AN ABSTRACT OF THE DISSERTATION OF

Gunnar Felix Schuppert for the degree of Doctor of Philosophy in Crop Science presented on October 21, 2004.

Title: Molecular Mechanisms Underlying the High Oleic Acid Phenotype in Sunflower

Abstract approved: Redacted for Privacy

Steven J. Knapp

Wildtype sunflower typically produces 12-24% oleic acid (18:1) and 70-82% linoleic acid (18:2). High-oleic sunflower, by contrast, produces up to 80-94% oleic acid. The monounsaturated oleic acid has a greater oxidative stability than the polyunsaturated linoleic acid, predominant in wildtype sunflower, and has greater nutritional benefits than polyunsaturated and saturated fatty acids. High oleic sunflower lines are based on high-oleic acid germplasm, which originated from an induced mutation (Ol_1). Ol_1 is necessary but often not sufficient for producing the high oleic phenotype, presumably because additional quantitative trait loci (QTL) segregate in some genetic backgrounds. The seed specific oleate desaturase (*FAD2-1*), which displays greatly reduced transcript levels in high oleic lines, has been established as the principal candidate gene affecting oleic acid content in sunflower kernels and cosegregates with Ol_1.

In this research, we demonstrated that the *FAD2-1* gene is tandemly duplicated in mutant high oleic sunflower lines, with the two copies being separated by 3.1 kb. We showed that reduced transcript levels of *FAD2-1* in developing kernels of high oleic lines were caused by the RNA interference (RNAi) pathway. Further evidence was obtained when sense and antisense transcripts of *FAD2-1* were detected in mutant lines; wildtype lines only produced transcripts in the sense direction. Bi-directional

transcripts formed double-stranded RNA, which serve as a trigger for RNAi machinery.

In addition, we compared the transcript levels of 48 glycerolipid biosynthetic genes in the developing kernels of four low and four high oleic sunflower lines using microarrays. The analyses revealed that the lipid transfer protein was the only additional gene besides *FAD2-1* with differing transcript levels in each of the four low versus high oleic comparisons. The microarray experiments also revealed no significant differences in transcript levels for genes directly involved in oleic acid synthesis for three of the four comparisons. Only one comparison showed differing transcript levels for about a quarter of the genes in the glycerolipid biosynthesis.

Sequence-based molecular markers were developed for eleven candidate genes involved in the oleic acid biosynthesis and used to identify candidate genes for the QTL underlying oleic acid content in sunflower seeds. A population of 262 recombinant inbred lines segregating for oleic acid content was analyzed and revealed that the oleic acid phenotype is caused by the intralocus and interlocus effects of several genes. These findings emphasize the complexity of the phenotype and indicate a limitation in the applicability of marker-assisted selection for high oleic acid content. Overall, our understanding of the molecular mechanisms underlying the high oleic acid phenotype in sunflower increased.

Molecular Mechanisms Underlying the High Oleic Acid Phenotype in Sunflower

by

Gunnar Felix Schuppert

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented October 21, 2004
Commencement June 2005

Doctor of Philosophy dissertation of Gunnar Felix Schuppert presented on October 21, 2004.

APPROVED:

Redacted for Privacy

Major Professor, representing Crop Science

Redacted for Privacy

Head of the Department of Crop and Soil Science

Redacted for Privacy

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of the Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Redacted for Privacy

Gunnar Felix Schuppert, Author

ACKNOWLEDGEMENTS

I am completely indebted to Dr. Steven J. Knapp for everything he has poured into this research and the development of my career. I am very grateful for his enthusiasm and encouragement, which have inspired me greatly. I feel very privileged to have been part of his lab, which constitutes an outstanding research environment.

I would also like to thank the members of my graduate committee, Dr. James C. Carrington, Dr. Patrick M. Hayes, Dr. Oscar Riera-Lizarazu and Dr. Jeffrey J. Morrell. I am very grateful for the time that each member of my graduate committee has invested in this process and ultimately in this thesis.

Special thanks and gratitude are extended to current and former members of the lab, who have supported me along the way. Dr. Mary B. Slabaugh has been an incredible resource regarding all aspects and intricacies of molecular biology. Dr. B. Shaun Bushman has not only shared his vast expertise in the aspects of RNA work with me, but has been a most helpful and reliable resource during the later stages of my research. In addition, the microarray experiments would not be part of this thesis without him. Jimmie Crane and Bob Brunick were kind enough to assure that the sunflowers were healthy and productive in the field and greenhouse. Thanks are due to the fellow graduate students for their help, support and friendship along the way: Sureeporn, Ju-Kyung, Sonali, Kishore, Cate, Carrie, Vid, Guillermo, Robin and Carol. In addition I would like to extend my thanks to Shunxue, Adam, Pablo, Judy and Sergei for their help and technical assistance. I wish to thank Caprice Rosato for her patience and unreserved help with the microarray experiments and Dr. Kristin D. Kasschau for introducing me to the world of small RNAs.

The funding and support from the USDA grant 'Genomics and Genetics of Nutritionally Important Traits in Sunflower' to Dr. Steven J. Knapp is also acknowledged.

I feel very grateful for the friends I have made during my time at Oregon State University. I am thankful for all their friendship and support and lasting memories,

which we share. I would also like to thank my friends back at home, who have invested much effort to keep our friendships alive despite great distance and rare visits.

Even though far from Corvallis, my work owes completely to my parents, brother and family. Without their love and support I could not have accomplished as much as I have in life.

Last, but not least, I thank Hayley for her patience, support and encouragement. It was her love that created a second home for me here in Oregon.

CONTRIBUTION OF AUTHORS

Dr. Steven J. Knapp proposed and guided every aspect of this research. He substantially assisted in formulating hypotheses and greatly contributed to the writing and polishing of this thesis. Dr. B. Shaun Bushman developed the microarray and contributed in all aspects of the RNA work. Mary B. Slabaugh helped and advised in regard to the molecular techniques and analyses. Dr. James C. Carrington and Dr. Kristin D. Kasschau provided initial input and guidance in the RNAi work. Shunxue Tang isolated the BAC clone carrying the *FAD2-2* gene. Adam Heesacker assisted with microsatellite screening and genotyping. Jimmie Crane and Robert Brunick helped with all technical aspects of growing sunflower in the field and greenhouse in Corvallis, OR. Glenn Cole and Eric Hoeft from Pioneer Hi-Bred International provided the RIL population and conducted the field trials in Woodland, CA. Caprice Rosato from the Central Services Laboratory of the Center for Gene Research and Biotechnology at OSU performed the microarray hybridizations and scans.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES (CONTINUED)

LIST OF TABLES

LIST OF APPENDICES

DEDICATION

To my grandfather

Dr. Carl-Ernst Büchting

Molecular Mechanisms Underlying the High Oleic Acid Phenotype in Sunflower

CHAPTER 1

INTRODUCTION

The recent awareness of the profound negative effects of trans fatty acids on the levels of low density lipoprotein cholesterol in humans has prompted a reassessment of the fats and oils in our diet (Ascherio and Willet, 1997, Lichtenstein et al., 1999). The main source for oils used in commercial food production are vegetable oils, such as corn (*Zea mays* L.), soybean (*Glycine max* L.) or sunflower (*Helianthus annuus* L.) oil. These oils are rich in polyunsaturated fatty acids, such as linoleic (18:2) and linolenic acid (18:3), and currently dominate the global marketplace (FAOSTAT data, 2004). Trans fatty acids are the result of hydrogenating the double bonds present in the polyunsaturated fatty acids, which is necessary to increase oil stability and thus the ability to process it. Oils mainly composed of monounsaturated fatty acids, such as oleic acid (18:1), do not require hydrogenation in order to be processed commercially. In addition research has shown that diets based on monounsaturated oils, like olive (*Olea europaea* L.) or canola (*Brassica napus* L.) oil, effectively decrease low density lipoprotein and increase high density lipoprotein cholesterol compared to diets based on saturated oils, thus lowering the risk of coronary heart disease (Grundy et al., 1986, Kris-Etherton et al., 1999).

Wildtype sunflower typically produces 12-24% oleic acid (18:1) and 70-82% linoleic acid (18:2) (Dorrel and Vick, 1997). However, mid-oleic (NuSun) and high-oleic sunflower lines have been developed, producing 55-75% (NuSun) and 80-94% oleic acid (Dorrel and Vick, 1997). High oleic sunflower oil is higher in monounsaturated fat than olive and canola oil and lower in saturated fat than olive oil (USDA National Nutrient Database). NuSun and high oleic hybrids, which account for about 65% of the total oilseed sunflower acreage in the US in 2003 (National Sunflower Association, 2004), are based on high-oleic acid germplasm, which

originated from an induced mutation (Ol_1) discovered by Soldatov (1976) and released as Pervenets, an open-pollinated cultivar descended from VNIIMK8931 (the open-pollinated M_0 population). Ol_1 is necessary but often not sufficient for producing the high oleic phenotype, presumably because additional quantitative trait loci (QTL) segregate in some genetic backgrounds (Miller et al., 1987a, Fernández-Martínez et al., 1989; Fernandez et al., 1999, Lacombe et al., 2002a).

Since the discovery of high oleic phenotypes (Soldatov, 1976), several models have been proposed to explain the genetics of the high oleic acid phenotype in sunflower (Urie, 1985; Miller et al., 1987a; Fernández-Martínez et al., 1989; Fernandez et al., 1999, Lacombe and Bervillé, 2001, Lacombe et al., 2002a, 2002b). Ol_1 has been the only common denominator in the different genetic models (Urie, 1985; Miller et al., 1987a; Fernández-Martínez et al., 1989; Fernandez et al., 1999, Lacombe and Bervillé, 2001, Lacombe et al., 2002a, 2002b). Progeny from wildtype x mutant (low oleic x high oleic) crosses either produce discrete, non-overlapping oleic acid distributions typical of the segregation of a single dominant mutation (Ol_1) or continuous, overlapping oleic acid distributions typical of the segregation of one or more QTL. Miller et al. (1987a) concluded that the high oleic phenotype was controlled by Ol_1 and a 'modifier' (*ml*). Besides *ml*, other putative modifiers have been proposed on the basis of phenotypic analyses, e.g. Ol_2, Ol_3, and *supole* (Fernández- Martínez et al., 1989; Lacombe et al., 2001); however, the assignment of progeny to putative genotypic classes has often been based on arbitrary breakpoints in the tails of complex overlapping oleic acid classes.

The principal candidate gene for QTL affecting oleic acid concentration is oleate desaturase (*FAD2*) (Hongtrakul et al. 1998a, Lacombe and Bervillé, 2001, Pérez-Vich et al., 2002). Hongtrakul et al. (1998a) isolated an oleate desaturase cDNA (*FAD2-1*) that is highly expressed in developing kernels. Using RT-PCR analysis, they found that *FAD2-1* was weakly expressed in developing kernels of mutant lines. The expression of *FAD2-1* was quantitatively different among three wildtype lines (HA89, HA292, and RHA274), differences that carried through to mutant BC_1F_4 lines (HA341, HA349, and RHA345) developed using a common donor (Pervenets).

Martínez-Rivas et al. (2001) described two additional members of the *FAD2* gene family (*FAD2-2* and *FAD2-3*). Using Northern analysis, they reproduced the results of Hongtrakul et al. (1998a) for *FAD2-1* and reported that the expression of *FAD2-2* and *FAD2-3* were unchanged in wildtype and mutant lines (HA89 and HAOL9, respectively). Southern analysis of wildtype and mutant lines, when *EcoRI* digested DNA was probed with *FAD2-1*, revealed two bands (5.3 and 7.2 kb) in high oleic acid lines and one band (5.3 kb) in low oleic acid lines (Hongtrakul et al., 1998a). When Southern analyses were based on digests with *HindIII*, wildtype lines produced a single 7.3 kb band (*FAD2-1* wildtype allele) and high oleic lines produced a single 14.3 kb band (*fad2-1* mutant allele). These results indicated that the *FAD2-1* locus is duplicated and rearranged in the mutant lines (Hongtrakul et al., 1998a).

A survey of 114 low and 125 high oleic sunflower accessions revealed complete correlation between the high oleic phenotype and the *FAD2-1* RFLP marker developed by Hongtrakul et al. (1998a) (Lacombe and Bervillé, 2001). This was further substantiated by cosegregation of the high oleic phenotype with the *fad2-1* allele in two independent populations segregating for high oleic acid content, indicating tight linkage between the *FAD2-1* locus and Ol_1 (Lacombe and Bervillé, 2001, Lacombe et al., 2002a). The BD40713 x BE78079 F_2 population, derived from a cross between a low and a high oleic acid line, showed a bimodal distribution with an observed 1:2:1 segregation ratio of the RFLP marker, an F_6 recombinant inbred line (RIL) population displayed a continuous distribution, based on a cross between low and high oleic acid lines (83HR4 x RHA345). In the RIL population, 43 of the 96 lines carrying the *fad2-1* allele revealed a low oleic phenotype (< 55% oleic acid), while no lines with a high oleic phenotype were found among the 90 lines with the *FAD2-1* allele (Lacombe et al., 2002a). Therefore, Lacombe et al. (2002a) postulated that an independent locus, termed supole, segregated among the 96 RILs carrying the *fad2-1* allele, which is necessary for producing high oleic acid.

Pérez-Vich et al. (2002) constructed an RFLP map using F_2 progeny from a cross between high oleic and high stearic acid lines (HAOL9 x CAS-3). The F_2 distribution for oleic acid concentration was quantitative (discontinuous). *FAD2-1*

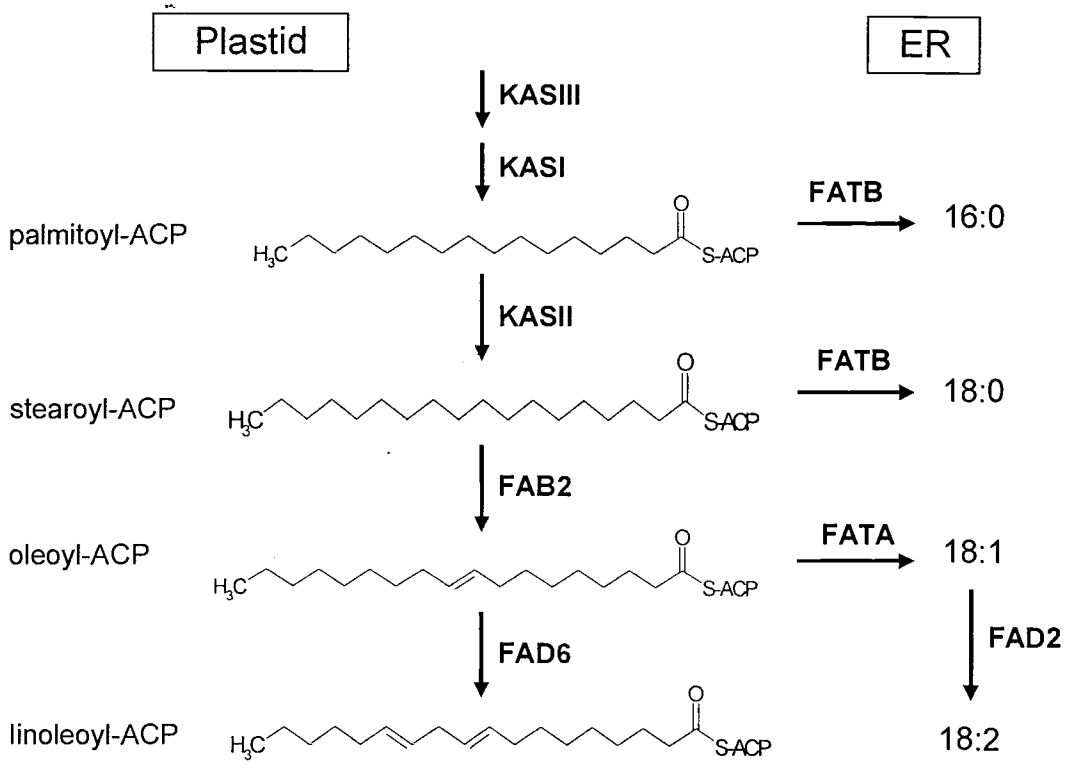
mapped to linkage group (LG) 14 coincident with the major QTL, which is presumed to be Ol_1 (Pérez-Vich et al. 2002). In addition to the QTL on LG 14, Pérez-Vich et al. (2002) detected a QTL on LG 1, spanning the map position of a stearoyl-ACP desaturase (*FAB2-1*), and a minor QTL on LG 8.

Mutations in two independent *FAD2* (*ahFAD2A* and *ahFAD2B*) loci were responsible for an increase in the oleic acid content in the peanut (*Arachis hypogaea* L.) seed oil (Jung et al., 2000a, 200b). One mutation caused a severe reduction in *ahFAD2B* transcript levels in high oleate peanut varieties. Another mutation resulted in a change of the DNA coding sequence in *ahFAB2A*, causing a change in the amino acid at a conserved residue, thereby causing a loss of oleyl-PC desaturase activity in the high oleic varieties (Jung et al., 2000b). In addition, transformation of cotton (*Gossypium hirsutum* L.) with constructs carrying an inverted repeat of *FAD2* resulted in a loss of *FAD2* transcripts in the developing seed, and an increase in oleic acid content in cottonseed (Liu et al., 2002). These observations further emphasize that *FAD2* is the principal candidate gene for QTL affecting oleic acid concentration.

The results obtained by Pérez-Vich et al. (2002) indicated that additional loci such as *FAB2-1*, which is directly involved in oleic acid biosynthesis, may determine the high oleic acid phenotype in sunflower. Hongtrakul et al. (1998b) isolated two distinct cDNAs coding for the stearoyl-ACP desaturase (*FAB2-1* and *FAB2-2*) in sunflower, which are both highly expressed in the developing kernels of wildtype and mutant lines. FAB2 is a plastidial desaturase, which introduces the first double bond in the 18:0-ACP, thereby producing 18:1-ACP (Figure 1.1) (Ohlrogge and Browse, 1995). Thus, FAB2 directly catalyzes the final step necessary in the synthesis of oleic acid.

The glycerolipid biosynthetic pathway has been studied in depth in model plant systems (*Arabidopsis thaliana*) during the last fifteen years. Since this pathway is essential and thus conserved among plants, it allowed us to apply the knowledge obtained in Arabidopsis to identify additional candidate genes for the high oleic phenotype (Figure 1.1). One group of likely candidates is the acyl-ACP thioesterases (FAT), which hydrolyze the acyl group from the fatty acid acyl-ACP. The free fatty

Figure 1.1. Schematic diagram of part of the glycerollipid biosynthetic pathway.



acids are then subsequently exported from the plastid into the cytosol and primarily used for glycerolipid biosynthesis at the endoplasmic reticulum (Browse and Somerville, 1991) (Figure 1.1). Arabidopsis carries two classes of FAT enzymes, FATA and FATB, with differing substrate preference. While FATA preferentially hydrolyzes unsaturated oleoyl-ACP, FATB shows highest in vitro activity on saturated acyl groups (Salas and Ohlrogge, 2002). Bonaventure et al. (2003) observed a drastic reduction in palmitate (16:0) and stearate (18:0) in glycerolipids in leaves, flowers, roots, and seeds of Arabidopsis plants carrying a disrupted *FATB* gene.

Stoutjesdijk et al. (2002) transformed Arabidopsis with an inverted repeat construct of *FAD2* driven by the seed specifc napin promoter, resulting in a drastic increase in oleic acid and a decrease in linoleic and linolenic acid in the seed. The transformed plants produced trace amounts of linoleic and linolenic acid, which was attributed to the activity of the plastidial oleate desaturase (FAD6) (Stoutjesdijk et al., 2002). A similar hypothesis has been put forward by Liu et al. (2002) to explain the residual levels of linoleic acid in high oleic cottonseed. Therefore, *FAD6* constitutes another candidate gene with possible effects on the high oleic phenotype in sunflower (Figure 1.1).

Three separate condensing enzymes are required to produce oleic acid. While KASIII catalyzes the condensation of acetyl-CoA and malonyl-ACP, which marks the first condensation step in fatty acid biosynthesis, KASI acts on the subsequent seven condensation reactions, which leads to the formation of palmitoyl-ACP (Ohlrogge and Browse, 1995). The final condensation step in the biosynthesis of oleic acid is performed by KASII, which is specific for this single enzymatic reaction (Figure 1.1) (Ohlrogge and Browse, 1995). The KAS enzymes have been associated with a role in the control of the flux through the fatty acid biosynthetic pathway (Ohlrogge and Browse, 1995). The direct influence of KASII activity on the ratio of the saturated fatty acids palmitate (16:0) and stearate (18:0), together with the putative control function of the KASs, added these genes to list of candidates involved in the high oleic trait.

The goal of this study was to gain a deeper understanding of the genetic basis for the high oleic phenotype in sunflower. In order to achieve this goal we formulated a number of specific objectives for our research, which are stated below:

- Determine the structure of the *FAD2-1* gene duplication and rearrangement in the mutant lines.
- Elucidate the mechanism underlying the reduced *FAD2-1* expression in the mutant lines.
- Assess the expression profiles of the glyceroplipid genes in four comparisons of low versus high oleic lines using microarray.
- Develop and map sequence based markers for the candidate genes involved in oleic acid biosynthesis and assess their diversity on a panel of sunflower lines.
- Test for association between the developed markers of the candidate genes and the oleic acid seed content in a RIL population segregating for oleic acid content.

Microarray Analysis of Developing Sunflower Kernels and Characterization of the *FAD2-1* Duplication in High Oleic Acid Sunflower

CHAPTER 2

INTRODUCTION

Recent advances in the study of the glycerolipid biosynthetic pathway in the model species *A. thaliana* have led to the identification and characterization of the key genes involved in de novo fatty acid synthesis (Ohlrogge and Browse, 1995). This discovery has greatly benefited and advanced from the vast amounts of publicly available sequences, contributed from the genome sequencing projects of Arabidopsis and rice (*Oryza sativa* L.) and various other plant EST projects (Mekhedov et al., 2000). Mekhedov et al. (2000) identified 65 plant polypeptides involved in the lipid metabolism in the Arabidopsis and rice genome, respectively. In addition, Girke et al. (2000) produced a microarray with 2,600 seed-expressed genes based on a cDNA library from developing Arabidopsis seed. Seed-specific microarrays facilitate the analysis of expression of a large number of genes simultaneously, thus revealing tissue specific expression patterns or identifying possible candidate genes for subsequent examination (Girke et al., 2000).

The development of the Composite Genome Project EST database (CGPDB) (http://cgpdb.ucdavis.edu) allowed us to identify sunflower EST sequeneces for nearly every gene involved in lipid metabolism (Kozik et al., 2002). These sequences were used to develop a miniature microarray for genes in the glycerolipid biosynthesic pathway in order to study the expression profiles of these genes during early kernel development in low oleic (wildtype) and high oleic (mutant) lines from different genetic backgrounds.

Independent of the genetic background of the cross, the oleate desaturase (*FAD2*) has been established as the principal candidate gene affecting oleic acid content in the sunflower seed (Hongtrakul et al., 1998a, Lacombe and Bervillé, 2001,

Pérez-Vich et al., 2002). Hongtrakul et al. (1998a) isolated an oleate desaturase cDNA (*FAD2-1*) that is highly expressed in developing kernels. Using RT-PCR analysis, they found that *FAD2-1* was weakly expressed in developing kernels of mutant lines. The expression of *FAD2-1* was quantitatively different among three wildtype lines (HA89, HA292, and RHA274), differences that carried through to mutant BC_1F_4 lines (HA341, HA349, and RHA345) developed using a common donor (Pervenets) and the respective wildtype lines. Observed expression differences for *FAD2-1* were independently confirmed by Martínez-Rivas et al (2001) and Lacombe et al. (2002b).

Southern analysis of wildtype and mutant lines, when *EcoRI* digested DNA was probed with *FAD2-1*, revealed two bands (5.3 and 7.2 kb) in high oleic acid lines and one band (5.3 kb) in low oleic acid lines (Hongtrakul et al., 1998a). When Southern analyses were based on digests with *HindIII*, wildtype lines produced a single 7.3 kb band (*FAD2-1* wildtype allele) and high oleic lines produced a single 14.3 kb band (*fad2-1* mutant allele). These results indicated that the *FAD2-1* locus was duplicated and rearranged in the mutant lines (Hongtrakul et al., 1998a, Lacombe and Bervillé, 2001). Lacombe et al. (2002b) isolated four phage clones carrying the *FAD2-1* locus, derived from genomic DNA of the high oleic line RHA345. Sequencing of two of the clones revealed a complete copy of *FAD2-1*, as previously reported by Hongtrakul et al. (1998a), including an intron in the 5' UTR of the gene.

Stoutjesdijk et al. (2002) transformed Arabidopsis with a construct carrying an inverted repeat of *FAD2* driven by the seed specific napin promoter, which resulted in a drastic increase in oleic acid and a reduction in linoleic and linolenic acid in the seed. The transcription of the inverted repeat of *FAD2* resulted in the formation of double-stranded RNA, which served as a trigger for RNA induced gene silencing (for reviews see Hannon, 2002, and Matzke and Matzke, 2004). The transformation of cotton with a similar inverted repeat construct of *FAD2* driven by the seed-specific soybean lectin promoter resulted in high oleic cotton lines (Liu et al., 2002). The increase in oleic acid content in cottonseed was accompanied by a loss of *FAD2* transcription in developing seeds.

In the following we demonstrate that the high oleic sunflower lines carry two complete copies of the *FAD2-1* gene in tandem array separated by 3.1 kb. Further we report that the greatly reduced transcript levels of *FAD2-1* in the developing kernels seem to be caused by RNA interference (RNAi) pathway. Northern analyses of developing kernels of wildtype and mutant lines, using *FAD2-1* specific probes, identified a 21 nt short-interfering RNA (siRNA) in mutant lines only. In addition, we show that the mutant lines produced *FAD2-1* transcripts in sense and antisense directions, whereas wildtype lines only produced transcripts in the sense direction. The bi-directional transcripts of *FAD2-1* result in double-stranded RNA, which in turn serves as a trigger for the RNAi machinery.

MATERIALS AND METHODS

Plant Materials and DNA/RNA Isolation

The wildtype lines HA89 and RHA274 are low oleic oilseed inbred lines. HA292 is a low oleic confectionary inbred line. The three mutant high oleic lines HA341, HA349 and RHA345 were derived by selecting for high oleic acid content among BC_1F_2 and BC_1F_3 progeny from crosses between the three recurrent parents and the high oleic donor cultivar Pervenets (HA89*2/Pervenets, HA292*2/Pervenets and RHA274*2/Pervenets), respectively (Miller et al., 1987b). PHC is a high oleic and PHD is a low oleic proprietary oilseed line developed by Pioneer Hi-Bred International (Johnston, Iowa). VNIIMK8931 and Pervenets are open pollinated oilseed cultivars developed in Russia and seed samples were acquired from Mary Brothers (USDA-ARS National Plant Germplasm System, North Central Plant Introduction Station, Ames, Iowa).

Leaves were harvested from individual 3-week-old greenhouse-grown plants and stored at -80°C until extraction. The frozen leaf samples were ground in liquid N_2, and DNA was isolated from the ground samples using a modified CTAB method (Webb and Knapp, 1990).

Developing kernels at 10, 14, 18 and 22 days after flowering (DAF) were harvested from bagged heads grown in Corvallis, Oregon, in 2002, and immediately frozen in liquid N_2 and stored at -80°C. One gram of frozen kernels was ground in liquid N_2, and total RNA was extracted by adding 10 ml of the TRIzol reagent (Invitrogen, Carlsbad, CA, USA). The samples were vortexed and incubated at RT for 5 min. After adding 2 ml of chloroform the samples were mixed vigorously and centrifuged at 10,000 x g for 20 min at 4°C. The top phase was transferred to a fresh tube; 5 ml of Isopropanol were added, and incubated at RT for 10 min. After centrifugation at 3,000 x g for 20 min the supernatant was discarded and the pellet washed with 10 ml of 75% Ethanol. The samples were spun at 3,000 x g for 10 min and the ethanol wash was repeated. The supernatant was discarded and the pellet air-

dried for 5 min. The RNA was resuspended in 500 µl of diethyl pyrocarbonate-treated (DEPC) deionized water. Low molecular weight RNA was isolated subsequently for the northern blot analysis using the RNA/DNA midi kit (Qiagen, Valencia, CA, USA) according to the instructions of the manufacturer.

Fatty Acid Analysis

Individual seeds were analyzed for the four fatty acids of palmitic (16:0), stearic (18:0), oleic (18:1) and linoleic acid (18:2) according to the method described by Brandt and Knapp (1993) with slight modifications. Briefly, single seeds or half seeds were crushed with a plastic rode in 1 ml of hexane, incubated at RT for 10 min, 0.5 ml was transferred to a fresh tube and evaporated to dryness at 50°C under a N_2 stream. Lipids solubilized in 0.1 ml ethyl ether were converted to fatty acid methyl esters (FAMEs) by adding 0.1 ml of 0.1 M KOH in methanol and incubating for 5 min at 50°C. Methylation reactions were stopped with 0.1 ml of 0.15 M HCl, and FAMEs were extracted with 1 ml Hexane. One-microliter samples were injected in a HP 6890 Series Gas Chromatograph (Hewlett-Packard, Wilmington, DE, USA), and fatty acid profiles were calculated using the HP ChemStation software (Hewlett-Packard, Wilmington, DE, USA).

Microarray Experiments and Data Analysis

The custom array comprised 43 sunflower genes of the glycerolipid biosynthetic pathway and 5 'housekeeping' genes. Sunflower sequences in GenBank (NCBI, Bethesda, MD, USA), in the Composite Genome Project EST database (CGPDB) (http://cgpdb.ucdavis.edu) (Kozik et al., 2002) and in an in-house developing kernel library were used for the design of 46 of the 48 oligonucleotides spotted in the array (Table 2.2). Targets Han22 and Han23 were based on sequence derived from *Carthamus tinctorius* GenBank accessions (Table 2.2). The 48 genes were used to design 70 nt long oligonucleotides, custom manufactured by Qiagen (Qiagen, Valencia, CA, USA). The oligos were spotted on amino silane coated UltraGAPS slides (Corning, Corning, NY, USA) with DNA at 25 µM in 3xSSC, 1.5

M Betaine buffer using a BioRobotics MicrogridII arrayer (Genomic Solutions, Ann Arbor, MI, USA), followed by a 1 hr bake (1st edition) or a 24 hr desiccation (2nd edition) and final crosslinking at 600 mJ in a UV Stratalinker 2400 (Stratagene, La Jolla, CA, USA). In addition to the 48 sunflower oligos (8 replicates/oligo) the array included 4 human and vector genes as negative controls (8 replicates/control), 32 buffer blank wells, 10 spiking controls of 70 oligonucleotides in length (2 replicates/spike in 1st edition, 8 replicates/spike in 2nd edition) of the SpotReport Alien Oligo Array Validation System (Stratagene, La Jolla, CA, USA) and 332 (1st ed.) or 192 (2nd ed.) empty wells.

Three arrays were probed and analyzed for the comparison of low versus high oleic sunflower lines HA89-HA341, HA292-HA349 and RHA274-RHA345. Each of these three comparisons included a dye-swap between the cy3 and cy5 labels from the same RNA extraction, and an additional RNA extraction as a biological replicate. Four arrays were probed and analyzed for the comparison of low versus high oleic sunflower line PHD-PHC, including a dye-swap experiment for two independent RNA extractions (biological replicates). Thus each comparison is represented by at least three array hybridizations.

Ten μg of total RNA from developing kernels at 18 DAF was used for each of the two dyes (cy3 and cy5) for cDNA synthesis, labeling and hybridization. All procedures, including the subsequent washes, were conducted according to the 3DNA Array 350 Expression Array Detection Kit for Microarrays (Genisphere, Hatfield, PA, USA). The arrays were scanned using a ScanArray 4000 (PerkinElmer, Boston, MA, USA) scanner with the accompanying software ScanArray Express version 1.1 (PerkinElmer, Boston, MA, USA). Measurements on the scanned image were made using QuantArray version 3.0 software (PerkinElmer, Boston, MA, USA). All arrays were conducted by the Central Services Laboratory of the Center for Gene Research and Biotechnology at Oregon State University (Corvallis, OR, USA).

Stringent control measures were applied for all steps of the data analysis. The QuantArray output was reduced to two measured variables for each spot and each dye: intensity of the spot and background of the spot. All intensities were adjusted by

subtracting the spot background. Measurement of the overall array background was based on the median of the adjusted intensities of the four negative controls, the buffer blanks and empty wells. Only arrays with similar median intensity for all three categories were analyzed further. General *actin* intensity was based on the average of the eight adjusted *actin* intensities. The adjusted intensities of the 48 genes were then standardized to *actin* by dividing through the general *actin* intensity. The resulting standardized intensities were used in the subsequent statistical analysis using the SAS version 8.02 software (SAS Institute Inc., Cary, NC, USA). The statistical model for the individual array analysis was Han## = genotype (1 or 2), considering the 8 replicates of each gene as a random effect in the model in order to obtain a more accurate error term. Mean intensities for each gene were based on least square means and the p-value for each test was adjusted using the Bonferroni adjustment to account for the 48 simultaneous comparisons (Bonferroni, 1936, Kuehl, 2000). The statistical model for the analysis of the combined array data from the 3(4) arrays (n) was the same with Han## = genotype (1 or 2) treating the array (n) and the array*genotype effects as random effects in the model. Mean intensities for each gene were based on least square means. Intensity differences were considered significantly different at the $\alpha = 0.10$ level in both analyses, but only accepted as real if significant in n-1 of the n arrays. In addition the average intensity of those genes had to be at least twice the median of the overall array background in order to be considered.

Real-time PCR Quantification of mRNAs

A 2 µg aliquot of total RNA from developing kernels at 10, 14, 18 and 22 DAF was used as a template for first strand cDNA synthesis using the SuperScript II RNase H^- Reverse Transcriptase kit (Invitrogen, Carlsbad, CA, USA) according to the instructions of the manufacturer. Genespecific primers (Table 2.5) were designed using either Primer3 software (Rozen and Skaletsky, 2000) or Primer Express software (Applied Biosystems, Foster City, CA, USA). Primers were tested by showing that the PCR-product produced a single band after gel-electrophoresis. Real-time PCR analysis was performed using the QuantiTect SYBR Green PCR Kit

(Qiagen, Valencia, CA, USA) with slight modifications to the instructions of the manufacturer on the ABI PRISM 7000 Sequence Detection System (Applied Biosystems, Foster City, CA, USA). The reaction contained in a final volume of 25 µl, 8 ng of reverse transcribed total RNA, 2.5 µM of the forward and reverse primer and 2x QuantiTect SYBR Green PCR Master Mix (Qiagen, Valencia, CA, USA). Each line assayed for the transcript levels of a particular gene was also assayed for *actin* transcript levels. In addition each real-time PCR run included two sets of standard curves (gene of interest and *actin*) with 32, 16, 8, 4 and 2 ng of reverse transcribed total RNA, respectively. All reactions were performed in triplicate. The comparative threshold (C_t) value was set using the ABI Sequence Detection Systems 1.1 software (Applied Biosystems, Foster City, CA, USA). Levels of the mRNA of the gene of interest were normalized to *actin* mRNA levels. The real-time PCR products were sequenced to confirm gene specificity of the primers.

Long-distance PCR and Sequencing

Long-distance PCR was performed according to Loeffert et al. (1998) with slight modifications. The initial denaturation step was performed at 94°C for 4 min, followed by 11 cycles of 94°C for 20 s, 60°C for 45 s and 68°C for 3 min 30 s. The extension time was increase by 10 s in each of the remaining 24 cycles with a final extension time of 20 min. Sequencing was performed by the Nevada Genomics Center (Reno, NV) using an ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA). Sequence alignments were made using the Vector NTI Suite 8 software programs AlignX and ContigExpress (Invitrogen, Carlsbad, CA, USA).

Northern Blot Analysis

Blot hybridization analysis was performed as described by Llave et al. (2000) with some modifications. Low molecular weight RNA (20 µg) was resolved on denaturing 17% polyacrylamide gels, electroblotted to Hybond-N+ membranes (Amersham Biosciences, Piscataway, NJ, USA) using a Genie Electrophoretic Blotter (Idea Scientific Company, Minneapolis, MN, USA) for 4 h at 200 mA, and UV

crosslinked using a UV Stratalinker 2400 (Stratagene, La Jolla, CA, USA). Radiolabeled probes for the *FAD2-1* coding sequence were made using the Random Primers DNA Labeling System (Invitrogen, Carlsbad, CA, USA) in the presence of α-^{32}P-dCTP (MP Biomedicals, Irvine, CA, USA). Blots used for the analysis of low molecular weight RNA were prehybridized and hybridized using PerfectHyb Plus buffer (Sigma, St. Louis, MS, USA). Hybridization of blots using random-primed *FAD2-1* probes was performed at 38°C for 16 h. Blots were washed at 50°C as described (Llave et al., 2000) and exposed to Sterling High Speed X-Ray Film (Bioworld, Dublin, OH, USA) at -80°C for 2 to 4 days.

Synthetic oligoribonucleotides of 21 and 24 nucleotides (nt) (Dharmacon, Dallas, TX, USA) were used as size standards during electrophoresis. Ethidium bromide staining and visualization of the 5S RNA/tRNA bands in low molecular weight RNA gels were used to monitor the loading of RNA samples. Synthetic oligoribonucleotides specific to *FAD2-1*, *FAD2-2* and *FAD2-3* (Dharmacon, Dallas, TX, USA) of 24 nt were designed using Primer3 software (Rozen and Skaletsky, 2000).

Strand Specific Reverse Transcription

Strand specific reverse transcription was performed as described by Volpe et al. (2002) with minor modifications. Purified high molecular weight RNA of developing kernels at 14 DAF was treated with RQ1 DNase (Promega, Madison, WI, USA) and then analyzed in the reverse transcription reaction with SuperScriptII (Invitrogen, Carlsbad, CA, USA) according to instructions of the manufacturer.

Mcr-PCR

Mcr-PCR was performed according to Lippman et al. (2003) with some modifications. M.SssI (New England Biolabs, Beverly, MA, USA), which methylates all cytosine residues within the double-stranded dinucleotide recognition sequence 5'...CG...3', was incubated with 2μg of genomic DNA for 10 hours according to the instructions of the manufacturer and 96μmol SAM was replenished after 3 and 6 hours

of incubation. Two μg of untreated genomic DNA or 1μg of genomic DNA pretreated with M.SssI were incubated with McrBC (New England Biolabs, Beverly, MA, USA), a methylation-dependent endonuclease that restricts purine-C_{methyl} half-sites separated between 80 bp and up to 3kb, according to the instructions of the manufacturer for 10 hours. One mM GTP was replenished after 3 and 6 hours of digestion. Recovery of the expected PCR fragment following the digest indicates lack of methylation. McrBC activity was verified using the supplied control plasmid. The subsequent PCR was carried out using 40 to 60 ng of template DNA, 0.2 μM primer, 200 μM dNTPs, and 0.8 U of Taq polymerase (New England Biolabs, Beverly, MA, USA) in 20 ul of supplied 1 x reaction buffer. PCR amplification was perfomed in 96-well plates heated to 94°C for 4 min, followed by 34 cycles of 20 sec at 94°C, 30 sec at 57°C, and 1 min 20 sec at 72°C with a final extension time of 10 min at 72°C after the last cycle.

RESULTS

Fatty Acid Profiles of Wildtype (Low Oleic) and Mutant (High Oleic) Lines

The fatty acid composition in the mature kernels of four wildtype (HA89, HA292, RHA274, and PHD) and four mutant (HA341, HA349, RHA345, and PHC) lines of sunflower were quantified using gas chromatography (GC) (Table 2.1). HA341, HA349, RHA345 and PHC produced significantly more oleic acid than HA89, HA292, RHA274 and PHD. The wildtype lines HA89, RHA274, and PHD produced 177.5 g kg^{-1} to 232.9 g kg^{-1}oleic acid and 642.1 g kg^{-1} to 734.9 g kg^{-1} linoleic acid, whereas HA292 produced 322.0 g kg^{-1} oleic acid and 561.8 g kg^{-1} linoleic acid. HA341, HA349, RHA345, and PHC produced 857.9 g kg^{-1} to 889.8 g kg^{-1} oleic acid and 27.8 g kg^{-1} to 70.1 g kg^{-1} linoleic acid in the kernels (Table 2.1).

Microarray of Glycerolipid Biosynthesis Genes

The expression profiles of the glycerolipid biosynthesis genes were assessed for four pairs of low and high oleic acid lines (HA89-HA341, HA292-HA349, RHA274-RHA345, and PHD-PHC) using microarrays, in order to identify any candidate genes besides *FAD2-1* affecting the high oleic acid phenotype across the differing genetic backgrounds. The microarray comprised 43 genes involved in glycerolipid biosynthesis (Table 2.2), as well as five housekeeping genes *α-tubulin* (Han37), *β-tubulin* (Han36), *glyceraldehydes-3-phosphate dehydrogenase* (Han38), *actin* (Han39), and *elongation factor-1* (Han47). All four microarray experiments were probed with RNA extracted from developing kernels at 18 DAF. The comparisons of HA89-HA341, HA292-HA349, and RHA274-RHA345 consisted of three microarrays, including a dye-swap experiment between the cy3 and cy5 label, and an independent second RNA extraction as a biological replicate. The PHC-PHD comparison included four microarrays with a dye-swap experiment for each of the two biological replicates.

Since each oligonucleotide was spotted in eight different locations on the slide, the intensity value for each gene was based on the average of eight individual intensity

Table 2.1. Composition of the four fatty acids palmitic (16:0), stearic (18:0), oleic (18:1) and linoleic acid (18:2) in the mature kernels of eight sunflower lines given in g kg^{-1} seed weight with the respective standard deviation.

Line	Palmitic	Stearic	Oleic	Linoleic
			g kg^{-1}	
HA89	69.3 ± 2.8	51.7 ± 6.0	181.1 ± 12.8	697.8 ± 15.8
HA341	31.8 ± 1.7	53.6 ± 8.8	886.8 ± 12.8	27.8 ± 10.8
HA292	74.0 ± 7.9	42.2 ± 14.6	322.0 ± 47.6	561.8 ± 35.1
HA349	27.8 ± 2.7	43.8 ± 1.2	877.9 ± 13.0	50.5 ± 12.6
RHA274	73.5 ± 3.6	51.5 ± 5.5	232.9 ± 21.2	642.1 ± 21.7
RHA345	35.1 ± 2.4	44.4 ± 6.1	889.8 ± 21.6	30.6 ± 17.2
PHD	71.7 ± 2.4	15.9 ± 0.6	177.5 ± 4.7	734.9 ± 3.3
PHC	35.8 ± 0.8	36.2 ± 4.4	857.9 ± 22.0	70.1 ± 17.7

Table 2.2. DNA sequence sources for the oligonucleotides on the sunflower microarray.

Gene	Array ID	Source
carboxyltransferase alpha chain 1	Han1	QHF6P19.yg.ab1*
carboxyltransferase alpha chain 2	Han2	HAK_A10010_D06042#
biotin-carboxyl carrier protein	Han3	sT3bHAK011A01001#
hydroxyacyl-ACP desaturase	Han5	QHA8N16.yg.ab1
enoyl-ACP reductase 1 (ENR1)	Han8	QHK3L23.yg.ab1
enoyl-ACP reductase 2 (ENR2)	Han7	QHL1P22.yg.ab1
enoyl-ACP reductase 3 (ENR3)	Han6	QHB19F10.yg.ab1
ketoacyl-ACP synthase I (KASI)	Han9	QH_CA_Contig1997*
ketoacyl-ACP synthase II (KASII)	Han10	HAK003B03_HAK013H 11
ketoacyl-ACP synthase III (KASIII)	Han11	QH_CA_Contig1814
stearoyl-ACP desaturase (FAB2-1)	Han13	U91340^
stearoyl-ACP desaturase (FAB2-2)	Han12	U91339
plastidial lysophosphatidic acyl-CoA:acyl-transferase (LPAAT)	Han14	QHK16A23.yg.ab1
acyl carrier protein (ACP)	Han17	sT3bHAK014C01001
keto-acyl-CoA reductase (KR)	Han18	QHK16I10.yg.ab1
keto-acyl-CoA synthase 1	Han19	QH_CA_Contig5178
keto-acyl-CoA synthase 2	Han20	QH_CA_Contig909
keto-acyl-CoA synthase 3	Han21	QH_CA_Contig2991
keto-acyl-CoA synthase 4	Han45	QH_CA_Contig909_1
putative glyceraldehyde-3 phosphate acyl-transferase 1	Han22	L33841 (Carthamus tinctorius)
putative glyceraldehyde-3 phosphate acyl-transferase 2	Han23	L33841 (Carthamus tinctorius)
cytosolic lysophosphatidic acyl-CoA:acyl-transferase (LPAAT)	Han24	QHF16E22.yg.ab1
phosphatidic acid phosphatase	Han25	QHF7M07.yg.ab1
oleosin 1	Han29	QH_CA_Contig3856
oleosin 2	Han28	QH_CA_Contig3447
oleosin 3	Han27	QH_CA_Contig3122
oleosin 4	Han46	QH_CA_Contig442
ER oleate desaturase (FAD2-1)	Han30	AF251842
ER oleate desaturase (FAD2-2)	Han31	AF251843
ER oleate desaturase (FAD2-3)	Han32	AF251844
plastidial oleate desaturase (FAD6)	Han16	QH_CA_Contig1636
ER linoleate desaturase (FAD3)	Han15	QH_CA_Contig2407
plastidial linoleate desaturase (FAD7/FAD8)	Han26	QHL9Co1.yg.ab1
acyl-ACP thioesterase FatB	Han33	AF036565
acyl-ACP thioesterase FatA	Han34	QH_CA_Contig2357

Table 2.2. (Continued)

Genes	Array ID	Source
putative CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase	Han35	QH_CA_Contig1538
β-tubulin	Han36	QH_CA_Contig2700
α-tubulin	Han37	QH_CA_Contig3007
glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	Han38	QHG17I18.yg.ab1
actin	Han39	AF282624
3-methycrontonyl-CoA carboxylase	Han40	QH_CA_Contig3275
homomeric acetyl-CoA carboxylase 1 (ACCase)	Han41	QH_CA_Contig2910
homomeric acetyl-CoA carboxylase 2 (ACCase)	Han42	QH_CA_Contig980
keto-acyl-ACP reductase 1	Han4	QH_CA_Contig629
keto-acyl-ACP reductase 2	Han43	QH_CA_Contig1493
keto-acyl-ACP reductase 3	Han44	QH_CA_Contig629
elongation factor-1	Han47	sT3BHAK014B04016
non specific lipid transfer protein (LTP)	Han48	AF529201

* Singleton/Contig from the CGPDB EST database
# Singleton/Contig from the ESTAP developing kernel library
^ Sequence from Genbank

readings. After subtraction of the background, each individual intensity reading was standardized by dividing through the average *actin* intensity in the array. These eight standardized intensities per oligonucleotide were then analyzed statistically for the individual array and the combined arrays of dye-swap and biological replication. While accounting for multiple comparisons in the individual arrays with a Bonferroni adjustment, we considered transcript differences significant at the 0.10 level, when significant in n-1 of the n arrays (Bonferroni, 1936, Kuehl, 2000). In addition, we only considered those genes in each array, which displayed intensities at least twice above the average background based on the buffer blanks, empty spots and *A. thaliana* aliens.

Results of the Microarray Analysis of HA89-HA341

The comparison of low oleic line HA89 versus high oleic line HA341 yielded only three genes which displayed consistently differential expression: the seed specific *oleate desaturase* (*FAD2-1*) of the endoplasmatic reticulum (Hongtrakul et al., 1998a), a *keto-acyl-CoA reductase* (*KR*), and a non-specific *lipid transfer protein* (*LTP*) (Table 2.3). *FAD2-1* shows significantly higher transcript levels in the low oleic line HA89 in two out of the three arrays with a standardized intensity reading of 3.008 and 3.301 compared to the lower intensities of 0.198 and 0.592 in HA341 (Table 2.3). This was a 15.2-fold difference in array 1 and a 5.7-fold expression difference in array 2 (Table 2.3). Even though intensities for *FAD2-1* were comparable in array 3 with 2.423 for HA89 and 0.1656 for HA341, the difference was not significant. In addition, transcripts levels for *FAD2-1* differed significantly in the combined analysis of all three arrays with a p-value of 0.020. FAD2 is a cytosolic desaturase, bound to the ER membrane, which catalyses the conversion of phoshatidylcholine bound oleic acid to linoleic acid. Thus an increase in transcript levels of the seed specific *oleate desaturase* likely results in increased enzyme levels, which would allow for elevated desaturation activity in the developing seed. This would in turn result in a higher conversion rate of oleic acid to linoleic acid, which is reflected in the fatty acid profiles of the mature kernels (Table 2.1). The fact that lines with greatly reduced

Table 2.3: Results of the four microarray comparisons of HA89-HA341, HA292-HA349, RHA274-RHA345 and PHD-PHC. The intensities for each gene are standardized to the equivalent *actin* intensity and ratios are shows as cy3/cy5, while fold differences are based on simple X/x ratios. The p-value for the individual arrays is adjusted for the 48 simultaneous comparisons using Bonferroni, i.e. 0.005 = 0.0.0001 x 48.

Array ID	Gene	Array 1			Array 2			Array 3			Com bined
		Std. Intensity.	Fold	p-value	Std. Intensity	Fold	p-value	Std. Intensity	Fold	p-value	p-value
HA89 vs. HA341		HA341/HA89			HA89/HA341			HA341/HA89			
Han18	*KR*	ns	ns	ns	0.070/0.154	2.2	0.091	0.103/0.047	2.2	0.029	0.024
Han30	*FAD2-1*	0.198/3.008	15.2	0.048	3.301/0.592	5.7	0.005	ns	ns	ns	0.020
Han48	*LTP*	2.161/1.366	1.6	0.010	3.557/5.095	1.7	0.034	13.135/7.637	1.7	0.014	ns
HA292 vs. HA349		HA292/HA349			HA349/HA292			HA349/HA292			
Han06	*ENR2*	0.165/0.101	1.6	<.005	0.116/0.178	1.5	0.091	0.164/0.592	3.6	0.029	ns
Han12	*FAB2-2*	0.814/0.565	1.5	0.019	ns	ns	ns	1.473/4.444	3.0	<.005	ns
Han14	*LPAAT* plastidial	0.921/0.626	1.5	<.005	0.339/0.867	2.6	<.005	ns	ns	ns	ns
Han17	*ACP*	1.807/1.330	1.4	0.010	ns	ns	ns	1.791/2.389	1.3	0.005	ns
Han18	*KR*	0.118/0.067	1.8	0.029	ns	ns	ns	0.112/0.282	2.5	<.005	ns
Han24	*LPAAT* cytosolic	0.221/0.162	1.4	<.005	0.169/0.237	1.4	0.034	ns	ns	ns	0.095
Han29	*Oleosin1*	7.673/6.384	1.2	<.005	4.781/7.887	1.7	<.005	6.213/8.500	1.4	0.014	0.051
Han30	*FAD2-1*	2.298/0.359	6.4	0.038	0.480/2.716	5.7	<.005	0.160/5.143	32.1	0.010	0.076
Han31	*FAD2-2*	0.136/0.036	3.8	<.005	0.043/0.091	2.1	0.043	0.096/0.217	2.3	0.038	0.054
Han32	*FAD2-3*	0.127/0.032	4.0	0.081	ns	ns	ns	0.085/0.699	8.2	<.005	ns
Han46	*Oleosin4*	0.951/0.706	1.3	0.029	0.741/1.050	1.4	0.019	ns	ns	ns	ns
Han48	LTP	9.209/4.483	2.1	<.005	3.720/9.614	2.6	<.005	7.010/14.782	2.1	<.005	0.020

Table 2.3. (Continued)

Array ID	Gene	Array 1			Array 2			Array 3			Combined
		Std. Intensity	Fold	p-value	Std. Intensity	Fold	p-value	Std. Intensity	Fold	p-value	p-value
RHA274 vs. RHA345		RHA274/RHA345			RHA345/RHA274			RHA345/RHA274			
Han13	*FAB2-1*	ns	ns	ns	1.644/2.432	1.1	0.029	1.698/3.089	1.8	<.005	ns
Han17	*ACP*	ns	ns	ns	1.466/1.801	1.2	0.024	2.356/2.894	1.2	<.005	ns
Han27	*Oleosin3*	4.204/6.089	1.4	0.010	9.972/6.549	1.5	0.062	9.797/7.675	1.3	<.005	0.076
Han30	*FAD2-1*	2.018/0.183	11.4	<.005	ns	ns	ns	0.191/2.614	13.7	<.005	ns
Han38	*GAPDH*	0.568/0.472	1.1	0.038	ns	ns	ns	1.191/1.538	1.3	<.005	ns
Han48	*LTP*	1.091/3.047	2.8	<.005	11.595/6.728	1.7	<.005	17.667/9.525	1.9	<.005	0.108
PHD vs. PHC		PHC/PHD			PHD/PHC			PHD/PHC			
Han27	*Oleosin3*	10.525/6.271	1.7	<.005	7.191/9.202	1.3	0.048	7.155/9.903	1.4	<.005	0.013
Han29	*Oleosin1*	5.731/3.867	1.5	<.005	5.851/9.697	1.7	<.005	7.019/11.000	1.7	<.005	0.021
Han30	*FAD2-1*	0.668/3.368	5.0	0.043	3.762/0.661	5.7	<.005	3.628/0.873	4.2	0.024	0.009
Han48	*LTP*	8.291/3.317	2.5	<.005	5.340/13.813	2.6	<.005	8.379/21.178	2.5	<.005	0.028

Table 2.3. (Continued)

Array ID	Gene	Array 4			Combined
		Std. Intensity	Fold	p-value	p-value
PHD vs. PHC		PHC/PHD			
Han27	*Oleosin3*	9.134/5.656	1.6	<.005	0.013
Han29	*Oleosin1*	5.489/3.884	1.4	0.043	0.021
Han30	*FAD2-1*	0.836/2.007	2.4	0.062	0.009
Han48	*LTP*	9.231/3.779	2.4	<.005	0.028

FAD2-1 mRNA levels all displayed substantially elevated oleic acid content further strengthens this point.

The *keto-acyl-CoA reductase* displayed significant differences in the levels of transcript accumulation in two of the three arrays. The intensities of 0.154 and 0.103 in HA341 compared to 0.070 and 0.047 in HA89 show that mRNA levels were 2.2-fold greater in the high oleic line HA341 in arrays 2 and 3 (Table 2.3). The intensities for *KR* in array 1 also indicate a more than 2-fold difference, but were not significant in the statistical analysis. In the combined analysis, expression differences were clearly significant, resulting in an overall p-value of 0.024 for *KR*. HA341 also revealed 1.6, 1.4- and 1.7-fold higher transcript levels for the lipid transfer protein than HA89 with significant intensity differences in all three arrays. The intensities readings of the *lipid transfer protein* were 2.161, 5.095 and 13.1353 for HA341 and 1.366, 3.557 and 7.636 for HA89 in array 1, 3 and 3, respectively (Table 2.3). These intensity readings indicate that *LTP* is expressed at a much higher rate than *actin* in both genotypes. Despite significant expression differences in the individual arrays, differences in *LTP* transcript levels were not significant in the combined analysis (p-value = 0.216).

Results of the Microarray Analysis of HA292-HA349

The microarray of the low oleic line HA292 versus high oleic line HA349 revealed 11 genes with significant differential expression in at least two of the three arrays (Table 2.3). Among those are the three genes identified in the earlier array of HA89-HA341. The observed intensities for *FAD2-1* were 0.359, 0.480 and 0.160 in HA349 compared to 2.298, 2.716 and 5.143 in the low oleic line, resulting in a 6.4-, 5.7- and 32.1-fold difference in transcript levels for arrays 1, 2 and 3, respectively (Table 2.3). *FAD2-1* was also significant in the combined array analysis with a p-value of 0.076. In addition to the seed specific copy the two remaining paralogues of *FAD2* in the sunflower genome also revealed differential expression. *FAD2-2* and *FAD2-3* displayed a greater number of transcripts in the low oleic line HA292. While *FAD2-2* exhibited 3.8-, 2.1- and 2.3-fold higher expression levels in the three arrays; *FAD2-3*

only displayed significant differences in two of the arrays varying from 4.0- to 8.2-fold in arrays 1 and 3. Both genes are transcribed at much lower levels than the seed specific *FAD2-1*, which is illustrated by the relatively low intensity readings for *FAD2-2* with 0.136, 0.091 and 0.217 in HA292 and 0.036, 0.043 and 0.096 in the high oleic line HA349 (Table 2.3). In addition *FAD2-3* showed intensities of 0.127 and 0.699 for HA292 and 0.032 and 0.085 for HA349. While *FAD2-2* proved also significant in the overall analysis with a p-value of 0.054, *FAD2-3* was not significant, when the three arrays were combined. The non-specific *LTP* was abundantly expressed in both lines compared to the standard *actin*. The intensity readings of 9.209, 9.614 and 14.782 for HA292 and of 4.483, 3.720 and 7.010 for HA349 resulted in a 2.1-, a 2.6- and a 2.1-fold transcript difference. These results were confirmed in the combined analysis resulting in a p-value of 0.020 for *LTP*. Transcript levels were 1.8- and 2.5-fold greater for *KR* in the low oleic line HA292 in arrays 1 and 3 and the differences are close to the 2.2-fold difference observed in HA89-HA341 (Table 2.3). But *KR* was not significant in the combined analysis.

In addition, two of four oligonucleotides designed to the oleosin gene family displayed significant differential expression. While *oleosin1* was expressed at very high levels in both lines, demonstrated by intensity readings of 7.673, 7.887 and 8.500 for HA292 and 6.384, 4.781 and 6.213 for HA349, the difference in transcript accumulation was fairly low with 1.2, 1.7 and 1.4 in all three arrays. *Oleosin4* displayed only a marginal change in transcript levels with a 1.3- and a 1.4-fold higher expression in HA292 for arrays 1 and 2. But *oleosin4* was expressed at very low levels resulting in intensities of 0.951 and 1.050 for HA292 and 0.706 and 0.741 in HA349 (Table 2.3). In the combined analysis of the three arrays transcript levels differed significantly for *oleosin1* (p-value = 0.051), while the differences were non-significant for *oleosin4*. With the *enoyl-ACP reductase2* (*ENR2*) another member of a distinct gene family displayed significant expression differences. The gene is generally expressed at low levels, showing 1.6-, 1.5- and 3.6-fold higher transcript levels in the low oleic line HA292 (Table 2.3). But *ENR2* was non-significant in the combined comparison. Both *lysophosphatidic acyl-CoA:acyl-transrerases* (*LPAAT*), the

plastidial and the cytoplasmic form, showed elevated transcript levels in the low oleic line HA292 in arrays 1 and 2. The plastidial *LPAAT* varied in transcription levels between 1.4- and 2.5-fold demonstrated by intensities of 0.921 and 0.867 for HA292 and 0.626 and 0.339 for HA349 (Table 2.3). The cytoplasmic counterpart showed a constant 1.4-fold difference with 0.221 and 0.237 for HA292 and 0.162 and 0.169 for HA349, respectively. While the plastidial gene was non-significant in the combined analysis, the cytoplasmic *LPAAT* proved marginally significant with a p-value of 0.0946.

The *acyl carrier protein* (*ACP*) exhibited intensities of 1.807 and 2.389 for HA292 and 1.330 and 1.791 for HA349, translating into a 1.4- and a 1.3-fold greater transcript accumulation in HA292. The intensities indicated that *ACP* was expressed at higher levels than the *actin* standard in the developing kernel. Even though the differences in transcript levels for *ACP* were significant in array 1 and 3, they were non-significant in the overall comparison. Last, but not least, the *stearoyl-ACP desaturase2* (*FAB2-2*) revealed significant discrepancies in its transcript levels ranging from an 1.5-fold increase in array 1 to a 3.0-fold higher expression rate in array 3 for HA292. The expression of *FAB2-2* was neither significantly different in array 2 nor in the combined analysis.

Results of the Microarray Analysis of RHA274-RHA345

The microarray comparison of the low oleic line RHA274 with the high oleic line RHA345 revealed six genes with significant differences in their transcript accumulation in the developing kernel at 18 DAF. Each of the six genes proved significant in at least two of the three arrays, which were the *stearoyl-ACP desaturase1* (*FAB2-1*), *ACP*, *Oleosin3*, the seed specific *FAD2-1*, the *glyceraldehydes-3-phosphate dehydrogenase* (*GAPDH*) and *LTP* (Table 2.3). *FAB2-1* showed a 1.1-fold difference in transcript levels in array 3 and 1.8-fold in array 1. The measured intensities of 2.432 and 3.709 for RHA274 and 1.644 and 3.407 for RHA345 revealed a highly expressed *FAB2-1* gene. Transcript levels did not differ significantly in array 2 and in the combined analysis of the three arrays. The *ACP* also

displayed a greater rate of expression than the *actin* standard demonstrated in the intensity readings of 1.801 and 2.389 for RHA274 and 1.466 and 1.791 for RHA345. This was equivalent to a 1.2-fold higher expression in the low oleic line RHA274. *ACP* proved neither significant in array 1 nor in the overall comparison. In addition *oleosin3* displayed significant differential expression. While *oleosin3* was expressed at very high levels in both lines, demonstrated by intensity readings of 6.089, 9.972 and 9.797 for RHA274 and 4.204, 6.549 and 7.675 for RHA345, the difference in transcript accumulation was fairly low with 1.4, 1.5 and 1.3 in all three arrays. Transcript levels of *oleosin3* were also significantly different in the combined analysis demonstrated by a p-value of 0.076.

As seen in the previous two arrays, the seed specific *FAD2-1* was highly expressed in the low oleic line RHA274, resulting in intensities of 2.018 and 5.143. *FAD2-1* also exhibited a substantial reduction in transcript accumulation in the high oleic line RHA345, which was 11.4-fold lower in array 1 and with 13.7-fold slightly lower in array 2 (Table 2.3). Despite substantial differences in the intensities for HA292 and HA349 in array 2, *FAD2-1* was not significant in the statistical analysis. The same was observed in the combined analysis of the three arrays with a p-value of 0.235. The differences in transcript levels were comparably small for *GAPDH* with a 1.1- to 1.3-fold increase in the low oleic line RHA274. *GAPDH* revealed no significant transcript differences in array 2 as well as in the combined dataset. As seen before, *LTP* displayed a relatively high rate of transcription demonstrated by intensities of 3.047, 11.595 and 17.667 for RHA345 and 1.091, 6.728 and 9.525 for RHA274 (Table 2.3). This was equivalent to a 2.8-, 1.7- and 1.9-fold increase in mRNA levels in the high oleic line RHA345. In the combined analysis *LTP* reached only borderline significance with a p-value of 0.108.

Results of the Microarray Analysis of PHD-PHC

The last microarray comparison between the low oleic line PHD and the high oleic line PHC yielded four genes with significant differences in mRNA accumulation in at least three of the four microarrays. Two of four microarray targets designed to the

oleosin gene family displayed significant differential gene expression. While *oleosin1* was expressed at very high levels in both lines, demonstrated by intensity readings of 5.713, 9.687, 11.000 and 5.489 for PHC and 3.867, 5.851, 7.019 and 3.884 for PHD, the difference in transcript accumulation was relatively low with 1.5, 1.7, 1.7 and 1.4 between the two lines in all four arrays. *Oleosin3* exhibited a much greater rate of transcription than the *actin* standard and displayed only a modest change in transcript levels as well, with a 1.7-, 1.3-, 1.4- and 1.6-fold higher expression in PHC for the arrays 1, 2, 3 and 4, respectively. Both were also significant in the combined analysis, reflected by the p-values of 0.013 and 0.021 for *oleosin1* and *3*.

FAD2-1 showed similar transcriptional differences as observed in the three earlier microarray experiments. Its expression was greatly reduced in the high oleic line PHC with intensity readings ranging from 0.661 to 0.837 compared to the highly expressed copy in PHD with a range from 2.007 to 3.762 in measured intensities (Table 2.3). This was equivalent to a 5.0-, 5.7-, 4.2- and 2.4-fold difference in the arrays 1, 2, 3 and 4, respectively. These results were confirmed in the analysis of the combined dataset with a p-value of 0.009. In addition, the previously identified *LTP* displayed 2.4- to 2.6-fold higher transcript levels in the high oleic line PHC. The gene was highly expressed in both lines with intensities of 8.291, 13.813, 21.178 and 9.231 for PHC and 3.317, 5.340, 8.379 and 3.779 for PHD when compared to the *actin* standard (Table 2.3). The transcriptional differences were also significant in the combined analysis (p-value = 0.022).

Overall the microarray experiments revealed only a few genes with significant differences in their transcript levels. The seed specific *FAD2-1* displayed a greatly reduced accumulation of mRNA transcripts in all four high oleic lines. The only other gene involved in the glycerolipid biosynthesis, which showed consistent differential expression across all four comparisons was the non-specific *LTP*. But the differences in gene expression of *LTP* were not associated with the high oleic acid phenotype, since HA292 was the only low oleic line revealing a higher transcription rate than the high oleic counterpart. *ACP*, *KR*, *oleosin1* and *oleosin3* showed differing rates of transcription in two of the four microarray comparisons. The remaining genes only

displayed significant differences in transcript levels in the individual low by high oleic acid comparisons.

Examination of Microarray Inconsistencies

We examined those genes, which revealed insignificant differences in transcript levels in one array, while displaying significant differences in the remaining arrays. Most microarray experiments have few replicates for each gene on the array slide (e.g. Schenk et al., 2000, Weber at al., 2004). Due to the limited number of genes in our array we were able to include eight replicates per gene on each slide. The eight intensity readings per gene enabled us to detect and discard inconsistencies within the array with high statistical stringency. While *FAD2-1* expression differences were highly significant in arrays 1 and 3 for RHA274-RHA345 (p-values of <.005 for both arrays), the gene proved not significant in array 2 and in the combined dataset (Table 2.3). A look at the averaged intensities (2.018/0.183, 6.409/0.532 and 2.614/0.191 in arrays 1, 2 and 3) of *FAD2-1* for RHA274/RHA345 suggested a significant difference for array 2 as well. When we examined the individual eight intensity readings of *FAD2-1* in each array, we observed that the standard deviation of the eight readings was 0.70 and 0.09 in array 1 and 0.80 and 0.04 in array 3 for RHA274 and RHA345. In contrast the standard deviations in array 2 were 5.95 and 0.43, respectively. Thus the variation within the intensity readings of *FAD2-1* for RHA274 in array 2 was so substantial that the subsequent statistical test deemed it insignificant. The same effect caused an insignificant result for *GAPDH* in this particular array. We also observed substantial variation within the eight intensity readings of *KR* in array 1 for the HA89-HA341 comparison as well as *FAD2-1* in array 3 for HA89-HA341. A moderate lack of consistency within the intensity octets was found for all the genes listed as non significant in one of the three arrays of HA292-HA349 and the two genes *FAB2-1* and *ACP* in array 1 of RHA274-RHA345 (Table 2.3). Thus the large number in replicates per gene allowed us to limit the number of significant genes to those with consistent effects within each array.

The large variation in the intensity values of *FAD2-1* in array 2 for RHA274-RHA345 impacted the analysis of the combined data for the three arrays of RHA274-RHA345 in a similar fashion. In the test of the combined dataset, *FAD2-1* displayed a substantial array*genotype interaction effect leading to an insignificant test result across arrays. In addition we observed large array*genotype interactions for *LTP* in the HA89-HA341 and the RHA274-RHA345 comparison, demonstrated by the large range of intensity values among the three arrays, and thus resulting in an overall insignificance of the expression differences (Table 2.3). Overall, the thorough examination of the raw data allowed us to identify the cause of the inconsistencies observed in the replicate microarrays.

Validation of Microarray Results

We selected a subset of the genes with differential transcript levels for the validation of the results of the microarray experiments. We used 2-step real-time PCR to assess the relative transcript levels of *FAD2-1*, *FAD2-2*, *FAD2-3*, *KR*, *LTP*, *FAB2-1*, *FAB2-2*, and *FATA* among the eight lines. We quantified the mRNA abundance of the gene of interest and of the *actin* gene simultaneously in the same real-time PCR reaction, thus allowing us to assess the relative transcript levels compared to *actin*.

Real-time PCR showed a 9.8-fold difference in transcript levels for *FAD2-1*, which fell between the two values of 5.7 and 15.2 in the arrays of HA89-HA341. While the microarray showed a range from 5.7 and 6.4 to 32.1 in HA292-HA349 for increased *FAD2-1* expression in HA292, the real-time PCR yielded a 13.4-fold higher expression rate in the same line (Table 2.4) (Primers in Table 2.5). Similar numbers were obtained in RHA274-RHA345 with an 11.4- and 13.7-fold higher expression rate in the low oleic line. With a 12.1-fold increase in transcript levels for RHA274 the real-time PCR result fully supports the findings of the array experiment. Real-time PCR displayed an elevation of 23.9-fold for *FAD2-1* mRNA accumulation in the developing kernel of PHD, while the values of the four microarrays of PHD-PHC only ranged from 2.4 to 5.7 for the low oleic line.

Table 2.4. Comparison of the real-time PCR results and microarray results for the purpose of validation. The relative amount is a ratio of the amounts of cDNA, normalized to *actin*, of the gene of interest in the respective lines.

Gene	Real-time PCR		Array 1	Array 2	Array 3	Array 4
	Gene/*actin*	Fold	Fold	Fold	Fold	Fold
HA89/HA341						
FAD2-1	2.554/0.260	9.8	15.2	5.7	ns	-
FAD2-2	5.297/7.662	0.7	ns	ns	ns	-
FAD2-3	8.906/6.994	1.3	ns	ns	ns	-
FAB2-1	1.680/1.099	1.5	ns	ns	ns	-
FATA	2.327/1.787	1.3	ns	ns	ns	-
KR	0.669/9.101	13.6	ns	2.2	2.2	-
LTP	0.596/1.085	1.8	1.6	1.7	1.7	-
HA292/HA349						
FAD2-1	1.262/0.094	13.4	6.4	5.7	32.1	-
FAD2-2	0.441/0.204	2.2	3.8	2.1	2.3	-
FAD2-3	6.757/3.362	2.0	4.0	ns	8.2	-
FAB2-2	100.650/64.891	1.6	1.5	ns	3.0	-
FATA	2.589/1.687	1.5	ns	ns	ns	-
KR	8.825/5.405	1.6	1.8	ns	2.5	-
LTP	0.803/0.204	3.9	2.6	2.1	2.6	-
RHA274/RHA345						
FAD2-1	1.943/0.161	12.1	11.4	ns	13.7	-
FAD2-2	0.088/0.077	1.1	ns	ns	ns	-
FAD2-3	2.457/3.042	0.8	ns	ns	ns	-
FAB2-1	1.033/1.013	1.0	ns	1.1	1.8	-
FAB2-2	1.217/2.162	0.6	ns	ns	ns	-
FATA	2.133/2.780	0.8	ns	ns	ns	-
KR	0.643/0.642	1.0	ns	ns	ns	-
LTP	0.401/1.025	2.6	2.8	1.7	1.9	-
PHD/PHC						
FAD2-1	7.516/0.314	23.9	5.0	5.7	4.2	2.4
FAD2-2	17.845/9.439	1.9	ns	ns	ns	ns
FAD2-3	13.611/5.675	2.4	ns	ns	ns	ns
FAB2-1	2.797/2.106	1.3	ns	ns	ns	ns
FATA	3.391/3.635	0.9	ns	ns	ns	ns
KR	0.800/0.463	1.7	ns	ns	ns	ns
LTP	0.667/0.324	2.1	-2.5	-2.6	-2.5	-2.4

Table 2.5. Oligonucleotide primer names and sequences used in the real-time PCR assay to asses transcript levels.

Name	Sequence forward primer	Sequence reverse primer
FAD2-1	5'-ACACGTCTGTGACCAACGAA-3'	5'-GACAGCGGTTATGGTGAGGT-3'
FAD2-2	5'-CAGGTTAGTGAACCATGGGTG-3'	5'-TGACAGACCGGTTGAAACAA-3'
FAD2-3	5'-GTAGGTCACTAAACAATGGGTGC-3'	5'-ACAGATCGCTTAAAGCAGTGG-3'
FAB2-1	5'-GGTGACGTGAAGCTGGCTCA-3'	5'-CGAGAACAGTGCCGTCCG-3'
FAB2-2	5'-AATCTGAACAACTGCCGACTGC-3'	5'-CCATCGCGAATTTAGGAGAT-3'
FATA	5'-ATCGATTCCGTTTCAATTCG-3'	5'-ACAATCCGTCTTCCGTCAAG-3'
KR	5'-ATTACGGCCGGGCTAGGT-3'	5'-CAATAGCAGCACCAGATCCA-3'
LTP	5'-ATGGCTCAACTCATGGTGCAC-3'	5'-CAGCACAACAAGCCGGAGTC-3'
Actin	5'-GCAAAAAGCAGCTCGTCTGT-3'	5'-AGCAGCTTCCATTCCAATCA-3'

FAD2-2 and *FAD2-3* only differed significantly in the HA292-HA349 comparison with higher transcript levels observed in the low oleic line HA292. The 3.8-, 2.1- and 2.3-fold elevated transcript levels for *FAD2-2* in the array were confirmed by the 2.2-fold expression difference observed in the real-time PCR (Table 2.4). Real-time PCR indicated an almost 2-fold difference in transcript levels between PHD and PHC, which was not found in the microarray experiments. While we reported a 4-fold and an 8.2-fold increase in transcripts of *FAD2-3* for HA292 in the arrays, we observed only a 2-fold difference in the quantitative PCR. Similarly to *FAD2-2*, quantitative PCR detected a 2.4-fold difference in PHD compared to PHC, while the transcript level differences were not significant in the arrays.

Real-time PCR and microarray findings were relatively similar for *FAB2-1* in HA89-HA341, RHA274-RHA345, and PHD-PHC. *FAB2-1* transcript levels were not significantly different in HA89-HA341 and PHD-PHC, which is in agreement with the 1.5-fold and 1.3-fold differences detected by real-time PCR, respectively. While the array yielded only slight differences of 1.1- and 1.8-fold higher expression for RHA345, the real-time PCR detected no change in expression between the two lines (Table 2.4). The *FAB2-1* results for HA292-HA349 were not repeatable and therefore omitted. *FAB2-2* showed a 1.5- and 3.0-fold discrepancy in HA292-HA349 with reduced accumulation in HA349, while real-time PCR indicated a 1.6-fold greater expression rate in the low oleic line HA292 (Table 2.4). RHA345 showed a 1.8-fold increase in *FAB2-2* transcript levels compared to RHA274 in the real-time PCR, while the arrays proved not significant. HA89-HA341 and PHD-PHC were not assessed by real-time PCR.

FATA transcript levels did not differ significantly in any of the four microarray experiments. This was confirmed by real-time PCR, which only detected very modest differences among the lines ranging from a 1.5-fold in HA292-HA349 to 1.1-fold in PHD-PHC (Table 2.4).

According to the results of the microarray experiment *KR* transcription rates differed for HA89-HA341 and HA292-HA349. The latter comparison exhibited a 1.8- and 2.5-fold increase in *KR* transcripts for HA292. This result was confirmed with a

1.6-fold increase for HA292 in the quantitative PCR (Table 2.4). But real-time PCR indicates a substantial 13.6-fold elevation in mRNA levels of *KR* in HA341. The microarrays displayed only a modest 2.2-fold increase in the high oleic line HA341 (Table 2.4). Differences in transcript levels for *KR* were not significant in the arrays of RHA274-RHA345 and PHD-PHC, which was confirmed by real-time PCR for RHA274-RHA345, but PHD-PHC displayed a modest 1.7-fold difference in *KR* transcript levels (Table 2.4).

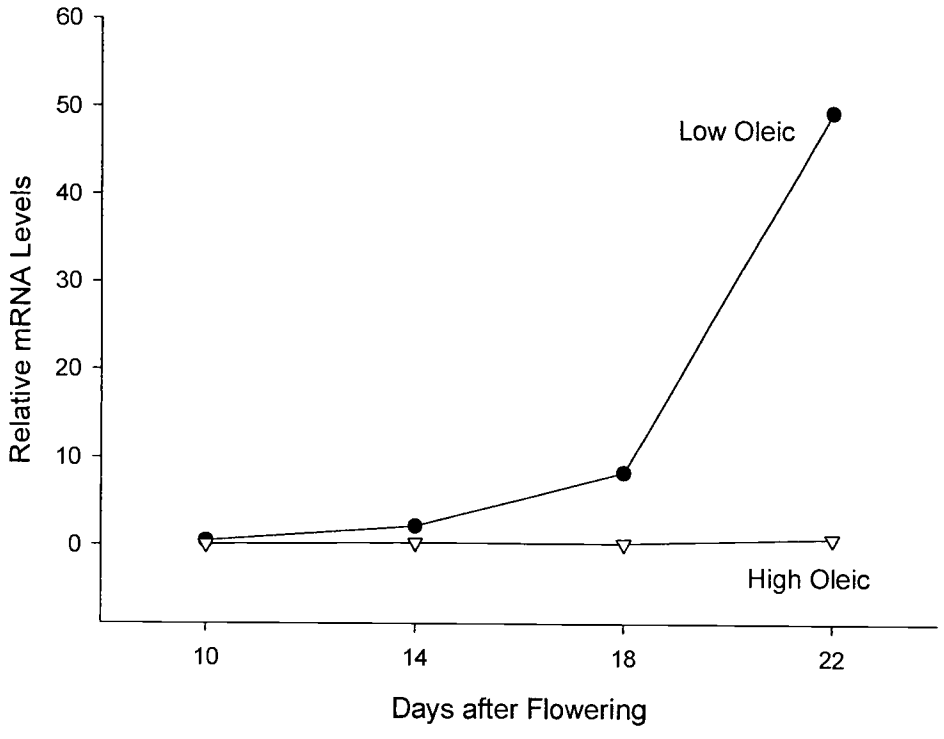
The microarray showed 1.6- to 1.7-fold higher transcript levels of *LTP* for HA341, when compared to HA89. This was confirmed with a 1.8-fold increase in *LTP* mRNA for HA341 in the real-time PCR (Table 2.4). While HA292 displayed a 2.1- to 2.6-fold increase in *LTP* transcripts, quantitative PCR detected slightly larger 3.9-fold increase. The transcript levels of LTP were 2.6-fold higher in RHA345 according to real-time PCR. This is in agreement with the 2.8-, 1.7- and 1.9-fold difference observed between RHA345 and RHA274. Even though transcript levels were almost identical among the four microarrays of PHD and PHC, with a 2.4- to 2.6- fold elevation of mRNA levels in the high oleic line PHC, real-time PCR failed to confirm these findings (Table 2.4).

Overall, the real-time PCR confirmed the observed trends in differential gene expression in the microarray experiments, thus validating the results obtained through the microarrays. The comparison of the four low and high oleic acid sunflower lines yielded only two glycerolipid biosynthetic genes (*FAD2-1* and *LTP*), which were consistently differentially expressed in the eight lines.

Transcript Levels of *FAD2-1* During Kernel Development

The relative transcript levels of *FAD2-1* were assayed in HA89 and HA341 using 2-step real-time PCR (Figure 2.1). The relative amounts of mRNA were quantified in the developing kernels at 10, 14, 18 and 22 DAF. The wildtype line HA89 with a low oleic acid phenotype shows increasing accumulation of *FAD2-1* transcripts in the seed during this time period compared to the *actin* transcript levels, which are used to standardize the data (Figure 2.1). At 10 DAF the ratio of *FAD2-1*

Figure 2.1. Relative transcript levels of *FAD2-1* in HA89 and HA341 during early kernel development.



transcripts over *actin* transcripts was 0.47 in HA89 and 0.07 in HA341. The transcript levels of *FAD2-1* increased in both lines compared to *actin* at 14 DAF to reach 2.15 in HA89 and 0.24 in HA341. The ratio further increased in HA89 to 8.31 at 18 DAF, but was slightly reduced in HA341 to 0.18. The largest gain in *FAD2-1* transcript accumulation was observed at 22 DAF, when the *FAD2-1/actin* ratio reached 48.23 in HA89. In HA341 this ratio also increased, but to a far lesser extent to reach 0.76 at 22 DAF. Thus at 22 DAF we observed a 64-fold difference in *FAD2-1* transcript levels between HA89 and HA341. The results of the real-time PCR experiment clearly show that *FAD2-1* expression is greatly reduced in the high oleic acid line compared to the low oleic acid line during the early stages of kernel development.

FAD2-1 Allele Sequencing

Hongtrakul et al. (1998a) showed that the coding sequence of *FAD2-1* was conserved between HA89 and HA341. We used primers F2 and R2, which were designed to the *FAD2-1* cDNA sequence from GenBank (U91341), to amplify the complete coding region of the gene (1,136 bp) from HA89, HA341, PHC, and PHD to both confirm the earlier results and to screen for DNA polymorphisms between PHC and PHD. The primers yielded a 1,226 bp fragment, which was subsequently cloned and sequenced. The coding sequence of *FAD2-1* was completely conserved among wildtype and mutant lines (Appendix A). Hence, differences in *FAD2-1* expression in developing kernels cannot be attributed to differences in the coding sequence of the gene.

The *FAD2* gene of *Arabidopsis thaliana* contains an 1,134 bp intron in the 5'UTR (Okuley at al., 1994). Since most housekeeping genes of the glycerolipid biosynthesis pathway tend to be fairly conserved across plant species, we considered the possibility that *FAD2-1* in sunflower might have an intron in the same position. The GenBank sequence U91341 contains the full 5' and 3' UTR of *FAD2-1*. We designed primers F1 and R1, where the forward primer F1 is on the very 5' end of the GenBank sequence and the reverse R2 is anchored in the coding sequence 44 bp

downstream of the start codon. First, we tested the primers on the cDNA of eight sunflower lines. The cDNA was derived from developing kernels at 14 DAF. The primers amplified the expected 192 bp fragment in all eight lines (Figure 2.2). We subsequently sequenced the eight fragments and they perfectly matched the *FAD2-1* 5' UTR sequence of U91341. Second, we used the same primer pair on genomic DNA of the same eight lines. The intron was difficult to amplify, but was ultimately amplified using long-distance PCR (LD-PCR). LD-PCR amplified a 1,877 bp fragment (Figure 2.2) in all eight lines. We sequenced the fragment and found that the ends matched the respective 5' UTR sequences of *FAD2-1* (GenBank AY802989, AY802990, AY802991, and AY802992) (Appendix A). Thus, we concluded that the *FAD2-1* gene of sunflower carries a 1,685 bp intron between base pairs 92-93 in the 5' UTR. In addition, HA89 produced a 192 bp fragment from genomic DNA. This fragment matched the fragment recovered from the cDNA. We have no explanation for the origin of this fragment in HA89. Southern analyses with the *FAD2-1* probe showed no indication of an additional *FAD2-1* copy in HA89 (Hongtrakul et al., 1998a, Martinez-Rivas et al., 2001, Lacombe and Bervillé, 2001).

Duplication of *FAD2-1* in High Oleic Acid Lines

Southern analyses of nine wildtype and three mutant lines revealed two bands (5.3 and 7.2 kb) in the high oleic acid lines when digested with *EcoRI* and probed with *FAD2-1*, while only producing one band in low oleic acid lines (5.3 kb) (Hongtrakul et al., 1998a, Martinez-Rivas et al., 2001, Lacombe and Bervillé, 2001). When the southern analysis was based on a digest with *HindIII*, the wildtype lines produced a single 7.3 kb band and the high oleic lines also revealed a single 14.3 kb band. Therefore, Hongtrakul et al. (1998a) concluded that *FAD2-1* is duplicated and rearranged in the high oleic lines. We designed primers F3 and R1 in order to amplify the region between the two copies of *FAD2-1* in the high oleic lines using long-distance PCR. We were able to recover a 3.3 kb fragment in the high oleic acid lines using an initial extension time of 3 minutes and 30 seconds (Figure 2.3). We subsequently cloned and sequenced the complete fragment (GenBank AY817416

Figure 2.2. Intron of *FAD2-1* in eight sunflower lines. A) PCR on genomic DNA B) RT-PCR on RNA from developing kernels at 18 DAF. A + B) First and last lane: 1kb+ DNA ladder; HA89, HA292, RHA274 and PHD are low oleic sunflower lines, HA341, HA349, RHA345 and PHC are high oleic sunflower lines.

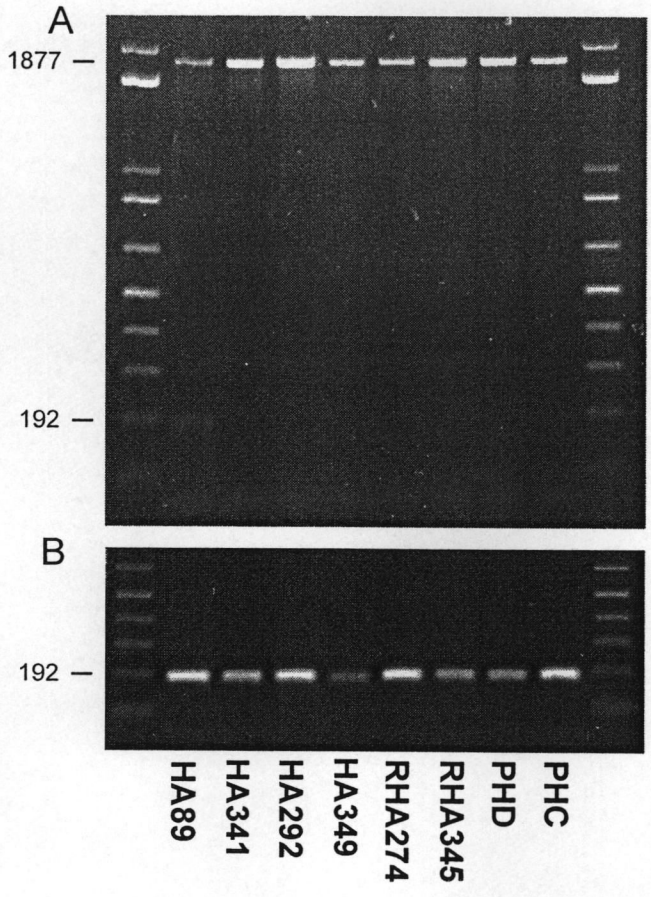
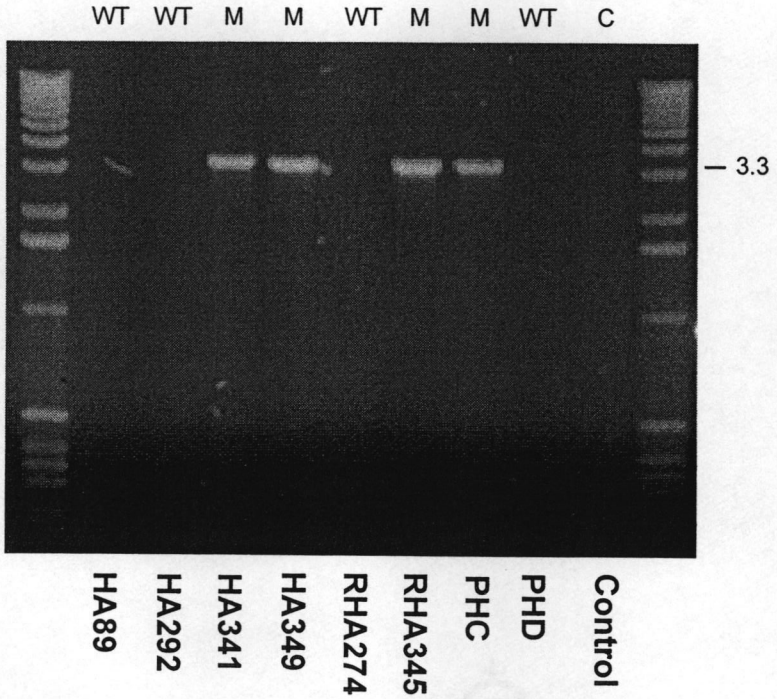


Figure 2.3. Intergenic region between the two tandemly duplicated copies of *FAD2-1* in the high oleic sunflower lines. First and last lane: 1kb DNA extension ladder, WT: wildtype low oleic lines, M: mutant high oleic lines, C: no template control of PCR using primers F3-R1.



and AY817417). When the sequence was blasted, the best match found in the GenBank database was the sunflower *FAD2-1* sequence (U91341). The 5' end of the 3.3 kb fragment was identical to the 3' UTR. In addition, the 3' end of the 3.3 kb fragment matched the last 32 bp of the 5' UTR and the first 71 bp of the coding sequence, including the reverse primer site; thus, we concluded that two copies of *FAD2-1* were in a tandem array separated by a 3.1 kb intergenic region (Figure 2.4). This was confirmed by testing for inverted repeats using different primer combinations with similar extension times (data not shown).

The next step was to further characterize the duplication in the high oleic lines. First, we aligned the 3.3 kb fragment (intergenic region) with the complete genomic sequence of *FAD2-1* (intron and exon). The alignment revealed complete sequence identity starting at bp 1,441 of the intron until bp 71 of the coding region (the 3' end of the 3.3 kb fragment). This indicated that the novel joint for the duplication might be in the intron of *FAD2-1*. Our earlier experiment however, showed a complete intron upstream of *FAD2-1* in all lines (Figure 2.2). Thus, primer F5 was designed to target the area of the intron of *FAD2-1* that did not align with the 3.3 kb fragment. We tested primer F5 with four different reverse primers (R1, R2, R3 and R4) to determine the location of the complete intron with regard to the duplication (Figure 2.5). Primer R1 is the same reverse primer that was used to amplify the intron in the first place. F5-R1 yielded a single band of 733 bp in all eight lines tested, thus confirming the presence of the complete intron in high oleic lines (Figure 2.5). Primer R2 was located 575 bp downstream of the start codon and produced a single band of 1,339 bp in combination with F5 in all eight lines. This indicated that at least a partial *FAD2-1* gene is downstream of the intron. The 1,862 bp fragment amplified by primers F5-R3 in the 3' UTR of *FAD2-1* confirmed that low and high oleic lines carry a complete copy of *FAD2-1*. Primer R4 was 542 bp downstream of the stop codon in the 3.3 kb fragment. Primers F5-R4 amplified a single band of 2,342 bp in all eight lines (Figure 2.5). This amplicon links the complete copy of *FAD2-1* carrying the intron to the region of the gene duplication. Thus, we can conclude that the 3.3 kb fragment is downstream of the original *FAD2-1* copy present in wildtype and mutant lines (Figure 2.4).

Figure 2.4. Schematic representation of the *FAD2-1* locus in the wildtype low oleic lines and of the *FAD2-1* duplication in the mutant high oleic lines.

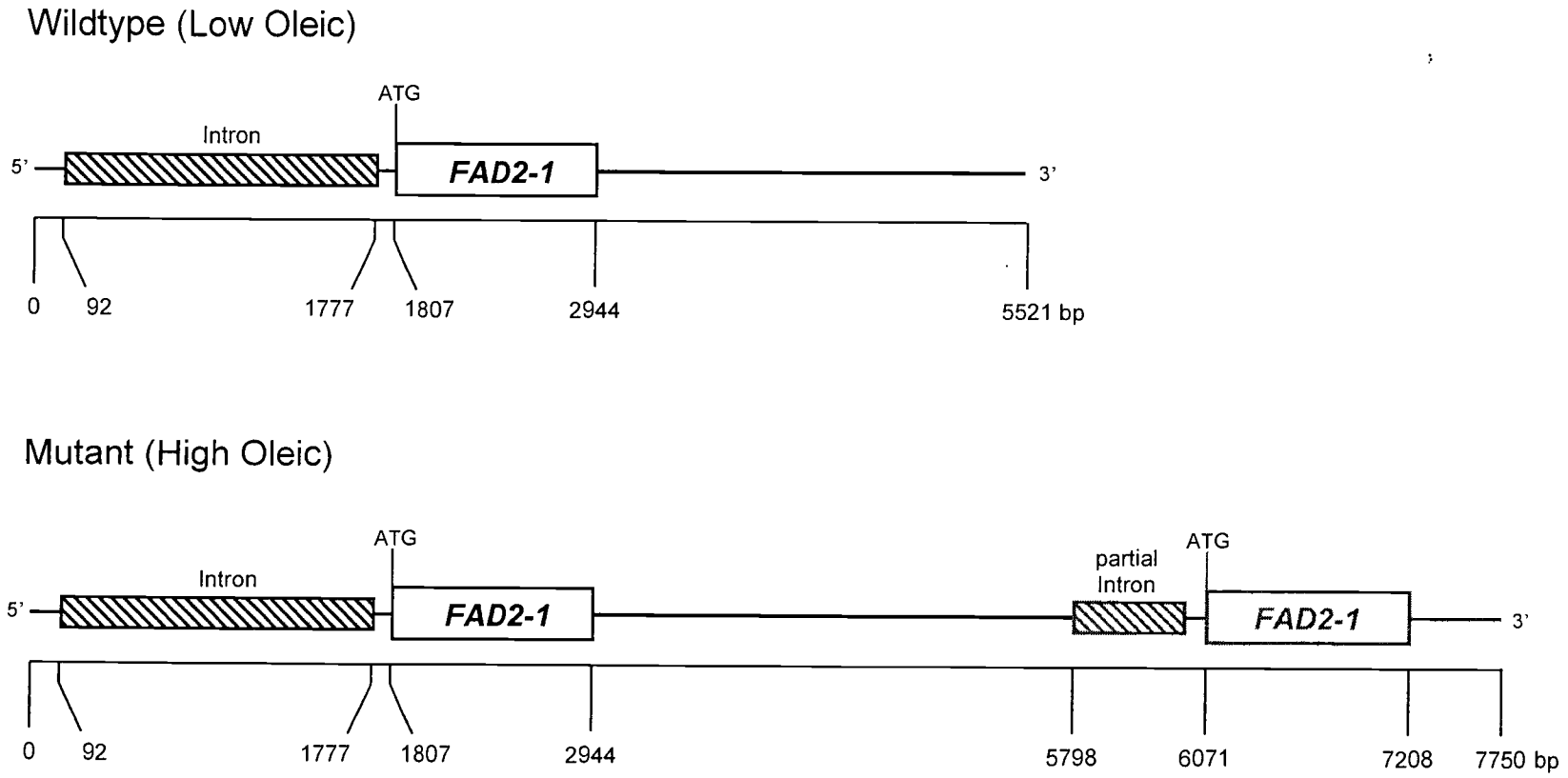
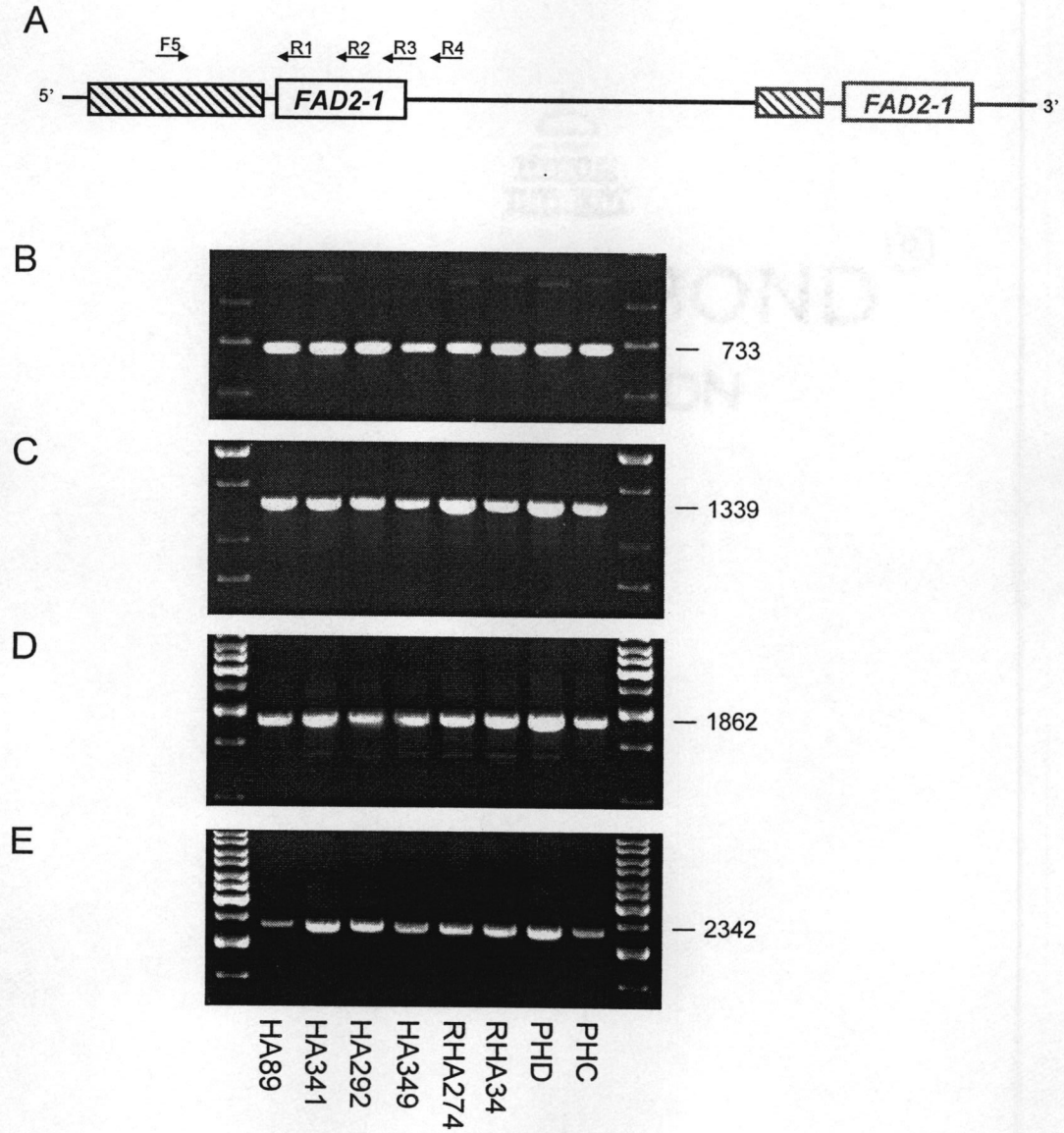


Figure 2.5. The first copy of the *FAD2-1* gene is complete in the mutant lines and corresponds to the wildtype allele of *FAD2-1*. A) Schematic diagram depicting the locations of the primers. B through E) First and last lanes: Generuler 1kb DNA ladder. B) The primers F5-R1 amplify the expected 733 bp band in all eight lines. C) The primers F5-R2 recover the same 1,339 bp fragment in all lines. D) The primers F5-R3 amplify an 1,862 bp fragment in all lines showing that the *FAD2-1* coding region is complete in the 1st copy of the mutant lines. E) The primers F5-R4 yield a 2,342 bp fragment showing that the full gene including 3' UTR is present in the high oleic lines.



Next, we determined what constitutes the bulk of the DNA present in the 3.1 kb intergenic region. From the initial alignment we knew that the 3' end of the sequence corresponded to the intron. The amplicon of F5-R4 indicated that at least part of the intergenic region was downstream of the *FAD2-1* wildtype copy. We designed two additional reverse primers located in the 3.1 kb intergenic region. R5 was 1,960 bp and R6 was 2,577 bp downstream of the stop codon. We used the forward primer F3, which was used to recover the 3.3 kb fragment, in conjunction with R4, R5 and R6. F3 and R4 yielded the expected 600 bp fragment in all lines (Figure 2.6). We confirmed the identity of the PCR fragment by sequencing. Primers F3 and R5 amplified a 2,018 bp fragment, not only as expected in all four high oleic lines, but also in all four low oleic lines (Figure 2.6). Thus, we concluded that most of the intergenic region is a sequence that is downstream of *FAD2-1* in the wildtype. This assumption was further strengthened by the fact that F3 and R6 recovered a 2,636 bp fragment in seven of the eight lines tested (Figure 2.6). Only HA292, the only confectionary sunflower line on the panel, did not produce a band. This is most likely due to sequence divergence between confectionary and oilseed lines.

The initial alignment of the 3.3 kb fragment indicated a tandem duplication of *FAD2-1*. The 3' end of the fragment aligned only with the latter part of the intron sequence and completely with the first 71 bp of the coding region. In order to test whether this point of sequence divergence constituted the novel joint, we designed the primer F4 just upstream of the putative novel joint and used the same reverse primer R1, which was used to recover the intergenic region. The primers amplified a single band of 650 bp in high oleic lines as expected (Figure 2.7), but did not yield any product in low oleic lines. In addition, we tested the forward primer F4 with the reverse primers R2, R3, and R4, which are located across the *FAD2-1* gene (Figure 2.7, Table 2.6), in order to determine if there is a complete copy of *FAD2-1* downstream of the novel joint. If the tandem copy of *FAD2-1* is complete, we expected PCR amplification products in all four high oleic lines, but none in the four low oleic lines. We observed the expected bands in all four high oleic lines with all three reverse primers (Figure 2.7). Primers F4-R2 yielded a 1,256 bp confirming at least a partial

Figure 2.6. The intergenic region between the two tandemly duplicated copies of *FAD2-1* is mainly comprised of the genomic sequence 3' of the gene in the wildtype lines. A) Schematic drawing depicting the locations of the primers. B through D) First and last lanes: Generuler 1kb DNA ladder. B) The primers F3-R4 amplified the expected 600bp band in all lines. C) The primers F3-R5 recover a 2,018 bp fragment in all lines D) The primers F3-R6 amplify the expected 2,636 bp fragment in all lines except for HA 292.

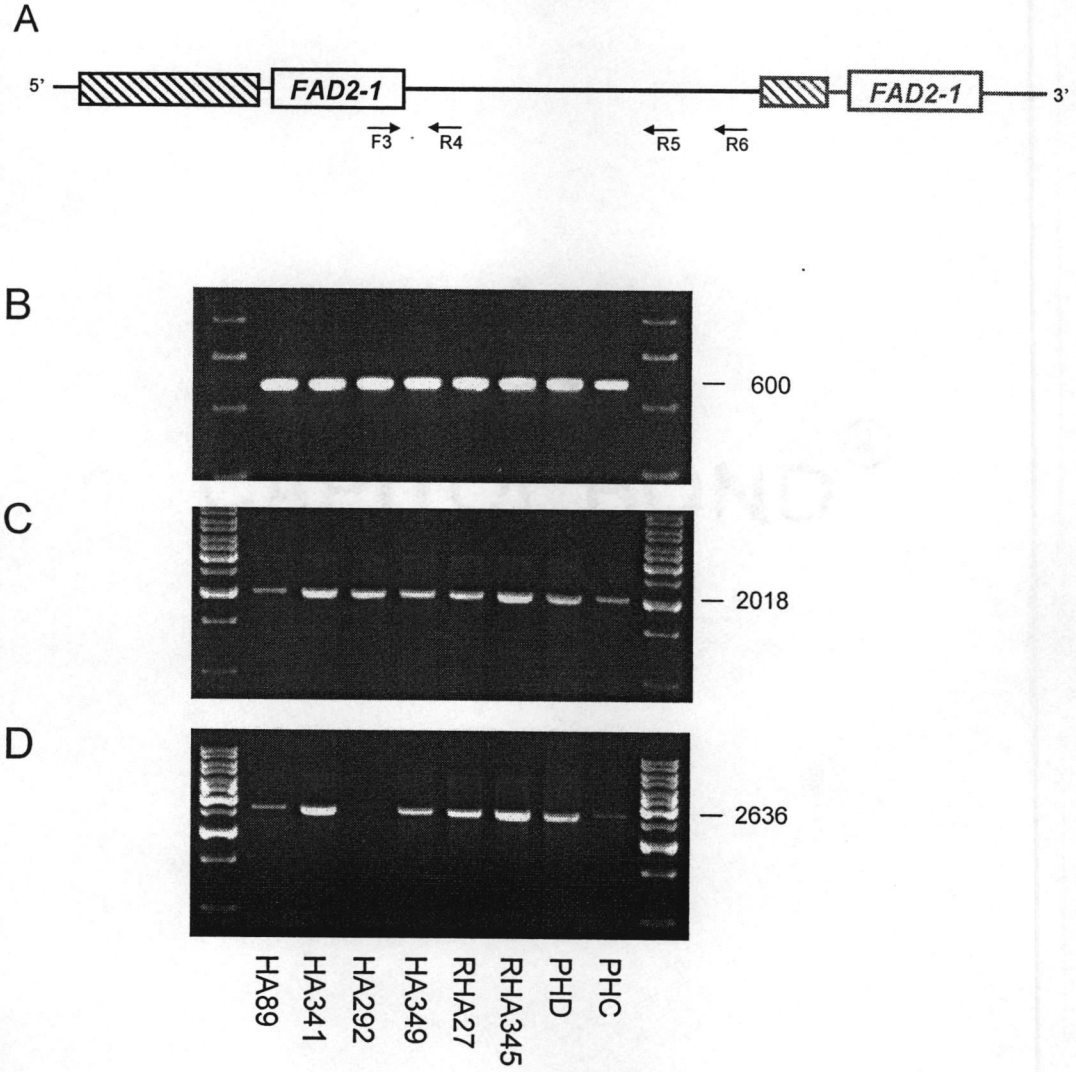


Figure 2.7. The duplicated copy of *FAD2-1* is complete and in tandem array. A) Schematic diagram depicting the location of the primers. B through E) First and last lane: 10 1kb+ DNA ladder, WT: wildtype low oleic lines, M: mutant high oleic lines. B) Only the high oleic lines yield a discrete band of 650 bp showing that the duplication is only present in these lines (F4-R1). C) The 1,256 bp band shows at least a partial duplication of *FAD2-1* in tandem array (F4-R2). D) Only the high oleic lines yield a discrete band of 1,779 bp showing that *FAD2-1* coding region is completely duplicated (F4-R3). E) The same four lines amplify a discrete 2,259 bp band indicating that the full gene including 3' UTR is duplicated (F4-R4).

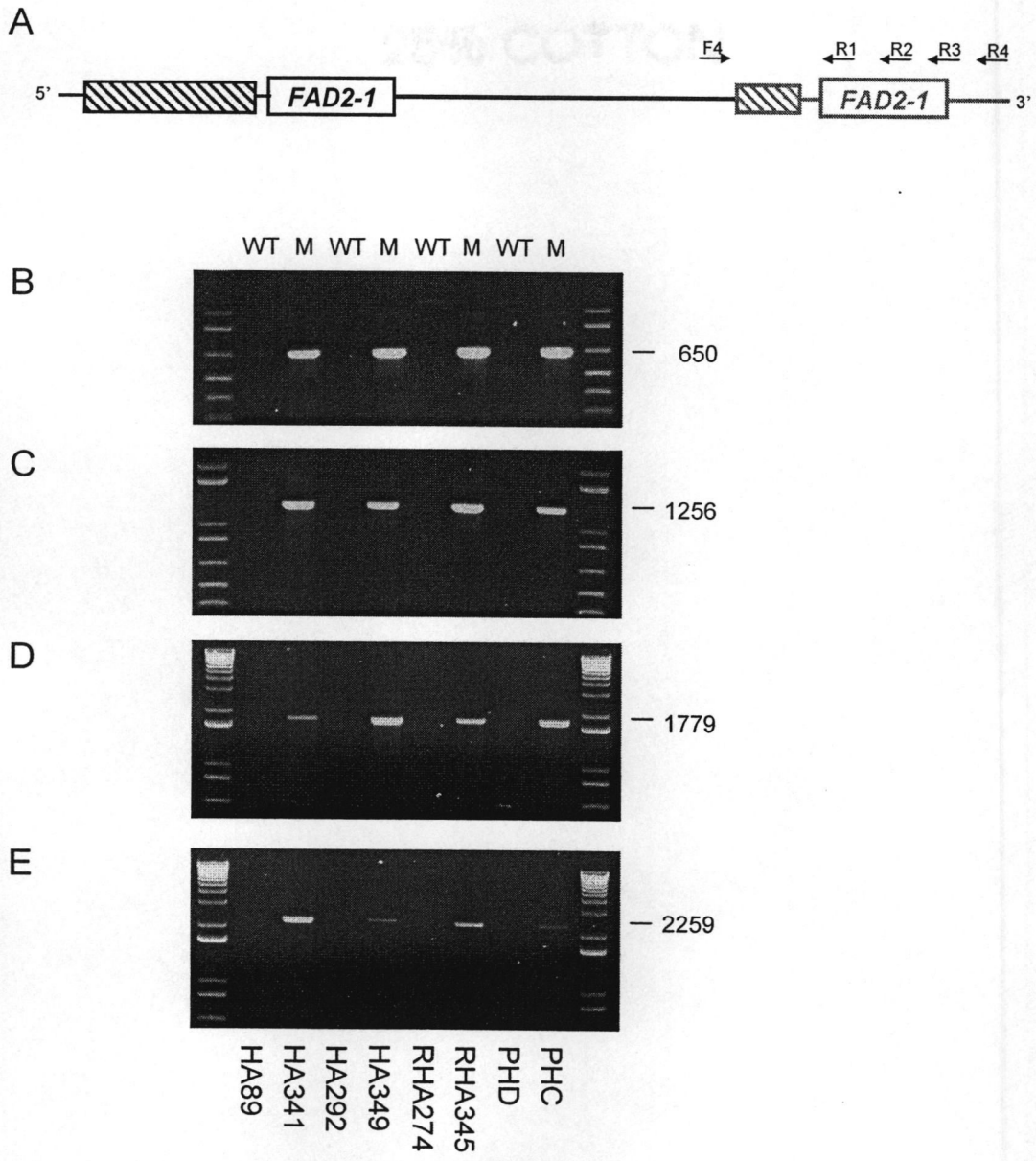


Table 2.6. Oligonucleotide primer names, sequences and their respective locations used for the analysis of the *FAD2-1* locus in the low and high oleic sunflower lines.

Name	Sequence	Location (bp)
F1	5'-CTCGTGCCGCGTCTGTCTCACTAG-3'	bp 1
F2	5'-GAAAAGTCTGGTCAAACAGTCAACAT-3'	bp 1781
F3	5'-GGAGCAAGATGATGAAGGGAAAGGAG-3'	bp 2912
F4	5'-GTAACGTCTGCGCGCTTGCAGACATCA-3'	bp 5521
F5	5'-GGGCAAAAACGCATTATGTC-3'	bp 1149
F6	5'-GATCACAGAGGAAGCCTTACCTACC-3'	-
F7	5'-AGGGTTGGTTTGGGTGATTT-3'	bp 2528
F8	5'-CAATGGGTCTTGTCCTGGTT-3'	bp 2983 and 7247
F9	5'-CACAGCTCGTTTTGACCTGA-3'	bp 3470 and 7734
R1	5'-GGTTTTGCATGAGGGACTCGATCGAGTG-3'	bp 1850 and 6114
R2	5'-CGAGAACCAGGACAACAGCCATTGTC-3'	bp 2381 and 6646
R3	5'-CCGATGTCGGACATGACTATC-3'	bp 3006 and 7271
R4	5'-TCAGGTCAAAACGAGCTGTG-3'	bp 3485 and 7750
R5	5'-GTAGTTTTGGAAAGCTAGAGACC-3'	bp 4804
R6	5'-CTGATGTCTGCAAGCGCGCAGACGTTA-3'	bp 5522
R7	5'-GCCCCGAAAAAGGTAGAGAT-3'	-
R8	5'-CACCTCCTTTCCCTTCATCA-3'	bp 2913
R9	5'-GTTTTCCGTCATTGGTTATGG-3'	bp 4017

duplication of *FAD2-1*. The discrete 1,779 bp band amplified by the primers F4-R3 indicated the presence of a complete copy of the gene downstream of the novel joint. Finally, the 2,259 bp fragment recovered by the primer combination F4-R4 indicated that the complete *FAD2-1* gene, including the 3' UTR, was tandemly duplicated in high oleic lines (Figure 2.7).

Origin of the Duplication in Pervenets

The next step in our investigation was to examine the cultivar Pervenets, which was the original source for the high oleic acid phenotype. Open-pollinated cultivar Pervenets was developed by selecting for high oleic acid among progeny in an DMS mutagenesis experiment (Soldatov, 1976). The open-pollinated cultivar VNIIMK8931 constitutes the M_0 generation of this experiment. We performed half-seed analysis using GC on 14 VNIIMK8931 individuals and six Pervenets individuals in order to quantify their fatty acid profiles. The individuals of VNIIMK8931 displayed oleic acid content in the seed ranging from 310.6 g kg^{-1} to 560.5 g kg^{-1} and linoleic acid ranging from 574.4 g kg^{-1} and 327.6 g kg^{-1} (Figure 2.8). The Pervenets individuals contained between 879.9 g kg^{-1} and 896.5 g kg^{-1} oleic and 12.2 g kg^{-1} to 25.2 g kg^{-1} linoleic acid. The remaining part of the seed was planted in the greenhouse. DNA was extracted from young leaves of each individual. The primers F4-R1 (a dominant marker for the presence of the *FAD2-1* duplication), which yielded a 650 bp fragment in the high oleic lines previously, were tested on the 20 individuals. The 14 VNIIMK8931 individuals produced no bands, while all six Pervenets samples amplified the expected band (Figure 2.9). Primers F6-R7 of the *FAB2-2* gene were used as a control for proper PCR amplification in all lines. This indicates that the *FAD2-1* duplication is not present in the M_0 material of the mutagenesis experiment, but is present in high oleic Pervenets individuals.

Regulation of *FAD2-1* by RNAi

The duplication of *FAD2-1* in the high oleic lines does not result in elevated transcript levels of this gene (Table micro). To the contrary, lines carrying the tandem

Figure 2.8. Oleic and linoleic acid concentrations (g kg^{-1}) of individuals within VNIIMK8931 and Pervenets.

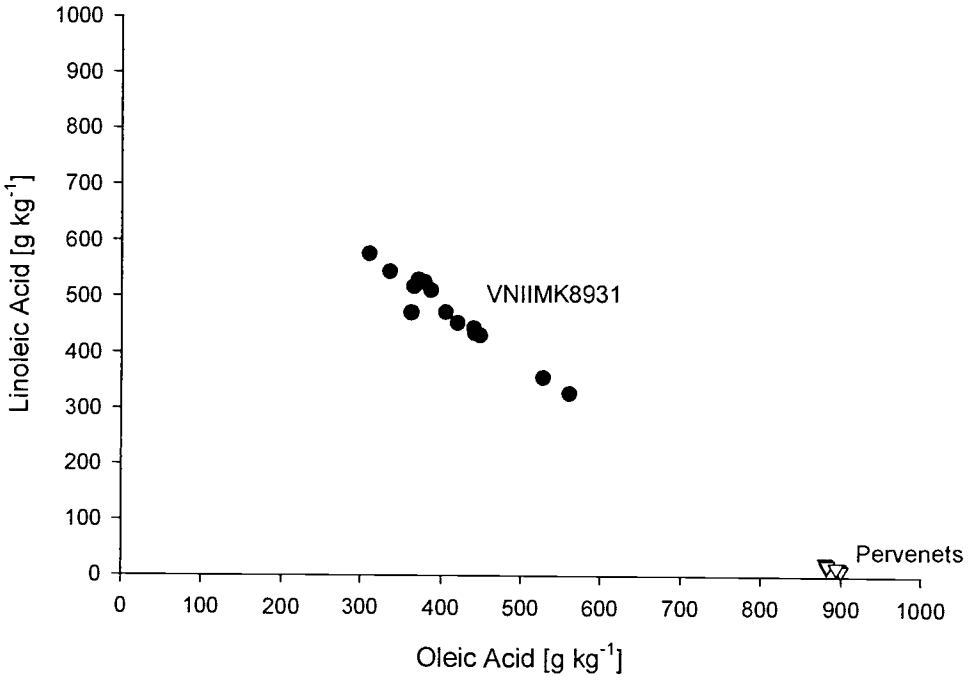
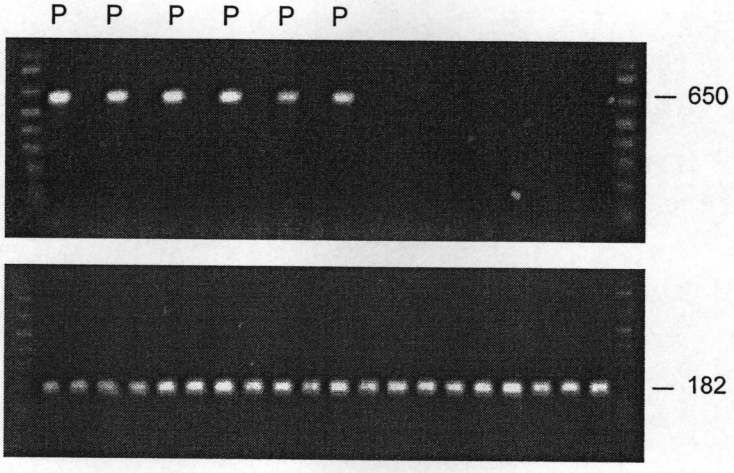


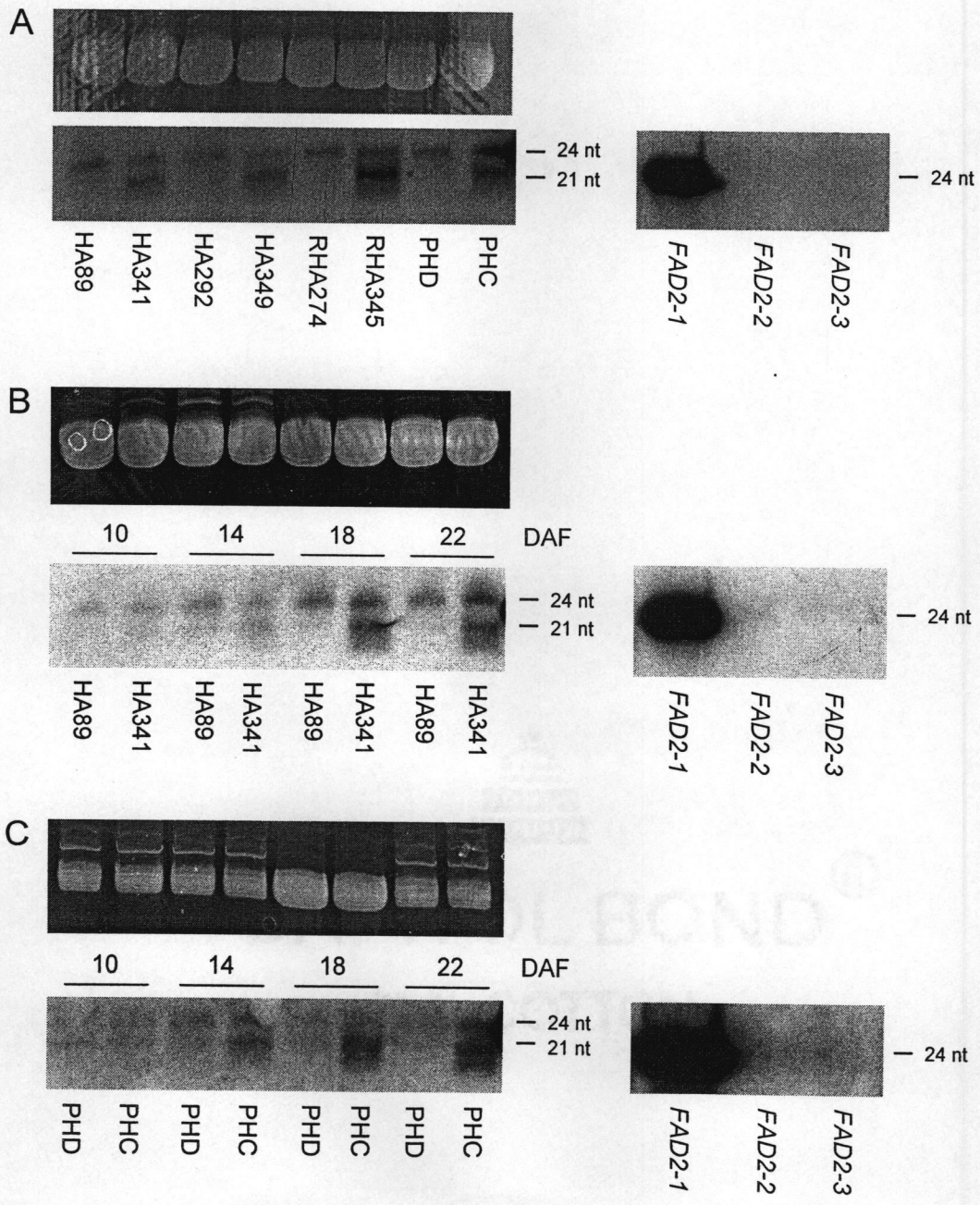
Figure 2.9. The duplication of *FAD2-1* is only present in the cultivar Pervenets (P) and not present in Soldatov's original starting material VNIIMK 8931 (M_0). The upper panel shows that the 650 bp of the dominant marker (F4-R1) only appears in the Pervenet individuals. The lower panel shows control PCR amplification of *FAB2-2* partial coding sequence (F6-R7) on all DNA samples.



duplication display drastically reduced transcript levels in developing kernels (Figure 2.1). We hypothesized that the reduction of *FAD2-1* mRNA was caused by the RNAi silencing pathway. This hypothesis was tested by a series of northern analyses. The low molecular weight RNA of eight sunflower lines extracted from developing kernels at 14 DAF was probed with a radioactively labeled probe of *FAD2-1*. The full length probe of the *FAD2-1* coding sequence proved difficult in the northern blot assays due to its length (1226 bp). Thus, we designed three shorter probes of about 400 bp spanning the complete coding sequence. The 385 bp probe (primers F7-R8) designed to the 3' part of the *FAD2-1* coding region gave the best hybridization results and was subsequently used in northern blot assays (Table 2.6). The probe hybridized to small RNAs of 24 nt in all eight lines. But only the four high oleic lines HA341, HA349, RHA345 and PHC displayed a distinct, 21 nt size band (Figure 2.10) when probed with *FAD2-1*. In addition, the probe hybridized to the *FAD2-1* 24 nt oligoribonucleotide control as expected, but did not hybridize to the *FAD2-2* nor the *FAD2-3* oligoribonucleotide (Figure 2.10).

We used the same probe to test for the presence of short-interfering RNAs (siRNAs) during early kernel development. The low molecular weight RNA of HA89, HA341, PHC, and PHD was extracted at 10, 14, 18, and 22 DAF. When hybridized with *FAD2-1*, all lines showed the previously observed 24 nt band. The intensity of the band increased as DAF increased in HA89 and HA341 (Figure 2.10). HA341 shows no 21 nt band at 10 DAF, but displays a very faint 21 nt band at 14 DAF. The intensity of the band increased at 18 and 22 DAF. There was no 21 nt band in the HA89 small RNAs. The same results were observed in PHC and PHD. The high oleic line PHC revealed a 21 nt band at 14, 18, and 22 DAF, while this band is not present at 10 DAF. The low oleic line PHD showed no bands of this size at any of the time points assayed. The results of the northern blot analyses clearly showed the presence of 21 nt short-interfering RNAs in the high oleic acid sunflower lines, thus indicating that the expression of *FAD2-1* is regulated by the RNAi machinery in these lines.

Figure 2.10. The *FAD2*-1 expression in the developing seed is regulated by RNAi. A through C) Top panel: Ethidium bromide stained t-RNAs and 5S RNAs prior to the transfer as an equal loading control. Lower panel: Northern blot of small RNAs. Right panel: 24nt oligonucleotide RNAs as hybridization control for *FAD2-1, FAD2-2* and *FAD2*-3. A) Small RNAs of eight sunflower lines at 14 DAF probed with *FAD2-1*. B) Small RNAs of HA89 and HA341 at 10, 14, 18 and 22 DAF probed with *FAD2-1*. C) Small RNAs of PHC and PHD at 10, 14, 18 and 22 DAF probed with *FAD2-1*.



Bi-directional Transcripts Induce Silencing

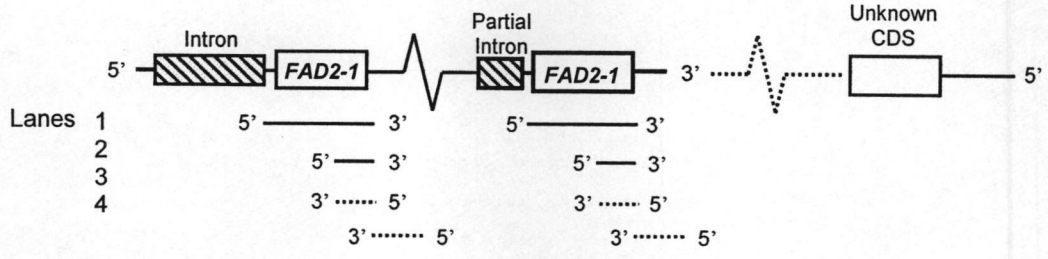
The presence of the *FAD2-1* specific siRNAs offered an explanation for reduced *FAD2-1* expression, but did not shed light on what triggers *FAD2-1* silencing. The fact that the best hybridization was observed with the probe designed to the 3' region of *FAD2-1*, led us to hypothesize that the duplication led to the disruption of a downstream gene. The elimination of the transcriptional termination signal of a gene in opposite orientation would lead to an extension of the mRNA transcript. Upon reaching the 3' end of *FAD2-1* the extended transcript would be complementary to the *FAD2-1* transcript, and thus form double-stranded RNA, which in turn would serve as a trigger for the siRNA machinery. We used a strand specific reverse transcription assay to test this hypothesis. The reverse transcription reaction with gene specific reverse primer R3 yielded the expected 1226 bp *FAD2-1* transcript in low and high oleic lines in the subsequent PCR (Figure 2.11 lane 1). In addition, reverse primer R8 yielded a transcript fragment complementary to the northern blot probe in conjunction with PCR primer F7 in the lines tested (Figure 2.11 Lane2). When we used the forward primer F7 in the strand specific reverse transcription reaction, the subsequent PCR with primer R8 only yielded the 386 bp fragment in the high oleic lines HA341 and PHC (Figure 2.11 Lane 3). To further strengthen the evidence we used forward primer F8 in the reverse transcription. This primer is located 39 bp downstream of the stop codon in the 3' UTR of *FAD2-1*. The following PCR with reverse primer R4 recovered a single fragment of the expected 503 bp in the two high oleic lines only (Lane 4). The control reaction with forward primer F7, which is devoid of the reverse transcriptase enzyme, showed no sign of DNA contamination (Lane C). This clearly shows that bi-directional transcripts are present in the mutant lines, which are thought to induce the silencing mechanism.

Methylation of *FAD2-1*

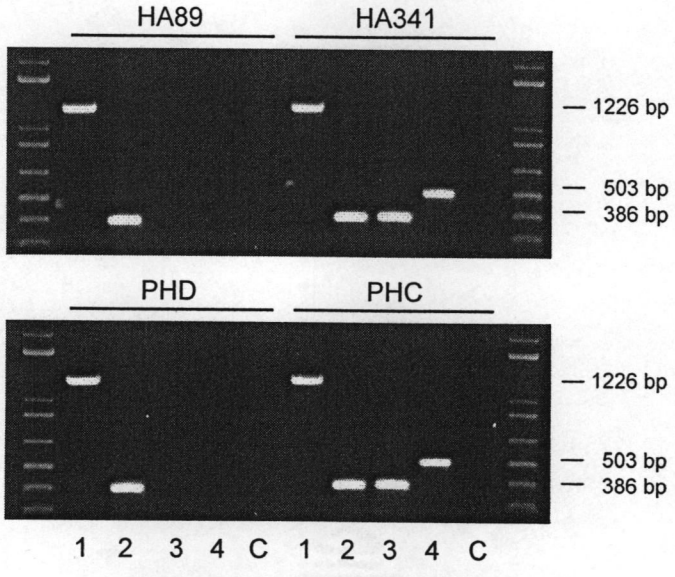
Gene silencing through the RNAi machinery is often accompanied by methylation of the target DNA or its promoter region. We assessed the methylation status of *FAD2-1* and its 3' region in low and high oleic lines using Mcr-PCR. We

Figure bi-di. Strand specific reverse transcription (RT) assay on a set of four sunflower lines. A) Schematic diagram displaying the location of the obtained fragments B) RT-PCR assay: Lane 1: Gene specific primer R3. Lane 2: Primer R8. Lane 3: Primer F7. Lane 4: Primer F8. Lane C: Primer F7 without reverse transcriptase. Top panel: Low oleic line HA89 and high oleic line HA341. Lower panel: Low oleic line PHD and high oleic line PHC.

A



B



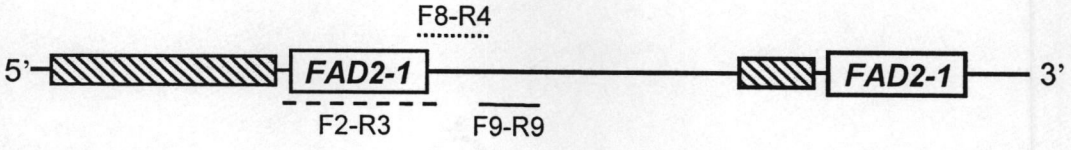
used primers F2-R3, covering the complete coding region of *FAD2-1*, as well as primers F8-R4 and F9-R9, all located 3' of the gene, for PCR on genomic DNA. The primers produced the expected size fragments in all lines tested (Figure 2.12 first lane). When genomic DNA was previously digested with the endonuclease McrBC, which cleaves DNA containing methylcytosine on one or both strands, primers F2-R3 failed to produce a fragment in all four lines. The control procedure involves two steps. First, the genomic DNA was incubated with the methylase M.SssI, which methylates all C_pG cytosine residues in double stranded DNA. The following PCR recovered the fragment in all four lines. Second, an aliquot of the methylated DNA was subsequently digested with McrBC. As expected, PCR performed on the processed DNA yielded no amplicon across all lines (Figure 2.12). These results indicated that the *FAD2-1* coding region was methylated in both low and high oleic lines.

PCR with primers F8-R4 yielded a 503 bp fragment from genomic DNA. After the digest with McrBC, a faint band of the expected size was observed in HA341, PHD, and PHC, but not in HA89 (Figure 2.12). The PCR on the genomic DNA after control incubation with M.SssI produced the expected band, while the PCR following the digest with McrBC recovered no products as expected (Figure 2.12). Thus methylation status differs between HA89 and HA341, but not between PHC and PHD, but was independent of the oleic acid phenotype.

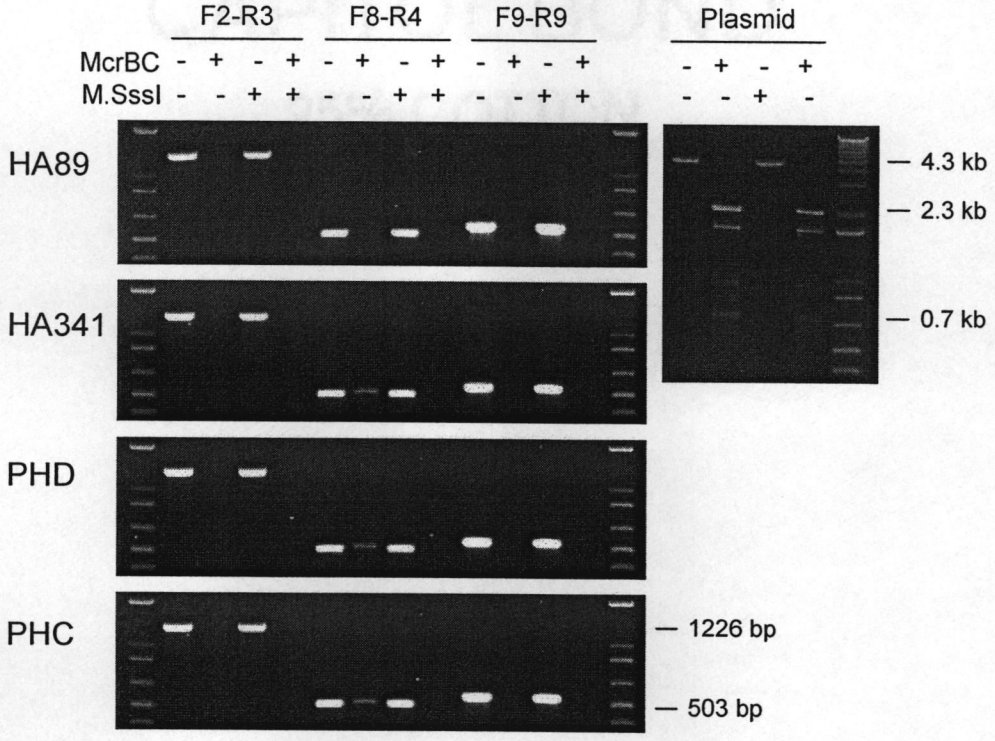
Primers F9-R9 produced a single 548 bp fragment from the genomic DNA of the four lines. The digest with McrBC resulted in a loss of the fragment in the subsequent PCR in all lines (Figure 2.12). PCR on the control DNA, incubated with M.SssI and digested with McrBC, recovered no fragment as well (Figure 2.12). Thus the 3' region of *FAD2-1* is methylated in HA89, HA341, PHC and PHD, but is again not associated with the high oleic phenotype. The control digests of the 4.3 kb plasmid confirmed proper endonuclease activity of McrBc by producing the expected fragments ranging from 700 bp to 2.3 kb. Overall the results of the Mcr-PCR indicate that *FAD2-1* and its 3' bordering sequence is at least partially methylated in both wild type and mutant lines. However, the assay is not sensitive to any putative differences in the amount of methylation between the wildtype and mutant lines.

Figure 2.12. Methylation of *FAD2-1* is assessed using Mcr-PCR on four sunflower lines. A) Schematic diagram of the assessed DNA fragments. B) Genomic DNA of low oleic lines HA89 and PHD and high oleic lines HA341 and PHC were digested with McrBC with and without prior incubation with M.SssI. Subsequently PCR was performed with primer pairs F2-R3, F8-R4 and F9-R9. McrBC endonuclease activity was monitored with a plasmid control.

A



B



DISCUSSION

We showed that transcript levels differ not only significantly for *FAD2-1* at 18 DAF as seen in the microarrays, but that *FAD2-1* mRNAs accumulate in the developing kernel of the HA89 between 10 and 22 DAF compared to high oleic line HA341 (Figure 2.1). Hongtrakul et al. (1998a) and Martinez-Rivas et al. (2001) showed that *FAD2-1* expression is seed specific and observed similar differences in the accumulation of *FAD2-1* transcripts between low and high oleic lines. The substantial increase in *FAD2-1* transcripts occurs at the time of rapid oil production in the developing kernel. The percent oil in the sunflower seed increases from about 4% at 7 DAF to about 40% at 20 DAF (Ganbhir and Anand, 1979). The importance of the FAD2 desaturation activity in this process is demonstrated by the fact that oleic acid is the major fatty acid contributing to lipid accumulation in the developing kernel during this time period in the high oleic lines (Garcés et al., 1989). Jung et al. (2000) isolated two distinct cDNA sequences from developing peanut seeds. While both *FAD2* genes are expressed in normal oleate phenotype lines (36%-67% oleic acid), one of the *FAD2* genes shows severely reduced expression in the high oleate phenotype (80%). In addition, reduced transcript levels of the *FAD2* gene have been associated with high oleic phenotypes of cotton and Arabidopsis (Liu et al., 2002 and Stoutjesdijk et al., 2002).

Even though the high oleic lines carry an additional copy of *FAD2-1*, the duplication is associated with a drastic reduction in *FAD2-1* transcription. Silencing of genes in the glycerolipid biosynthetic pathway has been successfully used to alter the seed oil composition in plants. Introduction of a construct carrying an inverted repeat of the *FAD2* gene in *A. thaliana* resulted in efficient and stable silencing of its endogenous *FAD2* gene (Stoutjesdijk et al., 2002). Transcription of the inverted repeat of *FAD2* leads to the formation of double-stranded RNA, which serves as a trigger for RNA induced silencing. The oleic acid content in the Arabidopsis seeds increased from 17% in the wildtype to about 55% in the transformed plants. Liu et al. (2002)

obtained high oleic cottonseed lines through the introduction of constructs carrying an inverted repeat of *FAD2*. The transformation resulted in a substantial down-regulation of *FAD2* expression in the seed and an increase in oleic acid in the seed oil from 13.2 % in the wildtype to 78.2 % in the transformed line. These numbers are similar to the shifts observed in the oleic acid content of the four low and high oleic sunflower lines (Table 2.1).

We showed that small RNAs of 21 nt, specific to *FAD2-1*, are present in the developing kernels at 14 DAF of the four high oleic lines, but are absent in the low oleic lines (Figure 2.10). The presence of gene specific 21 nt RNAs is a hallmark of RNA induced silencing, also referred to as RNA interference (RNAi) (for reviews see Zamore, 2002, Hannon, 2002, and Matzke and Matzke, 2004), demonstrating that the expression of *FAD2-1* is regulated through RNAi in the developing kernels of the high oleic lines. The fact, that the abundance of *FAD2-1* specific 21 nt RNAs increases at the same time in the developing kernels of high oleic lines, while transcript levels of *FAD2-1* increase in the low oleic lines, further strengthens this point. In addition, we demonstrated that bi-directional transcription of at least part of *FAD2-1* occurs in the developing kernels of the high oleic lines, resulting in dsRNA complementary to *FAD2-1*. Double-stranded RNA induced silencing in plants was first observed by Waterhouse et al. (1998), and dsRNA has been shown to be a potent inducer of gene specific silencing (for reviews see Hannon, 2002, and Matzke and Matzke, 2004). In addition, we observed 24 nt small RNAs in all surveyed genotypes. At this point we can only speculate as to their nature or function. But the fact that they are ubiquitous in all lines suggests that these small RNAs are not part of the RNAi regulation of *FAD2-1* expression.

RNA induced silencing is often accompanied by methylation of the target gene and/or its promoter region (Wassenegger et al. 1994, Mette et al., 2000). The results of our McrPCR assays indicated that *FAD2-1* was methylated in wildtype and mutant lines, based on the DNA extracted from young leaves. The fact, that the *FAD2-1* coding sequence is at least partially methylated in the low oleic lines, is in contradiction to its high expression observed in the developing seed. Rabinowicz et al.

surveyed the methylation status of genes and transposons in genomic DNA from maize leaves and mouse spleen cells (2003). In their report, 95% of all surveyed maize exons were unmethylated. However, tissue specific methylation has been shown for genes coding for two seed storage proteins in maize. The genes of the zeins and glutelins storage proteins exhibited a uniform methylation pattern in different somatic tissues, while the genes were characterized by an extensive decline in methylation in endosperm cells (Bianchi and Viotti, 1988). Since *FAD2-1* displayed strictly seed specific expression (Hongtrakul et al., 1998a), methylation differences might exist in the developing kernel for *FAD2-1* between wildtype and mutant lines, which were absent in our analysis of the *FAD2-1* methylation status in leaf DNA. Thus, further experiments with DNA from different tissues will be necessary to accurately assess the degree of *FAD2-1* methylation in low and high oleic lines.

Hongtrakul et al. (1998a) had previously reported no allelic differences between the HA89 and the HA341 *FAD2-1* coding sequence, but had shown that *FAD2-1* was duplicated and rearranged in the high oleic lines. The lack of allelic variation in *FAD2-1* was also observed in the coding sequences of PHC and PHD. We showed that the sunflower *FAD2-1* gene carries a 1685 bp intron in the 5'UTR, which was monomorphic in HA89, HA341, PHC, and PHD (Appendix A). High oleic lines carry at least one additional copy of *FAD2-1*, 3127 bp downstream of the original wildtype copy. The DNA between the two copies of *FAD2-1* was mainly comprised of wildtype downstream sequence, while the insertion of the duplicated copy resulted in a truncation of the large intron (Figure 2.4). The duplication was also present in the cultivar Pervenets, the original source of the high oleic phenotype. Individuals with a high oleic acid content were selected among the progeny of the cultivar VNIIMK8931, which had previously been mutagenized with dimethyl sulfate (DMS), resulting in the cultivar Pervenets (Soldatov 1976). We have confirmed that the duplication is absent in the starting material of the experiment, the open pollinated cultivar VNIIMK8931 (Figure 2.9), which had been indicated by Lacombe and Bervillé (2001). DMS is an alkylating agent, which has been shown to primarily induce single nucleotide mutations (Hoffmann, 1980). DMS shares the same alkylating properties as EMS,

inducing G/C to A/T single nucleotide transitions, which are not sufficient to explain the observed duplication (Ashburner 1990, Greene at al., 2003). Hoffmann (1980) reported that alkylation of guanine weakens the glycosidic linkage in the nucleotide, which can result in the rare loss of the purine base, which in turn can lead to a breakage in the double-stranded DNA. Patel et al. (2004) isolated two mutations in the peanut (*Arachis hypogaea* L.) *FAD2* gene (*ahFAD2B*) from independent high oleic lines, which were obtained by diethyl sulfate mutagenesis. Both mutations consisted of an insertion of a 205 bp miniature inverted-repeat transposable element (MITE) at different locations in the coding region (Patel et al., 2004). Diethyl sulfate (DES) belongs to the same class of mutagenic alkylating agents as DMS and EMS (Hoffmann, 1980). These findings indicate that the chemical mutagenesis with DMS or DES has an effect on DNA stability, which is not very well understood. The DMS mutagenesis experiment was performed on an open pollinated cultivar and not on an inbred or hybrid line, allowing for the possibility that the selected individual with the elevated oleic acid content acquired the gene duplication prior to the experiment. Therefore the origin of the gene duplication remains obscure.

The microarray experiments revealed only one gene (*LTP*) in addition to *FAD2-1*, which differed consistently in its transcript levels across the four comparisons. While *LTP* transcription was higher in the three high oleic lines HA341, RHA345 and PHC compared to their low oleic counterparts HA89, RHA274 and PHD, its expression was elevated in the low oleic line HA292 compared to HA349. Thus the differences in *LTP* transcript levels do not correlate with the high oleic phenotype (Table 2.1). The LTP has been shown to be involved in the plant stress response in pepper and rice, since expression of the *LTP* genes is induced by pathogen infection, wounding and abiotic stresses in aerial vegetative and reproductive tissues (Jung et al., 2003, Guiderdoni et al., 2002). A putative *LTP* cDNA from sunflower seeds was constitutively expressed in the seeds, while expression was lacking in the aerial tissues (Regente and Canal, 2003). Our understanding of the specific role and involvement of LTP in the cellular processes is still very limited. Therefore we are unable to make any assumptions about the role of *LTP* in kernel development. But the fact that this gene is

differential expressed in all four comparisons suggests a possible involvement of the gene in the high oleic phenotype. Further examination of *LTP* will be necessary to determine its role in the lipid metabolism of the developing kernels of sunflower.

Besides *FAD2-1*, HA292 is the only wildtype line that also displayed elevated transcript levels for the other two *FAD2* genes, *FAD2-2* and *FAD2-3* (Table 2.3), which was confirmed by real-time PCR. Both genes are not differentially expressed in the other three microarray comparisons, which are in accordance with the uniform expression observed in developing embryos, cotyledons, leaves, hypocotyls and roots of HA89 by Martinez-Rivas et al. (2001). The elevated transcript levels of *FAD2-2* and *FAD2-3* in HA292 were somewhat contradictory to the observed phenotype, as we would expect a high linoleic acid content in HA292. In contrast, HA292 displayed a noticeably higher oleic acid and lower linoleic acid content (Table 2.1) than the other three wildtype lines. The discrepancy between the phenotype and the elevated transcript levels of *FAD2-2* and *FAD2-3* can not be fully explained by the fact, that HA292 is the only confectionary sunflower line among the eight lines, and therefore resulting from the confectionary genetic background.

Three of the four oligonucleotides designed to the *oleosin* gene family displayed differential expression in one or more arrays. Oleosins are phospholipid membrane proteins, which represent the major protein constituent of the outer oil body matrix. Thus, transcripts are expected to be highly abundant in the developing kernel, demonstrated by the large intensity readings of *oleosin1* and *oleosin3*. We did not expect to see major changes in *oleosin* expression between lines, since the compared genotypes only differed for fatty acid composition and not oil concentration in the kernel. Even though the transcriptional differences were statistically significant, the difference in relative mRNA amounts only ranged from 1.2 to 1.7 between lines. The biological relevance of these subtle discrepancies in the highly expressed *oleosin* genes is likely negligible, given the overall abundance of oil bodies in the seed.

The *keto-acyl-CoA reductase* showed a 2.2-fold increase in transcripts in HA341 compared to HA89 and a 1.8- to 2.5-fold increase in HA292 compared to HA349. Real-time PCR indicated even greater expression differences for *KR* in

HA89-HA341. The fact, that *KR* expression is relatively low compared to the *actin* standard in the microarray, and the fact, that the real-time PCR assay is by far more sensitive to differences in RNA amounts, account for the observed discrepancies between microarray and real-time PCR results. KR is a membrane bound cytosolic enzyme, which is involved in the biosynthesis of very long chain fatty acids, such as cuticular waxes (C_{26} to C_{32}) (Xu et al., 2002, and therein). Since only saturated fatty acids, such as palmitic (C_{16}) and stearic (C_{18}) acid, are substrates for the production of long chain fatty acids, oleic and linoleic acid levels should not be affected by differential KR activity.

The *enoyl-ACP reductase2 (ENR2)* and the *acyl carrier protein* displayed increased transcriptional activity in HA292 compared to HA349. The latter one was also found to be slightly overexpressed in RHA274 compared to RHA345. The acyl carrier protein is a central protein cofactor, which is essential in the de novo fatty acid synthesis. After malonyl-transacetylation, ACP is involved in all subsequent reactions of the plastidial pathway leading up to the formation of C_{16} and C_{18} fatty acids (Ohlrogge and Browse, 1995). Therefore the observed minimal differences in *ACP* transcription rate are biologically not very meaningful in regard to fatty acid composition. The *ENR2* is one of three oligonucleotides of this gene family present on the array. ENR is also an essential enzyme in de novo fatty acid synthesis in the plastid. It is necessary to complete each round in the cyclic fatty acid elongation process resulting in a saturated fatty acid (Ohlrogge and Browse, 1995). Thus *ENR2* transcriptional differences are unlikely to affect the overall fatty acid composition in HA292 and HA349.

While *FAB2-1* showed slightly increased transcriptional activity in RHA274, it could not be confirmed in the quantitative PCR validation. The stearoyl-ACP desaturase is a plastidial, non-membrane bound enzyme, which is responsible for the introduction of the first double bond in the saturated fatty acid chain resulting in the formation of oleoyl-ACP. In addition the paralogue *FAB2-2* revealed elevated transcript levels for HA292 in two of the three arrays varying from a 1.5- to a 3.0-fold difference. Real-time PCR confirmed a modest 1.6-fold change in transcript levels.

Even though HA292 and HA349 differ in respect to their saturated versus unsaturated fatty acid ratio (Table 2.1), similar shifts in this ratio are observed in the other three low by high oleic acid comparisons, which do not display transcriptional differences in *FAB2* expression.

Both, plastidial and cytosolic, *LPAAT* genes revealed minor increases in mRNA levels in the low oleic line HA292. Both LPAAT enzymes are membrane bound in the plastid and ER, respectively, and responsible for the rapid conversion of lysophosphatidate (LPA) to phosphatidate (PA) by addition of either C_{16} or C_{18} fatty acids to LPA. Since oleic and linoleic are both 18-carbon fatty acids, differential *LPAAT* expression is unlikely to play a role in oleic/linoleic fatty acid composition.

The three microarray comparisons between the sunflower oilseed lines HA89-HA341, RHA274-RHA345 and PHD-PHC revealed very few genes with differential expression patterns, which indicate that transcriptional regulation of the genes in the glycerolipid biosynthesis is conserved among oilseed lines. This is further supported by the fact that 12 of the 48 genes tested in the array displayed significant differences in transcript levels for the comparison between the confectionary line HA292 and the oilseed line HA349. Furthermore the confectionary line HA292 displayed a distinctly different fatty acid profile compared to the wildtype lines HA89, RHA274, and PHD. The fact that we observed only modest transcriptional differences between HA292 and HA349 for most genes can be partially attributed to the differing genetic backgrounds of the two lines.

The *FAD2-1* Mutation Exposes Allelic Variability Among Candidate Genes for Oleic Acid QTL in Sunflower

CHAPTER 3

INTRODUCTION

Since the discovery of high oleic phenotypes (Soldatov, 1976), several models have been proposed to explain the genetics of high oleic acid in sunflower. Ol_1 has been the only common denominator in the different genetic models (Urie, 1985; Miller et al., 1987a; Fernández-Martínez et al., 1989; Fernandez et al., 1999, Lacombe and Bervillé, 2001, Lacombe et al., 2002a, 2002b). The *FAD2-1* mutation, which cosegregates with Ol_1, is necessary but not always sufficient for producing the high oleic acid phenotype in sunflower kernels (Lacombe and Bervillé, 2001, Lacombe et al., 2002a, 2002b). Oleic acid distributions among progeny segregating for the *FAD2-1* mutation (Ol_1) are typically complex or quantitative, although the signal-to-noise ratio is sufficiently low in some populations to identify two discrete genotypic classes (*FAD2-1*/_and *fad2-1/fad2-1*) (Fernández-Martínez et al., 1989; Fernandez et al., 1999, Lacombe and Bervillé, 2001, Lacombe et al., 2002a). Such populations presumably lack allelic variability for QTL underlying oleic acid variability in other genetic backgrounds. Several hypotheses have been formulated as to the number and nature of loci underlying oleic acid variability in sunflower populations segregating for the *FAD2-1* mutation. While Miller et al. (1987a) concluded that the high oleic phenotype was controlled by Ol_1 and a 'modifier' (*ml*), other putative modifiers have been proposed on the basis of phenotypic analyses, e.g. Ol_2, Ol_3, and *supole* (Fernández- Martínez et al., 1989; Lacombe et al., 2001); however, none of the underlying genes or genetic mechanisms have been identified..

The glycerolipid biosynthetic pathway has been studied in depth in the model plant systems *A. thaliana* during the last fifteen years. Since this pathway is essential and thus conserved among plants, it allows us to use the knowledge obtained in

Arabidopsis to identify additional candidate genes for the high oleic phenotype. The review of the current literature on oleic acid biosynthesis and high oleic oilseeds in other plant species identified several candidate genes possibly affecting the oleic acid content in the sunflower seed (Ohlrogge and Browse, 1995, Mekhedov et al., 2000, Liu et al., 2002, Stoutjesdijk et al., 2002, and Bonaventure et al., 2003). The genes in our study include *FAD2-1*, *FAD2-2*, *FAD2-3*, *FATA*, *FATB*, *FAB2-1*, *FAB2-2*, *FAD6*, *KASI*, *KASII*, and *KASIII*.

The goal of the present study was to identify genes, which are underlying the quantitative variability for seed oil fatty acids in a sunflower population segregating for the *FAD2-1* mutation. Candidate gene approaches were used to build genetic models and gain an understanding of the network of gene effects underlying the complex oleic acid distributions observed among RILs produced from a cross between low and high oleic acid inbred lines (PHC × PHD). We performed analyses of allelic diversity for several candidate genes among low and high oleic acid lines. We identified DNA polymorphisms in these genes, developed sequence-tagged-site (STS) markers, and genetically mapped several candidate genes in PHC × PHD and other mapping populations. In addition, we performed QTL analyses in PHC × PHD using the candidate gene loci as independent variables. Several of the loci were identified as candidates for oleic acid QTL in sunflower.

MATERIALS AND METHODS

Plant Materials and DNA Isolation

HA89 and RHA274 are wildtype, low oleic oilseed inbred lines, whereas HA292 is a wildtype, low oleic confectionary inbred line. HA341, HA349 and RHA345 are mutant, high oleic lines. They were derived by selecting for high oleic acid content among BC_1F_2 and BC_1F_3 progeny from crosses between the three recurrent parents and the high oleic donor cultivar Pervenets (HA89*2/Pervenets, HA292*2/Pervenets and RHA274*2/Pervenets), respectively (Miller et al., 1987b). PHC is a high oleic and PHD is a low oleic proprietary oilseed line, while PHA and PHB are proprietary oilseed restorer lines, all developed by Pioneer Hi-Bred International (Johnston, Iowa). RHA 280 is a confectionary fertility restorer line and RHA801 is an oilseed fertility restorer line (Fick et al., 1974, Roath et al., 1981). RHA373 and RHA377 are oilseed fertility restorer lines (Miller, 1992); NMS373, NMS377and NMS801 are nuclear male-sterile sunflower genetic stocks (Miller, 1992 and 1997). HA383 is an oilseed maintainer line (Miller and Gulya, 1995). Seeds samples of ANN1811, Havasupai, Hopi, and *H. argophyllus* were acquired from Mary Brothers (USDA-ARS, National Plant Germplasm System, North Central Plant Introduction Station, Ames, Iowa, USA) (http://www.ars-grin.gov) and described in Tang and Knapp (2003). ANN1238 and *H. deserticola* seed samples were obtained from Loren Rieseberg (Indiana University, Bloomington, IN, USA).

The F_7 RIL reference mapping population of RHA280 x RHA801 was described in Tang et al. (2002) and Yu et al. (2003). The F_5 RIL population of PHC x PHD was described in Tang et al. (2003). NMS373*2/ANN1811 is a backcross derived population developed in the summer of 2001 and 2002 in Corvallis, OR; NMS377*2/Havasupai and NMS801*2/*H. argophyllus* are backcross derived populations developed in the summer of 2003 and 2004 in Corvallis, OR (unpublished data).

Leaves were harvested from individual 3 to 6-week-old greenhouse-grown plants and stored at -80°C until extraction. The frozen leaf samples were ground in liquid N_2, and DNA was isolated from the ground samples using a modified CTAB method (Webb and Knapp, 1990). DNA was isolated from a bulked leaf sample of 10 BC_1 progeny for the 3 backcross derived populations.

Candidate Gene Identification

Sunflower sequences in GenBank (NCBI, Bethesda, MD, USA), in the Composite Genome Project EST database (CGPDB) (http://cgpdb.ucdavis.edu) (Kozik et al., 2002), and in a developing kernel EST database (unpublished) were identified for *FAD2-1*, *FAD2-2*, *FAD2-3*, *FATA*, *FATB*, *FAB2-1*, *FAB2-2*, *FAD6*, *KASI*, *KASII*, and *KASII* (Table 3.1). The selected contig and singleton sequences were aligned with the respective *A. thaliana* locus (Table 3.1) from the TAIR database (Huala et al., 2001) using Vector NTI Suite 8 software programs AlignX and ContigExpress (Invitrogen, Carlsbad, CA, USA). Sequences identified as credible matches to the candidate genes were subsequently used as templates for primer design using Primer3 software (Rozen and Skaletsky, 2000).

SSCP and Sequence Analysis

PCR products were subjected to single strand conformational polymorphism (SSCP) analysis according to Slabaugh et al. (1997) with slight modifications. 5 µl samples were run on 0.5 x MDE gels (Cambrex Bio Science Rockland, Rockland, ME, USA) using 0.6 x TBE running buffer. Gels (1mm thick x 50 cm wide x 22cm high) were run on a DASG-400-50 polyacrylamide gel apparatus (CBS Scientific Co., Del Mar, CA, USA) at 3.5 watts constant power for 14 to 20 h depending on fragment size at room temperature and stained with silver nitrate (Sanguinetti et al., 1994). One of the glass plates was treated with γ-methacryloxypropultrimethoxysilane (Sigma Chemical Co., St. Lois, MO, USA) so that the gel remained attached during silver-staining.

Sequencing was performed by the Nevada Genomics Center (Reno, NV) using an ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA). Sequence alignments were made using the Vector NTI Suite 8 software programs AlignX and ContigExpress (Invitrogen, Carlsbad, CA, USA).

Microsatellite and SNP Genotyping and Mapping

Microsatellite markers described by Tang et al. (2001) and Yu et al. (2003) were screened for polymorphism between PHC and PHD using the genotyping methods described by Tang et al. (2003). The F_5 RILs were genotyped for the microsatellite markers found to be polymorphic between PHC and PHD. The genotyping assays were performed using post-PCR multiplexing.

SNP markers were assayed and analyzed using the acycloprime genotyping method, essentially as described in Kolkman et al. (2004). Briefly, SNPs were scored using the fluorescence polarization-template directed incorporation assay (Chen et al., 1999, Kwok and Chen, 2003) and commercial kits (AcycloPrime-FP SNP Detection Kit, Perkin-Elmer Life Sciences, Boston, MA, USA). Target amplification and terminator incorporation reactions were performed as recommended by the kit manufacturer. SNP genotypes were read on a Wallac 1420 VISTOR3™ fluorescence polarization plate reader (Perkin-Elmer, Boston, MA, USA) and alleles were called using an EXCEL macro supplied by Perkin-Elmer for AcycloPrime-FP SNP genotyping.

Agarose, SSCP, SNP, or SSR markers for the candidate genes were developed and genotyped in one of two segregating populations for which reference genetic linkage maps have been developed, RHA280 x RHA801 (Tang et al., 2001, Yu et al., 2003) and NMS373 x ANN1811 (Gandhi et al. 2005). The markers were integrated into the aforementioned genetic linkage maps according to Kolkman et al. (2004). Genetic mapping analyses were performed using MAPMAKER (Lander et al., 1987), essentially as described by Tang et al. (2002). The RIL mapping function of MAPMAKER 3.0 (Lander et al., 1987) and G-MENDEL 3.0 (Holloway and Knapp, 1993) were used to construct genetic linkage maps for LG 1, 7 and 14 among the PHC

x PHD RILs. Loci were grouped using likelihood odds (LOD) threshold of 3.0 and 4.0 and map distances were calculated using the Kosambi (1944) mapping function (Tang et al., 2003).

Fatty Acid Analysis

The 262 F_5 RILs were grown in the summer of 2000 and 2001 in Corvallis, OR, and Woodland, CA, and 5 heads were bagged per RIL. A bulked sample of 20 seeds from each year and location was analyzed for the four fatty acids of palmitic (16:0), stearic (18:0), oleic (18:1), and linoleic acid (18:2) according to the method described by Brandt and Knapp (1993) with slight modifications. Briefly, bulked seeds were crushed using a polytron homogenizer (Kinematica, Luzern, Switzerland) in 10 ml of hexane, incubated at RT for 10 min, 0.5 ml was transferred to a fresh tube and evaporated to dryness at 50°C under a N_2 stream. Lipids solubilized in 0.2 ml ethyl ether were converted to fatty acid methyl esters (FAMEs) by adding 0.2 ml of 0.1 M KOH in methanol and incubating for 5 min at 50°C. Methylation reactions were stopped with 0.2 ml of 0.15 M HCl, and FAMEs were extracted with 1 ml Hexane. One-microliter samples were injected in a HP 6890 Series Gas Chromatograph (Hewlett-Packard, Wilmington, DE, USA), and fatty acid profiles were calculated using the HP ChemStation software (Hewlett-Packard, Wilmington, DE, USA).

Statistical Analyses

Heterozygostity scores were estimated using the following equation: $H = 1 - \sum_{i=1}^{l} p_i^2$ where l is the number of codominant alleles at the locus, p is the frequency of the i-th allele (Liu, 1998). The model selection procedure for the analysis of the oleic acid content in the F_5 RIL population included stepwise regression, performed using SAS version 8.02 PROC REG software (SAS Institute Inc., Cary, NC, USA). Probability levels were set at 0.05 for the test to enter into the model and for the test to remain in the model. In addition we performed manual model selection adding main and interaction effects individually, considering only effects at the 0.001 probability level. Subsequently the selected effects were combined and further reduced

in number. Statistical analysis of the effects of *FAD2-1*, ORS460 (*FATB-1*), ORS1256 (*FATB-2*), *KASII*, *KASIII*, *FAB2-1*, *FAB2-2* and CRT394 (*FATA*) and the two-way interaction effects between *FAD2-1* × ORS460 (*FATB-1*), ORS460 (*FATB-1*) × *KASII*, ORS460 (*FATB-1*) × *KASIII*, ORS460 (*FATB-1*) × *FAB2-2*, *KASII* × *KASIII*, *KASII* × ORS1256 (*FATB-2*), *KASII* × CRT394 (*FATA*), *KASIII* × ORS1256 (*FATB-2*) and *FAB2-1* × *FAB2-2* on oleic acid content were performed using SAS version 8.02 PROC MIXED software (SAS Institute Inc., Cary, NC, USA). Least square means were estimated for each genotypic class. The intralocus effects of *FAD2-1*, ORS460 (*FATB-1*), ORS1256 (*FATB-2*), *KASII*, *KASIII*, *FAB2-1*, *FAB2-2* and CRT394 (*FATA*) were estimated using linear contrasts between homozygous genotypes ($y_{AA} - y_{aa}$), where y_{AA} and y_{aa} are the least square means for *PHDPHD* and *phcphc* F_5 RILs, respectively. The contrasts were adjusted with the appropriate divisor (2). The interlocus effects of *FAD2-1* × ORS460 (*FATB-1*), ORS460 (*FATB-1*) × *KASII*, ORS460 (*FATB-1*) × *KASIII*, ORS460 (*FATB-1*) × *FAB2-2*, *KASII* × *KASIII*, *KASII* × ORS1256 (*FATB-2*), *KASII* × CRT394 (*FATA*), *KASIII* × ORS1256 (*FATB-2*) and *FAB2-1* × *FAB2-2* were estimated using the product of linear contrasts between the two loci ($y_{AABB} - y_{AAbb} - y_{aaBB} - y_{aabb}$), where y_{AABB}, y_{AAbb}, y_{aaBB} and y_{aabb} are the least square means for *PHDPHDPHDPHD*, *PHDPHDphcphc*, *phcphcPHDPHD*, *phcphcphcphc* F_5 RILs, respectively. The contrasts were adjusted with the appropriate divisor (4). Because the number of F_5 RILs was smaller than the number of possible genotypic classes (the data were unbalanced), the statistical significance of each contrast (genetic effect) was estimated using Type III F-tests. The heritability of the oleic acid content and the R^2 value was estimated using SAS version 8.02 PROC GLM software (SAS Institute Inc., Cary, NC, USA).

RESULTS

Developing Markers for the Candidate Genes

The CGPDB EST database, a developing kernel EST library, and GenBank revealed sunflower sequences for each of the glycerolipid biosynthesis genes of interest (Table 3.1). The primers designed to the obtained sequences were subjected to the following procedure (for primer sequences see Appendix H): First PCR was performed on genomic DNA of a panel of up to 15 sunflower lines. The panel consisted of the low oleic lines HA89, HA292, RHA274 and PHD and the corresponding high oleic counterparts HA341, HA349, RHA345 and PHC. RHA280 and RHA801 are the parental lines of our current F_7 RIL reference mapping population (Tang et al., 2001, Yu et al., 2003), while RHA373 is the recurrent parent to our second BC_1 reference population, represented by a bulked DNA sample of 10 BC_1 progeny from NMS373*2/ANN1811 (Gandhi et al., 2005). In addition the panel comprised the parental lines for our two future BC_1 mapping populations, the recurrent parent RHA377 and RHA801 and bulked DNA samples of 10 BC_1 progeny from the crosses of NMS377*2/Havasupai as well as NMS801*2/*H. argophyllus*. PCR amplification and amplicon length was recorded on agarose, compared to the known cDNA length to determine the presence of introns in the genomic sequence and screened for possible DNA fragment length polymorphism. Since the frequency of detectable polymorphism on agarose was low with 0 to 2 out of the 15 fragments, the PCR products were then further screened on single strand conformational polymorphism (SSCP) polyacrylamide gels. In the absence of relevant polymorphism the fragments were subsequently sequenced to screen for small insertion-deletion polymorphism (INDEL) and/or single nucleotide polymorphism (SNP) (Appendix I). Below we present the results for each of the candidate genes in the oleic acid biosynthetic pathway.

Table 3.1. List of the CGPDB ESTs and Contigs and GenBank accessions used in the marker development for the candidate genes in the oleic acid biosynthesis. The *Arabidopsis thaliana* locus number refers to the equivalent locus in the TAIR database.

Gene	Source	Reference	*A. thaliana* locus
FAD 2-1	U91341	Hongtrakul et al. 1998a	AT3G12120
FAD2-2	QH_CA_Contig0888	Kozik et al. 2002	AT3G12120
	AJ292275		AT3G12120
	AF251843	Martinez-Rivas et al. 2001	AT3G12120
FAD 2-3	AF251844	Martinez-Rivas et al. 2001	AT3G12120
FAT A	AY805124		AT3G25110
	AY805125		AT3G25110
	QH_CA_Contig2357	Kozik et al. 2002	AT3G25110
FAT B	AF036565		AT1G08510
	QH_CA_Contig6130	Kozik et al. 2002	AT1G08510
KAS I	QH_CA_Contig1997	Kozik et al. 2002	AT5G46290
KAS II	AY805139		AT2G04540
	AY805138		AT2G04540
KAS III	QH_CA_Contig1814	Kozik et al. 2002	AT1G62640
	CF079060	Kozik et al. 2002	AT1G62640
FAB 2-1	U91340	Hongtrakul et al. 1998b	AT2G43710
FAB 2-2	U91339	Hongtrakul et al. 1998b	AT2G43710
FAD 6	QH_CA_Contig1636	Kozik et al. 2002	AT4G30950

FAD2-1

The coding sequence of the *FAD2-1* was found to be monomorphic between the low oleic line HA89 and the high oleic line HA341 (Hongtrakul et al., 1998a). We confirmed these results in Chapter 1 by sequencing the exon and intron of the low oleic lines HA89 and PHD and of the high oleic lines HA341 and PHC without recovering any polymorphism. In addition, we sequenced the coding sequence of the confectionary line RHA280 and the oilseed line RHA801, the parental lines of our reference mapping population, and found no sequence differences (GenBank AY800244 and AY800245) (alignment in Appendix A). From our previous work in Chapter 1 we were able to use primers F4-R1 (sequences in Appendix H) as a dominant marker for *FAD2-1* in the low by high oleic cross of PHD x PHC, since the primers amplified a 650 bp fragment only in the high oleic lines. The fragment mapped to linkage group (LG) 14 in PHC x PHD (Figure 3.1). The dominant marker was also subsequently used for the *FAD2-1* locus in our QTL analysis for oleic acid content among the 262 F_5 RILs of the cross between PHD x PHC, segregating for oleic acid content.

In our search for a co-dominant marker and a polymorphism in our mapping population RHA280 x RHA801, we sequenced 854 bp downstream of the stop codon of *FAD2-1*, using the sequence of the intergenic region isolated in Chapter 1 as a template for primer design. The sequencing revealed two microsatellite repeats in the 3' UTR of the gene (Figure 3.2), an imperfect GT repeat 78 bp downstream of the stop codon and an imperfect TA repeat at the cutoff of the 3' UTR, 177 bp downstream of the stop codon. We used fluorescently-labeled primers to assess their allelic diversity on a panel of elite and wild sunflower lines. Both microsatellites displayed no allelic diversity among the elite germplasm (Table 3.2). SSR1 produced a 210 bp fragment, while SSR2 yielded a 199 bp fragment. In contrast, the two markers were highly polymorphic among wild and exotic sunflower germplasm. The allele length varied from 199 bp in Havasupai to 232 bp in *H. argophyllus* for the GT repeat and ranged from 185 bp in *H. argophyllus* to 209 bp in the same accession for the TA repeat. The average heterozygosity scores for SSR1 and SSR2 were 0.39 and 0.45 respectively

Figure 3.1. Position of *FAD2-1* on LG 14 of our reference map and on LG 14 of PHD x PDC. The map of PHC x PHD is on the left, the map of the reference population is on the right.

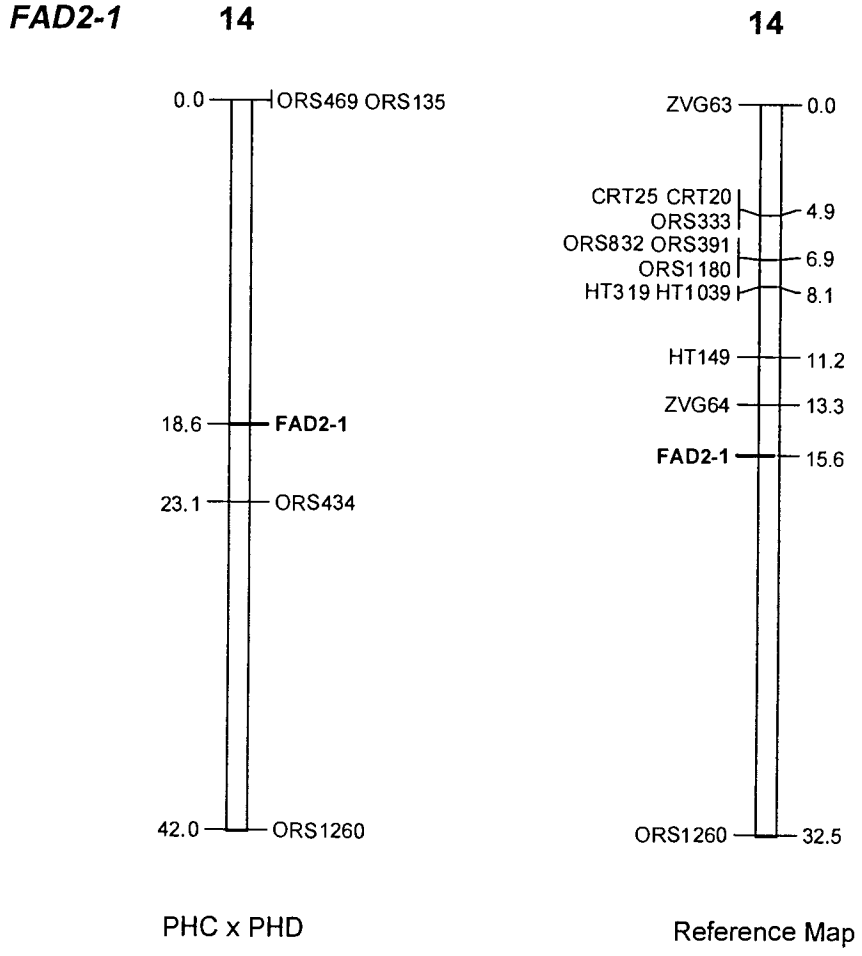


Figure 3.2. The 3'UTR of *FAD2-1* contains two microsatellites. The SSRs are marked in bold, uppercase letters and indicated below the alignment. The stop codon is marked in uppercase letters.

```
                    *         20          *         40          *
HA89    : ----------------------gaataagatgaatTAAtaagtgtttagt :  28
HA341   : -------tgtgttttggtacaagaataagatgaatTAAtaagtgtttagt :  43
PHC     : -------tgtgttttggtacaagaataagatgaatTAAtaagtgtttagt :  43
PHD     : -------tgtgttttggtacaagaataagatgaatTAAtaagtgtttagt :  43
FAD2-1  : aaggaggtgtgttttggtacaagaataagatgaatTAAtaagtgtttagt :  50
          Exon2                                    TAA       3'UTR

                   60          *         80          *         100
HA89    : ggatgtctatgcttattaagtctgtgacaatgggtcttgtcctggttctg :  78
HA341   : ggatgtctatgcttattaagtctgtgacaatgggtcttgtcctggttctg :  93
PHC     : ggatgtctatgcttattaagtctgtgacaatgggtcttgtcctggttctg :  93
PHD     : ggatgtctatgcttattaagtctgtgacaatgggtcttgtcctggttctg :  93
FAD2-1  : ggatgtctatgcttattaagtctgtgacaatgggtcttgtcctggttctg : 100

                    *        120          *        140          *
HA89    : ggtctgggtctggttaGTGTGTGTGGTTAGTGTGTGTtgtgggtcaagtg : 128
HA341   : ggtctgggtctggttaGTGTGTGTGGTTAGTGTGTGTtgtgggtcaagtg : 143
PHC     : ggtctgggtctggttaGTGTGTGTGGTTAGTGTGTGTtgtgggtcaagtg : 143
PHD     : ggtctgggtctggttaGTGTGTGTGGTTAGTGTGTGTtgtgggtcaagtg : 143
FAD2-1  : ggtctgggtctggttaGTGTGTGTGGTTAGTGTGTGTtgtgggtcaagtg : 150
                           GTGTGTGTGGTTAGTGTGTGT

                   160          *         180          *         200
HA89    : tatgtcagatggatgtaacgcagtatgtgttgccgtgttgggtgtgtttt : 178
HA341   : tatgtcagatggatgtaacgcagtatgtgttgccgtgttgggtgtgtttt : 193
PHC     : tatgtcagatggatgtaacgcagtatgtgttgccgtgttgggtgtgtttt : 193
PHD     : tatgtcagatggatgtaacgcagtatgtgttgccgtgttgggtgtgtttt : 193
FAD2-1  : tatgtcagatggatgtaacgcagtatgtgttgccgtgttgggtgtgtttt : 200

                    *        220          *        240          *
HA89    : agaactattaatgaaTATATTATATATATctcttatttgggtgaactcaa : 228
HA341   : agaactattaatgaaTATATTATATATATctcttatttgggtgaactcaa : 243
PHC     : agaactattaatgaaTATATTATATATATctcttatttgggtgaactcaa : 243
PHD     : agaactattaatgaaTATATTATATATATctcttatttgggtgaactcaa : 243
FAD2-1  : agaactattaatgaaTATAT---------------------------- : 220
          3'UTR end       TATATTATATATAT

                   260          *         280          *         300
HA89    : atcgggtaatgggtttaaatgggctagcaccactcacgactattgtttat : 278
HA341   : atcgggtaatgggtttaaatgggctagcaccactcacgactattgtttat : 293
PHC     : atcgggtaatgggtttaaatgggctagcaccactcacgactattgtttat : 293
PHD     : atcgggtaatgggtttaaatgggctagcaccactcacgactattgtttat : 293
FAD2-1  : ---------------------------------------------- :   -
```

Table 3.2. Allelic variation of the repeat length for the two microsatellites in the 3' UTR of *FAD2-1* gene.

Genotype	SSR1 in bp	SSR2 in bp
HA89	210	199
HA341	210	199
HA292	210	199
HA349	210	-
RHA274	210	199
RHA345	210	199
PHD	210	199
PHC	210	199
RHA280	210	199
RHA801	210	199
PHA	210	199
PHB	210	199
HA383	210	199
RHA373	210	199
NMS373 x ANN1811	210	-
ANN 1811-TX (PI 494567)	204/220	193/206
ANN 1238-NE	201/210	193
RHA 377	210	199
NMS377 x Havasupai	210	-
Havasupai (PI 369358)	199	181
Hopi (PI 369359)	210/220	199/206
NMS801 x *H. argophyllus*	204/210/220	-
H. argophyllus (PI 494582)	205/232	185/209
H. deserticola (Ames26094)	205	206
Mean heterozygosity (H)	0.39	0.45

(Table 3.2). Nevertheless we were unable to map *FAD2-1* in our reference populations.

In order to find a polymorphism allowing us to map *FAD2-1* in RHA280 x RHA801, we used the knowledge gained in Chapter 1. We knew that most of the intergenic region between the duplicated copies of *FAD2-1* in the high oleic lines is sequence downstream of the *FAD2-1* locus in the wildtype. When we screened a panel of eight lines for the 2,636 bp fragment in the intergenic region, the only confectionary line on the panel, HA292, failed to amplify. We were able to use these primers F3 and R6 to discriminate between the oilseed line RHA801 and the confectionary line RHA280, resulting in a dominant agarose marker, which mapped *FAD2-1* to LG 14 in our reference population (Figure 3.1).

FAD2-2

The CGPDB EST database contained one relevant hit to *FAD2-2*, contig 888 (Table 3.1). The GenBank search produced two sunflower *FAD2-2* matches, AJ292275 and AF251843 (Martínez-Rivas et al. 2001). Both GenBank sequences contained the complete coding sequence of the gene and were monomorphic when aligned. The contig contained sequences from both genotypes RHA280 and RHA801 and revealed eight synonymous SNPs between the two lines. We designed target primers FAD2-2-T to amplify a 143 bp fragment containing a G/A SNP and used primer FAD2-2-SNP in the template directed dye-terminator incorporation assay to map the SNP to LG 1 in RHA280 x RHA801 (Figure 3.3) (Table 3.3). In addition, we designed primers FAD2-2 spanning the complete coding region. We amplified and sequenced the alleles in the lines HA89, HA341, HA292, HA349, PHC and PHD (GenBank AY802993-AY802997). The six lines form two separate haplotypes. HA349 represents haplotype one with four synonymous SNPs compared to the second haplotype comprised of HA89, HA341, HA292, PHC and PHD (Appendix B). Due to the lack of allelic diversity in the coding region of PHC and PHD, we shifted our focus to the 5' and 3' genomic sequence flanking the gene.

Figure 3.3. Position of *FAD2-2* and the closest flanking marker on LG 1 of our reference map and on LG 1 of PHC x PHD. The flanking marker is underlined, the map of PHD x PHC is on the right, the map of the reference population on the left.

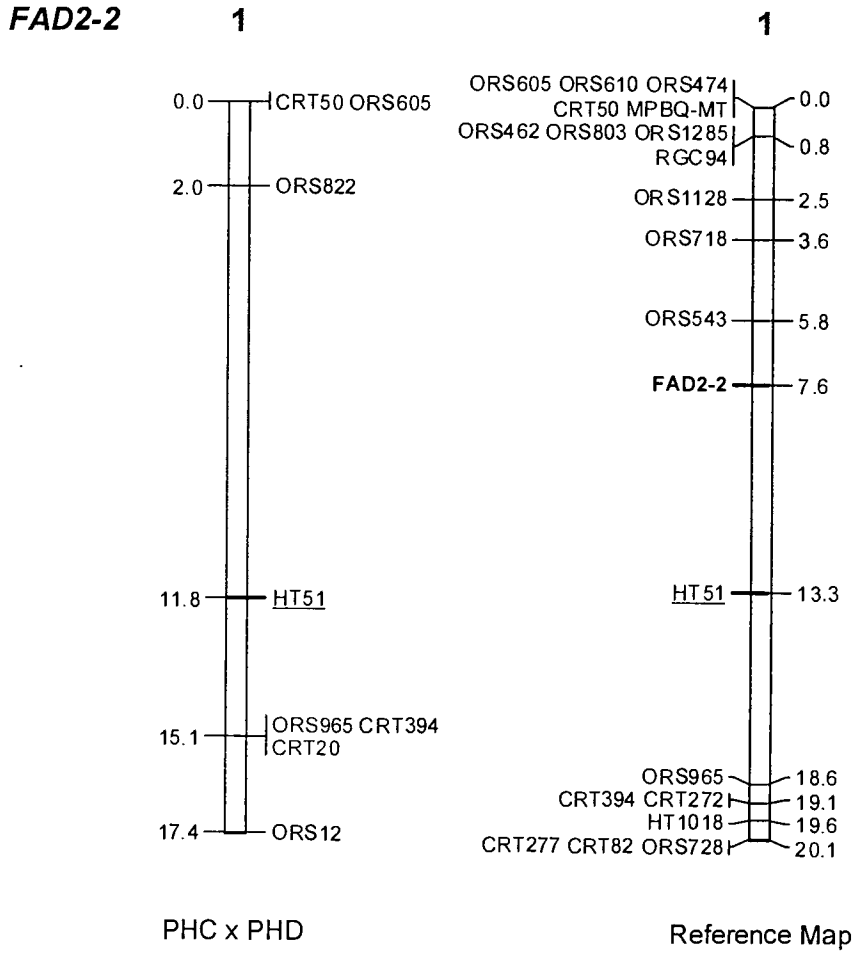


Table 3.3. Oligonucleotide primers used in the marker development process for the candidate genes of the oleic acid biosynthesis.

Name	Reference cDNA Sequence	Primer location (bp)	Type	HA89 allele length		Lines screened	Number of alleles	Heterozygosity
				cDNA	genomic			
FAB2-1 F/R	U91340	454 to 959	SSCP	505	~1300	13	4	0.49
FAB2-2 F/R	U91339	488 to 1,085	SSCP	597	~1200	13	2	0.14
KASI-1 F/R	QH_CA_Contig1997	28 to 380	SSCP/Sequencing	352	547	15/2	-/1	-/0
KASI-2 F/R	QH_CA_Contig1997	400 to 882	SSCP/Sequencing	482	1065	15/2	4/1	0.56/0
KASII F/R	sT3bHAK013H11095	344 to 640	SSCP	306	500	15	5	0.5
KASIII-1 F/R	QH_CA_Contig1814	6 to 431	SSCP	425	~1700	15	6	0.69
KASIII-2 F/R	QH_CA_Contig1814	148 to 354	SSCP	206	860	15	6	0.69
FATB A F/R	AF036565	71 to 441	SSCP	370	1135	15	1	0
FATB B F/R	AF036565	377 to 778	SSCP	401	483	15	1	0
FATB C F/R	AF036565	725 to 1,127	SSCP	402	581	15	1	0
FATB D F/R	AF036565	1,108 to 1,489	SSCP	381	1000	15	3	0.24
FATA-1 F/R	sT3bHAK010A07007	184 to 364	SSCP	180	180	15	2	0.12
FATA-2 F/R	sT3bHAK010A07007	350 to 630	SSCP	280	1221	15	3	0.55
FAD6-1 F/R	QH_CA_Contig1636	154 to 571	SSCP	417	-	15 (2)*	3 (2)*	0.24 (0.5)*
FAD6-2 F/R	QH_CA_Contig1636	516 to 918	SSCP	402	1948	15	5	0.52
FAD2-1 cds F/R	U91341	96 to 1,321	Sequencing	1226	1226	6	1	0
FAD2-2 cds F/R	AJ292275	5 to 1,295	Sequencing	1291	1291	8	3	0.41
FAD2-3 cds F/R	AF251844	4 to 1,215	SSCP	1212	1212	15	2	0.12
FAD2-2-T F/R	QH_CA_Contig888	328 to 470	SNP target	143	143	-	-	-
FAD2-2 -SNP	QH_CA_Contig888	376	SNP	-	-	2	2	0.5
HT51 F/R	QH_CA_Contig480	92 to 323	SNP target	231	231	-	-	-
HT51-SNP	QH_CA_Contig480	164	SNP	-	-	4	2	0.5

* The number in parenthesis refers to the number of PCR amplicons obtained.

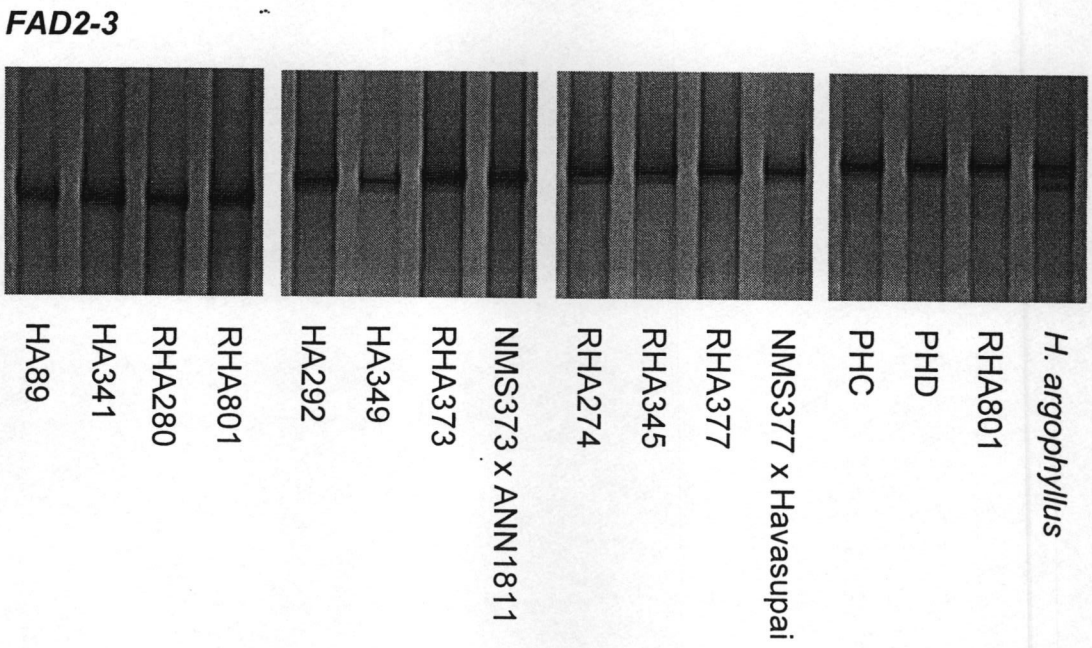
A single BAC clone from a genomic library of HA383 containing *FAD2-2* was previously isolated in our lab. We designed primers FAD2-2-5' and FAD2-2-3' to use a genome walking approach in order to obtain sequence upstream and downstream of *FAD2-2*. We recovered 1,067 bp of sequence upstream of the *FAD2-2* start codon, which did not contain a single nucleotide polymorphism between PHC and PHD (GenBank AY817414 and AY817415). Sequence quality deteriorated drastically beyond this point, thus we were not able to obtain further upstream sequence. In addition we were able to recover 436 bp of high quality sequence downstream of the *FAD2-2* stop codon, which was also monomorphic between PHC and PHD (GenBank AY817412 and AY817413). Overall we sequenced 2,656 bp at the *FAD2-2* locus without recovering any polymorphism between the two oilseed lines of the segregating population, further emphasizing the high degree of sequence conservation in the glycerolipid biosynthesis genes.

In order to be able to assess the effect of *FAD2-2* in the QTL analysis, we screened the markers flanking *FAD2-2* in RHA280 x RHA801 on the two parental lines PHD and PHC. The screening of 31 microsatellites surrounding *FAD2-2* yielded flanking markers ORS605 and CRT50 7.6 cM away. Rieseberg et al. mapped a SNP, derived from contig 450 in the CGPDB EST database and designated as HT51, 5.7 cM distant to *FAD2-2* using dHPLC (Figure 3.3). This method allows for mapping of a SNP without knowing neither the exact location nor the nucleotide configuration of the SNP. We used their primers HT51 to amplify a 231 bp fragment in RHA280, RHA801, PHC and PHD, which was subsequently sequenced (Table 3.3). It revealed that the G/T SNP between RHA280 and RHA801 is also polymorphic between PHC and PHD. We designed SNP primer HT51-SNP and genotyped HT51 on the 262 RILs using the aforementioned template directed dye-terminator incorporation assay (Figure 3.3).

FAD2-3

GenBank contained one sunflower sequence for *FAD2-3*, AF251844 (Table 3.1) (Martínez-Rivas at al., 2001), while the EST database search did not yield any

Figure 3.4. Screen for SSCP polymorphism in the *FAD2-3* gene on a panel of 15 sunflower lines.



matches for *FAD2-3*. The sequence obtained in GenBank covered the complete coding region of the gene, allowing us to design primers FAD2-3 spanning almost the entire exon (Table 3.3). We screened *FAD2-3* on the panel of 15 sunflower lines, and PCR produced a single, monomorphic fragment of 1,212 bp. The SSCP analysis revealed only two distinct alleles among the complete panel, whereas all 14 *H.annuus* lines carried the same allele (Figure 3.4). Sequence analysis of the fragments confirmed the results of the SSCP analysis. No quality sequence was available for NMS373 x ANN1811 and only low quality sequence was available for RHA377 and *H. argophyllus*. The latter one was the only sequence that displayed putative SNP polymorphism, which couldn't be assessed reliably due to the low sequence quality. In all other sunflower lines the coding sequence was completely conserved (GenBank AY802998-AY803008) (Appendix C). Like *FAD2-1* and *FAD2*-2, this locus carries no introns in the coding region, limiting any polymorphism to the far less likely exonic sequence. Due to the high degree of sequence conservation, we were unable neither to map the gene in any of our populations nor to develop a marker for it in order to include the locus in our QTL analysis.

FAB2

The GenBank search revealed two distinct sunflower *FAB2* sequences, U91339 and U91340 (Table 3.1) (Hongtrakul et al., 1998b), containing the complete coding sequence of each gene. Primers FAB2-1 and FAB2-2, specific to each sequence, revealed no agarose polymorphism (Table 3.3). The primers for *FAB2-1* amplified a ~1.3 kb fragment from genomic DNA, while the expected amplicon length from cDNA was 505 bp; thus the primers flanked an intron. The *FAB2-1* fragment confirmed the SSCP polymorphism between RHA280 and RHA801 observed by Hongtrakul et al. (1998b) (Figure 3.5). This polymorphism allowed us to map *FAB2-1* to LG 1 (Figure 3.6). *FAB2-1* was also genotyped and mapped in the 262 RILs of PHC x PHD. The four low oleic lines HA89, HA292, RHA274 and PHD carry the same allele as the high oleic lines HA341 and RHA345 and as RHA373 and NMS373 x ANN1811 (Figure 3.5). The two high oleic lines HA349 and PHC share a distinctly

Figure 3.5. Screen for SSCP polymorphism in the *FAB2* gene on a panel of 13 sunflower lines. Top panel: SSCP screen for *FAB2-1*. Lower panel: SSCP screen for *FAB2-2*.

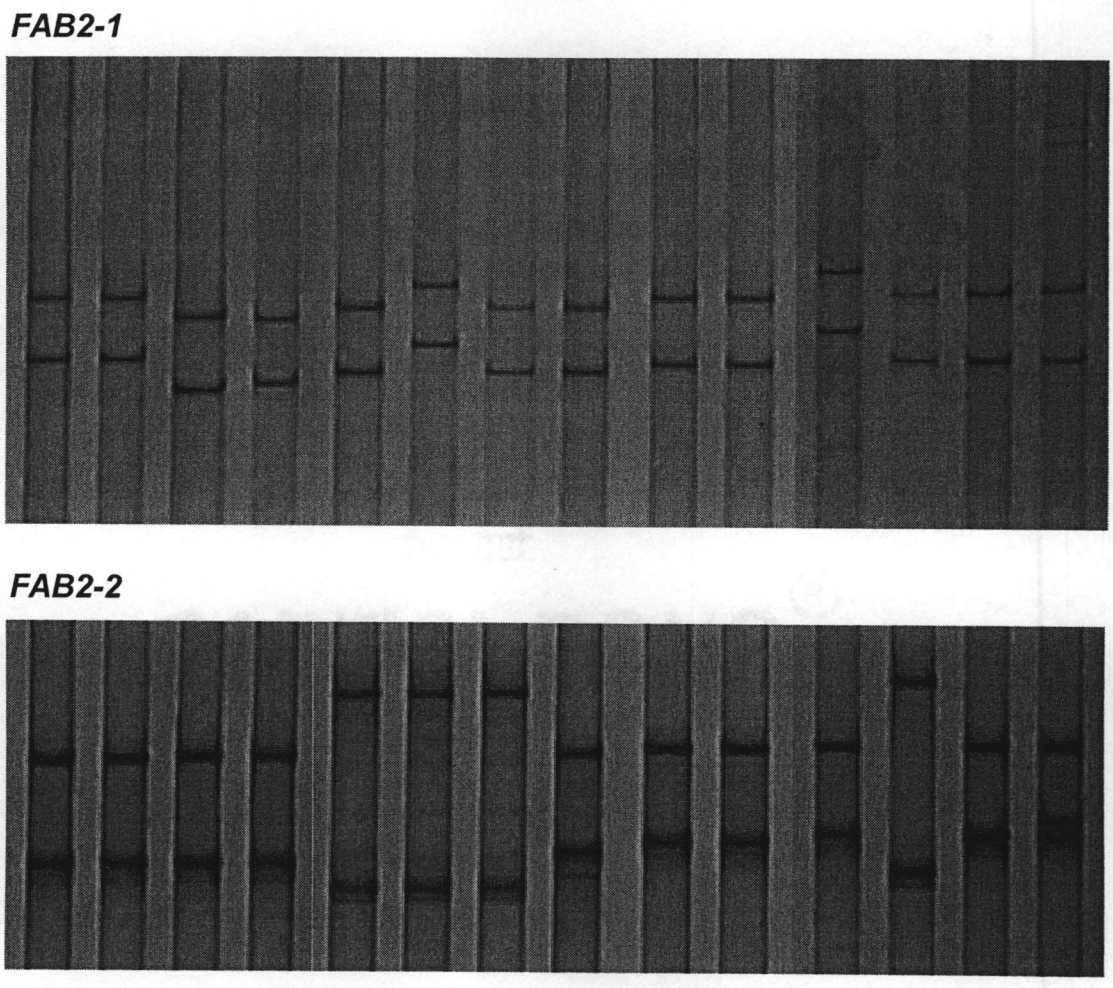
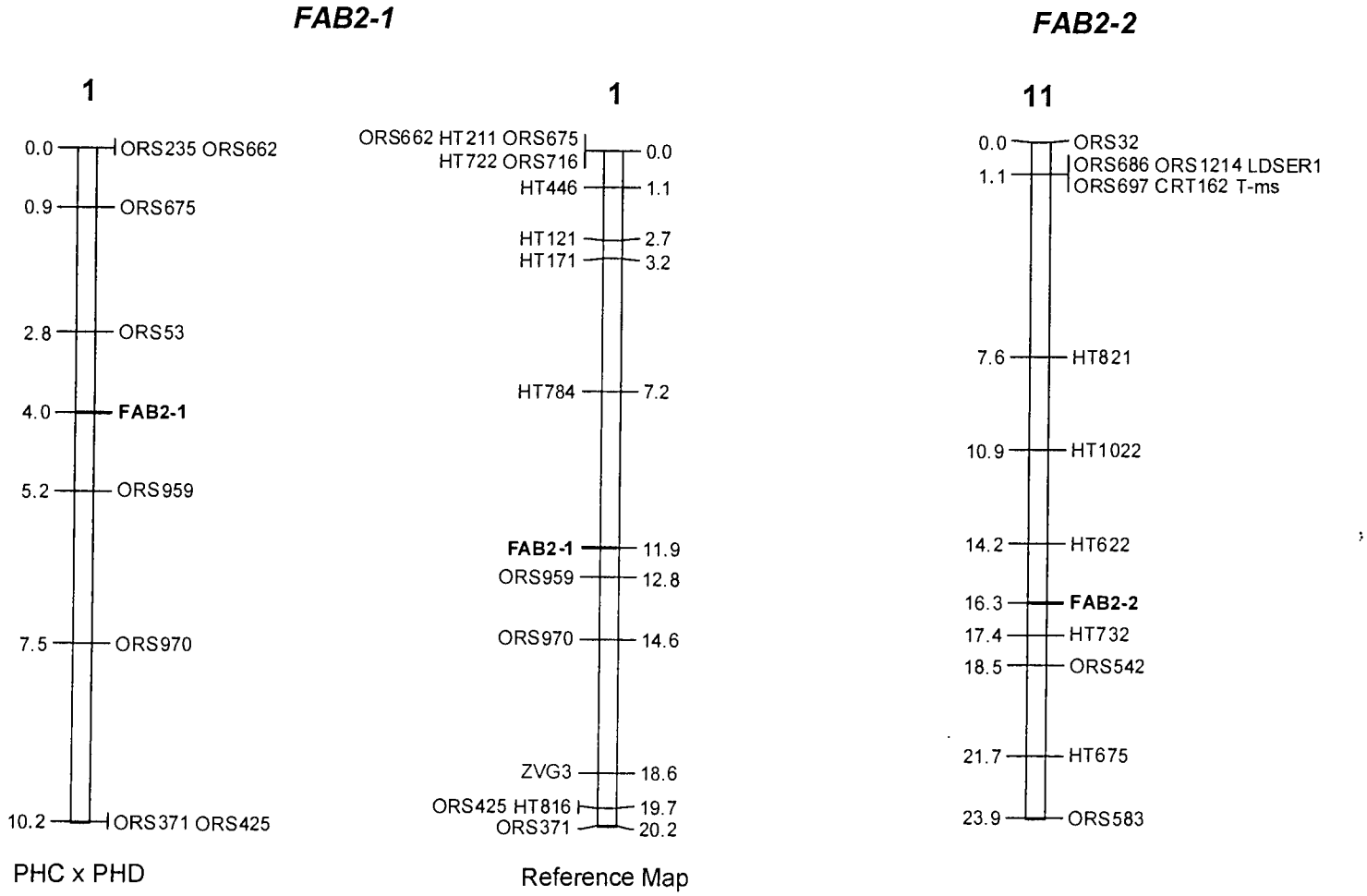


Figure 3.6. Position of *FAB2-1* and *FAB2-2* on LG 1 and LG 11, respectively. Position of *FAB2-1* in the PHC x PHD population (left) and in our reference map (right). Position of *FAB2-2* on LG 11 of our reference map (far right).



different allele as the latter eight lines, indicating that the *FAB2-1* allele is independent of the high oleic acid phenotype. *H. argophyllus* shows two additional bands compared to RHA801 and all other *H. annuus* lines screened, further demonstrating a great allelic diversity of the *FAB2-1* locus (Table 3.3).

The primers for *FAB2-2* are also flanking an intron, since the recovered genomic fragment is about 1.2 kb in length compared to the expected 597 bp from cDNA. The *FAB2-2* locus is overall less polymorphic than *FAB2-1*, which was reflected in a lower heterozygosity score (Figure 3.5). The SSCP assay reveals only two distinct alleles (Table 3.3). *FAB2-2* was mapped to LG 11 in NMS373 x ANN1811 and genotyped on the 262 RIL of PHC x PHD (Figure 3.6). PHC and PHD were the only high versus low oleic acid comparison displaying allelic variation at this locus. The 3 PILs HA89-HA341, HA292-HA349 and RHA274-RHA345 each carried identical alleles at this locus.

KASI

The CGPDB EST database contained one relevant hit to *KASI*, contig 1997 (Table 3.1). The GenBank search did not produce any sunflower *KASI* matches. When we compared the sequence of contig 1997 with the corresponding *A. thaliana KASI* locus AT5G46290, it revealed that the contig sequence comprised 5 of the 7 exons in addition to the 3'UTR of the Arabidopsis homologue. This allowed us to select primers flanking different introns in order to increase the likelihood to recover polymorphism. Primers KASI-1, designed to the 5' part of the contig sequence, produced one main band of 547 bp and several larger, faint bands on agarose, while primers KASI-2, designed to the 3' part of the contig, produced a single, monomorphic band of 1,065 bp on agarose (Table 3.3). The latter fragment was screened on SSCP revealing 4 distinct alleles on the panel of 15 lines (Figure 3.7). While the low oleic lines HA89, RHA274 and PHD shared the same allele as their high oleic counterparts HA341, RHA345 and PHC, low oleic line HA 292 carried the same allele as the high oleic line HA349. Thus each low and high oleic pair carried the same allele, but alleles differed between pairs. The different alleles in RHA280 and

Figure 3.7. Screen for SSCP polymorphism in *KASI*, *KASII* and *KASIII* on a panel of 15 sunflower lines.

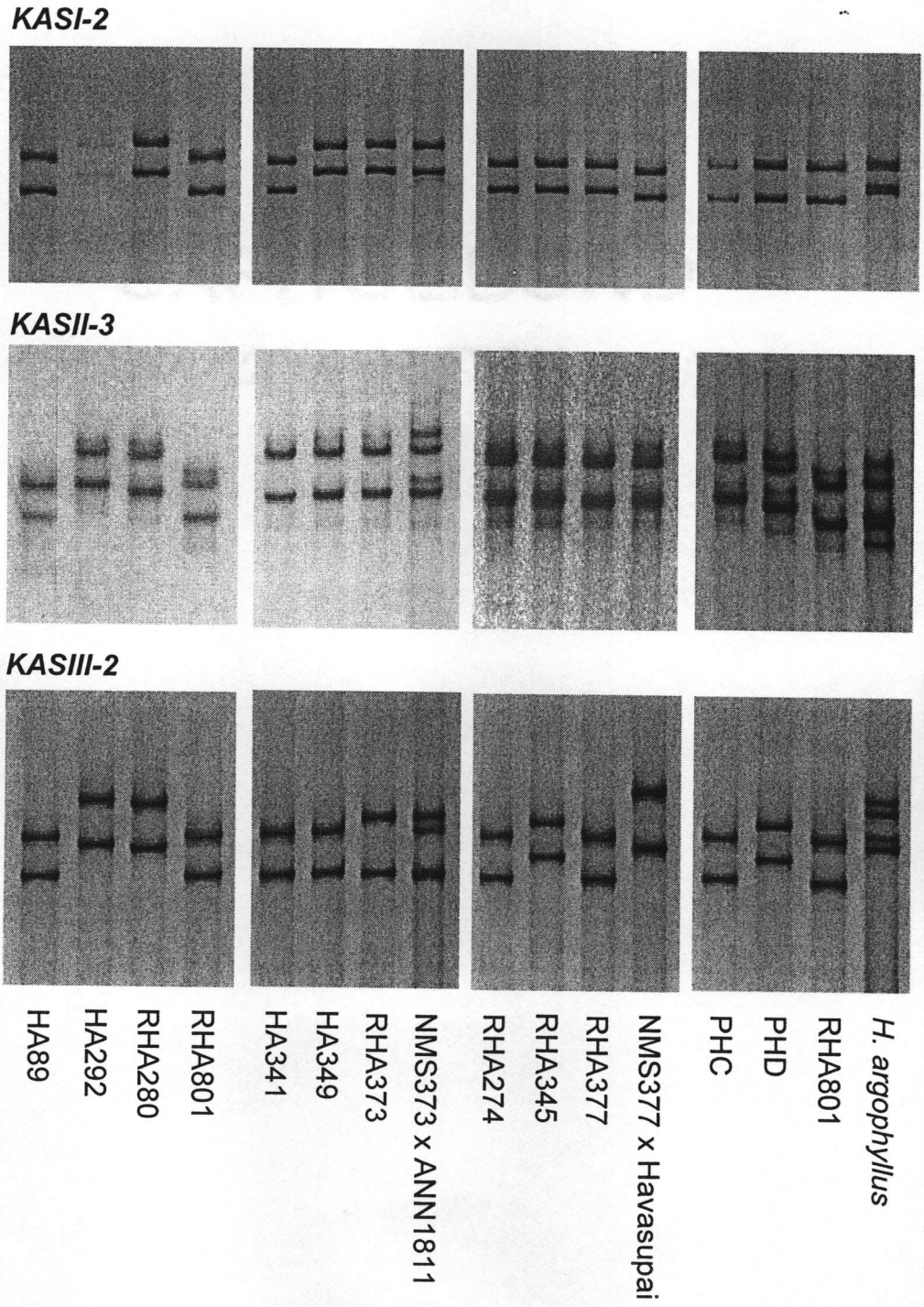
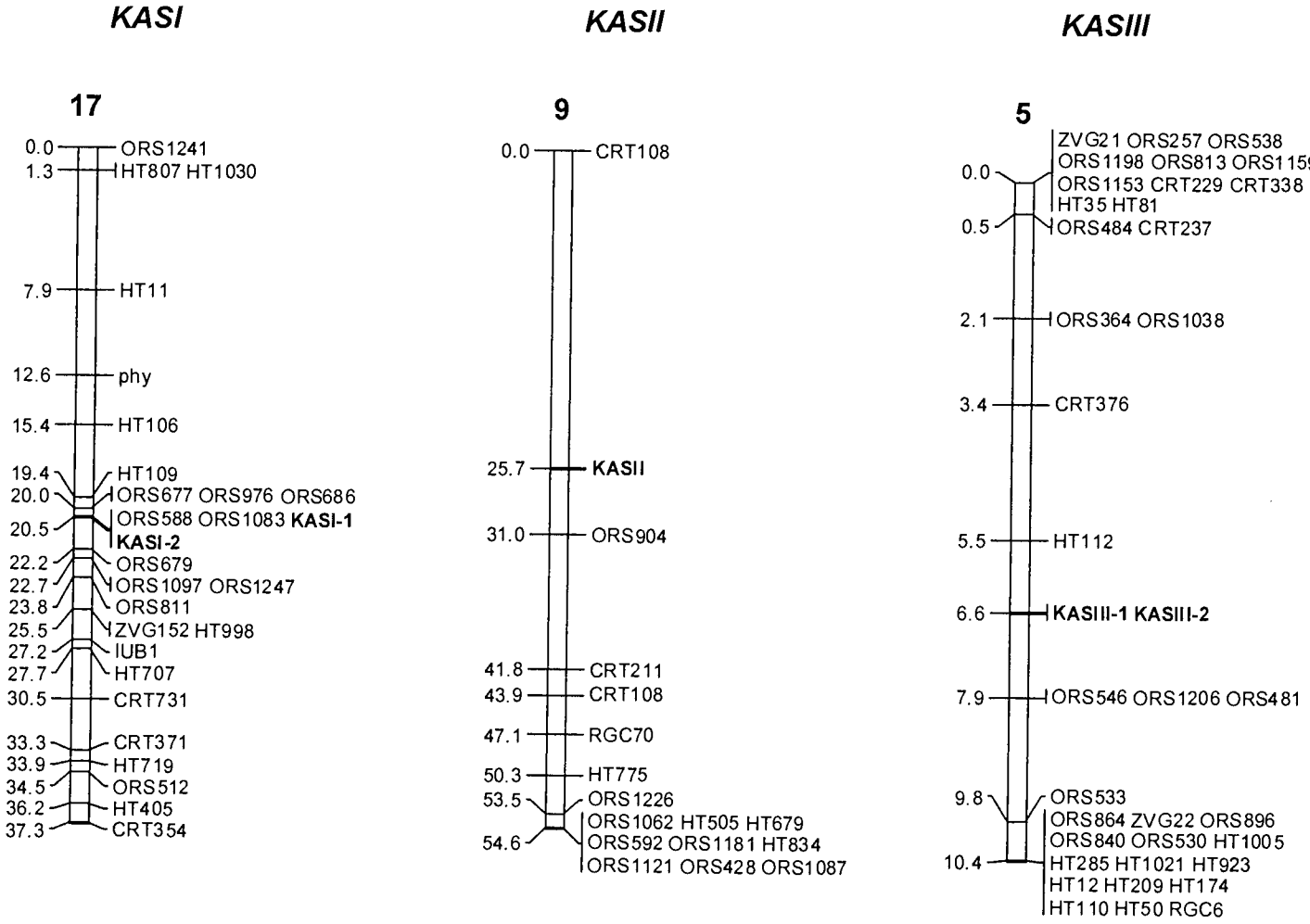


Figure 3.8. Position of *KASI*, *KASII* and *KASIII* on the LGs 17, 9 and 5 of our reference map, respectively.



RHA801 allowed us to map the gene to LG 17 (Figure 3.8). The bulked BC_1 progeny different allele as the latter eight lines, indicating that the *FAB2-1* allele is independent of NMS373 x ANN1811 carried an additional faint lower band, the allele from the wild species ANN1811, when compared to the elite recurrent parent RHA373 (Figure 3.7). The fourth allelic variant was only observed in *H. argophyllus*.

Due to the lack of polymorphisms in *KASI* between PHC and PHD we subsequently isolated and sequenced the main fragment obtained by KASI-1 and the fragment produced by the KASI-2 primers. There was no single polymorphic nucleotide in 1,631 bp of sequence between the two lines (Genbank AY805140-AY805143) (Appendix D). Thus, we decided to use the flanking marker approach for PHC and PHD, using the known location of *KASI* on our reference map to select flanking markers. The region of LG 17, which comprised *KASI*, was extraordinarily conserved between the two lines PHC and PHD. We screened 15 single locus microsatellite markers and 1 SNP marker covering an area of 34.9 cM around the *KASI* locus without recovering a single polymorphism, with the closest polymorphic marker (HT1030) being 19.2 cM distant to *KASI* (Figure 3.8). Thus, we were unable to include *KASI* in our candidate gene QTL analysis for the segregating RIL population.

KASII

The kernel EST database contained two relevant hits to *KASII*, deposited in Genbank as AY805138 and AY805139 (Table 3.1). GenBank and CGPDB EST database searches did not produce any sunflower *KASII* matches. The two sequences showed a substantial overlap of 373 bp with a sequence identity of 99.2% when aligned. Primers KASII were designed to cover 306 bp of the shared sequence (Table 3.3). The primers amplified a single ~500 bp fragment from genomic DNA of the 15 sunflower lines on the panel, which was monomorphic on agarose. SSCP analysis revealed 5 distinct alleles for *KASII* among the 15 lines (Figure 3.7) (Table 3.3). While HA292 and RHA274 shared the same allele as their high oleic counterparts HA349 and RHA345, the low oleic lines HA89 and PHD carried a distinctly different allele than the corresponding high oleic lines HA341 and PHC. The allelic variation among

PHC and PHD was used to genotype *KASII* in the 262 RILs. Even though RHA280 and RHA801 carried different alleles, the resulting polymorphism did not map to any of the 17 linkage groups in sunflower. The different alleles in RHA373 and NMS373 x ANN1811 allowed us to map *KASII* to LG 9 (Figure 3.8). Again, we observed an additional discrete allele in *H. argophyllus*.

KASIII

The CGPDB EST database contained two hits to *KASIII*, contig 1814 and singleton CF079060 (Table 3.1). The GenBank search did not produce any additional sunflower *KASIII* matches. When we compared the two sequences with the corresponding *A. thaliana KASIII* locus AT1G62640, it showed that the sequences comprised only part of the gene covering at least two introns, but lacking 5' and 3' coding sequence and UTRs of the Arabidopsis homologue. Primers KASIII-1 recovered a single, large ~1,700 bp fragment from genomic DNA compared to the expected cDNA size of 425 bp. Primers KASIII-2 produced a single, intron including fragment of about 860 to 890 bp, since the predicted length from cDNA was 206 bp (Table 3.3). The differences in fragment length for *KASIII-2* were not substantial enough, to map the fragment on agarose. The subsequent SSCP analysis showed at least 6 distinct alleles for *KASIII-2* among the 15 lines, displaying the highest heterozygosity of all screened loci (Figure 3.7). Low oleic lines HA89 and RHA 274 carried the same allele as high oleic lines HA341, HA349 and PHC, while the high oleic line RHA345 shared alleles with low oleic lines HA292 and PHD. The different alleles in RHA280 and RHA801 were again used to map this locus in our reference population of 94 F_7 RILs to LG 5 (Figure 3.8). The SSCP analysis of *KASIII-1* displayed the same haplotypes as *KASIII-2* and thus confirmed the location of *KASIII* by mapping to the same locus. Furthermore, polymorphisms were observed between the two BC_1 pools for NMS373 x ANN1811 and NMS377 x Havasupai when compared to their recurrent parent allele (Figure 3.7). As seen before with the other members of the *KAS* gene family, *H. argophyllus* displayed a distinctly different allele.

FATB

The GenBank search revealed a complete cDNA sequence for *FATB*, AF36565, while the CGPDB database search produced one significant hit, contig 6130 (Table 3.1). When we compared the two sequences with the corresponding *A. thaliana FATB* locus AT1G08510, it revealed that the Arabidopsis homologue carried 4 introns. This allowed us to select primers flanking different introns in order to increase the likelihood to recover polymorphism. Primers FATB A, FATB B, FATB C and FATB D produced each a single fragment and none revealed an agarose polymorphism (Table 3.3). All four fragments contained at least one intron, since fragment sizes were greater than expected from cDNA. Fragments A, B and C were also completely monomorphic among the 15 lines, when screened on SSCP (Figure 3.9). Only fragment D contained some limited allelic diversity. NMS373 x ANN1811 and *H. argophyllus* displayed two separate alleles (Figure 3.9). The additional band observed in the BC_1 bulk of NMS373 x ANN1811 allowed us to map *FATB* to LG 6 (Figure 3.10). This map position was in disagreement with the previously reported mapping of *FATB* to LG 7 (Pérez-Vich et al., 2002), thus, we designated the locus *FATB-2*. An RFLP-probe homologous to the GenBank sequence AF36565 was used to map *FATB* to LG 7 (Alberto Leon, personal communication). This locus is herein referred to as *FATB-1*.

We subsequently sequenced fragments A, B, C and D in all 15 lines. Low sequence quality in the forward and reverse sequencing pass led to the exclusion of fragment B in lines HA292 and RHA280 and fragments A, C and D in *H. argophyllus*. The obtained high quality sequences showed perfect homology with AF36565 and revealed a total of 6 introns in the sunflower *FATB* gene. Both sunflower and Arabidopsis contain an intron in the 5' UTR of *FATB*. The locations of introns 2, 3, 4 and 5 in the open reading frame are conserved between Arabidopsis and sunflower. The sunflower *FATB* gene carried an additional sixth intron in the last exon of the Arabidopsis homologue (Appendix E). The 3,026 bp of genomic sequence of the sunflower *FATB* did not reveal a single nucleotide polymorphism in neither exonic nor

Figure 3.9. Screen for SSCP polymorphism in the *FATB* gene on a panel of 15 sunflower lines.

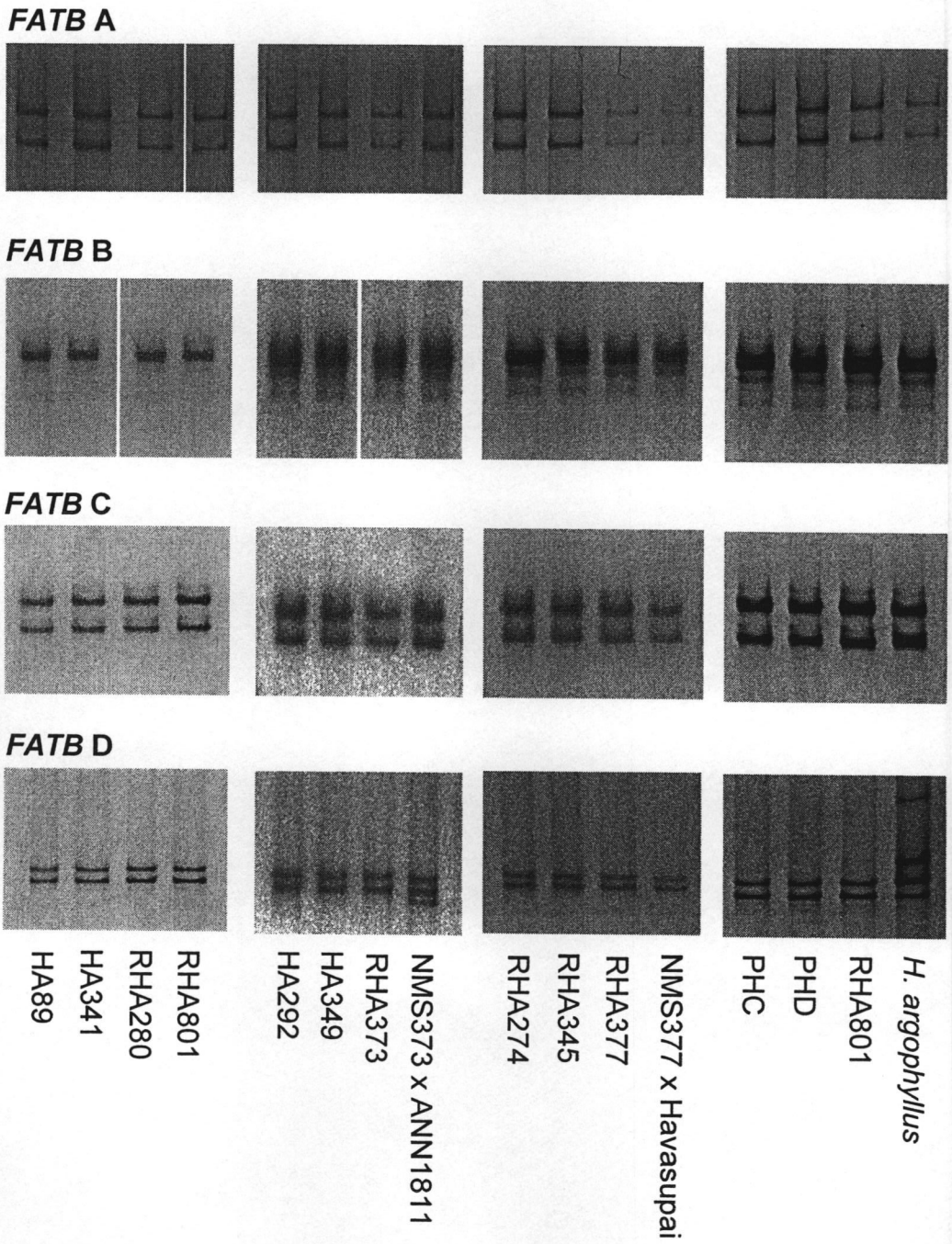
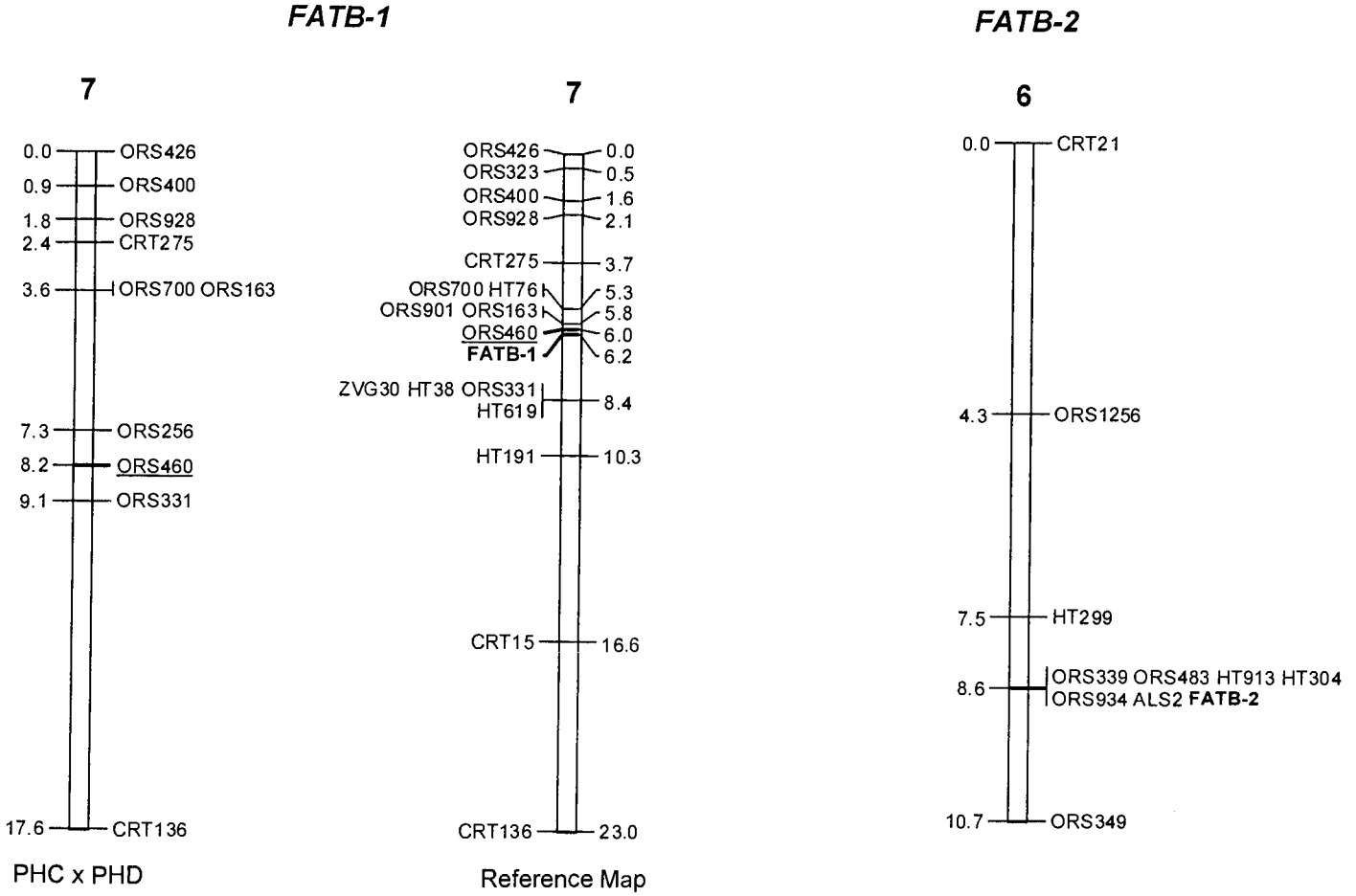


Figure 3.10. Position of *FATB-1* and *FATB-2* on LG 7 and LG 6, respectively. Position of *FATB-1* and flanking marker ORS460 (underlined) on linkage group 7 of our reference map (middle) and position of the flanking marker on linkage group 7 of PHC x PHD (left). Position of *FATB-2* and flanking marker ORS1256 on linkage group 6 of our reference map (far right).



intronic sequence, demonstrating a significant degree of sequence conservation for this gene across this diverse panel of genotypes (GenBank AY803009-AY803019) (Appendix E). In addition, the subsequent sequencing confirmed that the fragment mapped to LG 6 in NMS373 x ANN1811 was identical to the GenBank sequence and thus can be considered a potential *FATB* homologue. Therefore, we had to resort to the flanking marker approach for both *FATB-1* and *FATB-2*. ORS460, 0.2 cM apart from *FATB-1*, was the closest flanking marker available and proved polymorphic, when the microsatellitewas screened on PHC and PHD (Alberto Leon, personal communication) (Figure 3.10). The closest polymorphic marker in PHC and PHD for *FATB-2* was microsatellite ORS1256, 4.3 cM apart from *FATB-2* (Figure 3.10).

FATA

The kernel EST library search yielded two hits for *FATA*, deposited in GenBank as AY805124 and AY805125, while the CGPDB EST database search produced one significant hit, contig 2357 (Table 3.1). The two sequences from the kernel library showed complete overlap for all 638 bp with a sequence identity of 99.2 %. When aligned with contig 2357, the three sequences displayed substantial overlap for 486 bp in the 3' region with a sequence identity of 97.3%. The first 179 bp of contig 2357 did not overlap with the other two ESTs, which resulted in a sequence identity of only 34.7%. Primers FATA-1 were designed to the cover 180 bp of the shared sequence and recovered the same size fragment from genomic DNA (Table 3.3), where as primers FATA-2 spanned another 280 bp and yielded a 1,221 bp fragment from genomic DNA. Together the primers FATA-1 and FATA-2 covered 446 out of the 486 bp of shared EST sequence. Both primers amplified across the 15 lines and the obtained fragments displayed no polymorphism on agarose.

The SSCP analysis of *FATA-1* revealed only two alleles. All 14 *H. annuus* lines carried allele one and allele two was only present in *H. argophyllus* (Figure 3.11). The fragment of *FATA-2* contained three different alleles (Figure 3.11). While HA89 carried the same allele as PHC and PHD, HA341 shared alleles with HA292, HA349, RHA274 and RHA345. Thus, HA89 and HA341 was the only low by high oleic

Figure 3.11. Screen for SSCP polymorphism in the *FATA* gene on a panel of 15 sunflower lines. Top panel: SSCP screen for *FATA-1* Lower panel: SSCP screen for *FATA-2.*

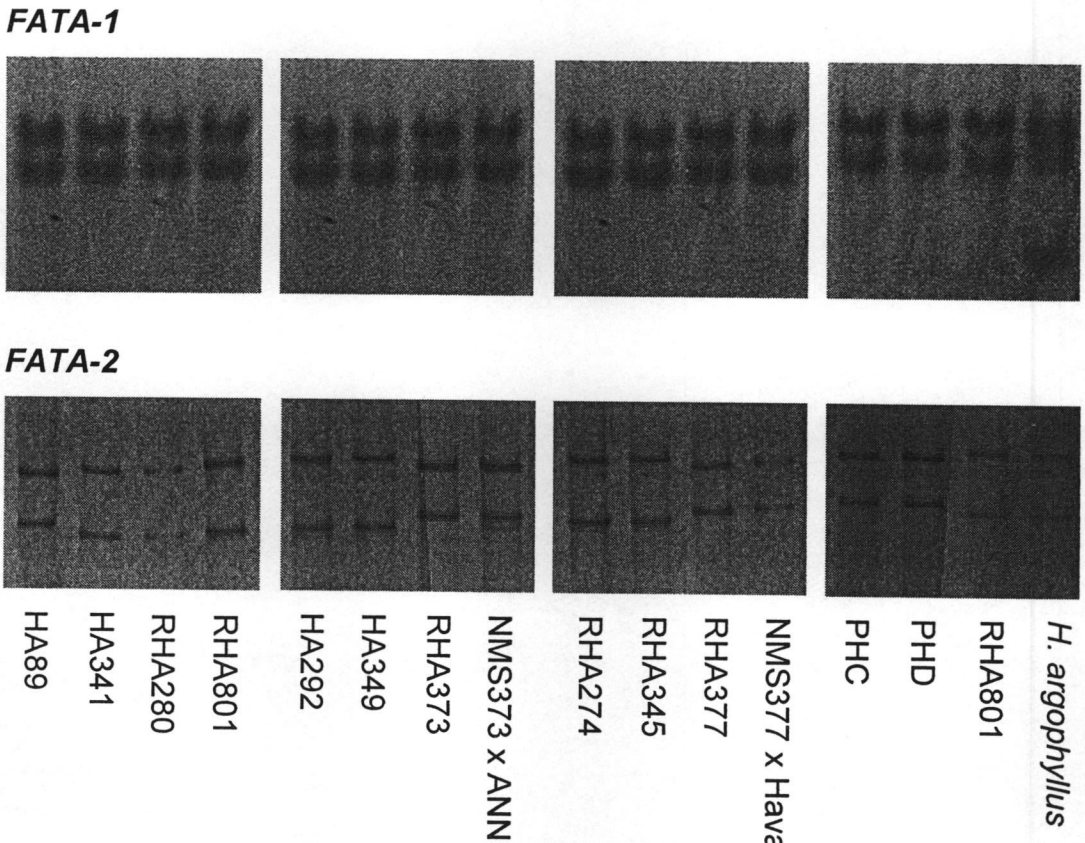
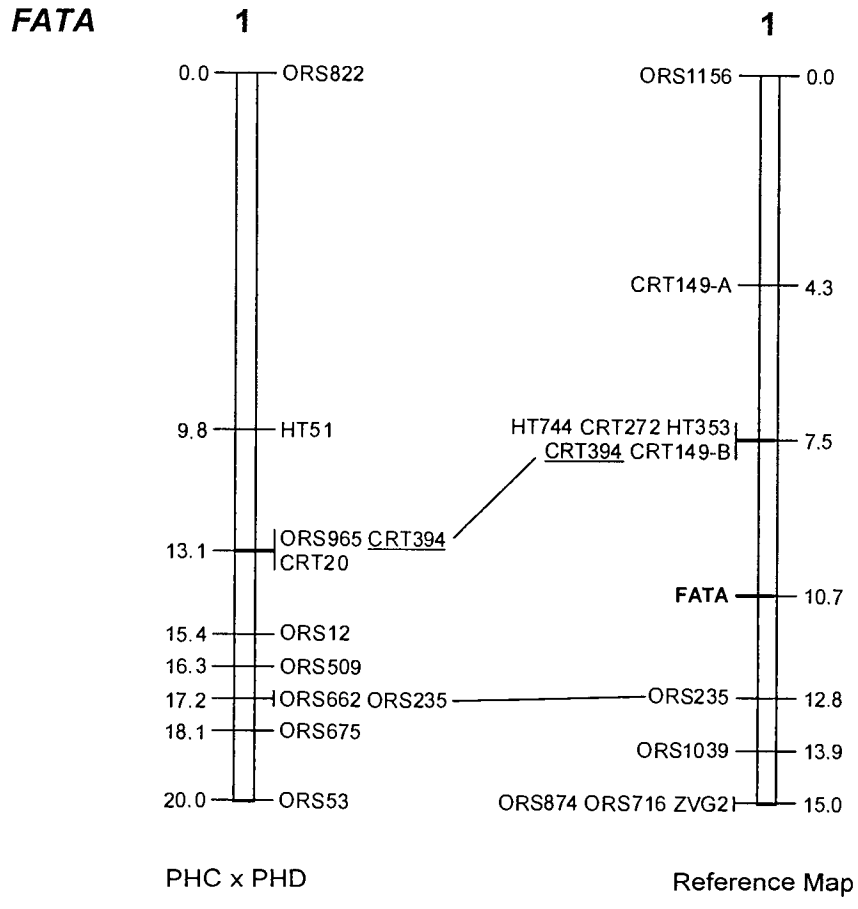


Figure 3.12. Position of *FATA* and flanking marker CRT394 on LG 1 of our reference map and on LG 1 of PHC x PHD. The flanking marker CRT394 is underlined. The map of PHD x PHC is on the left, the map of our reference population on the right.

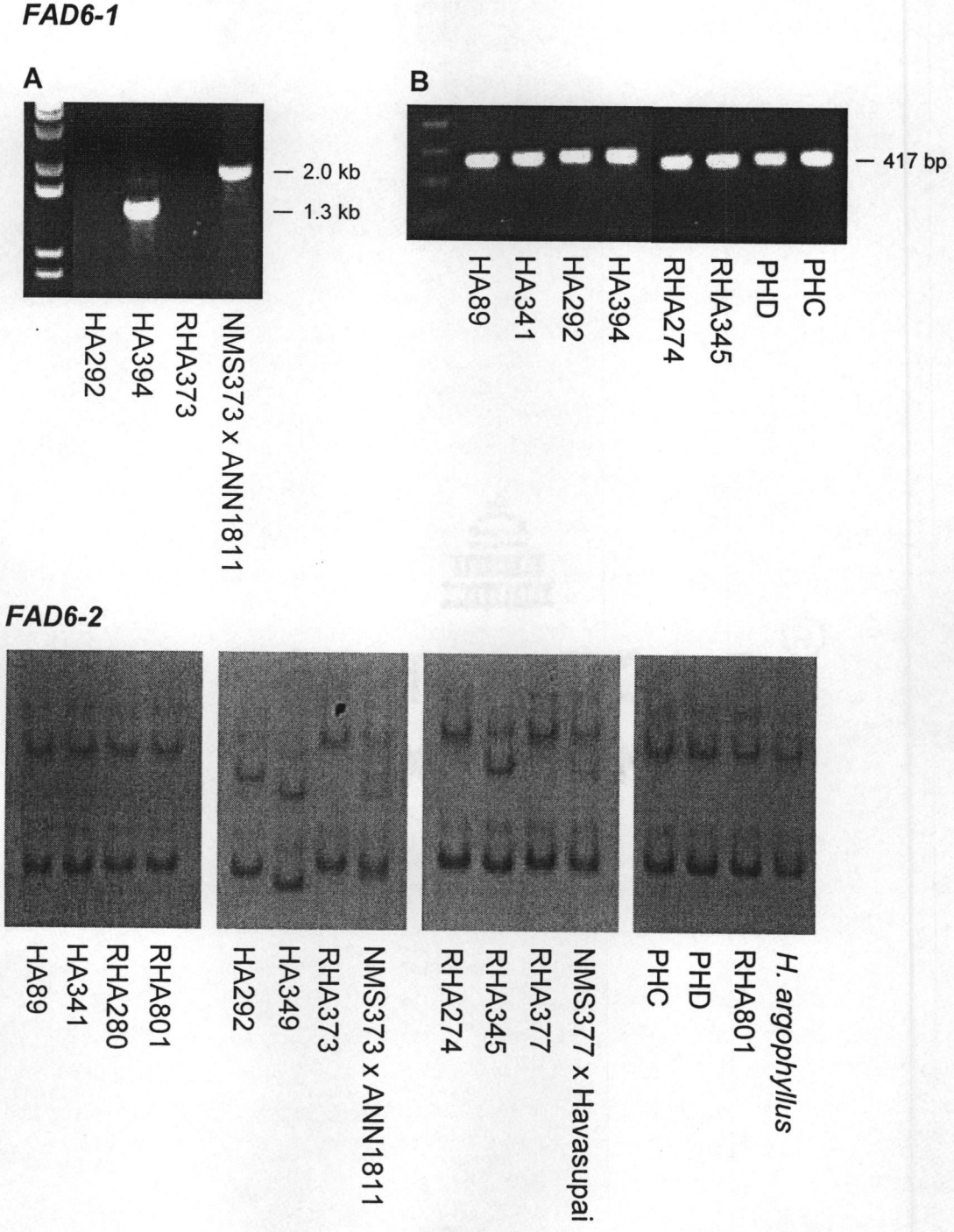


comparison showing allelic variation at this locus. The third allele was only present in NMS373 x ANN1811, allowing us to map this polymorphism in *FATA-2*, when compared to the recurrent parent RHA373, to LG 1 in this reference population (Figure 3.12). Since both fragments were monomorphic between PHC and PHD, we subsequently sequenced *FATA-2*. The high degree of sequence conservation revealed by previous sequencing led us to the conclusion that the exonic fragment *FATA-1* was highly unlikely to contain any useful polymorphism. Therefore we did not sequence the 180 bp fragment. However, we observed two distinct haplotypes in the 1221 bp fragment in 13 sunflower lines. No quality sequence data was available for NMS373 x ANN1811 and NMS377 x Havasupai. Haplotype one was comprised of lines HA89, RHA373, RHA377, PHC and PHD (GenBank AY805126-AY805130), while haplotype two consisted of lines HA341, HA292, HA349, RHA274, RHA345, RHA280, RHA801 and *H. argophyllus* (GenBank AY805131-AY805137) (Appendix F). This corresponds to the allelic diversity observed in the SSCP analysis. Since PHC and PHD revealed the same haplotype, we had to resort to the flanking marker approach. The closest flanking marker polymorphic in PHC x PHD was microsatellite CRT394, 3.2 cM apart from *FATA-2* (Figure 3.12).

FAD6

The CGPDB EST database contained one hit to *FAD6*, contig 1636 (Table 3.1). The GenBank search did not produce any sunflower *FAD6* matches. When we compared the sequence of contig 1636 with the corresponding *A. thaliana FAD6* locus AT4G30950, it revealed that the contig sequence comprised 5 of the 10 exons present in the Arabidopsis homologue. This allowed us to target the introns in our primer design process to increase the likelihood to recover polymorphism. Primers FAD6-1, designed to cover 417 bp of the 3' part of the contig sequence, amplified in only two lines out of the 15 on the panel. The primers produced a single band of about 1.3 kb in HA349 and a single band of about 2 kb in NMS373 x ANN1811 on agarose (Figure 3.13). We performed RT-PCR on RNA extracted from developing kernels at 18 DAF from lines HA89, HA341, HA292, HA349, RHA274, RHA345, PHC and PHD using

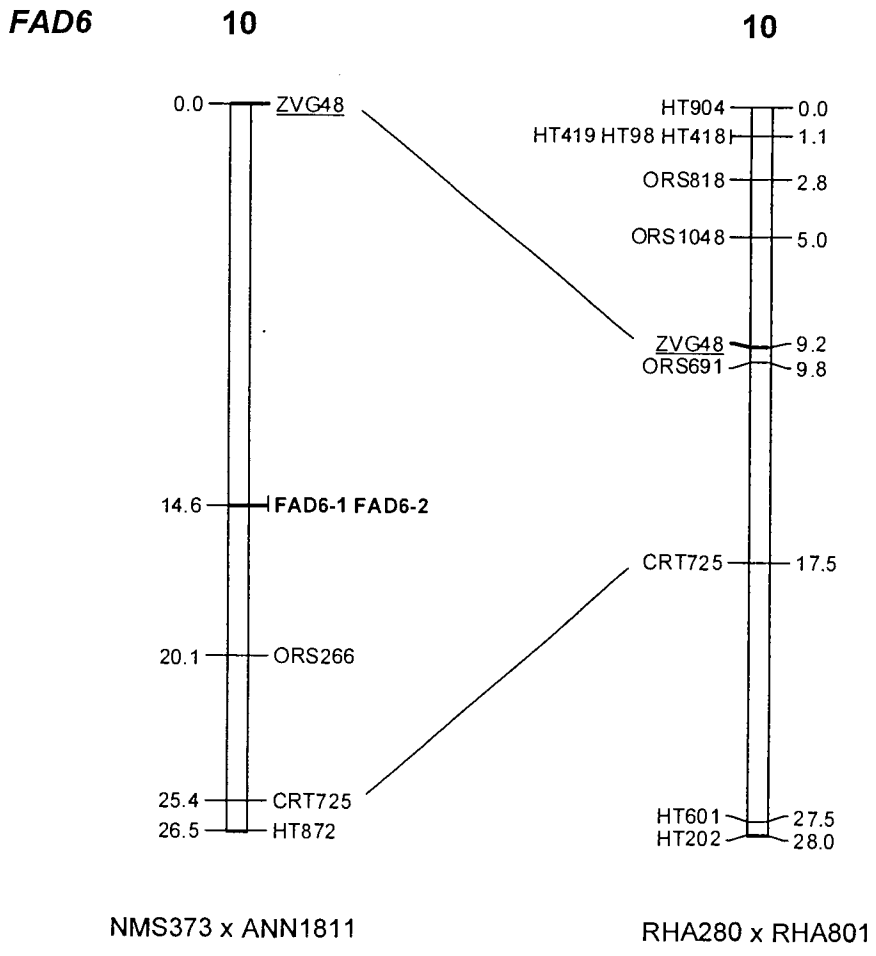
Figure 3.13. Screen for polymorphism in the *FAD6* gene. *FAD6-1* (top panel): A) Agarose polymorphism for *FAD6-1*on genomic DNA. B) RT-PCR of *FAD6-1* on a panel of 8 sunflower lines. *FAD6-2* (lower panel): SSCP screen for *FAD6-2* on a panel of 15 sunflower lines.



primers FAD6-1. All eight lines amplified the expected 417 bp fragments, confirming the presence of a very large intron in this part of *FAD6* (Figure 3.13). Despite several long distance PCR attempts we were unable to recover a fragment in any of these lines. In order to confirm that the bands observed in HA349 and NMS373 x ANN1811 were indeed from *FAD6* we sequenced the two fragments. The sequencing did not only reveal that the fragments contained sequence homologous to the 417 bp of the EST, but also that the sequence contained a second smaller intron of 105 bp, showing that intron locations were conserved when compared to the Arabidopsis homologue. The lack of PCR amplification with the primers FAD6-1 on genomic DNA of RHA373 allowed us to map *FAD6-1* to LG 10 in NMS373 x ANN1811 (Figure 3.14).

The primers FAD6-2, designed to the 5' part of the contig, produced a single, monomorphic band of 1,948 bp in all lines on agarose (Table 3.3). SSCP analysis revealed 5 distinct alleles among the 15 lines (Figure 3.13). HA89 shared alleles with HA341, RHA274, PHC and PHD as well as RHA280, RHA801, RHA373, RHA377 and *H .argophyllus*, while HA292 carried the same allele as RHA345. HA349 carried a distinct allele different from all other oilseed lines. Unique alleles were also present in NMS377 x Havasupai and NMS373 x ANN1811. The latter one was used to map *FAD6-2* to the same position on LG 10, confirming the previously obtained results with *FAD6-1*. Due to the monomorphic SSCP results for PHC and PHD, we subsequently sequenced the 1.9 kb fragment in the 15 lines. The fragment contained two separate introns of 71 and 1,361 bp length, which was expected from the comparison with the Arabidopsis homologue. No quality sequence data was available for NMS373 x ANN1811 and *H. argophyllus*. Among the remaining 13 lines we observed four distinct haplotypes. HA89, HA341, PHC, PHD, RHA280, RHA801 shared the same haplotype configuration (GenBank AY805150-AY805155), while RHA274, RHA373, RHA377 and NMS377 x Havasupai comprised haplotype two (GenBank AY805145-AY805147) (Appendix G). These two haplotypes were fairly similar with no sequence differences for most of the 1,948 bp. HA292 and RHA345 belonged to haplotype three (GenBank AY805148-AY805149), and HA349 formed

Figure 3.14. Position of *FAD6* and flanking marker ZVG48 on our reference maps. *FAD6* and flanking marker ZVG48 (underlined) on LG 10 of the BC_1 population of NMS373 x ANN1811 (left) and position of the flanking marker on LG 10 of our second reference map of RHA280 x RHA801 (right).



the remaining haplotype four (GenBank AY805144) (Appendix G). The SSCP analysis was not able to distinguish between haplotype one and two, but clearly separated the remaining haplotypes. Since PHC and PHD both belonged to the same haplotype we had to resort to the flanking marker strategy. The closest polymorphic marker between PHC and PHD was INDEL ZVG48 (fad6map), which served as a flanking marker in the QTL analysis.

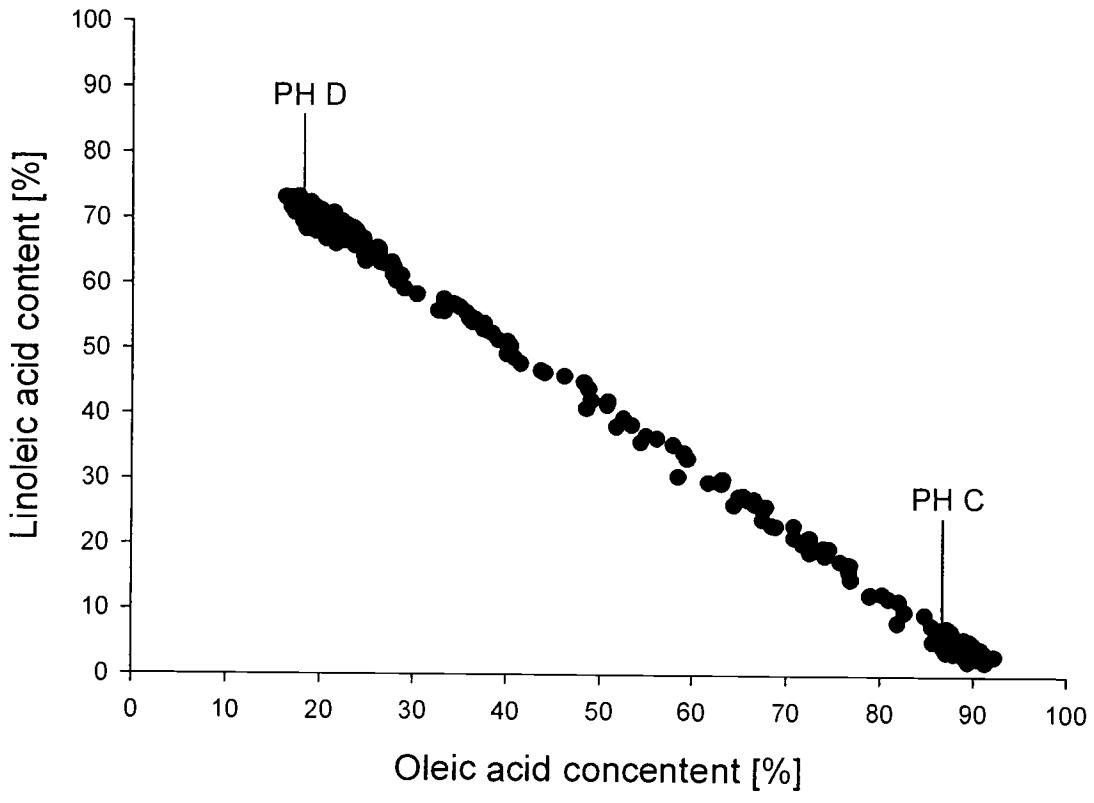
Fatty Acid Profiles of 262 Recombinant Inbred Lines

The fatty acid composition in the mature kernels of 262 recombinant inbred lines (RILs) and the two parental lines, PHC and PHD, was quantified using gas chromatography (GC). The phenotypic data consisted of four independent fatty acid measurements for each RIL, one collected from the field trial conducted in each location (Corvallis, OR, and Woodland, CA) and year (2000 and 2001). PHC produced an average of 85.1% (Corvallis, 2001) to 87.4% (Woodland, 2001) oleic acid and 4.7% (Woodland, 2001) to 7.1% (Corvallis, 2001) linoleic acid, whereas PHD contained on average between 17.8% (Corvallis, 2000) and 24.3% (Woodland, 2000) oleic acid and 66.1% (Woodland, 2000) to 73.5% (Corvallis, 2000) linoleic acid. The oleic acid content of the RILs ranged from 16.4% to 92.2% and the linoleic acid content varied between 2.2% and 73.2% (Figure 3.15), when based on the least square mean of the two years and two locations.

Effects of the Candidate Genes on Oleic Acid Content

Next, we combined the phenotypic data of the 262 RILs with the genotypic data of the ten candidate gene markers. Only one of the ten markers was dominant, the other nine markers were co-dominant. The fact that the 262 F_5 RILs were not fully homozygous at all loci sat limitations on the statistical analysis of the data. Even if the RILs were devoid of any heterozygosity, the number of possible genotypes (2^{10}) was still greater than the number of available RILs (1,024 vs. 262). We reduced the number of RILs to 217 for the analysis by eliminating all lines with missing genotypic data points. The 217 RILs contained 205 unique genotypes. Since the dataset did not

Figure 3.15. Distribution of the least square means of oleic acid and linoleic acid content in percent for the 262 RILs. The location of the parents PH C and PH D is indicated by the respective lines.



cover all possible genotypes, it was impossible to obtain estimates for all genotypic classes. Therefore we were not able to analyze the complete model including all main and interaction effects. Instead we used different model selection procedures to obtain the most informative and explanatory model. In order to keep the estimable effects to a reasonable size, we focused on main and two-way interaction effects. We only considered those effects for inclusion in our model, which proved highly significant in the different selection processes.

The phenotypic data consisted of four independent oleic acid measurements for each RIL, one collected from the field trial conducted in each location (Corvallis, OR, and Woodland, CA) and year (2000 and 2001). First, we evaluated if either the location or the year had an effect on the phenotype. The analysis revealed no differences between the two years, in which the field trials were conducted. But the analysis showed a significant difference between the oleic acid content of the RILs grown in Corvallis and Woodland. Further analysis exhibited no significant genotype by environment interactions, tested as gene x location and gene x gene x location interaction effects. Therefore we were able to obtain a location independent model for the high oleic phenotype. The heritability of oleic acid content was $\hat{h} = 0.96$. In addition, we calculated the genetic correlations between oleic acid content and the other three main fatty acids in the seed. The genetic correlation (r_G) between oleic acid and palmitic acid was -0.74, while r_G was -0.85 for stearic acid. The genetic correlation between oleic and linoleic acid content was -1.00.

Seven of ten candidate gene loci had significant intralocus effects on oleic acid content. In addition, nine two-way interactions were significant. Specifically, the significant intralocus effects were *FAD2-1*, ORS460 (*FATB-1*), ORS1256 (*FATB-2*), *KASIII*, *FAB2-1*, *FAB2-2* and CRT394 (*FATA*) and the significant epistatic effects were *FAD2-1* × ORS460 (*FATB-1*), ORS460 (*FATB-1*) × *KASII*, ORS460 (*FATB-1*) × *KASIII*, ORS460 (*FATB-1*) × *FAB2-2*, *KASII* × *KASIII*, *KASII* × ORS1256 (*FATB-2*), *KASII* × CRT394 (*FATA*), *KASIII* × ORS1256 (*FATB-2*) and *FAB2-1* × *FAB2-2* (Table 3.4) and accounted for 73% of the phenotypic variance based on the R^2. Even though *KASII* was not identified as a significant factor in the model selection process,

Table 3.4. Main and two-way interaction effects included in the model for the high oleic acid phenotype. DF = Degrees of freedon, SS = Sums of squares.

Source	DF	SS	F-value	Pr < F	R^2
FAD2-1	1	162243	724.4	<.0001	0.257
ORS460 (*FATB-1*)	2	15309	34.2	<.0001	0.024
ORS1256 (*FATB-2*)	2	2274	5.1	0.0065	0.004
KASII	2	104	0.2	0.7931	0.000
KASIII	2	6048	13.5	<.0001	0.010
FAB2-1	2	1434	3.2	0.0412	0.002
FAB2-2	2	1517	3.4	0.0343	0.002
CRT394 (*FATA*)	2	1914	4.3	0.0143	0.003
FAD2-1 x ORS460 (*FATB-1*)	2	9689	21.6	<.0001	0.015
ORS460 (*FATB-1*) × *KASII*	4	10891	12.2	<.0001	0.017
ORS460 (*FATB-1*) × *KASIII*	4	18638	20.8	<.0001	0.030
ORS460 (*FATB-1*) × *FAB2-2*	4	2283	2.6	0..0381	0.004
ORS1256 (*FATB-2*) × *KASII*	4	7236	8.1	<.0001	0.011
ORS1256 (*FATB-2*) × *KASIII*	4	7712	8.6	<.0001	0.012
KASII × *KASIII*	4	3990	4.5	0.0015	0.006
KASII × CRT394 (*FATA*)	4	7761	8.7	<.0001	0.012
FAB2-1 × *FAB2-2*	4	4084	4.6	0.0012	0.006

we included *KASII* as a main effect in the model, since four of the nine two-way interactions included *KASII*. The minimal impact of the *KASII* main effect was confirmed in the analysis and resulted in a p-value of 0.7931 (Table 3.4).

The most pronounced effect was observed for the *FAD2-1* gene. The additive effect for an allele substitution at the *FAD2-1* locus was estimated as a 20.7% change in oleic acid content (Table 3.5). The second largest main effect on oleic acid content was observed in ORS460 (*FATB-1*). The estimated additive effect of an ORS460 (*FATB-1*) allele substitution was –13.3%, i.e. a line homozygous at the ORS460 (*FATB-1*) locus carrying the PHC allele produced 26.5% less oleic acid on average than a comparable line homozygous for the PHD allele (Table 3.5). The additive effect of *KASIII* was –5.3%, indicating that the allele transmitted by the low oleic parent (PHD) caused an increase in oleic acid content.

ORS1256 (*FATB-2*) was the only allele besides *FAD2-1* from the high oleic parent PHC with an additive effect responsible for an increase in oleic acid content (Table 3.5). The additive effect of ORS1256 (*FATB-2*) was 4.1%. The other three significant main effects all showed that the low oleic line PHD was the source of the favorable allele.

FAB2-1 and *FAB2-2* displayed almost identical additive effects in the analysis. Both genes were responsible for a 2.7% and 3.0% drop in oleic acid content if the allele was derived from PHC. The additive effect of CRT394 (*FATA*) was slightly lower with –2.4%. Besides the earlier discussed insignificant main effect of *KASII*, *FAB2-1* and CRT394 (*FATA*) were the only two main effects found to have estimated effects only significant at the 0.05 probability level (Table 3.5).

The largest additive × additive effect was observed for the two-way interaction between the *FAD2-1* and the ORS460 (*FATB-1*) alleles, which was estimated at -3.8%, further emphasizing the importance of these two genes in regard to the oleic acid content (Table 3.5). The interaction between the ORS460 (*FATB-1*) and the *KASIII* allele caused the second largest additive × additive effect, estimated at 2.7%. The ORS460 (*FATB-1*) allele was present in a third interaction with *FAB2-2* and a fourth interaction term with *KASII*, with allele substitution effects of 1.5% and 1.1%,

Table 3.5. Additive and additive x additive effects for oleic acid in PHC x PHD.

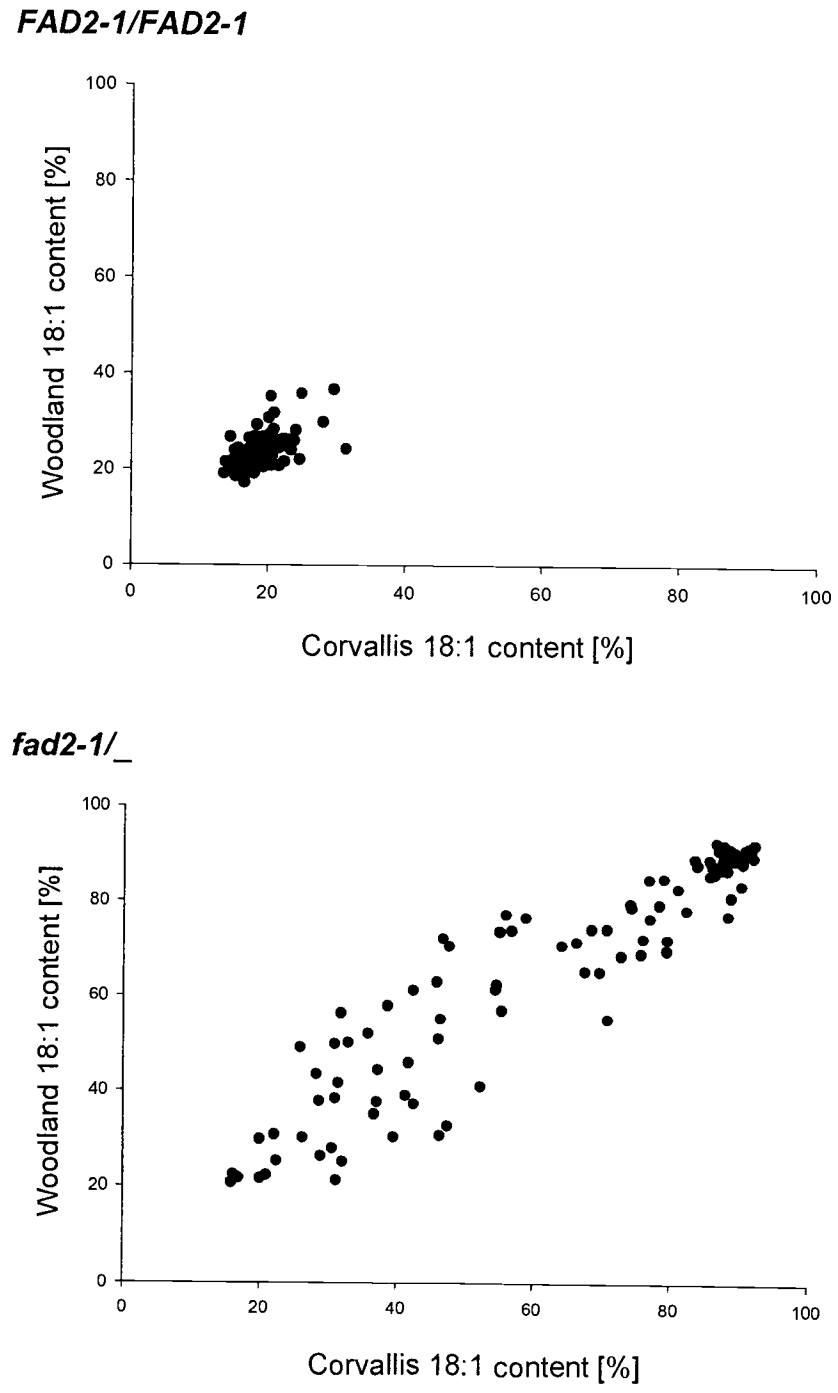
Main effects	Genotype		Effect	p-value
	C/C	D/D	α_k	
FAD2-1	65.8	23.8	20.7	<0.0001
ORS460 (*FATB-1*)	36.0	50.3	-13.3	<0.0001
ORS1256 (*FATB-2*)	40.1	44.7	4.1	0.0058
KASIII	40.4	45.5	-5.3	<0.0001
FAB2-1	41.1	43.3	-2.7	0.0122
FAB2-2	39.8	45.1	-3.0	0.0094
CRT394 (*FATA*)	44.1	41.0	-2.4	0.0483

Interaction effects	Genotype				Effect	p-value
	C/C	C/D	D/C	D/D	$\alpha_k \alpha_{k'}$	
FAD2-1 × ORS460 (*FATB-1*)	57.2	74.2	21.5	26.6	-3.8	<0.0001
ORS460 (*FATB-1*) × *KASII*	37.6	32.9	50.7	50.4	1.1	0.0825
ORS460 (*FATB-1*) × *KASIII*	34.0	37.4	45.4	54.4	2.7	<0.0001
ORS460 (*FATB-1*) × *FAB2-2*	36.7	34.8	46.0	53.9	1.5	0.0200
ORS1256 (*FATB-2*) × *KASII*	41.6	45.5	34.5	46.0	1.6	0.0130
ORS1256 (*FATB-2*) × *KASIII*	39.8	41.8	40.1	47.0	1.1	0.0863
KASII × *KASIII*	41.1	47.0	37.6	44.9	-1.5	0.0158
KASII × CRT394 (*FATA*)	43.4	44.1	45.5	37.4	-2.4	0.0001
FAB2-1 × *FAB2-2*	37.4	42.0	39.0	46.7	-2.3	0.0005

respectively. *KASII* was also present in significant two-way interactions with ORS1256 (*FATB-2*) with an estimated effect of 1.6%. The *KASII* × *KASIII* additive × additive effect was -1.5%, while the *KASII* × CRT394 (*FATA*) effect was -2.4% (Table 3.5). In addition to the significant interaction between *KASIII* and ORS460 (*FATB-1*), the 1.1% additive × additive effect of the *KASIII* × ORS1256 (*FATB-2*) interaction proved significant as well. Last, but nor least, the effect of the interaction between the two *FAB2* genes was significant, resulting in a -2.3% change in oleic acid.

The model clearly showed that the high oleic acid phenotype is driven by the actions and interaction of several genes in the glycerolipid biosynthetic pathway. While these genes and their epistatic interactions determined the amount of oleic acid in the seed, we found that *FAD2-1* has a paramount effect on the high oleic phenotype. This is best demonstrated by looking at the phenotypic distributions of the RILs carrying the wildtype allele from PHD (*FAD2-1, FAD2-1*) compared to the RILs with the PHC allele (*fad2-1,_*) (Figure 3.16). The two distributions clearly showed that the allelic constellation *fad2-1,_*, derived from the high oleic parent PHC, was absolutely necessary to obtain the high oleic phenotype, but was by no means fully sufficient. The allelic variation, which is revealed in the presence of the mutant *fad2-1* allele, is not visible in the lines carrying the wildtype allele *FAD2-1*.

Figure 3.16. Phenotypic distributions of the 262 RILs separated according to the allele present at the *FAD2-1* locus. The plot displays the oleic acid content of each RIL for the two locations Corvallis (x-axis) and Woodland (y-axis). *FAD2-1/FAD2-1* represents the lines carrying the allele from PHD, the lines with the allele from PHC are designated *fad2-1/fad2-1*.



DISCUSSION

Since the discovery of the high oleic mutation (Ol_1) (Soldatov 1976), sunflower breeders have observed a wide range of oleic acid distributions in populations segregating for Ol_1 and, from phenoytpic analyses, have proposed one, two, and polygene models for the regulation of oleic acid biosynthesis in sunflower kernels (Urie, 1985; Miller et al., 1987a; Fernández-Martínez et al., 1989; Fernandez et al., 1999, Lacombe and Bervillé, 2001, Lacombe et al., 2002a). Even in the segregating populations where simple genetic models have been formulated and tested, residual phenotypic variability has been present among progeny within phenotypic classes, and phenotypic classes for loci other than Ol_1 have often overlapped and been arbitrarily split into inferred genotypic classes (Fernández-Martínez et al., 1989; Fernandez et al., 1999, Lacombe and Bervillé, 2001, Lacombe et al., 2002a). These loci are more appropriately identified as QTL. Our hypothesis was that QTL identified across genetic backgrounds and independent analyses in sunflower, e.g., *ml*, Ol_2, Ol_3, and *supole*, could have been caused by one or more of the candidate gene loci targeted in our study (Miller et al., 1987a; Fernández-Martínez et al., 1989; Fernandez et al., 1999, Lacombe et al., 2002a). Because most of the earlier analyses were performed before candidate gene and QTL mapping approaches could be applied in sunflower, uncertainty surrounding the genetics of oleic acid has persisted, and no basis has existed for ascertaining the allelism of phenotypic loci or coincidence of QTL. Furthermore, until complete public genetic linkage maps were developed (Tang *et al.* 2002; Yu *et al.* 2003), comparative mapping of oleic acid QTL could not be performed.

By developing and screening sequence-based markers for several genes in the glycerolipid biosynthetic pathway, we have identified a series of candidates for *ml*, Ol_2, Ol_3, and *supole* and other oleic acid QTL in sunflower. The network of genes and allelic diversity of the QTL were exposed among progeny carrying the *FAD2-1* mutation (*fad2-1*/_). The phenotypic dispersion was negligible among wildtype homozygotes (*FAD2-1*/*FAD2-1*), but spanned the full phenotypic range among mutant

lines (Figure 3.16). We selected a cross that had repeatedly produced complex oleic acid distributions, primarily to increase the chance of identifying the maximum number of candidate genes possible in a single segregating population. While phenotypic complexity and gene number are not necessarily correlated (Falconer and Mackay, 1996; Lynch and Walsh, 1998), nearly every gene tested was found to significantly affect oleic acid concentration, either directly or indirectly.

Pérez-Vich et al. (2002) detected three loci on LGs 1, 8 and 14 in their QTL analysis of the high oleic acid trait. RFLP markers for *FAD2-1* and *FAB2-1* coincided with the QTL peaks found on LG 14 and LG 1, respectively. No association was found between the *FATB* locus on LG 7 and the high oleic acid content, which was a major factor in our analysis. In addition, Pérez-Vich et al. (2002) reported a minor QTL on LG 8. None of the candidate genes surveyed in our study map to LG 8. These findings indicate that the candidate genes tested in the present study play a prominent role in the fatty acid variability observed across the different genetic backgrounds in sunflower populations segregating for the *FAD2-1* mutation. The spectrum of genes regulating oleic acid variability undoubtedly varies across the genetic backgrounds as a function of allelic diversity present in the respective cross. It is the primary challenge in sunflower breeding programs to predict the oleic acid phenotypes of hybrids between low and high oleic or between high oleic inbred lines because of the latent genetic variability that is exposed in hybrids. The obvious, but not necessarily simple strategy for increasing predictability in mid- and high-oleic breeding programs is to reduce allelic diversity among QTL in both parents of a hybrid through MAS. But the feasibility of this approach is severely impacted by the number of loci targeted and the encountered lack of DNA polymorphism at these loci.

Whereas *FAD2-1* had previously been shown to be the major contributor to the high oleic acid phenotype, we knew that the mutation of the seed specific *FAD2-1* was necessary in order to obtain a high oleic line, it was not sufficient (Lacombe et al., 2002a, Pérez-Vich et al.; 2002). Instead, the analysis of the segregating RIL population revealed that a complex network of epistatic interactions was causing the phenotype. The highly significant main effect of *KASIII* and interaction effects

involving *KASII* and *KASIII* added a new perspective to the underlying genetics of the high oleic phenotype. Both genes code for essential condensing enzymes in the plastidial fatty acid synthesis. Since *KASIII* is only involved in the very first condensing step at the beginning of the glycerolipid biosynthesis, it has been suggested to be involved in the control of the flux through this metabolic pathway (Ohlrogge and Browse, 1995). The relative flux through the pathway could affect the overall fatty acid composition, since the enzymes in the pathway are known to have different catalytic efficiencies (Mekhedov et al., 2000). Similarly the condensing function of *KASII* is limited to the elongation of the 16-carbon palmitoyl-ACP to stearoyl-ACP, which would also be more likely to be overall rate limiting than influencing fatty acid composition directly. In addition, we have shown in microarray experiments earlier that there was no expression difference for the two genes in the developing kernels of PHC and PHD (Chapter 1). Nevertheless, the different parental alleles of *KASII* and *KASIII* displayed a profound impact on the oleic acid content in the RILs. At this point, we can only speculate as to the putative regulatory function of these genes in the observed epistatic interactions. Similar findings applied to the two *FAB2* genes, which catalyze the conversion of stearoyl-ACP to oleoyl-ACP, a direct prerequisite for oleic acid accumulation. While *FAB2-1* had previously been identified as having major effect on the high oleic acid phenotype, *FAB2-2* had not been identified as a contributor to the phenotype (Pérez-Vich et al., 2002).

FATB and *FATA* were more likely to have an effect on the phenotype, since both thioesterases hydrolyze the acyl group from acyl-ACP, thus allowing the fatty acids to leave the plastid (Browse and Somerville, 1991). While *FATB* displayed a higher affinity to the saturated acyl groups (palmitoyl-ACP and stearoyl-ACP), *FATA* preferentially hydrolyzed oleoyl-ACP (Salas and Ohlrogge, 2002). Therefore an increase in FATA activity would increase the amount of unsaturated fatty acids in the cytosol. A similar effect would be achieved by a decrease in FATB enzyme activity. Despite the fact, that we were unable to detect any differences in transcript levels in the developing kernels of the two parental lines, the two parental alleles at the two loci caused significant effects in the QTL analysis. In addition, we demonstrated that the

FATB locus is completely conserved between the two lines. Therefore the allelic effect had to be of a different regulatory nature, such as post-translation modification of the enzyme(s) or involvement in the regulatory network of the pathway.

Within the last decade the glycerolipid biosynthetic pathway in plants has been intensely studied. Model systems like *A. thaliana* provided insight into the enzymatic processes involved in de novo fatty acid synthesis, enabling us to study the molecular basis of the high oleic acid phenotype in sunflower. Despite these recent advances we are still lacking a thorough understanding of the regulation and control of this pathway. Recent experiments on gene expression during the seed development of two *A. thaliana* lines revealed that the genes of the glycerolipid synthesis could be divided into distinct groups depending on their expression patterns (Ruuska et al., 2002). The observed expression patterns indicated that only two major regulatory networks controlling gene expression were present during early seed development. In addition to our limited knowledge about the regulation of gene expression, we are also lacking a detailed insight into the spatial configuration of the respective enzymes. The individual enzymes are located in the stroma of the plastid, possibly organized into a large supramolecular complex (Ohlrogge and Browse, 1995). Such an enzymatic complex could reveal specific requirements in regard to the stoichiometry of the enzymes present. Gaining a better understanding of the regulatory network and the enzymatic configuration in the lipid biosynthesis will enable us in the future to determine the function of the epistatic interactions observed in our study.

The analysis of the phenotypic data revealed a significant location effect on the oleic acid content in the RILs. The lines with a wildtype fatty acid profile accumulated less oleic acid and more linoleic acid in the seed, when grown in Corvallis, OR, e.g. 18.1% and 18.7% oleic acid in 2000 and 2001 in Corvallis compared to 26.4% and 25.6% in Woodland in the same years. The observed effect can be attributed to the cooler nighttime temperatures of the Willamette Valley in Oregon compared to the respective temperatures of the Central Valley California's, since oleate desaturase (FAD2) activity has been shown to be temperature sensitive. Exposure of developing sunflower kernels to lower temperatures (20/10°C day/night) led to elevated enzyme

activity and increased linoleic acid content compared to 30/20°C (day/night) temperatures (Garcés et al., 1992, Sarmiento et al., 1998).

A second point revealed in our analysis was the limitation of the RIL population size on the estimation of the genetic effects of the candidate genes. These limitations have been widely discussed and accepted in regard to traditional QTL analyses, where the size of the population determined the ability to detect QTLs with minor effects in the surveyed genome (Melchinger et al., 1998, Utz et al., 2000). The candidate gene approach differed in the sense that we had a known amount of genetic main and interaction effects we wanted to assess. As mentioned earlier, even a fully inbred population of the same size would not have been large enough to assess all possible effects. Therefore, the obtained model might not be all encompassing and possibly lacking higher order interactions, important in the high oleic phenotype. Even though it might be of questionable value to estimate 5, 6 and 7-way interactions, it would certainly be desirable to estimate all main and two-way interactions simultaneously. This will in turn require the development of a very large segregating population consisting of purely homozygous lines, as any residual heterozygosity will increase the need for a large population unproportinally. The number of candidate genes to be included in such an analysis will likely increase in the future as availability of sequence data from EST databases and other large scale sequencing projects will facilitate marker development. In addition, new insights into the glycerolipid pathway will point to additional candidates to be included in such an analysis.

The task of developing sequence based markers for the aforementioned candidate genes in the segregating population proved much more difficult than anticipated due to the their monomorphic character. The genes of the glycerolipid pathway revealed two distinct patterns. Genes like *FAD2*, *FATB*, *FATA* or *FAD6* displayed a high degree of sequence conservation among the elite sunflower germplasm, especially within the oilseed lines, while genes of the *KAS*-family or *FAB2* revealed great allelic variability within the same germplasm. The most prominent example of sequence conservation found in our survey was the *FATB* locus. We showed that the gene contained 6 introns and spanned 3026 bp of genomic DNA.

The SSCP analysis and the subsequent sequencing revealed no single polymorphism among 12 elite sunflower lines, confectionary and oilseed, included in our screening panel. This extraordinary degree of sequence conservation and the associated lack of allelic variation at this locus suggested that the 12 lines derived their *FATB* allele from the same germplasm source. Due to the lack of polymorphism in this gene, future attempts to develop markers for the *FATB* locus in the elite sunflower germplasm, which could be used in marker assisted selection, will have to resort to the flanking genomic sequence. Such an approach would either entail genome walking or the construction of a BAC library with subsequent probing for *FATB*, isolation and sequencing of the respective clone. As we have demonstrated on the *FAD2-2* example for the two lines PHC and PHD, that polymorphism might not be found in direct vicinity to the locus of interest.

The inclusion of three wild sunflower lines, ANN1811, Havasupai and *H. argophyllus* in our germplasm screening panel revealed a much greater allelic diversity than the elite lines. This increased allelic variation allowed us to map those candidate genes, which proved monomorphic among the elite material. These findings were further supported by the diversity uncovered in the SSRs present in the 3' UTR of *FAD2-1*. The high degree of allelic variability of the two SSRs observed in the wild material reflected the extraordinary diversity found in native American land races and wild populations of cultivated sunflower (Tang and Knapp, 2003).

CHAPTER 4

CONCLUSIONS

Wildtype sunflowers typically produce 12-24% oleic acid (18:1) and 70-82% linoleic acid (18:2) (Dorrel and Vick, 1997). However, high-oleic sunflower lines have been developed, producing up to 90% oleic acid (Dorrel and Vick, 1997). The monounsaturated oleic acid has a greater oxidative stability than the polyunsaturated linoleic acid, predominant in the wildtype sunflowers, and has greater nutritional benefits than polyunsaturated and saturated fatty acids (Grundy et al., 1986, Kris-Etherton et al., 1999). High oleic sunflower lines are based on high-oleic acid germplasm, which originated from an induced mutation (Ol_1) discovered by Soldatov (1976) and released as the open- pollinated cultivar Pervenets. Ol_1 is necessary but often not sufficient for producing the high oleic phenotype, presumably because additional quantitative trait loci (QTL) segregate in some genetic backgrounds (Miller et al., 1987a). Since the discovery of high oleic phenotypes (Soldatov, 1976), several models have been proposed to explain the genetics of the high oleic acid phenotype in sunflower and Ol_1 has been the only common denominator in the different genetic models (Urie, 1985; Miller et al., 1987a; Fernández-Martínez et al., 1989; Fernandez et al., 1999, Lacombe and Bervillé, 2001, Lacombe et al., 2002a, 2002b).

The oleate desaturase (*FAD2*) has been established as the principal candidate gene affecting oleic acid content in the sunflower seed, since it has been shown to cosegregate with Ol_1 (Hongtrakul et al., 1998a, Lacombe and Bervillé, 2001, Pérez-Vich et al., 2002). Hongtrakul et al. (1998a) isolated an oleate desaturase cDNA (*FAD2-1*) that is highly expressed in developing kernels of wildtype lines, but displays greatly reduced expression in the mutant lines. In addition they showed that the *FAD2-1* locus is duplicated and rearranged in the mutant lines (Hongtrakul et al., 1998a, Lacombe and Bervillé, 2001).

In this research we identified and characterized the *FAD2-1* gene duplication in the high oleic sunflower lines. We showed that the high oleic lines carried a tandem duplication of *FAD2-1*. By inverting gene specific primers flanking the coding region and using long-distance PCR we recovered the 3.1 kb intergenic region spanning the 3' to 5'-UTRs in the duplicated copies. In addition, we found a 1.7 kb intron in the 5'-UTR of the first copy, which was also present in the wildtype. The novel joint of the duplication was discovered in a truncated 5'-UTR intron of the second copy, since 1.4 kb of the intron was missing. The *FAD2-1* locus was sequenced in the mutant, from the 5'UTR of the first copy to the 3'UTR of the second copy and was found to be 7.8 kb long. The tandem duplication was present in the cultivar Pervenets, the original source of the high oleic phenotype, but not in VNIIMK8931, the open-pollinated M_0 population.

We demonstrated that the reduced transcription of *FAD2-1* in the developing kernels of the high oleic lines was caused by the RNA interference (RNAi) pathway. Northern analyses of RNA from developing kernel at 14 DAF from wildtype and mutant lines with a 3'-UTR specific *FAD2-1* probe identified 21nt short-interfering RNA (siRNA) in mutant lines only. The *FAD2-1* duplication was hypothesized to have eliminated the transcription termination signal in a downstream gene oriented opposite of *FAD2-1*, thereby yielding an extended transcript complementary to *FAD2-1*, forming double-stranded RNA and triggering the RNAi machinery. When tested by bidirectional RT-PCR analyses, mutant lines produced transcripts in sense and antisense directions, whereas wildtype lines only produced transcripts in the sense direction.

In addition, we compared the transcript levels of 48 glycerolipid biosynthetic genes in the developing kernels of 4 low and high oleic lines using microarrays. It was our aim to identify any candidate genes with distinct transcriptional differences between the low and high oleic lines. The analyses revealed that the lipid transfer protein was the only additional gene besides *FAD2-1* with differing transcript levels in each of the four comparisons. The microarray experiments also revealed no significant differences in transcript levels for the genes directly involved in oleic acid synthesis

for the HA89-HA341, the RHA274-RHA345 and the PHD-PHC comparisons. In contrast, the HA292-HA349 array showed differing transcript levels for about a quarter of the genes in the glycerolipid biosynthesis, including all three members of the *FAD2* gene family as well as *FAB2-2*. This is most likely due to the unique genetic background of HA292.

The insights gained from the research of the lipid metabolism in Arabidopsis and the availability of sunflower EST sequences in the Composite Genome Project EST database allowed us to identify sunflower sequences homologous to the genes directly affecting the oleic acid synthesis and content in the seed. We obtained sequences for the following genes *FAD2-1*, *FAD2-2*, *FAD2-3*, *FATB*, *FATA*, *KASI*, *KASII*, *KASIII*, *FAB2-1*, *FAB2-2*and *FAD6*. We developed sequence-based agarose, SSCP, SSR and SNP markers for all of them, but *FAD2-3*, and subsequently mapped these candidate genes in one of our reference mapping populations. In the process of marker development, we discovered two distinct groups within the above mentioned genes. The first group is comprised of *FAD2-1*, *FAD2-2*, *FAD2-3*, *FATB*, *FATA*, *KASI* and *FAD6* and characterized by an extraordinary degree of sequence conservation in the elite sunflower germplasm. *KASII*, *KASIII*, *FAB2-1* and *FAB2-2* belong to the second group and displayed great allelic variability across the 15 surveyed sunflower lines.

We used the above mentioned markers to assess the effect of these candidate genes on the oleic acid content in sunflower seeds. The analysis of an F_5 RIL population segregating for oleic acid content revealed that the phenotype is based on a complex interaction of the different alleles at these loci. Our results demonstrate that ORS460 (*FATB-1*), ORS1256 (*FATB-2*), *KASIII*, *FAB2-1*, *FAB2-2* and CRT394 (*FATA*) as well as interactions between *FAD2-1* × ORS460 (*FATB-1*), ORS460 (*FATB-1*) × *KASII*, ORS460 (*FATB-1*) × *KASIII*, ORS460 (*FATB-1*) × *FAB2-2*, *KASII* × *KASIII*, *KASII* × ORS1256 (*FATB-2*), *KASII* × CRT394 (*FATA*), *KASIII* × ORS1256 (*FATB-2*) and *FAB2-1* × *FAB2-2* significantly contribute to the high oleic acid phenotype besides the known effect of *FAD2-1*. These findings emphasized the

complexity of the phenotype and indicated a limitation in the applicability of marker assisted selection for high oleic acid content.

Our research has uncovered the nature of *FAD2-1* gene duplication, which is necessary, but not sufficient, to obtain high oleic sunflower lines. In addition we have shown that the high oleic phenotype is regulated by a complex network of interactions between the genes of the glycerolipid biosynthesis, which is independent of regulation at the transcriptional level as demonstrated by the microarrays. Overall we have taken a big leap forward in understanding the molecular mechanisms underlying the high oleic acid phenotype in sunflower.

BIBLIOGRAPHY

Ascherio, A., and W.C. Willett. 1997. Health effects of trans fatty acids. Am. J. Clin. Nutr. 66:1006S-1010S.

Ashburner, M. 1990. Drosophila: A Laboratory Handbook. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Bianchi, M.W., and A. Viotti. 1988. DNA methylation and tissue-specific transcription of the storage protein gene of maize. Plant Mol. Biol.11:203-214.

Bonaventure, G., J.J. Salas, M.R. Pollard, and J.B. Ohlrogge. 2003. Disruption of the *FATB* gene in Arabidopsis demonstrates an essential role of saturated fatty acids in plant growth. Plant Cell 15:1020-33.

Bonferroni, C.E. 1936. Il calcolo delle assicurazioni si gruppi di teste, p. 13-60 *In* Studi in Onore del Professore Salvatore Ortu Carboni, Rome, Italy

Brandt, T., and S. J. Knapp. 1993. Gametic selection at fatty acid and allozyme marker loci and meiosis within *Cuphea viscosissima* x *Cuphea lanceolata* populations. Crop Science 33:1138-1143.

Browse, J., and C. Somerville. 1991. Glycerolipid synthesis: Biochemistry and regulation. Annu. Rev. Plant Physiol. Plant Mol. Biol. 42:467-506.

Chen, X., L. Levine, and P.Y. Kwok. 1999. Fluorescence polarization in homogeneous nucleic acid analysis. Genome Res. 9:492-8.

Dorrel, D.G., and B. A. Vick. 1997. Properties and processing of oilseed sunflower, p. 709-746, *In* A. A. Schneiter, ed. Sunflower technology and production. ASA, CSSA, SSSA, Madison, Wisconsin, USA.

Falconer, D.S., and Mackay, T.F.C. 1996 Introduction to quantitative genetics, 4th edition. Pearson Education Limited, Essex, England, UK.

FAOSTAT data. 2004. http://faostat.fao.org/faostat/collections?subset=agriculture

Fernandez, H., M. Baldini, and A. M. Olivieri. 1999. Inheritance of high oleic acid content in sunflower oil. Journal of Genetics & Breeding 53:99-103.

Fernández-Martínez, J., A. Jimenez, J. Dominguez, J. M. Garcia, R. Garcés, and M. Mancha. 1989. Genetic analysis of the high oleic acid content in cultivated sunflower (*Helianthus annuus* L.). Euphytica 41:39-51.

Fick, G.N., D.E. Zimmer, and M.L. Kinman. 1974. Registration of six sunflower parental lines. Crop Science 14:912.

Gambhir, P.N., and I. J. Anand. 1979. Oil and protein synthesis during grain development in sunflower. Indian journal of plant physiology 24:69-73.

Garcés, R., J. M. Garcia, and M. Mancha. 1989. Lipid characterization in seeds of a high oleic acid sunflower mutant. Phytochemistry 28:2597-2600.

Garcés, R., C. Sarmiento, and Mancha, M. 1992. Temperature regulation of oleate desaturase in sunflower (*Helianthus annuus* L.) seeds. Planta 186:461-465.

Ghandi, S.D., A.F. Heesacker, C.A. Freeman, J. Argyris, K. Bradford, and S.J. Knapp. 2005. The self-incompatibility locus (S) and quantitative trait loci for self-pollination and seed dormancy in sunflower. Theor. Appl. Genet. (in press).

Greene, E.A., C.A. Codomo, N.E. Taylor, J.G. Henikoff, B.J. Till, S.H. Reynolds, L.C. Enns, C. Burtner, J.E. Johnson, A.R. Odden, L. Comai, and S. Henikoff. 2003. Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. Genetics 164:731-40.

Grundy, S.M. 1986. Comparison of monounsaturated fatty acids and carbohydrates for lowering plasma cholesterol. N. Engl. J. Med. 314:745-8.

Guiderdoni, E., N. Ferriere, D. Meynard, M. Delseny, J. L. Notteghem, D. Tharreau, F. Vignols, M. J. Cordero, M. Lescot, and J. M. Garcia-Garrido. 2002. Inducibility by pathogen attack and developmental regulation of rice Ltp1 gene. Plant Mol. Biol. 49:683-699.

Hannon, G.J. 2002. RNA interference. Nature 418:244-251.

Hoffmann, G.R. 1980. Genetic effects of dimethyl sulfate, diethyl sulfate, and related compounds. Mutat. Res. 75:63-129.

Holloway, J.L., and S. J. Knapp. 1993. G-MENDEL 3.0 user's guide. Oregon State University, Corvallis, OR.

Hongtrakul, V., M.B. Slabaugh, and S.J. Knapp. 1998a. A seed specifc Δ-12 oleate desaturase gene is duplicated, rearranged, and weakly expressd in high oleic acid sunflower lines. Crop Science 38:1245-1249.

Hongtrakul, V., M.B. Slabaugh, and S.J. Knapp. 1998b. DFLP, SSCP, and SSR markers for Δ9-stearoyl-acyl carrier protein desaturases strongly expressed in developing seeds of sunflower: intron lengths are polymorphic among elite inbred lines. Mol. Breed. 4:195-203.

Huala, E., A.W. Dickerman, M. Garcia-Hernandez, D. Weems, L. Reiser, F. LaFond, D. Hanley, D. Kiphart, M. Zhuang, W. Huang, L.A. Mueller, D. Bhattacharyya, D. Bhaya, B.W. Sobral, W. Beavis, D.W. Meinke, C.D. Town, C. Somerville, and S.Y. Rhee. 2001. The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Res. 29:102-5.

Jung, H.W., B. K. Hwang, and W. Kim. 2003. Three pathogen-inducible genes encoding lipid transfer protein from pepper and differentially activated by pathogens, abiotic, and environmental stresses. Plant, cell and environment. 26:915-928.

Jung, S., G. Powell, K. Moore, and A. Abbott. 2000b. The high oleate trait in the cultivated peanut (*Arachis hypogaea* L). II. Molecular basis and genetics of the trait. Mol. Gen. Genet. 263:806-11.

Jung, S., D. Swift, E. Sengoku, M. Patel, F. Teule, G. Powell, K. Moore, and A. Abbott. 2000a. The high oleate trait in the cultivated peanut (*Arachis hypogaea* L.). I. Isolation and characterization of two genes encoding microsomal oleoyl-PC desaturases. Mol. Gen. Genet. 263:796-805.

Kolkman, J.M., M. B. Slabaugh, J. M. Bruniard, S. Berry, B. S. Bushman, C. Olungu, N. Maes, G. Abratti, A. Zambelli, J. F. Miller, A. Leon, and S. J. Knapp. 2004. Acetohydroxyacid Synthase Mutations conferring resistance to Imidazolinone or Sulfonylurea herbicides in sunflower. Theor. Appl. Genet. 109:1147-1159.

Kosambi, D.D. 1944. The estimation of map distance from recombination values. Ann. Eugen. 12:172-175.

Kozik, A., R.W. Michelmore, S.J. Knapp, M.S. Matvienko, L. Rieseber, H. Lin, M. van Damme, D. Lavelle, P. Chevalier, J. Ziegle, P. Ellison, J.M. Kolkman, M.B. Slabaugh, K. Livingston, L.Z. Zhou, S. Church, S. Edberg, L. Jackson, R. Kesseli, and K. Bradford. 2002. Sunflower and lettuce ESTs from the Composite Genome Project, http://compgenomics.ucdavis.edu

Kris-Etherton, P.M. 1999. AHA science advisory: monounsaturated fatty acids and risk of cardiovascular disease. J. Nutr. 129:2280-4.

Kuehl, R.O. 2000. Design of Experiments: Statistical principles of research design and analysis, p. 91-98, 2nd edition, Duxbury Press, Pacific Grove, CA, USA.

Kwok, P.Y., and X. Chen. 2003. Detection of single nucleotide polymorphisms. Curr. Issues Mol. Biol. 5:43-60.

Lacombe, S., and A. Bervillé. 2001. A dominant mutation for high oleic acid content in sunflower (*Helianthus annuus* L.) seed oil is genetically linked to a single oleate-desaturase RFLP locus. Mol. Breed. 8:129-137.

Lacombe, S., S. Léger, F. Kaan, and A. Bervillé. 2002a. Inheritance of oleic acid content in F_2 and a population of recombinant inbred lines segregating for the high oleic trait in sunflower. Helia 25:85-94.

Lacombe, S., A.G. Abbott, and A. Bervillé. 2002b. Repeats of an oleate desaturase region cause silencing of the normal gene explaining the high oleic Pervenets sunflower mutant. Helia 25:95-104.

Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln, and L. Newburg. 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-81.

Lichtenstein, A.H., L.M. Ausman, S.M. Jalbert, and E.J. Schaefer. 1999. Effects of different forms of dietary hydrogenated fats on serum lipoprotein cholesterol levels. N. Engl. J. Med. 340:1933-40.

Lippman, Z., B. May, C. Yordan, T. Singer, and R. Martienssen. 2003. Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS Biology 1:420-428.

Liu, B.H. 1998. Statistical Genomics CRC Press, New York.

Liu, Q., S. P. Singh, and A.G. Green. 2002. High-Stearic and high-oleic cottonseed oil produced by hairpin RNA-mediated post-transcriptional gene silencing. Plant Physiol. 129:1732-1743.

Llave, C., K.D. Kasschau, and J.C. Carrington. 2000. Virus-encoded suppressor of posttranscriptional gene silencing targets a maintenance step in the silencing pathway. Proc. Natl. Acad. Sci. U S A 97:13401-6.

Loeffert, D., S. Karger, N. Seip, and J. Kang. 1998. PCR optimization: longer products. QIAGEN News 3:1 & 18-20.

Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer, Sunderland, MA, USA.

Martínez-Rivas, J.M., P. Sperling, W. Luehs, and E. Heinz. 2001. Spatial and temporal regulation of three different microsomal oleate desaturase genes (*FAD2*) from normal-type and high-oleic varieties of sunflower (*Helianthus annuus* L.). Mol. Breed. 8:159-168.

Matzke, M.A., and A.J.M. Matzke. 2004. Planting the seeds of a new paradigm. PLoS Biology 2:0582-0586.

Melchinger, A.E., H.F. Utz, and C.C. Schön. 1998. Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149:383-403.

Mette, M.F., W. Aufsatz, J. van der Winden, and A.J. Matzke. 2000. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J. 19:5194-5201.

Miller, J.F., D.C. Zimmerman, and B.A. Vick. 1987a. Genetic Control of High Oleic Acid Content in Sunflower Oil. Crop Science 27:923-926.

Miller, J.F., D.C. Zimmerman, B.A. Vick, and W.W. Roath. 1987b. Registration of sixteen high oleic sunflower germplasm lines and bulk population. Crop Science 27:1323.

Miller, J.F. 1992. Registration of five oilseed sunflower germplasm restorer lines (RHA373 to 377) and two nuclear male-sterile populations (NMS274 and 801). Crop Science 32:1298.

Miller, J.F., and T. J. Gulya. 1995. Registration of four maintainer (HA382 to HA385) and four restorer (RHA386 to RHA389) sunflower germplasm lines. Crop Science 35:605.

Miller, J.F. 1997. Registration of smcHA 89(PEF1) cytoplasmic male-sterile, RPEF1 restorer, and two nuclear male-sterile (NMS 373 and 377) sunflower genetic stocks. Crop Science 37:1984.

National Sunflower Association. 2003. U.S. Sunflower Crop Quality Report. National Sunflower Association, Bismarck, ND, http://www.sunflowernsa.com

Ohlrogge, J., and J. Browse. 1995. Lipid Biosynthesis. Plant Cell 7:957-970.

Okuley, J., J. Lightner, K. Feldmann, N. Yadav, E. Lark, and J. Browse. 1994. Arabidopsis *FAD2* gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. Plant Cell 6:147-58.

Patel, M., S. Jung, K. Moore, G. Powell, C. Ainsworth, and A. Abbott. 2004. High-oleate peanut mutants result from a MITE insertion into the *FAD2* gene. Theor. Appl. Genet.108:1492-1502.

Pérez-Vich, B., J. M. Fernández-Martínez, M. Grondona, S.J. Knapp, and S.T. Berry. 2002. Stearoyl-ACP and oleoyl-PC desaturase genes cosegregate with quantitative trait loci underlying high stearic and high oleic acid mutant phenotypes in sunflower. Theor. Appl. Genet.104:338-349.

Rabinowicz, P.D., L.E. Palmer, B.P. May, M.T. Hemann, S.W. Lowe, W.R. McCombie, and R.A. Martienssen. 2003. Genes and transposons are differentially methylated in plants, but not in mammals. Genome Research. 13:2658-2664.

Regenet, M., and L. de la Canal. 2003. A cDNA encoding a putative lipid transfer protein expressed in sunflower seeds. J. Plant Physiol. 160:201-203.

Roath, W.W., J.F. Miller, and T.J. Gulya. 1981. Registration of RHA 801 sunflower germplasm. Crop Science 21:479.

Rozen, S., and H. J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers, p. 365-386, *In* S. Krawetz, and S. Misener, ed. Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa.

Ruuska, S.A., T. Girke, C. Benning, and J.B. Ohlrogge. 2002. Contrapuntal networks of gene expression during Arabidopsis seed filling. Plant Cell 14:1191-206.

Salas, J.J., and J.B. Ohlrogge. 2002. Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. Arch. Biochem. Biophys. 403:25-34.

Sanguinetti, C.J., E. Dias Neto, and A.J. Simpson. 1994. Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. BioTechniques 17:914-21.

Sarmiento, C., R. Garcés, and M. Mancha. 1998. Oleate desaturation and acyl turnover in sunflower (*Helianthus annuus* L.) seed lipids during rapid temperature adaptation. Planta 205:595-600.

Schenk, P.M., K. Kazan, I. Wilson, J.P. Anderson, T.Richmond, S.C. Somerville, and J.M. Manners. 2000. Coordinated plant defense responses in Arabidopsis reveled by microarray analysis. Proc. Natl. Acad. Sci. U S A 97:11655-11660.

Slabaugh, M.B., G. M. Huestis, J. Leonard, J. L. Holloway, C. Rosato, V. Hongtrakul, N. Martini, R. Toepfer, M. Voetz, J. Schell, and S. J. Knapp. 1997. Sequence-based genetic markers for genes and gene families: single-strand confirmational polymorphisms for the fatty acid synthesis genes of *Cuphea*. Theor. Appl. Genet. 94: 400-408

Soldatov, K.I. 1976.Chemical mutagenesis in sunflower breeding. 352-357. *In* Proc. 7th Int. Sunflower Conference, Krasnodar, USSR. International Sunflower Assoc., Vlaardingen, the Netherlands.

Stoutjesdijk, P.A., S.P. Singh, Q. Liu, C.J. Hurlstone, P.A. Waterhouse, and A.G. Green. 2002. hpRNA-Mediated targeting of the arabidopsis *FAD2* gene gives highly efficient and stable silencing. Plant Physiol. 129:1723-1731.

Tang, S., J.-K. Yu, M. B. Slabaugh, D. K. Shintani, and S. J. Knapp. 2002. Simple sequence repeat map of the sunflower genome. Theor. Appl. Genet. 105:1124-1136.

Tang, S., A. Heesacker, V. K. Kishore, A. Fernandez, E. S. Sadik, G. Cole, and S. J. Knapp. 2003. Genetic mapping of the Or_5 gene for resistance to *Orobanche* race E in sunflower. Crop Science 43:1021-1028.

Tang, S., and S.J. Knapp. 2003. Microsatellites uncover extraordinary diversity in native American land races and wild populations of cultivated sunflower. Theor. Appl. Genet. 106:990-1003.

U.S. Department of Agriculture, Agricultural Research Service. 2004. USDA National Nutrient Database for Standard Reference, Release 17. Nutrient Data Laboratory Home Page, http://www.nal.usda.gov/fnic/foodcomp.

Urie, L.A. 1985. Inheritance of high oleic acid in sunflower. Crop Science 25:986-989.

Utz, H.F., A. E. Melchinger, and C.C. Schön. 2000. Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics 154:1839-1849.

Volpe, T.A., C. Kidner, I. M. Hall, G. Teng, S. I. S. Grewal, and R. A. Martienssen. 2002. Regulation of heterochromatic silencing and histone H3 Lysine-9 methylation by RNAi. Science 297:1833-1837.

Wassenegger, M., S. Heimes, L. Riedel, and H.L. Sanger. 1994. RNA-directed de novo methylation of genomic sequences in plants. Cell 76:567-576.

Waterhouse, P.M., M.W. Graham, and M. Wang. 1998. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. Proc. Natl. Acad. Sci. U S A 95:13959-13964.

Webb, D.M., and S. J. Knapp. 1990. DNA extraction from a previously recalcitrant plant genus. Mol. Biol. Rep. 8:180-185.

Weber, M., E. Harada, C. Vess, E.V. Roepenack-Lahaye, and S. Clemens. 2004. Comparative microarray analysis of Arabidopsis thaliana and Arabidopsis halleri roots identifies nicotianamine synthase, a ZIP transporter and other genes as potential metal hyperaccumulation factors. Plant Journal 37:269-281.

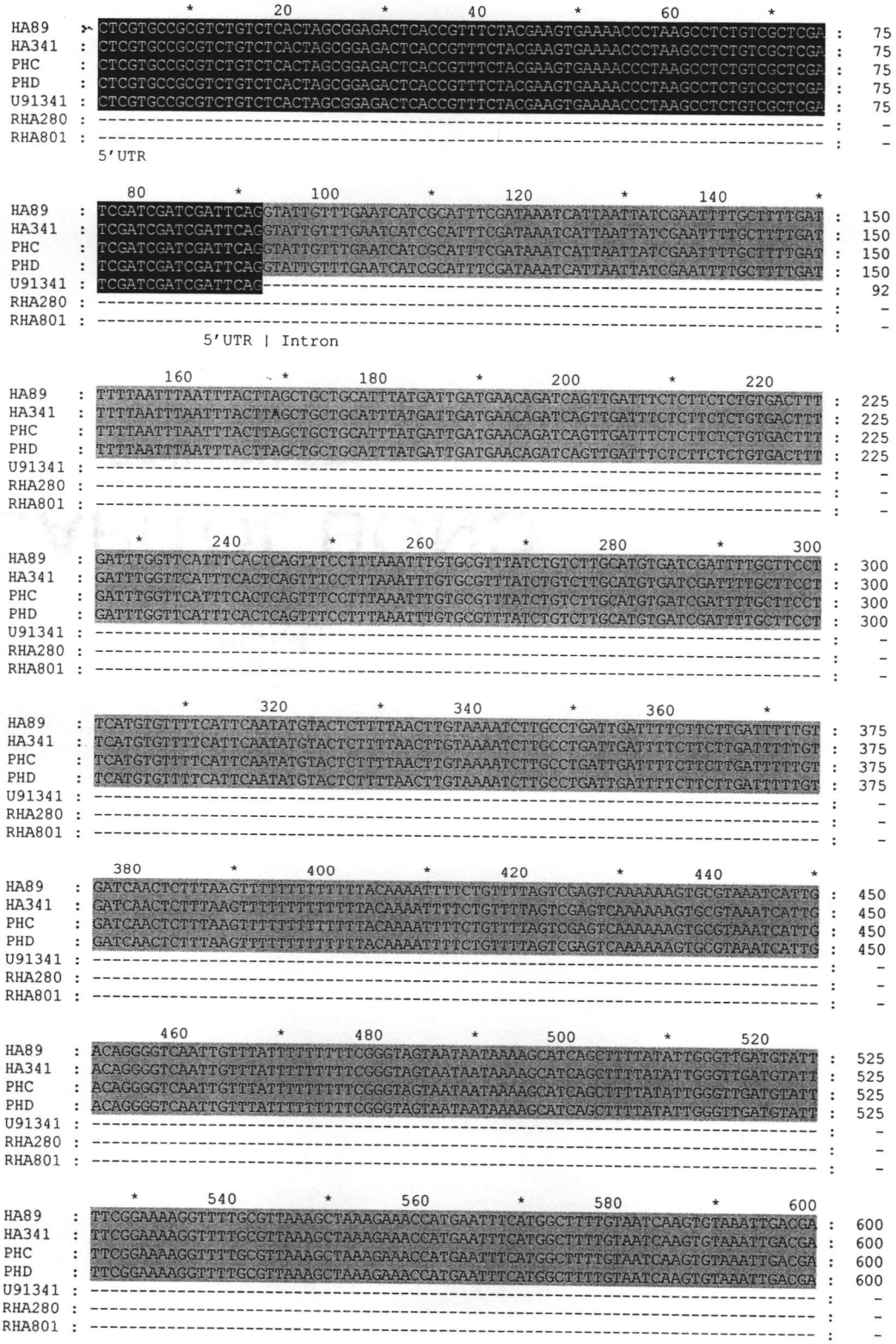
Xu, X., C.R. Dietrich, R. Lessire, B.J. Nikolau, and P.S. Schnable. 2002. The endoplasmic reticulum-associated maize GL8 protine is a component of the acyl-coenzyme A elongase involved in the production of cuticular waxes. Plant Physiol. 128:924-934.

Yu, J.-K., S. Tang, M. B. Slabaugh, A. Heesacker, G. Cole, M. Herring, J. Soper, F. Han, W.-C. Chu, D. M. Webb, L. Thompson, K. J. Edwards, S. Berry, A. J. Leon, M. Grondona, C. Olungu, N. Maes, and S. J. Knapp. 2003. Towards a Saturated Molecular Genetic Linkage Map for Cultivated Sunflower. Crop Science 43:367-387.

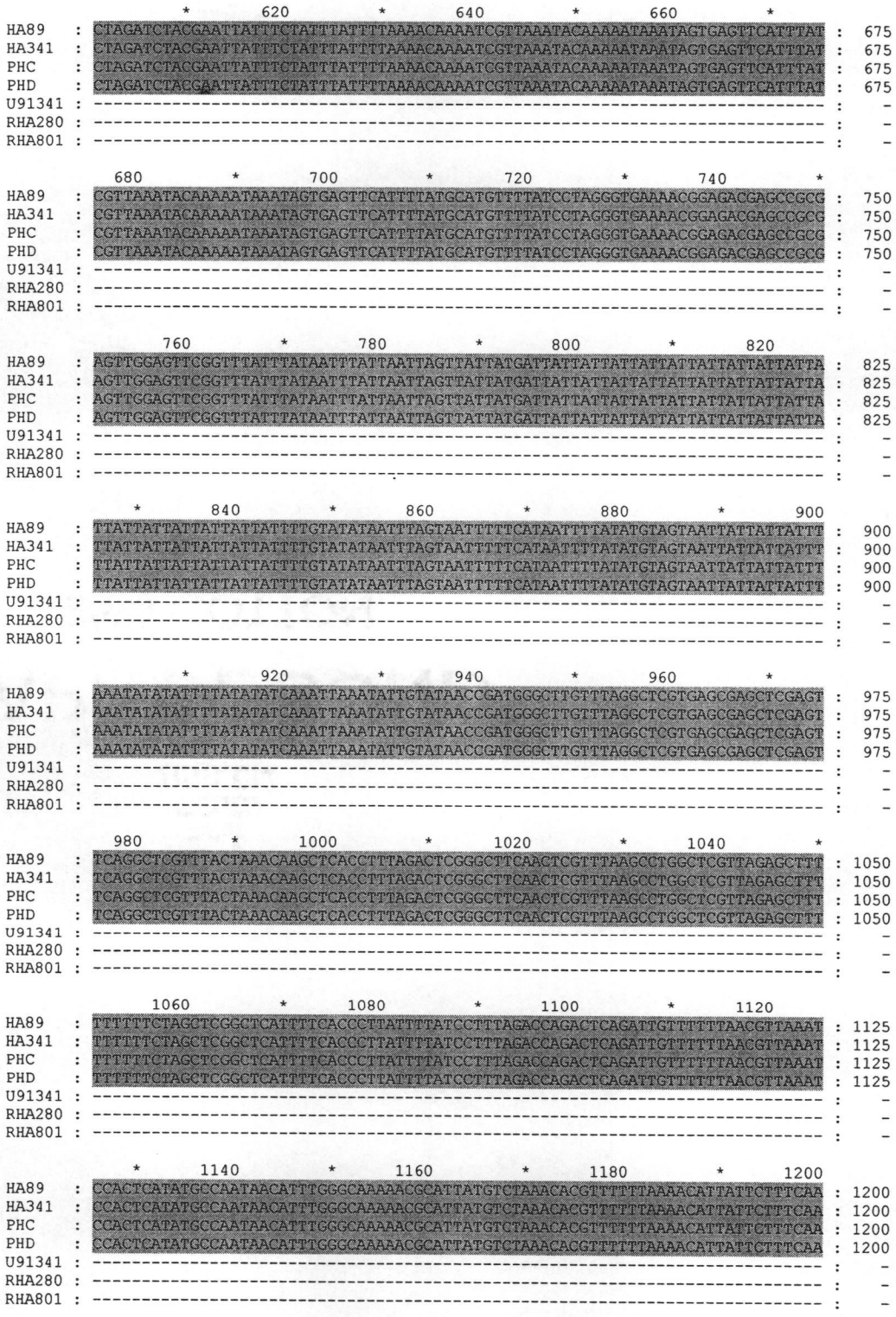
Zamore, P.D. 2002. Ancient pathways programmed by small RNAs. Science 296:1265-1269.

APPENDICES

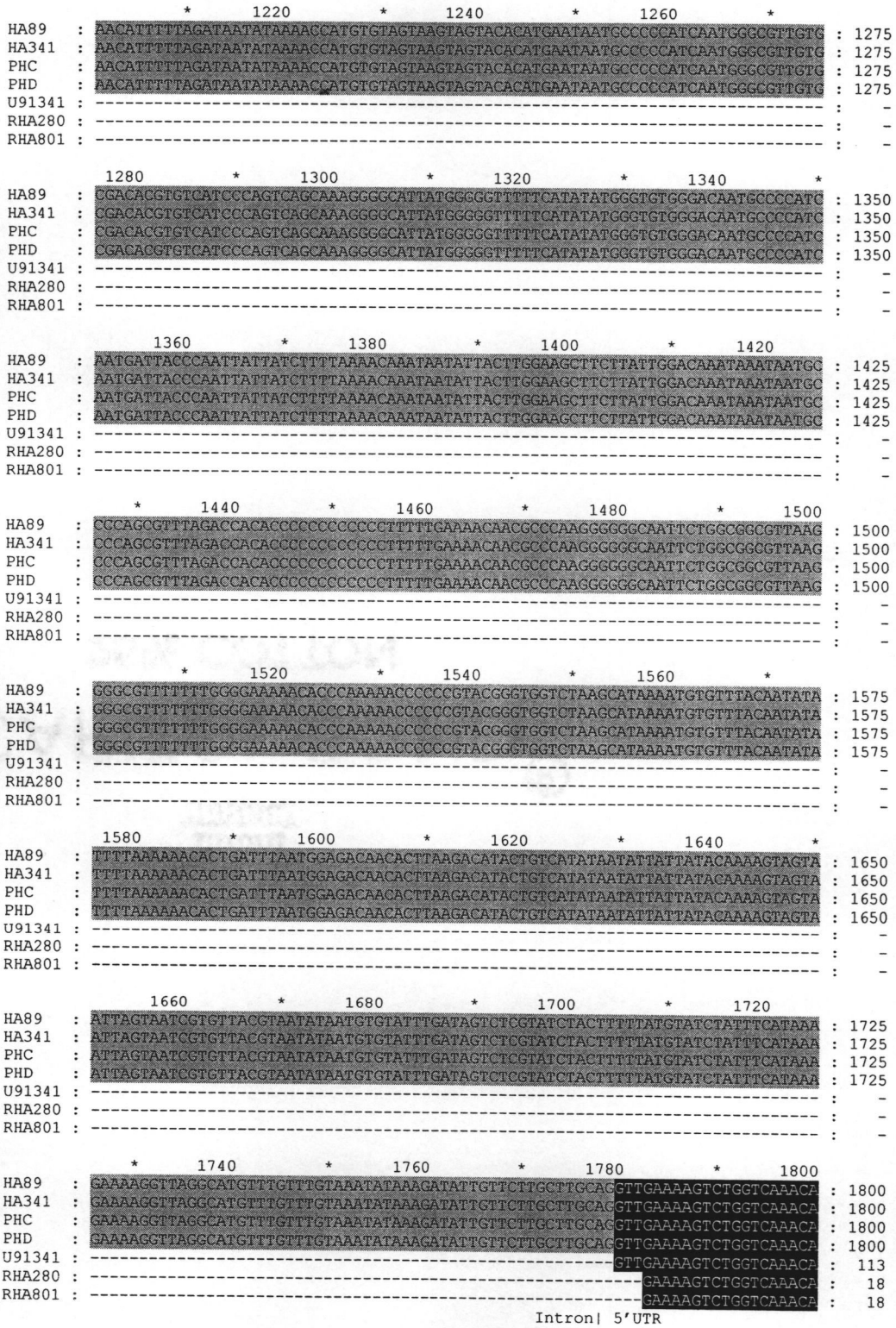
Appendix A. Alignment of the genomic nucleic acid sequence of *FAD2-1* for six sunflower lines with the GenBank cDNA sequence U91341. The alignment shows the 1685 bp intron in the 5' UTR. The start and stop codons are marked in bold.



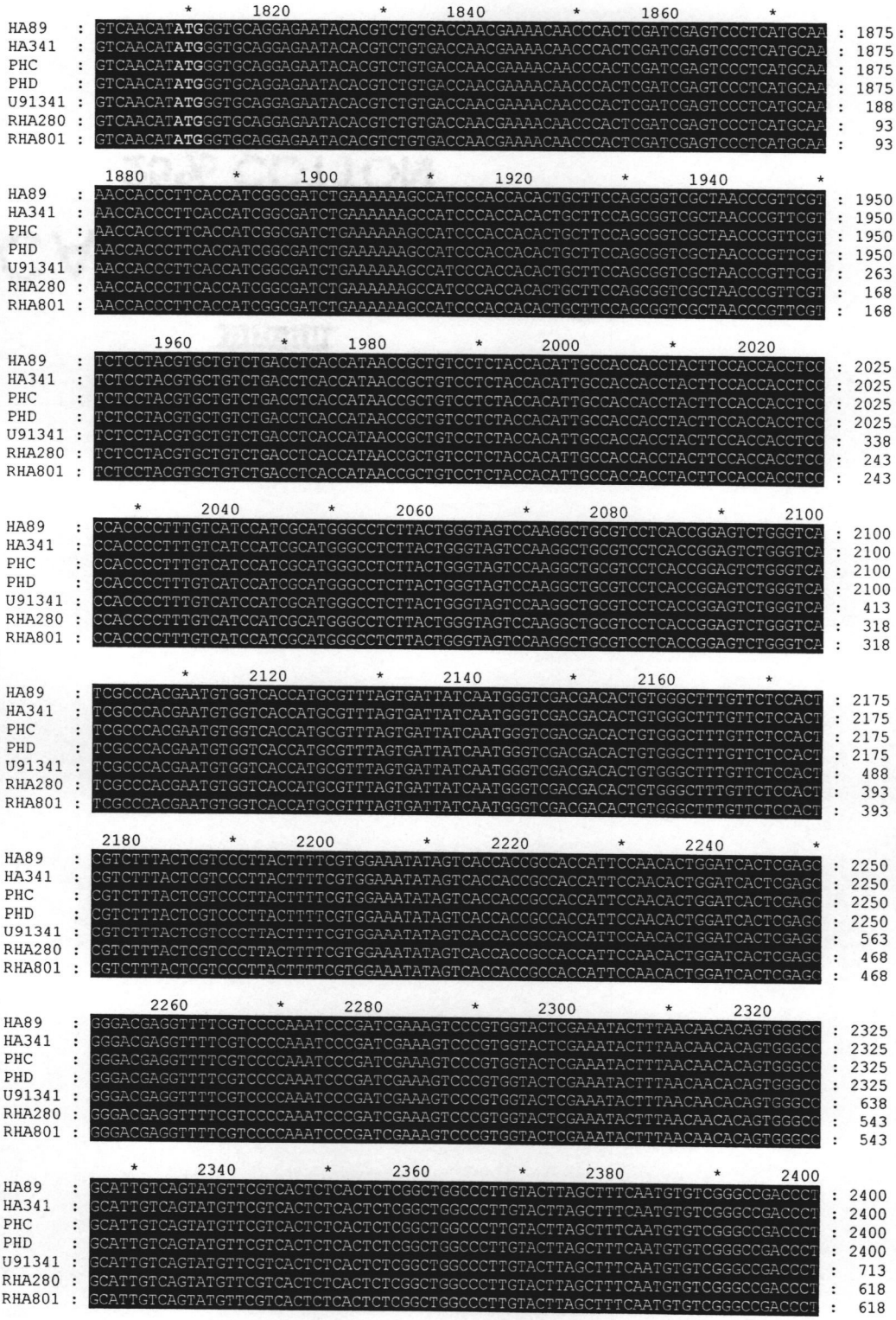
Appendix A. (Continued)



Appendix A. (Continued)



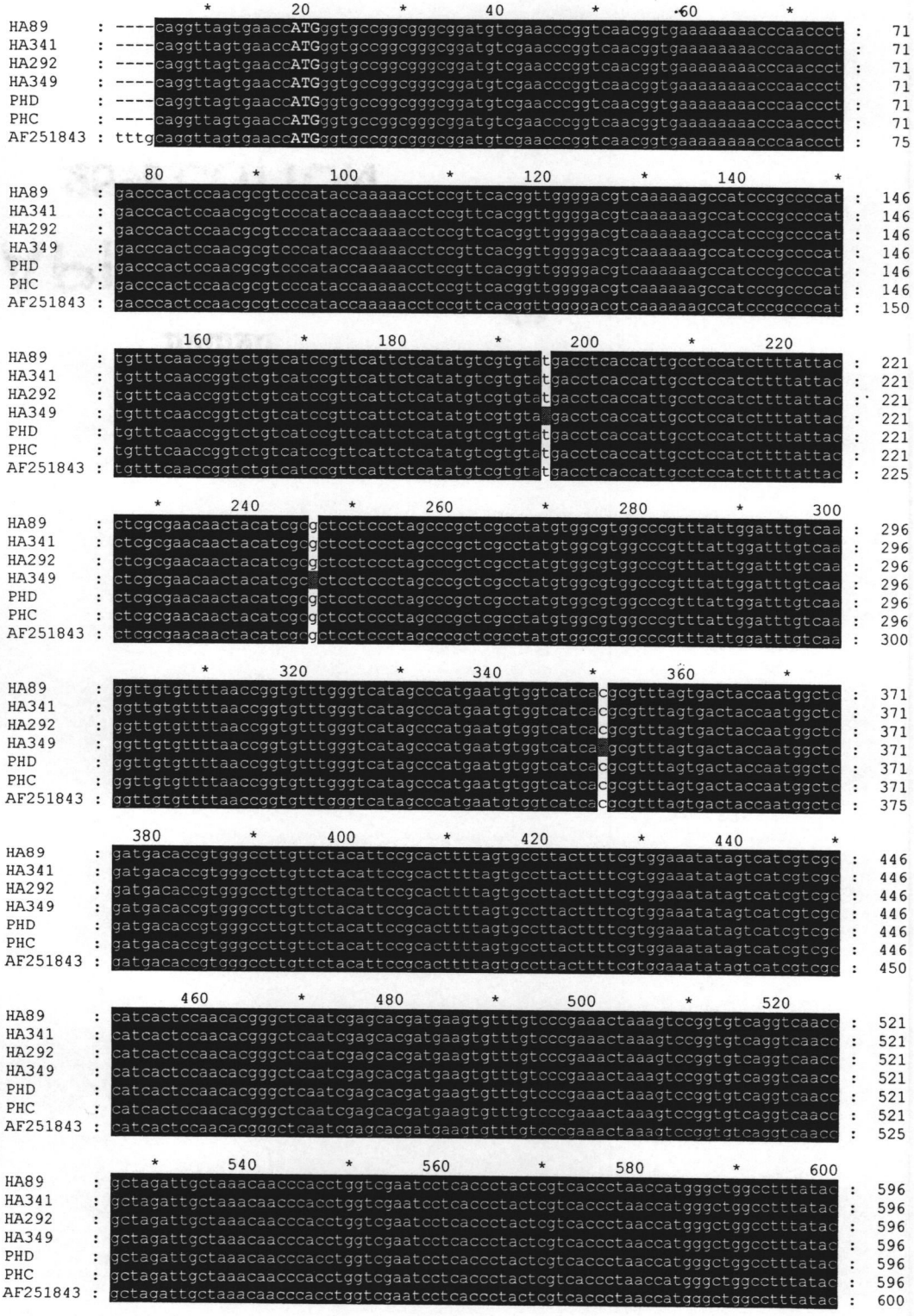
Appendix A. (Continued)



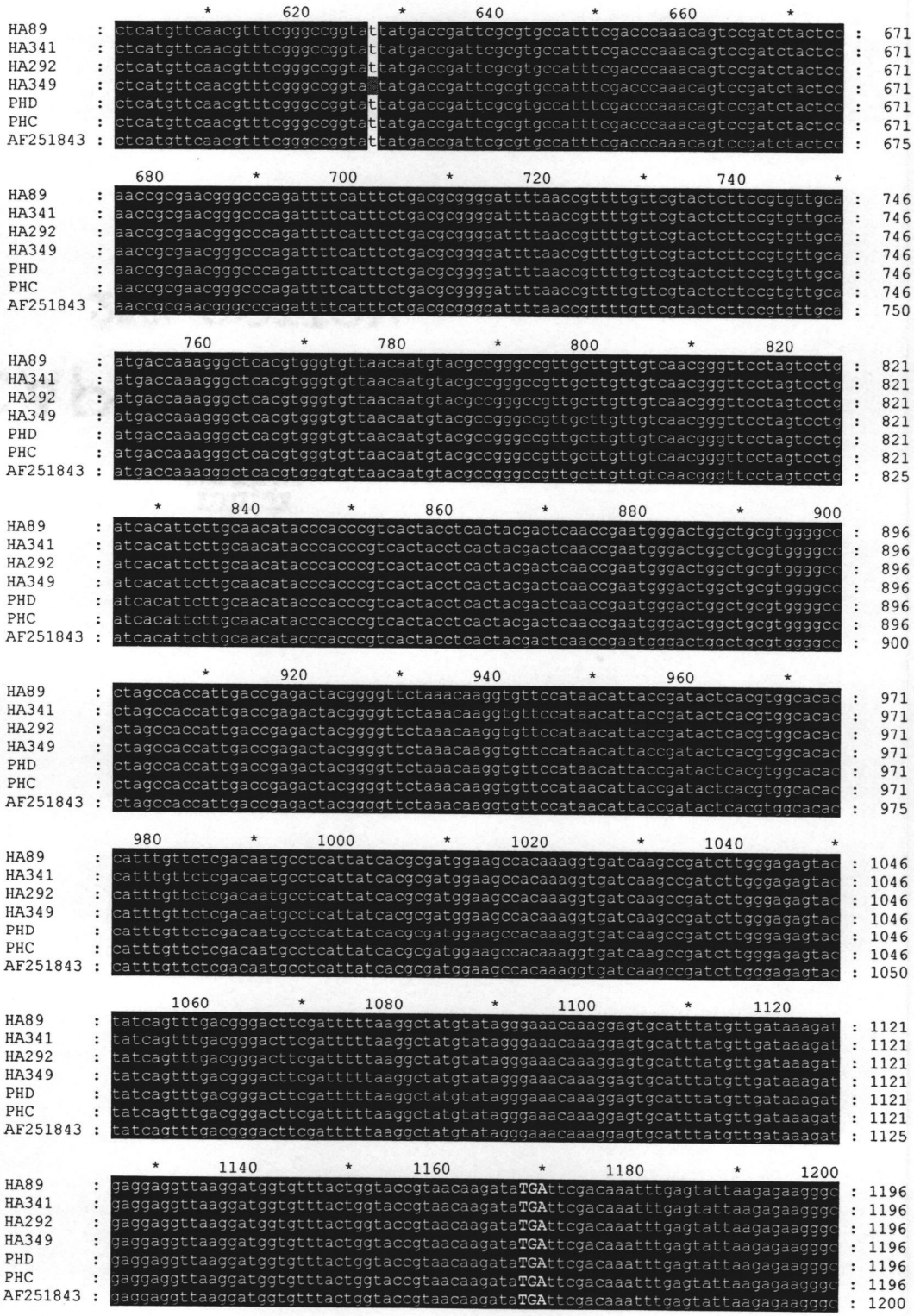
Appendix A. (Continued)

```
                    *       2420            *       2440            *       2460            *
HA89    : ATGACCGTTTCGCCTGCCACTACGTCCCAACCAGCCCTATGTACAATGAACGTAAACGTTACCAGATAGTCATGT : 2475
HA341   : ATGACCGTTTCGCCTGCCACTACGTCCCAACCAGCCCTATGTACAATGAACGTAAACGTTACCAGATAGTCATGT : 2475
PHC     : ATGACCGTTTCGCCTGCCACTACGTCCCAACCAGCCCTATGTACAATGAACGTAAACGTTACCAGATAGTCATGT : 2475
PHD     : ATGACCGTTTCGCCTGCCACTACGTCCCAACCAGCCCTATGTACAATGAACGTAAACGTTACCAGATAGTCATGT : 2475
U91341  : ATGACCGTTTCGCCTGCCACTACGTCCCAACCAGCCCTATGTACAATGAACGTAAACGTTACCAGATAGTCATGT :  788
RHA280  : ATGACCGTTTCGCCTGCCACTACGTCCCAACCAGCCCTATGTACAATGAACGTAAACGTTACCAGATAGTCATGT :  693
RHA801  : ATGACCGTTTCGCCTGCCACTACGTCCCAACCAGCCCTATGTACAATGAACGTAAACGTTACCAGATAGTCATGT :  693

            2480            *       2500            *       2520            *       2540            *
HA89    : CCGACATCGGGATTGTTATCACATCGTTCATCCTTTATCGTGTTGCTATGGCAAAAGGGTTGGTTTGGGTGATTT : 2550
HA341   : CCGACATCGGGATTGTTATCACATCGTTCATCCTTTATCGTGTTGCTATGGCAAAAGGGTTGGTTTGGGTGATTT : 2550
PHC     : CCGACATCGGGATTGTTATCACATCGTTCATCCTTTATCGTGTTGCTATGGCAAAAGGGTTGGTTTGGGTGATTT : 2550
PHD     : CCGACATCGGGATTGTTATCACATCGTTCATCCTTTATCGTGTTGCTATGGCAAAAGGGTTGGTTTGGGTGATTT : 2550
U91341  : CCGACATCGGGATTGTTATCACATCGTTCATCCTTTATCGTGTTGCTATGGCAAAAGGGTTGGTTTGGGTGATTT :  863
RHA280  : CCGACATCGGGATTGTTATCACATCGTTCATCCTTTATCGTGTTGCTATGGCAAAAGGGTTGGTTTGGGTGATTT :  768
RHA801  : CCGACATCGGGATTGTTATCACATCGTTCATCCTTTATCGTGTTGCTATGGCAAAAGGGTTGGTTTGGGTGATTT :  768

                2560            *       2580            *       2600            *       2620
HA89    : GCGTCTATGGGGTTCCGTTGATGGTTGTGAACGCGTTTCTGGTGTTGATCACTTATCTTCAACATACTCACCCTG : 2625
HA341   : GCGTCTATGGGGTTCCGTTGATGGTTGTGAACGCGTTTCTGGTGTTGATCACTTATCTTCAACATACTCACCCTG : 2625
PHC     : GCGTCTATGGGGTTCCGTTGATGGTTGTGAACGCGTTTCTGGTGTTGATCACTTATCTTCAACATACTCACCCTG : 2625
PHD     : GCGTCTATGGGGTTCCGTTGATGGTTGTGAACGCGTTTCTGGTGTTGATCACTTATCTTCAACATACTCACCCTG : 2625
U91341  : GCGTCTATGGGGTTCCGTTGATGGTTGTGAACGCGTTTCTGGTGTTGATCACTTATCTTCAACATACTCACCCTG :  938
RHA280  : GCGTCTATGGGGTTCCGTTGATGGTTGTGAACGCGTTTCTGGTGTTGATCACTTATCTTCAACATACTCACCCTG :  843
RHA801  : GCGTCTATGGGGTTCCGTTGATGGTTGTGAACGCGTTTCTGGTGTTGATCACTTATCTTCAACATACTCACCCTG :  843

            *       2640            *       2660            *       2680            *       2700
HA89    : GCTTGCCGCATTATGATAGCTCGGAATGGGAATGGTTAAAGGGAGCATTGGCGACAGTGGACCGTGACTATGGTG : 2700
HA341   : GCTTGCCGCATTATGATAGCTCGGAATGGGAATGGTTAAAGGGAGCATTGGCGACAGTGGACCGTGACTATGGTG : 2700
PHC     : GCTTGCCGCATTATGATAGCTCGGAATGGGAATGGTTAAAGGGAGCATTGGCGACAGTGGACCGTGACTATGGTG : 2700
PHD     : GCTTGCCGCATTATGATAGCTCGGAATGGGAATGGTTAAAGGGAGCATTGGCGACAGTGGACCGTGACTATGGTG : 2700
U91341  : GCTTGCCGCATTATGATAGCTCGGAATGGGAATGGTTAAAGGGAGCATTGGCGACAGTGGACCGTGACTATGGTG : 1013
RHA280  : GCTTGCCGCATTATGATAGCTCGGAATGGGAATGGTTAAAGGGAGCATTGGCGACAGTGGACCGTGACTATGGTG :  918
RHA801  : GCTTGCCGCATTATGATAGCTCGGAATGGGAATGGTTAAAGGGAGCATTGGCGACAGTGGACCGTGACTATGGTG :  918

                    *       2720            *       2740            *       2760            *
HA89    : TGTTGAACAAGGTGTTCCATCATATTACCGACACACATGTGGTGCACCATTTGTTTTCGACAATGCCTCATTATA : 2775
HA341   : TGTTGAACAAGGTGTTCCATCATATTACCGACACACATGTGGTGCACCATTTGTTTTCGACAATGCCTCATTATA : 2775
PHC     : TGTTGAACAAGGTGTTCCATCATATTACCGACACACATGTGGTGCACCATTTGTTTTCGACAATGCCTCATTATA : 2775
PHD     : TGTTGAACAAGGTGTTCCATCATATTACCGACACACATGTGGTGCACCATTTGTTTTCGACAATGCCTCATTATA : 2775
U91341  : TGTTGAACAAGGTGTTCCATCATATTACCGACACACATGTGGTGCACCATTTGTTTTCGACAATGCCTCATTATA : 1088
RHA280  : TGTTGAACAAGGTGTTCCATCATATTACCGACACACATGTGGTGCACCATTTGTTTTCGACAATGCCTCATTATA :  993
RHA801  : TGTTGAACAAGGTGTTCCATCATATTACCGACACACATGTGGTGCACCATTTGTTTTCGACAATGCCTCATTATA :  993

            2780            *       2800            *       2820            *       2840            *
HA89    : ATGCGATGGAAGCACAGAAGGCGCTGAGACCGGTGCTTGGGGAGTATTATCGGTTTGACAAGACCCCGTTTTATG : 2850
HA341   : ATGCGATGGAAGCACAGAAGGCGCTGAGACCGGTGCTTGGGGAGTATTATCGGTTTGACAAGACCCCGTTTTATG : 2850
PHC     : ATGCGATGGAAGCACAGAAGGCGCTGAGACCGGTGCTTGGGGAGTATTATCGGTTTGACAAGACCCCGTTTTATG : 2850
PHD     : ATGCGATGGAAGCACAGAAGGCGCTGAGACCGGTGCTTGGGGAGTATTATCGGTTTGACAAGACCCCGTTTTATG : 2850
U91341  : ATGCGATGGAAGCACAGAAGGCGCTGAGACCGGTGCTTGGGGAGTATTATCGGTTTGACAAGACCCCGTTTTATG : 1163
RHA280  : ATGCGATGGAAGCACAGAAGGCGCTGAGACCGGTGCTTGGGGAGTATTATCGGTTTGACAAGACCCCGTTTTATG : 1068
RHA801  : ATGCGATGGAAGCACAGAAGGCGCTGAGACCGGTGCTTGGGGAGTATTATCGGTTTGACAAGACCCCGTTTTATG : 1068

                2860            *       2880            *       2900            *       2920
HA89    : TAGCCATGTGGAGAGAGATGAAGGAATGTTTGTTTGTGGAGCAAGATGATGAAGGGAAAGGAGGTGTGTTTTGGT : 2925
HA341   : TAGCCATGTGGAGAGAGATGAAGGAATGTTTGTTTGTGGAGCAAGATGATGAAGGGAAAGGAGGTGTGTTTTGGT : 2925
PHC     : TAGCCATGTGGAGAGAGATGAAGGAATGTTTGTTTGTGGAGCAAGATGATGAAGGGAAAGGAGGTGTGTTTTGGT : 2925
PHD     : TAGCCATGTGGAGAGAGATGAAGGAATGTTTGTTTGTGGAGCAAGATGATGAAGGGAAAGGAGGTGTGTTTTGGT : 2925
U91341  : TAGCCATGTGGAGAGAGATGAAGGAATGTTTGTTTGTGGAGCAAGATGATGAAGGGAAAGGAGGTGTGTTTTGGT : 1238
RHA280  : TAGCCATGTGGAGAGAGATGAAGGAATGTTTGTTTGTGGAGCAAGATGATGAAGGGAAAGGAGGTGTGTTTTGGT : 1143
RHA801  : TAGCCATGTGGAGAGAGATGAAGGAATGTTTGTTTGTGGAGCAAGATGATGAAGGGAAAGGAGGTGTGTTTTGGT : 1143

                    *       2940            *       2960            *       2980            *       3000
HA89    : ACAAGAATAAGATGAATTAATAAGTGTTTAGTGGATGTCTATGCTTATTAAGTCTGTGACAATGGGTCTTGTCCT : 3000
HA341   : ACAAGAATAAGATGAATTAATAAGTGTTTAGTGGATGTCTATGCTTATTAAGTCTGTGACAATGGGTCTTGTCCT : 3000
PHC     : ACAAGAATAAGATGAATTAATAAGTGTTTAGTGGATGTCTATGCTTATTAAGTCTGTGACAATGGGTCTTGTCCT : 3000
PHD     : ACAAGAATAAGATGAATTAATAAGTGTTTAGTGGATGTCTATGCTTATTAAGTCTGTGACAATGGGTCTTGTCCT : 3000
U91341  : ACAAGAATAAGATGAATTAATAAGTGTTTAGTGGATGTCTATGCTTATTAAGTCTGTGACAATGGGTCTTGTCCT : 1313
RHA280  : ACAAGAATAAGATGAATTAATAAGTGTTTAGTGGATGTCTATGCTTATTAAGTCTGTGACAATGGGTCTTGTCCT : 1218
RHA801  : ACAAGAATAAGATGAATTAATAAGTGTTTAGTGGATGTCTATGCTTATTAAGTCTGTGACAATGGGTCTTGTCCT : 1218
```

Appendix B. Alignment of the genomic nucleic acid sequence of *FAD2-2* for 6 sunflower lines with the GenBank cDNA sequence AF251843. The start and stop codons are marked in uppercase and bold letters.



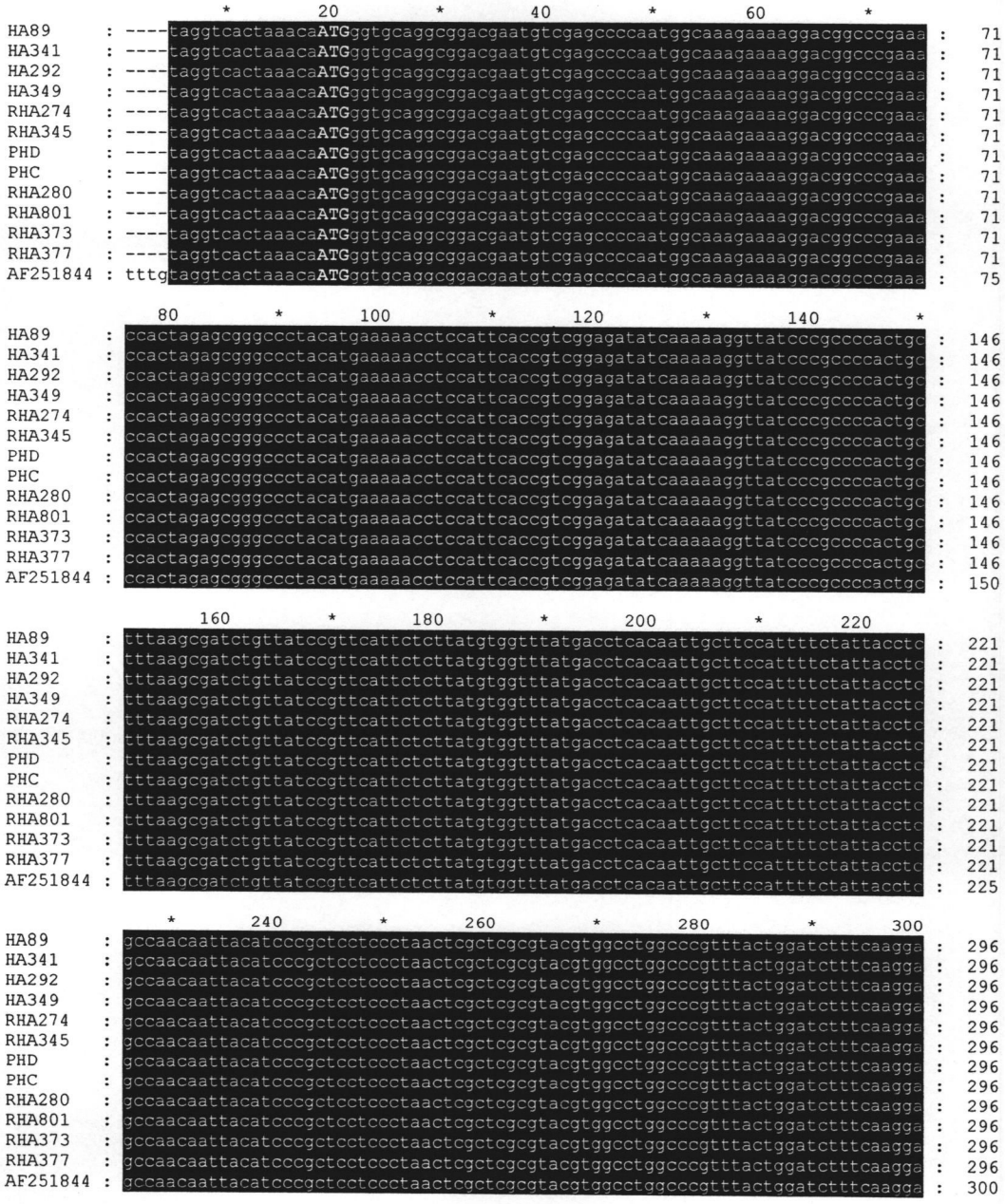
Appendix B. (Continued)



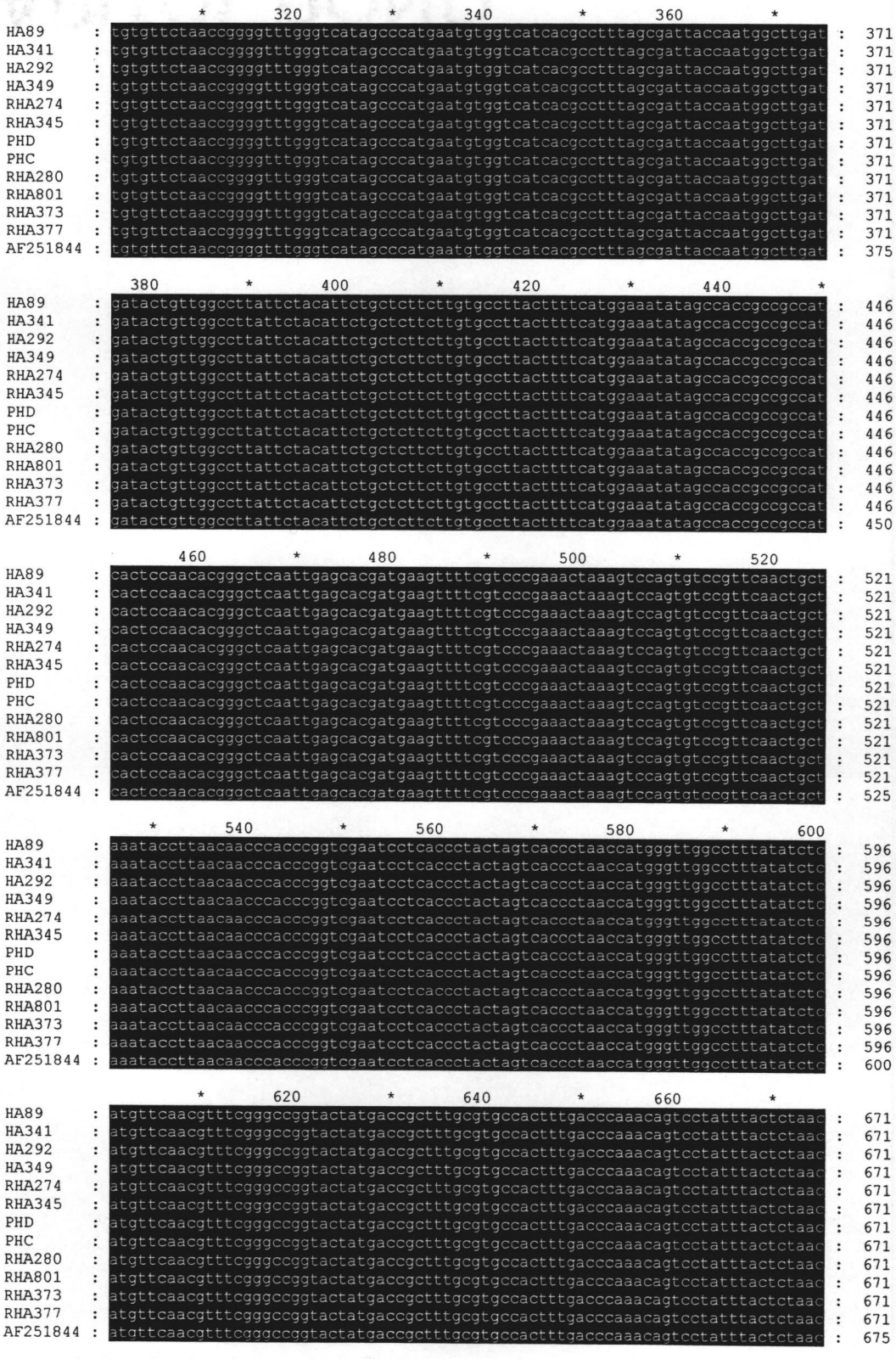
Appendix B. (Continued)

```
                  *      1220         *      1240         *      1260         *
HA89      : gaaattgtcttccttttgatgttatttaggttatgtttggtctttttatgtcagtttaatgaagttgtaataagt : 1271
HA341     : gaaattgtcttccttttgatgttatttaggttatgtttggtctttttatgtcagtttaatgaagttgtaataagt : 1271
HA292     : gaaattgtcttccttttgatgttatttaggttatgtttggtctttttatgtcagtttaatgaagttgtaataagt : 1271
HA349     : gaaattgtcttccttttgatgttatttaggttatgtttggtctttttatgtcagtttaatgaagttgtaataagT : 1271
PHD       : gaaattgtcttccttttgatgttatttaggttatgtttggtctttttatgtcagtttaatgaagttgtaataagt : 1271
PHC       : gaaattgtcttccttttgatgttatttaggttatgtttggtctttttatgtcagtttaatgaagttgtaataagt : 1271
AF251843  : gaaattgtcttccttttgatgttatttaggttatgtttggtctttttatgtcagtttaatgaagttgtaataagt : 1275

             1280         *      1300         *      1320         *      1340
HA89      : gttgtacctgaactattgtg---------------------------------------------------- : 1291
HA341     : gttgtacctgaactattgtg---------------------------------------------------- : 1291
HA292     : gttgtacctgaactattgtg---------------------------------------------------- : 1291
HA349     : gttgtacctgaactattgtg---------------------------------------------------- : 1291
PHD       : gttgtacctgaactattgtg---------------------------------------------------- : 1291
PHC       : gttgtacctgaactattgtg---------------------------------------------------- : 1291
AF251843  : gttgtacctgaactattgtgtttctagatcaccttcgttttcttttaagaagtttagcatcgttccaccc : 1345
```

Appendix C. Alignment of the genomic nucleic acid sequence of *FAD2-3* for 12 sunflower lines with the GenBank cDNA sequence AF251844. The start and stop codons are marked in uppercase and bold letters.



Appendix C. (Continued)



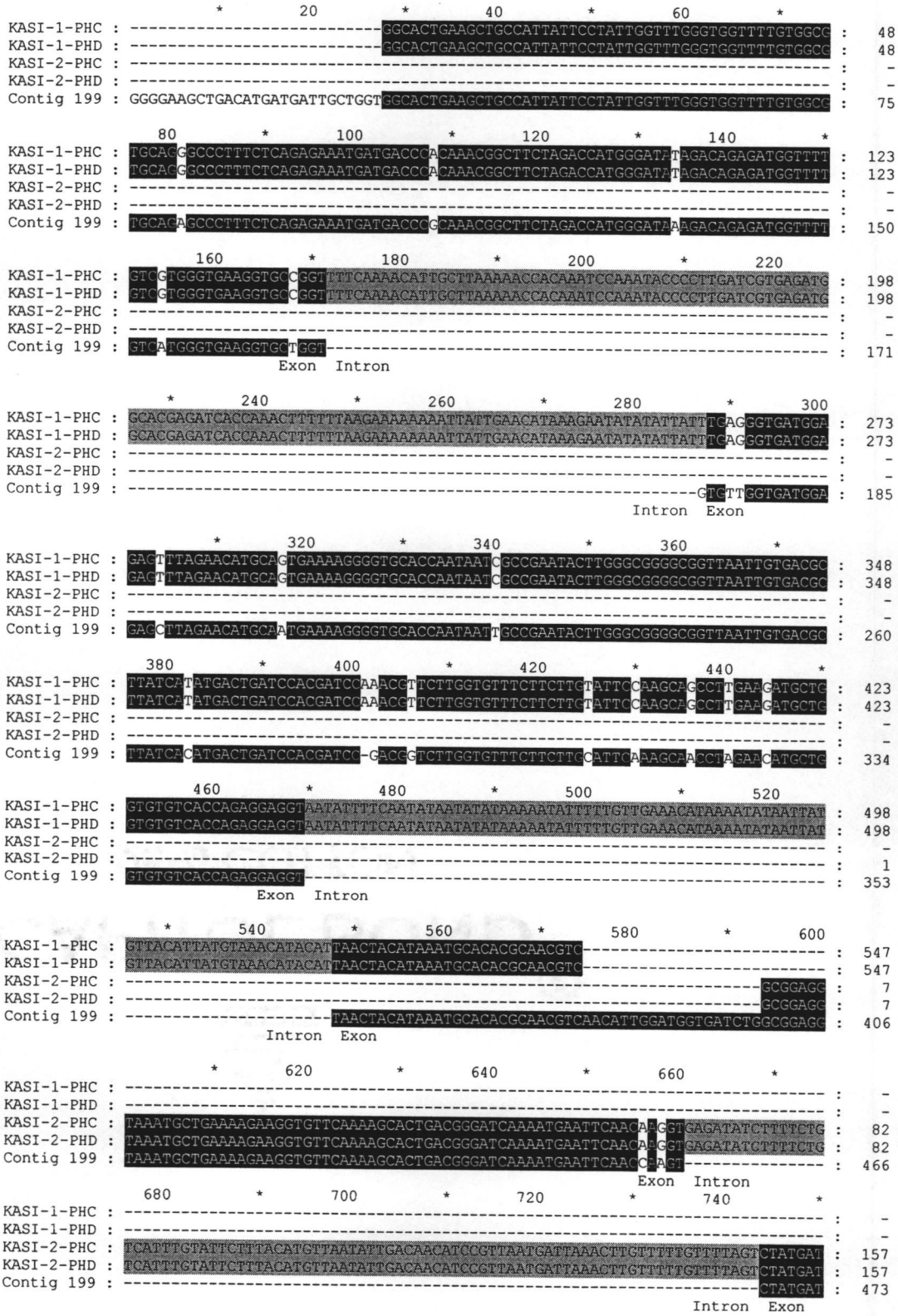
Appendix C. (Continued)

```
                680         *         700         *         720         *         740         *
HA89        : cgcgaacgggcccaaatattcatctccgatgccgggattttaaccgttttctacatacttttccgtctagcatcg : 746
HA341       : cgcgaacgggcccaaatattcatctccgatgccgggattttaaccgttttctacatacttttccgtctagcatcg : 746
HA292       : cgcgaacgggcccaaatattcatctccgatgccgggattttaaccgttttctacatacttttccgtctagcatcg : 746
HA349       : cgcgaacgggcccaaatattcatctccgatgccgggattttaaccgttttctacatacttttccgtctagcatcg : 746
RHA274      : cgcgaacgggcccaaatattcatctccgatgccgggattttaaccgttttctacatacttttccgtctagcatcg : 746
RHA345      : cgcgaacgggcccaaatattcatctccgatgccgggattttaaccgttttctacatacttttccgtctagcatcg : 746
PHD         : cgcgaacgggcccaaatattcatctccgatgccgggattttaaccgttttctacatacttttccgtctagcatcg : 746
PHC         : cgcgaacgggcccaaatattcatctccgatgccgggattttaaccgttttctacatacttttccgtctagcatcg : 746
RHA280      : cgcgaacgggcccaaatattcatctccgatgccgggattttaaccgttttctacatacttttccgtctagcatcg : 746
RHA801      : cgcgaacgggcccaaatattcatctccgatgccgggattttaaccgttttctacatacttttccgtctagcatcg : 746
RHA373      : cgcgaacgggcccaaatattcatctccgatgccgggattttaaccgttttctacatacttttccgtctagcatcg : 746
RHA377      : cgcgaacgggcccaaatattcatctccgatgccgggattttaaccgttttctacatacttttccgtctagcatcg : 746
AF251844    : cgcgaacgggcccaaatattcatctccgatgccgggattttaaccgttttctacatacttttccgtctagcatcg : 750

                760         *         780         *         800         *         820
HA89        : acaaaagggctcgtatgggtgttaaccatgtacggtggaccgttacttgtagttaacggtttcttagtccttatc : 821
HA341       : acaaaagggctcgtatgggtgttaaccatgtacggtggaccgttacttgtagttaacggtttcttagtccttatc : 821
HA292       : acaaaagggctcgtatgggtgttaaccatgtacggtggaccgttacttgtagttaacggtttcttagtccttatc : 821
HA349       : acaaaagggctcgtatgggtgttaaccatgtacggtggaccgttacttgtagttaacggtttcttagtccttatc : 821
RHA274      : acaaaagggctcgtatgggtgttaaccatgtacggtggaccgttacttgtagttaacggtttcttagtccttatc : 821
RHA345      : acaaaagggctcgtatgggtgttaaccatgtacggtggaccgttacttgtagttaacggtttcttagtccttatc : 821
PHD         : acaaaagggctcgtatgggtgttaaccatgtacggtggaccgttacttgtagttaacggtttcttagtccttatc : 821
PHC         : acaaaagggctcgtatgggtgttaaccatgtacggtggaccgttacttgtagttaacggtttcttagtccttatc : 821
RHA280      : acaaaagggctcgtatgggtgttaaccatgtacggtggaccgttacttgtagttaacggtttcttagtccttatc : 821
RHA801      : acaaaagggctcgtatgggtgttaaccatgtacggtggaccgttacttgtagttaacggtttcttagtccttatc : 821
RHA373      : acaaaagggctcgtatgggtgttaaccatgtacggtggaccgttacttgtagttaacggtttcttagtccttatc : 821
RHA377      : acaaaagggctcgtatgggtgttaaccatgtacggtggaccgttacttgtagttaacggtttcttagtccttatc : 821
AF251844    : acaaaagggctcgtatgggtgttaaccatgtacggtggaccgttacttgtagttaacggtttcttagtccttatc : 825

                *         840         *         860         *         880         *         900
HA89        : acattcttgcaacacacccacccgtcattaccacactacgactcaaccgaatgggactggttacgcggggctcta : 896
HA341       : acattcttgcaacacacccacccgtcattaccacactacgactcaaccgaatgggactggttacgcggggctcta : 896
HA292       : acattcttgcaacacacccacccgtcattaccacactacgactcaaccgaatgggactggttacgcggggctcta : 896
HA349       : acattcttgcaacacacccacccgtcattaccacactacgactcaaccgaatgggactggttacgcggggctcta : 896
RHA274      : acattcttgcaacacacccacccgtcattaccacactacgactcaaccgaatgggactggttacgcggggctcta : 896
RHA345      : acattcttgcaacacacccacccgtcattaccacactacgactcaaccgaatgggactggttacgcggggctcta : 896
PHD         : acattcttgcaacacacccacccgtcattaccacactacgactcaaccgaatgggactggttacgcggggctcta : 896
PHC         : acattcttgcaacacacccacccgtcattaccacactacgactcaaccgaatgggactggttacgcggggctcta : 896
RHA280      : acattcttgcaacacacccacccgtcattaccacactacgactcaaccgaatgggactggttacgcggggctcta : 896
RHA801      : acattcttgcaacacacccacccgtcattaccacactacgactcaaccgaatgggactggttacgcggggctcta : 896
RHA373      : acattcttgcaacacacccacccgtcattaccacactacgactcaaccgaatgggactggttacgcggggctcta : 896
RHA377      : acattcttgcaacacacccacccgtcattaccacactacgactcaaccgaatgggactggttacgcggggctcta : 896
AF251844    : acattcttgcaacacacccacccgtcattaccacactacgactcaaccgaatgggactggttacgcggggctcta : 900

                    *         920         *         940         *         960         *
HA89        : gccaccgtagaccgagactatgggatcttaaacaaggtgttccataacataaccgatacgcacgttacacaccat : 971
HA341       : gccaccgtagaccgagactatgggatcttaaacaaggtgttccataacataaccgatacgcacgttacacaccat : 971
HA292       : gccaccgtagaccgagactatgggatcttaaacaaggtgttccataacataaccgatacgcacgttacacaccat : 971
HA349       : gccaccgtagaccgagactatgggatcttaaacaaggtgttccataacataaccgatacgcacgttacacaccat : 971
RHA274      : gccaccgtagaccgagactatgggatcttaaacaaggtgttccataacataaccgatacgcacgttacacaccat : 971
RHA345      : gccaccgtagaccgagactatgggatcttaaacaaggtgttccataacataaccgatacgcacgttacacaccat : 971
PHD         : gccaccgtagaccgagactatgggatcttaaacaaggtgttccataacataaccgatacgcacgttacacaccat : 971
PHC         : gccaccgtagaccgagactatgggatcttaaacaaggtgttccataacataaccgatacgcacgttacacaccat : 971
RHA280      : gccaccgtagaccgagactatgggatcttaaacaaggtgttccataacataaccgatacgcacgttacacaccat : 971
RHA801      : gccaccgtagaccgagactatgggatcttaaacaaggtgttccataacataaccgatacgcacgttacacaccat : 971
RHA373      : gccaccgtagaccgagactatgggatcttaaacaaggtgttccataacataaccgatacgcacgttacacaccat : 971
RHA377      : gccaccgtagaccgagactatgggatcttaaacaaggtgttccataacataaccgatacgcacgttacacaccat : 971
AF251844    : gccaccgtagaccgagactatgggatcttaaacaaggtgttccataacataaccgatacgcacgttacacaccat : 975

                980         *         1000        *         1020        *         1040        *
HA89        : ttgttctccaccatgccacattatcatgcaatggaagcaacaaaggcgattaagccgattttgggagattattat : 1046
HA341       : ttgttctccaccatgccacattatcatgcaatggaagcaacaaaggcgattaagccgattttgggagattattat : 1046
HA292       : ttgttctccaccatgccacattatcatgcaatggaagcaacaaaggcgattaagccgattttgggagattattat : 1046
HA349       : ttgttctccaccatgccacattatcatgcaatggaagcaacaaaggcgattaagccgattttgggagattattat : 1046
RHA274      : ttgttctccaccatgccacattatcatgcaatggaagcaacaaaggcgattaagccgattttgggagattattat : 1046
RHA345      : ttgttctccaccatgccacattatcatgcaatggaagcaacaaaggcgattaagccgattttgggagattattat : 1046
PHD         : ttgttctccaccatgccacattatcatgcaatggaagcaacaaaggcgattaagccgattttgggagattattat : 1046
PHC         : ttgttctccaccatgccacattatcatgcaatggaagcaacaaaggcgattaagccgattttgggagattattat : 1046
RHA280      : ttgttctccaccatgccacattatcatgcaatggaagcaacaaaggcgattaagccgattttgggagattattat : 1046
RHA801      : ttgttctccaccatgccacattatcatgcaatggaagcaacaaaggcgattaagccgattttgggagattattat : 1046
RHA373      : ttgttctccaccatgccacattatcatgcaatggaagcaacaaaggcgattaagccgattttgggagattattat : 1046
RHA377      : ttgttctccaccatgccacattatcatgcaatggaagcaacaaaggcgattaagccgattttgggagattattat : 1046
AF251844    : ttgttctccaccatgccacattatcatgcaatggaagcaacaaaggcgattaagccgattttgggagattattat : 1050
```

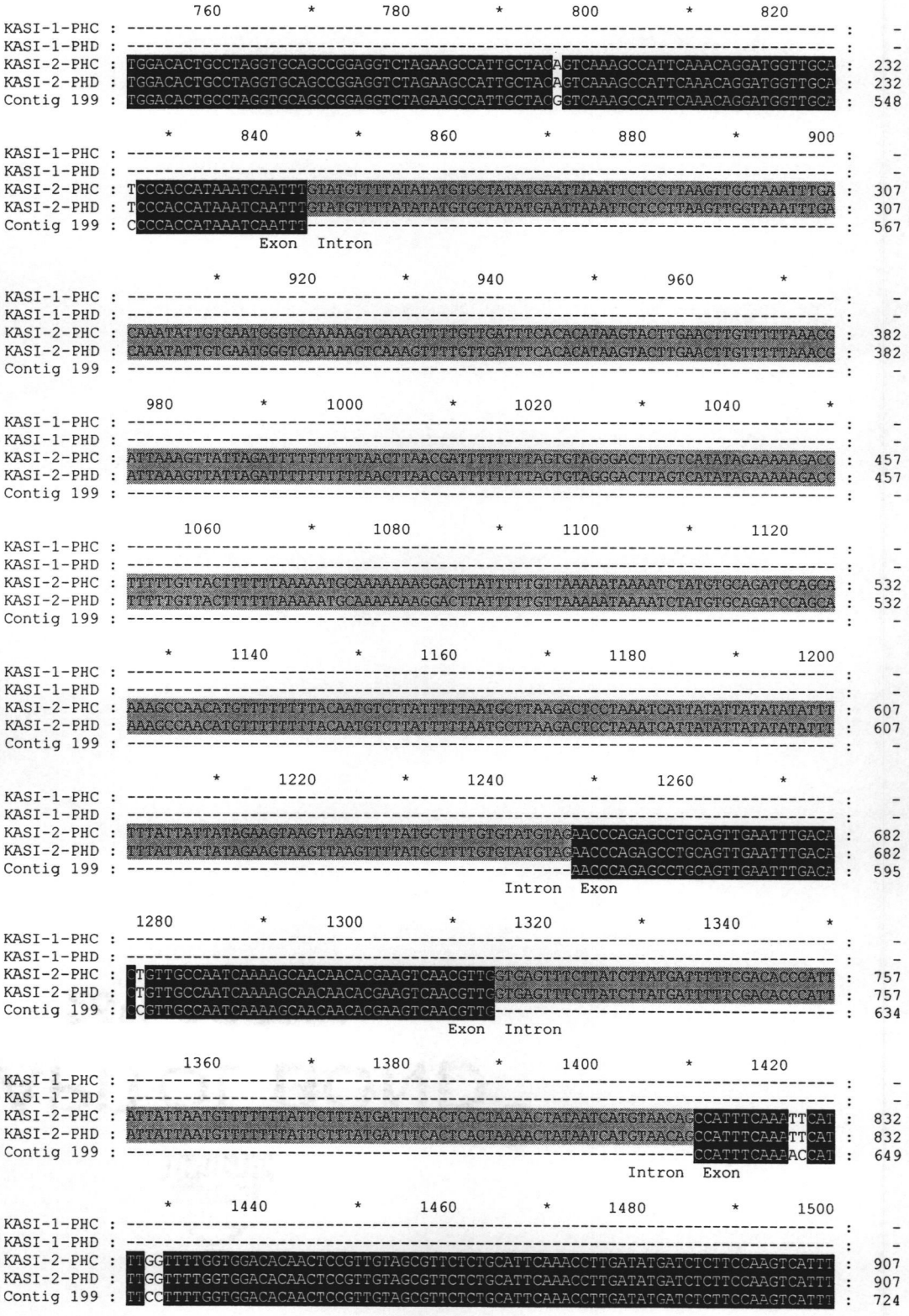
Appendix C. (Continued)

```
                1060         *       1080         *       1100         *       1120
HA89       : caattcgacgggacatcgatcttcaaggccatgtatagggagacaaaggaatgcatctatgttgataaagatgag : 1121
HA341      : caattcgacgggacatcgatcttcaaggccatgtatagggagacaaaggaatgcatctatgttgataaagatgac : 1121
HA292      : caattcgacgggacatcgatcttcaaggccatgtatagggagacaaaggaatgcatctatgttgataaagatgac : 1121
HA349      : caattcgacgggacatcgatcttcaaggccatgtatagggagacaaaggaatgcatctatgttgataaagatgac : 1121
RHA274     : caattcgacgggacatcgatcttcaaggccatgtatagggagacaaaggaatgcatctatgttgataaagatgac : 1121
RHA345     : caattcgacgggacatcgatcttcaaggccatgtatagggagacaaaggaatgcatctatgttgataaagatgac : 1121
PHD        : caattcgacgggacatcgatcttcaaggccatgtatagggagacaaaggaatgcatctatgttgataaagatgac : 1121
PHC        : caattcgacgggacatcgatcttcaaggccatgtatagggagacaaaggaatgcatctatgttgataaagatgac : 1121
RHA280     : caattcgacgggacatcgatcttcaaggccatgtatagggagacaaaggaatgcatctatgttgataaagatgac : 1121
RHA801     : caattcgacgggacatcgatcttcaaggccatgtatagggagacaaaggaatgcatctatgttgataaagatgac : 1121
RHA373     : caattcgacgggacatcgatcttcaaggccatgtatagggagacaaaggaatgcatctatgttgataaagatgac : 1121
RHA377     : caattcgacgggacatcgatcttcaaggccatgtatagggagacaaaggaatgcatctatgttgataaagatgac : 1121
AF251844   : caattcgacgggacatcgatcttcaaggccatgtatagggagacaaaggaatgcatctatgttgataaagatgac : 1125

              *       1140         *       1160         *       1180         *       1200
HA89       : gatgttaaggatggtgtctattggtaccgtaataagatcTGAtctgatctgatatgatacattcatgtttggtc- : 1195
HA341      : gatgttaaggatggtgtctattggtaccgtaataagatcTGAtctgatctgatatgatacattcatgtttggtc- : 1195
HA292      : gatgttaaggatggtgtctattggtaccgtaataagatcTGAtctgatctgatatgatacattcatgtttggtc- : 1195
HA349      : gatgttaaggatggtgtctattggtaccgtaataagatcTGAtctgatctgatatgatacattcatgtttggtc- : 1195
RHA274     : gatgttaaggatggtgtctattggtaccgtaataagatcTGAtctgatctgatatgatacattcatgtttggtc- : 1195
RHA345     : gatgttaaggatggtgtctattggtaccgtaataagatcTGAtctgatctgatatgatacattcatgtttggtc- : 1195
PHD        : gatgttaaggatggtgtctattggtaccgtaataagatcTGAtctgatctgatatgatacattcatgtttggtc- : 1195
PHC        : gatgttaaggatggtgtctattggtaccgtaataagatcTGAtctgatctgatatgatacattcatgtttggtc- : 1195
RHA280     : gatgttaaggatggtgtctattggtaccgtaataagatcTGAtctgatctgatatgatacattcatgtttggtc- : 1195
RHA801     : gatgttaaggatggtgtctattggtaccgtaataagatcTGAtctgatctgatatgatacattcatgtttggtc- : 1195
RHA373     : gatgttaaggatggtgtctattggtaccgtaataagatcTGAtctgatctgatatgatacattcatgtttggtc- : 1195
RHA377     : gatgttaaggatggtgtctattggtaccgtaataagatcTGAtctgatctgatatgatacattcatgtttggtc- : 1195
AF251844   : gatgttaaggatggtgtctattggtaccgtaataagatcTGAtctgatctgatatgatacattcatgtttggtca : 1200

                      *       1220         *       1240         *       1260
HA89       : ------------------------------------------------------------ :    -
HA341      : ------------------------------------------------------------ :    -
HA292      : ------------------------------------------------------------ :    -
HA349      : ------------------------------------------------------------ :    -
RHA274     : ------------------------------------------------------------ :    -
RHA345     : ------------------------------------------------------------ :    -
PHD        : ------------------------------------------------------------ :    -
PHC        : ------------------------------------------------------------ :    -
RHA280     : ------------------------------------------------------------ :    -
RHA801     : ------------------------------------------------------------ :    -
RHA373     : ------------------------------------------------------------ :    -
RHA377     : ------------------------------------------------------------ :    -
AF251844   : ctgatgtcgtttaagatggacatgtaacttatttagttaagatgaataagtgttgtact--- : 1259
```

Appendix D. Alignment of the partial genomic nucleic acid sequences of *KASI* for two sunflower lines with the corresponding *KASI* contig 1997.



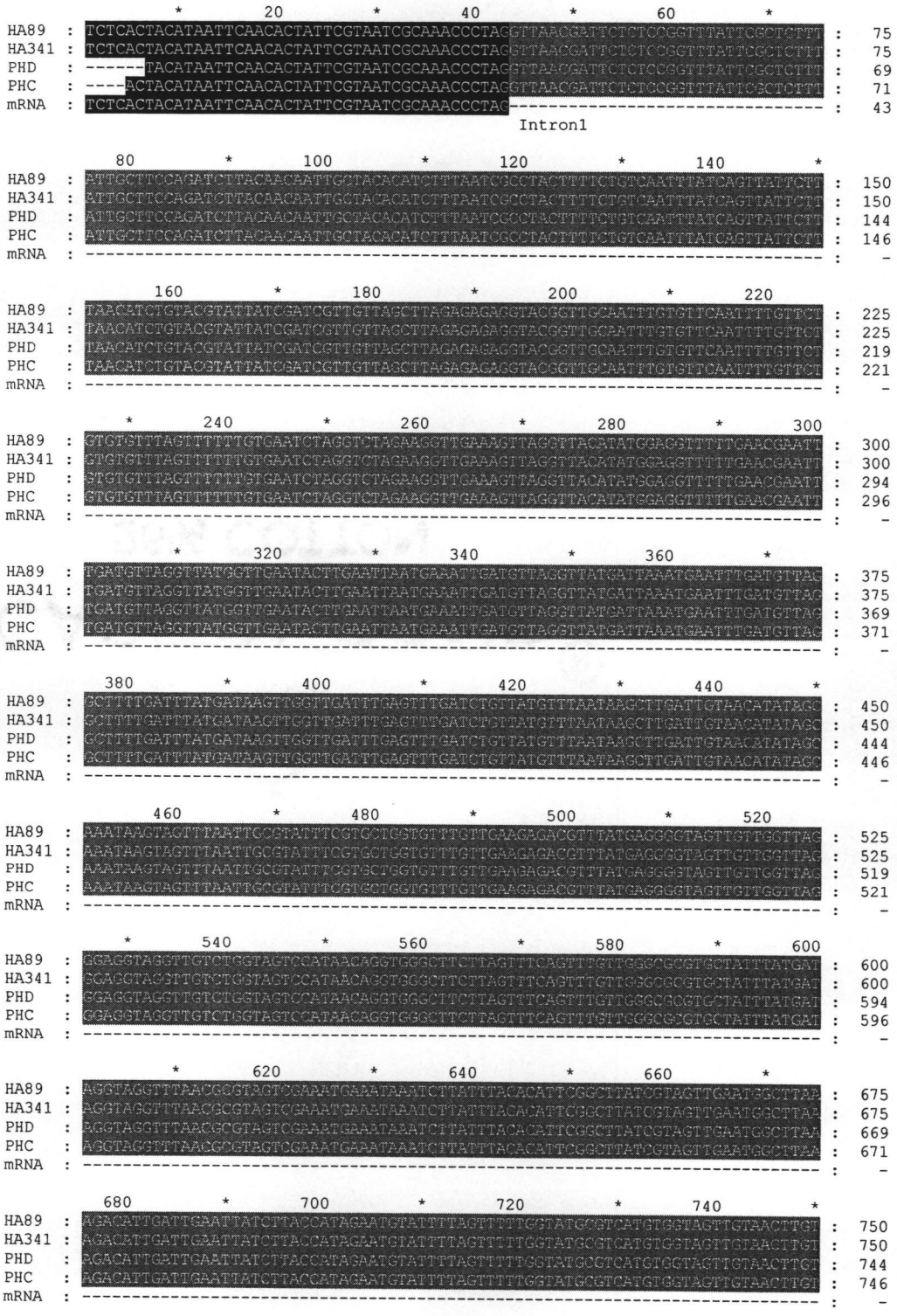
Appendix D. (Continued)



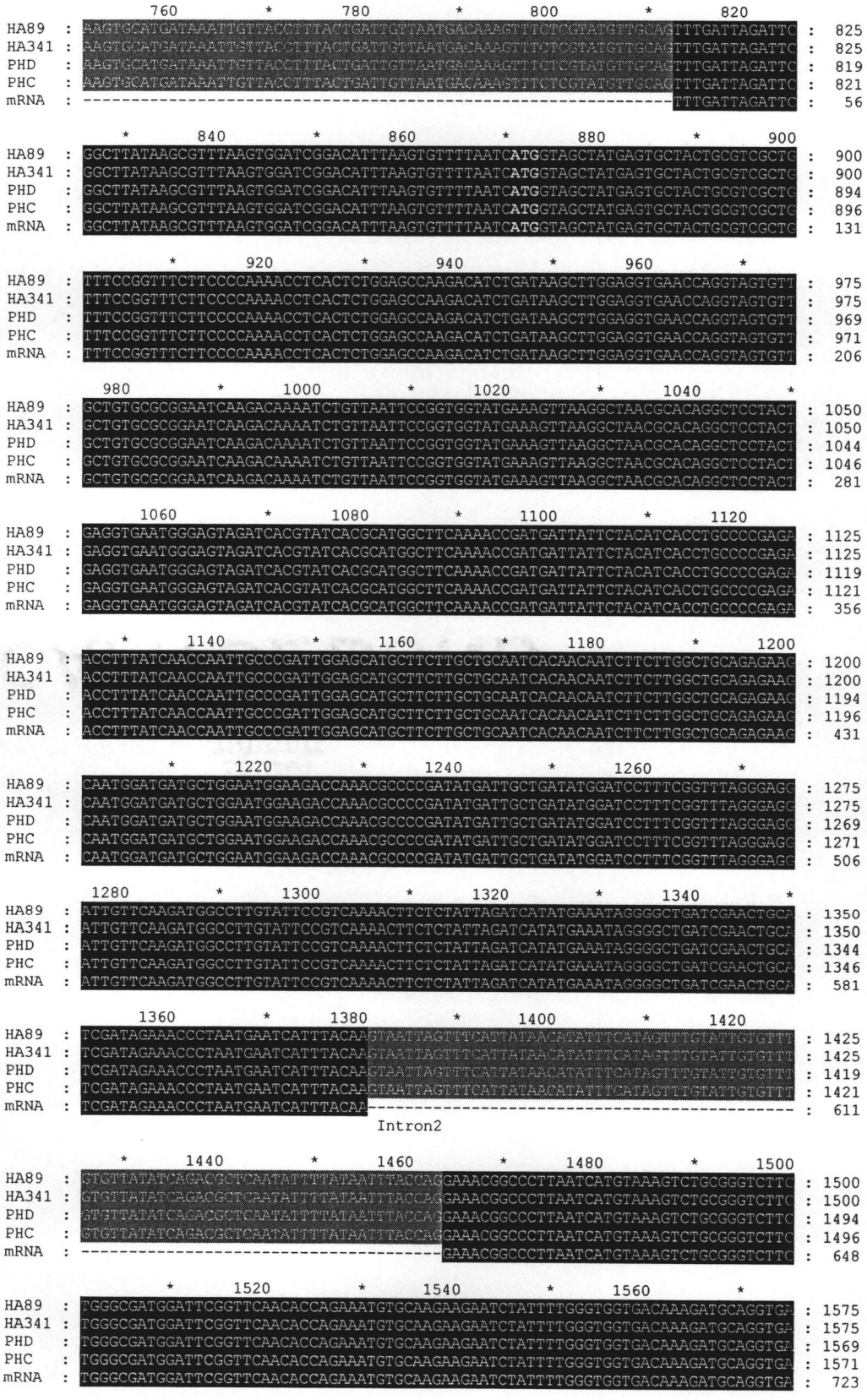
Appendix D. (Continued)

```
                    *      1520         *      1540         *      1560         *
KASI-1-PHC : ------------------------------------------------------------------------------------ :    -
KASI-1-PHD : ------------------------------------------------------------------------------------ :    -
KASI-2-PHC : TTGAGGCTTCATCAGTTGTTCTTGTTCTTTATTGTATGTCATTATAAGCGCGTTTTGATTTTGTAATCGTTGTTT :  982
KASI-2-PHD : TTGAGGCTTCATCAGTTGTTCTTGTTCTTTATTGTATGTCATTATAAGCGCGTTTTGATTTTGTAATCGTTGTTT :  982
Contig 199 : TTGAGGCTTCGTCAGTTGTTCTTGTTCTTTATTGTATGTCATTATAAGCGCGTTTTGATTTTGTAATCGCTGTTT :  799

                1580         *      1600         *      1620         *      1640         *
KASI-1-PHC : ------------------------------------------------------------------------------------ :    -
KASI-1-PHD : ------------------------------------------------------------------------------------ :    -
KASI-2-PHC : TGTATGTGATTCTGGTTTAAAAGACATCTACGACATTTGCATACCGATTTAGGCATTGAATTTTGAGTTGGGACA : 1057
KASI-2-PHD : TGTATGTGATTCTGGTTTAAAAGACATCTACGACATTTGCATACCGATTTAGGCATTGAATTTTGAGTTGGGACA : 1057
Contig 199 : TGAATGTGATTCTGGTTTAAAAGACATCTACGACATTTGCATACCGATTTAGGCATTGAATTTTGAGTTGGGACA :  874

KASI-1-PHC : -------- :    -
KASI-1-PHD : -------- :    -
KASI-2-PHC : AGTTTGGG : 1065
KASI-2-PHD : AGTTTGGG : 1065
Contig 199 : AGTTTGGG :  882
```

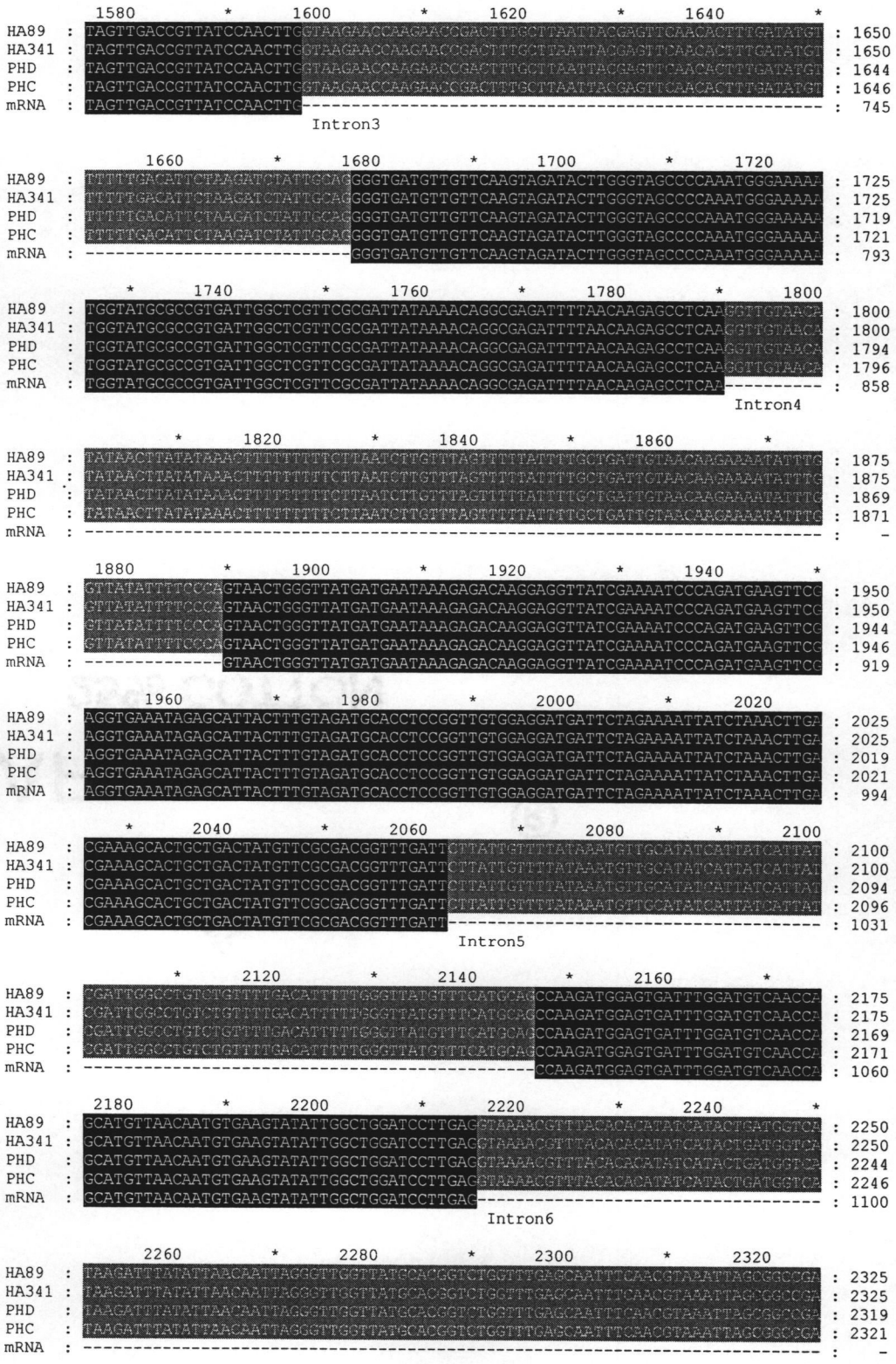
Appendix E. Alignment of the genomic nucleic acid sequence of *FATB* for four sunflower lines with the GenBank cDNA sequence AF036565. The introns are denoted in the alignment. The start and the stop codons are indicated in bold letters.



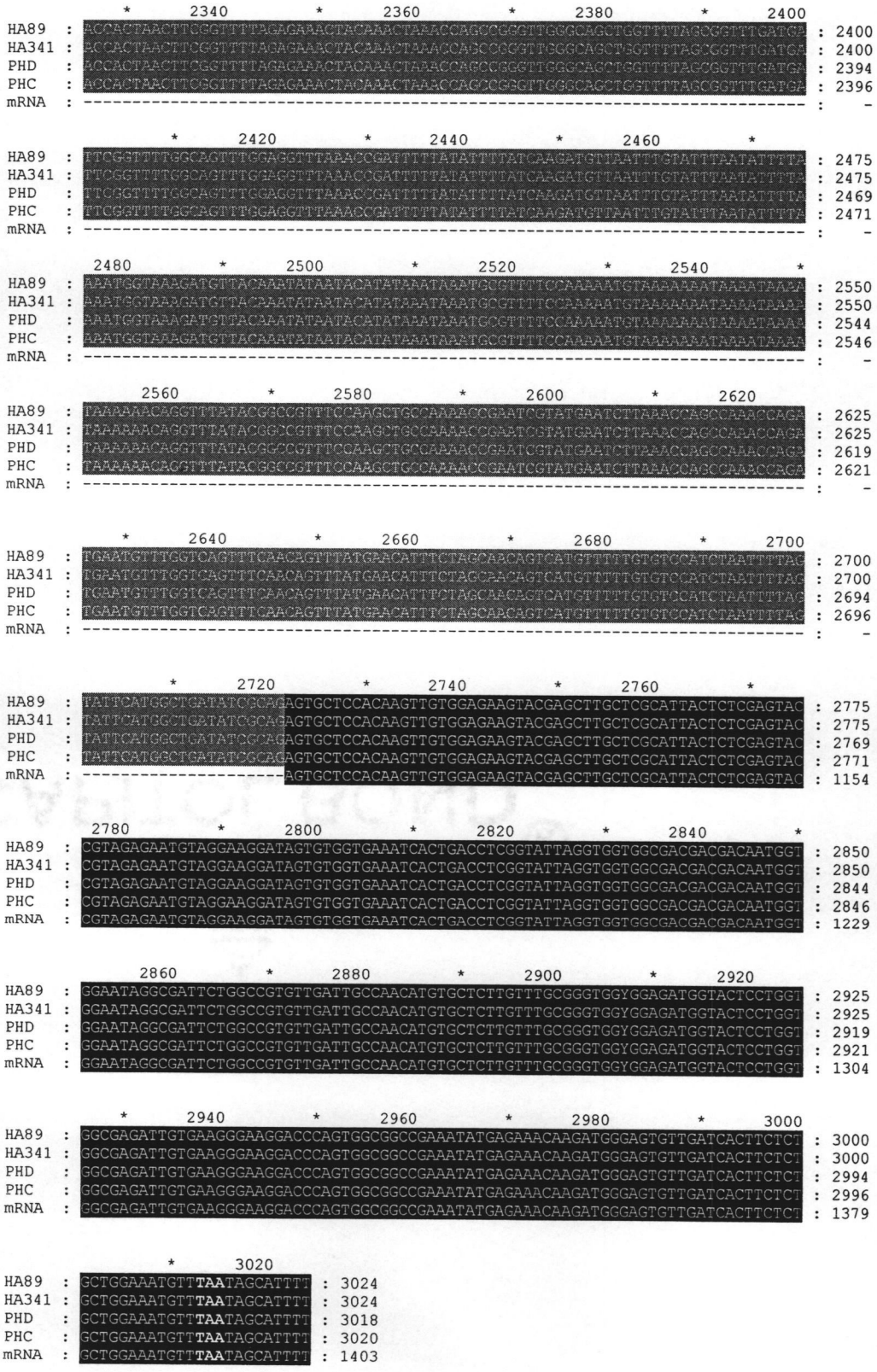
Appendix E. (Continued)



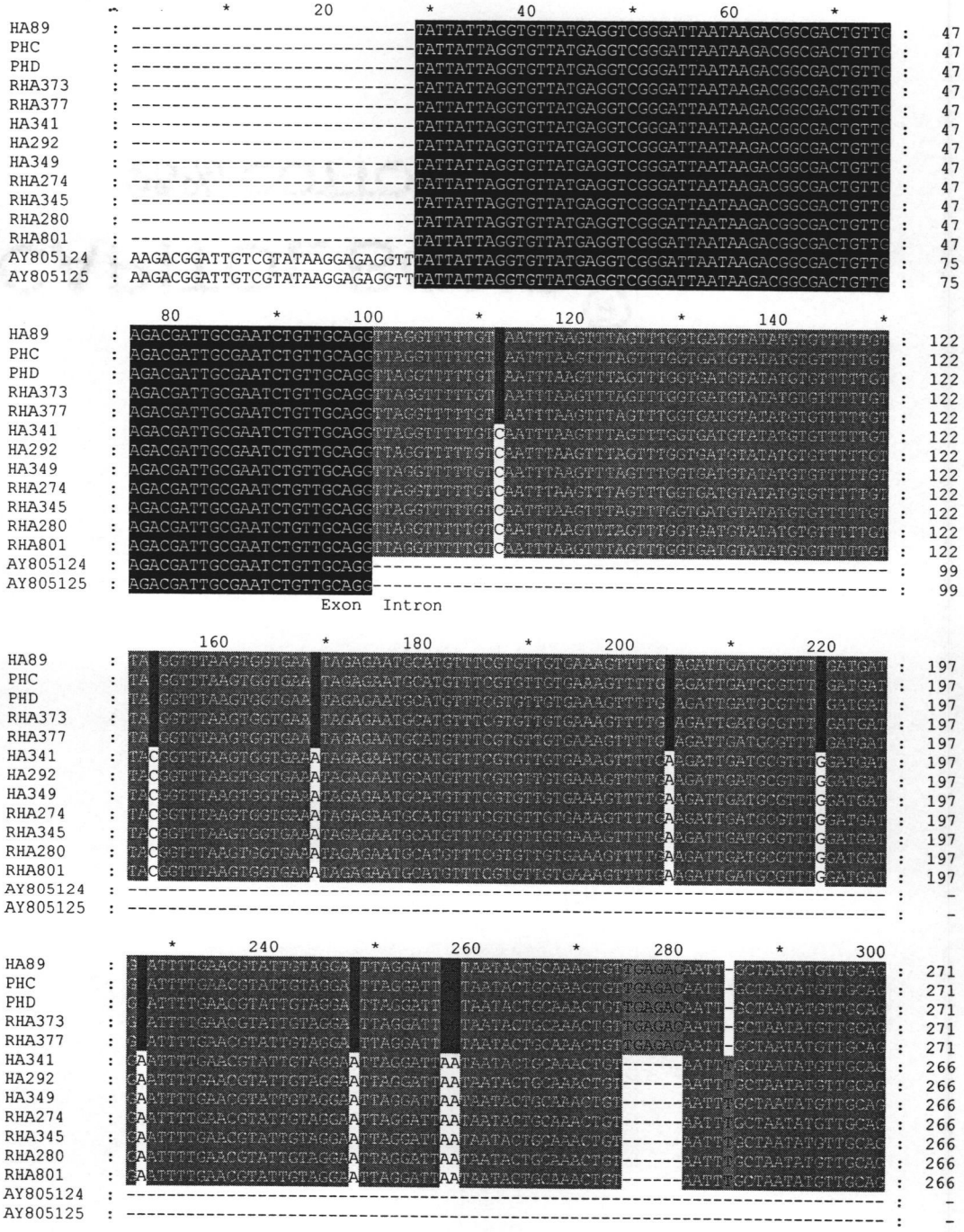
Appendix E. (Continued)



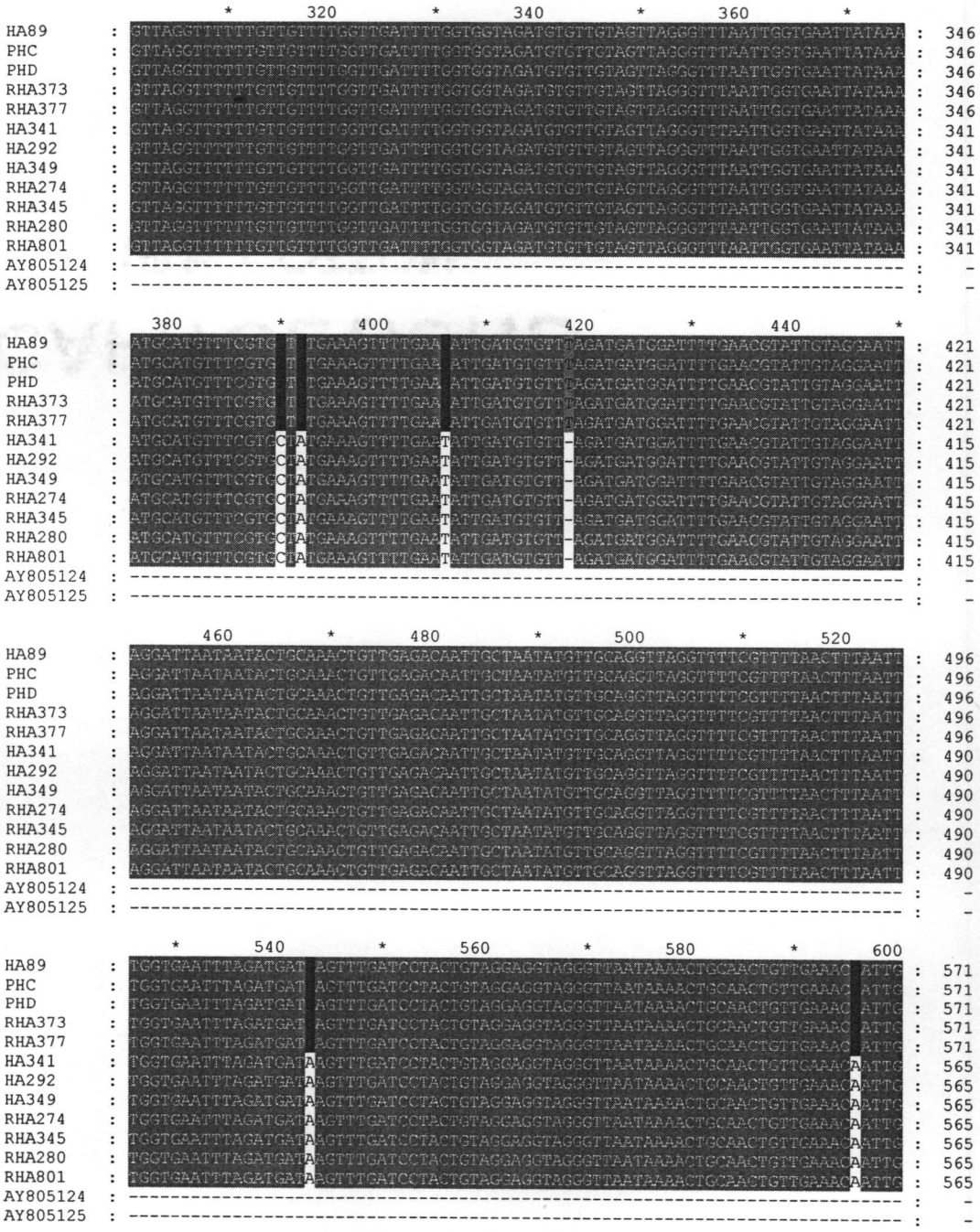
Appendix E. (Continued)



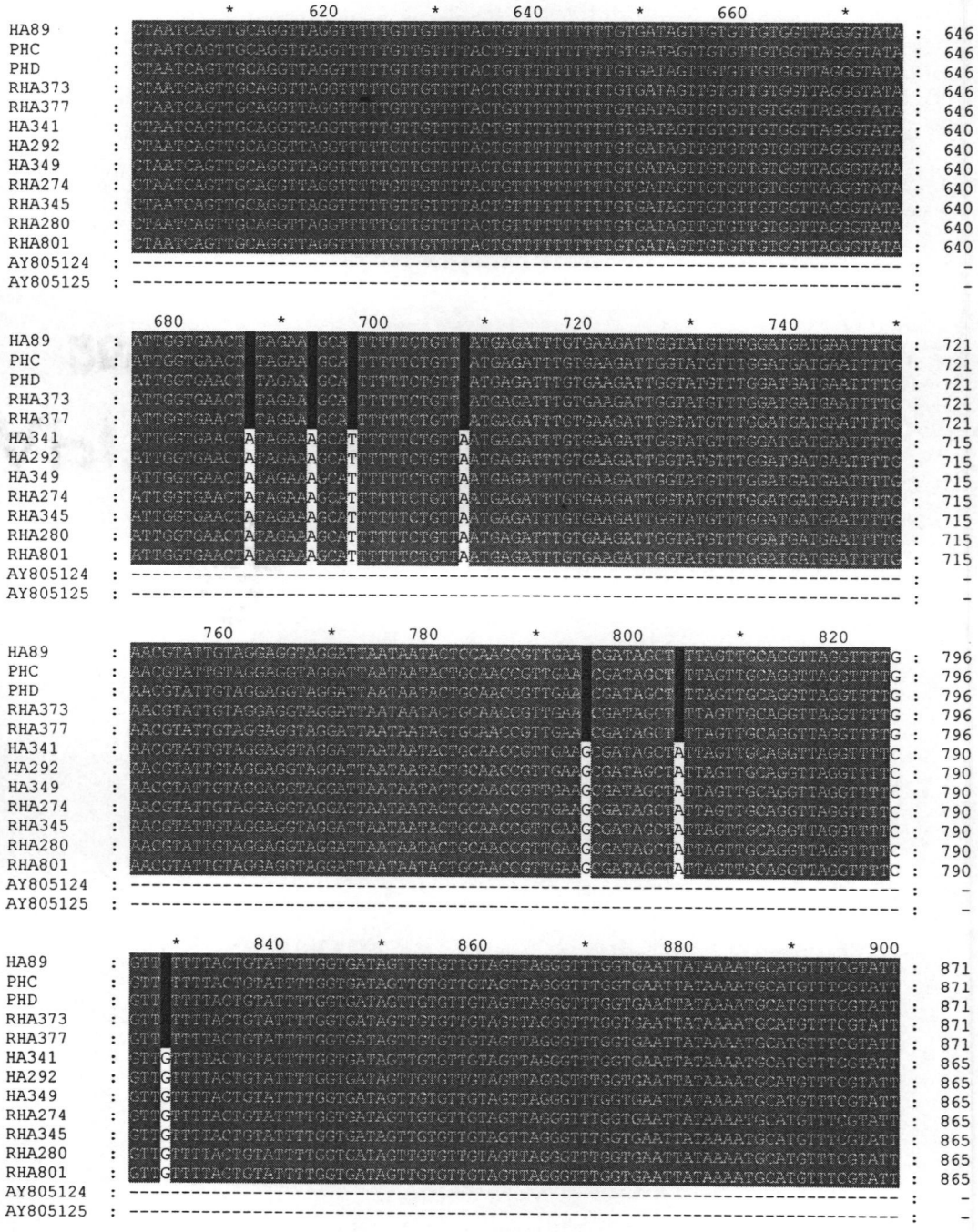
Appendix F. Alignment of the partial genomic nucleic acid sequences of *FATA* for 12 sunflower lines with the corresponding *FATA* EST sequences (AY805124 and AY805125).



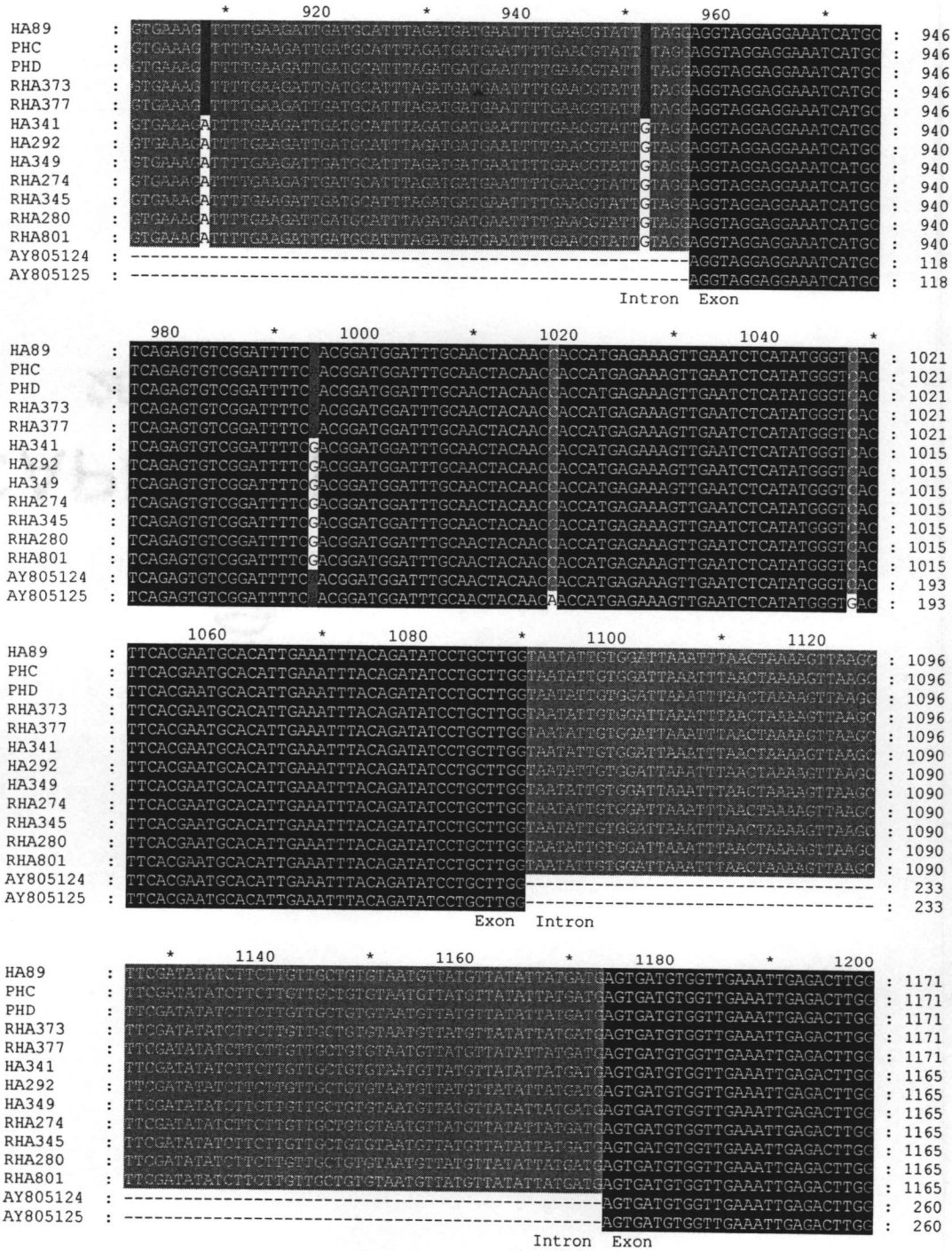
Appendix F. (Continued)



Appendix F. (Continued)



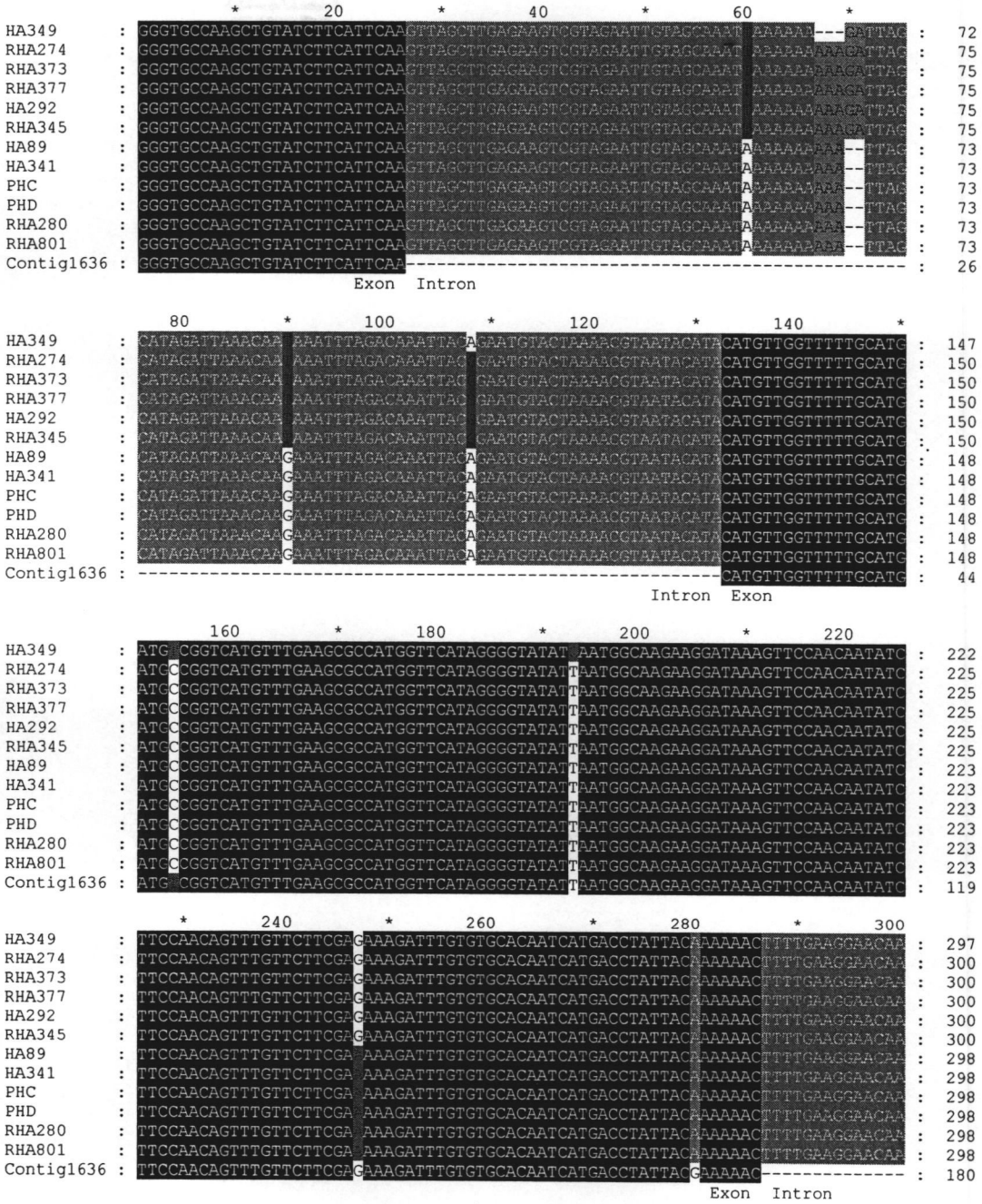
Appendix F. (Continued)



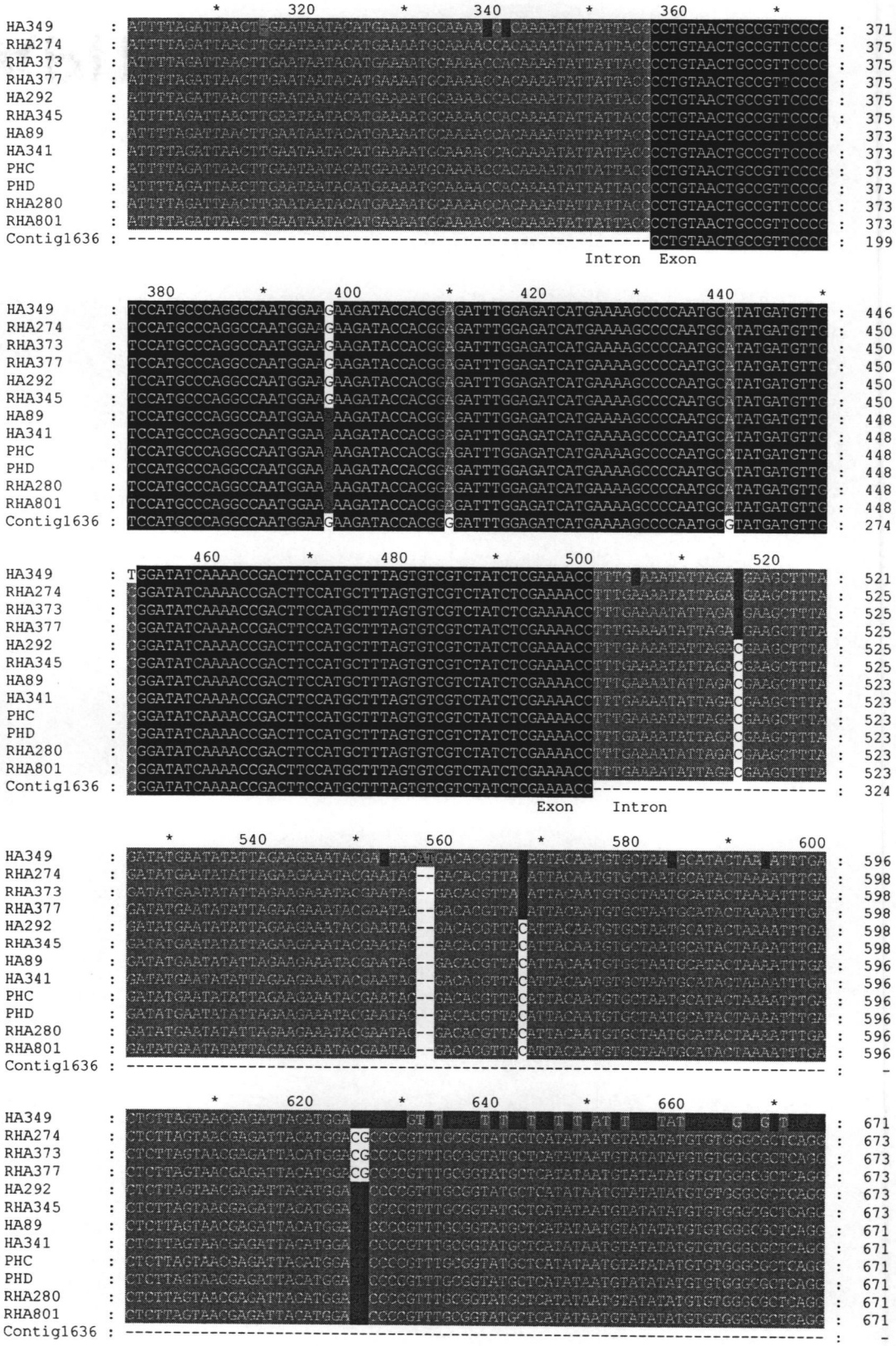
Appendix F. (Continued)

```
                    *      1220          *      1240
HA89        :  TGTCAAGGTGAAGGGAGA------------------------------ : 1189
PHC         :  TGTCAAGGTGAAGGGAGA------------------------------ : 1189
PHD         :  TGTCAAGGTGAAGGGAGA------------------------------ : 1189
RHA373      :  TGTCAAGGTGAAGGGAGA------------------------------ : 1189
RHA377      :  TGTCAAGGTGAAGGGAGA------------------------------ : 1189
HA341       :  TGTCAAGGTGAAGGGAGA------------------------------ : 1183
HA292       :  TGTCAAGGTGAAGGGAGA------------------------------ : 1183
HA349       :  TGTCAAGGTGAAGGGAGA------------------------------ : 1183
RHA274      :  TGTCAAGGTGAAGGGAGA------------------------------ : 1183
RHA345      :  TGTCAAGGTGAAGGGAGA------------------------------ : 1183
RHA280      :  TGTCAAGGTGAAGGGAGA------------------------------ : 1183
RHA801      :  TGTCAAGGTGAAGGGAGA------------------------------ : 1183
AY805124    :  TGTCAAGGTGAAGGGAGAATCGGGACTAGACGTGATTGGATTATCAAAG :  309
AY805125    :  TGTCAAGGTGAAGGGAGAATCGGGACTAGACGTGATTGGATTATCAAAG :  309
```

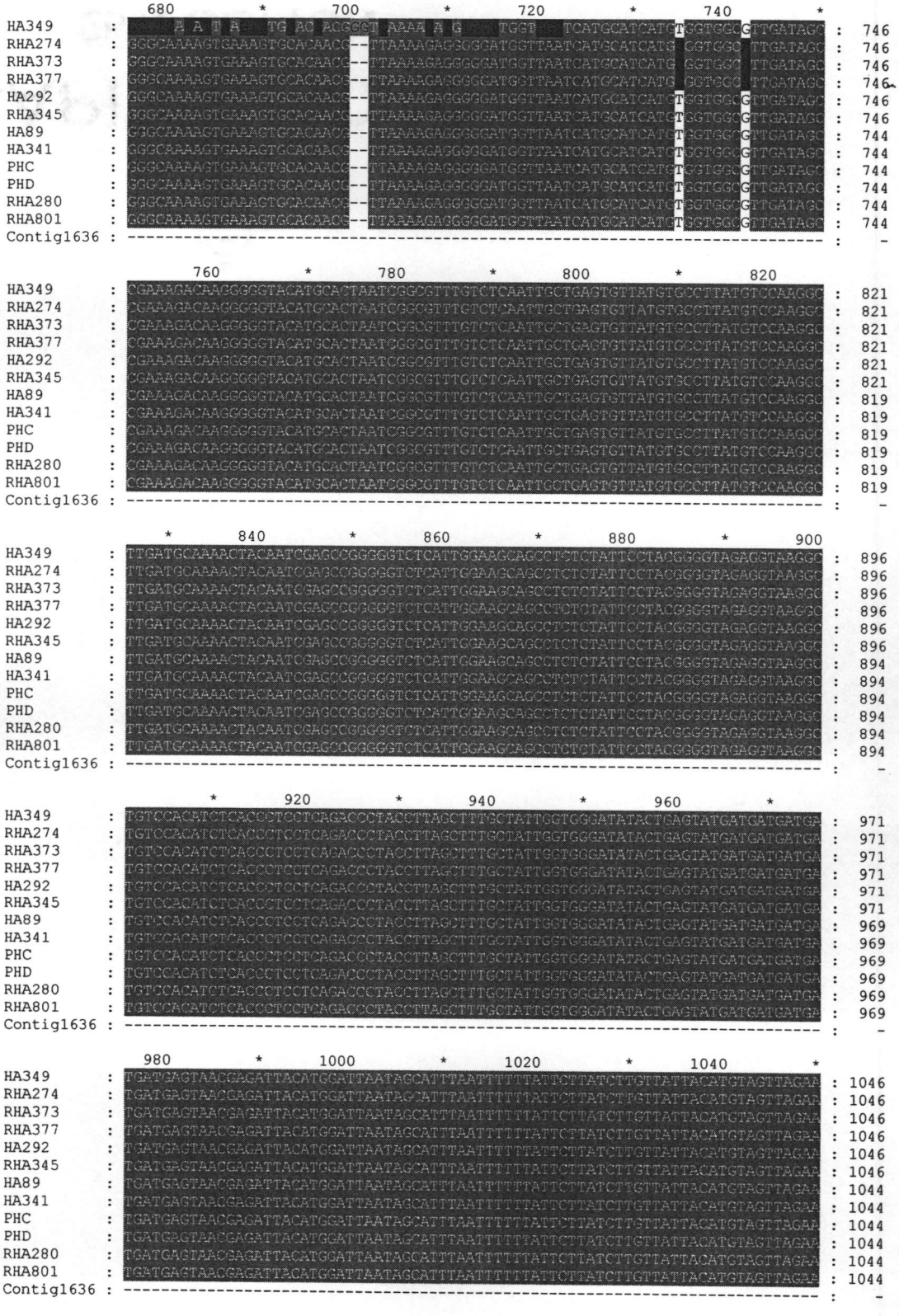
Appendix G. Alignment of the partial genomic nucleic acid sequences of *FAD6* for 12 sunflower lines with the corresponding *FAD6* contig 1636.



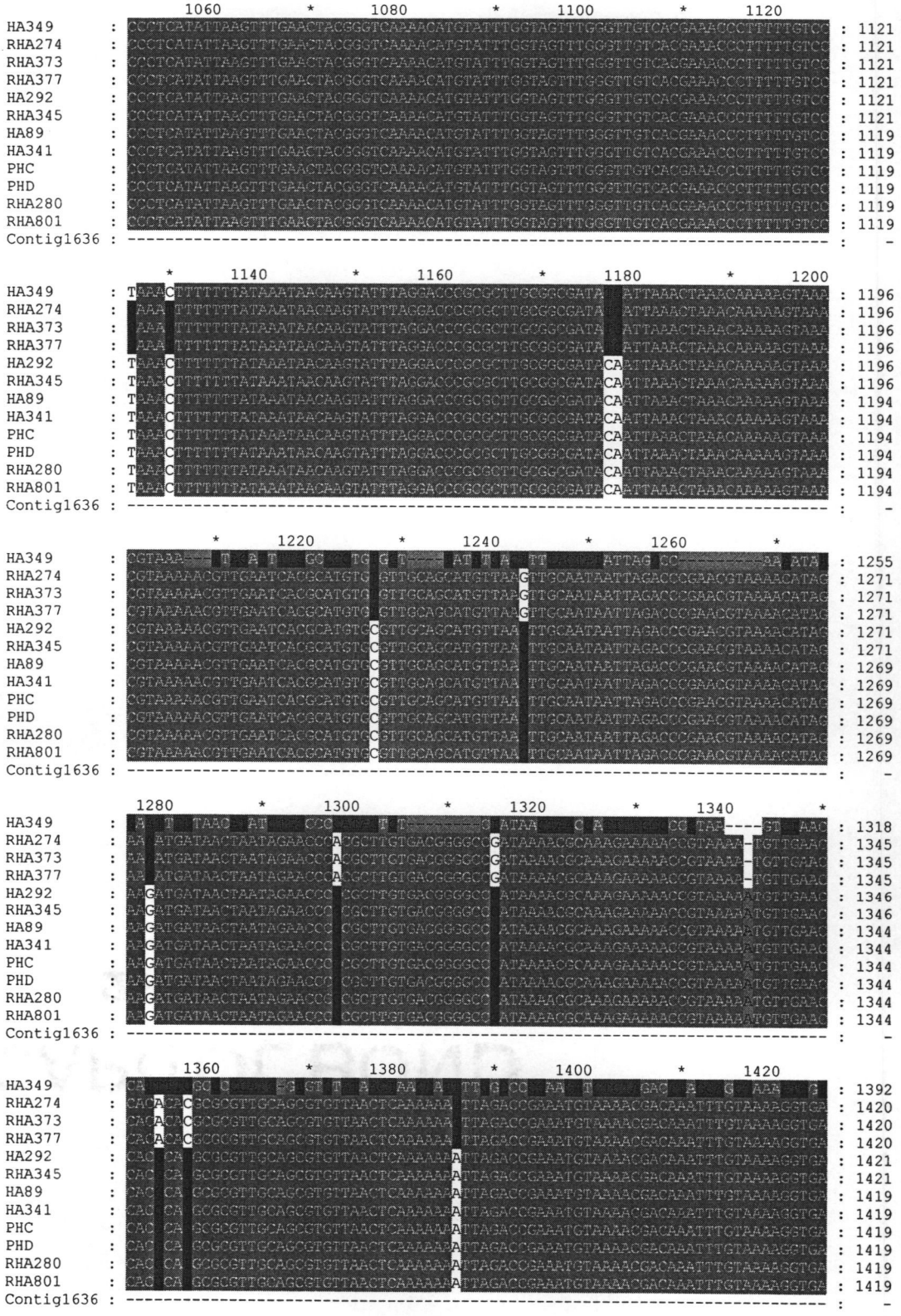
Appendix G. (Continued)



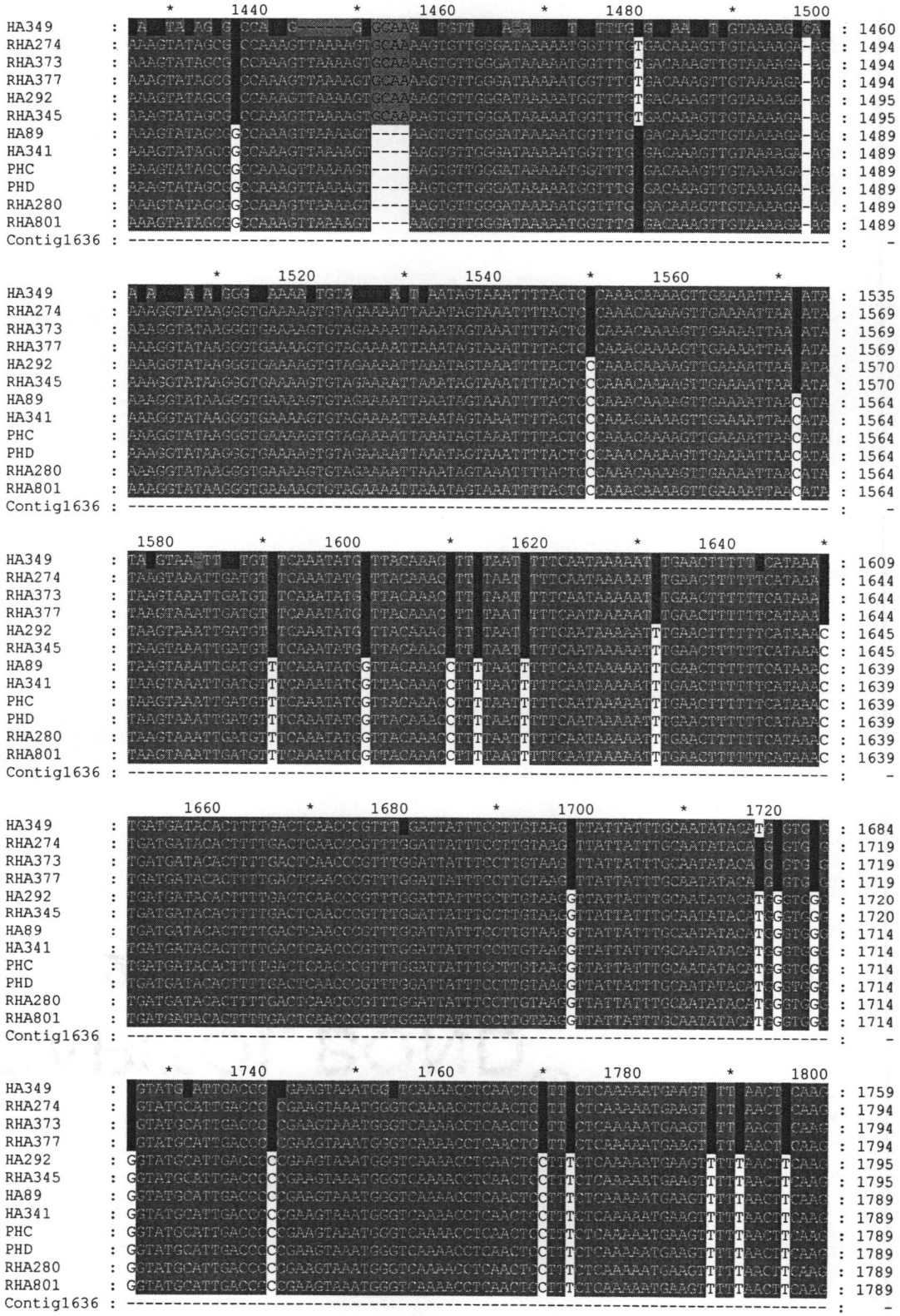
Appendix G. (Continued)



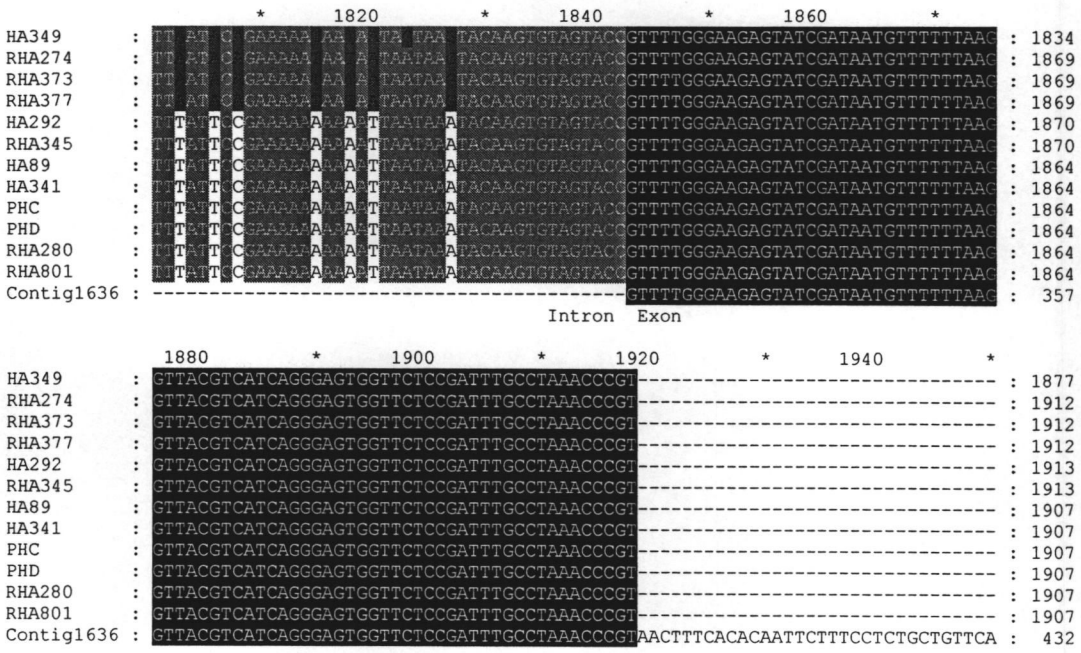
Appendix G. (Continued)



Appendix G. (Continued)



Appendix G. (Continued)



Appendix H. Oligonucleotide primer names and sequences.

Name	Sequence
FAB2-1 F	5'-GGGAGATATGATTACTGAAGAAGCAC-3'
FAB2-1 R	5'-GTGCAGGCATGGAGATCTTT-3'
FAB2-2 F	5'-GATCACAGAGGAAGCCTTACCTACC-3'
FAB2-2 R	5'-CAAAATGTCCGCATAGTCTTTCGCG-3'
KASI-1 F	5'-GGCACTGAAGCTGCCATTAT-3'
KASI-1 R	5'-GACGTTGCGTGTGCATTTAT-3'
KASI-2 F	5'-GCGGAGGTAAATGCTGAAAA-3'
KASI-2 R	5'-CCCAAACTTGTCCCAACTCA-3'
KASII F	5'-GAGTCCTCATTGGCTCTGCT-3'
KASII R	5'-CGACAAATCCTCCCAATCCT-3'
KASIII-1 F	5'-CTGTGCCTAAACTTGAGGTTTCT-3'
KASIII-1 R	5'-CCCTCTATCGGTCCAATCAA-3'
KASIII-2 F	5'-TGTGATTGGCAGCTGATTGT-3'
KASIII-2 R	5'-AGTGCTTGCTGCCAAAACC-3'
FATB A F	5'-GTCTCTCTCAACCTCTCTCTG-3'
FATB A R	5'-TTTTGAAGCCATGCGTGATA-3'
FATB B F	5'-GCTAACGCACAGGCTCCTAC-3'
FATB B R	5'-AATCCATCGCCCAGAAGAC-3'
FATB C F	5'-CAAGAAACGGCCCTTAATCA-3'
FATB C R	5'-GTCAGCAGTGCTTTCGTCAA-3'
FATB D F	5'-TGACGAAAGCACTGCTGACT-3'
FATB D R	5'-TGATCAACACTCCCATCTTGTT-3'
FATA-1 F	5'-ATCGATTCCGTTTCAATTCG-3'
FATA-1 R	5'-ACAATCCGTCTTCCGTCAAG-3'
FATA-2 F	5'-CGGAAGACGGATTGTCGTAT-3'
FATA-2 R	5'-TCTCCCTTCACCTTGACACC-3'
FAD6-1 F	5'-GTCCCATTAAGTTGGGCTTG-3'
FAD6-1 R	5'-GCATCATGCAAAAACCAACA-3'
FAD6-2 F	5'-GGGTGCCAAGCTGTATCTTC-3'
FAD6-2 R	5'-TTACGGGTTTAGGCAAATCG-3'
FAD2-1 cds F	5'-GAAAAGTCTGGTCAAACAGTCAACAT-3'
FAD2-1 cds R	5'-CGAGAACCAGGACAACAGCCATTGTC-3'
FAD2-2 cds F	5'-CAGGTTAGTGAACCATGGGTG-3'
FAD2-2 cds R	5'-CACAATAGTTCAGGTACAACAC-3'
FAD2-3 cds F	5'-GTAGGTCACTAAACAATGGGTGC-3'
FAD2-3 cds R	5'-CTTAAACGACATCAGTGACCAAACATG-3'
FAD2-1 SSR-TF	5'-GGAGCAAGATGATGAAGGGAAAGGAG-3'
FAD2-1 SSR-TR	5'-CTCGAAGGAGGTCCTACGTTC-3'
FAD2-1 SSR1 F	5'-GTTTGTGGAGCAAGATGATGAAG-3'
FAD2-1 SSR1 R	5'-CAACACATACTGCGTTACATCCA-3'

Appendix H. (Continued)

Name	Sequence
FAD2-1 SSR2 F	5'-TTAAGTCTGTGACAATGGGTCTTG-3'
FAD2-1 SSR2 R	5'-CCATTACCCGATTTGAGTTCAC-3'
FAD2-2 T F	5'-CACGCGATGGAAGCCACAAAG-3'
FAD2-2 T R	5'-CACCATCCTTAACCTCCTCATC-3'
FAD2-2 SNP	5'-CTATCAGTTTGACGGGACTTCG-3'
HT 51 F	5'-CGTTCATGCCCTTTTTGT-3'
HT 51 R	5'-AGGCTAAAGCCAGCAGGA-3'
HT51 SNP	5'-GCGTGACACAATTGATCTA-3'
F4	5'-GTAACGTCTGCGCGCTTGCAGACATCA-3'
R1	5'-GGTTTTGCATGAGGGACTCGATCGAGTG-3'
F3	5'-GGAGCAAGATGATGAAGGGAAAGGAG-3'
R6	5'-CTGATGTCTGCAAGCGCGCAGACGTTA-3'

Appendix I. Summary of the candidate gene sequencing results

Gene	5' Region	Open reading frame (ATG-TAA)		3' Region	Total	Number of lines	Number of Haplotypes
	bp	bp exonic	bp intronic	bp	bp		
FAD2-1	1807	1136	-	854	3797	4	1
FAD2-2	1067	1153	-	436	2656	2	1
FAD2-2	-	1153	-	-	1291	4	2
FDA2-3	-	1148	-	-	1195	12	1
KASI	-	854	777	-	1631	2	1
FATB	-	1418	1608	-	3026	12	1
FATA	-	280	941	-	1221	12	2
FAD6	-	402	1546	-	1948	12	4