 319bp and 473bp respectively) in all three cultivarsBarnea, Frantoio and Picual.

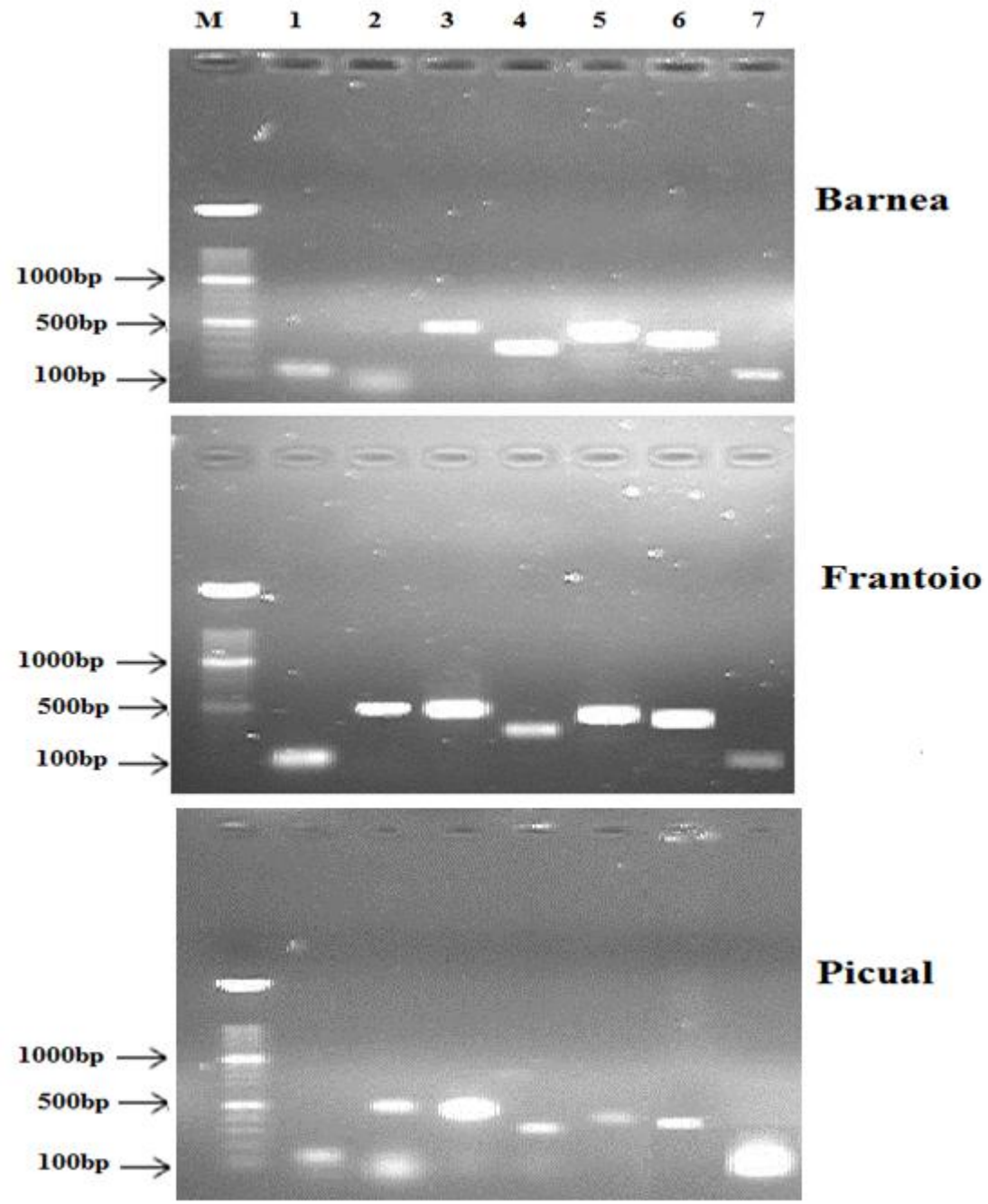


Figure 5.26 Confirming the specificity of *OeSMO2*alleles
The results of amplifying sections of the *SMO2* alleles from Barnea, Frantoio and Picual with the primers specific for the *OeSMO2-1g* (Lane 1),*OeSMO2-1h* (Lane 2), *OeSMO2-2a* (Lane 3), *OeSMO2-2b* (Lane 4),*OeSMO2-2c* (Lane 5),*OeSMO2-1e* (Lane 6) and*OeSMO2-1f* (Lane 7). M: molecular weight marker.

5.7.2 Sequencing of the *OeSMO2-2a* allele from Frantoio and Picual

The *OeSMO2-1a* and *OeSMO2-2a* coding sequencesisolated from Barneashare 89.65% similarity at nucleotide level with several key SNPs between the two distinct sequence classes. As the *OeSMO2-2a* allele were identified only from the sequencing data of Barnea and not in Frantoio and Picual, the AS-PCR products of this allele were directly sequenced (Section 2.6.6.4) to confirm their presence in these cultivars. The sequencing analysis of the *OeSMO2-2a* allele from Barnea, Frantoio and Picual revealed predicted size products (477bp) with the presence of SNPs characterisitic of the *OeSMO2-2a* in all cultivars(Figure 5.27).

5.8 DISCUSSION

In higher plants, the conversion of cycloartenol into functional phytosterols involves the removal of two methyl groups at C-4 and one methyl group at C-14 catalysed by two distinct SMOs, SMO1 and SMO2 in the sterol biosynthetic pathway.

Past research has shown that SMOactivity plays an important role in determining the relative content of the specific phytosterols in plants. As described in Section 1.9.3 evidence such as overexpression of the branch point enzyme SMT2 in tobacco mutant lacking SMO2led to the excessive accumulation of end product β-sitosterol at the expense of campesterol add weight to this hypothesis(Darnet *et al.* 2004). However, no information was available about the *SMO2* gene family, nor any allelic variation in these genes in olives. A BLASTn screen of the 454 sequencing data generated from developing olives(Alagna *et al.* 2009)revealed a 238bp sequence for the *SMO* gene, however the structure and organization of the *SMO2* gene family and any allelic variations between the olive cultivars, Barnea, Frantoio and Picual that may impact the total and relative sterol content in these olive oils have not been studied; some of these shortcomings have been addressed in the present chapter.

Towards this purposea better understanding of the gene structure of *SMO2* family in olives was required. To this end, *SMO2*cDNAs were isolated from the olive cultivar Barnea using a RACE-PCR based approach based on the partial 238bp *SMO* sequence identified from the 454 sequencing datagenerated from developing fruits. This was followed by subsequent amplification of full length *SMO2* coding sequences from Barnea and this data was used to

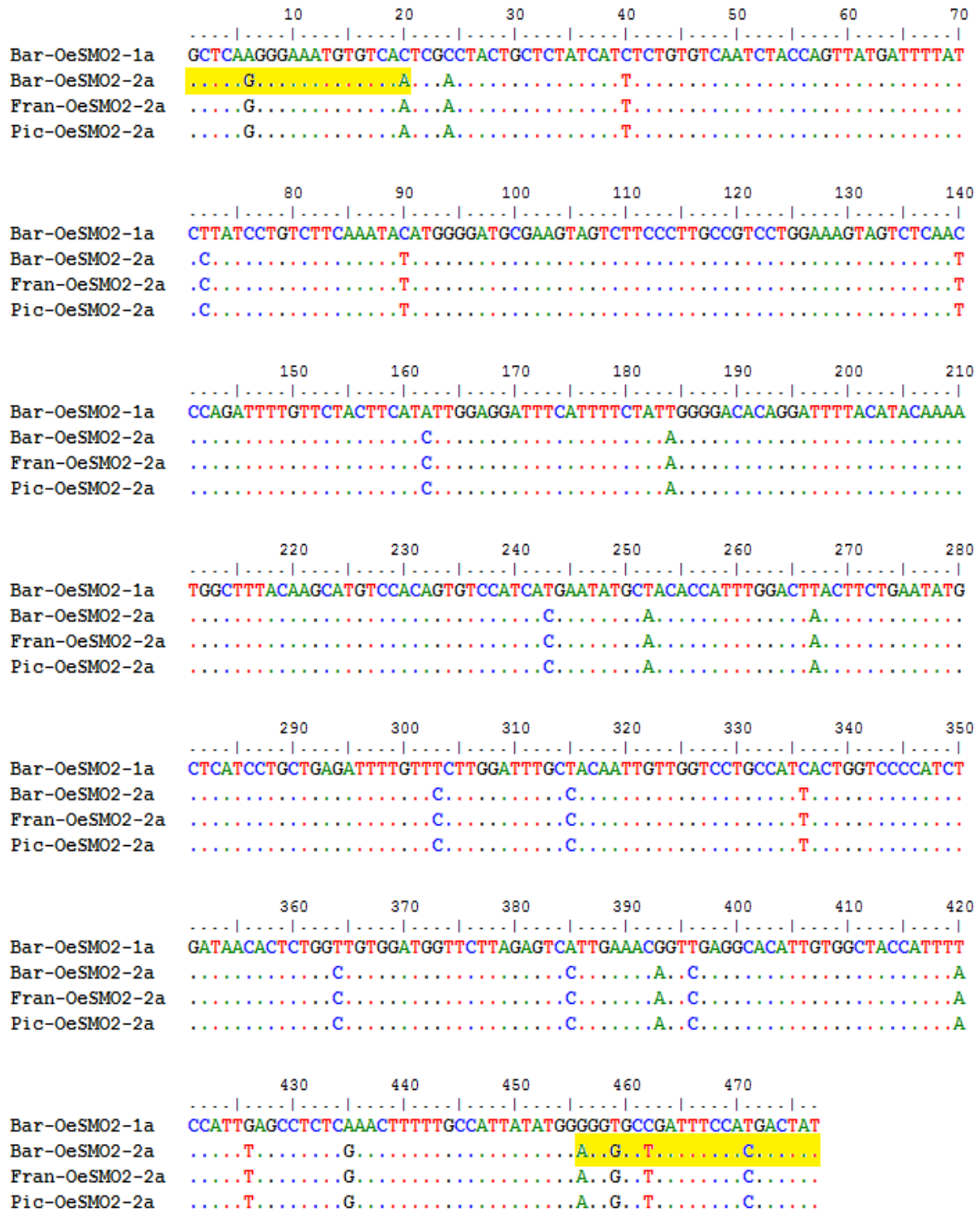


Figure 5.27Alignment of the Oe*SMO2-2a*sequences obtained from the sequencing of AS-PCR products from olive cultivars Barnea, Frantoio and Picual with the Oe*SMO2-1a*sequence deduced from cv. Barnea

Cultivars: Bar: Barnea; Fran: Fran; Pic: Picual; The numbering system relates to the nucleotide number in the sequences. The primer pair for the amplification of the AS-PCR products of the Oe*SMO2-2a* alleles is highlighted in yellow.

study the *SMO2* gene family in olives. Further to get an estimate of the expected copy number of the *SMO2* gene family in olives, genomic DNA was isolated from the leaves of olive cultivarsBarnea, Frantoio and Picualandwere digested with three different restriction enzymes *Eco*RI, *Hind*III and *Bam*H1 and Southern blotting experiment was conducted with a 671bp *SMO2* probe. Finally the full length *SMO2* genes were amplified from the other two Australian olive cultivars, Frantoio and Picual to identify any allelic differences between the cultivars and their allele specificity were assessed by AS-PCR.

5.8.1 Olive SMO2 is encoded byat least two loci

Two distinct *SMO2* cDNA sequences have been cloned from olive cultivar Barnea, designated as *OeSMO2-1*and *OeSMO2-2*. The 5' and 3' RACE sequencing data of Barnea identified two distinct sequence classes of *SMO2* clones in each with sequence similarity ranging from 94.3-97.2% and 73.3-81.2% respectively between each sequence class. The overlapping sections of the 5' and 3' RACE PCR products of the *SMO2*clones also revealed two distinct sequence classes where one sequence class showed >96% identity between each other while the second sequence class showed >98% identity between each other (Section 5.2.6). The use of this data to amplify and sequence the full length coding sequences (cds) of the *SMO2*clones from Barnea further confirmed the presence of two distinct *SMO2* cDNAs, namely *OeSMO2-1a* (Accession: KC862263)and *OeSMO2-2a* (Accession: KC862264), with91.24% identity to each other at amino acid level and 89.65% identity at nucleotide level.

The Southern data suggested that there are at least 4 loci within the Barnea genome that encode SMO2. In addition to the three distinct hybridizing fragments, the presence of an intensely hybridising fragment between 3-4.2kb in the *Eco*RI digest and 10-20kb in the *Hind*IIIdigest respectively, clouds the estimations of exact copy number of *SMO2* genes and it is plausible that additional *SMO2* genes/alleles may be present in Barnea. As olives are diploid in nature (Preedy 2010), the presence of these four hybridising fragments in the Southern blotting may represent four alleles of two*SMO2* genes.

The full length cDNA sequences of *OeSMO2-1* and *OeSMO2-2* have open reading frames of 822bp and 783bp respectively, and encode proteins of 274 and 261 residues with a predicted molecular mass of 31.8kDa and 30.8 kDa respectively. This is similar to the two

*SMO2*cDNAs cloned and isolated from *Arabidopsis* which contain ORFs of 801 and 783bp, encoding polypeptides of 267 and 261 amino acids, respectively (Darnet *et al.* 2001). According to previous reports (Bard *et al.* 1996; Darnet *et al.* 2001; Darnet *et al.* 2004; Wang *et al.* 2008; Oger *et al.* 2009), there are three conserved histidine-rich motifs in all SMO2 proteins that contribute to aniron-binding site required for the enzyme activity. The sequence alignment of OeSMO2-1a, OeSMO2-2a and other SMO2 homologuesfrom other plant species revealed that the two OeSMO2 sequences not only possess the three distinct histidine-rich regions (H^{127}RILH, H^{140}SVHH and H^{220}DYHH) but also the spacing and topology of the motifs align perfectly with the corresponding SMO sequence of AtSMO2 and other plant SMOs (Figure 5.13 and Figure 5.14).

The prediction of four possible hydrophobic transmembrane domains and a C-terminal hydrophilic regionsupports the notion that *SMO2* genes in olives are membrane bound as observed in *Saccharomyces cerevisiae*(Bard *et al.* 1996). Further, Pfam analysis suggested that *OeSMO2*belongs to the fatty acid hydroxylase superfamily which includes C-5 sterol desaturases and C-4 sterol methyl oxidases which is typical of all membrane-bound non-haem iron oxygenases (Bard *et al.* 1996; Darnet *et al.* 2001).

BLAST searches revealed high degree of sequence identity with SMO/SMO2 sequences isolated from various other plant species such as *Arabidopsis,* castor oil, tomato, cotton, tobacco, grapevine and soybean (75-85.5%) which suggests that OeSMO2-1 and OeSMO2-2 belongs to the SMO/SMO2 family.

As mentioned before, two distinct oxidation systems, SMO1 and SMO2 are involved in the removal of the first and second C4-methyls of phytosterol precursors in higher plants, where the first demethylation step is catalysed by SMO1 enzymatic complex whereas the second demethylation step is catalysed by SMO2 enzymatic complex (Section 1.9.3) (Benveniste 2004). As in this study we are interested in the SMO2 enzyme in olives (Section 5.1), the sequences of the two OeSMO2 proteins were aligned with *Arabidopsis thaliana*SMO1-1, SMO1-2 and SMO1-3 which revealed limited sequence identity of OeSMO2-1 and OeSMO2-2 proteins to *Arabidopsis thaliana*SMO1 (34.2-36.1%) strongly suggesting that these genes encode SMO2 enzymes, and not SMO1 (Figure 5.13-B).

5.8.2 Allelic differences between Barnea, Frantoio and Picual may be present

This chapter focused on characterising the *SMO2* gene family in olives and a comparison of the families in different olive cultivars. To this end, full length *SMO2* cDNAs were isolated from two additional olive cultivars, Frantoio and Picual that have contrasting phytosterol profiles(Mailer *et al.* 2008), and their sequences were compared against the Barnea *SMO2* clones to identify any allelic differences between the olive cultivars.

Comparison of the *SMO2* clones between olive cultivars Barnea, Frantoio and Picual revealed high sequence conservation between the cultivars (90.0-99.6%) (Figure IVB in Appendix IV).

The*OeSMO2-1*alleles which have been identified in Barnea, were also identified in Frantoio and Picual however *OeSMO2-2*alleles were absent in the sequencing data of Frantoio and Picual which warranted further investigation.

Apart from the two distinct *SMO2*sequence classes identified in Barnea, SNPs were identified at several places along the nucleotide sequences of the Barnea, Frantoio and Picual clones suggesting that there may be additional *OeSMO2-1* and/or *OeSMO2-2*alleles in these cultivars.

Based on their key SNPs, four *OeSMO2*-1 alleles namely*OeSMO2-1a,OeSMO2-1bOeSMO2-1c* and *OeSMO2-1d* were identified from the sequencing data of all three olive cultivars indicating these alleles are most likely present in all the cultivars.

Apart from these common alleles identified in the sequencing data of all three olive cultivars, seven cultivar specific alleles were identified which required further investigation (Figure 5.22 and Figure 5.23). Allele specific PCR was conducted under stringent conditions using primers that were specific for these alleles, to test for the presence or absence of these alleles in the other olive cultivars.

AS-PCR results suggested that *OeSMO2-1e* and *OeSMO2-1f* alleleswhich were found in Frantoio sequencing data are likely to be present in the other olive cultivars. *OeSMO2-1g* allele which was found in Picual sequencing datais likely to be present in Barnea and

Frantoio. The only allelic difference observed between the olive cultivars was that the allele*OeSMO2-1h* which was found in Picual sequencing dataappeared likely to be present in Frantoio but absent in Barnea. But none of the SNPs identified in the *OeSMO2-1h* allele lie within the conserved domains of the protein that might directly impact on its activity.

The *OeSMO2-2a, OeSMO2-2b* and *OeSMO2-2c* alleles which were found only in the Barnea sequencing data were likely to be present in Frantoio and Picual. To further confirm the presence of this second sequence class (*OeSMO2-2*) in Frantoio and Picual, the AS-PCR products of this allele were sequenced revealing this allele was present in all three olive cultivars under investigation.

Results from the allele specific PCR suggested that allelic differences may be present between Barnea and the other two cultivars. This was further supported by the Southern blotting results which identified distinct RFLPs between the olive cultivars.

The hybridising pattern of *SMO2* probe revealed the presence of four strong hybridising bandswhich taken together with the sequencing data are likely to represent the four alleles of the two distinct gene families.In addition, a 5.5kb digest which was identified in the Barnea *Eco*RI digest, was missing in Frantoio and Picual. In Picual, a slightly smaller fragment (2.8kb) was seen in place of the missing 5.5kb fragment in the *Eco*RI digest. The *Hind*III digest in the Southern blotting showed otherwise identical hybridizing pattern in Barnea and Picual, however a unique smaller fragment (2.5kb)was present in Frantoio in place of the largest clear band (~10 kb) in Picual and Barnea.In all three olive cultivars, the *Bam*HI digest produced an intensely strong hybridizing fragment ranging from 10-18kb which may be explained due to incomplete digestion of the gDNA.

In conclusion, the results from the isolation of full length *SMO2* genes from olive cultivars Barnea, Frantoio and Picual and the subsequent Southern blottingdata revealed the presence of atleast four alleles which may represent two *SMO2* genes in diploid olives. This is similar to the structure of the *SMO2* gene family in *Arabidopsis* which has been found to contain two distinct *SMO2* genes(Darnet *et al.* 2004). *SMO2*sequences from a few others plant species such as *Gossypium arboretum*, *Vitis Vinifera*, *Glycine max*,*Solanum lycopersicum* and *Ricinus communis* have been found in the Genbank/EMBL databases, however their functional characterisation has not been conducted to date (Section 1.9.3.3). Interestingly, in

this study, the Southern blotting analysis and AS-PCR results also suggestedthat allelic differences may be present between the olive cultivars. The *SMO2* allelic differences observed between the olive cultivars is interesting as this might play an important role in controlling the relative amounts of β-sitosterol and campesterol in olive oils, however none of the identified allelic differences lie within the conserved domains of the protein that might directly impact on its activity. It would be further interesting to check if any expression differences of the *SMO2* genes between the olive cultivars is seen which could have an impact on the campesterol levels of these oils.

CHAPTER 6
EXPRESSION ANALYSIS OF *SMT2* AND *SMO2* GENESUSING REAL TIME QUANTITATIVE PCR (qPCR)

List of Figures and Tables in Chapter 6

ABSTRACT

In chapter 4 and 5, the characterization of full length *SMT2* and *SMO2* cDNAs from three different olive cultivars Barnea, Frantoio and Picual has been described.In this chapterthe expression analysis of *SMT2* and *SMO2* gene familiesin the developing olive fruit has been described to identify any expression differences that may contribute to the specific sterol profiles in the Australian olive oils under investigation. However, to study the expression levels of *SMT2* and *SMO2* gene families in olives, the expression level of these target genes has to be normalised using stable internal control genes known as reference genes to derive relative changes in gene expression. To date, no published information about the identification of valid reference genes in olives is available. Therefore, a reference gene validation experiment was conducted to select stably expressed candidate reference genes. The analyses of expression stability of reference genes revealed that GAPDH, EF1-alpha and PP2A are suitable reference genes for expression analysis in developing olivemesocarp tissues and were used for normalisation against the target genes, *SMT2* and *SMO2*.The expression analysis of *SMT2* genes across the three different olive cultivars Barnea, Frantoio and Picual in the 2009 crop season showed that *SMT2* genes exhibitsimilar expression patterns in all three cultivars with no interesting variation. The expression analysis of *SMO2* genes across the three different olive cultivars Barnea, Frantoio and Picual in the 2009 crop season revealed that as lipids get synthesized later in the developmental stages of the olive fruit, *SMO2* expression in the fruit gradually increases, however the magnitude of this change in expression of *SMO2* does appear to vary between olive cultivars, in particular between Barnea and Picual. Significant difference in *SMO2* expression was observed between Barnea and Picual at 96 DAF and 116 DAF, with a robust upregulation in Picual at 136 DAF, suggesting differential expression pattern of *SMO2* in these two cultivars.

6.1 INTRODUCTION

The phytosterol levels inAustralian olive oilsderived from the cultivarBarnea (representing 41% of the olive crop in Australia) has been shown to contain up to 4.8% campesterol, which exceeds the IOOC standards for extra virgin olive oil that stipulate a campesterol level of less than 4%.The sterol content of the Barnea cultivar, on a year to year basis, has shown fluctuations in total and relative sterol levels, however campesterol levels have remained consistently high, strongly suggesting that the sterol levels are strongly influenced by genetic factors.The characterized sterol biosynthetic pathway in plants contains a bifurcation that leads to the formation of β-sitosterol or campesterol, with the flux controlled by the activity of two branch-point enzymes, SAM-24-methylene-lophenol-C-24-methyltransferase2 (SMT2) and C-4α-sterol-methy-oxidase2 (SMO2).

Experimental evidence has demonstrated that both SMT2 and SMO2 can influence campesterol and β-sitosterol levels in other plants including *Arabidopsis thaliana*and tobacco. Further, the expression analysis of *SMT2-1* and*SMT2-2* in leaf, roots and seeds and at various stages of growth of soybean showed *SMT2* was constitutively expressed in all tissues during the life history of soybean, however expression levels declined to very low levels in the later stages of seed maturation (Neelakandan *et al.* 2010). In cotton, highest expression levels were detected in 10 DPA (day post anthesis) fibers and the lowest expression levels were observed in leaves and cotyledons for both *GhSMT2-1* and *GhSMT2-2* genes (Luo *et al.* 2008). In cotton, expression profiles of the two *GhSMT2* genes in different organs and at various stages of fibre development showed that the expression level of *GhSMT2-1* was10 times higher than *GhSMT2-2* in all organs and tissues detected(Luo *et al.* 2008).Thus, it is conceivable that the relative activity or expression of these enzymes, SMT2 and SMO2 could play a pivotal role in determining the relative amounts of β-sitosterol and campesterol in Australian olive oils. Functional characterisation of these two proteins and the genes encoding them has been undertaken in several plants, fungal and bacterial species.

In chapter 4 and 5, characterization of full length *SMT2* and *SMO2* cDNAs from three different olive cultivars Barnea, Frantoio and Picual have been described.In this chapterthe expression analysis of *SMT2* and *SMO2* gene familiesin the developing olive fruit has been described to identify any expression differences that may contribute to the specific sterol

profiles in the Australian olive oils under investigation. However, to study the expression levels of *SMT2* and *SMO2* gene families in olives, the expression level of these target genes has to be normalised using stable internal control genes, known as reference genes,to derive relative gene expression levels. Few reference genes were suggested for the major tissues (leaves, fruits and pedicels) of olive (Dündar, 2013), however no firmly established information about the identification of valid reference genes in olive is available to date. Therefore, in order to normalise the expression data of target genes in olives, a reference genes validation experiment needed to be conducted to select stably expressed candidate reference genes.

In plants a number of reference gene validation attempts have been reported covering both model and crop species such as *Arabidopsis* (Czechowski *et al.* 2005), rice (Jain M *et al.* 2006), wheat (Long *et al.* 2011) grapevine (Gamm *et al.* 2011), barley (Jarošová *et al.* 2010), soybean(Hu *et al.* 2009) and cotton (Artico *et al.* 2010). Eight commonly used reference genes in plants were chosen from these previous reports for normalisation of RT-qPCR data in olives which included 60S ribosomal protein L18, protein phosphatase 2A, polypyrimidine tract-binding protein, tubulin alpha, aquaporin tonoplast intrinsic protein, ubiquitin carrier protein, glyceraldehyde 3-phosphate dehydrogenase and elongation factor 1 alpha.

The analysis of the expression stability of reference genes were compared using two different algorithms, GeNorm and BestKeeper and the identified stablemost reference genes were used for accurate normalization against the target genes, *SMT2* and *SMO2*.

6.2SELECTION OF CANDIDATE REFERENCE GENES

In this study, the expression stability of genes previously used as reference genes (Section 1.11) in relative quantification by qPCR, were tested to develop an assay for the quantification of two target genes, *SMT2* and *SMO2* genes in the developing olive fruit of thecultivars Barnea, Frantoio and Picual during the 2009 season. Eight commonly used reference genes in plants were chosen for normalisation of RT-qPCR data (Section 1.11) which included 60S ribosomal protein L18, protein phosphatase 2A, polypyrimidine tract-binding protein, tubulin alpha, aquaporin tonoplast intrinsic protein, ubiquitin carrier protein, glyceraldehyde 3-phosphate dehydrogenase and elongation factor 1 alpha.The forward and reverse primer pairs designed for the eight reference genes, along with their abbreviated names, function, accession numbers and expected amplicon length have been shown in Table 2.16. All samples used in this study, with their cultivar name, date and year of harvest, negative controls, standards and their abbreviations has been shown in Table 2.14.

6.2.1 Determination of purity of extracted RNA

Agarose gel electrophoresis images of total RNA extracted from olive fruit tissues harvested at individual timepoints in the 2009 (96, 109, 116 and 136 DAF) crop season from olive cultivars Barnea, Frantoio and Picual have previously been shown in Figure 4.1 and 4.16, revealing two distinct bands representing 18S- and 28S- RNA bands. To further verify RNA quality, the integrity of all RNA sampleswere assessed by their RNA integrity number (RIN) calculated by the Agilent 2100 Bioanalyzer (Section 2.5.3). This analysis revealed that all samples had RIN values ranging between 8.3-9.9 (Table 6.1).

BARNEA		**FRANTOIO**		**PICUAL**	
Samples	RIN values	Samples	RIN values	Samples	RIN values
B-1-09	8.5	F-1-09	9.8	P-1-09	9.9
B-2-09	8.7	F-2-09	9.3	P-2-09	9.3
B-3-09	9.1	F-3-09	9.7	P-3-09	9.1
B-4-09	9.6	F-4-09	8.3	P-4-09	8.9

Table 6.1 RNA integrity numbers (RIN) of various olive RNA samples analysed with the Agilent 2100 Bioanalyzer
Annotation of each sample has been shown in Table 2.15
RIN: RNA integrity number

6.2.2 Amplification of reference genes and target genes by PCR

PCR was conducted (Section 2.7.3) using the forward and reverse gene specific primers designed for each of the eight reference genes (Table 2.16) and two target genes (Table 2.17) with the pooled sample of Barnea RNA (Section 2.4.2) which resulted in the amplification of expected size products (Figure 6.1).

6.2.3 Verification of reference genes by sequencing

The PCR products of the eight reference genes amplified from Barnea were gel purified and sequenced directly as described in Section 2.6.6.4. The sequencing analysis revealed that all amplified fragments were between 77-96% identical to sequencesused for designing primers for the reference genes (Table 6.2) (data not shown).

Gene name	Amplicon length (bp)*	Identity %	Accession number/Cluster ID of reference sequences*
60S RBP	101	87	Olea database: Cluster ID-OLEEUCl011221
PP2A	189	79	Olea database: Cluster ID-OLEEUCl010038
PTB	156	82	Olea database: Cluster ID-OLEEUCl031691
TUBA	100	86	Olea database: Cluster ID-OLEEUCl051890
TIP2	104	77	Olea database: Cluster ID-OLEEUCl011159
UBQ	150	84	AF429430.1
GAPDH	210	96	EF506530.1
EF1-alpha	191	89	XM_002527974.1
SMO2	221	91	*OeSMO2-1a* (Accession: KC862263) *OeSMO2-2a* (Accession: KC862264)
SMT2	246	94	*OeSMT2-1a* (Accession: KC862262)

Table 6.2 Percentage sequence identity between the amplified products of the eight candidate reference genes and two target genes with their corresponding reference sequences* as derived from Genbank/RACE sequencing data

***** Primer details of each primer pair and their melting temperature has been shown in Table 2.16 and 2.17

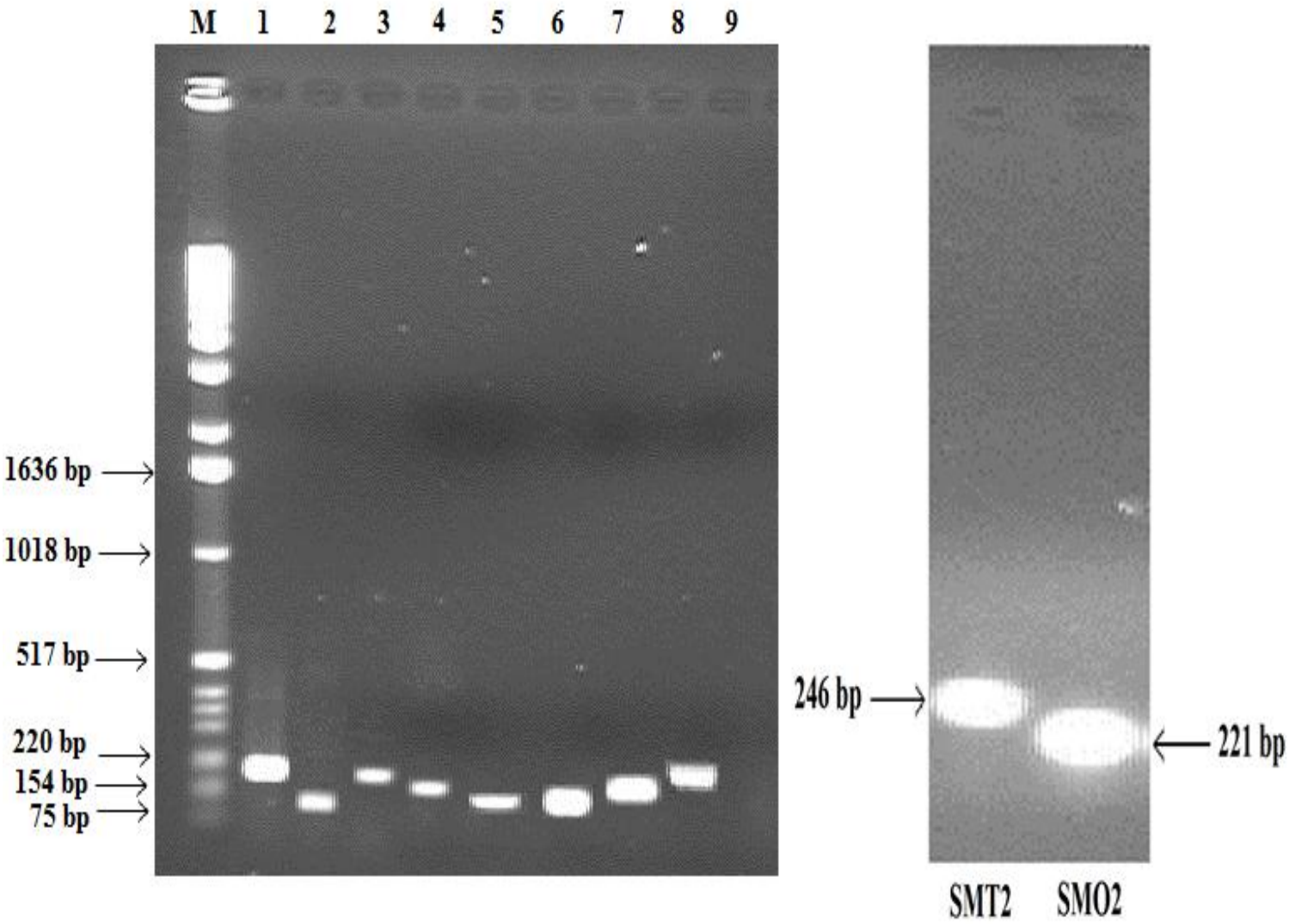


Figure 6.1 Agarose gel electrophoresis of PCR products of eight reference genes (left) and two target genes, *SMO2* and *SMT2* (right)
Lanes: M: 100bp molecular weight marker; Lane1: glyceraldehyde 3-phosphate dehydrogenase, Lane2: 60S ribosomal protein L18, Lane 3: protein phosphatase 2A, Lane 4: polypyrimidine tract-binding protein, Lane5: tubulin alpha, Lane6: aquaporin tonoplast intrinsic protein, Lane 7: ubiquitin carrier protein, Lane8: elongation factor 1 alpha, Lane9: water-only negative control.

6.2.4 Determination of PCR amplification efficiency

A calibration curve was generated for each reference and target gene using 5-fold serial dilutions of pooled cDNAas described in Section 2.10.2. The PCR efficiency (E) and standard error (SD) values characterising each standard curve are given in Table 6.3. The amplification efficiencies for the eight candidate reference genes ranged between 1.513-2.038, standard error between 0.005-0.081 and the R^2 value between 85.5-99.6%. These values for each gene were used in the analysis of their relative expression. For the target genes, *SMT2* and *SMO2*, the PCR amplification efficiency was 1.836 and 2.392, standard error of 0.020 and 0.050 and the R^2 value of 87.6% and 85.5% respectively.

6.2.5 Melting curve analysis

A melting curve analysis was performed at the end of each PCR cycle (Section 2.10.7). Gene specific amplification of all reference genes showed the presence a single peak in the melt curve analysis except TIP2 which showed the presence of 2 peaks, a large peak at 83˚C and a smaller peak at a lower temperature (~78.5˚C) (Figure 6.2). No signal was detected in the negative controls for all eight reference genes.

6.2.6 Expression levels of candidate reference genes

The eight reference genes displayed a wide expression range with Cq values ranging from 21 to 39 (Figure 6.3) (Table VA and Table VB inAppendix V). Highly expressed genes with Cq values between 21-25 cycles were EF1-alpha and UBQ. Genes with intermediate expression levels with Cq values ranging from 28-32 cycles were GAPDH, PP2A, 60S RBP and TUBA. Genes with lower expression levels with Cq values >34 cycles were PTB and TIP2.

Gene name	PCR efficiency value*	Standard error (SE)	R^2 (%)
60S RBP	1.984	0.014	97.8
PP2A	2.038	0.020	96.9
PTB	2.243	0.057	85.5
TUBA	1.704	0.069	87.0
TIP2	2.093	0.081	93.5
UBQ	1.513	0.054	89.0
GAPDH	2.111	0.018	97.1
EF1-alpha	1.908	0.005	99.6
SMO2	2.392	0.050	85.5
SMT2	1.836	0.020	87.6

Table 6.3 Details of parameters of selected candidate reference genes and target genes derived from qRT-PCR analysis
* Measure of the PCR amplification efficiency calculated from calibration curve derived from prepared standards (Section 2.10.2)

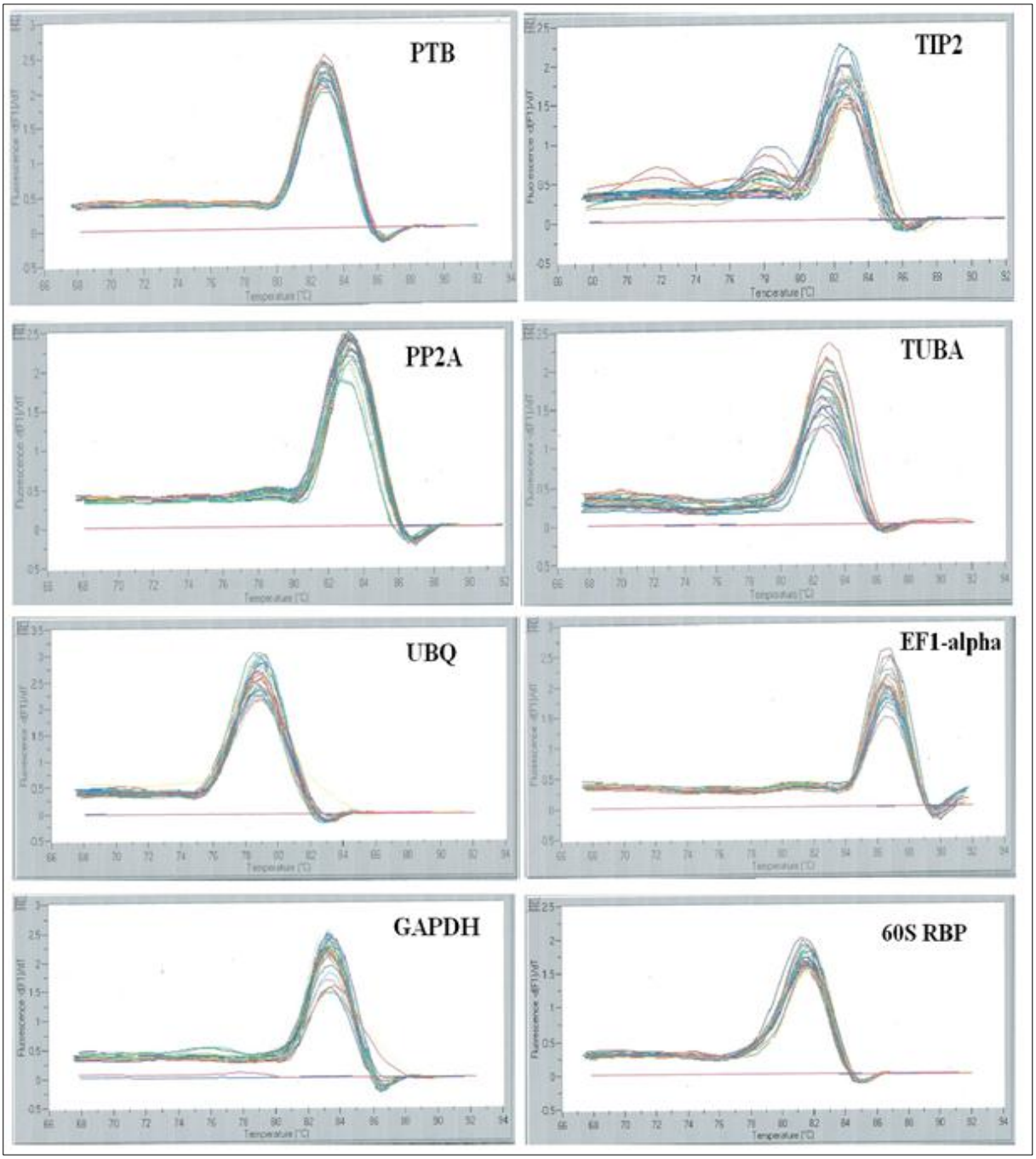


Figure 6.2 Melting curve peak of eight candidate reference genes
Flouroscence values were plotted against temperature (˚C) using the LightCycler Carousel (Roche)

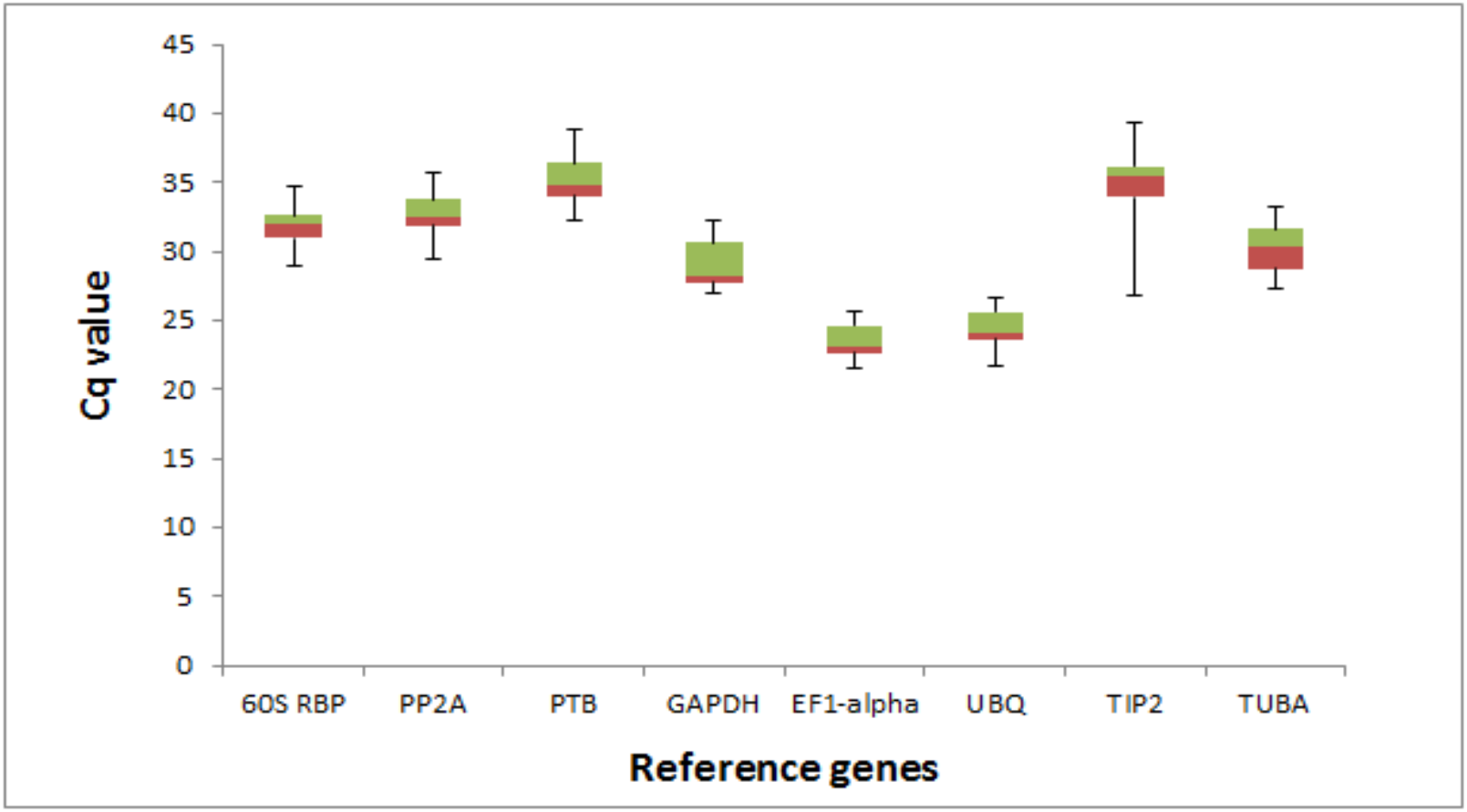


Figure 6.3 Expression levels of eight candidate reference genes
The values are given as real-time PCR quantification cycle (Cq) values in a total of 12 samples (in duplicate) from the 2009 crop season from olive cultivars, Barnea, Frantoio and Picual. The boxes represent the 25th and 75th percentiles and the line within the boxes represents the median. The whiskers indicate the range of Cq values of the data.

6.2.7 Selection of most stable reference genes

6.2.7.1 GeNorm analysis using qBase Plus

The expression stability of the eight reference genes studied varied dramatically with M values ranging from 1.85 to 0.85 (Figure 6.4A). TIP2 was the least stable gene with an M value of 1.85, whereas GAPDH was identified as the most stable gene, with an M value of 0.85. The ranking of the eight candidate reference genes according to their expression stability as calculated by the geNorm program in qBase Plus have been shown in Table 6.4. In addition to this, pairwise variation V_n/V_{n+1} between two sequential normalisation factors NF_n and NF_{n+1} was also calculated to determine the optimal number of reference genes to be used for normalisation. According to GeNorm V, V6/7 showed the lowest pairwise variation of 0.18 (Figure 6.4B) indicating that six reference genes with the lowest M values was the optimal number of reference genes which should be used for the most accurate normalisation.

	GeNorm M	
Ranking	**Genes**	**M value**
1	**GAPDH**	0.84
2	**EF1-alpha**	0.90
3	**PP2A**	0.92
4	**60S RBP**	1.06
5	**PTB**	1.15
6	**UBQ**	1.22
7	**TUBA**	1.30
8	**TIP2**	1.84

Table 6.4 Ranking of the eight candidate reference genes according to their expression stability as calculated by the geNorm program in qBase Plus

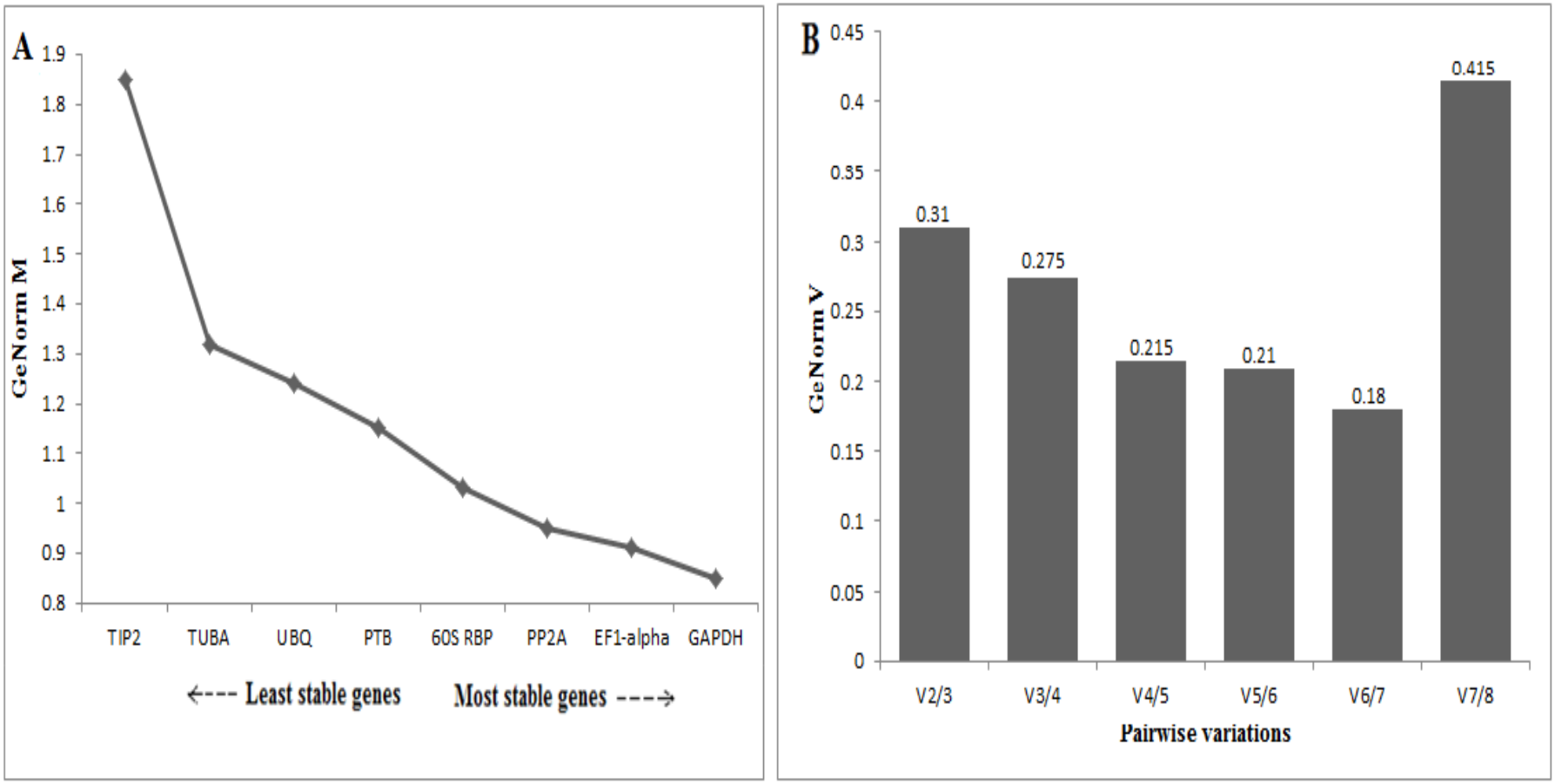


Figure 6.4Validation of candidate reference genes in olivesusing GeNorm algorithm in qBase Plus
(A) Average expression stability values (M) of the eight reference genes plotted from least stable (left) to most stable (right). The M value was calculated for each gene and the least stable gene with the highest M value was excluded from the next calculation round. (B) Pairwise variation analysis between the normalisation factors NFn and NFn+1 to determine the optimal number of reference genes to be used for normalisation against target genes.

6.2.7.2 BestKeeper analysis

BestKeeper was also used to calculate and compare the gene expression variation for the eight candidate reference genes based on the geometric mean of their Cq values and displayed as the standard deviation (SD) and coefficient of variance (CV) (Section 2.10.8.2) (Table 6.5). The variation in gene expression of three reference genes UBQ, TIP2 and TUBA was greater than two-fold (SD greater than 1) and thus considered as inconsistent and should not be used for normalisation. The other five reference genes had SD≤1 and thus are considered to be stably expressed (p≤0.05). The ranking of these five reference genes was based on their pairwise correlation with the BestKeeper index value which is indicated by the Pearson correlation coefficient (r) (Table 6.5). BestKeeper recommended PP2A as the most stable gene with a correlation coefficient of 0.805. The comparison of the five candidate reference genes with this r value resulted in a ranking as follows, from the least stable to the most stable: 60S RBP> PTB> GAPDH> EFI-alpha> PP2A. Though TIP2 has the highest correlation coefficient of 0.832, it cannot be used for normalisation because of its unacceptable SD value (SD>1).

Factors	**Reference genes**							
	60S RBP	**PP2A**	**PTB**	**GAPDH**	**EF1-alpha**	**UBQ**	**TIP2**	**TUBA**
n	24	24	24	24	24	24	24	24
GM [Cq]	31.45	32.32	34.51	28.57	23.28	24.35	34.29	30.17
AM [Cq]	31.47	32.35	34.53	28.59	23.31	24.39	34.48	30.23
Min [Cq]	28.89	29.43	32.17	26.99	21.41	21.67	26.87	27.27
Max [Cq]	34.22	35.5	37.36	30.86	25.45	26.68	39.31	33.27
SD [± Cq]	0.90	1.00	0.93	0.94	0.99	1.07	2.64	1.54
CV [% Cq]	2.87	3.08	2.69	3.29	4.26	4.41	7.66	5.11
r value	0.654	0.805	0.683	0.717	0.769	0.711	0.832	0.665
p value	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Ranking	5	1	4	3	2			

Table 6.5BestKeeper descriptive statistical analyses of eight reference genes in olivesbased on Cq values

Abbreviations: n: number of samples; GM[Cq]: geometric mean of Cq values; AM[Cq]: arithmetic mean of Cq values; Min[Cq] and Max[Cq]: extreme values of Cq; SD [±Cq]: standard deviation of Cq values; CV[% Cq]: coefficient of variance expressed as percentage of Cq values; r: coefficient of correlation, p: probability value. Genes with SD>1 have been highlighted in red box. PP2A showing the highest r value out of the five reference genes displaying acceptable SD values has been highlighted in green box.

6.3 EXPRESSION ANALYSIS OF *SMT2* GENES IN BARNEA, FRANTOIO AND PICUAL

In order to investigate the tissue specific developmental expression of *SMT2* genes, the expression levels of *SMT2*alleles were measured at different times (96, 109, 116 and 136 DAF) during olive fruit development and ripening in the mesocarp tissues from three olive cultivars, Barnea, Frantoio and Picual during the 2009 season using qPCR. As mentioned in Chapter 2 (Section 2.10.5), a single primer pair was designed based on consensus sequences between all of the identified *SMT2* alleles, therefore the expression analysis of all the *SMT2* alleles were measured in different samples together. The raw Cq values of *SMT2* genes across all samples and its amplification plot are shown in Table 6.6 and Figure 6.5. The results of the analysed data (after normalization with the most stable reference genes, discussed below) accompanying their p-value, fold change ratio and range of 95% confidence interval has been shown in Appendix VG and VH.

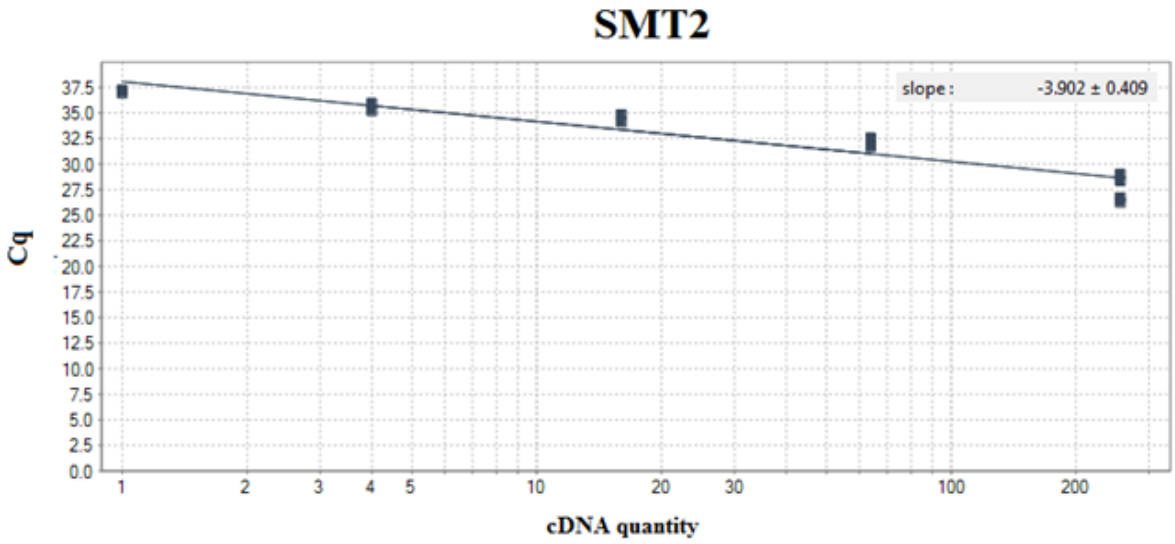


Figure 6.5 Representative efficiency curvesfor *SMT2* gene
Mean Cq values were plotted against the five four-fold cDNA serial dilutions (Section 2.10.3) using the qBase Plus software.
Slope obtained for each plot has been shown top right.

SMT2			
Samples	**Cq value**	**Average Cq**	**SD**
B-1-09	25.63	25.445	0.26
Repli. of B-1-09	25.26		
B-2-09	26.11	26.45	0.48
Repli. of B-2-09	26.79		
B-3-09	23.58	23.66	0.11
Repli. of B-3-09	23.74		
B-4-09	25.71	25.42	0.41
Repli. of B-4-09	25.13		
F-1-09	24.53	24.41	0.16
Repli. of F-1-09	24.29		
F-2-09	25.63	25.385	0.34
Repli. of F-2-09	25.14		
F-3-09	26.89	26.87	0.02
Repli. of F-3-09	26.85		
F-4-09	26.1	25.825	0.38
Repli. of F-4-09	25.55		
P-1-09	24.59	24.635	0.06
Repli. of P-1-09	24.68		
P-2-09	25.63	25.805	0.24
Repli. of P-2-09	25.98		
P-3-09	27.26	27.195	0.09
Repli. of P-3-09	27.13		
P-4-09	26	26.02	0.028
Repli. of P-4-09	26.04		

Table 6.6 The transcription profiles of target genes *SMT2* given as Cq values across all samples in Barnea, Frantoio and Picual

Average Cq values with the standard deviation (SD) for all samples shown.
Repli: Replicate
Annotation for each sample with their name of olive cultivar, timepoint and year has been given in Table 2.15.

6.3.1 Analysis of melting curve peak of *SMT2*in Barnea,Frantoio and Picual

The sequence of the *SMT2* probe used for qPCR analysis was analysed to determine any expected melting temperature differences between the *SMT2* amplicons. No melting temperature differences were expected from the four identified *SMT2* alleles (data not shown).

Amplification of *SMT2* in all samples resulted in 2-3 melting peaks at approximately 82˚C (Figure 6.6). In order to detect any in differences between the melting profiles of the *SMT2* amplicons between cultivars, the melting peaks of *SMT2* were plotted separately for the three olive cultivars Barnea, Frantoio and Picual (Figure 6.6 and Fgure 6.7).

Though only differing by 0.2-0.3˚C, melt curve data showed the presence of at least two melting peaks in Frantoio and Picual with Tm of 81.76-81.80˚C and 82.28-82.36˚C respectively. In Barnea, *SMT2* amplification showed the presence of three melting peaks with Tm of 81.76˚C, 82.20˚C and 82.68˚C.

6.3.2Relative *SMT2* expression during fruit development in Barnea

The real time expression profiling of *SMT2* gene during the developmental stages of Barnea fruit showed that *SMT2* is expressed at relatively low levels during the initial fruit developmental stage (96 DAF) with approximately a 9.3 fold increase at 109 DAF (p=0.068), 6 fold increase at 116 DAF (p=0.04) and 6.7 fold increase at 136 DAF (p=0.074) (Figure 6.8).

6.3.3Relative *SMT2* expression during fruit development in Frantoio

The real time expression profiling of *SMT2* gene during the developmental stages of Frantoio fruit showed that in comparison to 96 DAF, there is a slight (1.9 fold) decrease in expression at 116 DAF (p=0.02) and a slight (1.4 fold) increase in expression at 136 DAF (p=0.134). A significant increase (2.7 fold) in *SMT2* expression is observed between 116 DAF and 136 DAF (Figure 6.9).

6.3.4Relative *SMT2* expression during fruit development in Picual

In Picual, a significant 3.4 fold (p=0.03) and 3.6 fold (p=0.01) decrease in *SMT2* expression was observed at 116 DAF from 96 DAF and 109 DAF respectively (Figure 6.10). *SMT2* transcript levels reach a maximum at a later ripening stage of the Picual fruit (136 DAF) with approximately 2.5 fold (p=0.0006), 2.4 fold (p=0.03) and 8.6 fold (p=0.02) increase as compared to 96 DAF, 106 DAF and 116 DAF respectively.

6.3.5 Comparison of *SMT2* expression between olive cultivars Barnea, Frantoio and Picual

At 116 DAF, there was a significant decline in *SMT2* expression in Frantoio and Picual [2.6 fold (p=0.01) and 4.8 fold (p=0.01)] respectively] compared to Barnea (Figure 6.11). At 109 DAF, there was a 2.4 fold and 2.1 fold increase in *SMT2* expression in Barnea as compared to Frantoio and Picual respectively, however this difference was not statistically significant (Table VH in AppendixV). No such significant differences in *SMT2* expression were observed between the cultivars at 96 DAF and at 136 DAF.

Combinations of all four timepoints were also used to compare the overall expression of *SMT2* between Barnea, Frantoio and Picual using a one-way ANOVA test (Table VG in AppendixV). However the analysis did not show any statistically significant differences between the cultivars.

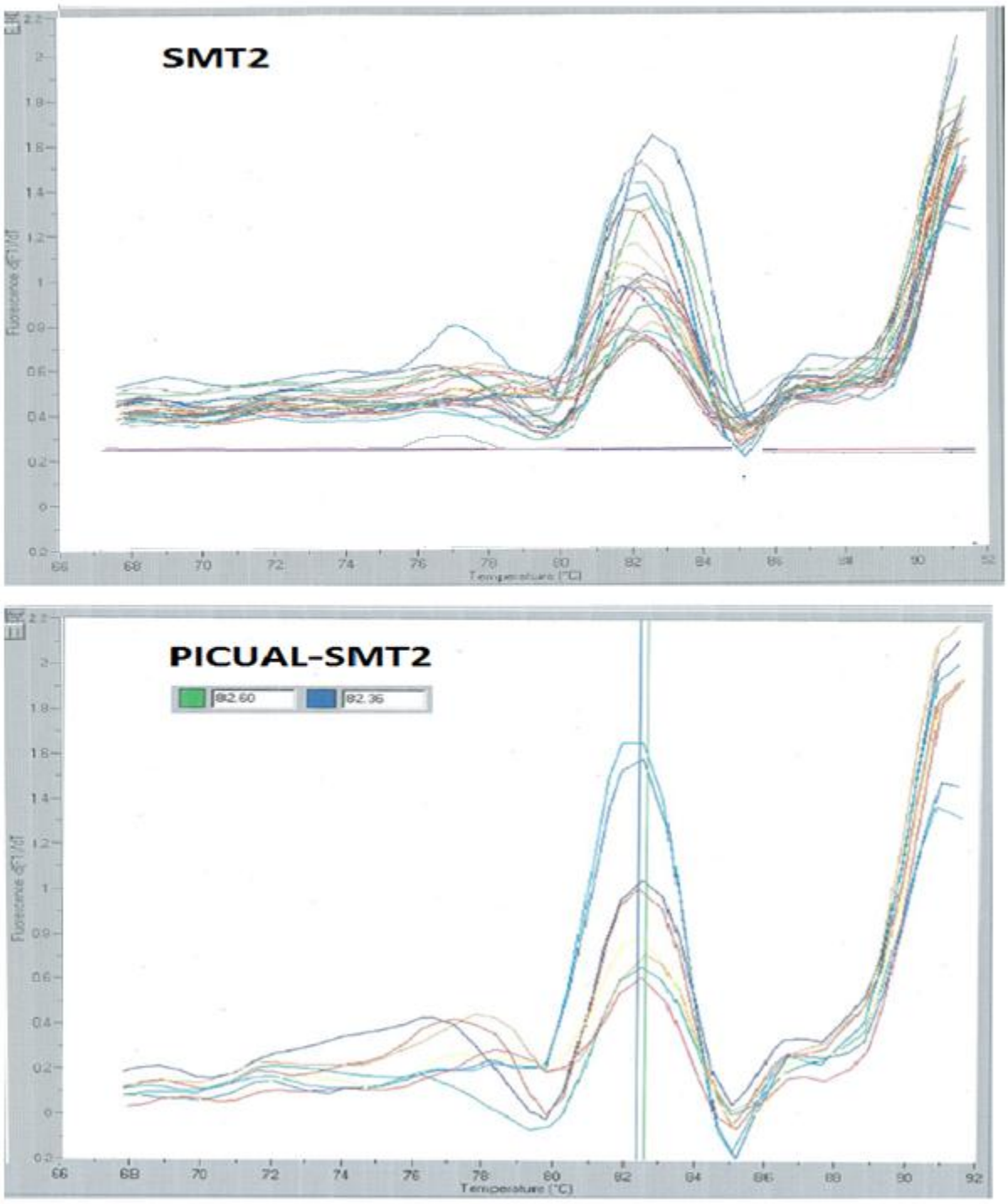


Figure 6.6 Melt curve peaks of *SMT2* in all samples and in Picual samples only
Fluorescence values were plotted against temperature (˚C) using the LightCycler Carousel (Roche)

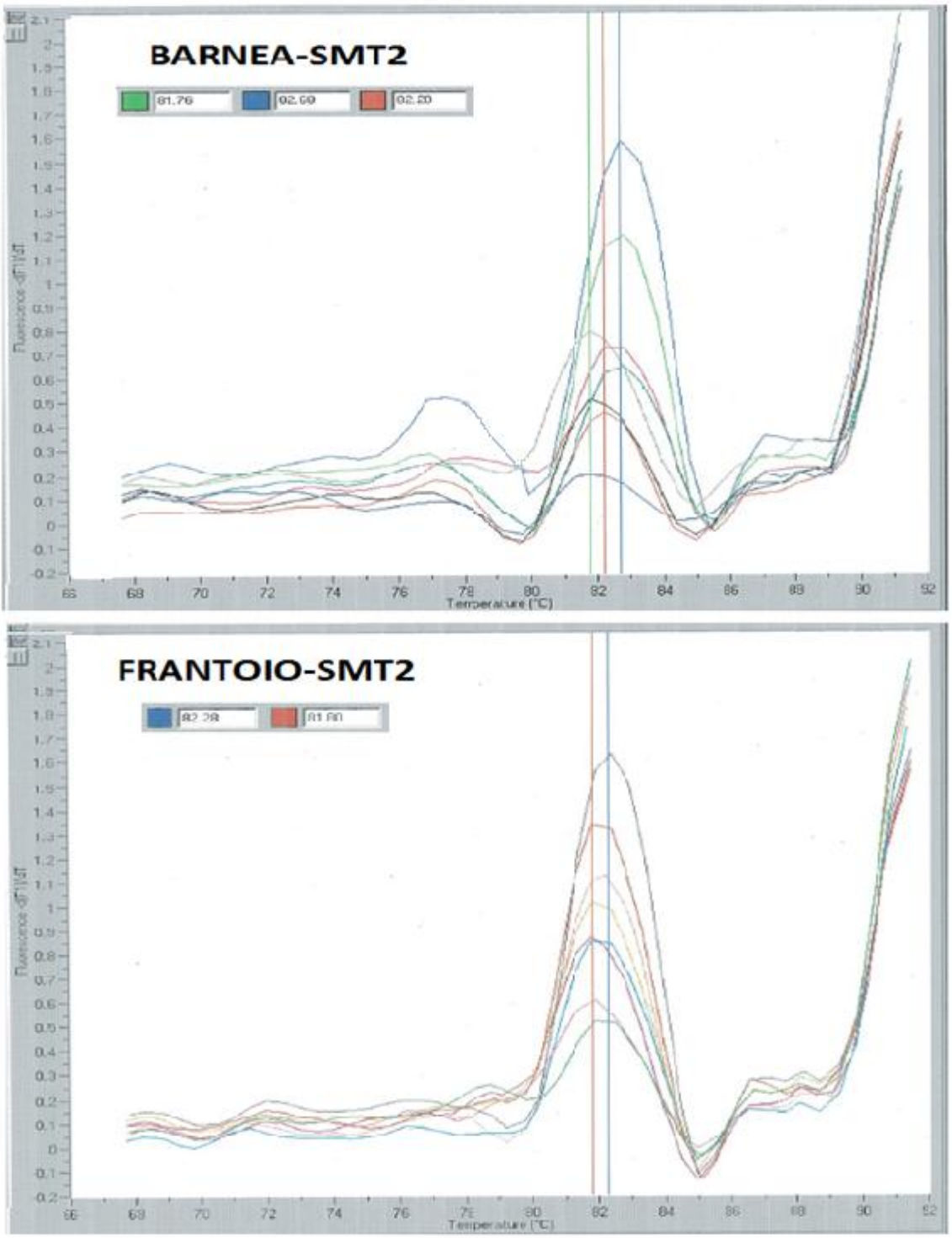


Figure 6.7 Melt curve peaks of *SMT2* in Barnea and Frantoio samples
Fluorescence values were plotted against temperature (˚C) using the LightCycler Carousel (Roche)

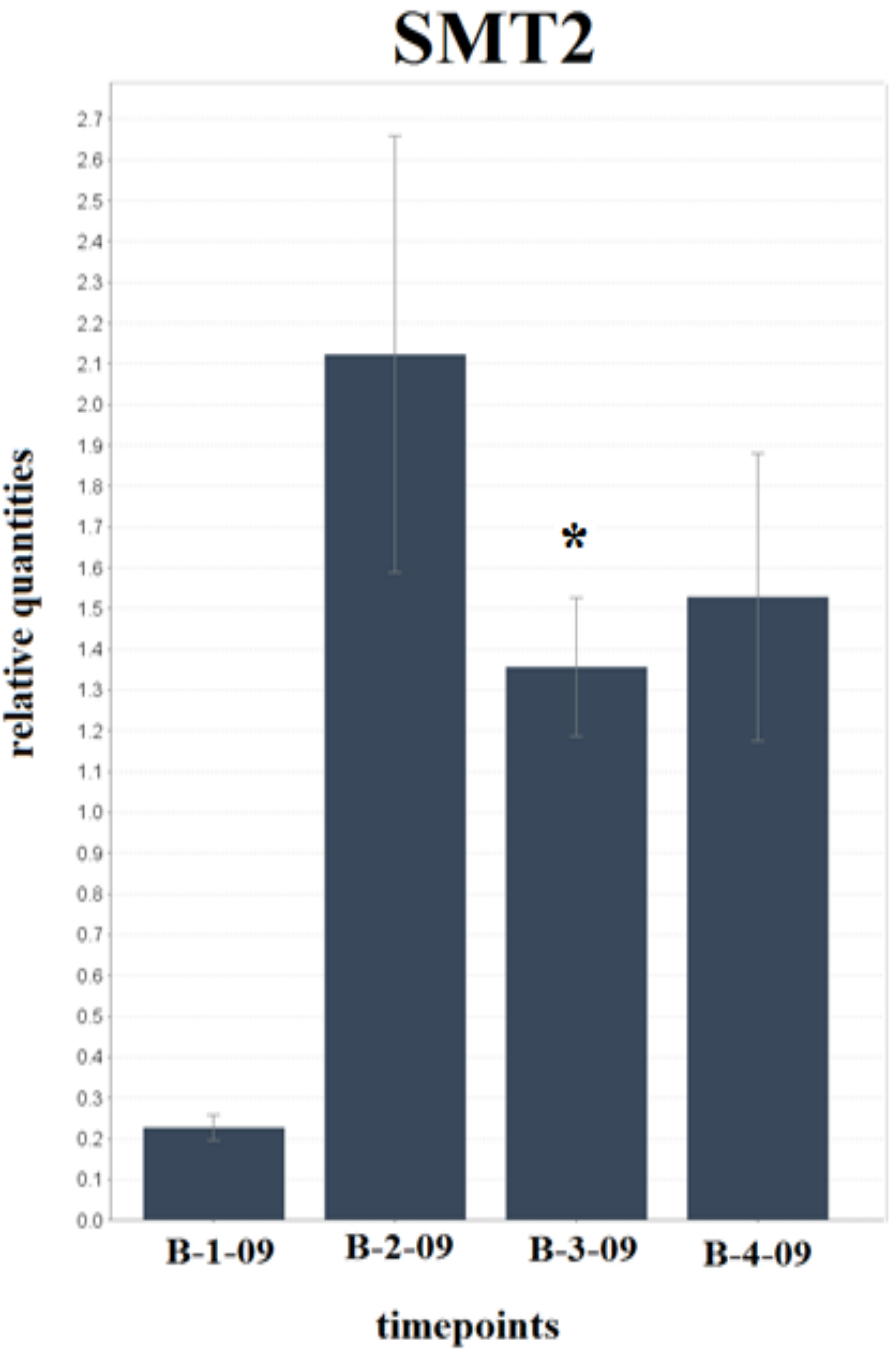


Figure 6.8 Relative expression levels of genes coding for SMT2 enzymes during fruit development in Barnea
Normalization was performed using the most stable combination of reference genes (Section 6.1.6.3). Each value is the mean± standard deviation of two independent measurements. The asterisk indicates significant difference at 95% level with p-value ≤0.05between B-1-09/B-3-09. B-1-09: 96 DAF; B-2-09: 109DAF; B-3-09: 116 DAF; B-4-09: 136 DAF.

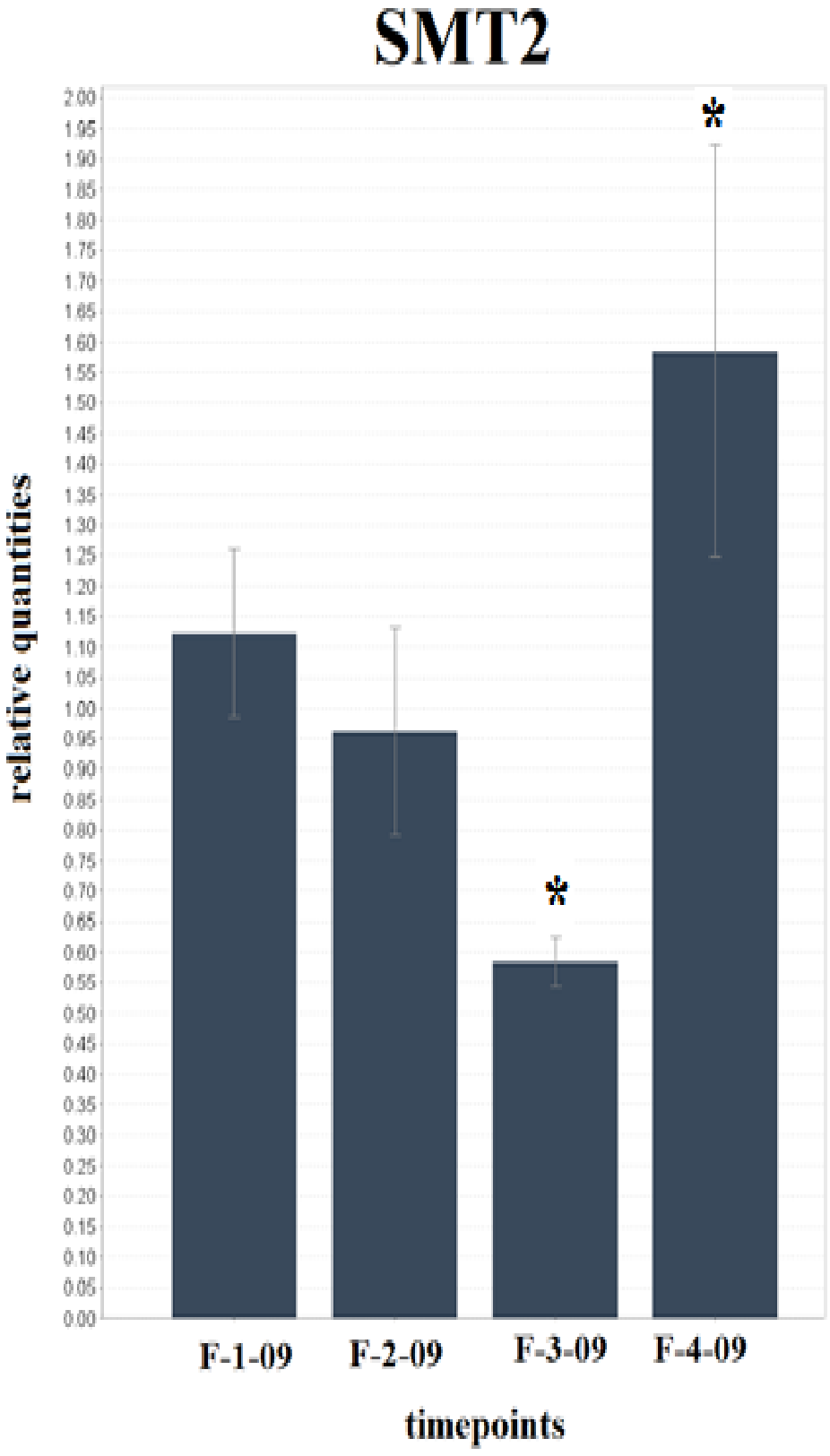


Figure 6.9 Relative expression levels of genes coding for SMT2 enzymes during fruit development in Frantoio Normalization was performed using the best combination of reference genes selected (Section 6.1.6.3). Each value is the mean± standard deviation of two independent measurements. The asterisk indicates significant difference at 95% level with p-value ≤0.05 between F-1-09/F-3-09 and F-3-09/F-4-09. F-1-09: 96 DAF; F-2-09: 109DAF; F-3-09: 116 DAF; F-4-09: 136 DAF.

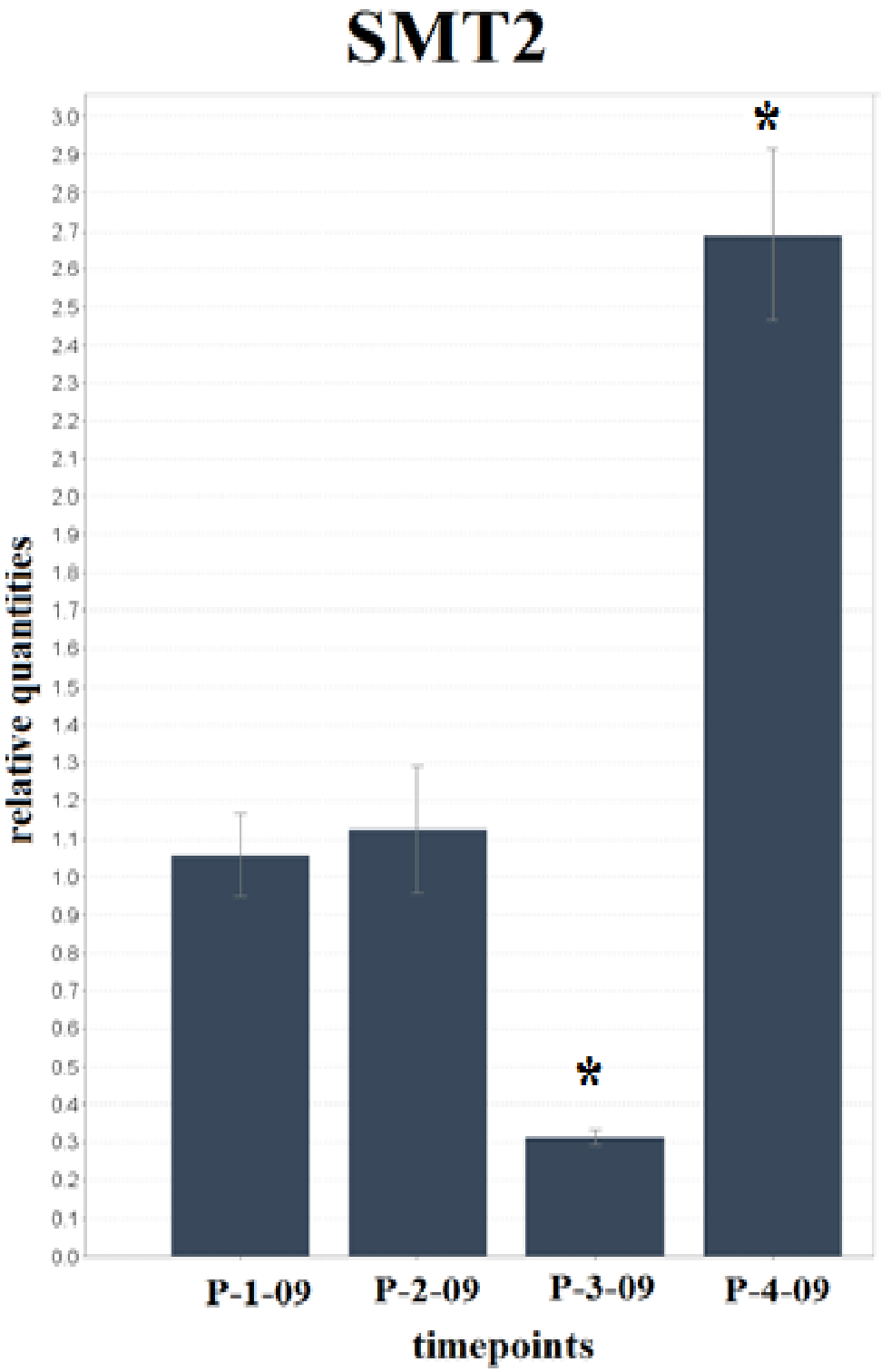


Figure 6.10 Relative expression levels of genes coding for SMT2 enzymes during fruit development in Picual
Normalization was performed using the best combination of reference genes selected (Section 6.1.6.3). Each value is the mean± standard deviation of two independent measurements. The asterisk indicates significant difference at 95% level with p-value ≤0.05 between P-1-09/P-3-09, P-1-09/P-4-09, P-2-09/P-3-09, P-2-09/P-4-09, P-3-09/P-4-09. P-1-09: 96 DAF; P-2-09: 109DAF; P-3-09: 116 DAF; P-4-09: 136 DAF.

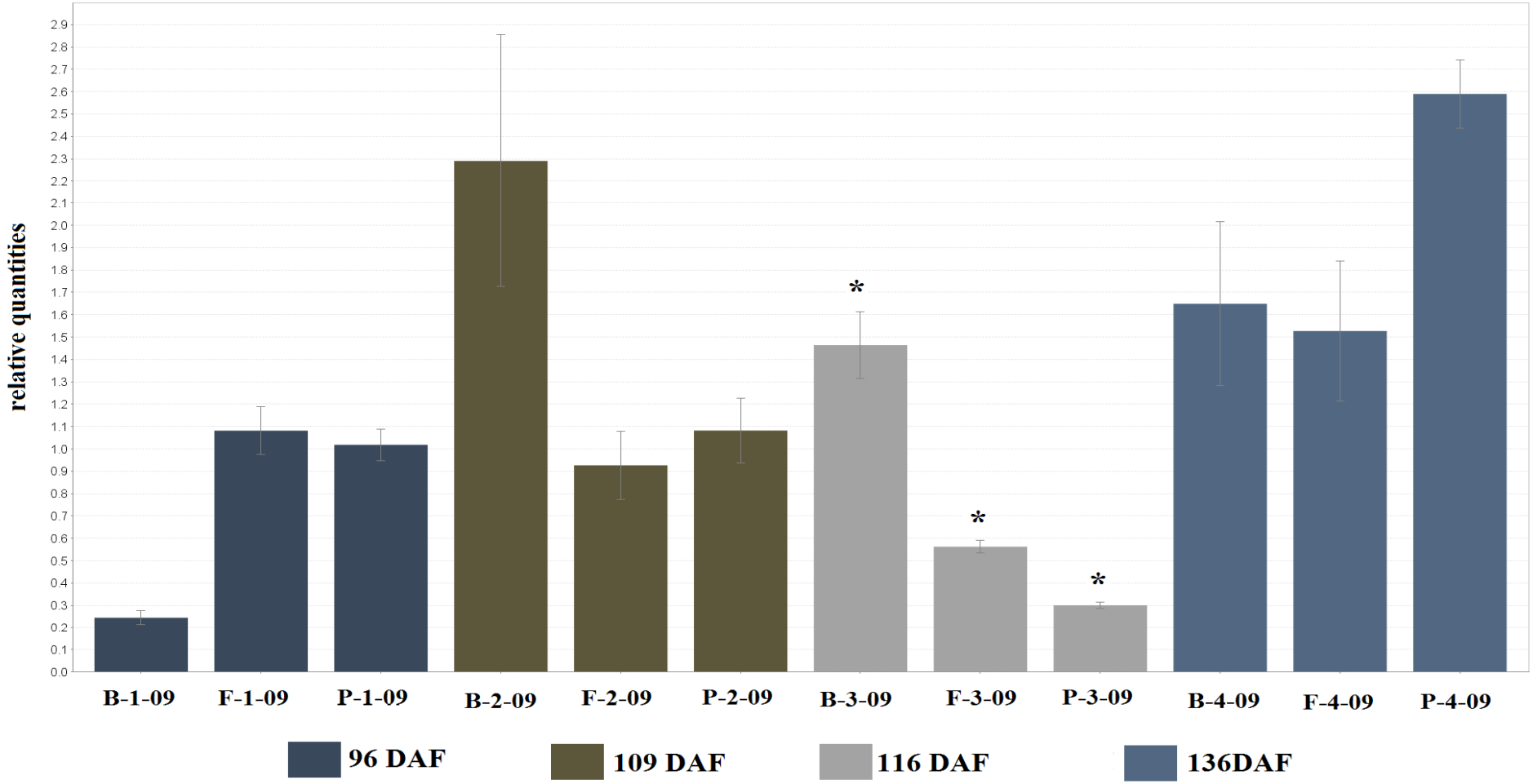


Figure 6.11 Relative expression levels of genes coding for SMT2 enzymes across Barnea, Frantoio and Picual in different timepoints
Normalization was performed using the best combination of reference genes selected (Section 6.1.6.3). Each value is the mean± standard deviation of two independent measurements. The asterisk indicates significant difference at 95% level with p-value ≤0.05 between B-3-09/F-3-09/P-3-09.DAF: days after flowering. B: Barnea; F: Frantoio; P: Picual; 1:96 DAF; 2: 109 DAF; 3: 116 DAF; 4: 136 DAF; 09: 2009 crop season.

6.4 EXPRESSION ANALYSIS OF *SMO2* GENES IN BARNEA, FRANTOIO AND PICUAL

In similar manner to that undertaken in the analysis of *SMT2*, the expression levels of *SMO2*genes were measured at different developmental periods (96, 109, 116 and 136 DAF) during olive fruit development in the mesocarp tissues from three olive cultivars, Barnea, Frantoio and Picual during the crop season 2009. A single primer pair was designed based on consensus sequences between all of the identified *SMO2* alleles in the three cultivars (Section 2.10.5), and therefore the expression pattern of all the *SMO2* alleles were measured in different olive samples as a whole. The raw Cq values of *SMO2* genes across all samples and its amplification plot are shown in Table 6.7 and Figure 6.12. The results of the analysed data (after normalization with stable most reference genes) accompanying their p-value, fold change ratio and range of 95% confidence interval has been shown in Appendix VE and VF.

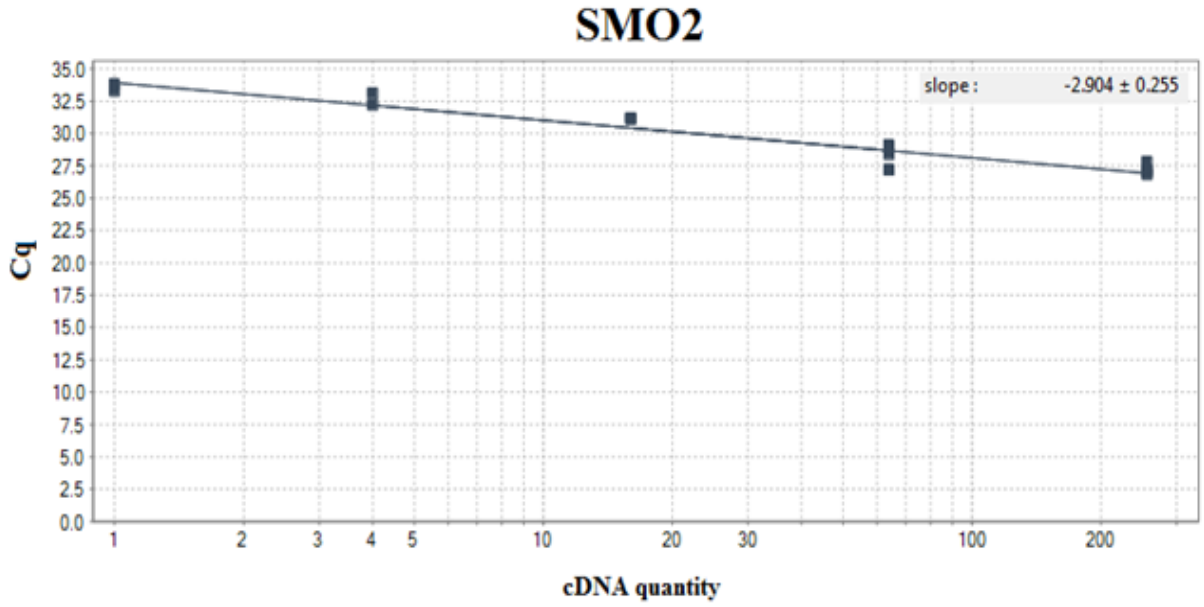


Figure6.12 Representative efficiency curvesfor *SMO2* gene
Mean Cq values were plotted against the five four-fold cDNA serial dilutions (Section 2.10.3) using the qBase Plus software.
Slope obtained for each plot has been shown top right.

SMO2			
Samples	**Cq value**	**Average Cq**	**SD**
B-1-09	26.58	26.54	0.05
Repli. of B-1-09	26.5		
B-2-09	26.79	26.945	0.21
Repli. of B-2-09	27.1		
B-3-09	26.91	26.925	0.02
Repli. of B-3-09	26.94		
B-4-09	25.41	25.525	0.16
Repli. of B-4-09	25.64		
F-1-09	26.88	26.825	0.07
Repli. of F-1-09	26.77		
F-2-09	26.43	26.43	0
Repli. of F-2-09	26.43		
F-3-09	26.79	26.615	0.24
Repli. of F-3-09	26.44		
F-4-09	26.77	26.835	0.09
Repli. of F-4-09	26.9		
P-1-09	26.21	26.335	0.17
Repli. of P-1-09	26.46		
P-2-09	26.8	26.765	0.04
Repli. of P-2-09	26.73		
P-3-09	26.52	26.585	0.09
Repli. of P-3-09	26.65		
P-4-09	24.55	24.48	0.09
Repli. of P-4-09	24.41		

Table 6.7 The transcription profiles of target gene*SMO2* given as Cq values across all samples in Barnea, Frantoio and Picual

Average Cq values with the standard deviation (SD) for all samples shown.
Repli: Replicate
Annotation for each sample with their name of olive cultivar, timepoint and year has been given in Table 2.15.

6.4.1 Analysis of melting curve peak of *SMO2*in Barnea,Frantoio and Picual

The sequence of the *SMO2* probe used for qPCR analysis was analysed to determine any expected melting temperature differences between the *SMO2* amplicons. The two distinct *SMO2* sequences identified in Barnea, *OeSMO2-1a* and *OeSMO2-1b* were expected to have two slightly different melting temperatures of 79˚C and 80˚C with a GC% of 43% and 42% respectively, however no melting temperature differences were expected from the other identified *SMO2* alleles (data not shown). Amplification of *SMO2* in all samples resulted in a broad melting peak at approximately 82˚C (Figure 6.13 and Figure 6.14). In order to detect any differences between the melting profiles of the *SMO2* amplicons between cultivars, the melting peaks of *SMO2* were plotted separately for the three olive cultivars Barnea, Frantoio and Picual (Figure 6.13 and Figure 6.14).

Melting curve data showed the presence of at least two melting peaks in Barnea, Frantoio and Picual with Tm of 81.97-82.05˚C and 82.21-82.40˚C respectively.

6.4.2Relative *SMO2* expression during fruit development inBarnea

The real time expression profiling of *SMO2* gene during the developmental stages of Barnea fruit showed that *SMO2* is expressed at low levels during the initial fruit developmental stage (96 DAF) with approximately a 12 fold increase at 109 DAF, though not significant (p=0.06) (Figure 6.15). In comparison to 96 DAF and 116 DAF, there is a significant upregulation [14.8 fold (p=0.0018) and 9.6 fold (p=0.02) respectively] of *SMO2* at later stages of fruit development (136 DAF).

6.4.3 Relative *SMO2* expression during fruit development inFrantoio

The real time expression profiling of *SMO2* gene during the developmental stages of Frantoio fruit showed that as compared to 96 DAF, there is a gradual increase in *SMO2* expression at later timepoints with 2 fold increase at 109 DAF, 2.6 fold at 116 DAF and 3.2 fold at 136 DAF, however none of the changes in expression are significant (p>0.05) (Figure 6.16).

6.4.4 Relative *SMO2* expression during fruit development inPicual

In Picual, no significant differences in *SMO2* expression were observed at the first three timepoints (96DAF, 109 DAF and 136 DAF) (Figure 6.17). However at 136 DAF, there was significant upregulation of *SMO2,* with a 25 fold, 16 fold and 22 fold increase as compared to 96 DAF, 109 DAF and 116 DAF respectively.

6.4.5 Comparison of *SMO2* expression between olive cultivars Barnea, Frantoio and Picual

SMO2 expression was relatively constant at 96 DAF, 109 DAF and 116 DAF in all three olive cultivars, with a slight (3.0 fold and 2.1 fold) increase in expression in Picual as compared to Barnea at 96 DAF ($p=0.02$) and 116 DAF ($p=0.01$) respectively (Figure 6.18).

At the fruit maturation stage (136 DAF), a significant difference in *SMO2* expression was observed between the three cultivars with Picual overexpressing *SMO2* by 5 fold as compared to Barnea ($p=0.01$) and by 12.3 fold as compared to Frantoio ($p=0.02$) (Table VF in Appendix V).

Combinations of all four timepoints were also used to compare the overall expression of *SMO2* between Barnea, Frantoio and Picual using the one-way ANOVA test (Table VE in Appendix V). However the analysis did not show any statistically significant differences between the cultivars.

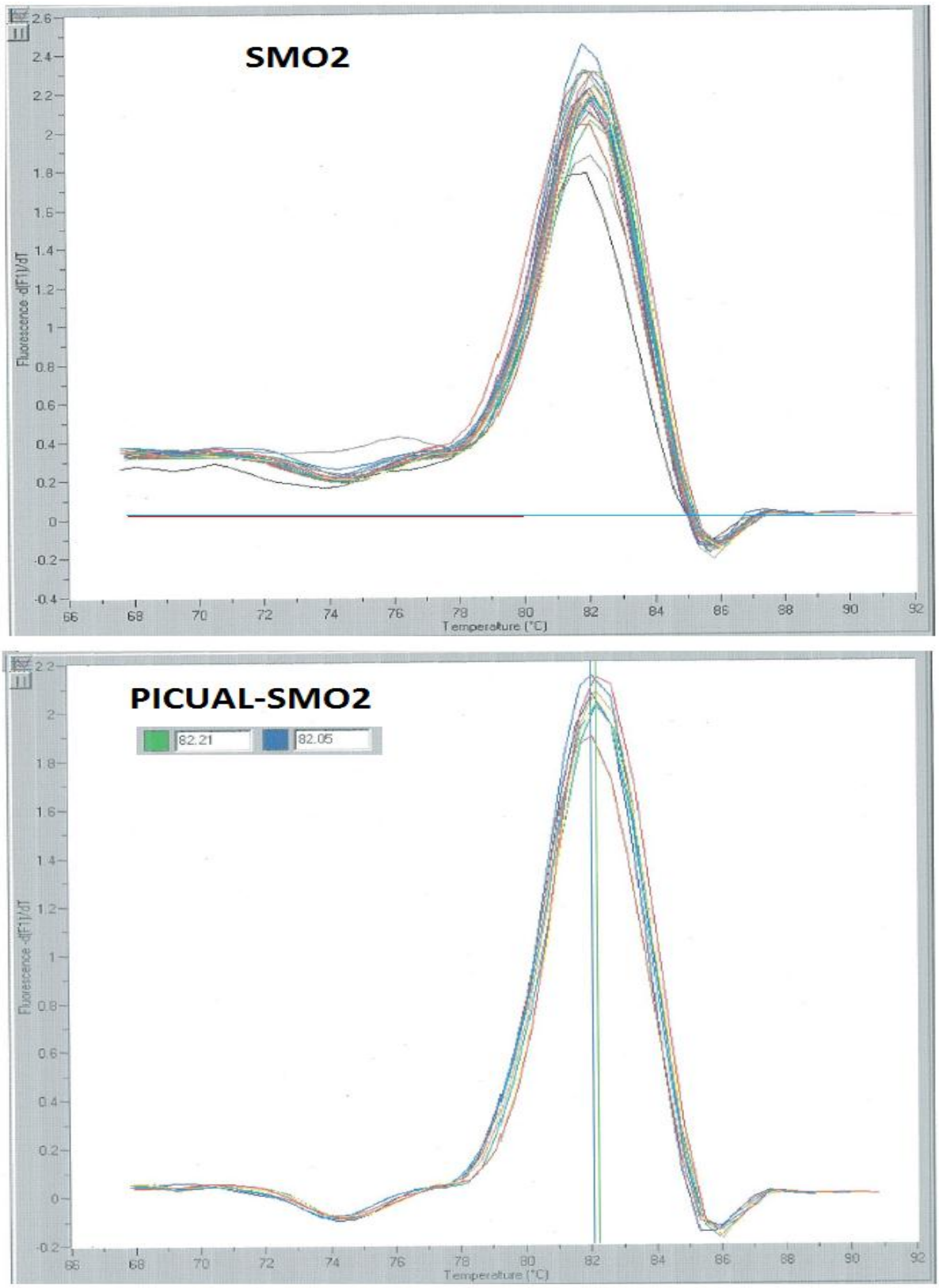


Figure 6.13 Melt curve peaks of *SMO2* in all samples and in Picual samples only
Flouroscence values were plotted against temperature (˚C) using the LightCycler Carousel (Roche)

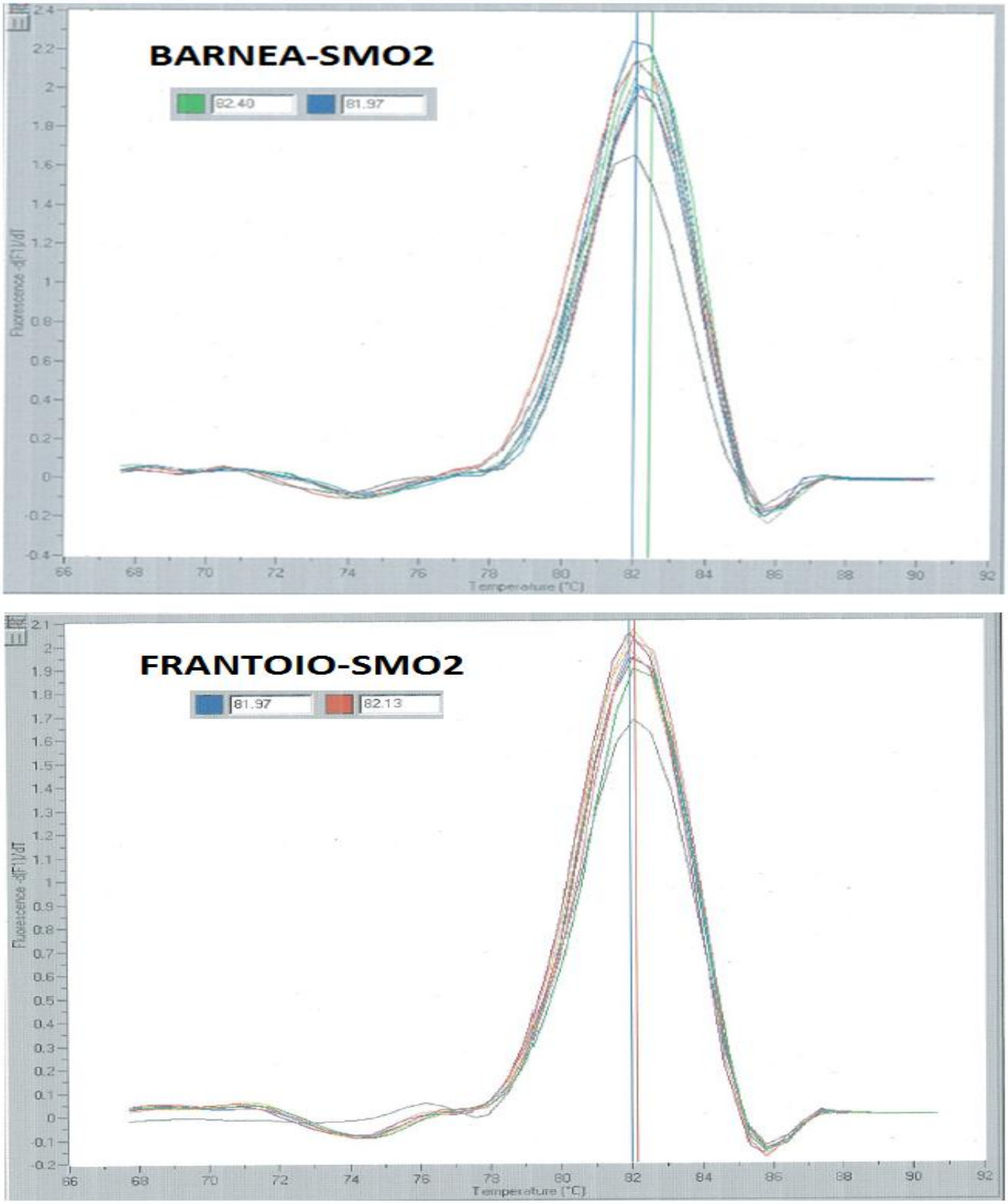


Figure 6.14 Melt curve peaks of *SMO2* in Barnea and Frantoio samples
Flouroscence values were plotted against temperature (˚C) using the LightCycler Carousel (Roche)

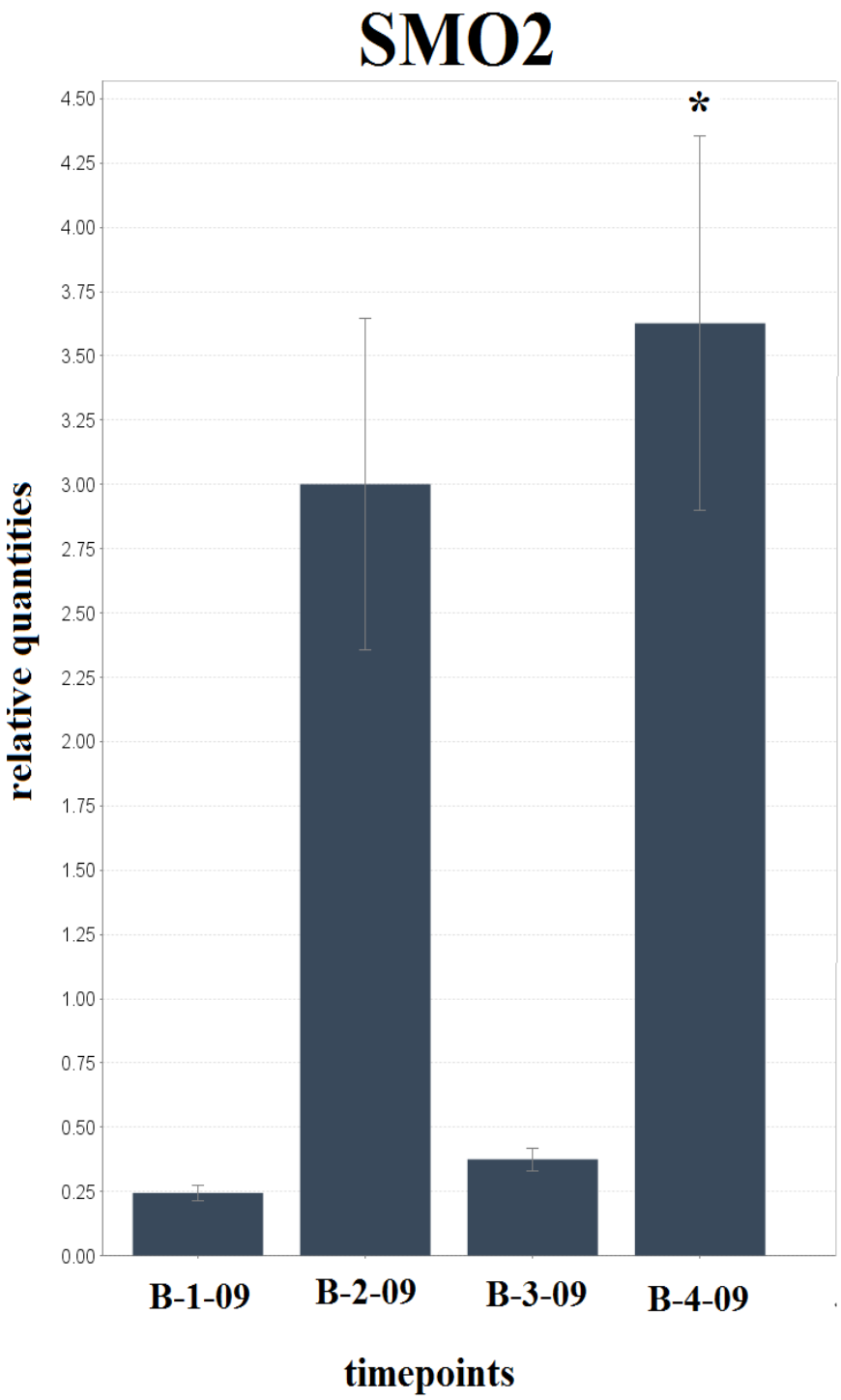


Figure 6.15Relative expression levels of genes coding for SMT2 enzymes during fruit development in Barnea Normalization was performed using the best combination of reference genes selected (Section 6.1.6.3). Each value is the mean± standard deviation of two independent measurements. The asterisk indicates significant difference at 95% level with p-value ≤0.05 between B-1-09/B-4-09 and B-3-09/B-4-09. B-1-09: 96 DAF; B-2-09: 109DAF; B-3-09: 116 DAF; B-4-09: 136 DAF.

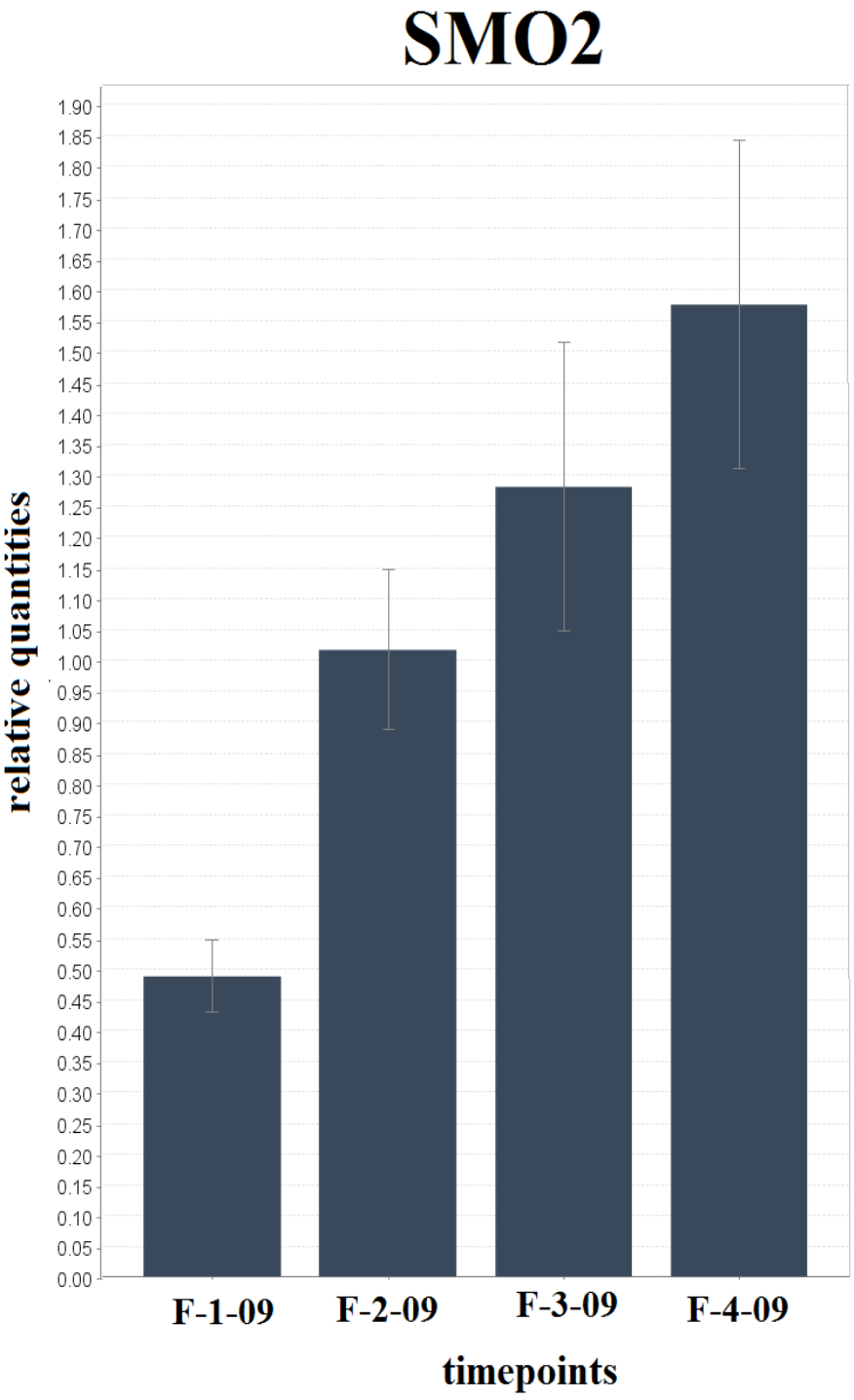


Figure 6.16 Relative expression levels of genes coding for SMT2 enzymes during fruit development in Frantoio Normalization was performed using the best combination of reference genes selected (Section 6.1.6.3). Each value is the mean± standard deviation of two independent measurements. F-1-09: 96 DAF; F-2-09: 109DAF; F-3-09: 116 DAF; F-4-09: 136 DAF.

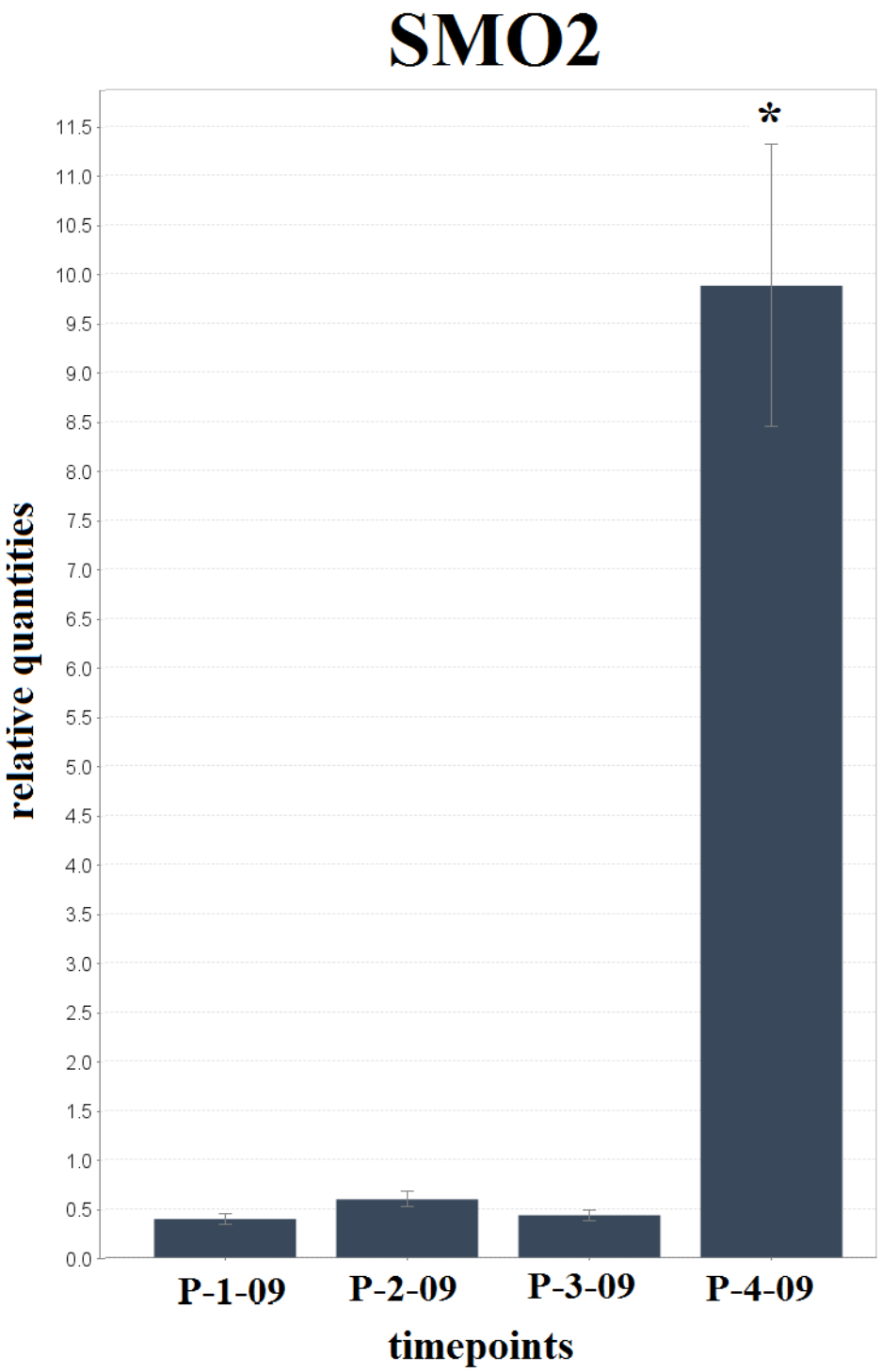


Figure 6.17 Relative expression levels of genes coding for SMT2 enzymes during fruit development in Picual
Normalization was performed using the best combination of reference genes selected (Section 6.1.6.3). Each value is the mean± standard deviation of two independent measurements. The asterisk indicates significant difference at 95% level with p-value ≤0.05 between P-1-09/P-4-09, P-2-09/P-4-09 and P-3-09/P-4-09. P-1-09: 96 DAF; P-2-09: 109DAF; P-3-09: 116 DAF; P-4-09: 136 DAF.

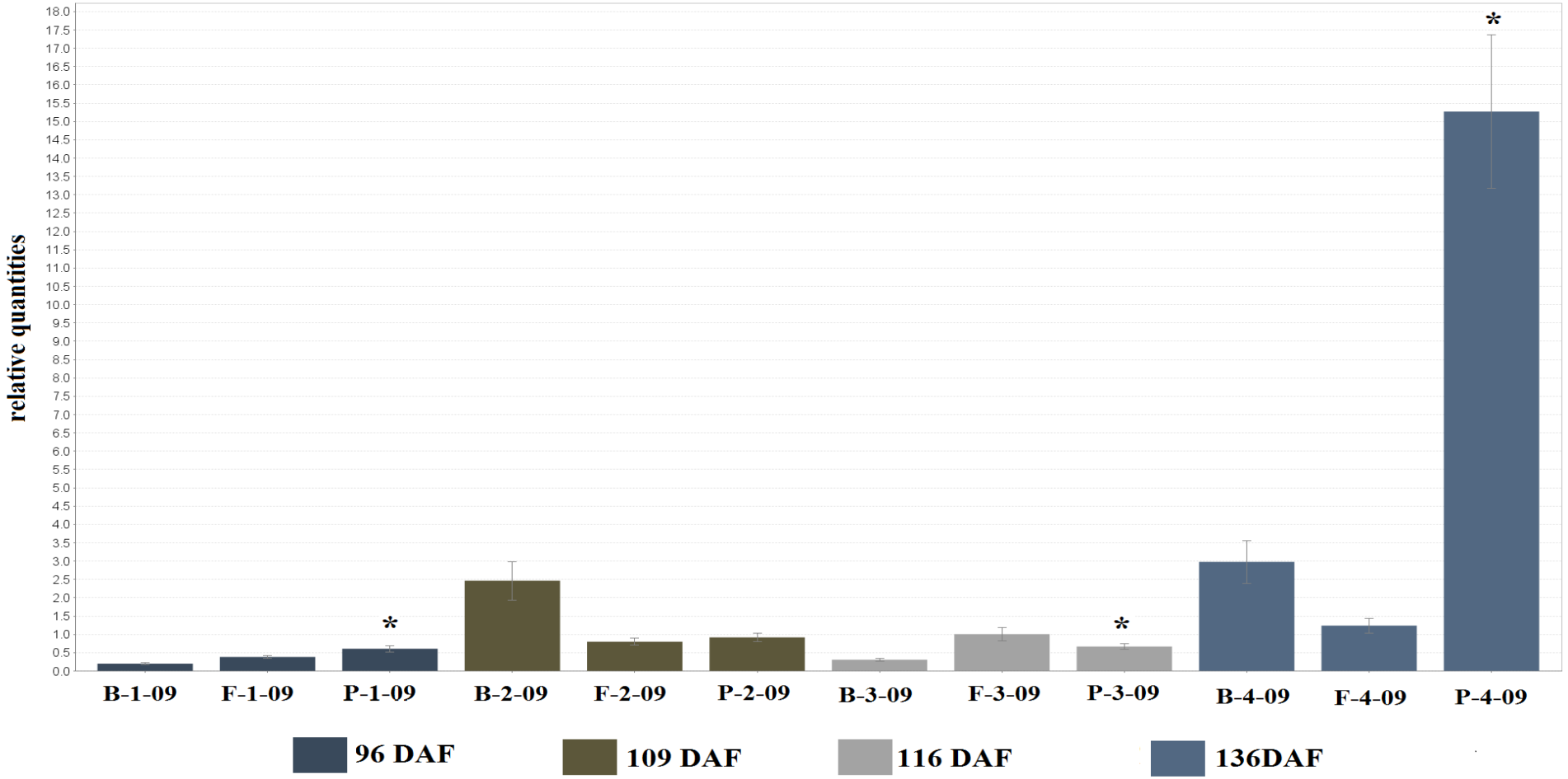


Figure 6.18Relative expression levels of genes coding for SMO2 enzymes across Barnea, Frantoio and Picual in different timepoints
Normalization was performed using the best combination of reference genes selected (Section 6.1.6.3). Each value is the mean± standard deviation of two independent measurements. The asterisk indicates significant difference at 95% level with p-value ≤ 0.05 between B-1-09/P-1-09, B-3-09/P-3-09, B-4-09/P-4-09 and F-4-09/P-4-09.DAF: days after flowering. B: Barnea; F: Frantoio; P: Picual; 1: 96 DAF; 2: 109 DAF; 3: 116 DAF; 4: 136 DAF; 09: 2009 crop season.

6.5 DISCUSSION

In this study, the expression levels of *SMT2* and *SMO2*were investigated at four different timepoints during the development of the olive fruit in three olive cultivars, Barnea, Frantoio and Picual using qPCR. This data provided an insight into the potentialmetabolic roles of *SMT2* and *SMO2*in olivesby identifying the expression pattern of these two genes during the developmental stages of the fruit. In addition, a comparison of the expression patterns of*SMT2* and *SMO2* between the three olive cultivarswas conducted which could possibly explain the presence of higher campesterol levels observed in olive oils extracted from the olive cultivarBarnea (Section 1.7 and 3.2.1).

For an accurate comparison of mRNA levels in different samples, it is crucial to normalize the expression of target genes against appropriate reference genes. An ideal reference gene should be expressed at constant level in all types of cells, at any time in cell cycle and differentiation and/or with any sample treatment (Hu *et al.* 2009; Artico *et al.* 2010; Uddin *et al.* 2011). Traditional reference genes such as EF1-alpha and UBQ (involved in basic cellular processes) or actin and tubulin (involved in cell structure maintenance) have been frequently used in qPCR experiments however recent research have shown that these genes may be inappropriate for normalization in qPCR experiments due to their expression variability under different experimental conditions (Section 1.11), implying the need to validate the expression stability of a control gene in given species and organs/tissues under specific experimental conditions. Normalization with multiple reference genes has become a gold standard in qPCR expression analysis(Hu *et al.* 2009; Artico *et al.* 2010; Gamm *et al.* 2011) and also a requisite according to MIQE(Minimum Information for Publication of Quantitative Real-Time PCR Experiments)guidelines (Bustin *et al.* 2009). To this end, a number of software packages have been developed to assess the stability of candidate reference genes in different biological experimental settings, including, geNorm, BestKeeper and NormFinder (Section 1.11). However, the validation of reference genes is not very common in plant research (Jain M *et al.* 2006; Cruz *et al.* 2009; Artico *et al.* 2010). Although some reference genes were suggested for the major tissues (leaves, fruits and pedicels) of olive (Dündar, 2013), no firmly established information about the identification of valid reference genes in olive hasbeen reported to date.

Therefore, in order to obtain a solid basis for the normalization of gene expression data during the developmental stages of the olive fruit, an evaluation of the expression stability of eight candidate reference genes (Table 2.16) at four different periods during mesocarp development across three olive cultivars during the 2009 season was conducted. Details of all eight candidate reference genes along with their selection criteria have been described in Section 2.10.3 and Table 2.16. In this study the combination of GeNorm (qBase Plus) and BestKeeper programs were used to validate the expression stability of the reference genes in olivesand the three most stable reference genes were used for normalisation against the target genes *SMT2* and *SMO2.*

6.5.1 GAPDH, EF1-alpha and PP2A are suitable reference genes for expression analysisin developing olivemesocarp tissues

This is one of the first reports conducted to assess the expression stability of candidate reference genes in different developmental stages of the olive fruit.

The primers that were designed for the eight reference genes amplified single PCR products of the expected size from the olive cDNA pools as shown by gel electrophoresis (Figure 6.1) and melt-curve analysis (Figure 6.2, 6.6, 6.13) suggested that single products were amplified and that only TIP2 showed the formation of some primer-dimers as indicated by Figure 6.2. This specificity was confirmed by sequencing all eight PCR products which showed 77-96% identity to sequences used for designing primers for the reference genes (Table 6.2).

Previous research has shown that analysis of expression stability of reference genes using different combinations of GeNorm, BestKeeper and Normfinder results in minor to substantial discrepancies in the final ranking of the suitable reference genes which is typically explained by the differences in the mathematical models associated with each program (Reid *et al.* 2006; Gamm *et al.* 2011). In this study the results obtained by the two algorithms did not show much discrepancy and both the programs were compared for the final choice of suitable reference genes. The ranking of the eight candidate reference genes based on their stability (M) and correlation coefficient values as calculated using GeNorm (qBase Plus) and BestKeeper algorithms respectively are shown in Table 6.8.

Ranking	GeNorm	BestKeeper
1	GAPDH	PP2A
2	EF1-alpha	EF1-alpha
3	PP2A	GAPDH
4	60S RBP	PTB
5	PTB	60S RBP
6	UBQ	
7	TUBA	
8	TIP2	

Table 6.8 Final ranking of the candidate reference genes

According to GeNorm analysis using qBase Plus, the two traditional reference genes GAPDH and EF1-alpha were the most stable with lowest M-values, followed by the reference gene PP2A. 60S RBP and PTB were placed in the middle of the ranking, while UBQ, TUBA and TIP2 displayed inappropriate expression stability with higher M-values and thus appear to be regulated in these tissues.

According to BestKeeper analysis software, PP2A was ranked as the most suitable reference gene with the highest correlation coefficient of 0.805, followed by EF1-alpha and GAPDH. PTB and 60SRBP were placed in the middle of the ranking, while UBQ, TUBA and TIP2 showed SD ≥1 and thus considered as inconsistent and were not used for the ranking of stable most reference genes.

Thus, GAPDH, EF1-alpha and PP2A were determined to be the three most stable reference genes analysed in the olive fruit, even if they were not in the exact same ranking order. Many studies with similar findings where different softwares showing the same reference genes as the most stable genes but not in the exact same ranking order have been reported in animals and plants (Maroufi *et al.* 2010; Uddin *et al.* 2011). Among these three genes, EF1-alpha had the highest expression levels in olives, while GAPDH and PP2A had intermediate expression

levels as determined by Cq values (Figure 6.3). It is important to note that the *Arabidopsis thaliana* Affymetrix ATH1 microarray data which has been used to identify several new reference genes superior to the conventionally applied reference genes has shown that these more stably expressed new reference genes are expressed at much lower levels than traditional reference genes (Czechowski *et al.* 2005). The four reference genes GAPDH, EF1-alpha, UBQ and PTB which were used in this study were also assessed in the *Arabidopsis* microarray data which revealed all four genes having high expression levels, however were less stable than other new reference genes used in the study (Czechowski *et al.* 2005).

GAPDH, a traditional reference gene, was considered the most suitable reference gene inthe major olive tissues (leaves, fruits and pedicels) as reported by Dündar, *et al*, 2013, which further supports our data. GAPDH was also chosen as a suitable reference gene in coffee leaves under drought-stress and in different cultivars(Cruz *et al.* 2009), however performed poorly across tissues and organs of tomato at different developmental stages (Expósito-Rodríguez *et al.* 2008).

EF1-alpha was found to be very stably expressed under conditions of biotic and abiotic stress in potato (Nicot *et al.* 2005), while EF1-beta was found to be the most stable in soybean (Jian *et al.* 2008). EF1-alpha was also found to be stable in expression across different tissues of rice (Jain M *et al.* 2006).

PP2A was identified as a suitable reference gene in cotton across different plant organs (Artico *et al.* 2010) as well as in *Arabidopsis*(Czechowski *et al.* 2005). PTB and 60S RBP were placed in the middle of the ranking order in olives by both software packages, again differing in their order (Table 6.8). PTB was identified as one of the most stable reference gene in cotton during fruit development (Artico *et al.* 2010). 60SRBP was identified as one of the suitable reference genes to normalize gene expression data in two different grapevine organs (leaves and berries) (Gamm *et al.* 2011).

Both GeNorm and BestKeeper ranked UBQ, TUBA and TIP2 as poor performers as reference genes in olive fruit development. UBQ14 was identified as one of the most stable reference gene across different plant organs in cotton (Artico *et al.* 2010). UBQ10 enjoyed very stable expression in *Arabidopsis* (Czechowski *et al.* 2005) however performed poorly as a reference in soybean (Hu *et al.* 2009) and in grapevine (Reid *et al.* 2006).TUBA was

identified as being very stably expressed across development in soybean while was highly unstable across tissues and organs of tomato at different developmental stages (Expósito-Rodríguez *et al.* 2008). While TIP2 outperformed several traditional reference genes in *Arabidopsis* across different tissues, organs and developmental stages (Czechowski *et al.* 2005), it performed poorly in grapevine (Reid *et al.* 2006) as well as in the present experiments in olives.

This analysis clearly suggested that reference genes are regulated differently in different plant species which may be partly explained by the fact that reference genes not only play an important role in cell metabolism but also have other cellular functions (Artico *et al.* 2010) and also highlights the importance of validating putative reference genes in different species/tissues/conditions.

GeNorm also provides a measure for the optimal number of stable controls that should be used for normalization based on pairwise variation analysis between subsequent normalisation factors (Section 2.10.8.1). According to Genorm V, a combination of six most stable reference genes was calculated as being optimal for gene expression studies in olive tissue samples with the lowest pairwise variation value of 0.18. According to Vandesompele (2002) the optimal cut-off V numbershould be around 0.15, however many other studies using this application have resulted in higher pairwise variations (Kuijk *et al.* 2007; Silveira *et al.* 2009; Jarošová *et al.* 2010; Hoenemann *et al.* 2011). GeNorm threshold is not a strict cut-off value but its an ideal value to provide guidance to researchers to determine the optimal number of reference genes and that the observed trend of changing pairwise variation values is equally informative (Hong *et al.* 2008; Hu *et al.* 2009).

The aim of this study was to analyse the expression of only two target genes in 12 samples. As this is a comparatively small study, it is impractical to use excessive numbers of reference genes for normalization and thus the minimal use of three most stable reference genes is recommended for calculating the normalization factors (Vandesompele *et al.* 2002; Gu *et al.* 2011; Uddin *et al.* 2011). Therefore though GeNorm V recommended using six reference genes for normalization, we decided to use three most stable reference genes GAPDH, EF1-alpha and PP2A recommended by both the programs for normalization against the target genes in the olive tissue samples.

In conclusion, GAPDH, EF1-alpha and PP2A were found to be the most stable reference genes in olive mesocarp tissues during fruit development. Nevertheless, the limited number of samples and the unavailability of biological replicates was a limitation of this study and this study was only conducted on olive mesocarp tissues. Artico. *et al*(2010) designed a study to identify new reference genes in cotton in 23 experimental samples consisting of six distinct plant organs (flower buds, fruits, leaves, stems, branches, roots and floral meristem), seven stages of flower development (flower buds with seven different diameter sizes), four stages of fruit development (fruits with different diameter sizes) and in flower verticils (petal, sepal, stamen, carpel and pedicel). Therefore to extend our observation in olives, a similar study may be designed by comparing the performance of the three stable reference genes identified in this study along with additional frequently used and new reference genes over a large set of biological samples representing not only different developmental stages, but different olive tissues/organs during flower and fruit development.

In summary, this is one of the first reports on the evaluation of candidate reference genes in *O. europaea*L. and this study will provide guidance to other researchers to select reference genes for normalization across tissues obtained from the mesocarp region of the olive fruit. However, if different treated samples or different tissues/organs of the olive plants are used, the expression stability of the reference genes should be re-evaluated under those conditions.

6.5.2 Olive SMT2 is encoded by atleast twogene loci

In chapter 4,we have described the isolation and characterization of the full length coding sequences of the *SMT2* genes from three different olive cultivars, Barnea, Frantoio and Picual along with an analysis of the allelic differences between the olive cultivars.

This study revealed four putative *SMT2* coding sequences identified in olive, which may represent four alleles of two *SMT2* genes in diploid olive(Section 4.8.1).Results from the allele specific PCR suggested that all four alleles are present in the three cultivars indicating no allelic differences between the cultivars which was further supported by the Southern blotting results which did not identify any distinct RFLPs between the olive cultivars (Section 4.8.2).

In order to further validate our results derivative melting peak data generated in the qPCR experiment was also analysed. A single primer pair that was designed to amplify all the *SMT2* alleles was used for qPCR analysis. Analysis of the sequencing data showed that the four*SMT2* sequences identified in Barnea, were expected to have no melting temperature differences between each other.

However, the melting curve data of *SMT2* showed the presence of at least two peaks in Frantoio and Picual cultivars although only differing by 0.2-0.6˚C. Barnea also showed the presence of the two melting peaks as observed in Frantoio and Picual, however there was an additional peak observed at 82.68˚C. This may indicate the presence of different *SMT2* alleles in olives, further indicating the presence of an additional *SMT2* allele in olives which might have escaped detection in our screeningprocedures due to more divergentDNA sequence composition.This would be similar to the *SMT2* gene families of other plant species such as *A. thaliana(Schaeffer et al. 2001), N. tabacum*(Schaeffer *et al.* 2000),*G.* max (Neelakandan *et al.* 2010) and *G. hirsutum*(Luo *et al.* 2008), which have all been found to contain two copies of the *SMT2* gene.

6.5.3*SMT2* genes exhibit similar expression patterns in Barnea, Frantoio and Picual

In Barnea, *SMT2* expression was lower in early stages of fruit development (96 DAF) with a 9.3 fold (109 DAF), 6 fold (116 DAF) and 6.7 fold (136 DAF) increase gradually as the fruit matured in comparison to 96 DAF. However, only the 6 fold increase in 116 DAF was statistically significant ($p \leq 0.05$).

In both Frantoio and Picual, there was a significant decrease in *SMT2* expression at 116 DAF (1.9 fold and 3.4 fold respectively) as compared to 96 DAF however the transcript levels of *SMT2* peaked at 136 DAF for both of the cultivars (2.7 fold and 8.6 fold respectively) as compared to the 116 DAF.

This analysis showed that in all three olive cultivars, *SMT2* is expressed at lower levels during the early developmental stages of the olive fruit with a slight decrease at 116 DAF for Frantoio and Picual followed by expression levels peaking at fruit maturation stage (136

DAF). In Barnea, *SMT2* expression peaked at 109 DAF however this change was not statistically significant ($p \geq 0.05$). This is in contrast with soybean where the expression analysis of *SMT2-1* and*SMT2-2* in leaf, roots and seeds and at various stages of growth of soybean showed *SMT2* was constitutively expressed in all tissues during the life cycle of soybean, however expression levels declined to very low levels in the later stages of seed maturation (Neelakandan *et al.* 2010). In cotton, highest expression levels were detected in 10 DPA (day post anthesis) fibers and the lowest expression levels were observed in leaves and cotyledons for both *GhSMT2-1* and *GhSMT2-2* genes (Luo *et al.* 2008). In this study the expression of all *SMT2* alleles as a whole were analysed in different olive cultivarsbut the individual expression patterns of each of the *SMT2* alleles is yet to be investigated. This is important as in other plants such as cotton and soybean (Luo *et al.* 2008; Neelakandan *et al.* 2009) significant differences in expression levels of the SMT2 isoforms have been identified. In cotton expression profiles of the two *GhSMT2* genes in different organs and at various stages of fibre development showed that the expression level of *GhSMT2-1* was10 times higher than *GhSMT2-2* in all organs and tissues detected(Luo *et al.* 2008).Unlike *Arabidopsis*, where the expression of *At*SMT2-2 was generally lower and thought to be functionally redundant, in soybean, *SMT2-2* was expressed generally at higher levels than *SMT2-1* or *SMT1* in all tissues, suggesting different requirements for the SMT2 isoforms in soybean physiology (Neelakandan *et al.* 2009). Further, the catalytic efficiency of *Gm*SMT2-2 for the natural substrate was slightly higher than *Gm*SMT2-1 whereas the catalytic efficiency of *Gm*SMT2-1 was more similar to the catalytic efficiency of *Gm*SMT1(Neelakandan, Song et al. 2009).

In this study the only statistical difference in *SMT2* expression between the three cultivars was observed at 116 DAF where there was a decline in *SMT2* expression in Frantoio and Picual (2.6 fold and 4.8 fold respectively] compared to Barnea. However this result is not especially interesting at this point considering low number of replicates and relatively small differences in expression observed between cultivars. Further studies with more biological replicates and comparison between seasons may provide more insight in this area.

6.5.4 Olive SMO2 is encoded by atleast two gene families

In chapter 5,we have described the isolation and characterization of the full length coding

sequences of the *SMO2* genes from three different olive cultivars, Barnea, Frantoio and Picual along with an analysis of the allelic differences between the olive cultivars.

The analysis of the sequencing data has revealed (Section 5.8.1) that olive SMO2 is likely encoded by two gene loci. Results from the allele specific PCR suggested that allelic differences may be present between cultivars which was further supported by the Southern blotting results which identified distinct RFLPs between the olive cultivars (Section 5.8.2).In order to further validate our results,derivative melting peak data generated in the qPCR experiment was also analysed. Only a single primer pair that would amplify all the *SMO2* alleles was used for qPCR analysis. Analysis of the sequencing data showed that the two distinct *SMO2* sequences identified in Barnea, *OeSMO2-1a* and *OeSMO2-1b* were expected to have two slightly different melting temperatures of 79˚C and 80˚C but no melting temperature differences were expected from the other identified *SMO2* alleles.

The melting curve data of *SMO2* showed the presence of two melting peaks in Barnea, Frantoio and Picual although only differing by 0.2-0.4˚C. This could possibly explain the presence of two *SMO2* gene loci in olives validating our sequencing data, which would also be similar to the number of *SMO2* genes identified in *Arabidopsis*(Darnet *et al.* 2001). The additional *SMO2* alleles that were identified in the olive cultivars(Section 5.6) could not be detected in the melting peak data due to the high sequence identity and, therefore, melting temperatures, between the alleles.

6.5.5*SMO2*has higher expression in Picual as compared to Barnea and Frantoio

In both Barnea and Picual, *SMO2* was expressed at lower levels during the early developmental stages (96 DAF, 109 DAF and 116 DAF) of the olive fruit with the expression levels peaking at fruit maturation stage (136 DAF), except in Barnea which showed a 12 fold increase at 109 DAF, though not statistically significant (p=0.06). Interestingly, in Picual there was a significant upregulation of *SMO2* at 136 DAF, with a 25 fold, 16 fold and 22 fold increase as compared to 96 DAF, 109 DAF and 116 DAF respectively (Figure 6.17).

In Frantoio, *SMO2* expression was lower in early stages of fruit development (96 DAF) with a 2 fold (109 DAF), 2.6 fold (116 DAF) and 3.2 fold (136 DAF) increase gradually as the fruit matured, however none of the changes in expression were statistically significant ($p>0.05$).

Comparison of *SMO2* expression between the three olive cultivars showed that *SMO2* expression was relatively constant at 96 DAF, 109 DAF and 116 DAF in all three olive cultivars, with a slight (3 fold and 2.1 fold) but significant increase in expression in Picual as compared to Barnea at 96 DAF and 116 DAF respectively.

The most interesting observation of this study was that at the fruit maturation stage (136 DAF), a significant difference in *SMO2* expression was observed between the three olive cultivars with Picual having a significantly higher level of expression of *SMO2* by 5 fold as compared to Barnea and by 12.3 fold as compared to Frantoio.

The analysis of this data clearly suggests that as lipids get synthesized later in the developmental stages of the olive fruit, *SMO2* expression in the fruit gradually increases, however the magnitude of this change in expression of *SMO2* does appear to vary between olive cultivars. It is important to note that, except at 109 DAF, there is a significant difference in *SMO2* expression observed between Barnea and Picual at 96 DAF and 116 DAF, with a robust upregulation in Picual at 136 DAF, suggesting differential expression pattern of *SMO2* in these two cultivars. As mentioned before (Section 6.5.3), the unavailability of biological replicates and samples from different seasons is a limitation of this study, however the magnitude of *SMO2* expression differences observed between Barnea and Picual at 136 DAF is quite interesting. Similar expression analysis of *SMO2* in different tissues/organs or at various stages of growth inother plant species is yet to be investigated.

6.5.6 The relationship between *SMT2* and *SMO2* expression and campesterol levels

Previous research has clearly shown that *SMT2*and *SMO2* play a crucial role in balancing the ratio of campesterol to sitosterol in *Arabidopsis* and tobacco (Section 1.9). In *Arabidopsis*, transgenic plants overexpressing*AtSMT2-1* showed a concomitant increase of the amount of sitosterol with a parallel drop in the amount of campesterol and brassicasterol and vice-versa

(Schaeffer *et al.* 2001). In tobacco, *SMO2* knock down transgenic plants showed increased accumulation of sitosterol and stigmasterol and decreased accumulation of campesterol (Darnet *et al.* 2004).

In this study we assessed the expression profiles of *SMO2* and *SMT2* genes in different olive cultivars during different developmental stages of the fruit. This expression study is just a preliminary study to get an overview of the expression profile of these genes in olives and is limited to determining the mRNA levels of these genes in different olive cultivarsbut the impact of these differences at protein level in the mature fruitis yet to be investigated. Thus, in order to understand the functional role of *SMO2* and *SMT2* genes in olives, olive mutants lacking these genes, or individual alleles, could be prepared and subsequently the sterol levels in the plants could be analysed to understand the impact of these enzymes on the olive oil.

Unlike *Arabidopsis* and tobaccomutants showing lower campesterol levels in overexpressed *SMT2* transgenic mutants, in soybean a set of transgenic *Arabidopsis* plants harbouring seed targeted overexpression of *GmSMT2-2* showed an increase in sitosterol and stigmasterol levels while campesterol levels were more or less the same(Neelakandan *et al.* 2010). Since the sitosterol:campesterol ratio was not drastically altered it seemed likely that the campesterol levels were tightly regulated in soybean seed tissues implying that the role of SMT2 in controlling the levels of campesterol differs between plant species. Therefore it will be interesting to see what impact does natural mutants of *SMT2* and *SMO2*,which can be screened from the olive stocks, have on the sterol levels in olive plants.

Interestingly in our study we found that *SMT2* genes exhibit similar expression patterns in Barnea, Frantoio and Picual throughout fruit development and *SMO2* gene expression varies significantly during late development of the fruit between Barnea/Frantoio and Picual. The most noteworthy difference observed was the dramatic upregulation of *SMO2* at the maturation stage of the olive fruit (136 DAF) in Picual. This may indicate that in Picual at later timepoints, when maximum amount of oil gets deposited in the olive fruit, *SMO2* expression increases significantly and this in turn may lead to the increased accumulation of 24-methylene lophenol and eventually increase in campesterol levels at the expense of sitosterol. Previous research has shown that in *Arabidopsis* mutants where *SMT2* was co-suppressed, an accumulation of 24-methylene-lophenol, the substrate of *SMT2* takes placeindicating that in sterol biosynthesis, the subsequent metabolisation of 24-methylene-

lophenol by *SMO2* is probably a slow step, (Bouvier-Navé *et al.* 1997)thus offering venues for “fine” control of sterol composition in phytosterol biosynthesis in *Arabidopsis.*

Our results in this study are interesting as Australian olive oils extracted from the cultivar Picual have shown to contain comparatively lower levels campesterol (3.53%) while olive oils extracted from the cultivar Barnea have shown consistently higher campesterol levels (4.8%) (Section 3.2.1). Thus, though we found significant differences in *SMO2* expression between the olive cultivars, we would expect *SMO2* to be overexpressed in Barnea to have higher campesterol levels. However in our study in Barnea, no expected upregulation of *SMO2* was observed. To explain such higher campesterol levels observed in Barnea oils, we may need to further investigate other genes families that are downstream of SMO2 enzyme in the sterol biosynthetic pathway (Figure 1.13) such as STE1/DWF7, DWF-5 and DIM. These enzymes are shared between the campesterol and sitosterol pathways, so any change in their activity will affect both products, but it would be interesting to see if there are any differences in the expressionof these downstream genes between Barnea, Frantoio and Picual. It would also be interesting to check the expression of BR-biosynthetic enzymes such as CYP in olives as this enzyme catalyses the conversion of end-product campesterol to brassicasterol. It is plausible if this enzyme is significantly downregulated in Barnea as compared to other cultivars, it might lead to excessive accumulation of campesterol in this cultivar.

In conclusion, in this study only subtle differences were observed in *SMT2* expression between the olive cultivarswhereas *SMO2* exhibit some interesting differences between Barnea, Frantoio and Picual but these would need to be further confirmed across biological replicates and seasons.

CHAPTER 7
CONCLUSIONS AND FUTURE DIRECTIONS

The increase in olive oil consumption as a result of its reported health benefits has led to a large increase in the cultivation of olives and production of high quality olive oils in Australia. As olive oil is an international commodity it is important to ensure proper development of the local olive industry in an international context. One of the major concerns for Australian olive oil producers is that much of their oils notably, those derived from the Barnea cultivar (representing 41% of the olive crop in Australia), have campesterol levels up to 4.8% of total phytosterol content, which exceeds the international standards for extra virgin olive oil that stipulate a campesterol level of less than 4% and as a result, these oils cannot be exported on an international basis. The biological basis for this difference remain unanswered, however it is not unique to Australia as other countries around the world have also reported similar concerns (Koutsaftakis *et al.* 1999; Salvador *et al.* 2001)

As the sterol composition andtotal sterol content of olive oil has been found to be affected by various geographical and technological factors, one of the major aims of this study was to investigate the impact of horticultural and processing practices such as fruit size, irrigation, fruitmaturity, malaxing time, malaxing temperature, delays between harvest and process and storagetime on sterol levels of olive oils produced from three olive cultivars Barnea, Frantoio and Picual.

Furthermore, the consistently high campesterol levels as observed in Australian olive oils, particularly those derived from the olive Barneacultivar, on a year to year basis, strongly implicates genetic factors as the cause of these different levels. The characterized sterol biosynthetic pathway in plants contains a bifurcation that leads to the formation of β-sitosterol or campesterol, with the flux controlled by the activity of two branch-point enzymes, SAM-24-methylene-lophenol-C-24-methyltransferase2 (SMT2) and C-4α-sterol-methy-oxidase2 (SMO2). There is clear evidence that the genes*SMT2*and *SMO2*play crucial roles in determining the ratio of campesterol to sitosterol in plants. It is plausible that allelic differences in these genes and/or differences in the expression patterns of these genes may impact on the sterol profiles observed in the oils derived from these cultivars. Thus one of the major aims of this study was to isolate and sequence *SMT2* and *SMO2* cDNAs from Barnea,

Frantoio and Picual to characterize these genes in olives and identify any allelic differences between these cultivars. Finally, this study also aimed to investigate the expression levels of the *SMT2* and *SMO2* genes throughout olive fruit development to identify any inter-cultivar differences that may also impact on the sterol profiles in these Australian olive oils.

However, to study the expression levels of *SMT2* and *SMO2* genes in olives, the expression level of these target genes should be normalised using stable internal control genes known as reference genes to derive relative changes in gene expression. To our knowledge, no widely established information about the identification of valid reference genes in olive is available to date, although some reference genes were suggested for the major tissues (leaves, fruits and pedicels) of olive (Dündar, 2013). Therefore, in order to normalise the expression data of target genes in olives, a reference - genes - validation experiment was further conducted to select and / or confirm stably expressed candidate reference genes.

CONCLUSIONS

7.1 Impact of technological factors affecting phytosterol levels in Australian olive oils

The horticultural and processing practices that may have an impact on the sterol content and profile of the olive oils extracted from the olive cultivars under investigation, Barnea, Frantoio and Picual were analysed (Chapter 3).

- The analysis of the sterol composition of olive oils extracted from three different cultivars and two consecutive years have shown that though most of the sterol compounds were significantly affected by the season, the cultivar has shown the most significant level of effect on the composition of different sterols, where particularly the cultivar Barnea showed exceptionally higher levels of campesterol which exceeded the IOOC limit of 4%.
- Horticultural practices such as irrigation and fruit characteristics showed that most of these compounds such as β-sitosterol, sitostanol, Δ5-avenasterol and Δ7-avenasterol are significantly affected however, the campesterol levels remained consistently high regardless of fruit size/maturity/irrigation regimes.

- Processing practices such as malaxing time, malaxing temperature and delays between harvest and processing, had asignificant impact on the concentrations of triterpenedialcohols and stigmasterol, howeverthere was no significant change in the campesterol content of these oils.

In conclusion, different horticultural and processing practices seem to have a significant impact on some aspects of the sterol content and profile of the most commonly cultivated cultivars of olive in Australia, however no evaluated management or processing practices seem to have contributed in affecting the relative campesterol levels in these olive oils. This is similar to other olive producing countries where olive oils extracted from olive cultivars such as Cornicabra, Barnea, Koroneiki and Arbequina have consistently shown to have high campesterol levels irrespective of variations in environmental parameters (Ceci *et al.* 2007; Pardo *et al.* 2011). These consistent levels of campesterol observed within these Australian olive cultivars across more than one season and despite various altered horticultural and processing practices strongly suggest that the genetic makeup of the cultivars is the most important determinant of relative sterol levels.

7.2 The characterization and expression pattern of *SMT2*and *SMO2* genes of olive

Previous research has shown that the orientation of the sterol biosynthetic flux towards sitosterol or campesterol is mainly controlled by the two branchpoint enzymes, SMT2 and SMO2 in plants. As experimental evidences have demonstrated that both SMT2 and SMO2 can influence campesterol and β-sitosterol levels in *Arabidopsis thaliana* and tobacco, it is conceivable that the relative activity or expression of these enzymes could play a pivotal role in determining the relative amounts of β-sitosterol and campesterol in Australian olive oils.

7.2.1 The characterization and expression pattern of *SMT2* gene of olive

Investigations into the *SMT2* gene of olive provided insights into the organisation of this gene's alleles in olive (Chapter 4). In addition, the expression pattern of this gene in olive was also studied (Chapter 6) and the following conclusions were drawn:

- The RACE sequencing data and subsequent sequencing of full length *SMT2* clones from Barnea revealed high sequence conservation (>98.9%) between the clones where clone F1 was selected as representative of all Barnea *SMT2* clones and was named as *OeSMT2-1a*. Further analyses of the Barnea clones revealed the plausibility of the presence of three additional alleles namely *OeSMT2-1b, OeSMT2-1c* and *OeSMT2-1d*based on the identification of key SNPs within the coding sequence.
- The full length *SMT2* cDNA sequenceof Barnea *OeSMT2-1a* had an open reading frame of 1083bp and encoded a protein of 360 residues with a predicted molecular mass of 40.1kDa.
- The typical conserved structures characterized by the sterol C-24 methyltransferase (Section 1.8.2.3), such as region I(YEWGWGQSFHF), region II (LDAGCGVGGPMRAI), and region III (YSIEATCHAP) and region IV (KPGSMYVSYEW) were present in olive SMT2-1a deduced protein. The sequence alignment of OeSMT2-1a and other SMT2 homologuesfrom few plant species revealed that the spacing of these signature motifs in OeSMT2-1a align with the corresponding SMT2 sequence of other plant SMT2s.
- Pfam analysis suggested that *OeSMT2* belongs to the family of SAM dependent methyltransferases which includes a group of sequence related proteins that catalyse the distinct patterns of 24-alkyl sterols.
- BLASTsearches revealed high degree of sequence identity with *SMT2* sequences isolated from various other plant species such as *Arabidopsis,* cotton, tobacco, castor oil and soybean (77.2-85.8%) which suggests that OeSMT2-1a belongs to the SMT2 family.
- The Southern blotting data suggested that there are at least 2 loci within the Barnea genome that encode SMT2.
- The sequencing results and Southern analysis data, taken together suggested the presence of two *SMT2* gene families in olives, where the presence of the four *SMT2* sequences may represent four alleles of two *SMT2* genes.
- The full length *SMT2* cDNAs were isolated from two additional olive cultivars, Frantoio and Picual and the comparison of the *SMT2* clones between olive cultivars Barnea, Frantoio and Picual revealed high sequence conservation between the cultivars (99.4-100%). Based on their key SNPs, all three *OeSMT2*-1 alleles except *OeSMT2-1d* were identified from the sequencing data of all three olive cultivars.

- Results from the allele specific PCR suggested that no *SMT2* allelic differences are present between the olive cultivars which was further supported by the Southern blotting results which failed to identify any distinct RFLPs between the olive cultivars.
- The *SMT2* expression analysis showed that in all three olive cultivars, *SMT2* is expressed at lower levels during the early developmental stages of the olive fruit with a slight decrease at 116 DAF for Frantoio and Picual followed by expression levels peaking at fruit maturation stage (136 DAF). Above all, the analysis showed that *SMT2* genes exhibit similar expression patterns in Barnea, Frantoio and Picual with no such interesting differences in the expression pattern between the cultivars.

In conclusion, the results from the isolation of full length *SMT2* genes from olive cultivars Barnea, Frantoio and Picual and the subsequent Southern blotting data revealed that there are two*SMT2* loci that are heterozygous with the same alleles in all three cultivars. Therefore it would appear that allelic differences in the *SMT2* gene family are not likely to be responsible for the contrasting phytosterol profiles observed in the olive oils derived from these cultivars (Section 1.7 and 3.2.1).The *SMT2* expression analysis revealed that the differences observed in *SMT2* expression among the olive cultivars are too subtle to suggest that they are real.

7.2.2 The characterization and expression pattern of *SMO2* genesof olive

Investigations into the *SMO2* gene of olive provided insights into the organisation of this gene in olive (Chapter 5). In addition, the expression pattern of this gene family in olive was also studied (Chapter 6) and the following conclusions were drawn:

- The RACE sequencing data and subsequent sequencing of full length *SMO2* clones from Barnea revealed two distinct *SMO2* cDNAs, designated as *OeSMO2-1* and *OeSMO2-2* with91.24% similarity to each other at amino acid level and 89.65% similarity at nucleotide level.
- The full length cDNA sequences of *OeSMO2-1* and *OeSMO2-2* have open reading frames of 822bp and 783bp respectively, and encode proteins of 274 and 261 residues with a predicted molecular mass of 31.8kDa and 30.8 kDa respectively.

- The sequence alignment of OeSMO2-1a, OeSMO2-2a and other SMO2 homologuesfrom few plant species revealed that the two OeSMO2 sequences possess the three distinct histidine-rich regions (H^{127}RILH, H^{140}SVHH and H^{220}DYHH) which are typical of all all SMO2 proteins.
- The prediction of four possible hydrophobic transmembrane domains and a C-terminal hydrophilic regionsupports the notion that *SMO2* genes in olives are membrane bound.
- Pfam analysis suggested that *OeSMO2*belongs to the fatty acid hydroxylase superfamily which includes C-5 sterol desaturases and C-4 sterol methyl oxidases which is typical of all membrane-bound non-haem iron oxygenases.
- BLAST searches revealed high degree of sequence identity with SMO/SMO2 sequences isolated from various other plant species such as *Arabidopsis,* castor oil, tomato, cotton, tobacco, grapevine and soybean (75-85.5%) which suggests that OeSMO2-1 and OeSMO2-2 belongs to the SMO/SMO2 family.
- The Southern data suggested that there are at least 2 loci with four alleles within the Barnea genome that encode SMO2.
- The full length *SMO2* cDNAs were isolated from two additional olive cultivars, Frantoio and Picual and the comparison of the *SMO2* clones between olive cultivars Barnea, Frantoio and Picual revealed high sequence conservation between the cultivars (90.0-99.6%). The *OeSMO2-1* alleles which have been identified in Barneacultivar, were also identified in Frantoio and Picual however *OeSMO2-2* alleles were absent in the sequencing data of Frantoio and Picual which may warrant further investigation in future studies (see below).
- Apart from the two distinct *SMO2* sequences identified in Barnea, SNPs were identified at several places along the nucleotide sequences of the Barnea, Frantoio and Picual clones suggesting that there may be additional *OeSMO2-1* and / or *OeSMO2-2* alleles in these cultivars. To validate the presence or absence of these cultivar specific alleles in the other olive cultivars, allele specific PCR was conducted under stringent conditions
- Results from the allele specific PCR suggested that allelic differences may be present between cultivars which was further supported by the Southern blotting results which identified distinct RFLPs between the olive cultivars.

- To further confirm the presence of this second sequence (*OeSMO2-2*) in Frantoio and Picual, the AS-PCR products of this allele were sequenced revealing this allele was present in all three olive cultivars under investigation.
- Comparison of *SMO2* expression between the three olive cultivars showed that *SMO2* expression was relatively constant at 96 DAF, 109 DAF and 116 DAF in all three olive cultivars, with a slight (3 fold and 2.1 fold) but significant increase in expression in Picual as compared to Barnea at 96 DAF and 116 DAF respectively. The most interesting observation of this study was that at the fruit maturation stage (136 DAF), a significant difference in *SMO2* expression was observed between the three olive cultivars with Picual having a significantly higher level of expression of *SMO2* by 5 fold as compared to Barnea and by 12.3 fold as compared to Frantoio. The analysis of this data clearly suggests that as lipids get synthesized later in the developmental stages of the olive fruit, *SMO2* expression in the fruit gradually increases, however the magnitude of this change in expression of *SMO2* does appear to vary between olive cultivars.

In conclusion, the isolation of full length *SMO2* genes from olive cultivars Barnea, Frantoio and Picual and the subsequent Southern blotting data revealed the presence of at least four alleles representing two *SMO2* genes in diploid olives. The Southern blotting analysis and AS-PCR results also suggested the presence of additional *SMO2* alleles in olives and allelic differences between the olive cultivars may be present. The *SMO2*expression analysis revealed that *SMO2* exhibit some interesting differences between Barnea, Frantoio and Picual but these would need to be confirmed across biological replicates and seasons.

7.3Selection of suitable reference genes for expression analysis in developing olivemesocarp tissues

As part of this research it was necessary to validate a set of reference genes for analysing relative gene expression levels in the developing olive fruit and these results are relevant to any research into gene expression in developing olive fruit.

To this end, eight commonly used reference genes in plants were chosen from previous reports for normalisation of RT-qPCR data in olives (Chapter 6) which included 60S ribosomal protein L18, protein phosphatase 2A, polypyrimidine tract-binding protein, tubulin alpha, aquaporin tonoplast intrinsic protein, ubiquitin carrier protein, glyceraldehyde 3-phosphate dehydrogenase and elongation factor 1 alpha and the analysis of the expression stability of reference genes were compared using two different algorithms, GeNorm (qBase Plus) and BestKeeper and the identified stablemost reference genes were used for accurate normalization against the target genes, *SMT2* and *SMO2*.

- GAPDH, EF1-alpha and PP2A were determined to be the three most stable reference genes analysed in the olive fruit by both algorithms, even if they were not in the exact same ranking order while UBQ, TUBA and TIP2 were determined to be the least stable genes.
- According to Genorm V, a combination of six most stable reference genes was calculated as being optimal for gene expression studies in olive tissue samples, however as this is a comparatively small study, it is impractical to use excessive numbers of reference genes for normalization, therefore the three most stable reference genes GAPDH, EF1-alpha and PP2A recommended by both the programs were used for normalization against the target genes in the olive tissue samples.

In summary, this study will provide guidance to other researchers to select reference genes for normalization across tissues obtained from the mesocarp region of the olive fruit. However, if different treated samples or different tissues/organs of the olive plants are used, the expression stability of the reference genes should be re-evaluated under those conditions.

FUTURE DIRECTIONS

The findings of the current study have provided much of the groundwork to permit investigations into functional associations between the *SMT2* and *SMO2* genes and sterol profiles of olive oils. In addition, this research has provided the Australian olive industry with the means to investigate the biological basis of phytosterol levels and some leads as to what may have an impact on these levels in olive oils. Identification of the genetic basis of sitosterol and campesterol levels may providesupportfor the use of alternative testing approaches to identify the adulteration of olive oil and may even provide such alternatives through a better understanding of how these compounds are synthesised. This may prove useful in selecting cultivars or marketing cultivars with specific sterol profiles in the future.

While providing some answers, this study also raises several research questions that warrant further investigation. The following research directions are suggested to gain further insight into these areas.

- The sequencing results of the full length *SMT2* clones and Southern analysis data, taken together suggested the presence of two *SMT2* loci in olive. However it is important to note that the two isolated *SMT2* genes share 78.38%, 89.0% and 86.2% sequence similarity between each other at nucleotide level in *Arabidopsis*, soybean and cotton respectively, however the Barnea *SMT2* alleles share >99% similarity at nucleotide level. This may indicate either *SMT2* alleles in olive are highly conserved unlike other plant species, or an additional *SMT2* allele may be present in olive which might have escaped detection in our screeningprocedures due to more divergent DNA sequence compositionand thus this possibility should not be excluded and would require further investigation.

- In this study the reference gene validation experiment was conducted using only eight reference genes on only 12 samples during one crop season. The limited number of samples and the unavailability of biological replicates was a limitation of this study and this study was only conducted on olive mesocarp tissues. Artico. *et al* (2010) designed a study to identify new reference genes in cotton in 23 experimental samples consisting of six distinct plant organs (flower buds, fruits, leaves, stems, branches,

roots and floral meristem), seven stages of flower development (flower buds with seven different diameter sizes), four stages of fruit development (fruits with different diameter sizes) and in flower verticils (petal, sepal, stamen, carpel and pedicel). Therefore to extend our observation in olives, a similar study may be designed by comparing the performance of the three stable reference genes identified in this study along with additional frequently used and new reference genes over a large set of biological samples representing not only different developmental stages, but different seasons and olive tissues/organs during flower and fruit development.

- In this study the expression of all *SMT2* and *SMO2* alleles as a whole, using conserved primers, were analysed in different olive cultivarsbut the individual expression of each of the *SMT2*and *SMO2* alleles is yet to be identified.This is important as in other plants such as cotton and soybean (Luo *et al.* 2008; Neelakandan *et al.* 2009) significant differences in expression levels of the SMT2 isoforms have been identified. In cotton expression profiles of the two *GhSMT2* genes in different organs and at various stages of fibre development showed that the expression level of *GhSMT2-1* was10 times higher than *GhSMT2-2* in all organs and tissues detected(Luo *et al.* 2008). Similarly, the expression of all the identified olive *SMT2* and *SMO2* alleles should be analysed to determine differences in expression between the alleles and/or cultivars.

- In this study we assessed the expression profiles of *SMO2* and *SMT2* genes in different olive cultivars during different developmental stages of the olive fruit. This expression study is just a preliminary study to get an overview of the expression profile of these genes in olives and is limited to determining the mRNA levels of these genes in different olive cultivarsbut the impact of these differences at protein level in the mature fruitis yet to be identified. Thus in order to understand the functional role of *SMO2* and *SMT2* genes in olives, olive genotypes lacking these specific alleles should be used and subsequently the sterol levels in the plants should be analysed to understand the impact of these enzymes on the olive oil. Unlike *Arabidopsis* and tobaccomutants showing lower campesterol levels in overexpressed *SMT2* transgenic mutants, in soybean a set of transgenic *Arabidopsis* plants harbouring seed targeted overexpression of *GmSMT2-2* showed an increase in

sitosterol and stigmasterol levels while campesterol levels were more or less same(Neelakandan *et al.* 2010). Since the sitosterol:campesterol ratio were not drastically altered it seemed likely that the campesterol levels were tightly regulated in soybean seed tissues implying that the role of SMT2 in controlling the levels of campesterol differs between plant species. Therefore it will be interesting to see what impact does *SMT2* and *SMO2* knockdown mutants have on the sterol levels in olive plants.

- The most noteworthy difference observed was the dramatic upregulation of *SMO2* at the maturation stage of the olive fruit (136 DAF) in Picual. This may indicate that in Picual at later timepoints, when maximum amount of oil gets deposited in the olive fruit, *SMO2* expression increases significantly. Our results in this study is interesting as Australian olive oils extracted from the cultivar Picual have shown to contain comparatively lower levels campesterol (3.53%) while olive oils extracted from the cultivar Barnea have shown consistently higher campesterol levels (4.8%) (Section 3.9). Thus, though we found significant differences in *SMO2* expression between the olive cultivars, we would expect *SMO2* to be overexpressed in Barnea to have higher campesterol levels. However in our study in Barnea, no expected upregulation of *SMO2* was observed. To explain such higher campesterol levels observed in Barnea oils, we may need to further investigate other genes that are downstream of SMO2 enzyme in the sterol biosynthetic pathway (Figure 1.14) such as STE1/DWF7, DWF-5 and DIM. These enzymes are shared between the campesterol and sitosterol pathways, so any change in their activity will affect both products, however it would be interesting to see if there are any differences in the expressionof these downstream genes between Barnea, Frantoio and Picual. It is critical to check the expression of BR-biosynthetic enzymes such as CYPs in olives as this enzyme catalyses the conversion of end-product campesterol to brassicasterol. It is logical to think that if this enzyme is significantly downregulated in Barnea as compared to other cultivars, it might lead to excessive accumulation of campesterol in this cultivars.

REFERENCES

Alagna, F., N. D'Agostino, L. Torchia, M. Servili, R. Rao, M. Pietrella, G. Giuliano, M. L. Chiusano, L. Baldoni and G. Perrotta (2009). "Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development." *BMC Genomics***10**(399): 1-15.

Alonso, A. and M. A. Martinez-Gonzalez (2004). "Olive oil consumption and reduced incidence of hypertension: the SUN study." *Lipids***39**: 1233-1238.

Alonso, A., V. Ruiz-Gutierrez and M. A. Martinez-Gonzalez (2006). "Monounsaturated fatty acids, olive oil and blood pressure: epidemiological, clinical, and experimental evidence." *Public Health Nutrition***9**(2): 251-257.

Andersen, C. L., J. L. Jensen and T. F. Ørntoft (2004). "Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets." *The Journal of Cancer Research***64**: 5245-5250.

Aparicio, R. and G. Luna (2002). "Characterization of monovarietal virgin olive oils." *European Journal of Lipid Science and Technology***104**: 614-627.

Aparicio, R., M. T. Morales and V. Alonso (1997). "Authentication of European virgin olive oils by their chemical compounds, sensory attributes, and consumers' attitudes." *Journal of Agricultural and Food Chemistry***45**: 1076-1083.

Artico, S., S. M. Nardeli, O. Brilhante, M. F. Grossi-de-Sa and M. Alves-Ferreira (2010). "Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data." *BMC Plant Biology***10**(49): 1-12.

Awad, B. A., A. Downie, C. S. Fink and U. Kim (2000). "Dietary phytosterol inhibits the growth and metastasis of MDA-MB-231 human breast cancer cells grown in SCID mice." *Anticancer Research***20**(2A): 821-824.

Ayton, J., R. J. Mailer, A. Haigh, D. Tronson and D. Conlan (2007). "Quality and oxidative stability of Australian olive oil according to harvest date and irrigation." *Journal of Food Lipids***14**: 138-156.

Bard, M., D. A. Bruner, C. A. Pierson, N. D. Lees, B. Biermann, L. Frye, C. Koegel and R. Barbuch (1996). "Cloning and characterization of *ERG25*, the *Saccharomyces cerevisiae* gene encoding C-4 sterol methyl oxidase." *Proceedings of the National Academy of Sciences of the United States of America***93**: 186-190.

Beardsell, D., J. Francis and D. Ridley (2001). "Health promoting constituents in plant derived edible oils." *Journal of Food Lipids***9**: 1-34.

Bennett, M. D. and J. B. Smith (1976). "Nuclear DNA amounts in Angiosperms." *Annals of Botany***2749**(933): 227-274.

Benveniste, P. (2002). Sterol metabolism. The *Arabidopsis* Book, American Society of Plant Biologists**:** 1-31.

Benveniste, P. (2004). "Biosynthesis and accumulation of sterols." *Annual Review of Plant Biology***55**: 429-457.

Berger, A., P. J. Jones and S. S. Abumweis (2004). "Plant sterols: factors affecting their efficacy and safety as functional food ingredients." *Lipids in Health and Disease***3**(5): 1-19.

Bohacenko, I. and Z. Kopicova (2001). "Detection of olive oils authenticity by determination of their sterol content using LC/GC." *Czech Republic Journal of Food Science***19**(3): 97-103.

Boskou, D. (2006). Olive Oil: Chemistry and Technology, 1st ed. Illinois, AOCS (American Oil's Chemist's Society) Publishing.

Boskou, D. (2006). Olive oil: Chemistry and technology, 2nd ed. Illinois, AOCS (American Oil's Chemist's Society) Publishing.

Boutte´, Y. and M. Grebe (2009). "Cellular processes relying on sterol function in plants." *Current Opinion in Plant Biology***12**: 705-713.

Bouvier-Navé, P., T. Husselstein and P. Benveniste (1998). "Two families of sterol methyltransferases are involved in the first and the second methylation steps of plant sterol biosynthesis." *European Journal of Biochemistry***256**: 88-96.

Bouvier-Navé, P., T. Husselstein, T. Desprez and P. Benveniste (1997). "Identification of cDNAs encoding sterol methyl-transferases involved in the second methylation step of plant sterol biosynthesis." *European Journal of Biochemistry***246**: 518-529.

Bradford, P. G. and A. B. Awad (2007). "Phytosterols as anticancer compounds." *Molecular Nutrition and Food Research***51**: 161-170.

Bustin, S. A., V. Benes, J. A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele and C. T. Wittwer (2009). "The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments." *Clinical Chemistry***55**(4): 611-622.

Bustos, L. (2010). "Work to begin on new Australian Standard for the olive oil industry: Media release: 1 July, 2010." Standards Australia. Retrieved 28th January, 2013, from http://www.standards.org.au/OurOrganisation/News/Documents/100701_Work%20to%20start%20on%20Olive%20Oil%20Standard.pdf.

Calpe-Berdiel, L., J. C. Escolà-Gil, S. Benítez, C. Bancells, F. González-Sastre, X. Palomer and F. Blanco-Vaca (2007). "Dietary phytosterols modulate T-helper immune response but do not induce apparent anti-inflammatory effects in a mouse model of acute, aseptic inflammation." *Life Sciences***80**(21): 1951-1956.

Carland, F., S. Fujioka and T. Nelson (2010). "The sterol methyltransferases SMT1, SMT2, and SMT3 influence *Arabidopsis* development through nonbrassinosteroid products." *Plant Physiology***153**: 741-756.

Ceci, L. N. and A. A. Carelli (2007). "Characterization of monovarietal Argentinian olive oils from new productive zones." *Journal of the American Oil Chemists' Society* **84**: 1125-1136.

Chan, A. P., J. Crabtree, H. Lorenzi, Q. Zhao, J. Orvis, D. Puiu, A. Melake-Berhan, K. M. Jones, J. Redman, G. Chen, E. B. Cahoon, M. Gedil, M. Stank, B. J. Haas, J. R. Wortman, C. M. Fraser-Liggett, J. Ravel and P. D. Rabinowicz (2010). "Draft genome sequence of the oilseed species *Ricinus communis*." *Nature Biotechnology***28**(9): 951-956.

Choudhury, N., L. Tan and A. S. Truswell (1995). "Comparison of palmolein and olive oil: effects on plasma lipids and vitamin E in young adults." *American Journal of Clinical Nutrition***61**: 1043-1051.

Chung, H. Y., S. Fujioka, S. Choe, S. Lee, Y. H. Lee, N. I. Baek and I. S. Chung (2010). "Simultaneous suppression of three genes related to brassinosteroid (BR) biosynthesis altered campesterol and BR contents, and led to a dwarf phenotype in *Arabidopsis thaliana*." *Plant Cell Reports***29**(4): 397-402.

Clouse, S. D. (2002). "*Arabidopsis* mutants reveal multiple roles for sterols in plant development." *American Society of Plant Biologists***14**(9): 1995-2000.

Covas, M. I., V. Ruiz-Gutierrez, R. Torre, A. Kafatos, R. M. Lamuela-Raventos, J. Osada, R. W. Owen and F. Visioli (2006). " Minor components of olive oil: Evidence to date of health benefits in humans." *Nutrition Reviews***64**: S20-S30.

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

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

Darnet, S., M. Bard, A. Hoeft and A. Rahier (2004). "Identification of two novel families of membrane-bound non heme oxygenases involved in plant sterol biosynthesis: A VIGS approach." *Institut de Biologie Moléculaire des Plantes*.

Darnet, S., M. Bard and A. Rahier (2001). "Functional identification of sterol-4α-methyl oxidase cDNAs from *Arabidopsis thaliana* by complementation of a yeast *ERG25* mutant lacking sterol-4α-methyl oxidation." *FEBS Letters***508**: 39-43.

Darnet, S. and A. Rahier (2003). "Enzymological properties of sterol-C4-methyl-oxidase of yeast sterol biosynthesis." *Biochimica et Biophysica Acta***1633**: 106-117.

Darnet, S. and A. Rahier (2004). "Plant sterol biosynthesis: Identification of two distinct families of sterol 4α-methyl oxidases." *Biochemical Society***378**: 889-898.

De Stefani, E., P. Brennan, P. Boffeta and A. L. Rono (2000). "Vegetables, fruits, related dietary antioxidants, and the risk of squamous cell carcinoma of the esophagus: a case-control study in Uruguay." *Nutrition and Cancer***38**: 23-29.

Diener, A. C., H. Li, W. Zhou, W. J. Whoriskey, D. W. Nes and G. R. Fink (2000). "Sterol methyl transferase-1 controls the level of cholesterol in plants." *The Plant Cell***12**: 853-870.

Dourtoglou, V. G., T. Dourtoglou, A. Antonopoulos, E. Stefanou, S. Lalas and C. Poulos (2003). "Detection of olive oil adulteration using principal component analysis applied on total and region FA content." *Journal of the American Oil Chemists' Society***80**(3): 203-208.

Dündar, E., Ö,Suakar, T. Unver and A, Dagdelen (2013)."Isolation and expression analysis of cDNAs that are associated with alternate bearing in *Olea europaea* L. cv. Ayvalık."*BMC Genomics***14**(219):1-14.

Engel, R. and H. Schuert (2005). "Formulation of phytosterols in emulsions for increased dose response in functional foods." *Innovative Food Science and Emerging Technologies***5**: 233-237.

Estruch, R., M. Martinez-Gonzalez, D. Corella, J. Salas-Salvado, V. Ruiz-Gutierrez and M. I. Covas (2006). "Effects of a Mediterranean style diet on cardiovascular risk factors: a randomized trial." *Annals of Internal Medicine* **145**: 1-11.

Expósito-Rodríguez, M., A. A. Borges, A. Borges-Pérez and J. A. Pérez (2008). "Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process." *BMC Plant Biology***1**(131): 1-12.

Fernandes, P. and J. M. S. Cabral (2007). "Phytosterols: applications and recovery methods." *Bioresource Technology***98**: 2335-2350.

Ferragut, V. (2007). The toxic oil syndrome in Spain. Case Studies in Food Safety and Environmental Health, Springer. **6:** 43-51.

Ferrara, L., A. S. Raimondi, L. d'Episcopo, L. Guida, A. Dello Russo and T. Marotta (2000). "Olive oil and reduced need for antihypertensive medications." *Archives of Internal Medicine* **160**: 837-842.

Field, S. (2009). "Olive business." Retrieved 28th January, 2013, from http://www.olivebusiness.com/.

Firestone, D. (2001). "Assuring the integrity of olive oils." *Journal of AOAC International***84**(1): 176-180.

Fito, M., M. Cladellas, R. de la Torre, J. Martı, D. Mun˜oz, H. Schro¨der, M. Alca´ntara, M. Pujadas-Bastardes, J. Marrugat, M. Lo´pez-Sabater, J. Bruguera and M. I. Covas (2007). "Anti-inflammatory effect of virgin olive oil in stable coronary disease patients: a

randomized, crossover, controlled trial." *European Journal of Clinical Nutrition***62**(4): 570-574.

Foxhall, L. (2007). Olive cultivation in ancient Greece: seeking the ancient economy, Oxford University Press.

Frankel, E. N. (2010). "Chemistry of extra virgin olive oil: Adulteration, oxidative stability and antioxidants." *Journal of Agricultural and Food chemistry***58**: 5991-6006.

Gamm, M., M. Héloir, J. Kelloniemi, B. Poinssot, D. Wendehenne and M. Adrian (2011). "Identification of reference genes suitable for qRT-PCR in grapevine and application for the study of the expression of genes involved in pterostilbene synthesis." *Molecular Genetics and Genomics***285**(4): 273-285.

García-Segovia, P., A. Sánchez-Villegas, J. Doreste, F. Santana and L. Serra-Majem (2006). "Olive oil consumption and risk of breast cancer in the Canary Islands: a population-based case-control study." *Public Health Nutrition***9**: 163-167.

Georgio, B. (2007). "OLEA Database." Retrieved 28th January, 2013, from http://www.oleadb.it/.

Grebenok, R. J., D. W. Galbraith and D. D. Penna (1997). "Characterization of *Zea mays* endosperm C-24 sterol methyltransferase: one of two types of sterol methytransferase in higher plants." *Plant Molecular Biology***34**: 891-896.

Gu, Y., M. Li, K. Zhang, L. Chen, A. Jiang, J. Wang and X. Li (2011). "Evaluation of endogenous control genes for gene expression studies across multiple tissues and in the specific sets of fat and muscle-type samples of the pig." *Journal of Animal Breeding and Genetics***128**: 319-325.

Guillaume, C., L. Ravetti, D. L. Ray and J. Johnson (2011). "Technological Factors Affecting Sterols in Australian Olive Oils." *Journal of the American Oil Chemists' Society* **89**(1): 29-39.

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

Hall, T. A. (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." *Nucleic Acids Symposium Series***41**: 95-98.

Halvadakis, C. P. and M. Niaounakis (2006). Olive processing waste management: Literature review and patent survey, Elsevier.

Harwood, J. L. and R. Aparicio (2000). Handbook of olive oil: Analysis and properties. New York, Springer.

Harwood, J. L. and P. Yaqoob (2002). "Nutritional and health aspects of olive oil." *European Journal of Lipid Science and Technology***104**: 685-697.

Hearn, S. (2002). R&D plan for the Australian olive industry 2003-2008, Rural Industries Reseach and Developmental Corporation.

Hellemans, J., G. Mortier, A. D. Paepe, F. Speleman and J. Vandesompele (2007). "qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data." *Genome Biology***8**(2).

Hoenemann, C. and A. Hohe (2011). "Selection of reference genes for normalization of quantitative real-time PCR in cell cultures of *Cyclamen persicum*." *Electronic Journal of Biotechnology***14**(1): 1-8.

Holmberg, N., M. Harker, C. L. Gibbard, A. D. Wallace, J. C. Clayton, S. Rawlins, A. Hellyer and R. Safford (2002). "Sterol C-24 methyltransferase type 1 controls the flux of carbon into sterol biosynthesis in tobacco weed." *Plant Physiology* **130**: 303-311.

Hong, S., P. Seo, M. Yang, F. Xiang and C. Park (2008). "Exploring valid reference genes for gene expression studies in *Brachypodium distachyon* by real-time PCR." *BMC Plant Biology***8**(112): 1-11.

Hu, R., C. Fan, H. Li, Q. Zhang and Y. Fu (2009). "Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR." *BMC Molecular biology***10**(93): 1-12.

Husselstein, T., D. Gachotte, T. Desprez, M. Bard and P. Benveniste (1996). "Transformation of *Saccharomyces cerevisiae* with a cDNA encoding a sterol C-methyltransferase from *Arabidopsis thaliana* results in the synthesis of 24-ethyl sterols." *FEBS (Federation of European Biochemical Societies) Letters***381**: 87-92.

International Olive Oil Council (2003). Trade standard applying to olive oils and olive pomace oils. COI/T.15/NC no. 3/Rev. 1. Madrid, Spain.

International Olive Oil Council (2006). Trade standard applying to olive oils and olive pomace oils. COI/T.15/NC no.3/Rev. 2 Madrid, Spain.

International Olive Oil Council (2009). Trade standard applying to olive oils and olive pomace oils. COI/T.15/NC No 3/Rev. 4. Madrid, Spain.

International Olive Oil Council (2010). Olive product market report summary No 35.

International Olive Oil Council (2010). Olive product market report summary No 39.

International Olive Oil Council (2011). Trade standard applying to olive oils and olive pomace oils. COI/T.15/NC No 3/Rev. 6. Madrid, Spain.

Jain M, Nijhawan A, Tyagi AK and K. JP. (2006). "Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR." *Biochemical and Biophysical Research Communications***345**: 646-651.

Jang, M. S., K. S. Han and S. K. Kim (2000). "Identification of brassinosteroids and their biosynthetic precursors from seeds of pumpkin." *Bulletin of the Korean Chemical Society***21**(2): 161-164.

Jarošová, J. and J. K. Kundu (2010). "Validation of reference genes as internal control for studying viral infections in cereals by quantitative real-time RT-PCR." *BMC Plant Biology***10**(146): 1-9.

Jian, B., B. Liu, Y. Bi, W. Hou, C. Wu and T. Han (2008). "Validation of internal control for gene expression study in soybean by quantitative real-time PCR." *BMC Molecular biology***9**(59): 1-14.

Jiménez-Escrig, A., A. B. Santos-Hidalgo and F. Saura-Calixto (2006). "Common sources and estimated intake of plant sterols in the Spanish diet." *Journal of Agricultural and Food Chemistry***54**: 3462-3471.

Kailis, S. and D. Harris (2007). Producing table olives. Australia, CSIRO Publishing.

Karlen, Y., A. McNair, S. Perseguers, C. Mazza and N. Mermod (2007). "Statistical significance of quantitative PCR." *BMC Bioinformatics***8**(131): 1-16.

Katan, M. B., S. M. Grundy and P. Jones (2005). "Efficacy and safety of plant stanols and sterols in the management of blood cholesterol levels." *Nutrition Research***25**: 413-428.

Kiritsakis, K. A. (1999). Composition of olive oil and its nutritional and health effect. Retrieved from http://www.regional.org.au/au/gcirc/1/205.htm#. Greece, The Regional Institute Ltd.

Koutsaftakis, A., F. Kotsifaki and E. Stefanoudaki (1999). "Effect of extraction system, stage of ripeness and kneading temperature on the sterol composition of virgin olive oils." *Journal of American Oil Chemist's Society***76**(12): 1477-1481.

Krichene, D., A. Allalout, B. Baccouri, G. Q. Fregapane, M. D. Salvador and M. Zarrouk (2009). "Territorial investigation based on the chemical composition of chemlali virgin olive oils." *Asian Journal of Biochemistry***4**(1): 1-12.

Kris-Etherton, P. M., T. A. Pearson, Y. Wan, Rebecca L Hargrove, K. Moriarty, V. Fishell and T. D. Etherton (1999). "High-monounsaturated fatty acid diets lower both plasma cholesterol and triacylglycerol concentrations." *American Society for Clinical Nutrition***70**(6): 953-954.

Kritchevsky, D. and S. C. Chen (2005). "Phytosterol-health benefits and potential concerns: A review." *Nutrition Research***25**: 413-428.

Kuijk, E., L. Puy, H. Tol, H. Haagsman, B. Colenbrander and B. Roelen (2007). "Validation of reference genes for quantitative RT-PCR studies in porcine oocytes and preimplantation embryos." *BMC Developmental Biology***7**(58): 1-8.

Landa, M., N. Frago and A. Tres (1994). "Diet and the risk of breast cancer in Spain." *European Journal of Cancer***3**: 313-320.

Lazzara, P. and G. Rana (2010). The crop coefficient (Kc) values of the major crops grown under Mediterranean climate. Retrieved from http://www.miraproject.eu/MELIA/Members/p35isa/CropCoeff.pdf/view. CRA- Research Unit for Agricultural in Dry Environments. Italy**:** 1-28.

Long, X. Y., Y. X. Liu, H. Rocheleau, T. Ouellet and G. Y. Chen (2011). " Identification and validation of internal control genes for gene expression in wheat leaves infected by strip rust." *International Journal of Plant Breeding and Genetics***5**(3): 255-267.

López-Miranda, J., F. Pérez-Jiménez, E. Ros, R. De Caterina, L. Badimón, M. Covas, E. Escrich and J. Ordovás (2010). "Olive oil and health: summary of the II international conference on olive oil and health consensus report, Jaen and Cordoba (Spain) 2008." *Nutrition, Metabolism, and Cardiovascular Diseases* **20**(4): 284-294.

Loureiro, J., E. Rodriguez, J. Doležel and C. Santos (2007). "Two new nuclear isolation buffers for plant DNA flow cytometry: A test with 37 species." *Annals of Botany***100**(4): 875-888.

Luo, M., K. Tan, Z. Xiao, M. Hu, P. Liao and K. Chen (2008). "Cloning and expression of two sterol C-24 methyltransferase genes from upland cotton (*Gossypium hirsuturm L.*)." *Journal of Genetics and Genomics***35**: 357-363.

Luo, M., Y. Xiao, X. Li, X. Lu, W. Deng, D. Li, L. Hou, M. Hu, Y. Li and Y. Pei (2007). "GhDET2, a steroid 5alpha-reductase, plays an important role in cotton fiber cell initiation and elongation." *Plant Journal***51**: 419-430.

Mailer, R. (2005). "Variation in oil quality and fatty acid composition in Australian olive oil." *Australian Journal of Experimental Agriculture***45**: 115-119.

Mailer, R. (2007). Australian olive oil and international market. International News on Fats, Oils and Related Materials. Australia.

Mailer, R. (2007). The natural chemistry of Australian extra virgin olive oil, Rural Research and Development Corporation (RIRDC).

Mailer, R. (2008). Olive oil chemistry in Australia: Analysis and quality parameters. 99th AOCS Annual Meeting & Expo, Seattle, Washington USA, American Oil Chemists' Society.

Mailer, R. (2012). Cultivation of Olives in Australia. Olive oil - constituents, quality, health properties and bioconversions. D. Boskou. Australia, InTech.

Mailer, R. and J. Ayton (2008). A survey of Australian olive cultivars to determine compliance with international standards, Rural Industries Research and Development Corporation (RIRDC).

Mailer, R. and C. Beckingham (2006). "Testing olive oil quality: chemical and sensory methods." *Primefacts***231**: 1-5.

Maroufi, A., E. V. Bockstaele and M. D. Loose (2010). "Validation of reference genes for gene expression analysis in chicory (*Cichorium intybus*) using quantitative real-time PCR." *BMC Molecular biology***11**(15): 1-12.

Martin-Moreno, J., W. Willett, L. Gorgojo, J. Banegas, F. Rodriguez-Artalejo, J. Fernandez-Rodriguez, P. Maisonneuve and P. Boyle (1994). "Dietary fat, olive oil intake and breast cancer risk." *International Journal of Cancer* **58**: 774-780.

Mendilaharsu, M., E. De Stefani, H. Deneo-Pellegrini, J. Carzoglio and A. Ronco (1998). "Phytosterols and risk of lung cancer: a case-control study in Uruguay." *Lung Cancer***21**: 37-45.

Miettinen, T. A., P. Puska, H. Gylling, H. Vanhanen and E. Vartiainen (1995). "Reduction of serum cholesterol with sitostanol-ester margarine in a mildly hypercholesterolemic population." *The New England Journal of Medicine***333**: 1308-1312.

Moore, J. T. and J. L. Gaylor (1969). "Isolation and purification of an S-adenosylmethionine: delta-24 sterol methyltransferase from yeast." *The Journal of Biological Chemistry***244**(33): 3334-3340.

Murray, M. G. and W. F. Thompson (1980). "Rapid isolation of high molecular weight plant DNA." *Nucleic Acids Research***8**(19): 4321-4325.

Nashed, B., B. Yeganeh, K. T. HayGlass and M. H. Moghadasian (2005). "Antiatherogenic effects of dietary plant sterols are associated with inhibition of proinflammatory cytokine production in apo E-KO mice." *The Journal of Nutrition***135**(10): 2438-2444.

Neelakandan, A. K., H. T. Nguyen, R. Kumar, L. S. Tran, S. K. Guttikonda, T. N. Quach, D. L. Aldrich and W. D. Nes (2010). "Molecular characterization and functional analysis of *Glycine max sterol methyl transferase 2* genes involved in plant membrane sterol biosynthesis." *Plant Molecular Biology***74**(4-5): 503-518.

Neelakandan, A. K., Z. Song, J. Wang, M. H. Richards, X. Wu, B. Valliyodan, H. T. Nguyen and W. D. Nes (2009). "Cloning, functional expression and phylogenetic analysis of plant sterol 24C-methyltransferases involved in sitosterol biosynthesis." *Phytochemistry***70**(17-18): 1982-1998.

Nes, W. D. (2000). "Sterol methyl transferase: enzymology and inhibition." *Biochem Biophys Acta***1529**(1-3): 63-88.

Nes, W. D., P. Jayasimha, W. Zhou, R. Kanagasabai, C. Jin, T. T. Jaradat, R. W. Shaw and J. M. Bujnicki (2004). "Sterol methyltransferase: Functional analysis of highly conserved residues by site-directed mutagenesis." *Biochemistry***43**: 569-576.

Niaounkis, M. and C. P. Halvadakis (1992). Olive processing waste management: Literature review and patent survey, 2nd ed. Greece, Elsevier.

Nicolaiew, N., N. Lemort and L. Adorni (1998). "Comparison between extra virgin olive oil and oleic acid rich sunflower oil: effects on postprandial lipemia and LDL susceptibility to oxidation." *Annals of Nutrition and Metabolism* **42**: 251-260.

Nicot, N., J. F. Hausman, L. Hoffmann and D. Evers (2005). "Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress." *Journal of Experimental Botany***56**(421).

Normén, L. A., H. A. M. Brants, L. E. Voorrips, H. A. Andersson, P. A. Brandt and R. A. Goldbohm (2001). "Plant sterol intakes and colorectal cancer risk in the Netherlands cohort study on diet and cancer." *American Journal of Clinical Nutrition***74**: 141-148.

O' Brien, P. (2008). Guide to efficient olive harvesting, Rural Industries Research and Development Corporation (Australia).

Oger, E., S. Ghignone, E. Campagnac, J. Fontaine, A. Grandmougin-Ferjani and L. Lanfranco (2009). "Functional characterization of a C-4 sterol methyl oxidase from the endomycorrhizal fungus *Glomus intraradices*." *Fungal Genetics and Biology***46**: 486-495.

Olson, P. E. (2006). Tasting and classifying virgin olive oil-An international course for panel supervisors. University of Imperia, Italy, Rural Industries Research and Development Corporation.

Ozen, B. F. and L. J. Mauer (2002). "Detection of hazelnut oil adulteration using FT-IR spectroscopy." *Journal of Agricultural and Food chemistry.***50**(14): 3898-3901.

Pardo, J., M. lvarez-Ortı, A. M. Rey, M. Cuesta and A. Silvestre (2011). "Purity, quality and stability of Argentinean virgin olive oils." *European Journal of Lipid Science and Technology***113**: 597-608.

Parker, J. N. and P. M. Parker (2004). Olive oil-A medical dictionary, bibliography, and annotated research guide to internet references. United States, ICON Group International.

Pascal, S., M. Taton and A. Rahier (1993). "Plant sterol biosynthesis: Identification and characterization of two distinct microsomal oxidative enzymatic systems involved in sterol C4-demethylation." *The Journal of Biological Chemistry***268**(16): 11639-11654.

Patel, M. D. and P. D. Thompson (2006). "Phytosterols and vascular disease." *Artherosclerosis***186**: 12-19.

Pegel, K. H. (1997). "The importance of sitosterol and sitosterolin in human and animal nutrition." *South African Journal of Science***93**: 263-268.

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

Piironen, V., D. G. Lindsay, T. A. Miettinen, J. Toivo and A.-M. Lampi (2000). "Plant sterols: biosynthesis, biological function and their importance to human nutrition." *Journal of the Science of Food and Agriculture***80**: 939-966.

Pollak, O. (1953). "Reduction of blood cholesterol in man." *Circulation***7**: 702-706.

Preedy, V. R. (2010). Olives and olive oil in health and disease prevention. United States, Elsevier.

Psaltopoulou, T., R. I. Kosti, D. Haidopoulos, M. Dimopoulos and D. B. Panagiotakos (2011). "Olive oil intake is inversely related to cancer prevalence: a systematic review and a meta-analysis of 13800 patients and 23340 controls in 19 observational studies." *Lipids in Health and Disease***10**(127): 1-16.

Purkayastha, A. and I. Dasgupta (2009). "Virus-induced gene silencing: A versatile tool for discovery of gene functions in plants." *Plant Physiology and Biochemistry***47**(11-12): 967-976.

Quiles, J. L., C. Ramirez-Tortosa and P. Yaqoob (2006). Olive oil and health. United States, CAB International.

Ranalli, A., G. De Mattia and M. L. Ferrante (1997). "Evaluating a new olive oil processing method." *International Journal of Food Science and Technology***32**: 289-297.

Ranalli, A., L. Pollastri, S. Contento, G. D. Loreto, E. lannucci, L. Lucera and F. Russi (2002). "Sterol and alcohol components of seed, pulp and whole olive fruit oils. Their use to characterise olive fruit cultivar by multivariates." *Journal of the Science of Food and Agriculture***82**(8): 854-859.

Ravetti, L. (2006). Horticultural review of Australian olive industry. Canberra, Australian Olive Association Field Days.

Ravetti, L., C. Guillaume and J. Johnson (2010). Sterols in Australian olive oils: The effects of biological and technological factors. R. I. R. a. D. Corporation.

Ravetti, L., C. Guillaume and J. Johnson (2010). Sterols in Australian olive oils: The effects of biological and technological factors, Rural Industries Research and Developmental Corporation.

Reid, K., N. Olsson, J. Schlosser, F. Peng and S. Lund (2006). "An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development." *BMC Plant Biology***14**(6): 1-11.

Ronco, A., E. De Stefani, P. Boffetta and H. Deneo-Pellegrini (1999). "Vegetables, fruits, and related nutrients and risk of breast cancer." *Nutrition and Cancer***35**: 111-119.

Rudzinska, M., W. Uchman and E. Wasowicz (2005). "Plant sterols in food technology." *ACTA Scientarium Polonorum Technology***4**(1): 147-156.

Ruiz-Canela, M. and M. A. Martínez-González (2011). "Olive oil in the primary prevention of cardiovascular disease." *Maturitas***68**: 245-250.

Rural Industries Reseach and Developmental Corporation (2010). Australian olive industry research, development &extention plan 2010-2015.

Sakouhi, F., C. Absalon, S. Harrabi, C. Vitry, K. Sebei, S. Boukhchina, E. Fouquet and H. Kallel (2009). "Dynamic accumulation of 4-desmethylsterols and phytostanols during ripening of Tunisian Meski olives (*Olea europea L.*)." *Food Chemistry***112**(4): 897-902.

Salvador, M. D., F. Aranda, G. S. Alonso and G. Fregapane (2001). "Cornicabra virgin olive oil: A study of five crop seasons. Composition, quality and oxidative stability." *Food Chemistry***74**: 267-274.

Salvador, M. D., F. Aranda, S. Gómez-Alonso and G. Fregapane (2003). "Influence of extraction system, production year and area on Cornicabra virgin olive oil: a study of five crop seasons " *Food Chemistry***80**: 359-366.

Sambrook, J., E. Fritsch and T. Maniatis (1989). Molecular cloning: a laboratory manual, 2nd ed. New York, Cold Spring Harbor Laboratory Press.

Sanger, F., S. Nicklen and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors." *Proceedings of the National Academy of Sciences of the United States of America***74**(12): 5463-5467.

Schaeffer, A., P. Bouvier-Navé, P. Benveniste and H. Schaller (2000). "Plant sterol-C24-methyl transferases: Different profiles of tobacco transformed with SMT1 or SMT21." *Lipids***35**(3): 1-7.

Schaeffer, A., R. Bronner, P. Bebveniste and H. Schaller (2001). "The ratio of campesterol to sitosterol that modulates growth in Arabidopsis is controlled by sterol methyltransferase 2;1." *The Plant Journal***25**(6): 605-615.

Schaller, H. (2004). "New aspects of sterol biosynthesis in growth and development of higher plants." *Plant Physiology and Biochemistry***42**: 465-476.

Scharlaken, B., D. C. D. Graaf, K. Goossens, M. Brunain, L. J. Peelman and F. J. Jacobs (2008). "Reference gene selection for insect expression studies using quantitative real-time PCR: The head of the honeybee, *Apis mellifera*, after a bacterial challenge." *Journal of Insect Science***8**(33): 1-10.

Schmittgen, T. D. and B. A. Zakrajsek (2000). "Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR." *Journal of Biochemistry and Biophysiology***46**(1-2): 69-81.

Schrick, K., U. Mayer, G. Martin, C. Bellini, C. Kuhnt, J. Schmidt and G. Ju¨ rgens (2002). "Interactions between sterol biosynthesis genes in embryonic development of *Arabidopsis*." *Plant Journal***31**: 61-73.

Schroeder, A., O. Mueller, S. Stocker, R. Salowsky, M. Leiber, M. Gassmann, S. Lightfoot, W. Menzel, M. Granzow and T. Ragg (2006). "The RIN: an RNA integrity number for assigning integrity values to RNA measurements." *BioMedCentral Molecular Biology***7**(3): 1-14.

Shahtahmasebi, S. and S. Shahtahmasebi (2003). "A case report of possible health benefits of extra virgin olive oil." *The Scientific World***3**: 1265-1271.

Sheppard, S. (2008). Consumer attitudes to Australian extra virgin olive oil, Rural Industries Research and Development Corporation.

Shi, J., R. A. Gonzales and M. K. Bhattacharyya (1996). "Identification and characterization of an S-adenosyl-L-methionine: Δ24-sterol-C-methyltransferase cDNA from soybean." *The American Society for Biochemistry and Molecular Biology***271**: 9384-9389.

Sierksma, A., J. A. Weststrate and G. W. Meijer (1999). "Spreads enriched with plant sterols, either esterified 4,4-dimethylsterols or free 4-desmethylsterols, and plasma total and LDL-cholesterol concentrations." *British Journal of Nutrition***82**: 273-282.

Silveira, É. D., M. Alves-Ferreira, L. A. Guimarães, F. R. d. Silva and V. T. Carneiro (2009). "Selection of reference genes for quantitative real-time PCR expression studies in the apomictic and sexual grass *Brachiaria brizantha*." *BMC Plant Biology***9**(84): 1-10.

Song, Z. and W. D. Nes (2007). "Sterol biosynthesis inhibitors: Potential for transition state analogs and mechanism-based inactivators targeted at sterol methyltransferase." *Lipids***42**: 15-33.

Souter M, Topping J, Pullen M, Friml J, Palme K, Hackett R, Grierson D and L. K (2002). "Hydra mutants of *Arabidopsis* are defective in sterol profiles and auxin and ethylene signaling." *Plant Cell***14**: 1017-1031.

Stark, A. H. and Z. Madar (2002). "Olive oil as a functional food: epidemiology and nutritional approaches." *Nutrition Reviews***60**(6): 170-176.

Stefani, E. D., P. Boffetta, A. L. Ronco, P. Brennan, H. Doneo-Pellegrini, J. C. Carzoglio and M. Mendilaharsu (2000). "Plant sterols and risk of stomach cancer: A case-control study in Uruguay." *Nutrition and Cancer***37**: 140-144.

Tay, A., R. K. Singh, S. S. Krishnan and J. P. Gore (2002). "Authentication of olive oil adulterated with vegetable oils using fourier transform infrared spectroscopy." *Lebensmittel-Wissenschaft und-Technologie***35**(1): 99-103.

Therios, I. (2008). Olives. Greece, CAB (Centre for Agricultural Bioscience) International.

Trautwein, E. A. and I. Demonty (2007). "Phytosterols: natural compounds with established and emerging health benefits." *Dossier***14**: 259-265.

Uddin, M. J., M. U. Cinar, D. Tesfaye, C. Looft, E. Tholen and K. Schellander (2011). "Age-related changes in relative expression stability of commonly used housekeeping genes in selected porcine tissues." *BioMedCentral Research Notes***4**(441): 1-13.

Vandesompele, J., K. D. Preter, F. Pattyn, B. Poppe, N. V. Roy, A. D. Paepe and F. Speleman (2002). "Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes." *Genome Biology***3**(7): 1-12.

Vincent-Baudry, S., C. Defoort, M. Gerber, M.-C. Bernard, P. Verger, O. Helal, H. Portugal, R. Planells, P. Grolier, M.-J. Amiot-Carlin, P. Vague and D. Lairon (2005). "The Medi-RIVAGE study: reduction of cardiovascular disease risk factors after a 3-mo intervention with a Mediterranean-type diet or a low-fat diet." *American Society for Clinical Nutrition***82**(5): 964-971.

Vossen, P. (2007). "Olive oil: History, production, and characteristics of the world's classic oils." *American Society for Horticultural Science***42**(5): 1093-1100.

Wang, F., Y. Zhao, M. Dai, G. Zheng, J. Liu, Z. Ren and J. He (2008). "Cloning and functional identification of C-4 methyl sterol oxidase genes from the penicillin-producing fungus *Penicillium chrysogenum*." *FEMS Microbiology Letters***287**: 91-99.

Webster, L., P. Simpson and A. M. Shanks (2001). Adulteration of olive oil with hazelnut oil: to enable detection of unrefined and refined hazelnut oil in virgin and refined olive oil. United Kingdom, Fisheries Research Services**:** 1-40.

Weisman, Z. (2009). Desert olive oil cultivation. United States, Elsevier.

Weststratea, J. A. and G. W. Meijer (1998). "Plant sterol-enriched margarines and reduction of plasma total- and LDL-cholesterol concentrations in normocholesterolaemic and mildly hypercholesterolaemic subjects." *European Journal of Clinical Nutrition***52**(5): 334-343.

Xu, F., S. D. Rychnovsky, J. D. Belani, H. H. Hobbs, J. C. Cohen and R. B. Rawson (2005). "Dual roles for cholesterol in mammalian cells." *Proceeding of the National Acedemy of Sciences***102**(41): 14551-14556.

Yanik, H., M. Turktas, E. Dundar, P. Hernandez, G. Dorado and T. Unver (2013). "Genome-wide identification of alternate bearing-associated microRNAs (miRNAs) in olive (*Olea europaea L.*)." *BMC Plant Biology***13**(10): 1-21.

Yokota, T., T. Nomura and M. Nakayama (1997). "Identification of brassinosteroids that appear to be derived from campesterol and cholesterol in tomato shoots." *Plant Cell Physiology***38**(11): 1291-1294.

Zou, M. Q., X. F. Zhang, X. H. Qi, H. L. Ma, Y. Dong, C. W. Liu, X. Guo and H. Wang (2009). "Rapid authentication of olive oil adulteration by raman spectrometry." *Journal of Agricultural and Food Chemistry***57**(14): 6001-6006.

APPENDICES

APPENDIX I

CHEMICALS AND REAGENTS

A. General buffers and solutions

- 2X CTAB
  2%(w/v) CTAB, 100mM Tris-HCl (pH8.5), 20mM EDTA (pH8), 1.4M NaCl, 1% PVP and 1% DTT. These were dissolved in ddH_2O and autoclaved.
- 5% CTAB
  5% (w/v) CTAB, 0.35M NaCl.
- 1% CTAB
  1%(w/v) CTAB, 50mM Tris-HCl (pH8), 10mM EDTA (pH8).
- REB solution
  10mM Tris-HCl (pH7.5), 1mM EDTA (pH8), 1M NaCl. These were dissolved in ddH_2O and autoclaved.
- 3M Sodium Acetate (pH 5.2)
  40.8g of sodium acetate-3H_2O (Ajax Chemicals) were dissolved in ddH_2O and the pH was adjusted to 5.2 with glacial acetic acid and made up to 100 mL final volume of ddH_2O prior to autoclaving. For RNA work, sodium acetate was dissolved in DEPC treated water instead of ddH_2O and stored in RNase free glass bottle at room temperature.
- TE Buffer
  10mM Tris-Cl (pH 8), 1mM EDTA (pH 8).
- Ethidium Bromide (10mg/mL) (Sigma)
- TBE electrophoresis buffer (10X)
  Tris base (0.89M), boric acid (0.89M) and 20 mM EDTA (pH 8)were dissolved in ddH_2O and autoclaved. The 10X stock solution was dilutedwith ddH_2O prior to use.

- TAE electrophoresis buffer (50X)

  242g Tris base, 57.1mL Glacial acetic acid, 100mL 0.5M EDTA were dissolved in DEPC treated water. The pH was adjusted to 8 with glacial acetic acid prior to autoclaving. The 50X stock solution was diluted with DEPC treated water prior to use with RNA.

- Tris (1M)

  Tris base (121.1g) was dissolved in ddH_2O and the pH wasadjusted to 7.5 with glacial acetic acid. The volume was adjusted to 1 L and the solution was autoclaved.

- SSC Buffer (20X)

  3M NaCl, 0.3M sodium citrate.

- Denaturation solution

  1.5M NaCl, 0.5M NaOH.

- Neutralisation solution

  1.5M NaCl, 0.5M Tris-Cl (pH 7.2).

- Washing buffer

  0.1M Maleic acid, 0.15M NaCl (pH 7.5), 0.3% (v/v) Tween 20.

- Detection Buffer

  100mM Tris base (pH 9.5), 100 mM NaCl, 50 mM $MgCl_{2.}$

- Maleic Acid Buffer

  0.1M Maleic Acid, 0.15 M NaCl (pH 7.5).

- 100X Denhardt's solution (prehybridisation buffer)

  2g BSA, 2g ficoll 400 and 2g polyvinylpyrrolidone was dissolved in 50mL ddH_2O and the solution was made up to a final volume of 100mL and stored at -20˚C for upto three months.

- 10% SDS (electrophoresis grade, BioRad)

  10g sodium dodecyl sulfate (SDS) wasdissolved in 90mL ddH_2O and heated at 68˚C until the SDS was completelydissolved. This solution was filter sterilised with a 0.45µm filter andstored at room temperature.

- 0.5M EDTA (Ajax Chemicals)

  186.1g of disodium ethylenediaminetetra-acetate-2H_2O was added to 800mL of ddH_2O and the pH was adjusted to 8.0 by the addition of NaOH. The solution was made up to a final volume of 1000mL with ddH_2O and autoclaved.

- 10X CIP Buffer

0.5 M Tris-HCl, pH (8.5) (20°C), 1 mM EDTA

- 10X TAP Buffer

  0.5 M sodium acetate (pH 6.0), 10 mM EDTA, 1% β-mercaptoethanol, 0.1% Triton® X-100

- Ampicillin (50mg/mL) (Sigma)

  200mg of ampicillin was dissolved in ddH2O to a final volume of 4mL, and filer sterilised with a 0.22µm filter and stored at -20˚C.

- Kanamycin (50mg/mL) (Sigma)
- 200mg of kanamycin was dissolved in ddH2O to a final volume of 4mL, and filer sterilised with a 0.22µm filter and stored at -20˚C.
- BigDye sequencing dilution buffer

  [1.4mM MgCl2, 60mM Tris-Cl (pH 7.5), 3M sodium acetate (pH 5.2)]

- Chloroform: Isoamyl alcohol (24:1)

  Chloroform (24 mL) and isoamyl alcohol (1 mL) added together and mixed well.

- Phenol: Chloroform: Isoamyl alcohol (25:24:1)

  Phenol (25mL), Chloroform (24 mL) and isoamyl alcohol (1 mL) added together and mixed well.

- DEPC water

  0.1% DEPC and ddH2O water were mixed well, allowed to stand overnight under the fumehood, and then autoclaved.

- Buffered phenol (for DNA isolation):

  Special Grade phenol (Wako Pure Chemical Industries limited)was used. Equal volumes of phenol and 50 mM Tris.HCl buffer (see above) were mixed with 8-hydroxyquinoline (0.8% w/v) in a brown (light proof) baked bottle with a stirring bar for 10 minutes. The phases were allowed to separate and the aqueous top phase was removed and replaced with an equal volume of 50 mM Tris.HCl buffer. This procedure was repeated using 50 mM Tris.HCl buffer until the aqueous top phase was at pH 8.0. The buffered phenol was stored at 4˚Cin a dark container for two-three months.

B. Enzymes

A number of enzymes were used in this project especially for PCR, cloning and sequencing reactions, many of which were supplied with the kits. These enzymes included:

- RNase A (20mg/mL) which was used to prepare RNA free genomic DNA was supplied by Invitrogen.
- Platinum *Taq* polymerase and platinum *Taq* polymerase high fidelity for PCR amplifications were provided by Invitrogen.
- DNA polymerases for sequencing reactions were provided by Applied Biosystems in the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit v.3.
- Various restriction endonucleases such as Bam H1, Hind III and EcoR1, for Southern blotting and digestion of TOPO vectors, were provided by Promega, MBI Fermentas and New England Biolabs.
- The Klenow fragment of *E. coli* DNA polymerase I was supplied by Promega, which were used in the radiolabelling of probes for Southern blotting experiments along with the Prime-a-gene Labelling System.
- Alkaline Protease solution was used to purify plasmid DNA and was provided with the Wizard® SV Gel and PCR Clean-Up System from Promega.
- Calf Intestinal Phosphatase (CIP) was used to remove the 5′ phosphates from total RNA to eliminate truncated and non-mRNA in RACE experiments. This enzyme was supplied along with the gene Racer kit from Invitrogen.
- Tobacco Acid Pyrophosphatase (TAP) was used to remove the 5′ cap structure from intact, full-length mRNA and was supplied with the gene Racer kit from Invitrogen.
- T4 RNA ligase was used for the ligation of the GeneRacer™ RNA Oligo to the 5′ end of the mRNA and was supplied with the gene Racer kit from Invitrogen.

C. Microbiological media

- LB broth (Luria-Bertani broth)
  1% w/v bacto-tryptone (Oxoid),
  0.5% bacto-yeast extract (Oxoid),
  1% w/v NaCl.
  The LB media was sterilised by autoclaving at 121˚C for 15 minutes. To prepare solid LB agar plates, agar was added (20g/L) to the LB Media prior to autoclaving.

- LB-ampicillin broth
  1% w/v bacto-tryptone (Oxoid),
  0.5% bacto-yeast extract (Oxoid),
  1% w/v NaCl.
  100 μg/ml ampicillin
  The LB media was sterilised by autoclaving at 121˚C for 15 minutes, allowed to cool and then ampicillin (100μg/mL) was added and stored at +4°C. To prepare solid LB agar plates, agar was added (20g/L) to the LB Media prior to autoclaving.

- SOC Media
  2% Tryptone
  0.5% Yeast Extract
  10mM NaCl
  2.5mM KCl
  10mM $MgCl_2$
  10mM $MgSO_4$
  20mM glucose
  The SOC Media was provided with the TOPO TA cloning kit from Invitrogen and was stored at +4°C.

APPENDIX II

MAPS OF VECTORS USED IN THIS PROJECT

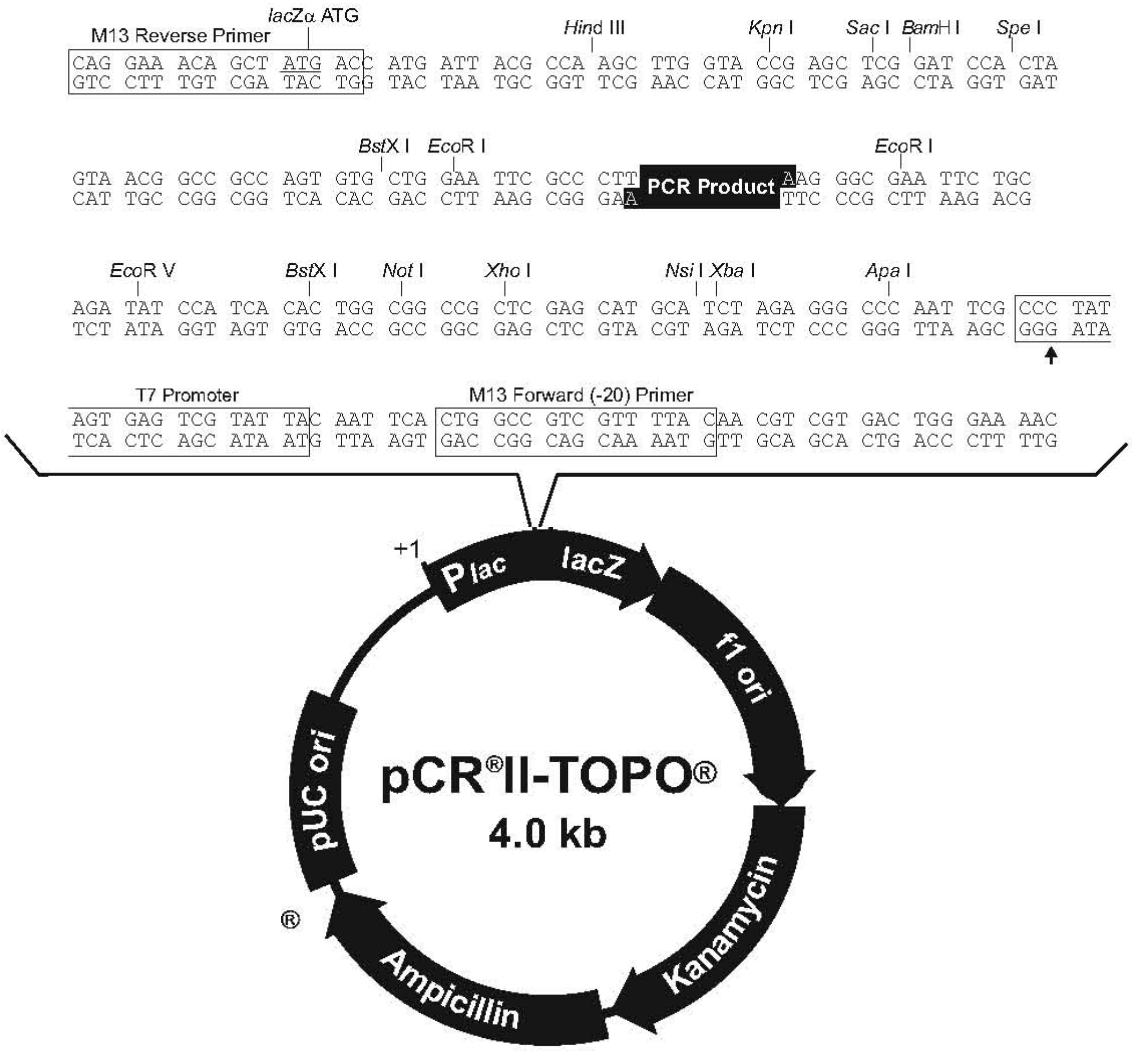
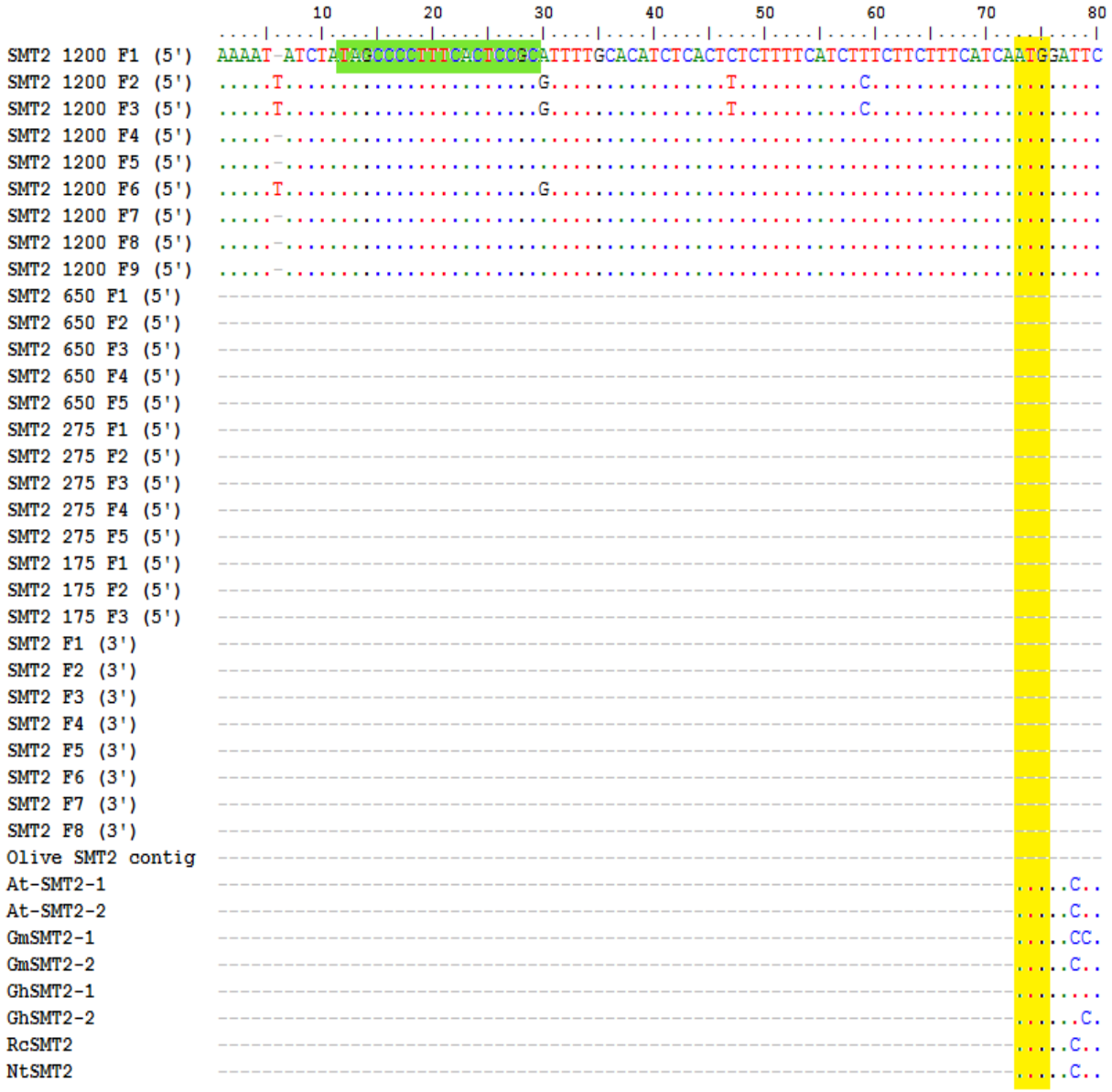
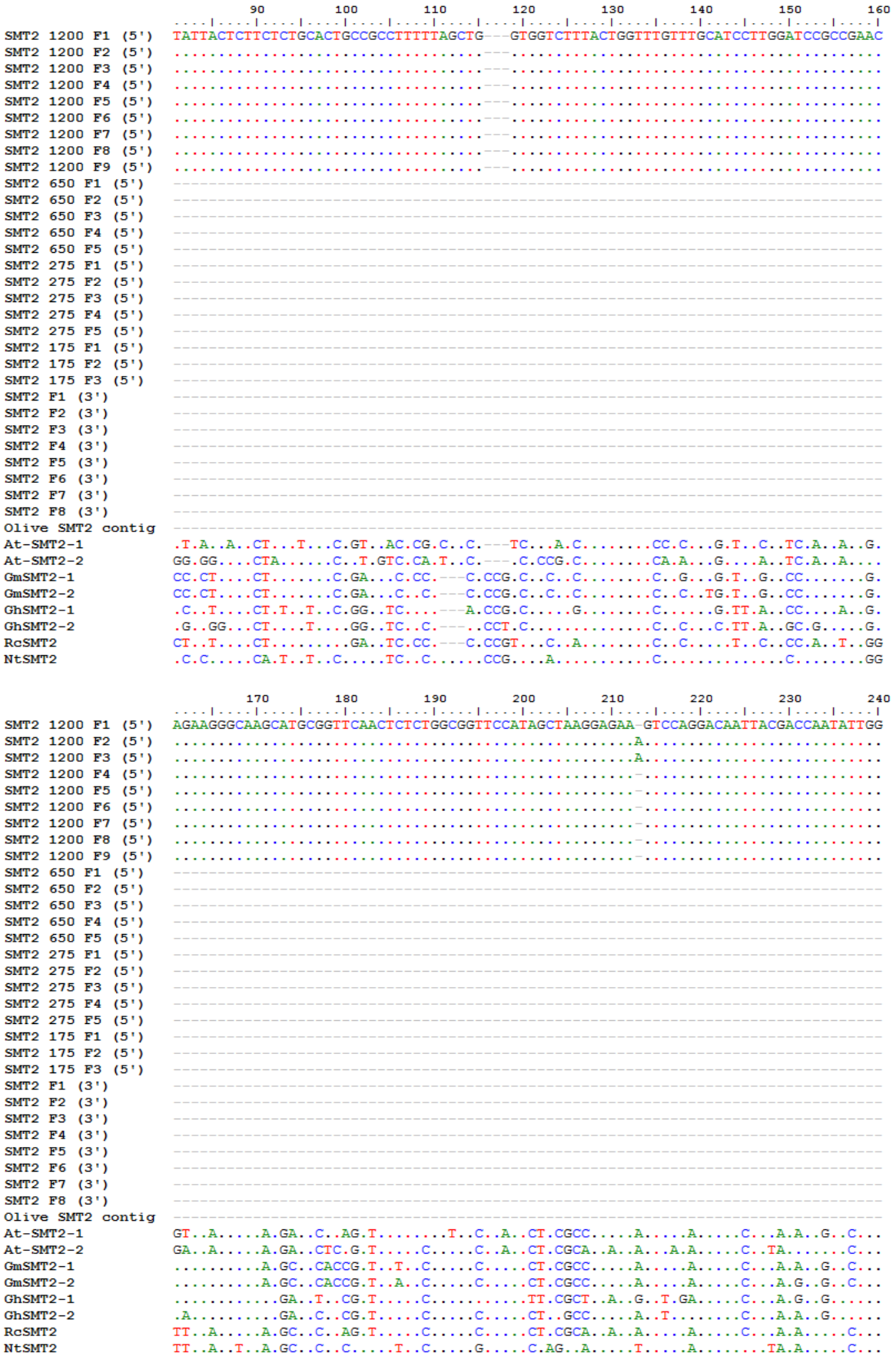


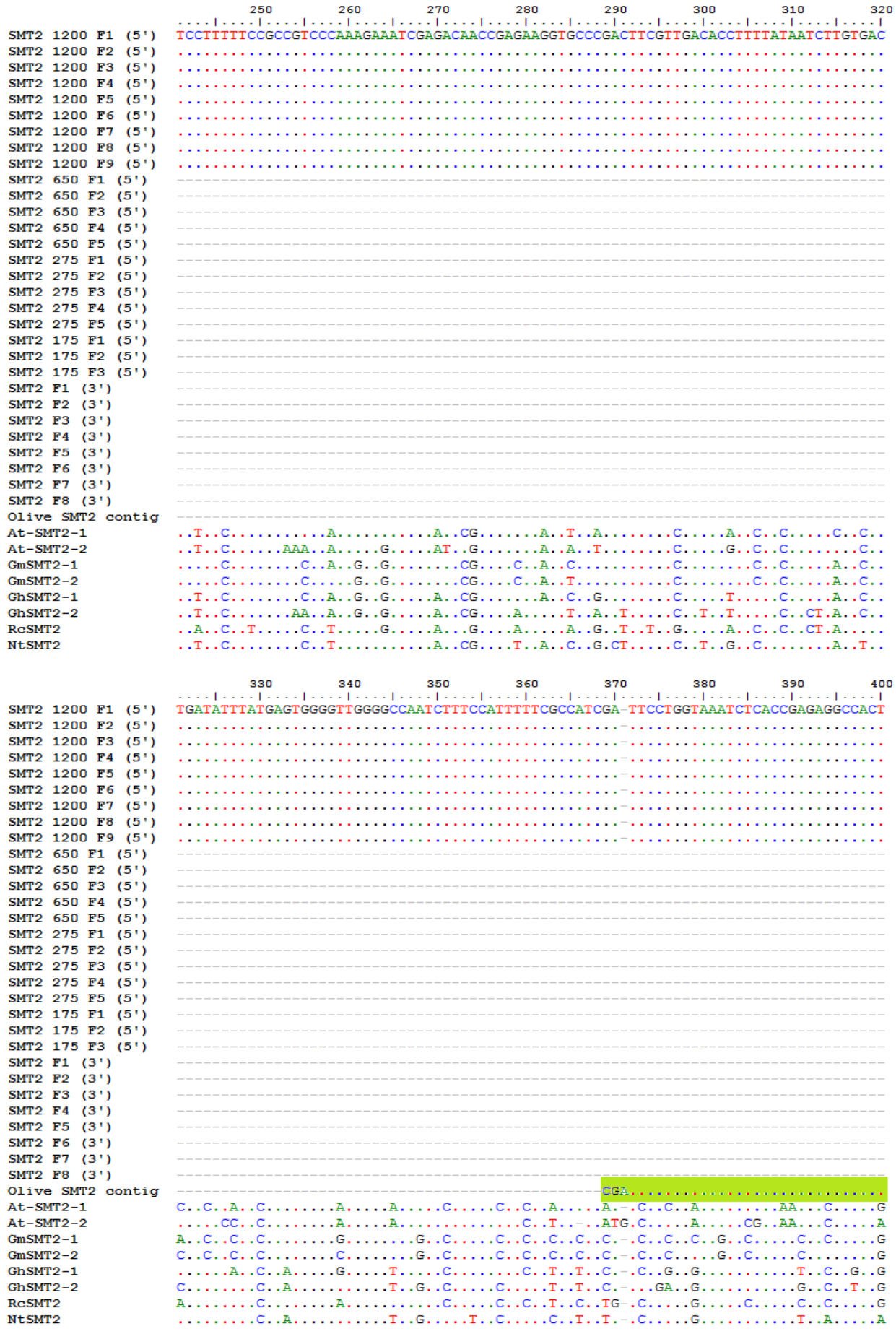
Figure IIA Map of pCR II-TOPO Vector
The pCR II-TOPO Vector was used for the cloning of *4SMO*and SMT2 gene fragments amplified by PCR. The vector map shows the following regions: LacZα gene**,** f1 origin, pUC origin, kanamycin resistance ORF, ampicillin resistance ORF and multiple cloning site containing M13 forward priming site, M13 reverse priming site, T7 promoter/priming site, SP6 promoter/priming site and various restriction enzyme sites.

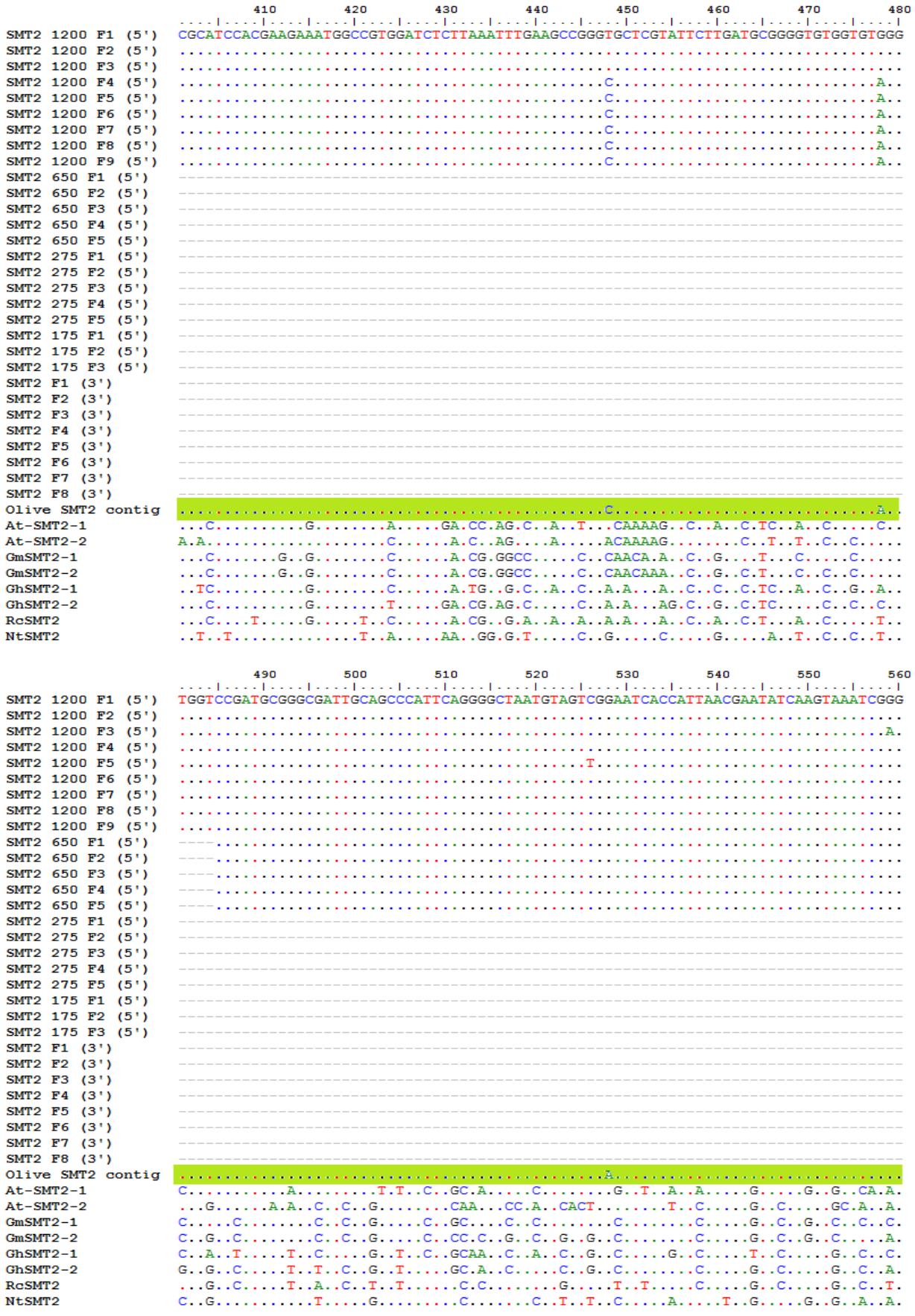
APPENDIX III

THE DNA SEQUENCES OF *SMT2* GENES FROM BARNEA, FRANTOIO AND PICUAL



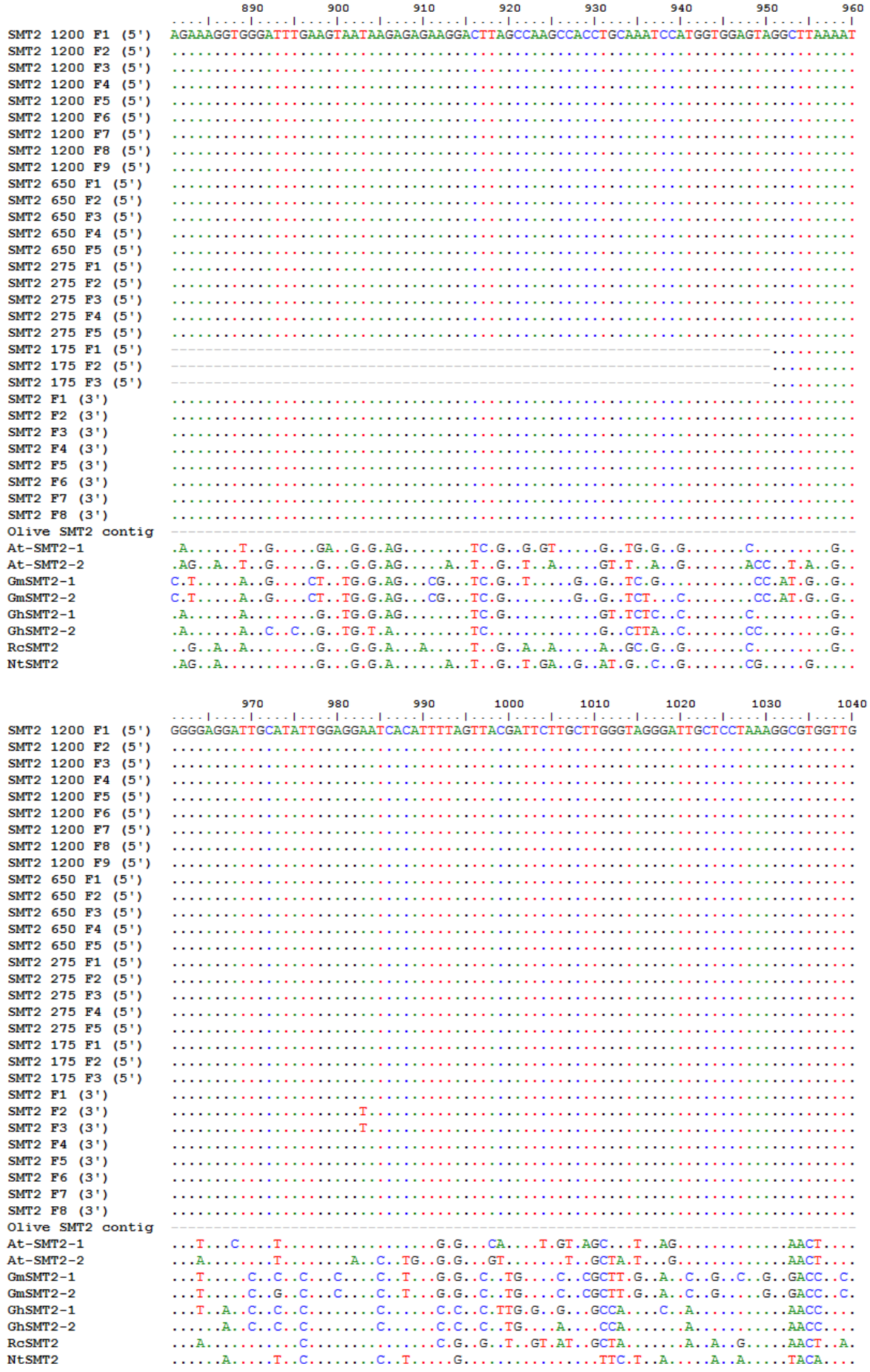

90 100 110 120 130 140 150 160
SMT2 1200 F1 (5') TATTACTCTTCTCTGCACTGCCGCCTTTTTAGCTG---GTGGTCTTTACTGGTTTGTTTGCATCCTTGGATCCGCCGAAC
SMT2 1200 F2 (5')
SMT2 1200 F3 (5')
SMT2 1200 F4 (5')
SMT2 1200 F5 (5')
SMT2 1200 F6 (5')
SMT2 1200 F7 (5')
SMT2 1200 F8 (5')
SMT2 1200 F9 (5')
SMT2 650 F1 (5')
SMT2 650 F2 (5')
SMT2 650 F3 (5')
SMT2 650 F4 (5')
SMT2 650 F5 (5')
SMT2 275 F1 (5')
SMT2 275 F2 (5')
SMT2 275 F3 (5')
SMT2 275 F4 (5')
SMT2 275 F5 (5')
SMT2 175 F1 (5')
SMT2 175 F2 (5')
SMT2 175 F3 (5')
SMT2 F1 (3')
SMT2 F2 (3')
SMT2 F3 (3')
SMT2 F4 (3')
SMT2 F5 (3')
SMT2 F6 (3')
SMT2 F7 (3')
SMT2 F8 (3')
Olive SMT2 contig
At-SMT2-1 .T.A..A..CT...T...C.GT..AC.CG.C..C.---TC...A.C.....CC.C...G.T..C..TC.A..A..G.
At-SMT2-2 GG.GG....CTA.....C..T.GTC.CA.T..C.----.C.CCG.C.....CA.A...G....A..TC.A..A....
GmSMT2-1 CC.CT....CT.....C.GA...C.CC.---C.CCG.C..C..C.....C..G...G.T..G..CC.....G.
GmSMT2-2 CC.CT....CT.....C.GA...C..C.---C.CCG.C..C..C.....C..C..TG.T..G..CC.....G.
GhSMT2-1 .C..T....CT.T..T..C.GG..TC....---A.CCG.C.....G.....C.....G.TT.A..CC....A..G.
GhSMT2-2 .G..GG...CT....T....GG..TC..C.----..CCT.C.....C..C...C.TT.A..GC.G.....G.
RcSMT2 CT..T....CT.....GA..TC.CC.---C.CCGT...C..A.....C..C.....T..C..CC.A..T..GG
NtSMT2 .C.C.....CA.T..T..C.....TC..C.....CCG....A.....C.....C.....GG
170 180 190 200 210 220 230 240
SMT2 1200 F1 (5') AGAAGGGCAAGCATGCGGTTCAACTCTCTGGCGGTTCCATAGCTAAGGAGAA-GTCCAGGACAATTACGACCAATATTGG
SMT2 1200 F2 (5')
SMT2 1200 F3 (5')
SMT2 1200 F4 (5')
SMT2 1200 F5 (5')
SMT2 1200 F6 (5')
SMT2 1200 F7 (5')
SMT2 1200 F8 (5')
SMT2 1200 F9 (5')
SMT2 650 F1 (5')
SMT2 650 F2 (5')
SMT2 650 F3 (5')
SMT2 650 F4 (5')
SMT2 650 F5 (5')
SMT2 275 F1 (5')
SMT2 275 F2 (5')
SMT2 275 F3 (5')
SMT2 275 F4 (5')
SMT2 275 F5 (5')
SMT2 175 F1 (5')
SMT2 175 F2 (5')
SMT2 175 F3 (5')
SMT2 F1 (3')
SMT2 F2 (3')
SMT2 F3 (3')
SMT2 F4 (3')
SMT2 F5 (3')
SMT2 F6 (3')
SMT2 F7 (3')
SMT2 F8 (3')
Olive SMT2 contig
At-SMT2-1 GT..A.....A.GA..C..AG.T.....T..C..A..CT.CGCC.....A.....A.....C...A.A..G..C...
At-SMT2-2 GA..A.....A.GA..CTC.G.T.....C.....C..A..CT.CGCA..A..A...A.A.....C..TA.....C...
GmSMT2-1 ..........A.GC..CACCG.T..T..C.....C.....CT.CGCC.....A.....A.....C...A.A..G..C...
GmSMT2-2 ..........A.GC..CACCG.T..A..C.....C.....CT.CGCC.....A.....A.....C...A.G..G..C...
GhSMT2-1 ...........GA..T..CG.T.....C..........TT.CGCT..A..G..T.GA.....C...A.G..G......
GhSMT2-2 .A.........GA..C..CG.T.....C.....C.....CT..GCC.....A..T.....C...A.A..G......
RcSMT2 TT..A.....A.GC..C..AG.T.....C.....C.....CT.CGCA..A..A.....A.....C...A.A.....C...
NtSMT2 TT..A..T..A.GC..C..C.....T..C.....G.....C.AG..A.....T.....A.....TA.A.....C...

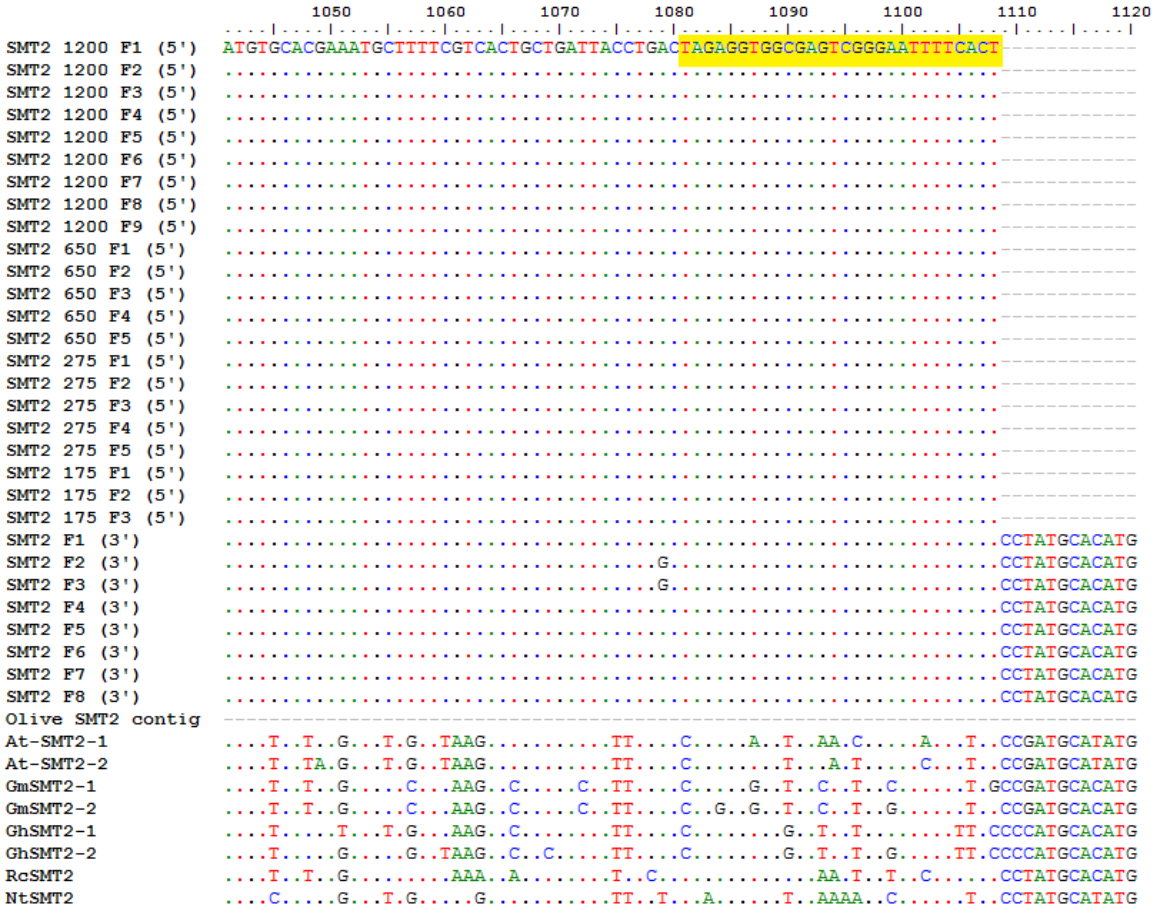

250 260 270 280 290 300 310 320
SMT2 1200 F1 (5') TCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGTGCCCGACTTCGTTGACACCTTTTATAATCTTGTGAC
SMT2 1200 F2 (5')
SMT2 1200 F3 (5')
SMT2 1200 F4 (5')
SMT2 1200 F5 (5')
SMT2 1200 F6 (5')
SMT2 1200 F7 (5')
SMT2 1200 F8 (5')
SMT2 1200 F9 (5')
SMT2 650 F1 (5')
SMT2 650 F2 (5')
SMT2 650 F3 (5')
SMT2 650 F4 (5')
SMT2 650 F5 (5')
SMT2 275 F1 (5')
SMT2 275 F2 (5')
SMT2 275 F3 (5')
SMT2 275 F4 (5')
SMT2 275 F5 (5')
SMT2 175 F1 (5')
SMT2 175 F2 (5')
SMT2 175 F3 (5')
SMT2 F1 (3')
SMT2 F2 (3')
SMT2 F3 (3')
SMT2 F4 (3')
SMT2 F5 (3')
SMT2 F6 (3')
SMT2 F7 (3')
SMT2 F8 (3')
Olive SMT2 contig
At-SMT2-1
At-SMT2-2
GmSMT2-1
GmSMT2-2
GhSMT2-1
GhSMT2-2
RcSMT2
NtSMT2
330 340 350 360 370 380 390 400
SMT2 1200 F1 (5') TGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGA-TTCCTGGTAAATCTCACCGAGAGGCCACT
SMT2 1200 F2 (5')
SMT2 1200 F3 (5')
SMT2 1200 F4 (5')
SMT2 1200 F5 (5')
SMT2 1200 F6 (5')
SMT2 1200 F7 (5')
SMT2 1200 F8 (5')
SMT2 1200 F9 (5')
SMT2 650 F1 (5')
SMT2 650 F2 (5')
SMT2 650 F3 (5')
SMT2 650 F4 (5')
SMT2 650 F5 (5')
SMT2 275 F1 (5')
SMT2 275 F2 (5')
SMT2 275 F3 (5')
SMT2 275 F4 (5')
SMT2 275 F5 (5')
SMT2 175 F1 (5')
SMT2 175 F2 (5')
SMT2 175 F3 (5')
SMT2 F1 (3')
SMT2 F2 (3')
SMT2 F3 (3')
SMT2 F4 (3')
SMT2 F5 (3')
SMT2 F6 (3')
SMT2 F7 (3')
SMT2 F8 (3')
Olive SMT2 contig CGA
At-SMT2-1
At-SMT2-2
GmSMT2-1
GmSMT2-2
GhSMT2-1
GhSMT2-2
RcSMT2
NtSMT2

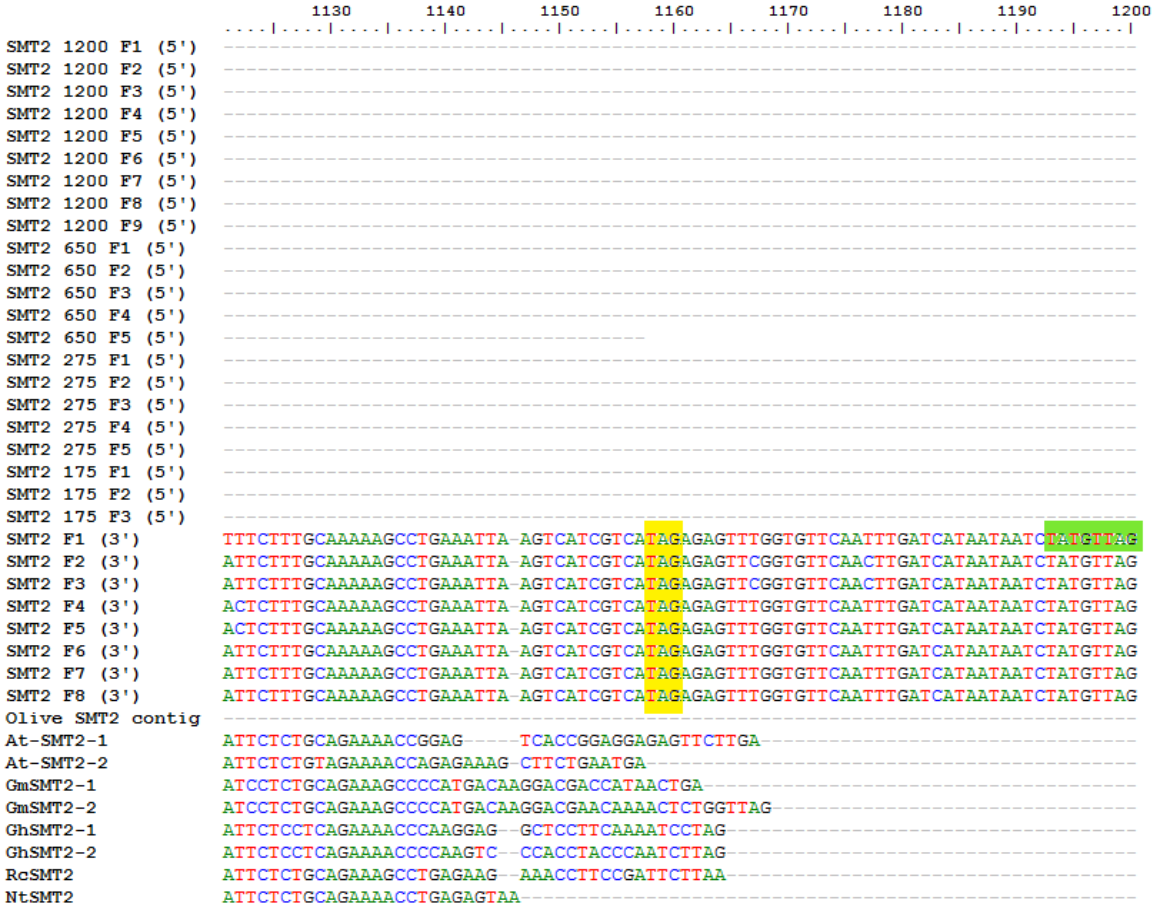

410 420 430 440 450 460 470 480
SMT2 1200 F1 (5') CGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGTGCTCGTATTCTTGATGCGGGGTGTGGTGTGGG
SMT2 1200 F2 (5')
SMT2 1200 F3 (5')
SMT2 1200 F4 (5')
SMT2 1200 F5 (5')
SMT2 1200 F6 (5')
SMT2 1200 F7 (5')
SMT2 1200 F8 (5')
SMT2 1200 F9 (5')
SMT2 650 F1 (5')
SMT2 650 F2 (5')
SMT2 650 F3 (5')
SMT2 650 F4 (5')
SMT2 650 F5 (5')
SMT2 275 F1 (5')
SMT2 275 F2 (5')
SMT2 275 F3 (5')
SMT2 275 F4 (5')
SMT2 275 F5 (5')
SMT2 175 F1 (5')
SMT2 175 F2 (5')
SMT2 175 F3 (5')
SMT2 F1 (3')
SMT2 F2 (3')
SMT2 F3 (3')
SMT2 F4 (3')
SMT2 F5 (3')
SMT2 F6 (3')
SMT2 F7 (3')
SMT2 F8 (3')
Olive SMT2 contig
At-SMT2-1
At-SMT2-2
GmSMT2-1
GmSMT2-2
GhSMT2-1
GhSMT2-2
RcSMT2
NtSMT2
490 500 510 520 530 540 550 560
SMT2 1200 F1 (5') TGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAATCGGG
SMT2 1200 F2 (5')
SMT2 1200 F3 (5')
SMT2 1200 F4 (5')
SMT2 1200 F5 (5')
SMT2 1200 F6 (5')
SMT2 1200 F7 (5')
SMT2 1200 F8 (5')
SMT2 1200 F9 (5')
SMT2 650 F1 (5')
SMT2 650 F2 (5')
SMT2 650 F3 (5')
SMT2 650 F4 (5')
SMT2 650 F5 (5')
SMT2 275 F1 (5')
SMT2 275 F2 (5')
SMT2 275 F3 (5')
SMT2 275 F4 (5')
SMT2 275 F5 (5')
SMT2 175 F1 (5')
SMT2 175 F2 (5')
SMT2 175 F3 (5')
SMT2 F1 (3')
SMT2 F2 (3')
SMT2 F3 (3')
SMT2 F4 (3')
SMT2 F5 (3')
SMT2 F6 (3')
SMT2 F7 (3')
SMT2 F8 (3')
Olive SMT2 contig
At-SMT2-1
At-SMT2-2
GmSMT2-1
GmSMT2-2
GhSMT2-1
GhSMT2-2
RcSMT2
NtSMT2

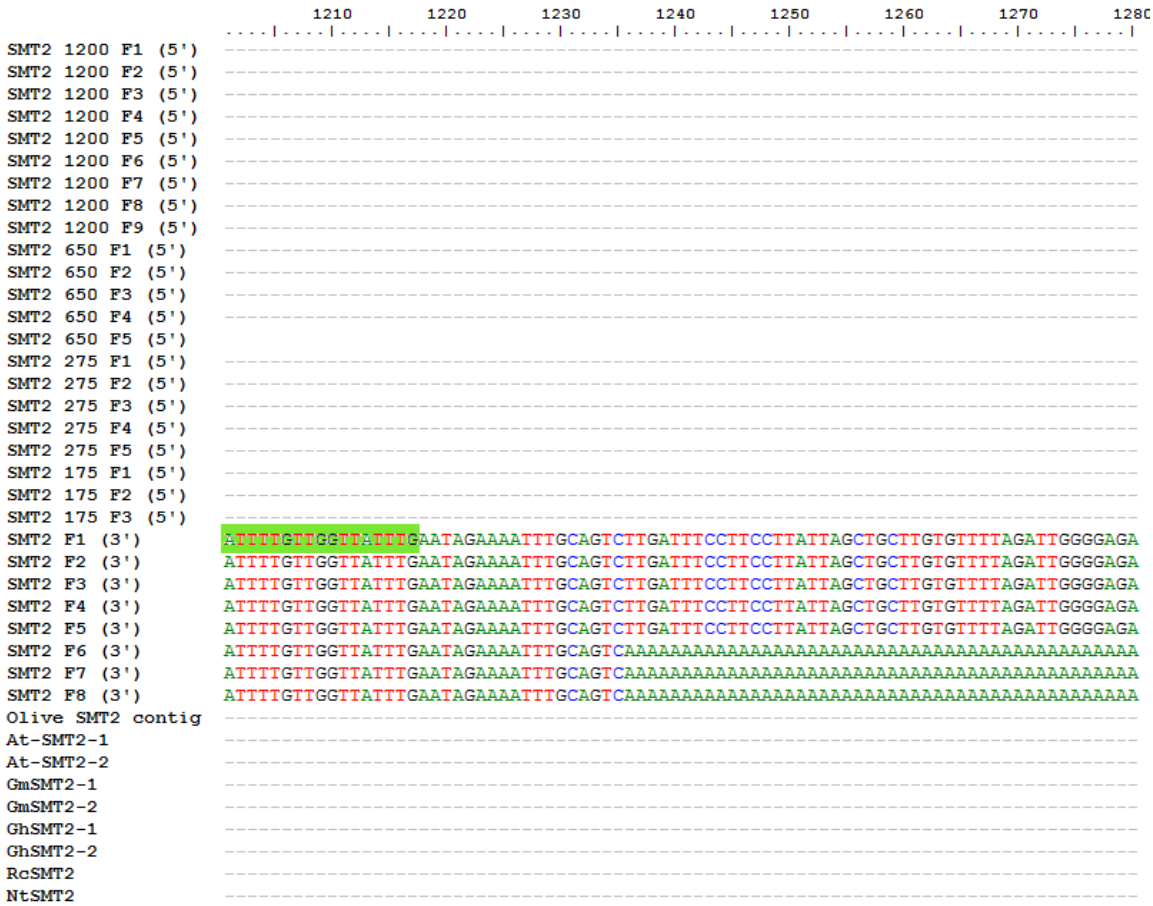

570 580 590 600 610 620 630 640
SMT2 1200 F1 (5') CGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGT
SMT2 1200 F2 (5')
SMT2 1200 F3 (5')
SMT2 1200 F4 (5')
SMT2 1200 F5 (5')
SMT2 1200 F6 (5')
SMT2 1200 F7 (5')
SMT2 1200 F8 (5')
SMT2 1200 F9 (5')
SMT2 650 F1 (5')
SMT2 650 F2 (5')
SMT2 650 F3 (5')
SMT2 650 F4 (5')
SMT2 650 F5 (5')
SMT2 275 F1 (5')
SMT2 275 F2 (5')
SMT2 275 F3 (5')
SMT2 275 F4 (5')
SMT2 275 F5 (5')
SMT2 175 F1 (5')
SMT2 175 F2 (5')
SMT2 175 F3 (5')
SMT2 F1 (3')
SMT2 F2 (3')
SMT2 F3 (3')
SMT2 F4 (3')
SMT2 F5 (3')
SMT2 F6 (3')
SMT2 F7 (3')
SMT2 F8 (3')
Olive SMT2 contig
At-SMT2-1 .TC.TCTC.....T.....A..T..T..C..CGCG.TT.....G.....T..T..C.....C.....G....A.
At-SMT2-2 .C.A.CTT.....A..T..A.....TCT.TC...A.C.....T..T..T..C..TT.AA.....G....A.
GmSMT2-1 .A...ATG.....A..GT.G..GTCC.TC.....G.....G.....T...A.....CCC
GmSMT2-2 .A...ATG.....T..GT.G..CTCT.TC.....G.G.....T..G.....TA.....G..T.TC
GhSMT2-1 .CC..ATG.....A.....G..A...TCT.TT.....T..T..T..G..C.....G.....G..TCAA
GhSMT2-2 .C...CTT.....A..T..G..A...TCG.TT..T.....G..T.....G..C..TT.G.....A...AAA
RcSMT2 ..C..CTG.....A..A..T..A...TC..TC.....T.....T..A.....G.....A..TCCA
NtSMT2 .CC..ATG.....A..A..T.....C...TCCTT.....T..T.....GC.A.....TCCA
650 660 670 680 690 700 710 720
SMT2 1200 F1 (5') GATAACAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTA
SMT2 1200 F2 (5')
SMT2 1200 F3 (5')
SMT2 1200 F4 (5')
SMT2 1200 F5 (5')
SMT2 1200 F6 (5')
SMT2 1200 F7 (5')
SMT2 1200 F8 (5')
SMT2 1200 F9 (5')
SMT2 650 F1 (5')
SMT2 650 F2 (5')
SMT2 650 F3 (5')
SMT2 650 F4 (5')
SMT2 650 F5 (5')
SMT2 275 F1 (5')
SMT2 275 F2 (5')
SMT2 275 F3 (5')
SMT2 275 F4 (5')
SMT2 275 F5 (5')
SMT2 175 F1 (5')
SMT2 175 F2 (5')
SMT2 175 F3 (5')
SMT2 F1 (3')
SMT2 F2 (3')
SMT2 F3 (3')
SMT2 F4 (3')
SMT2 F5 (3')
SMT2 F6 (3')
SMT2 F7 (3')
SMT2 F8 (3')
Olive SMT2 contig
At-SMT2-1 ..C.....T..T.....C..C.....G..T..C..G..G..G.....A..G...GCA.....
At-SMT2-2 ..A....CG..T.....T..C..G..C.....T..G..T..C..T..T..G..C.....A.....TCG.....T
GmSMT2-1 ..C.....C.....C..G..C..C..C..G.....G..T..C..G..G..G.....G..A..G...GC...A....T
GmSMT2-2 ..C.....C.....G..C..C..C..G.....G..T..C..T..C..G.....A..G...GC...A....T
GhSMT2-1 ..C.....C.....C.....C..C..C.....A..A.....C..G..G.....C...GC....G...T
GhSMT2-2 ..C....CG.....T..T..G..C..G..A..G..G..G.....C..G..G.....A..G...GCG...G.G.T
RcSMT2 ..C.....C.....G..C.....C..G..A..A..T..C..G..G..G..A.....G.....TGCA.....T
NtSMT2 ..C.....C.....T.....T..G..T..C..T..T..G..T.....A..G..T.AG.....


730 740 750 760 770 780 790 800
SMT2 1200 F1 (5') CCGGGTTTTAAAGCCCGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAAC
SMT2 1200 F2 (5')
SMT2 1200 F3 (5')
SMT2 1200 F4 (5')
SMT2 1200 F5 (5')
SMT2 1200 F6 (5')
SMT2 1200 F7 (5')
SMT2 1200 F8 (5')
SMT2 1200 F9 (5')
SMT2 650 F1 (5')
SMT2 650 F2 (5')
SMT2 650 F3 (5')
SMT2 650 F4 (5')
SMT2 650 F5 (5')
SMT2 275 F1 (5')
SMT2 275 F2 (5')
SMT2 275 F3 (5')
SMT2 275 F4 (5')
SMT2 275 F5 (5')
SMT2 175 F1 (5')
SMT2 175 F2 (5')
SMT2 175 F3 (5')
SMT2 F1 (3')
SMT2 F2 (3')
SMT2 F3 (3')
SMT2 F4 (3')
SMT2 F5 (3')
SMT2 F6 (3')
SMT2 F7 (3')
SMT2 F8 (3')
Olive SMT2 contig
At-SMT2-1
At-SMT2-2
GmSMT2-1
GmSMT2-2
GhSMT2-1
GhSMT2-2
RcSMT2
NtSMT2
810 820 830 840 850 860 870 880
SMT2 1200 F1 (5') ACCTGGAAGTAATCCAGGGAATTGAGAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCG
SMT2 1200 F2 (5')
SMT2 1200 F3 (5')
SMT2 1200 F4 (5')
SMT2 1200 F5 (5')
SMT2 1200 F6 (5')
SMT2 1200 F7 (5')
SMT2 1200 F8 (5')
SMT2 1200 F9 (5')
SMT2 650 F1 (5')
SMT2 650 F2 (5')
SMT2 650 F3 (5')
SMT2 650 F4 (5')
SMT2 650 F5 (5')
SMT2 275 F1 (5')
SMT2 275 F2 (5')
SMT2 275 F3 (5')
SMT2 275 F4 (5')
SMT2 275 F5 (5')
SMT2 175 F1 (5')
SMT2 175 F2 (5')
SMT2 175 F3 (5')
SMT2 F1 (3')
SMT2 F2 (3')
SMT2 F3 (3')
SMT2 F4 (3')
SMT2 F5 (3')
SMT2 F6 (3')
SMT2 F7 (3')
SMT2 F8 (3')
Olive SMT2 contig
At-SMT2-1
At-SMT2-2
GmSMT2-1
GmSMT2-2
GhSMT2-1
GhSMT2-2
RcSMT2
NtSMT2


890 900 910 920 930 940 950 960
SMT2 1200 F1 (5') AGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGACTTAGCCAAGCCACCTGCAAATCCATGGTGGAGTAGGCTTAAAAT
SMT2 1200 F2 (5')
SMT2 1200 F3 (5')
SMT2 1200 F4 (5')
SMT2 1200 F5 (5')
SMT2 1200 F6 (5')
SMT2 1200 F7 (5')
SMT2 1200 F8 (5')
SMT2 1200 F9 (5')
SMT2 650 F1 (5')
SMT2 650 F2 (5')
SMT2 650 F3 (5')
SMT2 650 F4 (5')
SMT2 650 F5 (5')
SMT2 275 F1 (5')
SMT2 275 F2 (5')
SMT2 275 F3 (5')
SMT2 275 F4 (5')
SMT2 275 F5 (5')
SMT2 175 F1 (5')
SMT2 175 F2 (5')
SMT2 175 F3 (5')
SMT2 F1 (3')
SMT2 F2 (3')
SMT2 F3 (3')
SMT2 F4 (3')
SMT2 F5 (3')
SMT2 F6 (3')
SMT2 F7 (3')
SMT2 F8 (3')
Olive SMT2 contig
At-SMT2-1 .A.....T..G.....GA..G.G.AG.....TC.G..G.GT.....G..TG.G..G.....C.....G..
At-SMT2-2 .AG..A..T..G.....G...G.G.AG.....A..T..G..T..A.....GT.T..A..G.....ACC..T.A..G..
GmSMT2-1 C.T.....A..G....CT..TG.G.AG...CG...TC.G..T.....G..G..TC.G.....CC.AT.G..G..
GmSMT2-2 C.T.....A..G....CT..TG.G.AG...CG...TC.G.....G..G..TCT...C.....CC.AT.G..G..
GhSMT2-1 .A.....A.....G..TG.G.AG.....TC.G.....GT.TCTC..C.....C.....G..
GhSMT2-2 .A.....A..C..C..G..TG.T.A.....TC.....G..CTTA..C.....CC.....G..
RcSMT2 ..G..A..A.....G...G.G.A...A.....T..G..A..A.....A..GC.G..G.....C.....G..
NtSMT2 .AG..A.....G...G.G.A.....A..T..G..T.GA..G..AT.G..C..G.....CG.....G.....
970 980 990 1000 1010 1020 1030 1040
SMT2 1200 F1 (5') GGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCTTGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTG
SMT2 1200 F2 (5')
SMT2 1200 F3 (5')
SMT2 1200 F4 (5')
SMT2 1200 F5 (5')
SMT2 1200 F6 (5')
SMT2 1200 F7 (5')
SMT2 1200 F8 (5')
SMT2 1200 F9 (5')
SMT2 650 F1 (5')
SMT2 650 F2 (5')
SMT2 650 F3 (5')
SMT2 650 F4 (5')
SMT2 650 F5 (5')
SMT2 275 F1 (5')
SMT2 275 F2 (5')
SMT2 275 F3 (5')
SMT2 275 F4 (5')
SMT2 275 F5 (5')
SMT2 175 F1 (5')
SMT2 175 F2 (5')
SMT2 175 F3 (5')
SMT2 F1 (3')
SMT2 F2 (3') .....T.....
SMT2 F3 (3') .....T.....
SMT2 F4 (3')
SMT2 F5 (3')
SMT2 F6 (3')
SMT2 F7 (3')
SMT2 F8 (3')
Olive SMT2 contig
At-SMT2-1 ...T...C....T.....G.G...CA....T.GT.AGC...T..AG.....AACT....
At-SMT2-2 ...A.....T.....A..C..TG..G.G...GT.....T..GCTA.T...G.....AACT....
GmSMT2-1 ...T.....C..C..C...C....C..T...G.G..C..TG....C..CGCTT.G..A..C..G..C..G..GACC..C.
GmSMT2-2 ...T.....C..G..C...C....C..T...G.G..C..TG....C..CGCTT.G..A..C..G.....G..GACC..C.
GhSMT2-1 ...T..A..C..C..C.....C.....C.C..C.TTG.G..G...GCCA....C..A.....AACC....
GhSMT2-2 ......A..C..C..C.....C.....C.C..C..TG....A....CCA.....A.....AACC....
RcSMT2 ...A.....C.....C.G..G..T..GT.AT..GCTA.....A..A..G.....AACT..A.
NtSMT2 ......A.....T..C.....C..T.....G.....TTC.T..A.....A..A.....TACA....


1050 1060 1070 1080 1090 1100 1110 1120
SMT2 1200 F1 (5') ATGTGCACGAAATGCTTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTCGGGAATTTTCACT
SMT2 1200 F2 (5')
SMT2 1200 F3 (5')
SMT2 1200 F4 (5')
SMT2 1200 F5 (5')
SMT2 1200 F6 (5')
SMT2 1200 F7 (5')
SMT2 1200 F8 (5')
SMT2 1200 F9 (5')
SMT2 650 F1 (5')
SMT2 650 F2 (5')
SMT2 650 F3 (5')
SMT2 650 F4 (5')
SMT2 650 F5 (5')
SMT2 275 F1 (5')
SMT2 275 F2 (5')
SMT2 275 F3 (5')
SMT2 275 F4 (5')
SMT2 275 F5 (5')
SMT2 175 F1 (5')
SMT2 175 F2 (5')
SMT2 175 F3 (5')
SMT2 F1 (3') CCTATGCACATG
SMT2 F2 (3') G CCTATGCACATG
SMT2 F3 (3') G CCTATGCACATG
SMT2 F4 (3') CCTATGCACATG
SMT2 F5 (3') CCTATGCACATG
SMT2 F6 (3') CCTATGCACATG
SMT2 F7 (3') CCTATGCACATG
SMT2 F8 (3') CCTATGCACATG
Olive SMT2 contig
At-SMT2-1 ....T..T..G...T.G..TAAG..........TT....C.....A..T..AA.C.....A...T..CCGATGCATATG
At-SMT2-2 ....T..TA.G...T.G..TAAG..........TT....C.....T...A.T.....C...T..CCGATGCATATG
GmSMT2-1 ....T..T..G.....C...AAG..C.....C..TT....C.....G..T..C..T..C.....T.GCCGATGCACATG
GmSMT2-2 ....T..T..G.....C...AAG..C.....C..TT....C..G..G..T..C..T..G.....T..CCGATGCACATG
GhSMT2-1 ....T.....T...T.G...AAG..C.....TT....C.....G..T..T.....TT.CCCCATGCACATG
GhSMT2-2 ....T.....G.....G..TAAG..C..C.....TT....C.....G..T..T..G.....TT.CCCCATGCACATG
RcSMT2 ....T..T..G.....AAA..A.....T..C.....AA.T..T..C.....CCTATGCACATG
NtSMT2 ....C.....G...T.G.....G..........TT..T...A.....T..AAAA..C.....T..CCTATGCATATG


1130 1140 1150 1160 1170 1180 1190 1200
SMT2 1200 F1 (5')
SMT2 1200 F2 (5')
SMT2 1200 F3 (5')
SMT2 1200 F4 (5')
SMT2 1200 F5 (5')
SMT2 1200 F6 (5')
SMT2 1200 F7 (5')
SMT2 1200 F8 (5')
SMT2 1200 F9 (5')
SMT2 650 F1 (5')
SMT2 650 F2 (5')
SMT2 650 F3 (5')
SMT2 650 F4 (5')
SMT2 650 F5 (5')
SMT2 275 F1 (5')
SMT2 275 F2 (5')
SMT2 275 F3 (5')
SMT2 275 F4 (5')
SMT2 275 F5 (5')
SMT2 175 F1 (5')
SMT2 175 F2 (5')
SMT2 175 F3 (5')
SMT2 F1 (3') TTTCTTTGCAAAAAGCCTGAAATTA-AGTCATCGTCATAGAGAGTTTGGTGTTCAATTTGATCATAATAATCTATGTTAG
SMT2 F2 (3') ATTCTTTGCAAAAAGCCTGAAATTA-AGTCATCGTCATAGAGAGTTCGGTGTTCAACTTGATCATAATAATCTATGTTAG
SMT2 F3 (3') ATTCTTTGCAAAAAGCCTGAAATTA-AGTCATCGTCATAGAGAGTTCGGTGTTCAACTTGATCATAATAATCTATGTTAG
SMT2 F4 (3') ACTCTTTGCAAAAAGCCTGAAATTA-AGTCATCGTCATAGAGAGTTTGGTGTTCAATTTGATCATAATAATCTATGTTAG
SMT2 F5 (3') ACTCTTTGCAAAAAGCCTGAAATTA-AGTCATCGTCATAGAGAGTTTGGTGTTCAATTTGATCATAATAATCTATGTTAG
SMT2 F6 (3') ATTCTTTGCAAAAAGCCTGAAATTA-AGTCATCGTCATAGAGAGTTTGGTGTTCAATTTGATCATAATAATCTATGTTAG
SMT2 F7 (3') ATTCTTTGCAAAAAGCCTGAAATTA-AGTCATCGTCATAGAGAGTTTGGTGTTCAATTTGATCATAATAATCTATGTTAG
SMT2 F8 (3') ATTCTTTGCAAAAAGCCTGAAATTA-AGTCATCGTCATAGAGAGTTTGGTGTTCAATTTGATCATAATAATCTATGTTAG
Olive SMT2 contig
At-SMT2-1 ATTCTCTGCAGAAAACCGGAG-----TCACCGGAGGAGAGTTCTTGA
At-SMT2-2 ATTCTCTGTAGAAAACCAGAGAAAG-CTTCTGAATGA
GmSMT2-1 ATCCTCTGCAGAAAGCCCCATGACAAGGACGACCATAACTGA
GmSMT2-2 ATCCTCTGCAGAAAGCCCCATGACAAGGACGAACAAAACTCTGGTTAG
GhSMT2-1 ATTCTCCTCAGAAAACCCAAGGAG--GCTCCTTCAAAATCCTAG
GhSMT2-2 ATTCTCCTCAGAAAACCCCAAGTC--CCACCTACCCAATCTTAG
RcSMT2 ATTCTCTGCAGAAAGCCTGAGAAG--AAACCTTCCGATTCTTAA
NtSMT2 ATTCTCTGCAGAAAACCTGAGAGTAA


1210
1220
1230
1240
1250
1260
1270
SMT2 1200 F1 (5')
SMT2 1200 F2 (5')
SMT2 1200 F3 (5')
SMT2 1200 F4 (5')
SMT2 1200 F5 (5')
SMT2 1200 F6 (5')
SMT2 1200 F7 (5')
SMT2 1200 F8 (5')
SMT2 1200 F9 (5')
SMT2 650 F1 (5')
SMT2 650 F2 (5')
SMT2 650 F3 (5')
SMT2 650 F4 (5')
SMT2 650 F5 (5')
SMT2 275 F1 (5')
SMT2 275 F2 (5')
SMT2 275 F3 (5')
SMT2 275 F4 (5')
SMT2 275 F5 (5')
SMT2 175 F1 (5')
SMT2 175 F2 (5')
SMT2 175 F3 (5')
SMT2 F1 (3')
SMT2 F2 (3')
SMT2 F3 (3')
SMT2 F4 (3')
SMT2 F5 (3')
SMT2 F6 (3')
SMT2 F7 (3')
SMT2 F8 (3')
Olive SMT2 contig
At-SMT2-1
At-SMT2-2
GmSMT2-1
GmSMT2-2
GhSMT2-1
GhSMT2-2
RcSMT2
NtSMT2

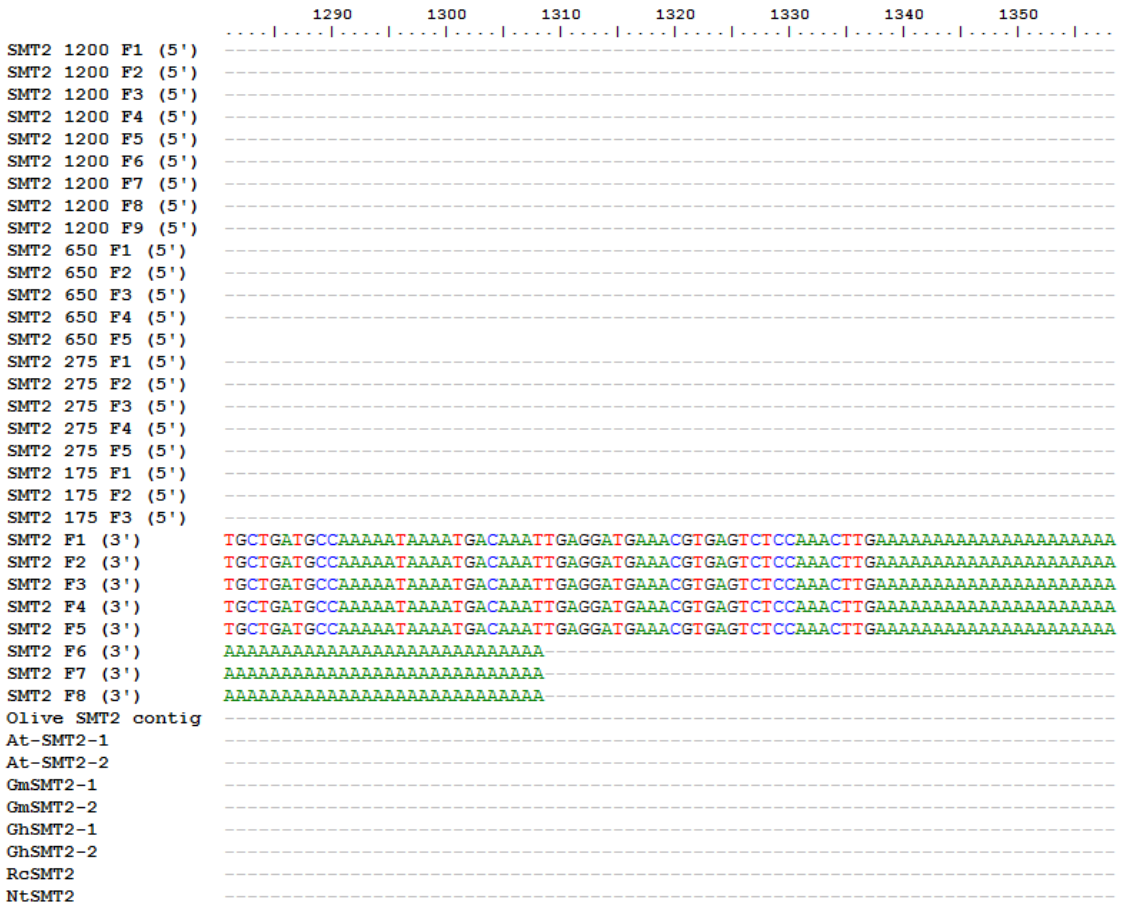
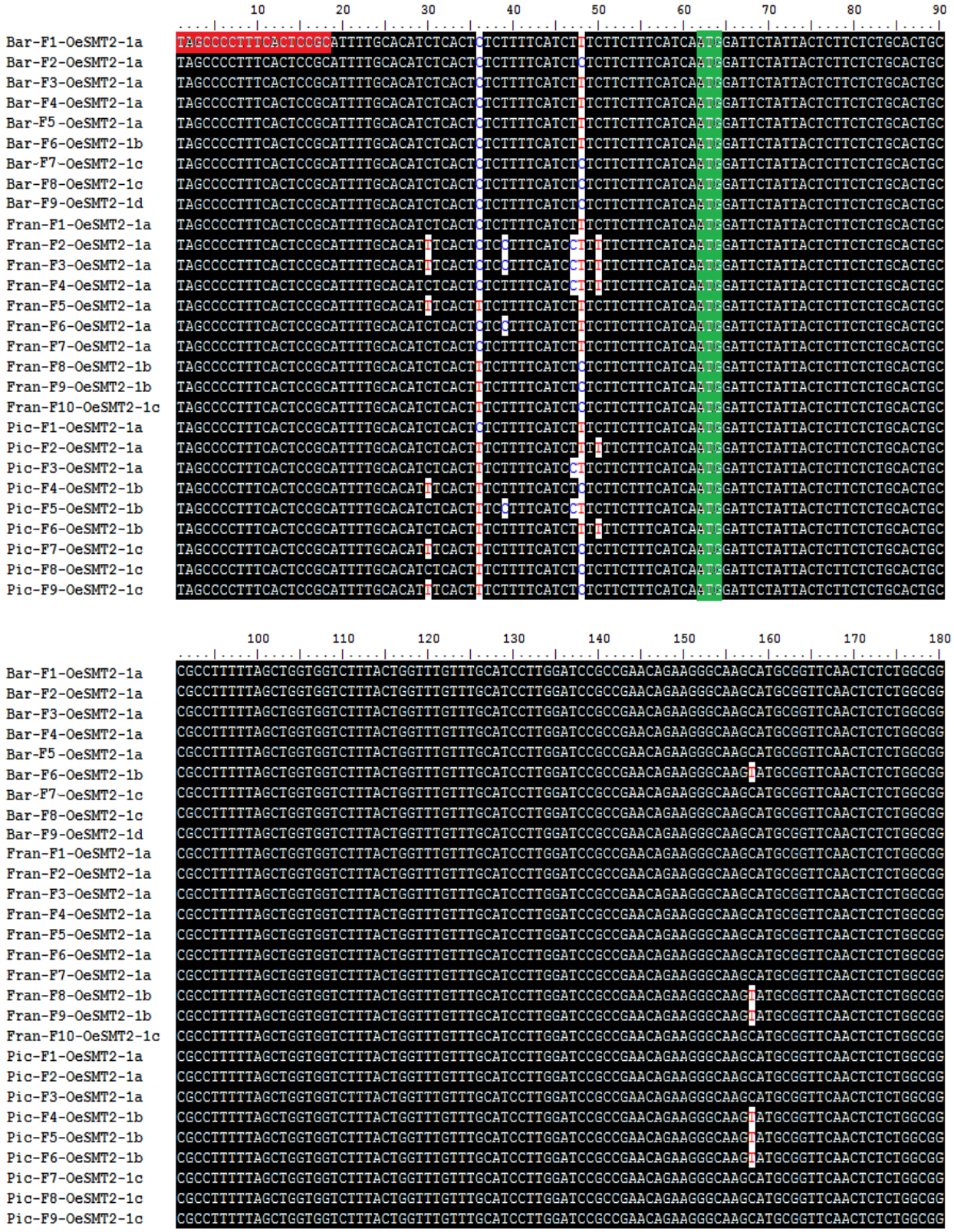
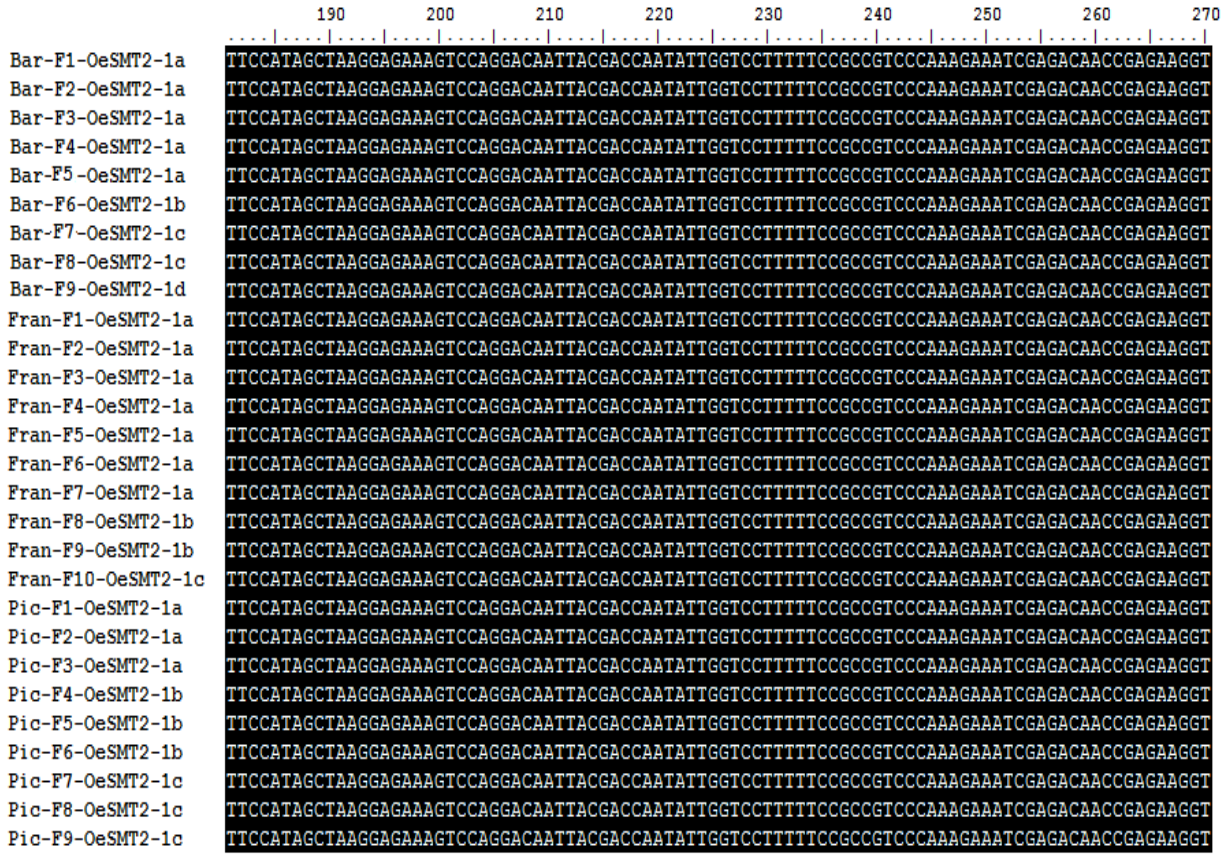
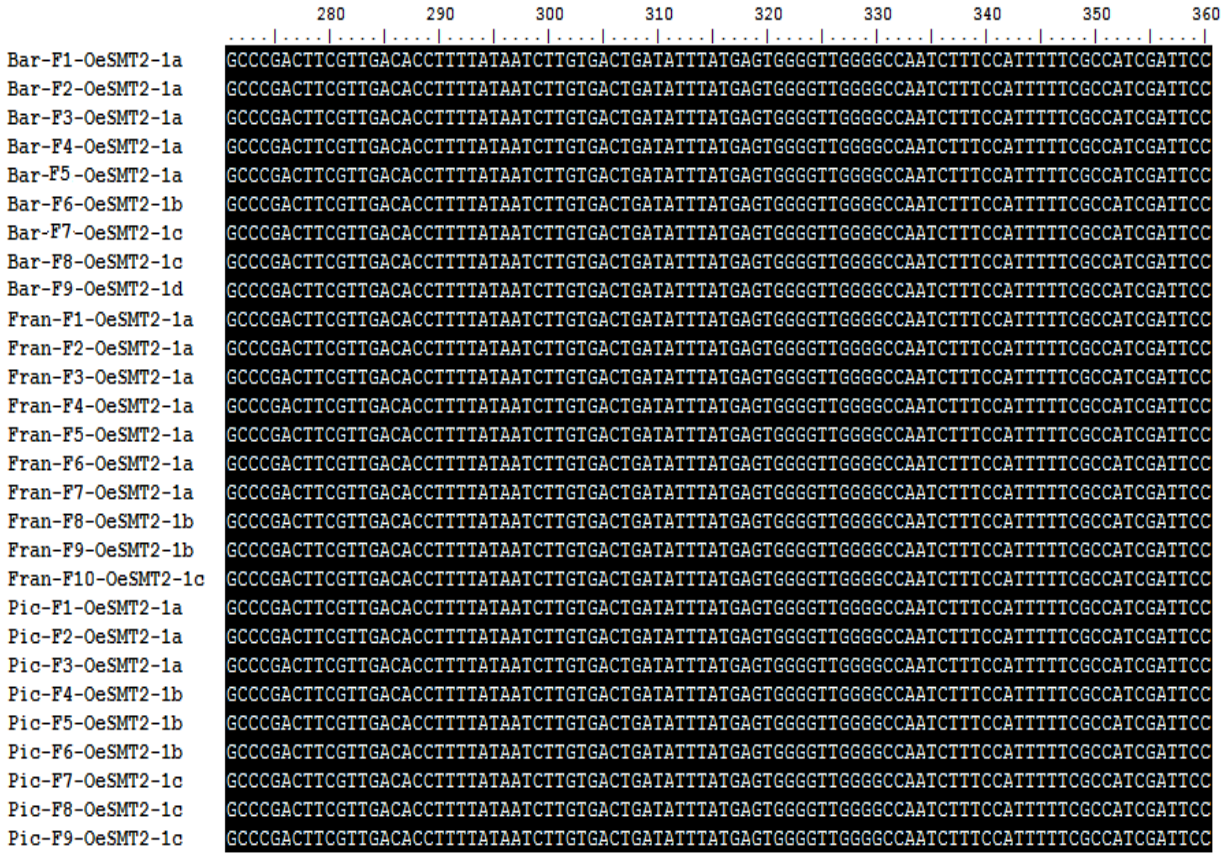

1290
1300
1310
1320
1330
1340
1350
SMT2 1200 F1 (5')
SMT2 1200 F2 (5')
SMT2 1200 F3 (5')
SMT2 1200 F4 (5')
SMT2 1200 F5 (5')
SMT2 1200 F6 (5')
SMT2 1200 F7 (5')
SMT2 1200 F8 (5')
SMT2 1200 F9 (5')
SMT2 650 F1 (5')
SMT2 650 F2 (5')
SMT2 650 F3 (5')
SMT2 650 F4 (5')
SMT2 650 F5 (5')
SMT2 275 F1 (5')
SMT2 275 F2 (5')
SMT2 275 F3 (5')
SMT2 275 F4 (5')
SMT2 275 F5 (5')
SMT2 175 F1 (5')
SMT2 175 F2 (5')
SMT2 175 F3 (5')
SMT2 F1 (3')
SMT2 F2 (3')
SMT2 F3 (3')
SMT2 F4 (3')
SMT2 F5 (3')
SMT2 F6 (3')
SMT2 F7 (3')
SMT2 F8 (3')
Olive SMT2 contig
At-SMT2-1
At-SMT2-2
GmSMT2-1
GmSMT2-2
GhSMT2-1
GhSMT2-2
RcSMT2
NtSMT2

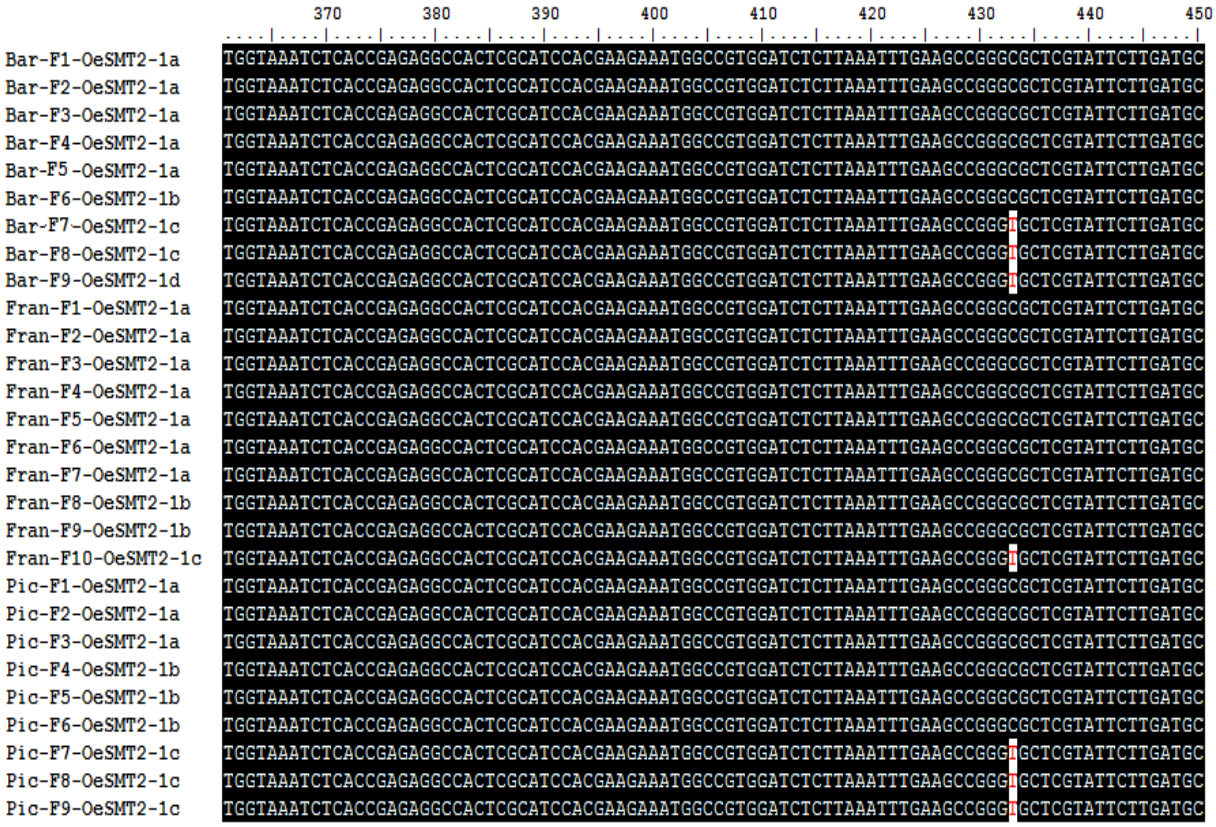
Figure IIIA Master alignment of *SMT2* 5'and3' RACE ten clones deduced from olive cultivarBarnea and previously published *SMT2* coding sequences from other plants.

5' RACE cDNA- 9 clones of ~1200bp products (FI to F9); 5 clones of ~650bp products (FI to F5); 5 clones of ~275bp products (FI to F5); 3 clones of ~175bp products (FI to F3); 3' RACE cDNA- 8 clones (F1 to F8). The olive *SMT2*contig (Cluster Id-OLEEUCl064665) derived from the Alagna *et al* data (2009) which was used to isolate full length *SMT2* from Barnea is highlighted in light green. Primers used for RACE PCR are highlighted in yellow and primers used for full length *SMT2* amplification are highlighted in deep green. The START codon (ATG) and STOP codon (TAG) are highlighted in yellow. Rc: *Riccinus communis* SMT (Accession: XM_002510554.1),Gm1: *Glycine max* SMT2-1(Accession:FJ483973.1), Gm2: *Glycine max* SMT2-2(Accession:FJ483974.1), Gh1: *Gossipum hirsutum* SMT2-1 (Accession:EU308589.1), Gh2: *Gossipum hirsutum* SMT2-2 (Accession:EU308590.1), At2-1: *Arabidopsis thaliana* SMT2-1 (Accession: NM101884.3), At2-2: *Arabidopsis thaliana* SMT2-2 (Accession: U71400.1).


10 20 30 40 50 60 70 80 90
Bar-F1-OeSMT2-1a
Bar-F2-OeSMT2-1a
Bar-F3-OeSMT2-1a
Bar-F4-OeSMT2-1a
Bar-F5-OeSMT2-1a
Bar-F6-OeSMT2-1b
Bar-F7-OeSMT2-1c
Bar-F8-OeSMT2-1c
Bar-F9-OeSMT2-1d
Fran-F1-OeSMT2-1a
Fran-F2-OeSMT2-1a
Fran-F3-OeSMT2-1a
Fran-F4-OeSMT2-1a
Fran-F5-OeSMT2-1a
Fran-F6-OeSMT2-1a
Fran-F7-OeSMT2-1a
Fran-F8-OeSMT2-1b
Fran-F9-OeSMT2-1b
Fran-F10-OeSMT2-1c
Pic-F1-OeSMT2-1a
Pic-F2-OeSMT2-1a
Pic-F3-OeSMT2-1a
Pic-F4-OeSMT2-1b
Pic-F5-OeSMT2-1b
Pic-F6-OeSMT2-1b
Pic-F7-OeSMT2-1c
Pic-F8-OeSMT2-1c
Pic-F9-OeSMT2-1c
100 110 120 130 140 150 160 170 180
Bar-F1-OeSMT2-1a
Bar-F2-OeSMT2-1a
Bar-F3-OeSMT2-1a
Bar-F4-OeSMT2-1a
Bar-F5-OeSMT2-1a
Bar-F6-OeSMT2-1b
Bar-F7-OeSMT2-1c
Bar-F8-OeSMT2-1c
Bar-F9-OeSMT2-1d
Fran-F1-OeSMT2-1a
Fran-F2-OeSMT2-1a
Fran-F3-OeSMT2-1a
Fran-F4-OeSMT2-1a
Fran-F5-OeSMT2-1a
Fran-F6-OeSMT2-1a
Fran-F7-OeSMT2-1a
Fran-F8-OeSMT2-1b
Fran-F9-OeSMT2-1b
Fran-F10-OeSMT2-1c
Pic-F1-OeSMT2-1a
Pic-F2-OeSMT2-1a
Pic-F3-OeSMT2-1a
Pic-F4-OeSMT2-1b
Pic-F5-OeSMT2-1b
Pic-F6-OeSMT2-1b
Pic-F7-OeSMT2-1c
Pic-F8-OeSMT2-1c
Pic-F9-OeSMT2-1c

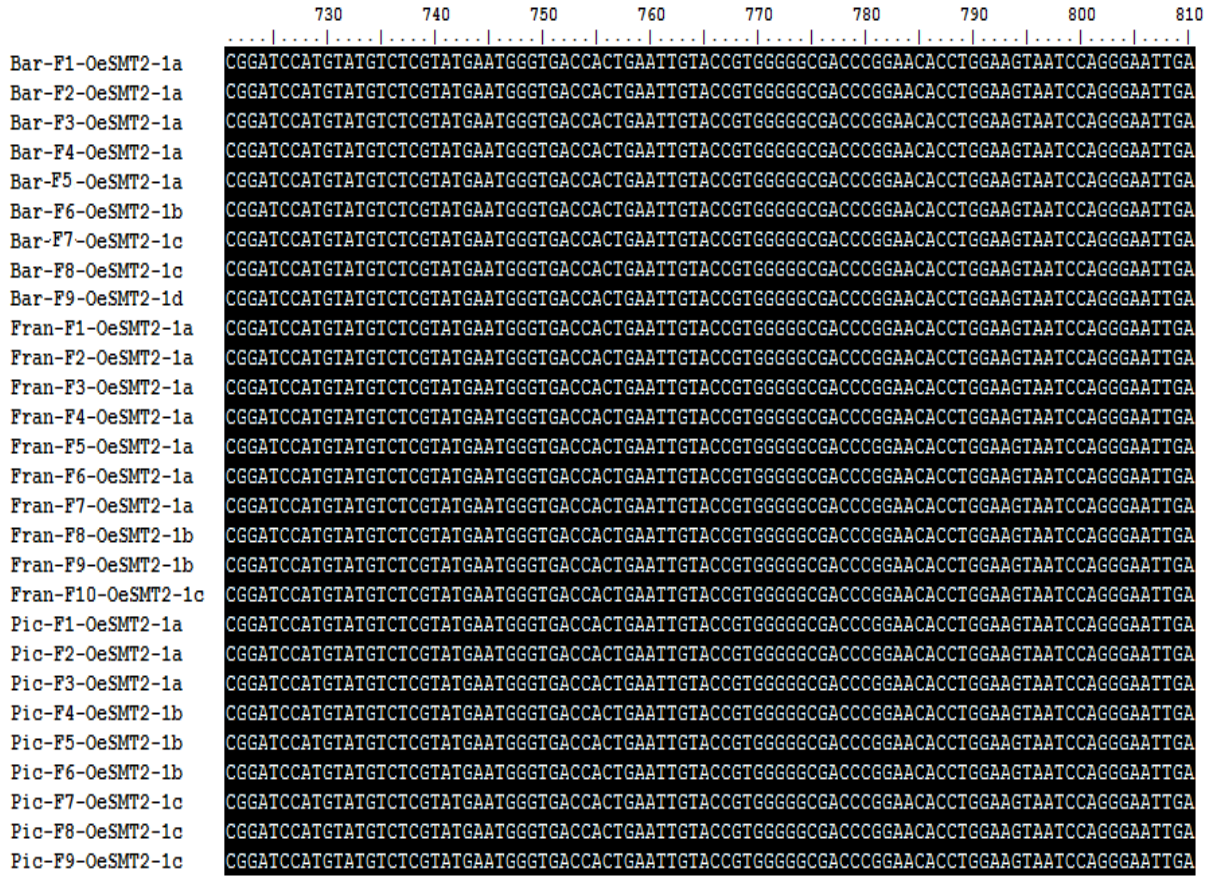

190 200 210 220 230 240 250 260 270
Bar-F1-OeSMT2-1a TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Bar-F2-OeSMT2-1a TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Bar-F3-OeSMT2-1a TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Bar-F4-OeSMT2-1a TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Bar-F5-OeSMT2-1a TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Bar-F6-OeSMT2-1b TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Bar-F7-OeSMT2-1c TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Bar-F8-OeSMT2-1c TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Bar-F9-OeSMT2-1d TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Fran-F1-OeSMT2-1a TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Fran-F2-OeSMT2-1a TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Fran-F3-OeSMT2-1a TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Fran-F4-OeSMT2-1a TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Fran-F5-OeSMT2-1a TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Fran-F6-OeSMT2-1a TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Fran-F7-OeSMT2-1a TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Fran-F8-OeSMT2-1b TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Fran-F9-OeSMT2-1b TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Fran-F10-OeSMT2-1c TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Pic-F1-OeSMT2-1a TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Pic-F2-OeSMT2-1a TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Pic-F3-OeSMT2-1a TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Pic-F4-OeSMT2-1b TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Pic-F5-OeSMT2-1b TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Pic-F6-OeSMT2-1b TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Pic-F7-OeSMT2-1c TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Pic-F8-OeSMT2-1c TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT
Pic-F9-OeSMT2-1c TTCCATAGCTAAGGAGAAAGTCCAGGACAATTACGACCAATATTGGTCCTTTTTCCGCCGTCCCAAAGAAATCGAGACAACCGAGAAGGT

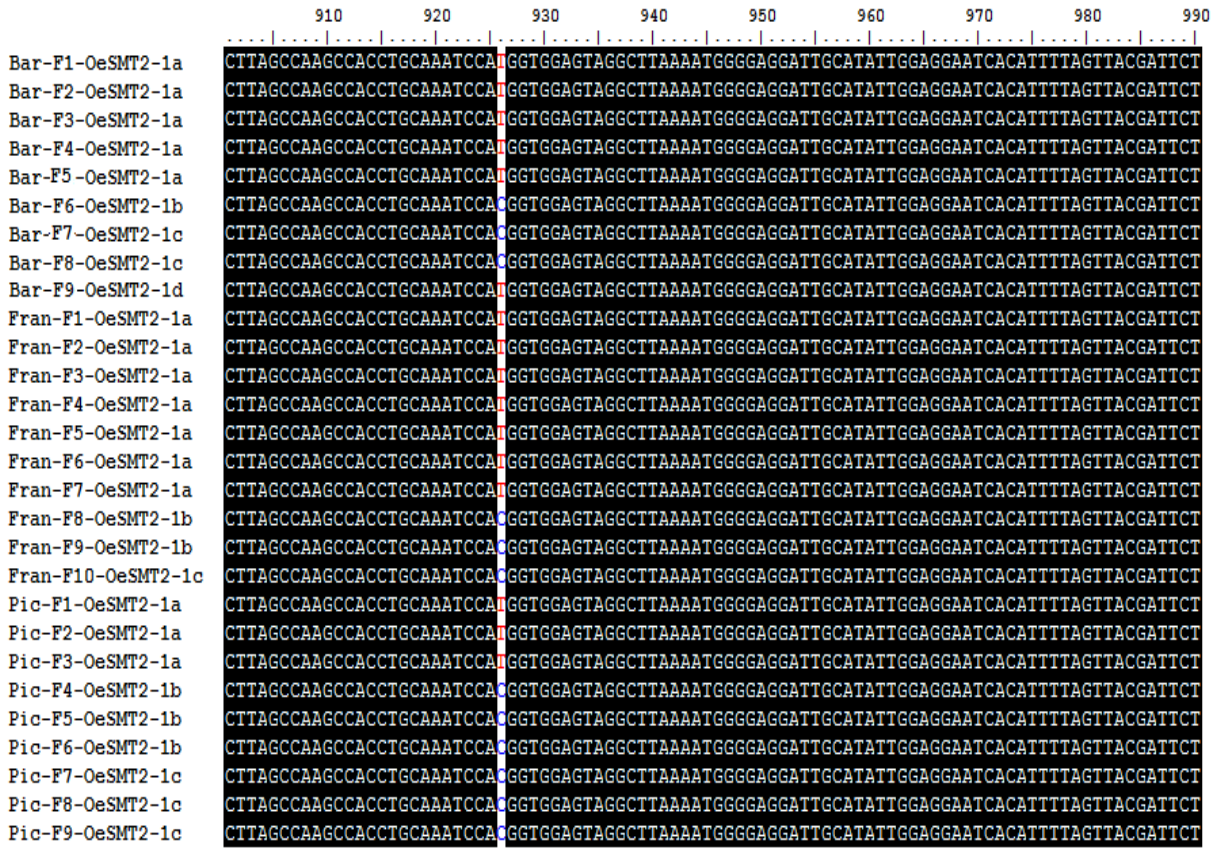

280 290 300 310 320 330 340 350 360
Bar-F1-OeSMT2-1a GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Bar-F2-OeSMT2-1a GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Bar-F3-OeSMT2-1a GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Bar-F4-OeSMT2-1a GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Bar-F5-OeSMT2-1a GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Bar-F6-OeSMT2-1b GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Bar-F7-OeSMT2-1c GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Bar-F8-OeSMT2-1c GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Bar-F9-OeSMT2-1d GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Fran-F1-OeSMT2-1a GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Fran-F2-OeSMT2-1a GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Fran-F3-OeSMT2-1a GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Fran-F4-OeSMT2-1a GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Fran-F5-OeSMT2-1a GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Fran-F6-OeSMT2-1a GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Fran-F7-OeSMT2-1a GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Fran-F8-OeSMT2-1b GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Fran-F9-OeSMT2-1b GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Fran-F10-OeSMT2-1c GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Pic-F1-OeSMT2-1a GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Pic-F2-OeSMT2-1a GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Pic-F3-OeSMT2-1a GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Pic-F4-OeSMT2-1b GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Pic-F5-OeSMT2-1b GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Pic-F6-OeSMT2-1b GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Pic-F7-OeSMT2-1c GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Pic-F8-OeSMT2-1c GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC
Pic-F9-OeSMT2-1c GCCCGACTTCGTTGACACCTTTTATAATCTTGTGACTGATATTTATGAGTGGGGTTGGGGCCAATCTTTCCATTTTTCGCCATCGATTCC

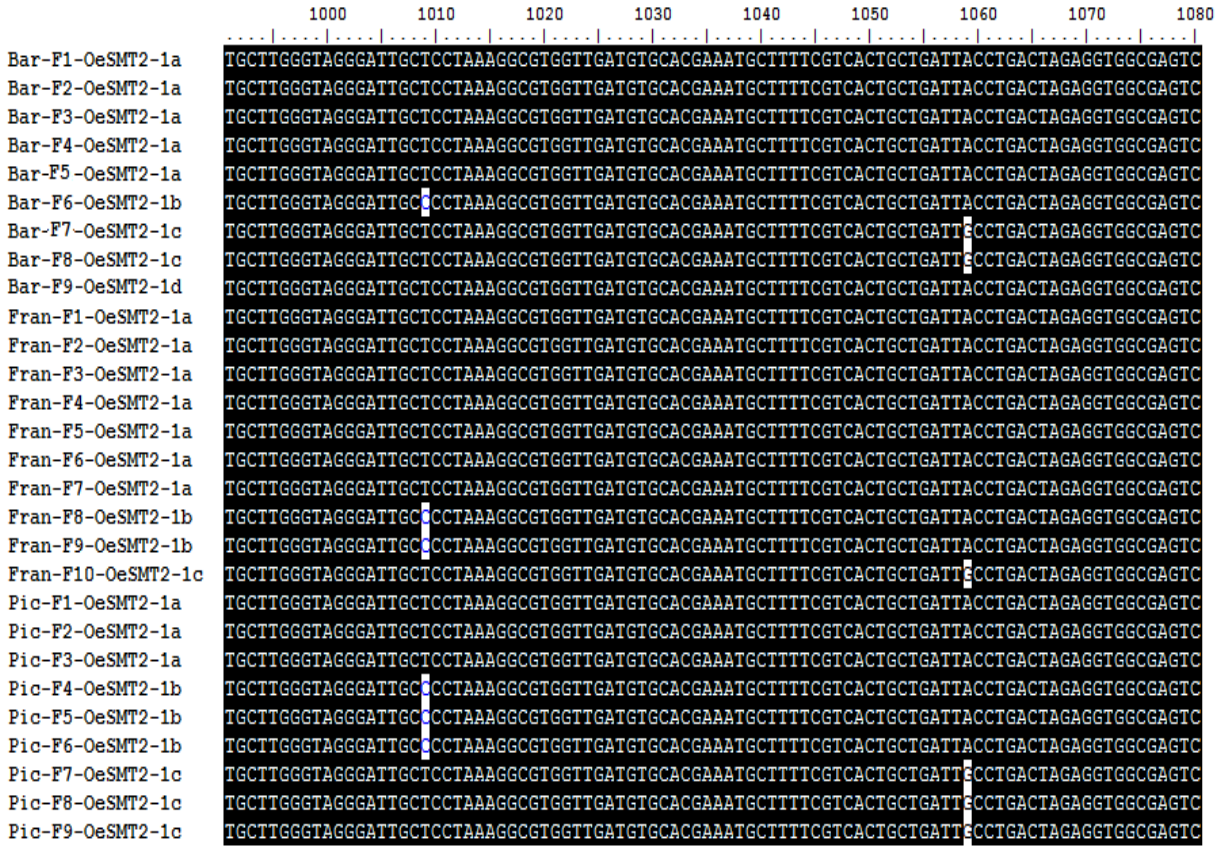

370 380 390 400 410 420 430 440 450
Bar-F1-OeSMT2-1a TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Bar-F2-OeSMT2-1a TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Bar-F3-OeSMT2-1a TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Bar-F4-OeSMT2-1a TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Bar-F5-OeSMT2-1a TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Bar-F6-OeSMT2-1b TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Bar-F7-OeSMT2-1c TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGTGCTCGTATTCTTGATGC
Bar-F8-OeSMT2-1c TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGTGCTCGTATTCTTGATGC
Bar-F9-OeSMT2-1d TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGTGCTCGTATTCTTGATGC
Fran-F1-OeSMT2-1a TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Fran-F2-OeSMT2-1a TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Fran-F3-OeSMT2-1a TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Fran-F4-OeSMT2-1a TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Fran-F5-OeSMT2-1a TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Fran-F6-OeSMT2-1a TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Fran-F7-OeSMT2-1a TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Fran-F8-OeSMT2-1b TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Fran-F9-OeSMT2-1b TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Fran-F10-OeSMT2-1c TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGTGCTCGTATTCTTGATGC
Pic-F1-OeSMT2-1a TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Pic-F2-OeSMT2-1a TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Pic-F3-OeSMT2-1a TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Pic-F4-OeSMT2-1b TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Pic-F5-OeSMT2-1b TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Pic-F6-OeSMT2-1b TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGCGCTCGTATTCTTGATGC
Pic-F7-OeSMT2-1c TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGTGCTCGTATTCTTGATGC
Pic-F8-OeSMT2-1c TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGTGCTCGTATTCTTGATGC
Pic-F9-OeSMT2-1c TGGTAAATCTCACCGAGAGGCCACTCGCATCCACGAAGAAATGGCCGTGGATCTCTTAAATTTGAAGCCGGGTGCTCGTATTCTTGATGC
460 470 480 490 500 510 520 530 540
Bar-F1-OeSMT2-1a GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Bar-F2-OeSMT2-1a GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Bar-F3-OeSMT2-1a GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Bar-F4-OeSMT2-1a GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Bar-F5-OeSMT2-1a GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Bar-F6-OeSMT2-1b GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Bar-F7-OeSMT2-1c GGGGTGTGGTGTGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Bar-F8-OeSMT2-1c GGGGTGTGGTGTGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Bar-F9-OeSMT2-1d GGGGTGTGGTGTGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Fran-F1-OeSMT2-1a GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Fran-F2-OeSMT2-1a GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Fran-F3-OeSMT2-1a GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Fran-F4-OeSMT2-1a GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Fran-F5-OeSMT2-1a GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Fran-F6-OeSMT2-1a GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Fran-F7-OeSMT2-1a GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Fran-F8-OeSMT2-1b GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Fran-F9-OeSMT2-1b GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Fran-F10-OeSMT2-1c GGGGTGTGGTGTGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Pic-F1-OeSMT2-1a GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Pic-F2-OeSMT2-1a GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Pic-F3-OeSMT2-1a GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Pic-F4-OeSMT2-1b GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Pic-F5-OeSMT2-1b GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Pic-F6-OeSMT2-1b GGGGTGTGGTGTAGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Pic-F7-OeSMT2-1c GGGGTGTGGTGTGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Pic-F8-OeSMT2-1c GGGGTGTGGTGTGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA
Pic-F9-OeSMT2-1c GGGGTGTGGTGTGGTGGTCCGATGCGGGCGATTGCAGCCCATTCAGGGGCTAATGTAGTCGGAATCACCATTAACGAATATCAAGTAAA

```
                          550        560        570        580        590        600        610        620        630
                          |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Bar-F1-OeSMT2-1a          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Bar-F2-OeSMT2-1a          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Bar-F3-OeSMT2-1a          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Bar-F4-OeSMT2-1a          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Bar-F5-OeSMT2-1a          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Bar-F6-OeSMT2-1b          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Bar-F7-OeSMT2-1c          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Bar-F8-OeSMT2-1c          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Bar-F9-OeSMT2-1d          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Fran-F1-OeSMT2-1a         TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Fran-F2-OeSMT2-1a         TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Fran-F3-OeSMT2-1a         TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Fran-F4-OeSMT2-1a         TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Fran-F5-OeSMT2-1a         TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Fran-F6-OeSMT2-1a         TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Fran-F7-OeSMT2-1a         TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Fran-F8-OeSMT2-1b         TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Fran-F9-OeSMT2-1b         TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Fran-F10-OeSMT2-1c        TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Pic-F1-OeSMT2-1a          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Pic-F2-OeSMT2-1a          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Pic-F3-OeSMT2-1a          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Pic-F4-OeSMT2-1b          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Pic-F5-OeSMT2-1b          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Pic-F6-OeSMT2-1b          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Pic-F7-OeSMT2-1c          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Pic-F8-OeSMT2-1c          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA
Pic-F9-OeSMT2-1c          TCGGGCGAGGGCACACAACAAGAAGGCCGGCCTTGATAAACAATGCGAAGTCGTGTGCGGCAATTTCCTCGAGATGCCCTTCGGTGATAA

                          640        650        660        670        680        690        700        710        720
                          |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Bar-F1-OeSMT2-1a          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Bar-F2-OeSMT2-1a          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Bar-F3-OeSMT2-1a          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Bar-F4-OeSMT2-1a          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Bar-F5-OeSMT2-1a          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Bar-F6-OeSMT2-1b          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Bar-F7-OeSMT2-1c          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGGGATCTACCGGGTTTTAAAGCC
Bar-F8-OeSMT2-1c          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGGGATCTACCGGGTTTTAAAGCC
Bar-F9-OeSMT2-1d          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Fran-F1-OeSMT2-1a         CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Fran-F2-OeSMT2-1a         CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Fran-F3-OeSMT2-1a         CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Fran-F4-OeSMT2-1a         CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Fran-F5-OeSMT2-1a         CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Fran-F6-OeSMT2-1a         CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Fran-F7-OeSMT2-1a         CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Fran-F8-OeSMT2-1b         CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Fran-F9-OeSMT2-1b         CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Fran-F10-OeSMT2-1c        CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGGGATCTACCGGGTTTTAAAGCC
Pic-F1-OeSMT2-1a          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Pic-F2-OeSMT2-1a          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Pic-F3-OeSMT2-1a          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Pic-F4-OeSMT2-1b          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Pic-F5-OeSMT2-1b          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Pic-F6-OeSMT2-1b          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGAGATCTACCGGGTTTTAAAGCC
Pic-F7-OeSMT2-1c          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGGGATCTACCGGGTTTTAAAGCC
Pic-F8-OeSMT2-1c          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGGGATCTACCGGGTTTTAAAGCC
Pic-F9-OeSMT2-1c          CAGTTTCGACGGAGCCTATTCAATTGAAGCCACTTGCCATGCACCAAAACTGGAAGATGTATACAGCGGGATCTACCGGGTTTTAAAGCC
```


730 740 750 760 770 780 790 800 810
Bar-F1-OeSMT2-1a CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Bar-F2-OeSMT2-1a CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Bar-F3-OeSMT2-1a CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Bar-F4-OeSMT2-1a CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Bar-F5-OeSMT2-1a CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Bar-F6-OeSMT2-1b CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Bar-F7-OeSMT2-1c CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Bar-F8-OeSMT2-1c CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Bar-F9-OeSMT2-1d CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Fran-F1-OeSMT2-1a CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Fran-F2-OeSMT2-1a CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Fran-F3-OeSMT2-1a CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Fran-F4-OeSMT2-1a CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Fran-F5-OeSMT2-1a CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Fran-F6-OeSMT2-1a CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Fran-F7-OeSMT2-1a CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Fran-F8-OeSMT2-1b CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Fran-F9-OeSMT2-1b CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Fran-F10-OeSMT2-1c CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Pic-F1-OeSMT2-1a CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Pic-F2-OeSMT2-1a CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Pic-F3-OeSMT2-1a CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Pic-F4-OeSMT2-1b CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Pic-F5-OeSMT2-1b CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Pic-F6-OeSMT2-1b CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Pic-F7-OeSMT2-1c CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Pic-F8-OeSMT2-1c CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
Pic-F9-OeSMT2-1c CGGATCCATGTATGTCTCGTATGAATGGGTGACCACTGAATTGTACCGTGGGGGCGACCCGGAACACCTGGAAGTAATCCAGGGAATTGA
820 830 840 850 860 870 880 890 900
Bar-F1-OeSMT2-1a GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Bar-F2-OeSMT2-1a GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Bar-F3-OeSMT2-1a GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Bar-F4-OeSMT2-1a GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Bar-F5-OeSMT2-1a GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Bar-F6-OeSMT2-1b GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGNAGGA
Bar-F7-OeSMT2-1c GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Bar-F8-OeSMT2-1c GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Bar-F9-OeSMT2-1d GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Fran-F1-OeSMT2-1a GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Fran-F2-OeSMT2-1a GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Fran-F3-OeSMT2-1a GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Fran-F4-OeSMT2-1a GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Fran-F5-OeSMT2-1a GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Fran-F6-OeSMT2-1a GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Fran-F7-OeSMT2-1a GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Fran-F8-OeSMT2-1b GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGNAGGA
Fran-F9-OeSMT2-1b GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGNAGGA
Fran-F10-OeSMT2-1c GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Pic-F1-OeSMT2-1a GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Pic-F2-OeSMT2-1a GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Pic-F3-OeSMT2-1a GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Pic-F4-OeSMT2-1b GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGNAGGA
Pic-F5-OeSMT2-1b GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGNAGGA
Pic-F6-OeSMT2-1b GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGNAGGA
Pic-F7-OeSMT2-1c GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Pic-F8-OeSMT2-1c GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA
Pic-F9-OeSMT2-1c GAGAGGCGATGCACTACCCGGGCTAAGGAGTTACAAGGATATCGCAGAGGTTGCGAGAAAGGTGGGATTTGAAGTAATAAGAGAGAAGGA


910 920 930 940 950 960 970 980 990
Bar-F1-OeSMT2-1a CTTAGCCAAGCCACCTGCAAATCCAT GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Bar-F2-OeSMT2-1a CTTAGCCAAGCCACCTGCAAATCCAT GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Bar-F3-OeSMT2-1a CTTAGCCAAGCCACCTGCAAATCCAT GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Bar-F4-OeSMT2-1a CTTAGCCAAGCCACCTGCAAATCCAT GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Bar-F5-OeSMT2-1a CTTAGCCAAGCCACCTGCAAATCCAT GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Bar-F6-OeSMT2-1b CTTAGCCAAGCCACCTGCAAATCCAC GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Bar-F7-OeSMT2-1c CTTAGCCAAGCCACCTGCAAATCCAC GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Bar-F8-OeSMT2-1c CTTAGCCAAGCCACCTGCAAATCCAC GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Bar-F9-OeSMT2-1d CTTAGCCAAGCCACCTGCAAATCCAT GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Fran-F1-OeSMT2-1a CTTAGCCAAGCCACCTGCAAATCCAT GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Fran-F2-OeSMT2-1a CTTAGCCAAGCCACCTGCAAATCCAT GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Fran-F3-OeSMT2-1a CTTAGCCAAGCCACCTGCAAATCCAT GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Fran-F4-OeSMT2-1a CTTAGCCAAGCCACCTGCAAATCCAT GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Fran-F5-OeSMT2-1a CTTAGCCAAGCCACCTGCAAATCCAT GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Fran-F6-OeSMT2-1a CTTAGCCAAGCCACCTGCAAATCCAT GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Fran-F7-OeSMT2-1a CTTAGCCAAGCCACCTGCAAATCCAT GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Fran-F8-OeSMT2-1b CTTAGCCAAGCCACCTGCAAATCCAC GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Fran-F9-OeSMT2-1b CTTAGCCAAGCCACCTGCAAATCCAC GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Fran-F10-OeSMT2-1c CTTAGCCAAGCCACCTGCAAATCCAC GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Pic-F1-OeSMT2-1a CTTAGCCAAGCCACCTGCAAATCCAT GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Pic-F2-OeSMT2-1a CTTAGCCAAGCCACCTGCAAATCCAT GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Pic-F3-OeSMT2-1a CTTAGCCAAGCCACCTGCAAATCCAT GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Pic-F4-OeSMT2-1b CTTAGCCAAGCCACCTGCAAATCCAC GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Pic-F5-OeSMT2-1b CTTAGCCAAGCCACCTGCAAATCCAC GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Pic-F6-OeSMT2-1b CTTAGCCAAGCCACCTGCAAATCCAC GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Pic-F7-OeSMT2-1c CTTAGCCAAGCCACCTGCAAATCCAC GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Pic-F8-OeSMT2-1c CTTAGCCAAGCCACCTGCAAATCCAC GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT
Pic-F9-OeSMT2-1c CTTAGCCAAGCCACCTGCAAATCCAC GGTGGAGTAGGCTTAAAATGGGGAGGATTGCATATTGGAGGAATCACATTTTAGTTACGATTCT


1000 1010 1020 1030 1040 1050 1060 1070 1080
Bar-F1-OeSMT2-1a TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Bar-F2-OeSMT2-1a TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Bar-F3-OeSMT2-1a TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Bar-F4-OeSMT2-1a TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Bar-F5-OeSMT2-1a TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Bar-F6-OeSMT2-1b TGCTTGGGTAGGGATTGCCCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Bar-F7-OeSMT2-1c TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTCCCTGACTAGAGGTGGCGAGTC
Bar-F8-OeSMT2-1c TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTCCCTGACTAGAGGTGGCGAGTC
Bar-F9-OeSMT2-1d TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Fran-F1-OeSMT2-1a TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Fran-F2-OeSMT2-1a TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Fran-F3-OeSMT2-1a TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Fran-F4-OeSMT2-1a TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Fran-F5-OeSMT2-1a TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Fran-F6-OeSMT2-1a TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Fran-F7-OeSMT2-1a TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Fran-F8-OeSMT2-1b TGCTTGGGTAGGGATTGCCCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Fran-F9-OeSMT2-1b TGCTTGGGTAGGGATTGCCCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Fran-F10-OeSMT2-1c TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTCCCTGACTAGAGGTGGCGAGTC
Pic-F1-OeSMT2-1a TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Pic-F2-OeSMT2-1a TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Pic-F3-OeSMT2-1a TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Pic-F4-OeSMT2-1b TGCTTGGGTAGGGATTGCCCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Pic-F5-OeSMT2-1b TGCTTGGGTAGGGATTGCCCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Pic-F6-OeSMT2-1b TGCTTGGGTAGGGATTGCCCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTACCTGACTAGAGGTGGCGAGTC
Pic-F7-OeSMT2-1c TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTCCCTGACTAGAGGTGGCGAGTC
Pic-F8-OeSMT2-1c TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTCCCTGACTAGAGGTGGCGAGTC
Pic-F9-OeSMT2-1c TGCTTGGGTAGGGATTGCTCCTAAAGGCGTGGTTGATGTGCACGAAATGCTTTCGTCACTGCTGATTCCCTGACTAGAGGTGGCGAGTC


1090 1100 1110 1120 1130 1140 1150 1160 1170
Bar-F1-OeSMT2-1a GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATAGAGAGTTTGGTGTTCAATTTGATCATA
Bar-F2-OeSMT2-1a GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATAGAGAGTTTGGTGTTCAATTTGATCATA
Bar-F3-OeSMT2-1a GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATAGAGAGTTTGGTGTTCAATTTGATCATA
Bar-F4-OeSMT2-1a GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATAGAGAGTTTGGTGTTCAACTTGATCATA
Bar-F5-OeSMT2-1a GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATAGAGAGTTTGGTGTTCAACTTGATCATA
Bar-F6-OeSMT2-1b GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCAYCGTCATAGAGAGTTYGGTGTTCAACTTGATCATA
Bar-F7-OeSMT2-1c GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATARAGAGTTTGGTGTTCAACTTGATCATA
Bar-F8-OeSMT2-1c GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATARAGAGTTTGGTGTTCAACTTGATCATA
Bar-F9-OeSMT2-1d GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCAYCGTCATAGAGAGTTTGGTGTTCAACTTGATCATA
Fran-F1-OeSMT2-1a GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATAGAGAGTTTGGTGTTCAATTTGATCATA
Fran-F2-OeSMT2-1a GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATAGAGAGTTTGGTGTTCAATTTGATCATA
Fran-F3-OeSMT2-1a GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATAGAGAGTTTGGTGTTCAATTTGATCATA
Fran-F4-OeSMT2-1a GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATAGAGAGTTTGGTGTTCAATTTGATCATA
Fran-F5-OeSMT2-1a GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATAGAGAGTTTGGTGTTCAATTTGATCAYA
Fran-F6-OeSMT2-1a GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATAGAGAGTTTGGTGTTCAATTTGATCAYA
Fran-F7-OeSMT2-1a GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATAGAGAGTTTGGGGTTCAATTTGATCATA
Fran-F8-OeSMT2-1b GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCAYCGTCATAGAGAGTTTGGTGTTCAATTTGATCATA
Fran-F9-OeSMT2-1b GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCAYCGTCATAGAGAGTTTGGTGTTCAATTTGATCATA
Fran-F10-OeSMT2-1c GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATARAGAGTTTGGGGTTCAATTTGATCATA
Pic-F1-OeSMT2-1a GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATAGAGAGTTTGGGGTTCAATTTGATCATA
Pic-F2-OeSMT2-1a GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATAGAGAGTTTGGGGTTCAATTTGATCATA
Pic-F3-OeSMT2-1a GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATAGAGAGTTTGGGGTTCAATTTGATCAYA
Pic-F4-OeSMT2-1b GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCAYCGTCATAGAGAGTTTGGGGTTCAATTTGATCATA
Pic-F5-OeSMT2-1b GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCAYCGTCATAGAGAGTTTGGGGTTCAATTTGATCATA
Pic-F6-OeSMT2-1b GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCAYCGTCATAGAGAGTTTGGGGTTCAATTTGATCATA
Pic-F7-OeSMT2-1c GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATARAGAGTTTGGGGTTCAATTTGATCATA
Pic-F8-OeSMT2-1c GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATARAGAGTTTGGGGTTCAATTTGATCATA
Pic-F9-OeSMT2-1c GGGAATTTTCACTCCTATGCACATGATTCTTTGCAAAAAGCCTGAAATTAAGTCATCGTCATARAGAGTTTGGGGTTCAATTTGATCATA
1180 1190 1200
Bar-F1-OeSMT2-1a ATAATCTATGTTAGATTTTGTTGGTTATTTG
Bar-F2-OeSMT2-1a ATAATCTATGTTAGATTTTGTTGGTTATTTG
Bar-F3-OeSMT2-1a ATAATCTATGTTAGATTTTGTTGGTTATTTG
Bar-F4-OeSMT2-1a ATAATCTATGTTAGATTTTGTTGGTTATTTG
Bar-F5-OeSMT2-1a ATAATCTATGTTAGATTTTGTTGGTTATTTG
Bar-F6-OeSMT2-1b ATAATCTATGTTAGATTTTGTTGGTTATTTG
Bar-F7-OeSMT2-1c ATAATCTATGTTAGATTTTGTTGGTTATTTG
Bar-F8-OeSMT2-1c ATAATCTATGTTAGATTTTGTTGGTTATTTG
Bar-F9-OeSMT2-1d ATAATCTATGTTAGATTTTGTTGGTTATTTG
Fran-F1-OeSMT2-1a ATAATCTATGTTAGATTTTGTTGGTTATTTG
Fran-F2-OeSMT2-1a AWAATCTATGTTAGATTTTGTTGGTTATTTG
Fran-F3-OeSMT2-1a ATAATCTATGTTAGATTTTGTTGGTTATTTG
Fran-F4-OeSMT2-1a ATAATCTATGTTAGATTTTGTTGGTTATTTG
Fran-F5-OeSMT2-1a AWAATCTATGTTAGATTTTGTTGGTTATTTG
Fran-F6-OeSMT2-1a ATAATCTATGGTAWATTTTGGTGGTTATTTG
Fran-F7-OeSMT2-1a ATAATCTATGGTAWATTTTGGTGGTTATTTG
Fran-F8-OeSMT2-1b ATAATCTATGTTAGATTTTGTTGGTTATTTG
Fran-F9-OeSMT2-1b ATAATCTATGTTAGATTTTGTTGGTTATTTG
Fran-F10-OeSMT2-1c ATAATCTATGTTAGATTTTGTTGGTTATTTG
Pic-F1-OeSMT2-1a ATAATCTATGTTAGATTTTGTTGGTTATTTG
Pic-F2-OeSMT2-1a ATAATCTATGGTAWATTTTGGTGGTTATTTG
Pic-F3-OeSMT2-1a AWAATCTATGGTAWATTTTGGTGGTTATTTG
Pic-F4-OeSMT2-1b ATAATCTATGGTAWATTTTGGTGGTTATTTG
Pic-F5-OeSMT2-1b ATAATCTATGGTAWATTTTGGTGGTTATTTG
Pic-F6-OeSMT2-1b ATAATCTATGGTAWATTTTGTTGGTTATTTG
Pic-F7-OeSMT2-1c ATAATCTATGTTAGATTTTGTTGGTTATTTG
Pic-F8-OeSMT2-1c ATAATCTATGTTAGATTTTGTTGGTTATTTG
Pic-F9-OeSMT2-1c ATAATCTATGTTAGATTTTGTTGGTTATTTG

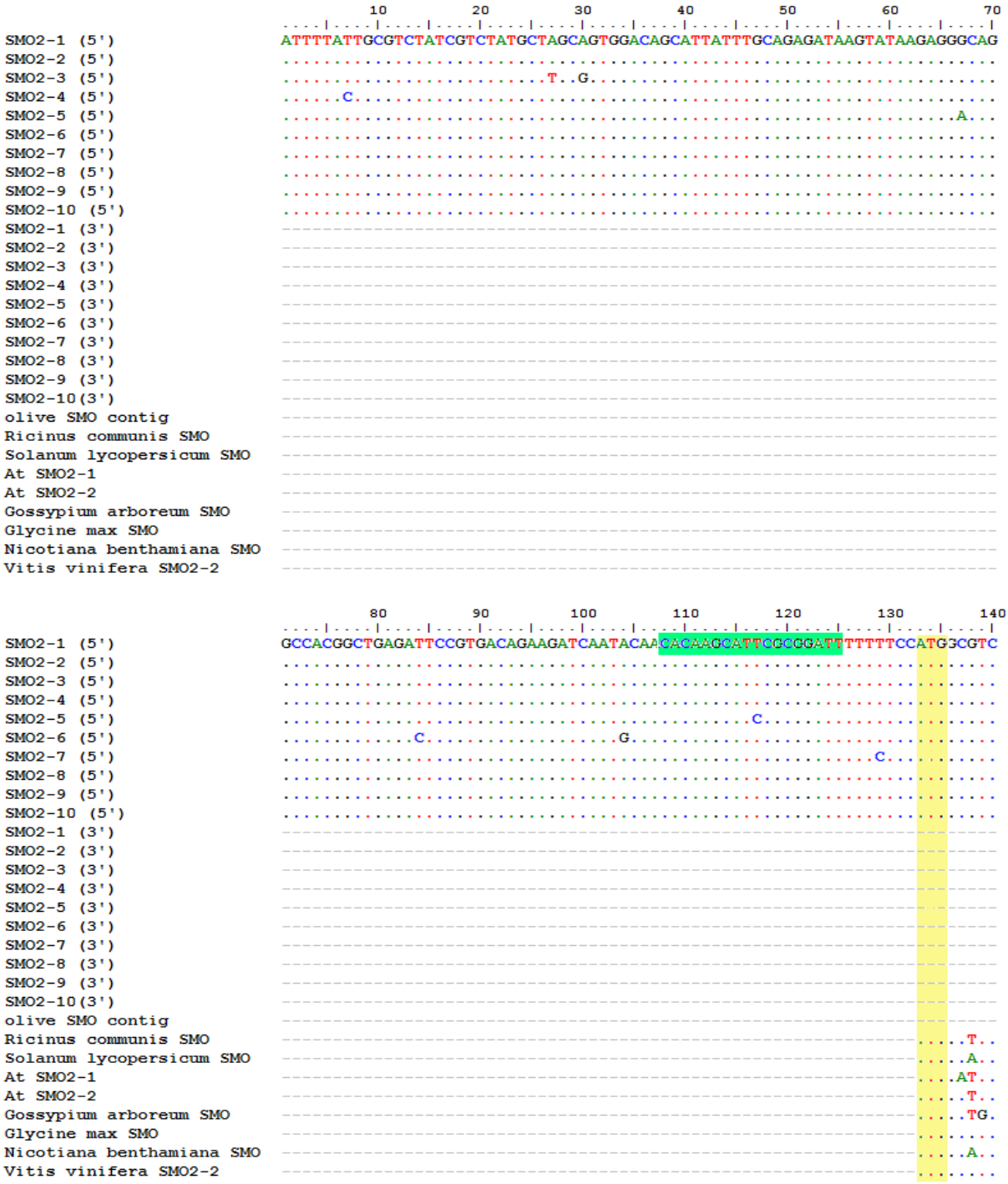
Figure IIIB Alignment of full length *SMT2* sequences deduced from olive cultivars Barnea, Frantoio and Picual

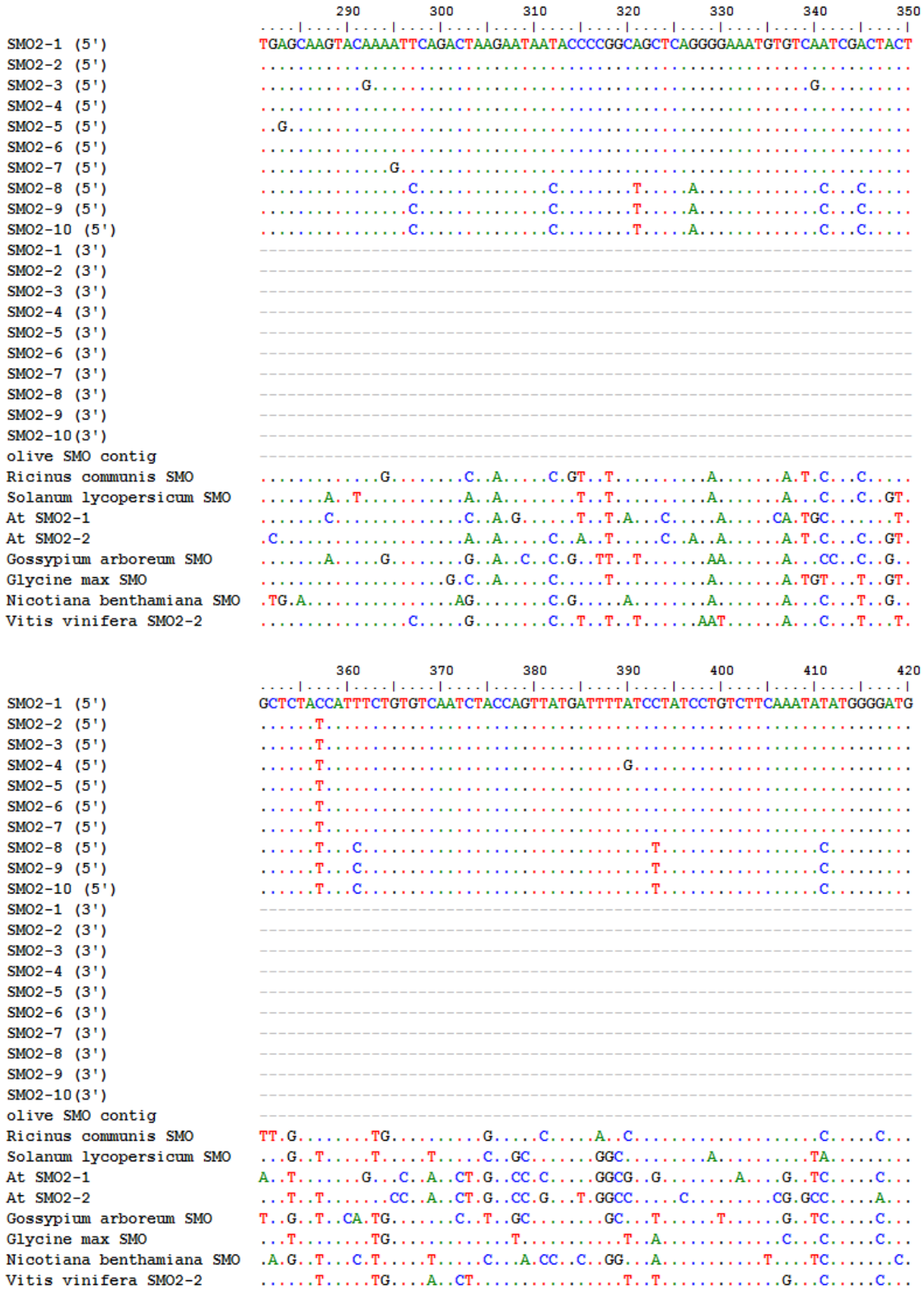
Cultivars: Bar: Barnea; Fran: Fran; Pic: Picual; The numbering system relates to the nucleotide number in the sequences. The START (ATG) and STOP (TAG/TAA) codon are highlighted in green. The primer pair for the isolation of full length *SMT2* genes are highlighted in red. F1-F10 represents the clones isolated from olive cultivars Barnea, Frantoio and Picual.

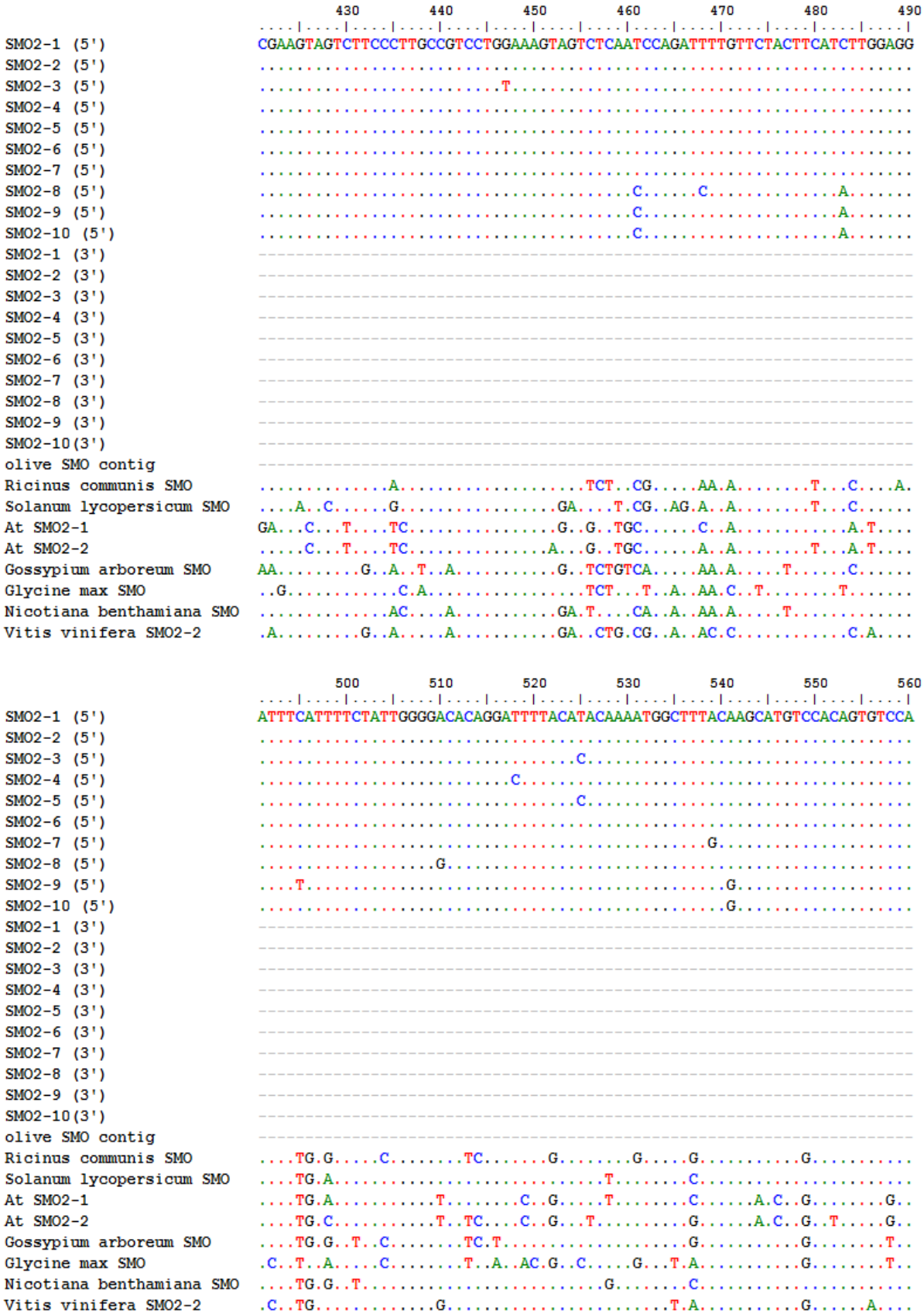
Seq->	1(5')	2(5')	3(5')	4(5')	5(5')	6(5')	7(5')	8(5')	9(5')	1(3')	2(3')	3(3')	4(3')	5(3')	6(3')	7(3')	8(3')
SMT2 F1 (5')	ID	0.998	0.998	1.000	0.998	0.998	0.990	0.992	1.000	0.998	0.994	0.994	0.998	0.998	0.998	0.998	0.998
SMT2 F2 (5')		ID	1.000	0.998	0.996	0.996	0.992	0.994	0.998	0.996	0.992	0.992	0.996	0.996	0.996	0.996	0.996
SMT2 F3 (5')			ID	0.998	0.996	0.996	0.992	0.994	0.998	0.996	0.992	0.992	0.996	0.996	0.996	0.996	0.996
SMT2 F4 (5')				ID	0.998	0.998	0.990	0.992	1.000	0.998	0.994	0.994	0.998	0.998	0.998	0.998	0.998
SMT2 F5 (5')					ID	1.000	0.992	0.990	0.998	0.996	0.992	0.992	0.996	0.996	0.996	0.996	0.996
SMT2 F6 (5')						ID	0.992	0.990	0.998	0.996	0.992	0.992	0.996	0.996	0.996	0.996	0.996
SMT2 F7 (5')							ID	0.986	0.990	0.988	0.984	0.984	0.988	0.988	0.988	0.988	0.988
SMT2 F8 (5')								ID	0.992	0.990	0.986	0.986	0.990	0.990	0.990	0.990	0.990
SMT2 F9 (5')									ID	0.998	0.994	0.994	0.998	0.998	0.998	0.998	0.998
SMT2 F1 (3')										ID	0.992	0.992	0.996	0.996	0.996	0.996	0.996
SMT2 F2 (3')											ID	1.000	0.992	0.992	0.992	0.992	0.992
SMT2 F3 (3')												ID	0.992	0.992	0.992	0.992	0.992
SMT2 F4 (3')													ID	1.000	0.996	0.996	0.996
SMT2 F5 (3')														ID	0.996	0.996	0.996
SMT2 F6 (3')															ID	1.000	1.000
SMT2 F7 (3')																ID	1.000
SMT2 F8 (3')																	ID

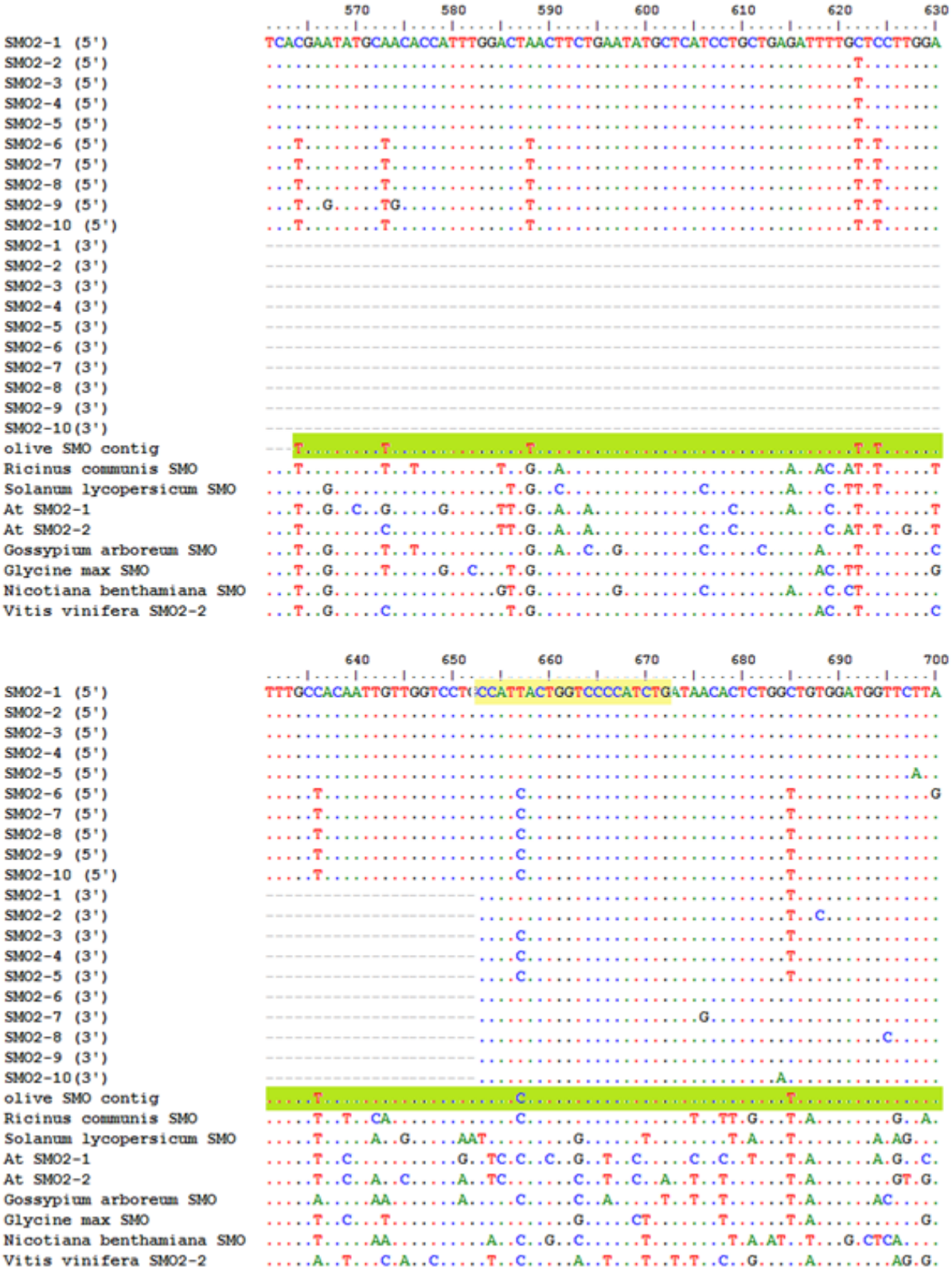
Figure IIICNucleotide sequence similarity of the overlapping sections of *SMT2* 5'and3' RACE clones deduced from olive cultivarBarnea. [1- SMT2-1, 2- SMT2-2, 3-SMT2-3, 4-SMT2-4, 5-SMT2-5, 6-SMT2-6, 7-SMT2-7, 8-SMT2-8, 9-SMT2-9. ID-100% identical, seq-sequences].

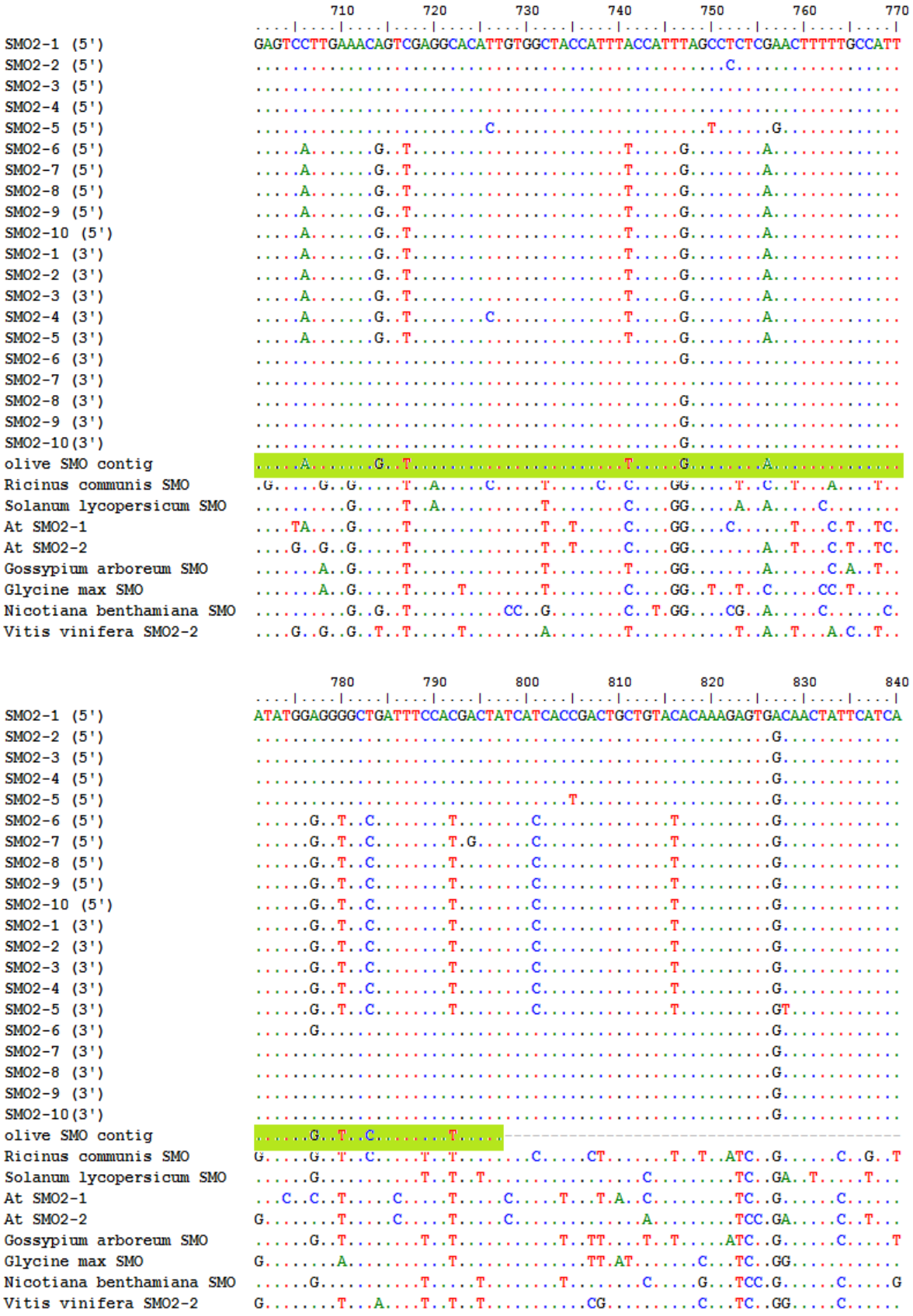
APPENDIX IV
THE DNA SEQUENCES OF *SMO2* GENES FROM BARNEA, FRANTOIO AND PICUAL

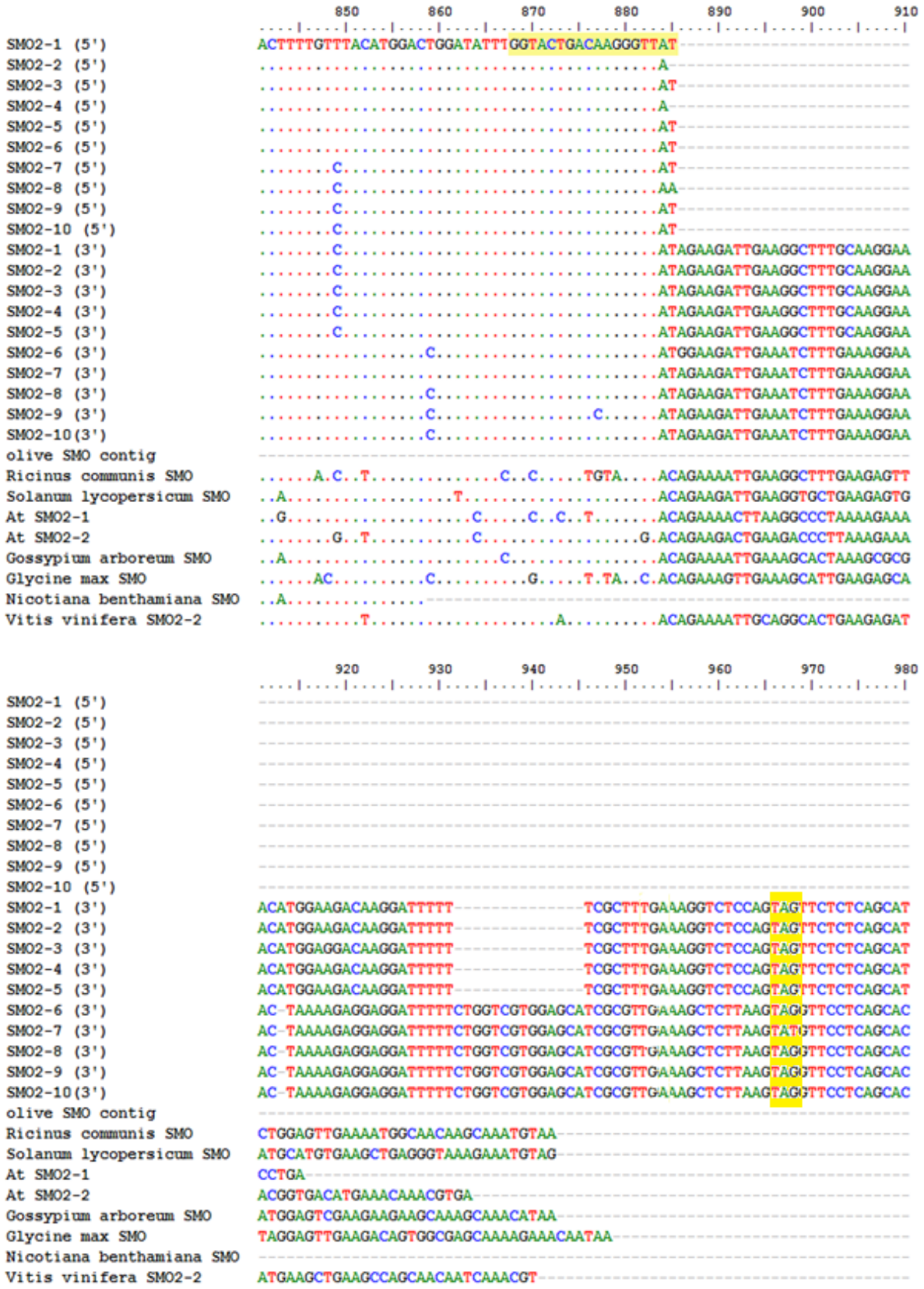


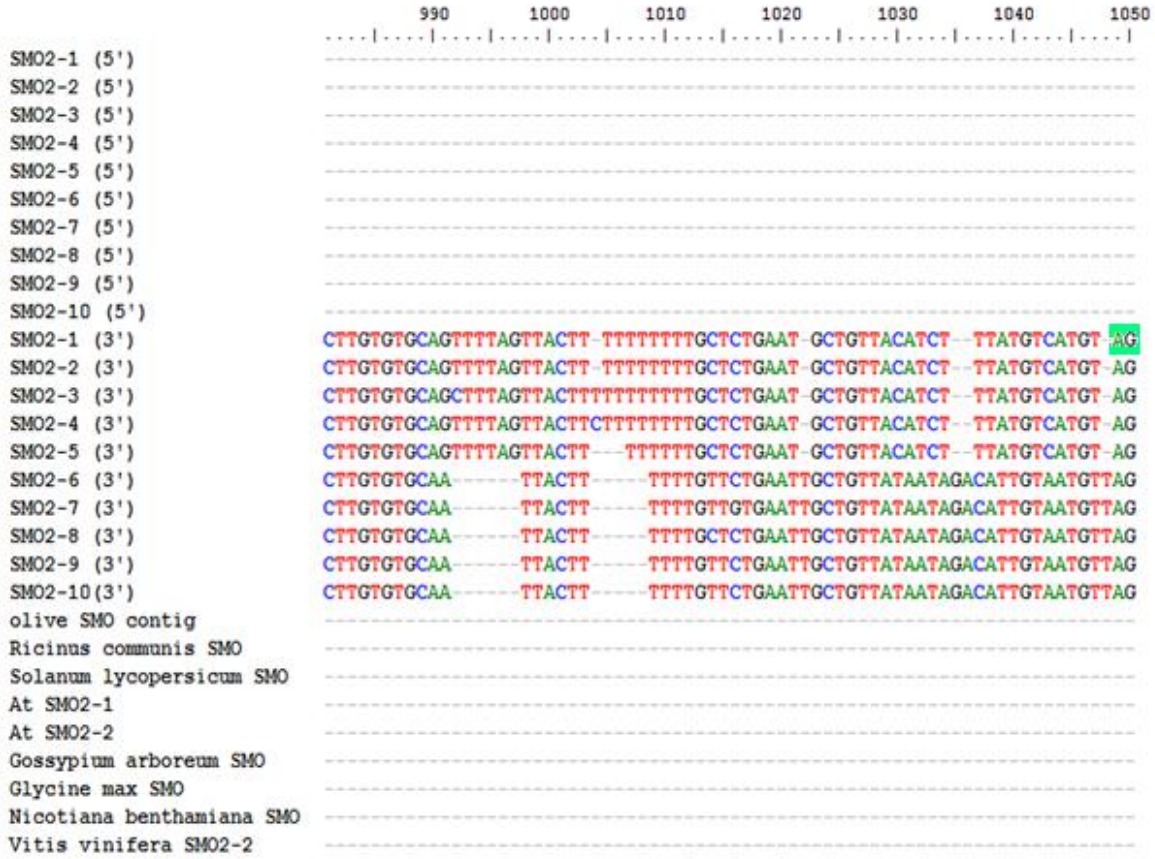

290 300 310 320 330 340 350
SMO2-1 (5') TGAGCAAGTACAAAATTCAGACTAAGAATAATACCCCGGCAGCTCAGGGGAAATGTGTCAATCGACTACT
SMO2-2 (5')
SMO2-3 (5')
SMO2-4 (5')
SMO2-5 (5')
SMO2-6 (5')
SMO2-7 (5')
SMO2-8 (5')
SMO2-9 (5')
SMO2-10 (5')
SMO2-1 (3')
SMO2-2 (3')
SMO2-3 (3')
SMO2-4 (3')
SMO2-5 (3')
SMO2-6 (3')
SMO2-7 (3')
SMO2-8 (3')
SMO2-9 (3')
SMO2-10(3')
olive SMO contig
Ricinus communis SMO
Solanum lycopersicum SMO
At SMO2-1
At SMO2-2
Gossypium arboreum SMO
Glycine max SMO
Nicotiana benthamiana SMO
Vitis vinifera SMO2-2
360 370 380 390 400 410 420
SMO2-1 (5') GCTCTACCATTTCTGTGTCAATCTACCAGTTATGATTTTATCCTATCCTGTCTTCAAATATATGGGGATG
SMO2-2 (5')
SMO2-3 (5')
SMO2-4 (5')
SMO2-5 (5')
SMO2-6 (5')
SMO2-7 (5')
SMO2-8 (5')
SMO2-9 (5')
SMO2-10 (5')
SMO2-1 (3')
SMO2-2 (3')
SMO2-3 (3')
SMO2-4 (3')
SMO2-5 (3')
SMO2-6 (3')
SMO2-7 (3')
SMO2-8 (3')
SMO2-9 (3')
SMO2-10(3')
olive SMO contig
Ricinus communis SMO
Solanum lycopersicum SMO
At SMO2-1
At SMO2-2
Gossypium arboreum SMO
Glycine max SMO
Nicotiana benthamiana SMO
Vitis vinifera SMO2-2

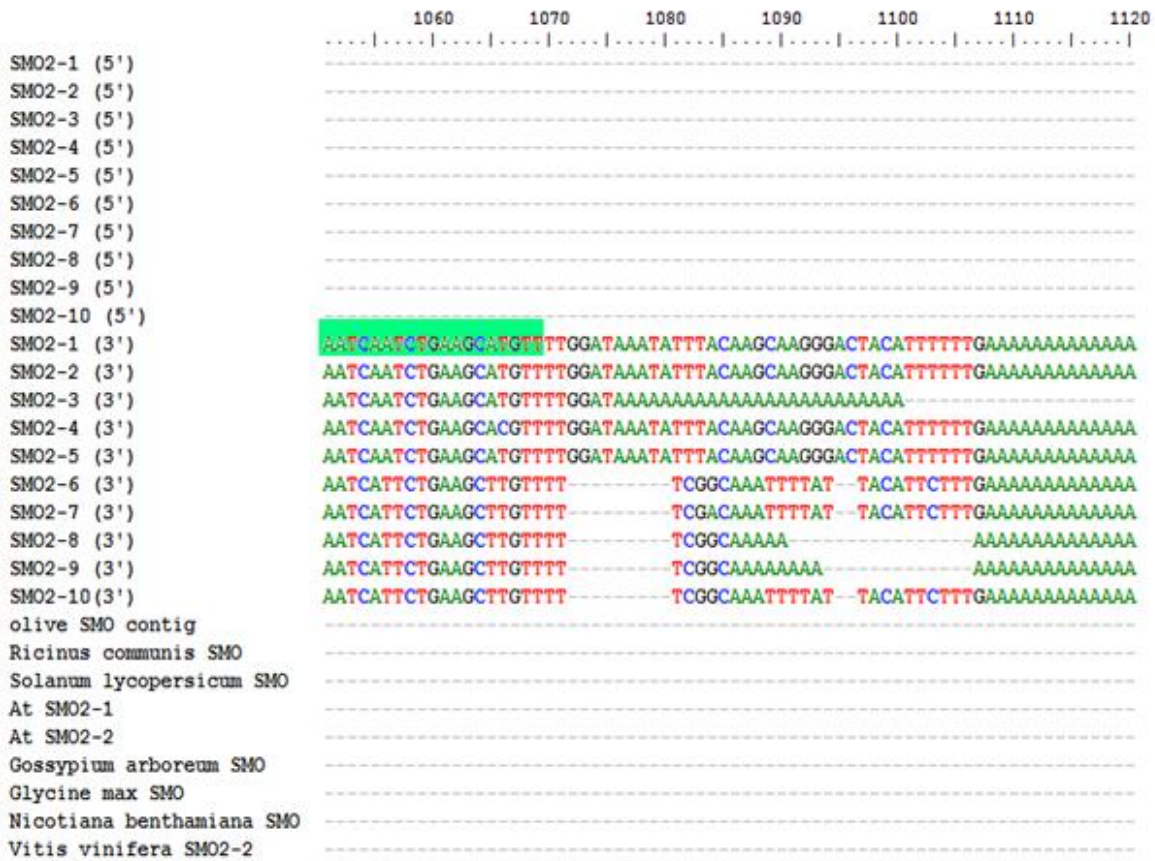

430 440 450 460 470 480 490
SMO2-1 (5') CGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCTCAATCCAGATTTTGTTCTACTTCATCTTGGAGG
SMO2-2 (5')
SMO2-3 (5')
SMO2-4 (5')
SMO2-5 (5')
SMO2-6 (5')
SMO2-7 (5')
SMO2-8 (5')
SMO2-9 (5')
SMO2-10 (5')
SMO2-1 (3')
SMO2-2 (3')
SMO2-3 (3')
SMO2-4 (3')
SMO2-5 (3')
SMO2-6 (3')
SMO2-7 (3')
SMO2-8 (3')
SMO2-9 (3')
SMO2-10(3')
olive SMO contig
Ricinus communis SMO
Solanum lycopersicum SMO
At SMO2-1
At SMO2-2
Gossypium arboreum SMO
Glycine max SMO
Nicotiana benthamiana SMO
Vitis vinifera SMO2-2
500 510 520 530 540 550 560
SMO2-1 (5') ATTTCATTTTCTATTGGGGACACAGGATTTTACATACAAAATGGCTTTACAAGCATGTCCACAGTGTCCA
SMO2-2 (5')
SMO2-3 (5')
SMO2-4 (5')
SMO2-5 (5')
SMO2-6 (5')
SMO2-7 (5')
SMO2-8 (5')
SMO2-9 (5')
SMO2-10 (5')
SMO2-1 (3')
SMO2-2 (3')
SMO2-3 (3')
SMO2-4 (3')
SMO2-5 (3')
SMO2-6 (3')
SMO2-7 (3')
SMO2-8 (3')
SMO2-9 (3')
SMO2-10(3')
olive SMO contig
Ricinus communis SMO
Solanum lycopersicum SMO
At SMO2-1
At SMO2-2
Gossypium arboreum SMO
Glycine max SMO
Nicotiana benthamiana SMO
Vitis vinifera SMO2-2


570 580 590 600 610 620 630
SMO2-1 (5') TCACGAATATGCAACACCATTTGGACTAACTTCTGAATATGCTCATCCTGCTGAGATTTTGCTCCTTGGA
SMO2-2 (5')
SMO2-3 (5')
SMO2-4 (5')
SMO2-5 (5')
SMO2-6 (5')
SMO2-7 (5')
SMO2-8 (5')
SMO2-9 (5')
SMO2-10 (5')
SMO2-1 (3')
SMO2-2 (3')
SMO2-3 (3')
SMO2-4 (3')
SMO2-5 (3')
SMO2-6 (3')
SMO2-7 (3')
SMO2-8 (3')
SMO2-9 (3')
SMO2-10(3')
olive SMO contig
Ricinus communis SMO
Solanum lycopersicum SMO
At SMO2-1
At SMO2-2
Gossypium arboreum SMO
Glycine max SMO
Nicotiana benthamiana SMO
Vitis vinifera SMO2-2
640 650 660 670 680 690 700
SMO2-1 (5') TTTGCCACAATTGTTGGTCCTNCCATTACTGGTCCCCATCTGATAACACTCTGGCTGTGGATGGTTCTTA
SMO2-2 (5')
SMO2-3 (5')
SMO2-4 (5')
SMO2-5 (5')
SMO2-6 (5')
SMO2-7 (5')
SMO2-8 (5')
SMO2-9 (5')
SMO2-10 (5')
SMO2-1 (3')
SMO2-2 (3')
SMO2-3 (3')
SMO2-4 (3')
SMO2-5 (3')
SMO2-6 (3')
SMO2-7 (3')
SMO2-8 (3')
SMO2-9 (3')
SMO2-10(3')
olive SMO contig
Ricinus communis SMO
Solanum lycopersicum SMO
At SMO2-1
At SMO2-2
Gossypium arboreum SMO
Glycine max SMO
Nicotiana benthamiana SMO
Vitis vinifera SMO2-2


710 720 730 740 750 760 770
SMO2-1 (5') GAGTCCTTGAAACAGTCGAGGCACATTGTGGCTACCATTTACCATTTAGCCTCTCGAACTTTTTGCCATT
SMO2-2 (5')
SMO2-3 (5')
SMO2-4 (5')
SMO2-5 (5')
SMO2-6 (5')
SMO2-7 (5')
SMO2-8 (5')
SMO2-9 (5')
SMO2-10 (5')
SMO2-1 (3')
SMO2-2 (3')
SMO2-3 (3')
SMO2-4 (3')
SMO2-5 (3')
SMO2-6 (3')
SMO2-7 (3')
SMO2-8 (3')
SMO2-9 (3')
SMO2-10(3')
olive SMO contig
Ricinus communis SMO
Solanum lycopersicum SMO
At SMO2-1
At SMO2-2
Gossypium arboreum SMO
Glycine max SMO
Nicotiana benthamiana SMO
Vitis vinifera SMO2-2
780 790 800 810 820 830 840
SMO2-1 (5') ATATGGAGGGGCTGATTTCCACGACTATCATCACCGACTGCTGTACACAAAGAGTGACAACTATTCATCA
SMO2-2 (5')
SMO2-3 (5')
SMO2-4 (5')
SMO2-5 (5')
SMO2-6 (5')
SMO2-7 (5')
SMO2-8 (5')
SMO2-9 (5')
SMO2-10 (5')
SMO2-1 (3')
SMO2-2 (3')
SMO2-3 (3')
SMO2-4 (3')
SMO2-5 (3')
SMO2-6 (3')
SMO2-7 (3')
SMO2-8 (3')
SMO2-9 (3')
SMO2-10(3')
olive SMO contig
Ricinus communis SMO
Solanum lycopersicum SMO
At SMO2-1
At SMO2-2
Gossypium arboreum SMO
Glycine max SMO
Nicotiana benthamiana SMO
Vitis vinifera SMO2-2


850 860 870 880 890 900 910
SMO2-1 (5')
SMO2-2 (5')
SMO2-3 (5')
SMO2-4 (5')
SMO2-5 (5')
SMO2-6 (5')
SMO2-7 (5')
SMO2-8 (5')
SMO2-9 (5')
SMO2-10 (5')
SMO2-1 (3')
SMO2-2 (3')
SMO2-3 (3')
SMO2-4 (3')
SMO2-5 (3')
SMO2-6 (3')
SMO2-7 (3')
SMO2-8 (3')
SMO2-9 (3')
SMO2-10(3')
olive SMO contig
Ricinus communis SMO
Solanum lycopersicum SMO
At SMO2-1
At SMO2-2
Gossypium arboreum SMO
Glycine max SMO
Nicotiana benthamiana SMO
Vitis vinifera SMO2-2
920 930 940 950 960 970 980
SMO2-1 (5')
SMO2-2 (5')
SMO2-3 (5')
SMO2-4 (5')
SMO2-5 (5')
SMO2-6 (5')
SMO2-7 (5')
SMO2-8 (5')
SMO2-9 (5')
SMO2-10 (5')
SMO2-1 (3')
SMO2-2 (3')
SMO2-3 (3')
SMO2-4 (3')
SMO2-5 (3')
SMO2-6 (3')
SMO2-7 (3')
SMO2-8 (3')
SMO2-9 (3')
SMO2-10(3')
olive SMO contig
Ricinus communis SMO
Solanum lycopersicum SMO
At SMO2-1
At SMO2-2
Gossypium arboreum SMO
Glycine max SMO
Nicotiana benthamiana SMO
Vitis vinifera SMO2-2


990 1000 1010 1020 1030 1040 1050
SMO2-1 (5')
SMO2-2 (5')
SMO2-3 (5')
SMO2-4 (5')
SMO2-5 (5')
SMO2-6 (5')
SMO2-7 (5')
SMO2-8 (5')
SMO2-9 (5')
SMO2-10 (5')
SMO2-1 (3') CTTGTGTGCAGTTTTAGTTACTT TTTTTTTTGCTCTGAAT GCTGTTACATCT TTATGTCATGT AG
SMO2-2 (3') CTTGTGTGCAGTTTTAGTTACTT TTTTTTTTGCTCTGAAT GCTGTTACATCT TTATGTCATGT AG
SMO2-3 (3') CTTGTGTGCAGCTTTAGTTACTTTTTTTTTTTTGCTCTGAAT GCTGTTACATCT TTATGTCATGT AG
SMO2-4 (3') CTTGTGTGCAGTTTTAGTTACTTCTTTTTTTTGCTCTGAAT GCTGTTACATCT TTATGTCATGT AG
SMO2-5 (3') CTTGTGTGCAGTTTTAGTTACTT TTTTTTGCTCTGAAT GCTGTTACATCT TTATGTCATGT AG
SMO2-6 (3') CTTGTGTGCAA TTACTT TTTTGTTCTGAATTGCTGTTATAATAGACATTGTAATGTTAG
SMO2-7 (3') CTTGTGTGCAA TTACTT TTTTGTTGTGAATTGCTGTTATAATAGACATTGTAATGTTAG
SMO2-8 (3') CTTGTGTGCAA TTACTT TTTTGCTCTGAATTGCTGTTATAATAGACATTGTAATGTTAG
SMO2-9 (3') CTTGTGTGCAA TTACTT TTTTGTTCTGAATTGCTGTTATAATAGACATTGTAATGTTAG
SMO2-10(3') CTTGTGTGCAA TTACTT TTTTGTTCTGAATTGCTGTTATAATAGACATTGTAATGTTAG
olive SMO contig
Ricinus communis SMO
Solanum lycopersicum SMO
At SMO2-1
At SMO2-2
Gossypium arboreum SMO
Glycine max SMO
Nicotiana benthamiana SMO
Vitis vinifera SMO2-2


1060 1070 1080 1090 1100 1110 1120
SMO2-1 (5')
SMO2-2 (5')
SMO2-3 (5')
SMO2-4 (5')
SMO2-5 (5')
SMO2-6 (5')
SMO2-7 (5')
SMO2-8 (5')
SMO2-9 (5')
SMO2-10 (5')
SMO2-1 (3') AATCAATCTGAAGCATGTTTTGGATAAATATTTACAAGCAAGGGACTACATTTTTTGAAAAAAAAAAAAA
SMO2-2 (3') AATCAATCTGAAGCATGTTTTGGATAAATATTTACAAGCAAGGGACTACATTTTTTGAAAAAAAAAAAAA
SMO2-3 (3') AATCAATCTGAAGCATGTTTTGGATAAAAAAAAAAAAAAAAAAAAAAAAA
SMO2-4 (3') AATCAATCTGAAGCACGTTTTGGATAAATATTTACAAGCAAGGGACTACATTTTTTGAAAAAAAAAAAAA
SMO2-5 (3') AATCAATCTGAAGCATGTTTTGGATAAATATTTACAAGCAAGGGACTACATTTTTTGAAAAAAAAAAAAA
SMO2-6 (3') AATCATTCTGAAGCTTGTTTT TCGGCAAATTTTAT TACATTCTTTGAAAAAAAAAAAAA
SMO2-7 (3') AATCATTCTGAAGCTTGTTTT TCGACAAATTTTAT TACATTCTTTGAAAAAAAAAAAAA
SMO2-8 (3') AATCATTCTGAAGCTTGTTTT TCGGCAAAAA AAAAAAAAAAAAA
SMO2-9 (3') AATCATTCTGAAGCTTGTTTT TCGGCAAAAAAAA AAAAAAAAAAAAA
SMO2-10(3') AATCATTCTGAAGCTTGTTTT TCGGCAAATTTTAT TACATTCTTTGAAAAAAAAAAAAA
olive SMO contig
Ricinus communis SMO
Solanum lycopersicum SMO
At SMO2-1
At SMO2-2
Gossypium arboreum SMO
Glycine max SMO
Nicotiana benthamiana SMO
Vitis vinifera SMO2-2

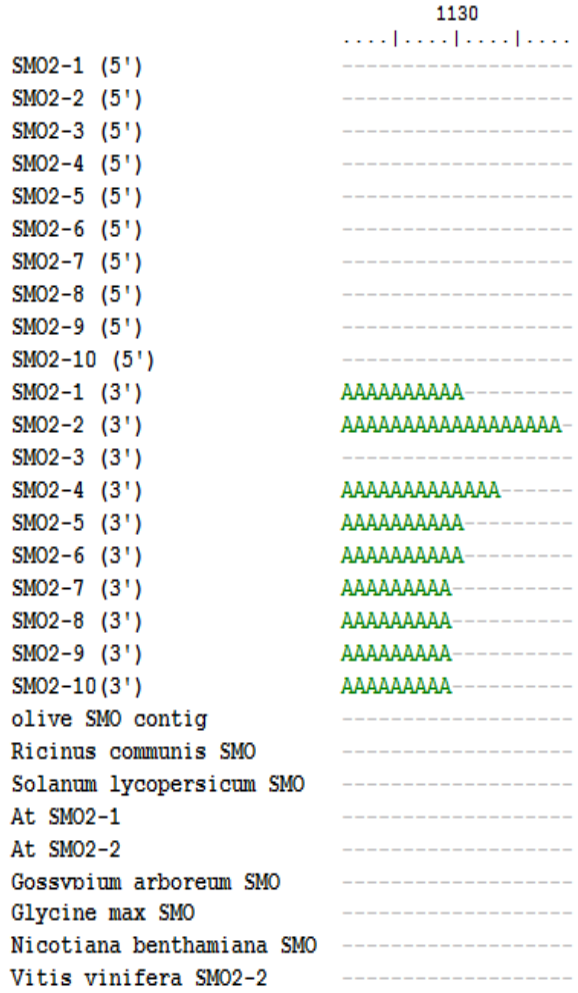
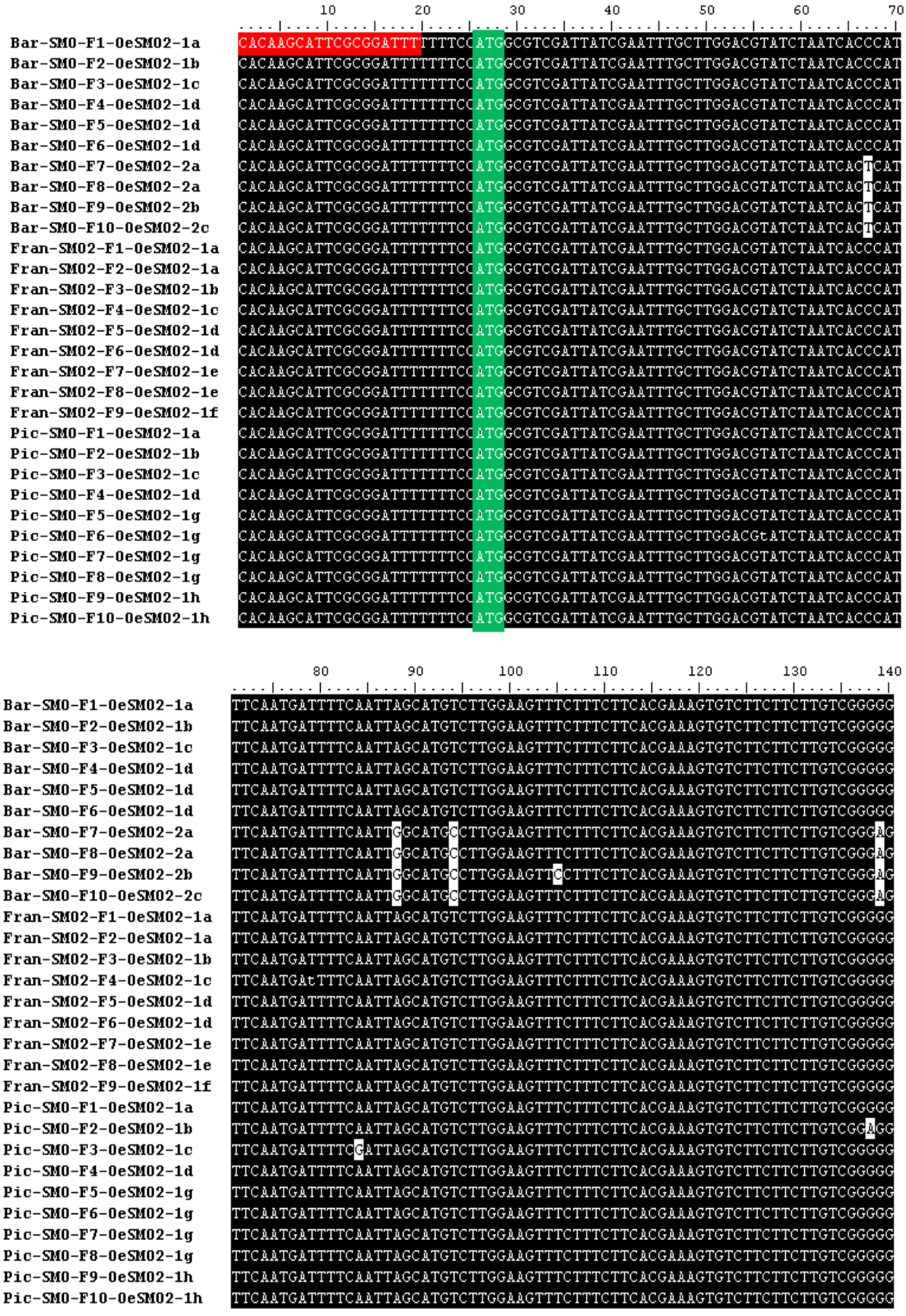
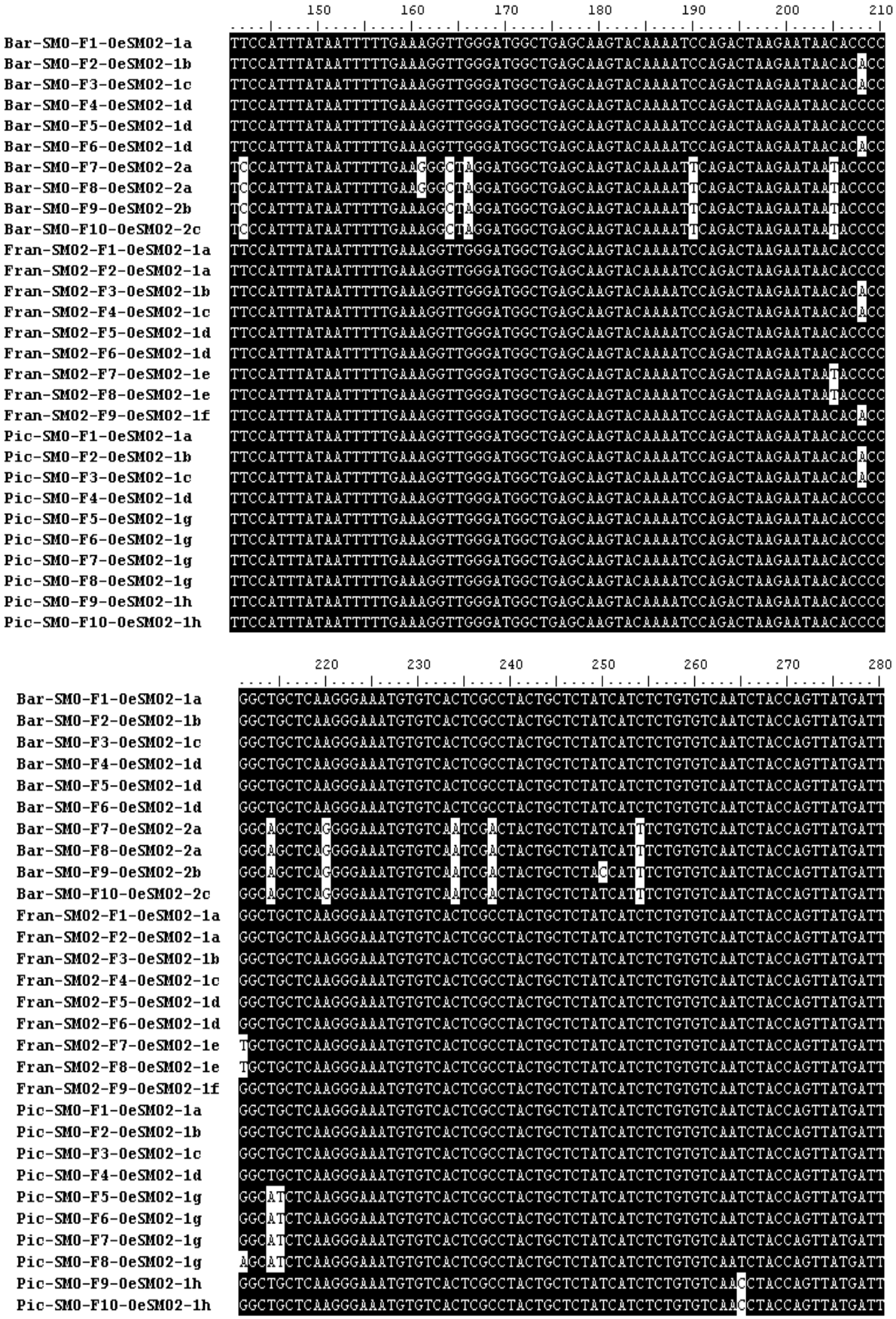
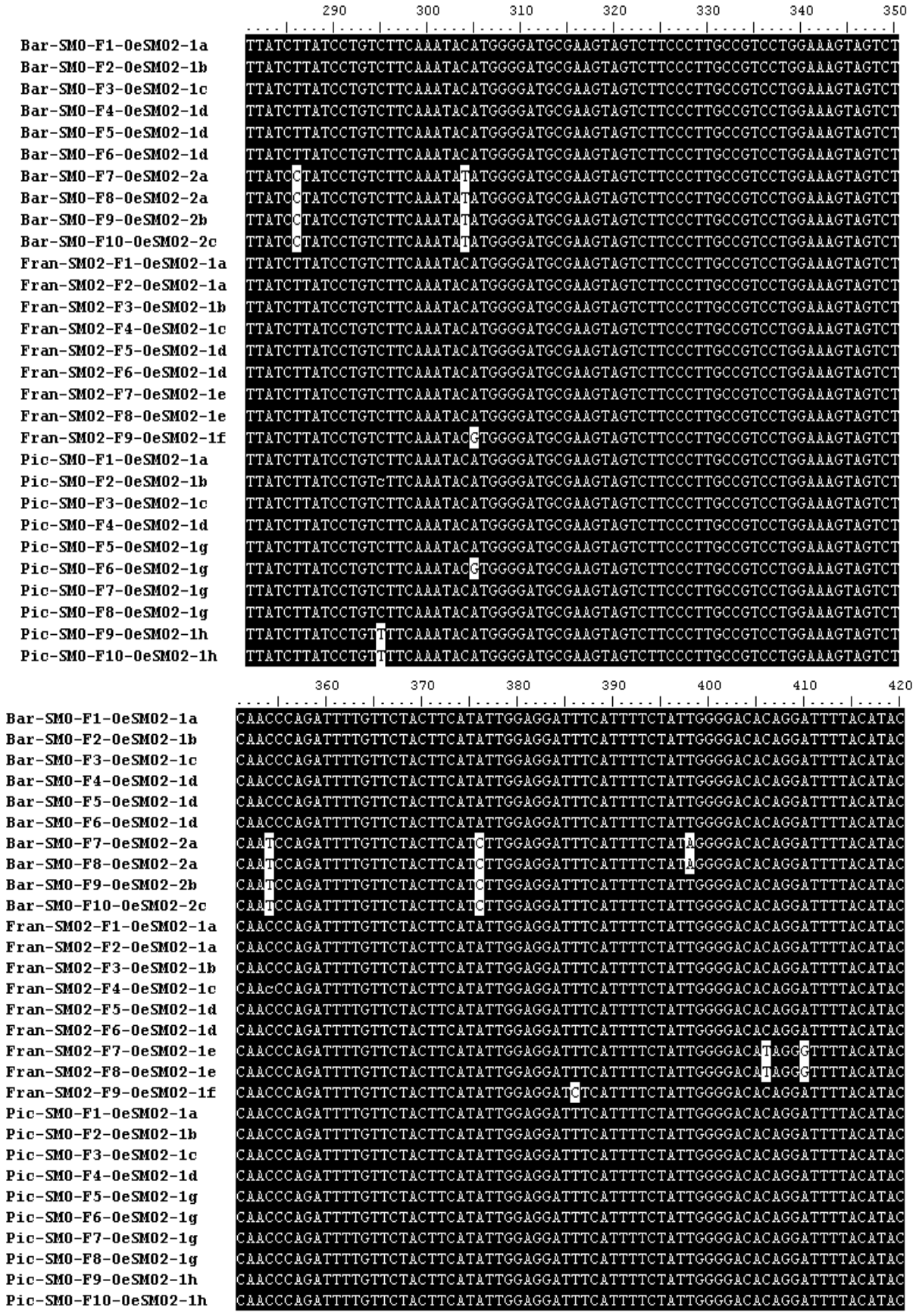


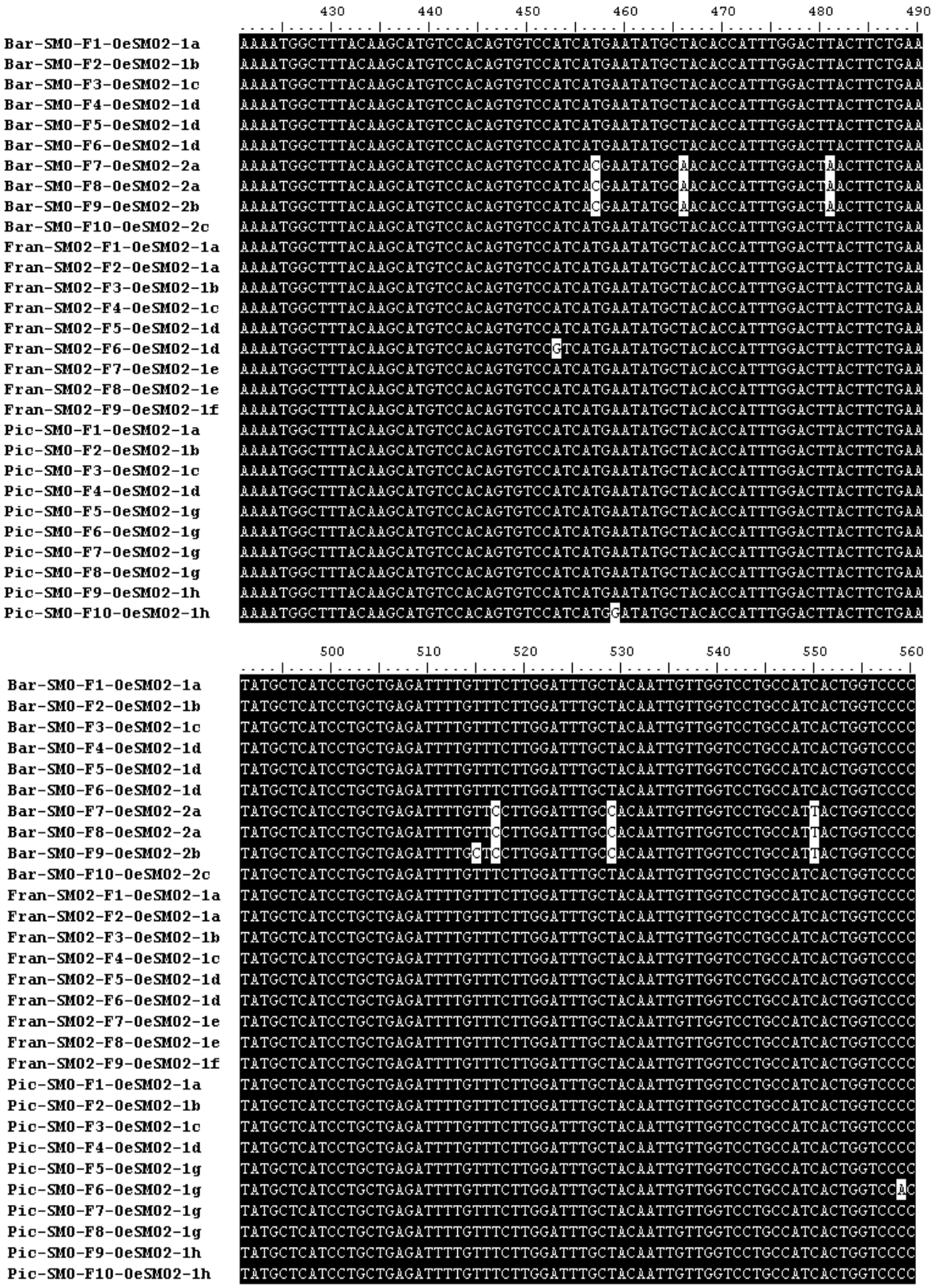
Figure IVAAlignment of *SMO2* 5'and3' RACE ten clones deduced from olive cultivarsBarnea and previously published *SMO* coding sequences from other plants.

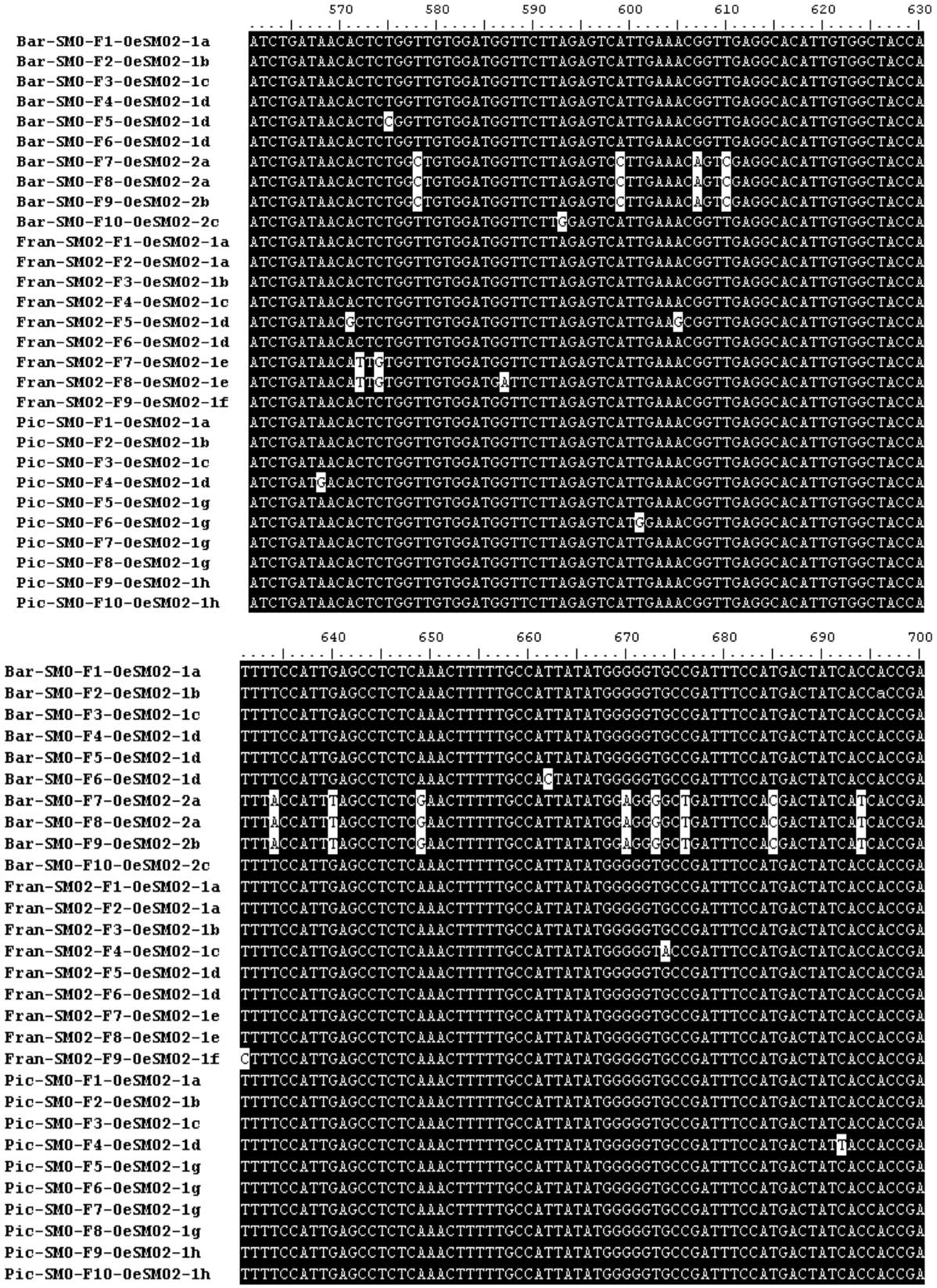
5'RACE cDNA -10 clones (SMO2-1 to SMO2-10); 3'RACE cDNA -10 clones (SMO21-SMO2-10). The olive *SMO*contig (Cluster Id-OLEEUCl011741) derived from the Alagna *et al* data (2009) which was used to isolate full length *SMO2* from Barnea is highlighted in light green. Primers used for RACE PCR are highlighted in yellow and primers used for full length *SMO2* amplification are highlighted in deep green. The START codon (ATG) and STOP codon (TAG) are highlighted in yellow. *Riccinus communis SMO* (Accession: XM_002520459.1),*Solanum lycopersicumSMO* (Accession:NM_001246951.1), *Gossipum arboreum SMO* (Accession: AF352575.1),*Vitis vinifera* partial *SMO2-2* (Accession: XM_002282617.1), *Glycine maxSMO* (Accession: NM_001253115.1), *Nicotiana benthamiana* partial *SMO* cds (Accession:AY321104.1, At SMO2-1: *Arabidopsis thaliana SMO2-1*(Accession: AF327853), At SMO2-2: *Arabidopsis thaliana SMO2-2*(Accession: AF346734)

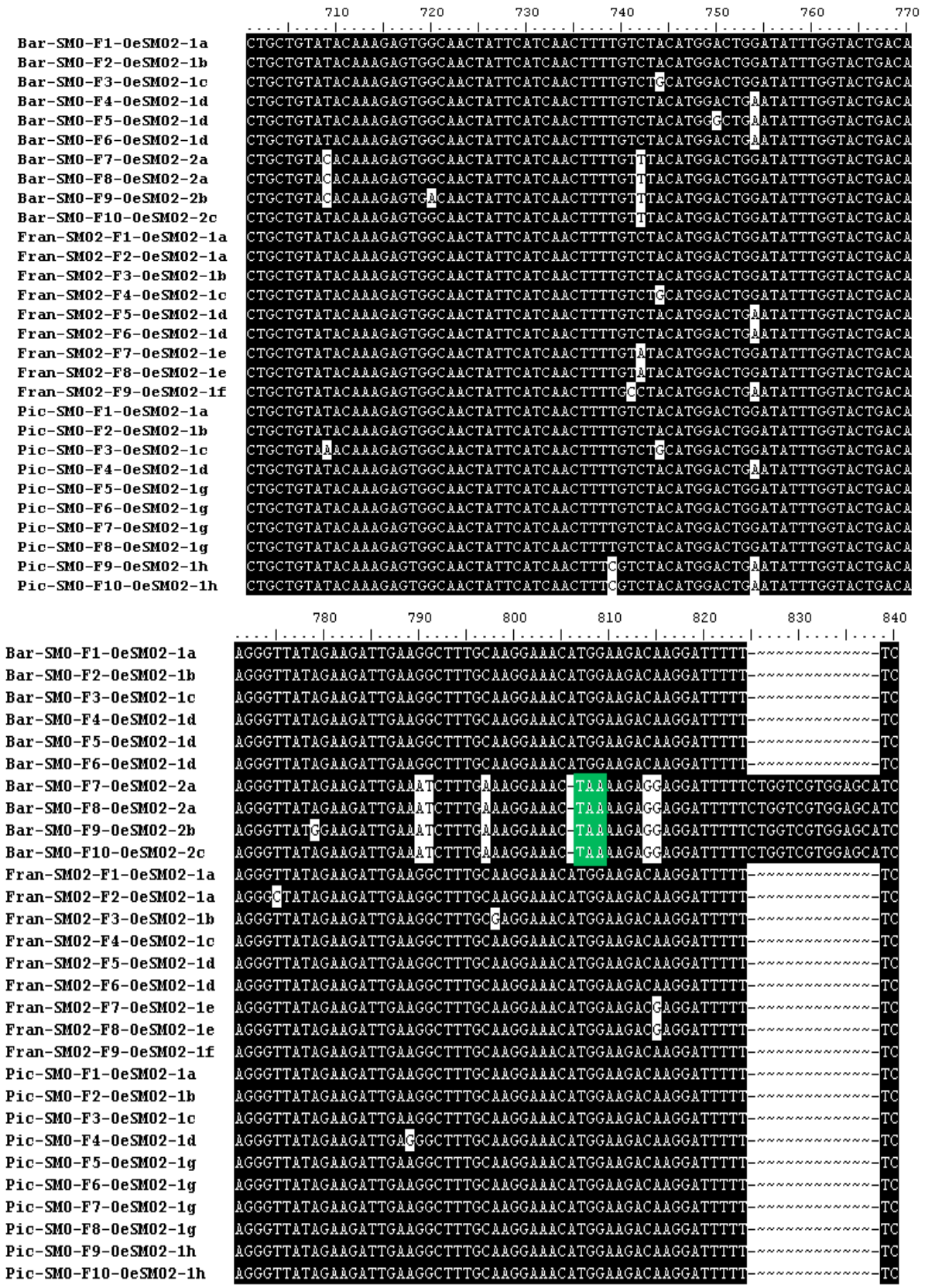

10 20 30 40 50 60 70
Bar-SM0-F1-0eSM02-1a CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Bar-SM0-F2-0eSM02-1b CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Bar-SM0-F3-0eSM02-1c CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Bar-SM0-F4-0eSM02-1d CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Bar-SM0-F5-0eSM02-1d CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Bar-SM0-F6-0eSM02-1d CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Bar-SM0-F7-0eSM02-2a CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACTCAT
Bar-SM0-F8-0eSM02-2a CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACTCAT
Bar-SM0-F9-0eSM02-2b CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACTCAT
Bar-SM0-F10-0eSM02-2c CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACTCAT
Fran-SM02-F1-0eSM02-1a CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Fran-SM02-F2-0eSM02-1a CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Fran-SM02-F3-0eSM02-1b CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Fran-SM02-F4-0eSM02-1c CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Fran-SM02-F5-0eSM02-1d CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Fran-SM02-F6-0eSM02-1d CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Fran-SM02-F7-0eSM02-1e CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Fran-SM02-F8-0eSM02-1e CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Fran-SM02-F9-0eSM02-1f CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Pic-SM0-F1-0eSM02-1a CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Pic-SM0-F2-0eSM02-1b CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Pic-SM0-F3-0eSM02-1c CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Pic-SM0-F4-0eSM02-1d CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Pic-SM0-F5-0eSM02-1g CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Pic-SM0-F6-0eSM02-1g CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGtATCTAATCACCCAT
Pic-SM0-F7-0eSM02-1g CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Pic-SM0-F8-0eSM02-1g CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Pic-SM0-F9-0eSM02-1h CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
Pic-SM0-F10-0eSM02-1h CACAAGCATTCGCGGATTTTTTTTCCATGGCGTCGATTATCGAATTTGCTTGGACGTATCTAATCACCCAT
80 90 100 110 120 130 140
Bar-SM0-F1-0eSM02-1a TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Bar-SM0-F2-0eSM02-1b TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Bar-SM0-F3-0eSM02-1c TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Bar-SM0-F4-0eSM02-1d TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Bar-SM0-F5-0eSM02-1d TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Bar-SM0-F6-0eSM02-1d TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Bar-SM0-F7-0eSM02-2a TTCAATGATTTTCAATTGGCATGCCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGAG
Bar-SM0-F8-0eSM02-2a TTCAATGATTTTCAATTGGCATGCCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGAG
Bar-SM0-F9-0eSM02-2b TTCAATGATTTTCAATTGGCATGCCTTGGAAGTTCCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGAG
Bar-SM0-F10-0eSM02-2c TTCAATGATTTTCAATTGGCATGCCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGAG
Fran-SM02-F1-0eSM02-1a TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Fran-SM02-F2-0eSM02-1a TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Fran-SM02-F3-0eSM02-1b TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Fran-SM02-F4-0eSM02-1c TTCAATGAtTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Fran-SM02-F5-0eSM02-1d TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Fran-SM02-F6-0eSM02-1d TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Fran-SM02-F7-0eSM02-1e TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Fran-SM02-F8-0eSM02-1e TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Fran-SM02-F9-0eSM02-1f TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Pic-SM0-F1-0eSM02-1a TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Pic-SM0-F2-0eSM02-1b TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGAGG
Pic-SM0-F3-0eSM02-1c TTCAATGATTTTCGATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Pic-SM0-F4-0eSM02-1d TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Pic-SM0-F5-0eSM02-1g TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Pic-SM0-F6-0eSM02-1g TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Pic-SM0-F7-0eSM02-1g TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Pic-SM0-F8-0eSM02-1g TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Pic-SM0-F9-0eSM02-1h TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG
Pic-SM0-F10-0eSM02-1h TTCAATGATTTTCAATTAGCATGTCTTGGAAGTTTCTTTCTTCACGAAAGTGTCTTCTTCTTGTCGGGGG


150 160 170 180 190 200 210
Bar-SM0-F1-0eSM02-1a
Bar-SM0-F2-0eSM02-1b
Bar-SM0-F3-0eSM02-1c
Bar-SM0-F4-0eSM02-1d
Bar-SM0-F5-0eSM02-1d
Bar-SM0-F6-0eSM02-1d
Bar-SM0-F7-0eSM02-2a
Bar-SM0-F8-0eSM02-2a
Bar-SM0-F9-0eSM02-2b
Bar-SM0-F10-0eSM02-2c
Fran-SM02-F1-0eSM02-1a
Fran-SM02-F2-0eSM02-1a
Fran-SM02-F3-0eSM02-1b
Fran-SM02-F4-0eSM02-1c
Fran-SM02-F5-0eSM02-1d
Fran-SM02-F6-0eSM02-1d
Fran-SM02-F7-0eSM02-1e
Fran-SM02-F8-0eSM02-1e
Fran-SM02-F9-0eSM02-1f
Pic-SM0-F1-0eSM02-1a
Pic-SM0-F2-0eSM02-1b
Pic-SM0-F3-0eSM02-1c
Pic-SM0-F4-0eSM02-1d
Pic-SM0-F5-0eSM02-1g
Pic-SM0-F6-0eSM02-1g
Pic-SM0-F7-0eSM02-1g
Pic-SM0-F8-0eSM02-1g
Pic-SM0-F9-0eSM02-1h
Pic-SM0-F10-0eSM02-1h
220 230 240 250 260 270 280
Bar-SM0-F1-0eSM02-1a
Bar-SM0-F2-0eSM02-1b
Bar-SM0-F3-0eSM02-1c
Bar-SM0-F4-0eSM02-1d
Bar-SM0-F5-0eSM02-1d
Bar-SM0-F6-0eSM02-1d
Bar-SM0-F7-0eSM02-2a
Bar-SM0-F8-0eSM02-2a
Bar-SM0-F9-0eSM02-2b
Bar-SM0-F10-0eSM02-2c
Fran-SM02-F1-0eSM02-1a
Fran-SM02-F2-0eSM02-1a
Fran-SM02-F3-0eSM02-1b
Fran-SM02-F4-0eSM02-1c
Fran-SM02-F5-0eSM02-1d
Fran-SM02-F6-0eSM02-1d
Fran-SM02-F7-0eSM02-1e
Fran-SM02-F8-0eSM02-1e
Fran-SM02-F9-0eSM02-1f
Pic-SM0-F1-0eSM02-1a
Pic-SM0-F2-0eSM02-1b
Pic-SM0-F3-0eSM02-1c
Pic-SM0-F4-0eSM02-1d
Pic-SM0-F5-0eSM02-1g
Pic-SM0-F6-0eSM02-1g
Pic-SM0-F7-0eSM02-1g
Pic-SM0-F8-0eSM02-1g
Pic-SM0-F9-0eSM02-1h
Pic-SM0-F10-0eSM02-1h


290 300 310 320 330 340 350
Bar-SM0-F1-0eSM02-1a TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Bar-SM0-F2-0eSM02-1b TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Bar-SM0-F3-0eSM02-1c TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Bar-SM0-F4-0eSM02-1d TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Bar-SM0-F5-0eSM02-1d TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Bar-SM0-F6-0eSM02-1d TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Bar-SM0-F7-0eSM02-2a TTATCCTATCCTGTCTTCAAATATATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Bar-SM0-F8-0eSM02-2a TTATCCTATCCTGTCTTCAAATATATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Bar-SM0-F9-0eSM02-2b TTATCCTATCCTGTCTTCAAATATATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Bar-SM0-F10-0eSM02-2c TTATCCTATCCTGTCTTCAAATATATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Fran-SM02-F1-0eSM02-1a TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Fran-SM02-F2-0eSM02-1a TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Fran-SM02-F3-0eSM02-1b TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Fran-SM02-F4-0eSM02-1c TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Fran-SM02-F5-0eSM02-1d TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Fran-SM02-F6-0eSM02-1d TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Fran-SM02-F7-0eSM02-1e TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Fran-SM02-F8-0eSM02-1e TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Fran-SM02-F9-0eSM02-1f TTATCTTATCCTGTCTTCAAATACGTGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Pic-SM0-F1-0eSM02-1a TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Pic-SM0-F2-0eSM02-1b TTATCTTATCCTGTcTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Pic-SM0-F3-0eSM02-1c TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Pic-SM0-F4-0eSM02-1d TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Pic-SM0-F5-0eSM02-1g TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Pic-SM0-F6-0eSM02-1g TTATCTTATCCTGTCTTCAAATAGCTGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Pic-SM0-F7-0eSM02-1g TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Pic-SM0-F8-0eSM02-1g TTATCTTATCCTGTCTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Pic-SM0-F9-0eSM02-1h TTATCTTATCCTGTTTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
Pic-SM0-F10-0eSM02-1h TTATCTTATCCTGTTTTCAAATACATGGGGATGCGAAGTAGTCTTCCCTTGCCGTCCTGGAAAGTAGTCT
360 370 380 390 400 410 420
Bar-SM0-F1-0eSM02-1a CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Bar-SM0-F2-0eSM02-1b CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Bar-SM0-F3-0eSM02-1c CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Bar-SM0-F4-0eSM02-1d CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Bar-SM0-F5-0eSM02-1d CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Bar-SM0-F6-0eSM02-1d CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Bar-SM0-F7-0eSM02-2a CAATCCAGATTTTGTTCTACTTCATCTTGGAGGATTTCATTTTCTATAGGGGACACAGGATTTTACATAC
Bar-SM0-F8-0eSM02-2a CAATCCAGATTTTGTTCTACTTCATCTTGGAGGATTTCATTTTCTATAGGGGACACAGGATTTTACATAC
Bar-SM0-F9-0eSM02-2b CAATCCAGATTTTGTTCTACTTCATCTTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Bar-SM0-F10-0eSM02-2c CAATCCAGATTTTGTTCTACTTCATCTTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Fran-SM02-F1-0eSM02-1a CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Fran-SM02-F2-0eSM02-1a CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Fran-SM02-F3-0eSM02-1b CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Fran-SM02-F4-0eSM02-1c CAAcCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Fran-SM02-F5-0eSM02-1d CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Fran-SM02-F6-0eSM02-1d CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Fran-SM02-F7-0eSM02-1e CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACATAGGGTTTTACATAC
Fran-SM02-F8-0eSM02-1e CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACATAGGGTTTTACATAC
Fran-SM02-F9-0eSM02-1f CAACCCAGATTTTGTTCTACTTCATATTGGAGGATCTCATTTTCTATTGGGGACACAGGATTTTACATAC
Pic-SM0-F1-0eSM02-1a CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Pic-SM0-F2-0eSM02-1b CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Pic-SM0-F3-0eSM02-1c CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Pic-SM0-F4-0eSM02-1d CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Pic-SM0-F5-0eSM02-1g CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Pic-SM0-F6-0eSM02-1g CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Pic-SM0-F7-0eSM02-1g CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Pic-SM0-F8-0eSM02-1g CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Pic-SM0-F9-0eSM02-1h CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC
Pic-SM0-F10-0eSM02-1h CAACCCAGATTTTGTTCTACTTCATATTGGAGGATTTCATTTTCTATTGGGGACACAGGATTTTACATAC


430 440 450 460 470 480 490
Bar-SM0-F1-0eSM02-1a
Bar-SM0-F2-0eSM02-1b
Bar-SM0-F3-0eSM02-1c
Bar-SM0-F4-0eSM02-1d
Bar-SM0-F5-0eSM02-1d
Bar-SM0-F6-0eSM02-1d
Bar-SM0-F7-0eSM02-2a
Bar-SM0-F8-0eSM02-2a
Bar-SM0-F9-0eSM02-2b
Bar-SM0-F10-0eSM02-2c
Fran-SM02-F1-0eSM02-1a
Fran-SM02-F2-0eSM02-1a
Fran-SM02-F3-0eSM02-1b
Fran-SM02-F4-0eSM02-1c
Fran-SM02-F5-0eSM02-1d
Fran-SM02-F6-0eSM02-1d
Fran-SM02-F7-0eSM02-1e
Fran-SM02-F8-0eSM02-1e
Fran-SM02-F9-0eSM02-1f
Pic-SM0-F1-0eSM02-1a
Pic-SM0-F2-0eSM02-1b
Pic-SM0-F3-0eSM02-1c
Pic-SM0-F4-0eSM02-1d
Pic-SM0-F5-0eSM02-1g
Pic-SM0-F6-0eSM02-1g
Pic-SM0-F7-0eSM02-1g
Pic-SM0-F8-0eSM02-1g
Pic-SM0-F9-0eSM02-1h
Pic-SM0-F10-0eSM02-1h
500 510 520 530 540 550 560
Bar-SM0-F1-0eSM02-1a
Bar-SM0-F2-0eSM02-1b
Bar-SM0-F3-0eSM02-1c
Bar-SM0-F4-0eSM02-1d
Bar-SM0-F5-0eSM02-1d
Bar-SM0-F6-0eSM02-1d
Bar-SM0-F7-0eSM02-2a
Bar-SM0-F8-0eSM02-2a
Bar-SM0-F9-0eSM02-2b
Bar-SM0-F10-0eSM02-2c
Fran-SM02-F1-0eSM02-1a
Fran-SM02-F2-0eSM02-1a
Fran-SM02-F3-0eSM02-1b
Fran-SM02-F4-0eSM02-1c
Fran-SM02-F5-0eSM02-1d
Fran-SM02-F6-0eSM02-1d
Fran-SM02-F7-0eSM02-1e
Fran-SM02-F8-0eSM02-1e
Fran-SM02-F9-0eSM02-1f
Pic-SM0-F1-0eSM02-1a
Pic-SM0-F2-0eSM02-1b
Pic-SM0-F3-0eSM02-1c
Pic-SM0-F4-0eSM02-1d
Pic-SM0-F5-0eSM02-1g
Pic-SM0-F6-0eSM02-1g
Pic-SM0-F7-0eSM02-1g
Pic-SM0-F8-0eSM02-1g
Pic-SM0-F9-0eSM02-1h
Pic-SM0-F10-0eSM02-1h


570 580 590 600 610 620 630
Bar-SM0-F1-0eSM02-1a ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Bar-SM0-F2-0eSM02-1b ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Bar-SM0-F3-0eSM02-1c ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Bar-SM0-F4-0eSM02-1d ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Bar-SM0-F5-0eSM02-1d ATCTGATAACACTCCGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Bar-SM0-F6-0eSM02-1d ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Bar-SM0-F7-0eSM02-2a ATCTGATAACACTCTGGCTGTGGATGGTTCTTAGAGTCCTTGAAACAGTCGAGGCACATTGTGGCTACCA
Bar-SM0-F8-0eSM02-2a ATCTGATAACACTCTGGCTGTGGATGGTTCTTAGAGTCCTTGAAACAGTCGAGGCACATTGTGGCTACCA
Bar-SM0-F9-0eSM02-2b ATCTGATAACACTCTGGCTGTGGATGGTTCTTAGAGTCCTTGAAACAGTCGAGGCACATTGTGGCTACCA
Bar-SM0-F10-0eSM02-2c ATCTGATAACACTCTGGTTGTGGATGGTTCTTGGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Fran-SM02-F1-0eSM02-1a ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Fran-SM02-F2-0eSM02-1a ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Fran-SM02-F3-0eSM02-1b ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Fran-SM02-F4-0eSM02-1c ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Fran-SM02-F5-0eSM02-1d ATCTGATAAGGCTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAGCGGTTGAGGCACATTGTGGCTACCA
Fran-SM02-F6-0eSM02-1d ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Fran-SM02-F7-0eSM02-1e ATCTGATAACATTGTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Fran-SM02-F8-0eSM02-1e ATCTGATAACATTGTGGTTGTGGATGATTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Fran-SM02-F9-0eSM02-1f ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Pic-SM0-F1-0eSM02-1a ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Pic-SM0-F2-0eSM02-1b ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Pic-SM0-F3-0eSM02-1c ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Pic-SM0-F4-0eSM02-1d ATCTGATGACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Pic-SM0-F5-0eSM02-1g ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Pic-SM0-F6-0eSM02-1g ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATGGAAACGGTTGAGGCACATTGTGGCTACCA
Pic-SM0-F7-0eSM02-1g ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Pic-SM0-F8-0eSM02-1g ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Pic-SM0-F9-0eSM02-1h ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
Pic-SM0-F10-0eSM02-1h ATCTGATAACACTCTGGTTGTGGATGGTTCTTAGAGTCATTGAAACGGTTGAGGCACATTGTGGCTACCA
640 650 660 670 680 690 700
Bar-SM0-F1-0eSM02-1a TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Bar-SM0-F2-0eSM02-1b TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCaCCGA
Bar-SM0-F3-0eSM02-1c TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Bar-SM0-F4-0eSM02-1d TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Bar-SM0-F5-0eSM02-1d TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Bar-SM0-F6-0eSM02-1d TTTTCCATTGAGCCTCTCAAACTTTTTGCCACTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Bar-SM0-F7-0eSM02-2a TTTACCATTTAGCCTCTCGAACTTTTTGCCATTATATGGAGGGGCTGATTTCCACGACTATCATCACCGA
Bar-SM0-F8-0eSM02-2a TTTACCATTTAGCCTCTCGAACTTTTTGCCATTATATGGAGGGGCTGATTTCCACGACTATCATCACCGA
Bar-SM0-F9-0eSM02-2b TTTACCATTTAGCCTCTCGAACTTTTTGCCATTATATGGAGGGGCTGATTTCCACGACTATCATCACCGA
Bar-SM0-F10-0eSM02-2c TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Fran-SM02-F1-0eSM02-1a TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Fran-SM02-F2-0eSM02-1a TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Fran-SM02-F3-0eSM02-1b TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Fran-SM02-F4-0eSM02-1c TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTACCGATTTCCATGACTATCACCACCGA
Fran-SM02-F5-0eSM02-1d TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Fran-SM02-F6-0eSM02-1d TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Fran-SM02-F7-0eSM02-1e TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Fran-SM02-F8-0eSM02-1e TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Fran-SM02-F9-0eSM02-1f CTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Pic-SM0-F1-0eSM02-1a TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Pic-SM0-F2-0eSM02-1b TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Pic-SM0-F3-0eSM02-1c TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Pic-SM0-F4-0eSM02-1d TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATTACCACCGA
Pic-SM0-F5-0eSM02-1g TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Pic-SM0-F6-0eSM02-1g TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Pic-SM0-F7-0eSM02-1g TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Pic-SM0-F8-0eSM02-1g TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Pic-SM0-F9-0eSM02-1h TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA
Pic-SM0-F10-0eSM02-1h TTTTCCATTGAGCCTCTCAAACTTTTTGCCATTATATGGGGGTGCCGATTTCCATGACTATCACCACCGA


710 720 730 740 750 760 770
Bar-SM0-F1-0eSM02-1a CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGACTGGATATTTGGTACTGACA
Bar-SM0-F2-0eSM02-1b CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGACTGGATATTTGGTACTGACA
Bar-SM0-F3-0eSM02-1c CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTGCATGGACTGGATATTTGGTACTGACA
Bar-SM0-F4-0eSM02-1d CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGACTGAATATTTGGTACTGACA
Bar-SM0-F5-0eSM02-1d CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGGCTGAATATTTGGTACTGACA
Bar-SM0-F6-0eSM02-1d CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGACTGAATATTTGGTACTGACA
Bar-SM0-F7-0eSM02-2a CTGCTGTACACAAAGAGTGGCAACTATTCATCAACTTTTGTTTACATGGACTGGATATTTGGTACTGACA
Bar-SM0-F8-0eSM02-2a CTGCTGTACACAAAGAGTGGCAACTATTCATCAACTTTTGTTTACATGGACTGGATATTTGGTACTGACA
Bar-SM0-F9-0eSM02-2b CTGCTGTACACAAAGAGTGACAACTATTCATCAACTTTTGTTTACATGGACTGGATATTTGGTACTGACA
Bar-SM0-F10-0eSM02-2c CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTTTACATGGACTGGATATTTGGTACTGACA
Fran-SM02-F1-0eSM02-1a CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGACTGGATATTTGGTACTGACA
Fran-SM02-F2-0eSM02-1a CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGACTGGATATTTGGTACTGACA
Fran-SM02-F3-0eSM02-1b CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGACTGGATATTTGGTACTGACA
Fran-SM02-F4-0eSM02-1c CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTGCATGGACTGGATATTTGGTACTGACA
Fran-SM02-F5-0eSM02-1d CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGACTGAATATTTGGTACTGACA
Fran-SM02-F6-0eSM02-1d CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGACTGAATATTTGGTACTGACA
Fran-SM02-F7-0eSM02-1e CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTATACATGGACTGGATATTTGGTACTGACA
Fran-SM02-F8-0eSM02-1e CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTATACATGGACTGGATATTTGGTACTGACA
Fran-SM02-F9-0eSM02-1f CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGCCTACATGGACTGAATATTTGGTACTGACA
Pic-SM0-F1-0eSM02-1a CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGACTGGATATTTGGTACTGACA
Pic-SM0-F2-0eSM02-1b CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGACTGGATATTTGGTACTGACA
Pic-SM0-F3-0eSM02-1c CTGCTGTAAACAAAGAGTGGCAACTATTCATCAACTTTTGTCTGCATGGACTGGATATTTGGTACTGACA
Pic-SM0-F4-0eSM02-1d CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGACTGAATATTTGGTACTGACA
Pic-SM0-F5-0eSM02-1g CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGACTGGATATTTGGTACTGACA
Pic-SM0-F6-0eSM02-1g CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGACTGGATATTTGGTACTGACA
Pic-SM0-F7-0eSM02-1g CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGACTGGATATTTGGTACTGACA
Pic-SM0-F8-0eSM02-1g CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTTGTCTACATGGACTGGATATTTGGTACTGACA
Pic-SM0-F9-0eSM02-1h CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTCGTCTACATGGACTGAATATTTGGTACTGACA
Pic-SM0-F10-0eSM02-1h CTGCTGTATACAAAGAGTGGCAACTATTCATCAACTTTCGTCTACATGGACTGAATATTTGGTACTGACA
780 790 800 810 820 830 840
Bar-SM0-F1-0eSM02-1a AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Bar-SM0-F2-0eSM02-1b AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Bar-SM0-F3-0eSM02-1c AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Bar-SM0-F4-0eSM02-1d AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Bar-SM0-F5-0eSM02-1d AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Bar-SM0-F6-0eSM02-1d AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Bar-SM0-F7-0eSM02-2a AGGGTTATAGAAGATTGAAATCTTTGAAAGGAAAC-TAAAAGAGGAGGATTTTTCTGGTCGTGGAGCATC
Bar-SM0-F8-0eSM02-2a AGGGTTATAGAAGATTGAAATCTTTGAAAGGAAAC-TAAAAGAGGAGGATTTTTCTGGTCGTGGAGCATC
Bar-SM0-F9-0eSM02-2b AGGGTTATGGAAGATTGAAATCTTTGAAAGGAAAC-TAAAAGAGGAGGATTTTTCTGGTCGTGGAGCATC
Bar-SM0-F10-0eSM02-2c AGGGTTATAGAAGATTGAAATCTTTGAAAGGAAAC-TAAAAGAGGAGGATTTTTCTGGTCGTGGAGCATC
Fran-SM02-F1-0eSM02-1a AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Fran-SM02-F2-0eSM02-1a AGGGCTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Fran-SM02-F3-0eSM02-1b AGGGTTATAGAAGATTGAAGGCTTTGCGAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Fran-SM02-F4-0eSM02-1c AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Fran-SM02-F5-0eSM02-1d AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Fran-SM02-F6-0eSM02-1d AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Fran-SM02-F7-0eSM02-1e AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACGAGGATTTTT-~~~~~~~~~~~~~~-TC
Fran-SM02-F8-0eSM02-1e AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACGAGGATTTTT-~~~~~~~~~~~~~~-TC
Fran-SM02-F9-0eSM02-1f AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Pic-SM0-F1-0eSM02-1a AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Pic-SM0-F2-0eSM02-1b AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Pic-SM0-F3-0eSM02-1c AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Pic-SM0-F4-0eSM02-1d AGGGTTATAGAAGATTGAGGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Pic-SM0-F5-0eSM02-1g AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Pic-SM0-F6-0eSM02-1g AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Pic-SM0-F7-0eSM02-1g AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Pic-SM0-F8-0eSM02-1g AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Pic-SM0-F9-0eSM02-1h AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC
Pic-SM0-F10-0eSM02-1h AGGGTTATAGAAGATTGAAGGCTTTGCAAGGAAACATGGAAGACAAGGATTTTT-~~~~~~~~~~~~~~-TC

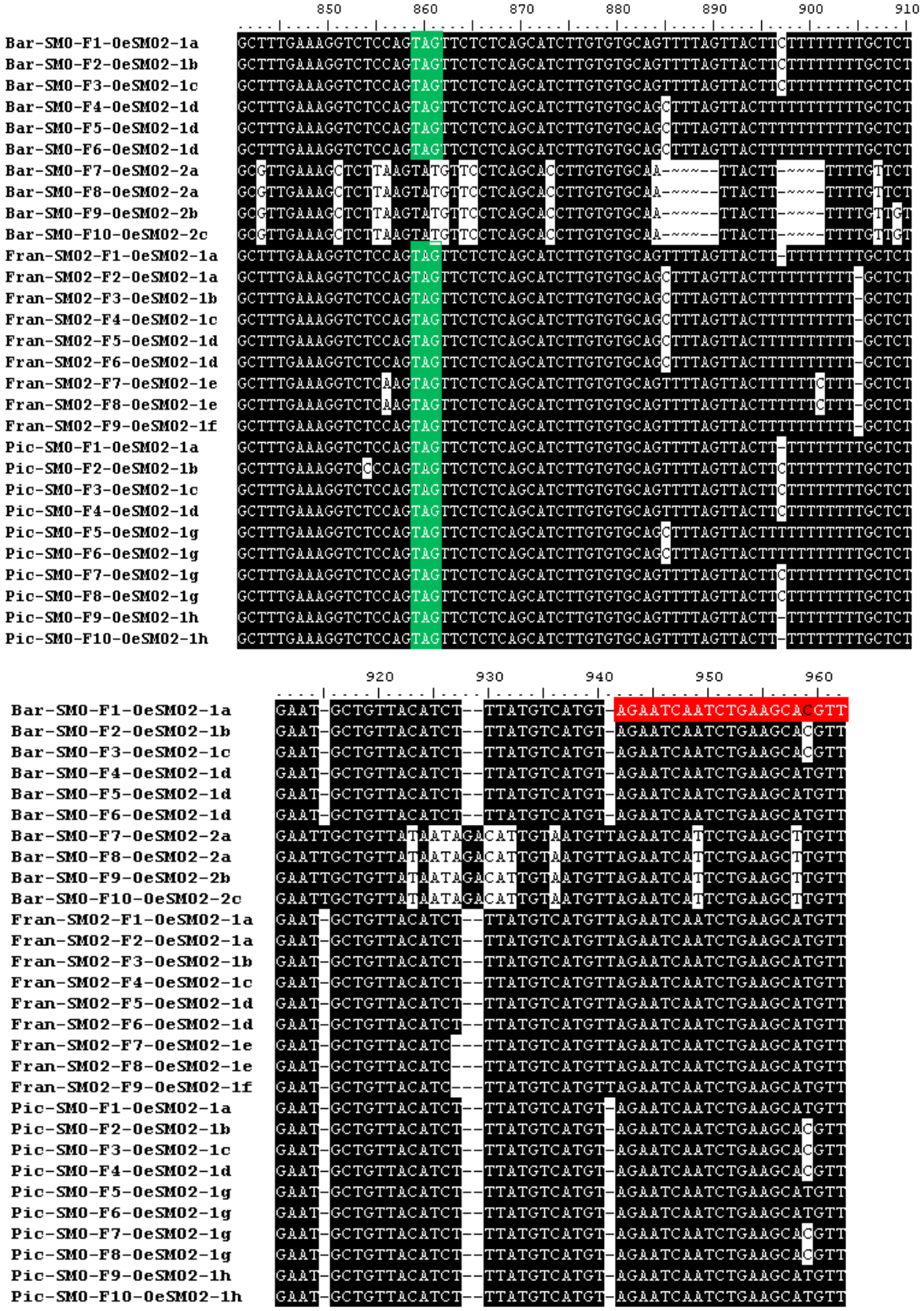

850 860 870 880 890 900 910
Bar-SM0-F1-0eSM02-1a
Bar-SM0-F2-0eSM02-1b
Bar-SM0-F3-0eSM02-1c
Bar-SM0-F4-0eSM02-1d
Bar-SM0-F5-0eSM02-1d
Bar-SM0-F6-0eSM02-1d
Bar-SM0-F7-0eSM02-2a
Bar-SM0-F8-0eSM02-2a
Bar-SM0-F9-0eSM02-2b
Bar-SM0-F10-0eSM02-2c
Fran-SM02-F1-0eSM02-1a
Fran-SM02-F2-0eSM02-1a
Fran-SM02-F3-0eSM02-1b
Fran-SM02-F4-0eSM02-1c
Fran-SM02-F5-0eSM02-1d
Fran-SM02-F6-0eSM02-1d
Fran-SM02-F7-0eSM02-1e
Fran-SM02-F8-0eSM02-1e
Fran-SM02-F9-0eSM02-1f
Pic-SM0-F1-0eSM02-1a
Pic-SM0-F2-0eSM02-1b
Pic-SM0-F3-0eSM02-1c
Pic-SM0-F4-0eSM02-1d
Pic-SM0-F5-0eSM02-1g
Pic-SM0-F6-0eSM02-1g
Pic-SM0-F7-0eSM02-1g
Pic-SM0-F8-0eSM02-1g
Pic-SM0-F9-0eSM02-1h
Pic-SM0-F10-0eSM02-1h
920 930 940 950 960
Bar-SM0-F1-0eSM02-1a
Bar-SM0-F2-0eSM02-1b
Bar-SM0-F3-0eSM02-1c
Bar-SM0-F4-0eSM02-1d
Bar-SM0-F5-0eSM02-1d
Bar-SM0-F6-0eSM02-1d
Bar-SM0-F7-0eSM02-2a
Bar-SM0-F8-0eSM02-2a
Bar-SM0-F9-0eSM02-2b
Bar-SM0-F10-0eSM02-2c
Fran-SM02-F1-0eSM02-1a
Fran-SM02-F2-0eSM02-1a
Fran-SM02-F3-0eSM02-1b
Fran-SM02-F4-0eSM02-1c
Fran-SM02-F5-0eSM02-1d
Fran-SM02-F6-0eSM02-1d
Fran-SM02-F7-0eSM02-1e
Fran-SM02-F8-0eSM02-1e
Fran-SM02-F9-0eSM02-1f
Pic-SM0-F1-0eSM02-1a
Pic-SM0-F2-0eSM02-1b
Pic-SM0-F3-0eSM02-1c
Pic-SM0-F4-0eSM02-1d
Pic-SM0-F5-0eSM02-1g
Pic-SM0-F6-0eSM02-1g
Pic-SM0-F7-0eSM02-1g
Pic-SM0-F8-0eSM02-1g
Pic-SM0-F9-0eSM02-1h
Pic-SM0-F10-0eSM02-1h

Figure IVB Alignment of full length *SMO2* sequences deduced from olive cultivarsBarnea, Frantoio and Picual

Cultivars: Bar: Barnea; Fran: Fran; Pic: Picual; The numbering system relates to the nucleotide number in the sequences. The START (ATG) and STOP (TAG/TAA) codon are highlighted in green. The primer pair for the isolation of full length *SMO2* genes are highlighted in red. F1-F10 represents the clones isolated from olive cultivars Barnea, Frantoio and Picual.

APPENDIX V GENE EXPRESSION ANALYSIS USING qPCR TECHNIQUE

		60S RBP			PP2A			EFI-alpha			GAPDH	
Samples	**Cq value**	**Average Cq**	**SD**	**Cq value**	**Average Cq**	**SD**	**Cq value**	**Average Cq**	**SD**	**Cq value**	**Average Cq**	**SD**
B-1-09	28.89	29.015	0.17	29.51	29.47	0.05	21.41	21.615	0.28	26.99	27.04	0.07
Repli. of B-1-09	29.14			29.43			21.82			27.09		
B-2-09	31.24	31.14	0.14	35.5	34.915	0.82	24.76	24.74	0.02	30.6	30.505	0.13
Repli. of B-2-09	31.04			34.33			24.72			30.41		
B-3-09	30.32	30.175	0.20	31.43	31.26	0.24	22.26	22.34	0.11	27.56	27.655	0.13
Repli. of B-3-09	30.03			31.09			22.42			27.75		
B-4-09	33.86	34.04	0.25	33.06	33.58	0.73	24.44	24.43	0.01	28.07	28.285	0.30
Repli. of B-4-09	34.22			34.1			24.42			28.5		
F-1-09	30.69	30.71	0.02	32.16	32.06	0.14	21.97	21.935	0.04	27.85	27.805	0.06
Repli. of F-1-09	30.73			31.96			21.9			27.76		
F-2-09	32.15	32.185	0.04	32.12	32.37	0.35	23.15	23.17	0.02	28.03	28.13	0.14
Repli. of F-2-09	32.22			32.62			23.19			28.23		
F-3-09	31.98	31.965	0.02	32.97	32.945	0.03	22.81	22.815	0.00	29.23	29.405	0.24
Repli. of F-3-09	31.95			32.92			22.82			29.58		
F-4-09	31.46	31.315	0.20	34.24	34.465	0.31	23.59	23.11	0.67	29.33	29.235	0.13
Repli. of F-4-09	31.17			34.69			22.63			29.14		
P-1-09	31.1	31.065	0.04	31.6	31.665	0.09	22.93	22.88	0.07	27.59	27.655	0.09
Repli. of P-1-09	31.03			31.73			22.83			27.72		
P-2-09	32.2	32.095	0.14	32.42	32.355	0.09	24.57	24.555	0.02	28.27	28.565	0.41
Repli. of P-2-09	31.99			32.29			24.54			28.86		
P-3-09	31.79	31.86	0.09	32.53	32.535	0.00	22.74	22.765	0.03	28.03	28.095	0.09
Repli. of P-3-09	31.93			32.54			22.79			28.16		
P-4-09	32.09	32.12	0.04	33.54	33.605	0.09	25.17	25.31	0.19	30.86	30.735	0.17
Repli. of P-4-09	32.15			33.67			25.45			30.61		

Table VA The transcription profiles of individual reference genes given as Cq values across all samples in Barnea, Frantoio and Picual
Average Cq values with the standard deviation (SD) for all samples shown.
Repli: Replicate
Annotation for each sample with their name of olive cultivar, timepoint and year has been given in Table 2.15.

		UBQ			TIP2			PTB			TUBA	
Samples	**Cq value**	**Average Cq**	**SD**	**Cq value**	**Average Cq**	**SD**	**Cq value**	**Average Cq**	**SD**	**Cq value**	**Average Cq**	**SD**
B-1-09	23.03	23.485	0.64	26.87	27.54	0.94	32.37	32.27	0.14	27.51	27.49	0.02
Repli. of B-1-09	23.94			28.21			32.17			27.47		
B-2-09	26.28	25.705	0.81	35.97	36.07	0.14	37.63	37.42	0.29	30.47	30.455	0.02
Repli. of B-2-09	25.13			36.17			37.21			30.44		
B-3-09	25.88	26.28	0.56	39.31	38.82	0.69	32.63	33.15	0.73	33.04	32.5	0.76
Repli. of B-3-09	26.68			38.33			33.67			31.96		
B-4-09	23.94	24.075	0.19	35.86	35.695	0.23	37.04	37.2	0.22	31.34	31.425	0.12
Repli. of B-4-09	24.21			35.53			37.36			31.51		
F-1-09	24.12	24.04	0.11	35.12	35.31	0.26	34.64	34.375	0.37	31.19	30.95	0.33
Repli. of F-1-09	23.96			35.5			34.11			30.71		
F-2-09	23.25	23.41	0.22	34.63	34.47	0.22	34.58	34.355	0.31	28.13	27.7	0.60
Repli. of F-2-09	23.57			34.31			34.13			27.27		
F-3-09	23.45	23.685	0.33	36.21	36.16	0.07	35.15	34.875	0.38	29.83	30.12	0.41
Repli. of F-3-09	23.92			36.11			34.6			30.41		
F-4-09	24.21	23.945	0.37	32.21	32.445	0.33	34.19	34.345	0.21	32.14	32.22	0.11
Repli. of F-4-09	23.68			32.68			34.5			32.3		
P-1-09	21.8	21.735	0.09	27.55	27.67	0.16	33.71	33.38	0.46	28.2	28.3	0.14
Repli. of P-1-09	21.67			27.79			33.05			28.4		
P-2-09	25.92	26.205	0.40	38.92	38.45	0.66	34.84	34.68	0.22	29.15	28.99	0.22
Repli. of P-2-09	26.49			37.98			34.52			28.83		
P-3-09	24.22	24.4	0.25	35.42	35.66	0.33	36.16	36.09	0.09	29.7	29.58	0.16
Repli. of P-3-09	24.58			35.9			36.02			29.46		
P-4-09	25.94	25.73	0.29	35.33	35.42	0.12	34.83	35.26	0.60	32.69	32.98	0.41
Repli. of P-4-09	25.52			35.51			35.69			33.27		

Table VB The transcription profiles of individual reference genes given as Cq values across all samples in Barnea, Frantoio and Picual
Average Cq values with the standard deviation (SD) for all samples shown.
Repli: Replicate
Annotation for each sample with their name of olive cultivar, timepoint and year has been given in Table 2.15.

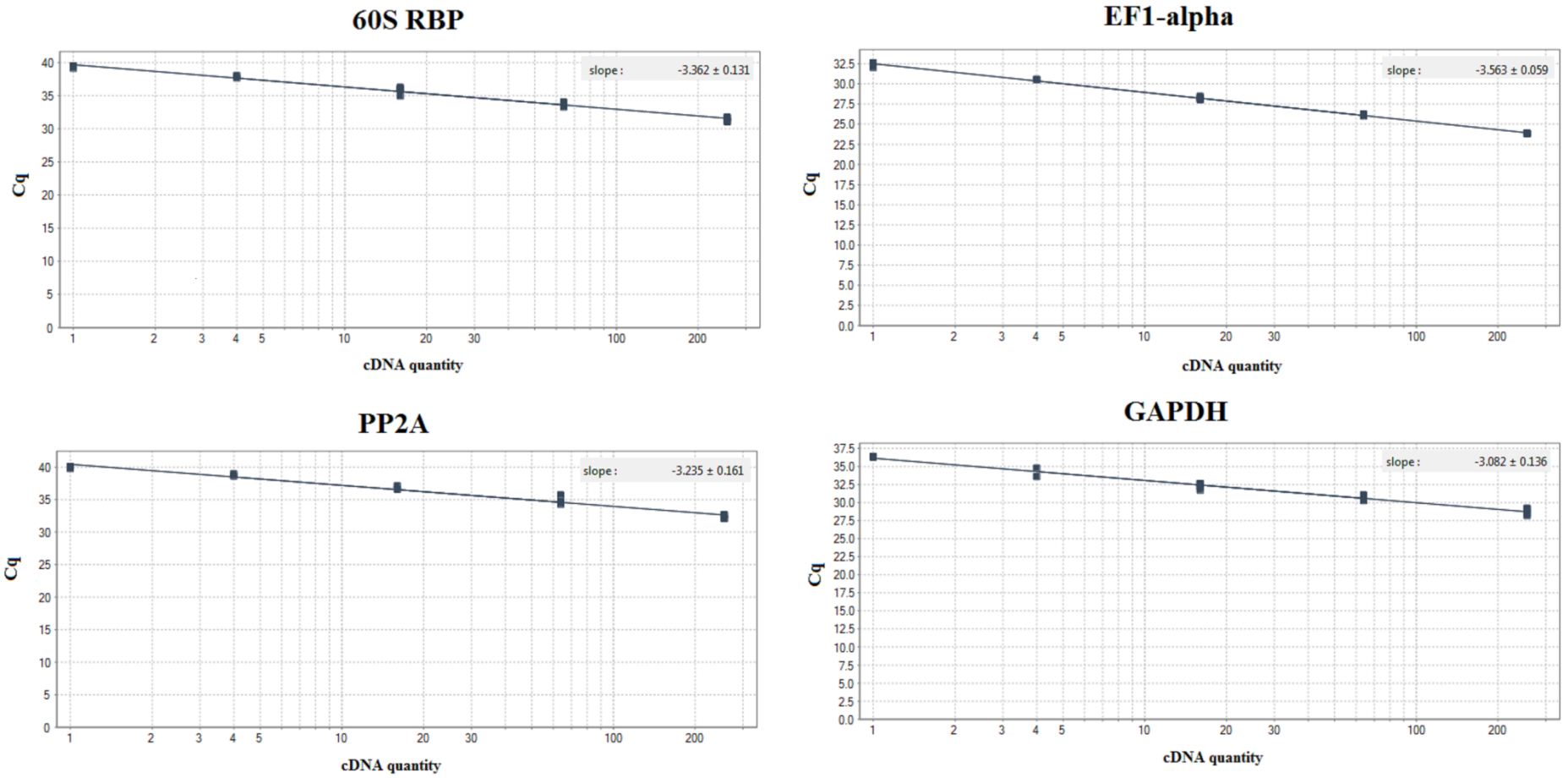


Figure VC Representative efficiency curvesfor individual reference genes
Mean Cq values were plotted against the five four-fold cDNA serial dilutions (Section 2.10.2) using the qBase Plus software. Slope obtained for each plot has been shown top right.

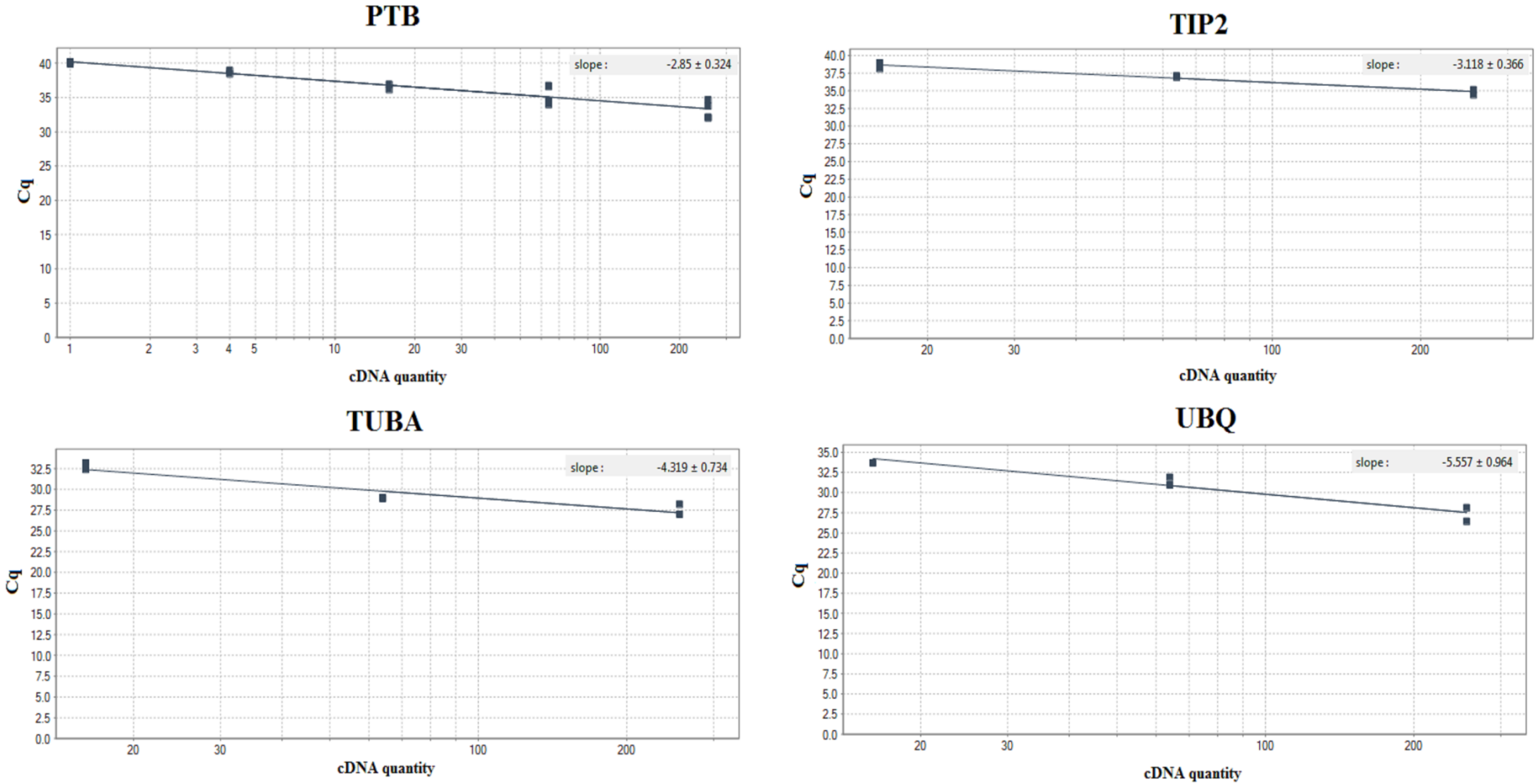


Figure VD Representative efficiency curvesfor individual reference genes
Mean Cq values were plotted against the five four-fold cDNA serial dilutions (Section 2.10.2) using the qBase Plus software.
Slope obtained for each plot has been shown top right.
For reference genes TIP2, TUBA and UBQ only three four-fold cDNA serial dilutions were used.

Target	Cultivar	Comparison	p-value	Fold change	95% CI low	95% CI high
SMO2	Barnea	B-1-09/B-2-09	0.06	12.26	0.005	1.383
		B-1-09/B-3-09	0.11	1.53	0.259	1.637
		B-1-09/B-4-09	0.001	14.82	0.041	0.111
		B-2-09/B-3-09	0.088	9.23	0.267	0.356
		B-2-09/B-4-09	0.623	1.20	0.053	13.036
		B-3-09/B-4-09	0.02	9.64	0.039	0.278
	Frantoio	F-1-09/F-2-09	0.076	2.08	0.156	1.478
		F-1-09/F-3-09	0.116	2.625	0.04	3.633
		F-1-09/F-4-09	0.134	3.2	0.063	1.519
		F-2-09/F-3-09	0.397	1.26	0.232	2.716
		F-2-09/F-4-09	0.117	1.55	0.309	1.346
		F-3-09/F-4-09	0.449	1.23	0.286	2.306
	Picual	P-1-09/P-2-09	0.139	1.50	0.406	1.084
		P-1-09/P-3-09	0.427	1.09	0.473	1.754
		P-1-09/P-4-09	0.00	25.07	0.03	0.06
		P-2-09/P-3-09	0.132	1.37	0.673	2.801
		P-2-09/P-4-09	0.00	16.63	0.04	0.1
		P-3-09/P-4-09	0.01	22.8	0.025	0.076

Table VE Statistical analysis of *SMO2* expression across different timepoints in Barnea, Frantoio and Picual
Experiments were performed in duplicate and the data was analysed using the unpaired-t tests (two-tailed), showing the p-value, fold change ratio and range of 95% confidence interval (CI). Samples showing significant differences in *SMO2* expression with a p-value ≤ 0.05 have been highlighted in green.

Target	Timepoint	Comparison	p-value	Fold change	95% CI low	95% CI high
SMO2	**96 DAF**	Barnea/Frantoio	0.077	1.91	0.192	1.423
		Barnea/Picual	0.02	3.02	0.205	0.532
		Frantoio/Picual	0.214	1.58	0.234	1.703
	109 DAF	Barnea/Frantoio	0.136	3.08	0.211	45.016
		Barnea/Picual	0.16	2.68	0.168	42.983
		Frantoio/Picual	0.374	1.14	0.516	1.471
	116 DAF	Barnea/Frantoio	0.09	3.26	0.033	2.81
		Barnea/Picual	0.01	2.16	0.366	0.581
		Frantoio/Picual	0.25	1.50	0.198	11.502
	136 DAF	Barnea/Frantoio	0.077	2.40	1.137	5.076
		Barnea/Picual	0.01	5.12	0.12	0.316
		Frantoio/Picual	0.02	12.31	0.036	0.183
	(96+109+116+139) DAF	Barnea/Frantoio	1	1.04	0.17	6.398
		Barnea/Picual	0.565	1.88	0.042	6.716
		Frantoio/Picual	0.456	1.96	0.049	5.26

Table VF Statistical analysis of *SMO2* expression across Barnea, Frantoio and Picual in different timepoints
Experiments were performed in duplicate and the data was analysed using the unpaired-t tests (two-tailed) or Mann Whitney 1 tests, showing the p-value, fold change ratio and range of 95% confidence interval (CI). Samples showing significant differences in *SMO2* expression with a p-value ≤ 0.05 have been highlighted in green. DAF: days after flowering.

Target	Timepoint	Comparison	p-value	Fold change	95% CI low	95% CI high
SMT2	**Barnea**	B-1-09/B-2-09	0.068	9.34	0.006	1.782
		B-1-09/B-3-09	0.04	5.97	0.04	0.724
		B-1-09/B-4-09	0.074	6.73	0.012	1.902
		B-2-09/B-3-09	0.435	1.56	0.02	123.985
		B-2-09/B-4-09	0.582	1.38	0.157	12.274
		B-3-09/B-4-09	0.781	1.12	0.015	52.695
	Frantoio	F-1-09/F-2-09	0.611	1.16	0.072	18.983
		F-1-09/F-3-09	0.02	1.92	1.386	2.666
		F-1-09/F-4-09	0.134	1.41	0.318	1.579
		F-2-09/F-3-09	0.267	1.64	0.138	19.674
		F-2-09/F-4-09	0.249	1.64	0.09	4.109
		F-3-09/F-4-09	0.04	2.71	0.209	0.65
	Picual	P-1-09/P-2-09	0.426	1.06	0.526	1.683
		P-1-09/P-3-09	0.03	3.38	1.601	7.159
		P-1-09/P-4-09	0.00	2.54	0.377	0.41
		P-2-09/P-3-09	0.01	3.59	0.673	2.801
		P-2-09/P-4-09	0.03	2.39	0.227	0.767
		P-3-09/P-4-09	0.02	8.60	0.054	0.251

Table VG Statistical analysis of *SMT2* expression across different timepoints in Barnea, Frantoio and Picual
Experiments were performed in duplicate and the data was analysed using the unpaired-t tests (two-tailed), showing the p-value, fold change ratio and range of 95% confidence interval (CI). Samples showing significant differences in *SMT2* expression with a p-value ≤ 0.05 have been highlighted in green.

Target	Timepoint	Comparison	p-value	Fold change	95% CI low	95% CI high
SMT2	**96 DAF**	Barnea/Frantoio	0.059	4.41	0.039	1.3
		Barnea/Picual	0.07	4.15	0.033	1.734
		Frantoio/Picual	0.27	1.06	0.78	1.445
	109 DAF	Barnea/Frantoio	0.196	2.47	0.264	23.19
		Barnea/Picual	0.286	2.11	0.027	164.155
		Frantoio/Picual	0.614	1.16	0.069	10.564
	116 DAF	Barnea/Frantoio	0.01	2.60	1.932	3.504
		Barnea/Picual	0.01	4.86	3.416	6.932
		Frantoio/Picual	0.03	1.87	1.314	2.662
	136 DAF	Barnea/Frantoio	0.859	1.07	0.034	34.622
		Barnea/Picual	0.416	1.57	0.008	51.541
		Frantoio/Picual	0.106	1.69	0.192	1.81
	96+109+116+139) DAF	Barnea/Frantoio	0.971	1.12	0.296	4.241
		Barnea/Picual	0.871	1.12	0.216	5.812
		Frantoio/Picual	0.999	1.00	0.266	3.765

Table VH Statistical analysis of *SMT2* expression across Barnea, Frantoio and Picual in different timepoints

Experiments were performed in duplicate and the data was analysed using the unpaired-t tests (two-tailed) or Mann Whitney 1 tests, showing the p-value, fold change ratio and range of 95% confidence interval (CI). Samples showing significant differences in *SMT2* expression with a p-value ≤ 0.05 have been highlighted in green. DAF: days after flowering.