Winter 2010

Genomic resource development for a diploid mint: Mentha longifolia

Zahra Hadadian
University of New Hampshire, Durham

GENOMIC RESOURCE DEVELOPMENT FOR A DIPLOID MINT: *MENTHA LONGIFOLIA*

BY

Zahra Hadadian

Bachelor of Science, Shiraz Azad University, Iran, 1996

Master of Science, Tehran University, Iran, 2001

THESIS

Submitted to the University of New Hampshire

In Partial Fulfillment of

the requirements for the degree of

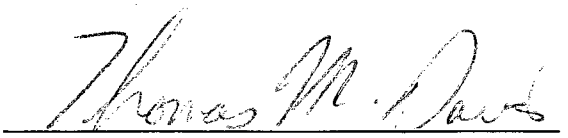
Master of Science

In

Plant Biology

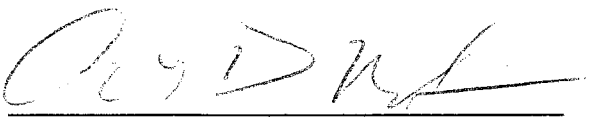
December, 2010

This dissertation has been examined and approved



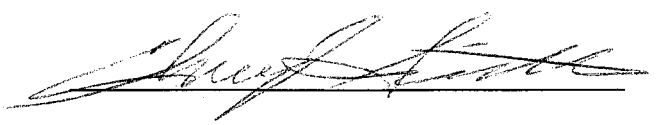
Thesis Director, Thomas M. Davis

Professor of Plant Biology and Genetics



Chris Neefus

Professor of Plant Biology, Chair, Department of Biological Sciences



Cheryl Smith

Extension Professor of Plant Biology

12.09.10

Date

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

GENOMIC RESOURCE DEVELOPMENT FOR A DIPLOID MINT: *MENTHA LONGIFOLIA*

by

Zahra Hadadian

University of New Hampshire, December 2010

This research project aimed to develop genomic resources needed to enable construction of a genetic linkage map of the diploid mint species *Mentha longifolia*. Such a map would facilitate identification of plant genes involved in resistance to *Verticillium* fungal infection. For this purpose, a small genomic library was constructed from germplasm accession CMEN 585, 279 genomic inserts were sequenced and annotated and 19 PCR primer pairs were designed and tested on two resistant and two susceptible accessions. The Cleaved Modified Polymorphic Sequence (CAPS) method of molecular marker genotyping was found to detect little variation between crossing parents CMEN 585 (resistant) and CMEN 584 (susceptible). Comparative sequencing of PCR products from two European and two South African accessions revealed greater diversity between than within geographic locations. Future efforts should focus on assessing more sensitive genotyping methods, and developing a mapping population from a cross between European and South African accessions.

INTRODUCTION

Background and Literature Review

The genus Mentha: taxonomy, botanical and cytological features, distribution, uses.

The mints (*Mentha* sp.) belong to the family Lamiaceae. Previously this family was called Labiateae in reference to the bilabiate (2-lipped) flowers exhibited by most of its species. This name was changed to Lamiaceae based on the legitimately established genus *Lamium* by the International Code of Botanical Nomenclature. Lamiaceae is a large family with about 250 genera and 6700 species. Most of them are herbaceous and the rest are shrubs or rarely trees. This flowering family is easily recognized by its opposite leaves, characteristically square stems, and emission of pungent and often very pleasant smell after crushing (1).

Mints can be annual or perennial herbs or shrubs. They have opposite leaves, with the square stems which is typical of the Lamiaceae family, and abundant and quite attractive flowers. Sepals and corollas are variously fused. Corollas are strongly 2-lipped, rarely 1-lipped with 2 or 4 stamens, superior ovaries which have 4 lobes and the style arises from the middle of these 4 lobes. Fruits are comprised of 4 nutlets, each containing one single seed (1). Mint plants are seldom more than 100cm in height. Leaves may be petiolate or sessile (lacking a petiole), and can be up to 7.5 cm in length. Flowers are produced in clusters in leaf axils or on terminal spikes in white, blue, or purple colors (3).

Plants of the Lamiaceae family mainly grow in sub-temperate areas having

a 6 to 27ºC temperature range, annual rainfall of more than 300 mm, and a soil pH of 4.5 to 8.3 (3). Their distribution in the U.S. covers almost all parts of the country except the southeast (7).

The practical uses of the mint plant go back to the ancient Egyptian, Greek and Roman cooks, who used it as a flavoring for wines and sauces. Today, this plant is used in the food industry as an ingredient in teas, alcoholic drinks, chewing gum, desserts, and candies. Mint is also famous as a treatment herb for stomach ache or chest pain. Mostly mints are grown commercially for their essential oils. Menthol and other mint essential oils are used in cosmetics and the aromatic (perfume) industry. Health companies use essential oil and menthol extensively as flavorings in breath fresheners, toothpastes (1).

Cultivation of commercial mint plants is done by vegetative propagation with both stolons and rhizomes. They need proper irrigation and nitrogen fertilizer for a good amount and quality of oil production. Mint oil increases with the age of the plant. Menthol is a major component of mint oil, contributing a pepperminty smell. Most mint oils are however chemically complex. The major oil component in different mints is variable. For example, the major essential oil component in peppermint is menthol, while in spearmint it is carvone (3). *Mentha longifolia* has piperitone oxide as a main essential oil compound; however, at least one accession (CMEN 584) with carvone as the principle oil component has been reported (5).

One of the most important characteristics of commercial mint and some other crop plants is polyploidy. Polyploid somatic cells contain more than two sets

of chromosomes. The basic chromosome number in *Mentha* is 12, so a diploid mint has two sets of 12 chromosomes, and has a chromosome number of 2n = 2x = 24, while a hexaploid mint would have 2n = 6x = 72 chromosomes. Polyploid plants are known for their larger leaf area compared to diploid plants. They have dark green color because of their greater amount of chlorophyll and thus, their potential for photosynthesis and accumulatation of essential oil in their cells is higher than diploid plants (4).

The most common and popular mints for cultivation and essential oil are spearmint (*Mentha spicata*) and peppermint (*Mentha piperita*), both of which are polyploid. Even though polyploid plants have advantages for oil production and commercial uses, they present problems for work on transmission genetic analyses and conventional breeding (5). However, a close relative of the commercial mints, *Mentha longofolia*, provides an attractive option for genetic research in *Mentha*, due in part to its diploid (2N = 2X = 24) chromosome constitution (2). Commercial mints are vegetatively-propagated clones and very little genetic diversity exists among them. However, considerable diversity exists in the available germplasm of *M. longifolia*, adding to its attractiveness as a subject for the research reported here.

Mentha germplasm is collected and maintained by the National Clonal Germplasm Repository (NCGR) in Corvallis, Oregon. Within the overall Mentha collection at NCGR, approximately 10% of these *Mentha* accessions have been defined as the Mentha "Core Collection", representing the maximum genetic diversity of the genus worldwide. This research utilizes accessions of *Mentha*

longifolia and other *Mentha* germplasm obtained from the NCGR Core Collection. Phylogenetic relationships in *Mentha* infered from DNA sequences of the chloroplast (cp) *rpl16* intron and *trnL-trnF* region suggest that *M. longifolia* may be the maternal parents of *M. canadensis* and *M. spicata* (6).

Verticillium wilt

Verticillium wilt is a common disease throughout the world, and causes damage to over 300 crop species throughout the United States and worldwide (8), including a wide variety of vegetables and herbaceous plants. Breeding programs have reduced the prevalence of this disease on many plants. For instance, in the Momor genotype of tomato (*Lycopersicon*), the *Ve* resistance gene was mapped in 1998 and was used as a resistant parent in the backcross breeding against the pathogen *Verticillium dahlia* (13 and 14).

Verticillium wilt is caused by two closely related soil born fungi, *Verticillium dahlia* and *V. albo-atrum.* Plants weakened by root damage from waterlogged soil, drought or other environmental stresses become very susceptible to this disease. Virulence and pathogenicity of these two Verticillium species varies in different hosts and different environments. Symptom development also depends on environmental conditions, which also vary. Sometimes symptoms are not obvious until flowering or fruiting or after the stressful condition such as hot or dry weather. In the infected plants the first symptoms usually develop on the older leaves. Symptoms include yellowing, wilting, drying and dropping of leaves, necrotic, V-shapes lesions at the tips of leaves. Additional symptoms on

mint leaves can be bronzing and twisting of the top leaves, which gives them a half-moon appearance (19).

A diagnostic characteristic of this disease is discoloration or streaking in the vascular system. Vascular tissue starts to become brown while the pith remains white. Discoloration usually occurs in all sections of the plant except leaf petioles (9).

Verticillium is a soilborn fungus and can persist for a long time in the soil by producing resting structures called microsclerotia. Microsclerotia can germinate and grow in a humid and moderate tempreture condition. The fungus can enter the root xylem and grow through that. It produces ovoid conidia, from specialized structures called phialides borne in a whorl around a conidiophore. As long as this fungus continues developing through the xylem, it restricts water and nutrients movements as well as producing enzymes and toxins. The plant reacts to this pathogen attack, as a means of blocking the penetrating fungus. The result of this defense is plugging and gumming of water-conducting vessels, which further restricts water in the host and wilt symptoms will be increased (9).

Mint fields infected by *Verticillium* suffer progressive decreases in annual oil yield and become economically unviable within 3 to 5 years. As long as this fungus produces resting spores and these spores can guarantee survival of this fungus in the field, managing Verticillium wilt can be difficult. The best and first strategy is to avoid planting susceptible species or varieties in areas known to be infected. Division of resistance and susceptible plants to *Verticillium*, mint has placed into the susceptible plants (9).

In the late 1940s, growers moved most of the mint production from Midwestern to western states in an effort to escape Verticillium wilt but, unfortunately they brought infected roots to the new areas and contaminated the soil.

Commercial peppermint varieties such as 'Black Mitcham' are susceptible. Resistance gene cannot be bred into commercial peppermint because it is a sexually sterile plant, probably due to its polyploidy. This research was performed to find a solution to this problem by identifying genes for resistance to Verticillium using the genetically favorable diploid species, *M. longifolia*. Once such resistance factors have been identified, perhaps they can be transferred into peppermint via a genetic engineering approach.

With this goal, *Mentha longifolia* accessions from NCGR collection were screened and classified as highly resistance, intermediate, or susceptible (5). CMEN 585, CMEN 17, CMEN 501 and CMEN 81 were resistant, while CMEN 584 and CMEN 516 were highly susceptible to *Verticillium* wilt. Crosses between the highly resistant CMEN 585 and susceptible CMEN 584 accessions were performed to develop progeny populations segregating for resistance vs. susceptibility (5).

Molecular markers

A genetic or molecular marker is a gene or specific fragment of DNA that exists in variant forms (alleles) and has a known location on the chromosome. Molecular markers can be identified and located throughout the whole genome, and used as pinpoints for the location of desirable genetic traits or indication of

specific genetic variation which may arise due to mutation or alteration in the genomic loci. A genetic marker may be a short DNA sequence, such as a sequence surrounding a single base-pair change (single nucleotide polymorphism (SNP)), or a longer one, like a microsatellite or simple sequence repeat (SSR).

To successfully use a specific marker to "mark" a specific trait in the progeny of a genetic cross, the molecular marker must be linked with the gene/trait of interest so that variations (alleles) of both the marker and the gene can be inherited together. Thus, individuals can be selected in which the molecular marker is present, since the marker indicates the presence of the desired trait (10).

Variations (alleles) of the same gene in a population of plants that can be detected through molecular markers are called polymorphisms. Differences between the DNA sequences of the alleles of a gene can be responsible for making a plant susceptible or resistance to a particular disease. DNA fragments associated with genes involved in disease resistance can also have sequences difference which can be used as markers to locate and track the gene in progeny (10).

One of the convenient marker technologies available in genetic research today is the cleaved amplified polymorphic sequence method (CAPS). This technique is based on identifying polymorphisms at a particular locus by the use of restriction enzyme digestion of polymerase chain reaction (PCR) products. The locus is amplified by PCR, followed by digestion with restriction enzymes.

Sequence polymorphisms may result in different sized restriction fragments. To apply this method, however it is necessary to have some genomic sequence data upon which to base the design of PCR primers. Very little genomic sequence information is available for mint.

Goal and Objectives

The goal of this research project was to develop genomic resources for the model mint diploid species, *Mentha longifolia*, that would ultimately enable the construction of a genetic linkage map and the identification of molecular markers linked to genes for resistance to verticillium wilt.

One specific objective of this research was to evaluate an existing F2 population derived from the resistant (R) x susceptible (S) cross CMEN 585 x CMEN 584, to determine whether it could be used as a basis for construction of the first genetic linkage map of *Mentha longifolia*. Such a map would provide a much needed context and resource for identifying plant genetic loci associated with disease resistance.

A second objective was added during the project to assess additional germplasm as potential parents for future use in developing a more suitable mapping population.

Approach

To acquire DNA sequence data needed as a basis for PCR primer design, my approach was to construct a small genomic library from *M. longifolia* DNA. A genomic library is a set of DNA clones. Approximately, 200 of these clones were sequenced, providing the needed DNA sequence information. PCR was

performed to amplify segments of the mint genome corresponding to the sequences that were cloned. The sequences between two mint accessions, CMEN 585 and CMEN 584 were compared. Polymorphisms were identified by digesting the PCR products with restriction enzymes and looking for different fragment sizes on electrophoresis gels (11).

MATERIALS AND METHODS

Plant materials

Germplasm used for this research was obtained from the NCGR. These accessions were from the family *Lamiaceae* (alt. *Labiatae*), subfamily *Nepetoideae*, and tribe: *Mentheae* (15, 16, 17). Two accessions, *Mentha longifolia* subsp. *polyadenia* (CMEN 584) and *M. longifolia* subsp. *capensis* (CMEN 585), had been collected from South Africa, while two other, *M. longifolia* CMEN 17 and CMEN 516, had been collected from an unknown European country or countries. CMEN 585 and CMEN 17 were resistant, while CMEN 584 and CMEN 516 were highly susceptible to *Verticillium* wilt. Two of the accessions (CMEN 584 and CMEN 585) had been crossed in earlier research to develop progeny populations (F1 and F2) segregating for resistance vs. susceptibility (5).

DNA isolation

DNA was isolated from fresh young mint leaf tissues using the CTAB miniprep method (12), with the following modifications: sodium bisulphate wasn't used, chloroform/octanol was added after incubation at 65°C, and the ethanol wash was performed with 70% ethanol without sodium acetate. The DNA concentration was measured by fluorometry and diluted to 5-40 ng/µl with sterile water (12).

Cloning/ library

A small genomic library was made by the following procedure:

Double digestion: genomic DNA (33-44 ng) was digested with two selected restriction enzymes (*Bam*HI and *Bgl*II) in the recommended buffer with Bovine Serum Albumin (BSA) (100 µg/ ml). Digests were incubated at 37°C overnight. Samples were mixed with Ficoll-based loading buffer (15% Ficoll polymer in distilled water, 0.25% Bromophenol Blue, 0.25% Xylene cyanol FF) and loaded in a 1.2% agarose gel in 1x TBE buffer (40 mM Tris- Borate, 1 mM EDTA, pH 8.0-8.5), and run at 7 v/cm. The gel was stained in ethidium bromide and exposed to UV light for DNA visualization and photo documentation.

Dephosphorylation: Dephosphorylation was done by adding Antarctic phosphatase (Catalog # M0289, New England BioLabs (NEB), Ipswich, MA) in its buffer to the digested DNA. The digested DNA was incubated in the thermocycler (Eppendorf Mastercycler ep gradient (Hamburg, Germany)) with the following program: 37°C for 15 min and then, 65°C for 5 min.

End repair: End repair (total volume, 24 µl) contained dephosphorylated DNA, 1.0 unit Econo-Taq™ DNA polymerase, 1X Econo-Taq PCR Buffer, 200 µM of each dNTP and 25 µM of $MgCl_2$. The final reaction was incubated in the thermocycler at 72°C for 30 min.

TOPO cloning of chemically competent E. coli: TOPO cloning reaction was made based on the chemically competent protocol for PCR /GW/TOPO by adding 1 µl of salt solution, 0.7 µl of TOPO vector and 2.25 µl of water to 2µl of fresh DNA from the end repair step (total volume, 6 µl). The reaction was incubated at room temperature for 1h.

Transformation of chemically competent E. coli: Transformation of chemically competent *E. coli* TOPO was achieved gently by mixing 2 µl TOPO cloning reaction with the bacteria cells, incubating on ice for 10 min, and placing in a water bath at 42°C for 30 sec as a heat shock. Then adding 250 µl SOC medium (0.5% Yeast Extract, 2% Tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4, 20 mM Glucose). The cells were incubated at 37°C with vigorous shaking for 60 min, spread on plates containing LB agar (0.1% NaCl, 0.1% tryptone, 0.5% yeast, 1.5% Agar, pH 7.0) plus Spectinomycin (100 µg/ ml), and incubated at 37°C overnight.

Colony PCR: A typical PCR product contained 0.2 µM of each primer, 0.2 mM each NTP, colony as a template DNA, I x Econo-Taq PCR buffer, 25 µM of $MgCl_2$, and 1.0 unit Econo-Taq™ DNA polymerase. For colony PCR, the reaction was incubated in the thermocycler with the following program: initial 5 min denaturation at 95°C, followed by 30 cycles of 1 min

at 95°C, 1 min 30 sec at 54 °C annealing temperatures for each designed primer, and 1 min at 72°C. The final extension step was 5 min at 72°C.

DNA sequencing: colonies of transformed bacteria were sent to the Purdue University Genome Center for sequencing.

Bioinformatics

Alignment of the obtained sequences (from Purdue University) with GenBank® nr database was performed using blastx and tblastx algorithms (Altschul et sl., 1990).

Primer Design

Some sequences from the genomic library which were matched with specific known genes were chosen as the basis for marker development. Forward and reverse primers were designed using the Primer Select program (DNA Star). PCR was performed using these primers with various plant DNA templates.

Polymerase Chain Reaction

Polymerase chain reaction (total volume, 25 µl) contained 40 ng of genomic DNA, 1.0 unit Econo-Taq™ DNA polymerase, 1X Econo-Taq PCR Buffer, 200 µM of each dNTP, 25 µM of $MgCl_2$, and 0.2 µM of each primer. The final reaction was incubated in the thermocycler with the following program: initial

2 min denaturation at 94°C, followed by 20 cycles of 30 sec at 94°C, 45 sec at different annealing temperatures for each designed primer, and 3 min at 72°C. The final extension step was 10 min at 72°C.

Markers

Marker development was accomplished by the following procedure: PCR products were cleaned with Promega SV Gel and PCR Clean Up Kit (Catalog # A9282, Promega). Cleaned PCR products were sent to New Hampshire's Hubbard Genome Center for sequencing by using DYEnamic ^{TM} Terminator ET cycle sequencing premix (Amersham Biosciences, Piscataway, NJ).

Sequences from both CMEN 584 and CMEN 585 were examined to find restriction site polymorphisms using the SeqBuilder (DNA Star) program , andto find polymorphisms such as SNPs using the SeqMan (DNA Star) program.

Restriction Digestion and Agarose Gel Electrophoresis

PCR products were quantified using fluorometry. 300 ng aliquot were digested with selected restriction enzymes, recommended buffer and BSA (100 µg/ ml). Digests were incubated at 37°C overnight. Samples were electrophoresed in 2% agarose TBE gels at 4 v/cm.

RESULTS

For the purpose of constructing a small genomic library, genomic DNA from CMEN 585 was successfully digested with *Bam*HI and *Bgl*II restriction enzymes, by comparing aliquots of undigested and digested DNA samples in an electrophoretic gel (Figure1). The undigested DNA (Lane 2) is visible as a band higher on the gel than the 12 kb upper band of the DNA ladder (Lane 1). In contrast, the digested DNA is visible only as a light smear in the 100 bp to 6,000 bp range (Lane 3).

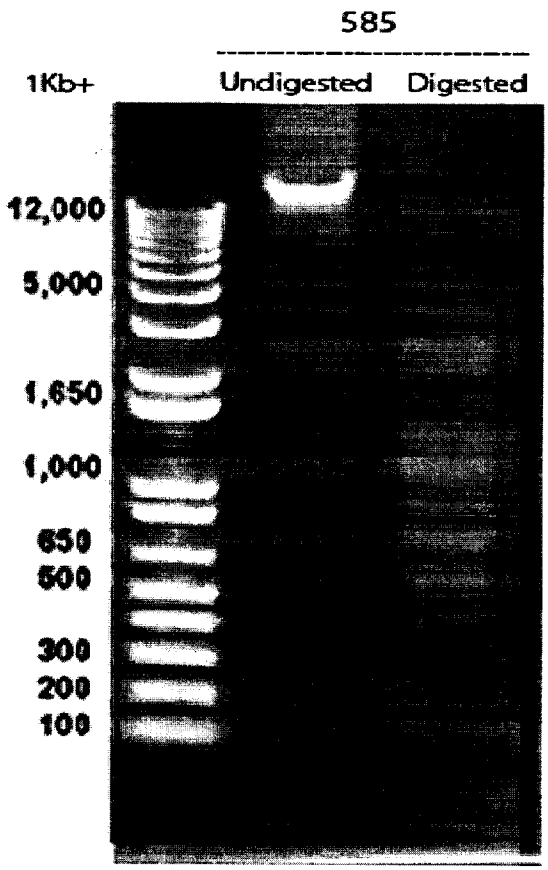


Figure 1: Gel electrophoresis of CMEN 585: **genomic DNA in 1.2% agarose TBE gel. Lane 1: 1kb+ ladder. Lane 2: undigested. Lane 3: digested.**

TOPO cloning

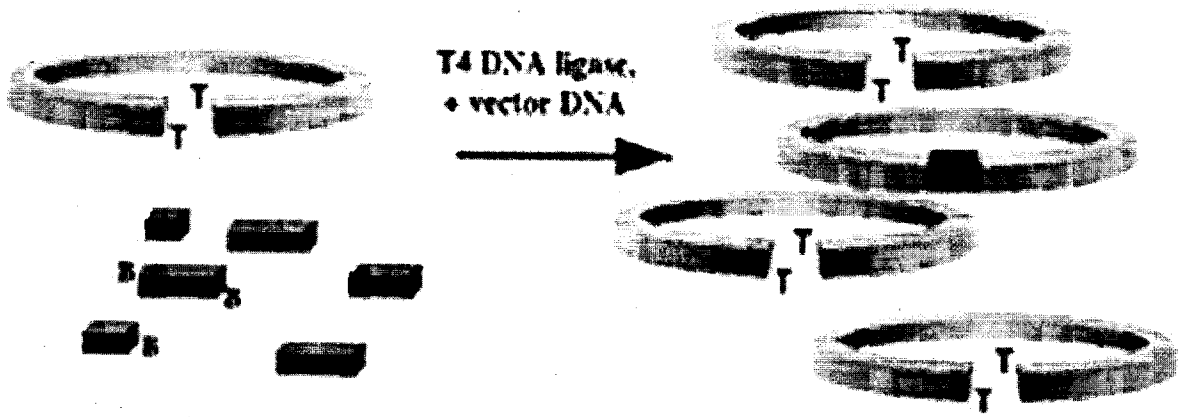


Figure 2: Transformation of digested CMEN 585 genomic DNA to the TOPO cloning of Chemically Competent *E. coli.*

When an aliquot of the digested genomic DNA was utilized for cloning into the TOPO-TA vector (Figure 2) and for transformation into chemically competent *E. coli* cells, plating of the transformed cells on selective medium resulted in the growth of bacterial colonies. Approximately 90 colonies were obtained from each of 6 plates.

Colony insert sizes were checked by colony PCR, followed by gel electrophoresis, as shown in Figure3. This gel shows a total of 96 colony PCR products. Colony insert sizes were between 200 to 3000 bp with an average size of 1420 bp.

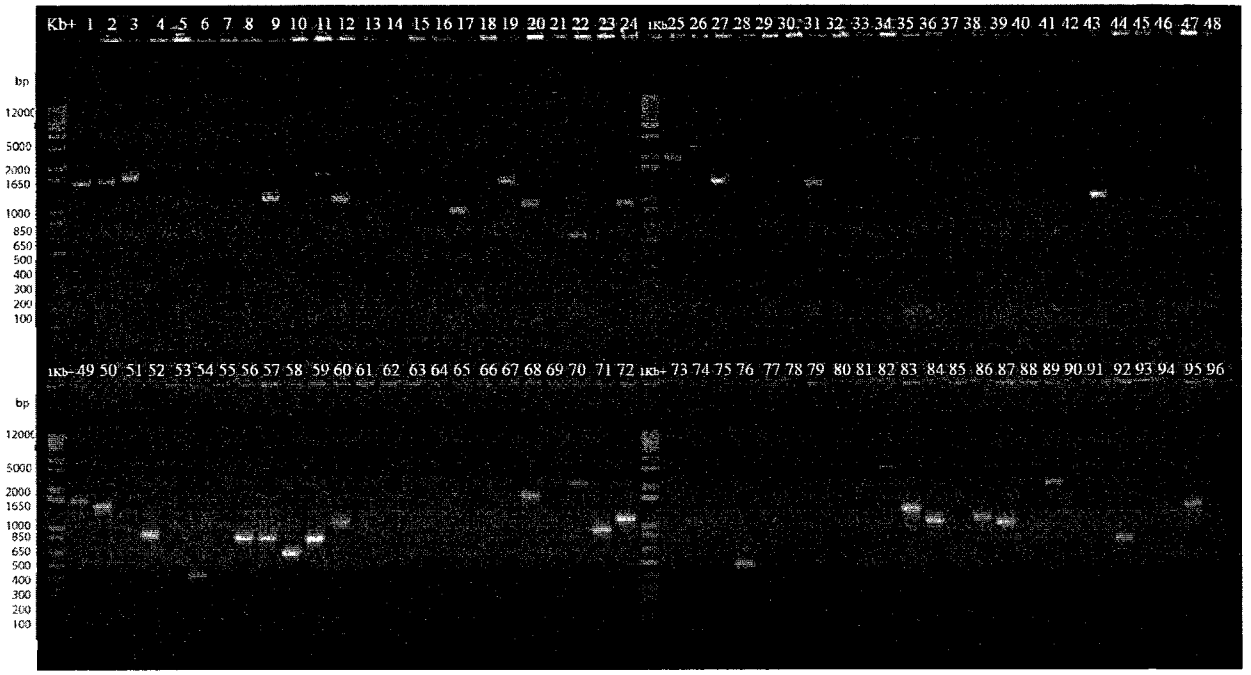


Figure 3: Gel electrophoresis of colony PCR products in a 2% agarose gel: 1kb+ ladder is represented four times on the gel, sample Lanes are numbered through 1 to 96 representing colony PCR products of 96 colonies from a single culture plate.

A total of 289 good quality sequences were obtained from the Purdue University Genome Center, representing 425 clones. Of these, both forward and reverse reads were obtained for 142 clones, while only forward reads were obtained for 18 clones, and only reverse reads were obtained for 41 clones. When both forward and reverse reads were obtained, the reads were long enough for the forward and reverse reads to overlap, allowing 88 contigs to be assembled. When the total 95,287 bp length of these contigs plus the singleton forward and reverse reads are added in, a combined total of 264,762 bp of genomic sequence was obtained. These sequences have been deposited in the GenBank GSS database under accession numbers in Appendix B.

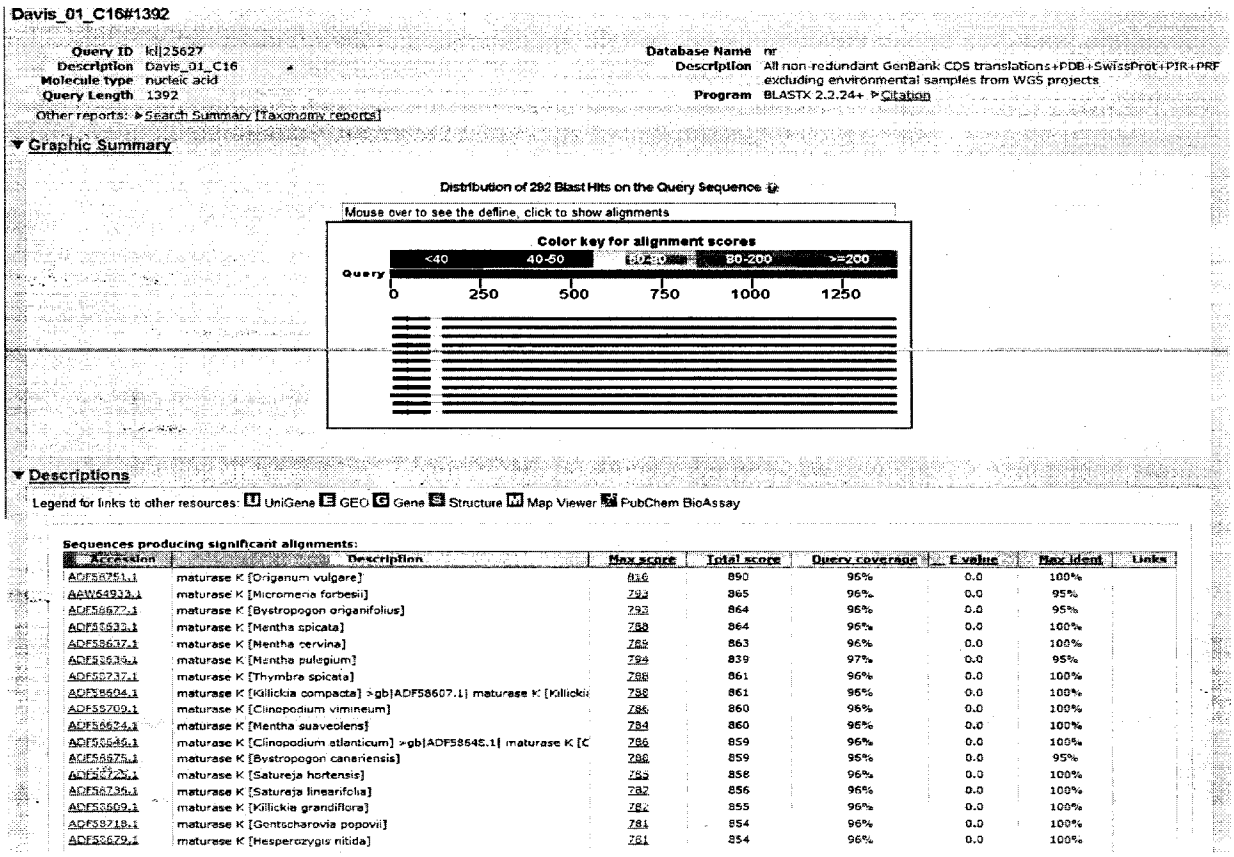


Figure 4: Example of a blast (blastx) output for one library clone sequence: The output indicates that this sequence corresponds to the chloroplast genomic sequence for the Maturase K gene.

The results of Blastx searches of the GenBank nonreduncant (nr) protein database (delimited to Viridiplantae) showed that 17.6% of sequences were derived from the choloroplast genome, 1.72% from the mitochondrial genome, 3.45% from both choloroplast and mitochondrial genomes, and 77.24% from the nuclear genome. In terms of putative gene identities, 51% corresponded to unknown or hypothetical proteins, 37% to transposons and 12% to known proteins (Appendix A).

The best quality sequences from the genomic library which were matched with the specific genes in the nucleus were chosen for use in marker

development, and for diversity analysis. Using these sequences as the basis for primer design, 19 primer pairs were designed by using primer select program (DNA Star) (Table 1).

Primer Name	Forward Primer Sequences	Reverse Primer Sequences	Annealing Temperature
I20	GAAAACGATGATAGTCCCAACCA	CCGCACAGATATAGCAAGCAGA	51.4℃
D07	TTTGAAAGAAATGAGGTATGGT	AAACCACCATAATCTGAATC	50.4℃
M18	TTCTTGAGGTTGGAAATGTT	CTCGAGGAATAGTGAAGAAATGTA	51.9℃
I04	GCTGCCATACGGGTGATTTAGTT	CCGGAGATGGCAGAGGAGGT	55.0℃
F03	CCTTCTACTGCGTGTGGAG	TGCTGCTATTTCACTATTATCTTA	50.9℃
L16	GCCATAATAACGCAAGCAT	GGAATTAACCAAGCCACTGT	52.2℃
P05	CAGAAGCTATGTGACCGCA	AAAGCTCCATTATCCTATCGTA	52.3℃
O22	TCGGAGGTTACGGCAATA	CACGAATCCCGCTATCAC	51.4℃
D03	GTGTCGGTAGATGAAGTAGT	CTCGAATAGAAATAGCACA	49.1℃
A06	TGTAGAGAAAGATAGAGAAAAATG	CAGTTGCCGATGCAGA	50.7℃
E06	ATGCCTGCCCTATGTGAGC	GCCGATGCAGAAGAAGAAGAT	52.9℃
C08	TCGCGACGTAGAGAAAGA	GTCGACGTGCTCCCTAAA	51.4℃
B09	ACGATGTCTCCAGAAATGATGTTA	TTGATTGATGAAGATTCCACGACT	55.2℃
N08	TCTTGAGATCGTCGGCACAG	GGACCTTGGCAAACTTCGTG	57.0℃
M06	CCAGGGGTAGGGAGGTAAT	TGAGGTCAATAGCAGGAAAGTT	53.4℃
N07	CAGGGGATATGGAGGTAAT	TGACAAAGCACAACAGACAT	51.9℃
C04	CAGGTGGTGCTTGGTATGATGGA	AGAGCCTGAGGGTGAACGAGAAG	54.7℃
J13	GCAAGCATAATACTCAAACAT	TCCGACGCAATCTTCTT	52.3℃
N15	TGGCAAGCATAATACTCAA	TTCCGACGCAATCTTCT	52.2℃

Table 1: Designed primers with DNA Star/primer select program for use in marker development and/or diversity analysis.

When PCR was performed with these primers, using genomic DNA from CMEN 584 and CMEN 585 as templates, products were obtained using 11 of the

primer pairs (D07, M18, I04, A06, J13, E06, C04, L16, M06, N08, and N07). The results for nine primer pairs were shown in Figure 5. No obvious PCR product length polymorphisms were observed when comparing CMEN 584 with CMEN 585.

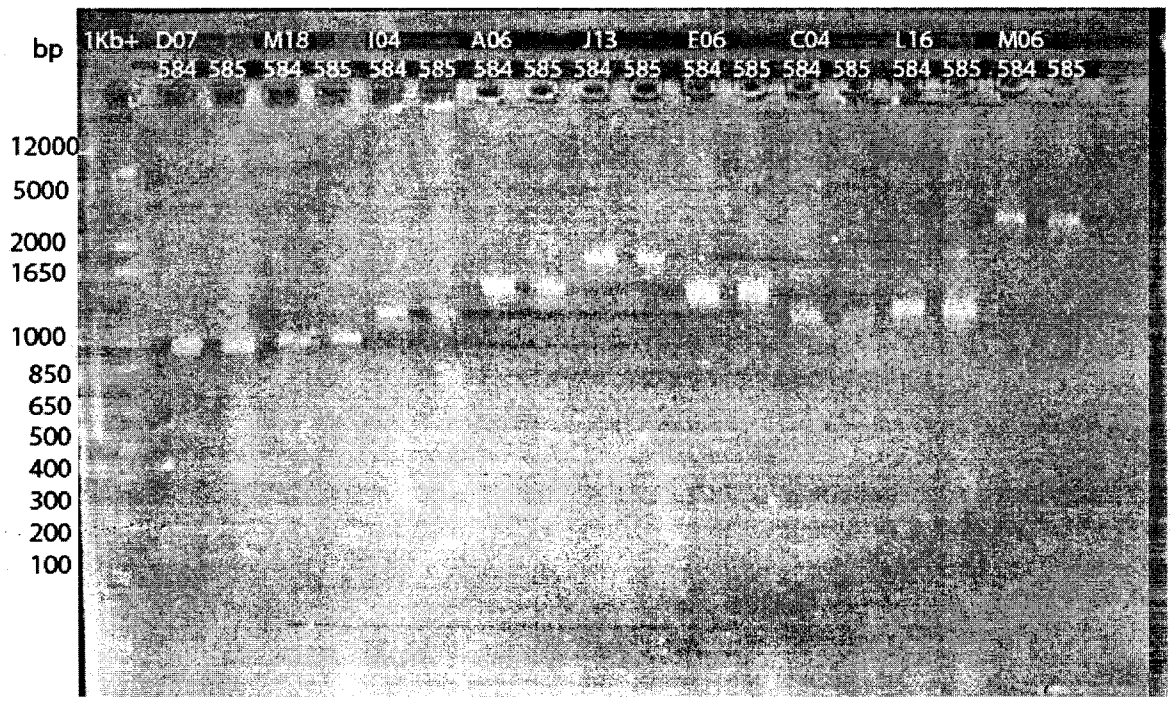


Figure 5: Gel electrophoresis of CMEN 584 and CMEN 585 PCR products: first lane: 1Kb+, the other lanes left to right: PCR products with the primer pairs of D07, M18, I04, A06, J13, E06, C04, L16, and M06. Electrophoresis was in a 2% agarose TBE gel.

Because PCR product polymorphisms were not observed in any CMEN 585 and CMEN 584, the PCR products were sent for sequencing to the UNH genomic center. Sequencing of parent PCR products revealed polymorphisms, and some of these polymorphisms could be used as a basis for marker development. Aligning sequences of these two accessions showed polymorphisms as SNPs or indels in both accessions. Sequencing revealed restriction sites which were present in only one parent (Figures 6 and 5).

Figure 6 is the alignment of parents -CMEN 585 and CMEN 584- PCR products which amplified with E06 primer. CMEN 585 (934 bp) and CMEN 584 (925 bp) PCR product sequences were compared with the CMEN 585 contig (1449bp) from the Purdue University. CMEN 585 and CMEN 584 PCR products of this primer revealed 4 bp indel in CMEN 584, and5 bp indel in CMEN 585 (Figure 6).

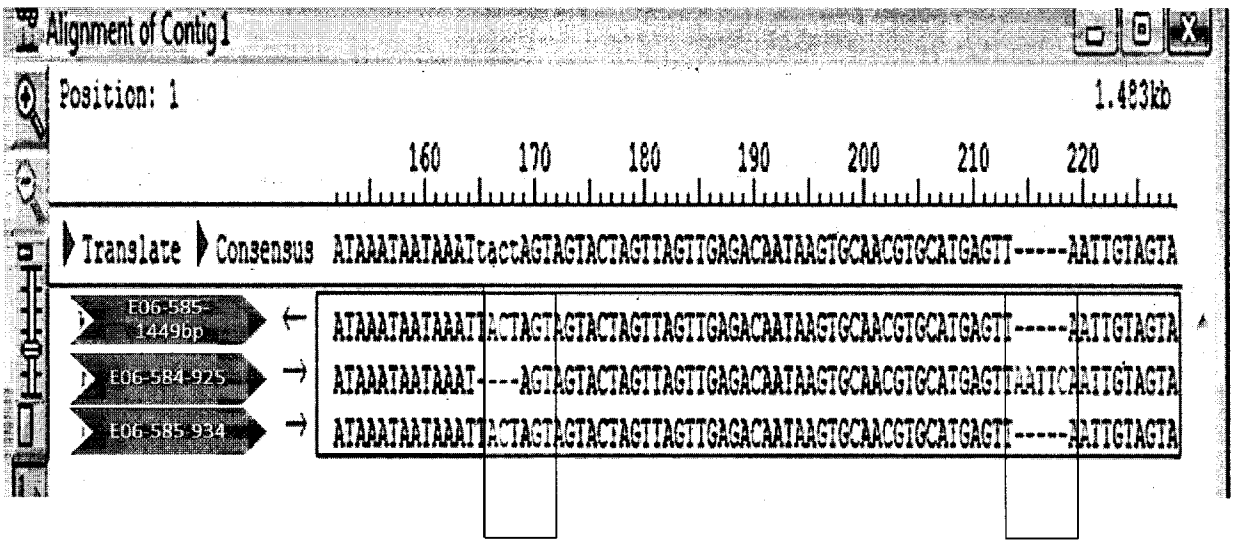


Figure 6: Sequence alignment of CMEN 584 and CMEN 585 PCR products with the E06 designed primer from the genomic library. Box on the left contains an indel in CMEN 584 and restriction site in the other parent CMEN 585. Box on the right contains an indel in CMEN 58

Box on left side of figure 6 contains a restriction site (ACTAGT) that is only present in the CMEN 585 parent. This restriction site was useful for finding the best enzyme (here: *Spe*I) for digestion.

PCR products of CMEN 584 and CMEN 585 were digested with different 4 bp restriction sites enzymes such as *Taq*αI, *Rsa*I, *Hae*III, *Bgl*I, and *Alu*I for M18, A06, E06, C04, L16, M06, N08, and N07 primers with the aim of finding polymorphisms between the two genotypes. As can be seen in the gel, CMEN 584 and CMEN 585 digested PCR products didn't show any polymorphism in

their fragment sizes. With the *Rsa*I and *Hae*III restriction enzymes, only the digested PCR products from primer pairs J13 displayed polymorphism between CMEN 584 and CMEN 585. *Rsa*I digested fragment sizes of CMEN 584 were 400, 500, 550, and 750 bp. *Rsa*I digested fragment sizes of CMEN 585 were 450, 550, and 750 bp, respectively (Lanes 3 and 4). The *Hae*III digested fragment sizes of CMEN 584 and CMEN 585 were 1500 and 1650 bp, respectively (Lanes 19 and 20) (Figure7 and Table 2).

1. N08- 584-*TaqαI*
2. N08- 585- *TaqαI*
3. J13- 584- *RsaI*
4. J13- 585- *RsaI*
5. M06- 584- *RsaI*
6. M06- 585- *RsaI*
7. E06- 584- *RsaI*
8. E06- 585- *RsaI*
9. A06- 584- *RsaI*
10. A06- 585- *RsaI*
11. N08- 584- *RsaI*
12. N08- 585- *RsaI*
13. N07- 584- *RsaI*
14. N07- 585- *RsaI*
15. L16- 584- *RsaI*
16. L16- 585- *RsaI*
17. C04- 584- *RsaI*
18. C04- 585- *RsaI*
19. J13- 584- *HaeIII*
20. J13- 585- *HaeIII*
21. M18- 584- *HaeIII*
22. M18- 585- *HaeIII*
23. No7- 584- *HaeIII*
24. No7- 585- *HaeIII*
25. A06- 584- *HaeIII*
26. A06- 585- *HaeIII*
27. E06- 584- *HaeIII*
28. E06- 585- *HaeIII*
29. C04- 584- *HaeIII*
30. C04- 585- *HaeIII*
31. L16- 584- *HaeIII*
32. L16- 585- *HaeIII*
33. N08- 584- *HaeIII*
34. N08- 585- *HaeIII*
35. A06- 584- *AluI*
36. A06- 585- *AluI*

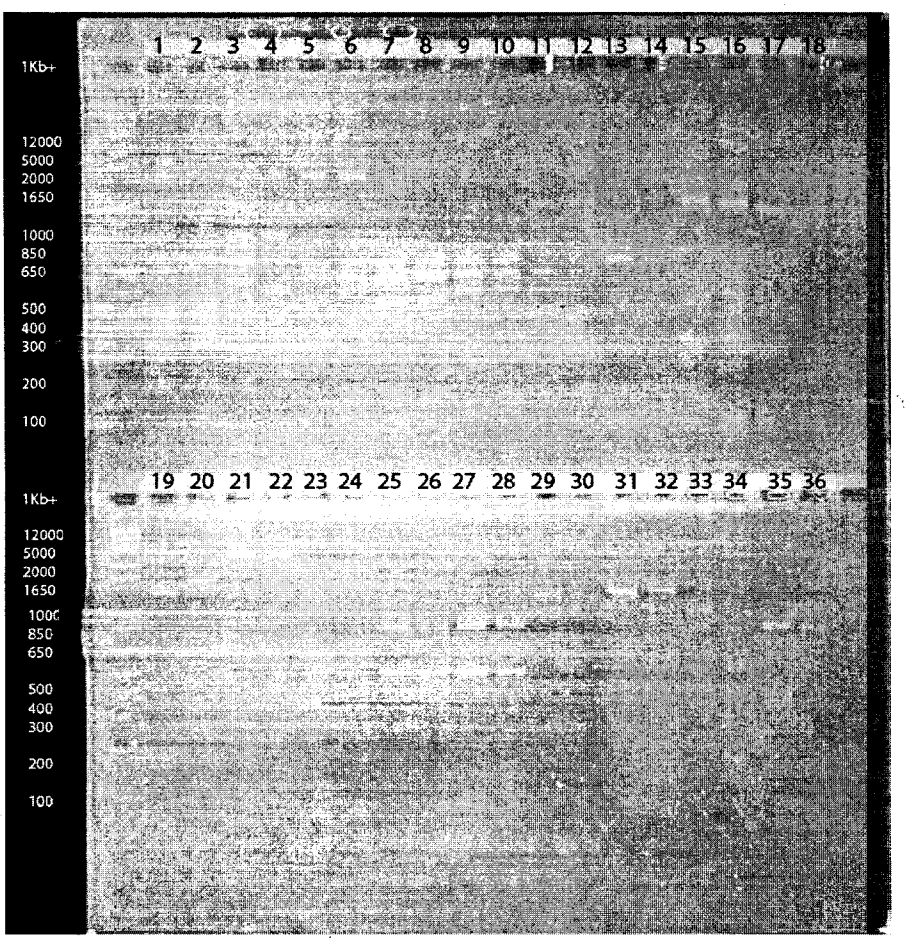


Figure 7: Gel electrophoresis of digested CMEN 584 and CMEN 585 PCR product: Gel electrophoresis of digested CMEN 584 and CMEN 585 PCR product in 2% Agarose gel. The contents of lanes 1 to 36 are written to the left of the gel.

Primer pairs	Digested enzymes	Expected fragment size (bp) in CMEN 584 and CMEN 585	
		CMEN 584	CMEN 585
E06	*Hae*III	500,1000	500, 1000
	*Rsa*I	400, 500, 550	400, 500, 550
N08	*Hae*III	850, 1000	850, 1000
	*Rsa*I	300, 400, 850, 1000	300, 400, 850, 1000
	*Taqα*I	250, 310	250, 310
M18	*Hae*III	3000	3000
N07	*Hae*III	500, 550	550
	*Rsa*I	500	500
A06	*Hae*III	500, 1000	500, 1000
	*Rsa*I	400, 470, 550	400, 470, 550
	*Alu*I	300, 850	300, 850
J13	*Hae*III	1500	1650
	*Rsa*I	400, 500, 550, 750	450, 550, 750
C04	*Hae*III	1650	1650
	*Rsa*I	1000	1000
L16	*Hae*III	1650	1650
	*Rsa*I	200, 1000	200, 1000
M06	*Rsa*I	300, 500, 1650	300, 500, 1650

Table 2: Digest fragment sizes of CMEN 584 and CMEN 585 PCR products: Digest fragment sizes of CMEN 584 and CMEN 585 PCR products for primer pairs E06, N08, M18, N07, A06, J13, C04, L16, and M06 digested with the *TaqαI*, *RsaI*, *HaeIII* , and *AluI* restriction enzymes.

If there was any polymorphism between CMEN 584 and CMEN 585, F2 generation template DNAs were also amplified and then digested by the same restriction enzymes. Polymorphism was observed the digested PCR products for the J13 and I20 primers.

The I20 primer pair provides an example of polymorphism detected via restriction digestion with the *Alu*I restriction enzyme (Figure 8, 9, 10). The I20-PCR product of CMEN 584 and 585 digested with *Alu*I had 3 bands. Digested fragment sizes of CMEN 584 were estimated about 150, 350 and 800 bp. Digested fragment sizes of CMEN 585 were estimated about 150, 250 and 800 bp. SAF1-1 had 4 bands with the estimated fragment sizes of about 150, 250,

350 and 800 bp. Thus, the SAF1-1 plant was heterozygous for two co-dominant alleles, designated higher arrow (350 bp band) and lower arrow (250 bp band).

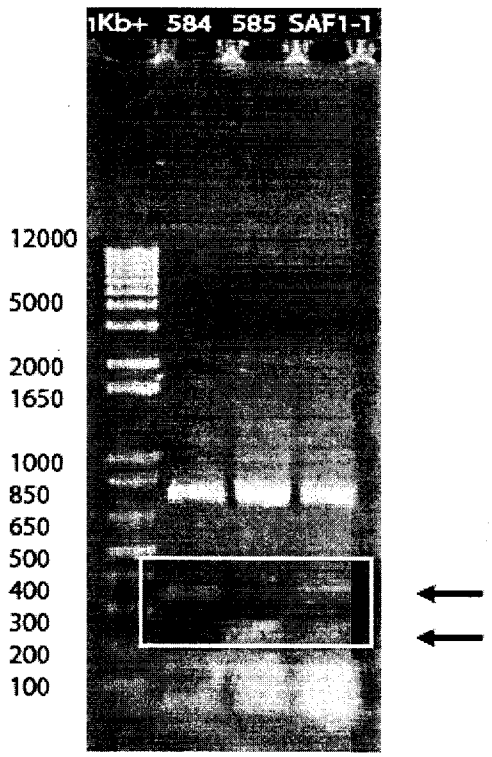


Figure 8: Gel electrophoresis of digested CMEN 584, CMEN 585 and SAF1-1 PCR product for I20 primer with *Alu*I in 2% agarose gel.

The undigested PCR products of CMEN 584, 585, SAF1-1 and some F2 generation individuals were electrophoresis on the gel. All parents, SAF1-1 and F2 generation estimated fragment sizes were approximately 1600bp.

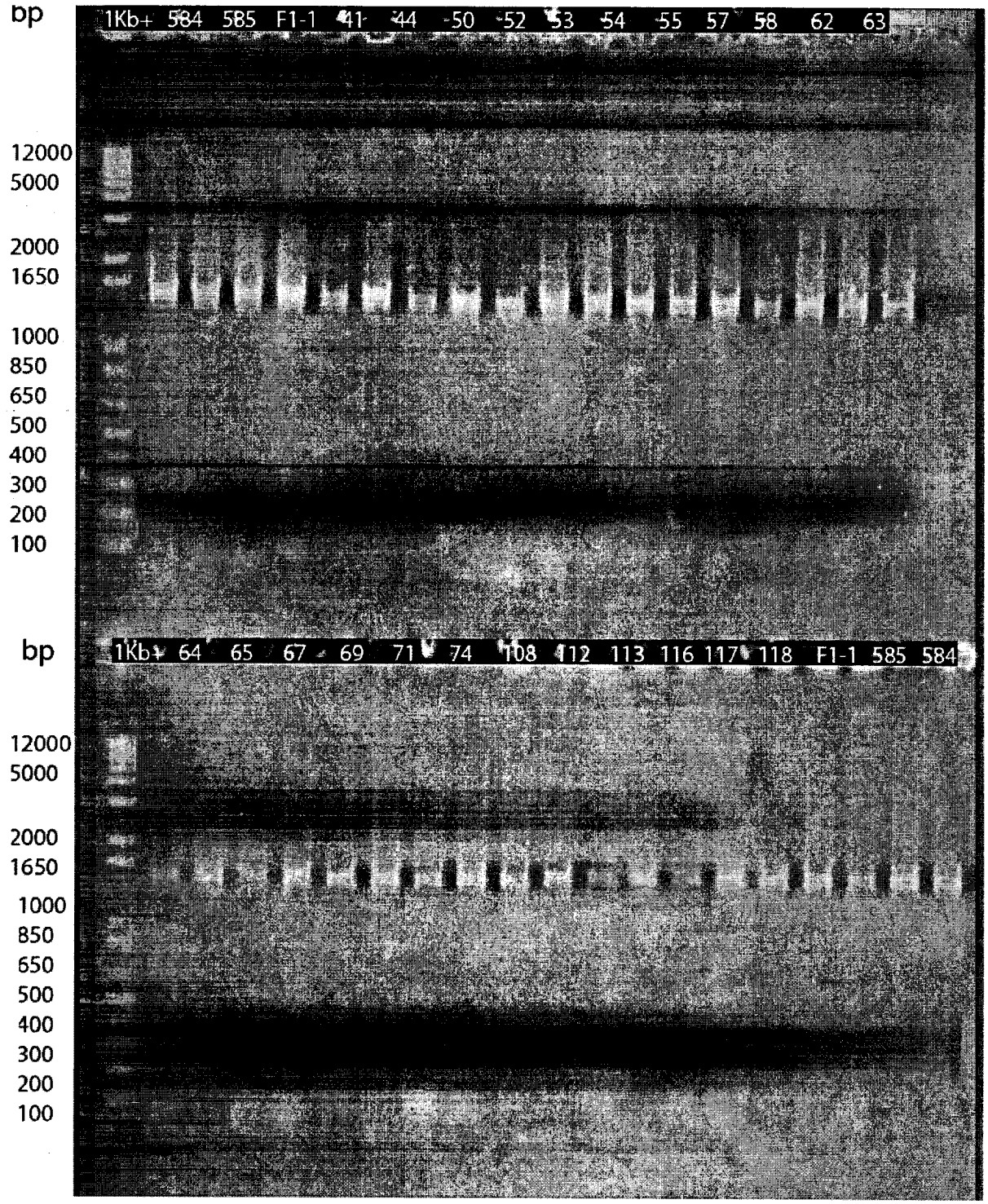


Figure 9: Gel electrophoresis of CMEN 584, CMEN 585, SAF1-1 and F2 generation PCR product for I20 in 2% agarose gel.

F2 generation PCR products were digested with the *Alu*I restriction enzyme. Digested F2 generation was loaded on the agarose gel. (Figure

10)

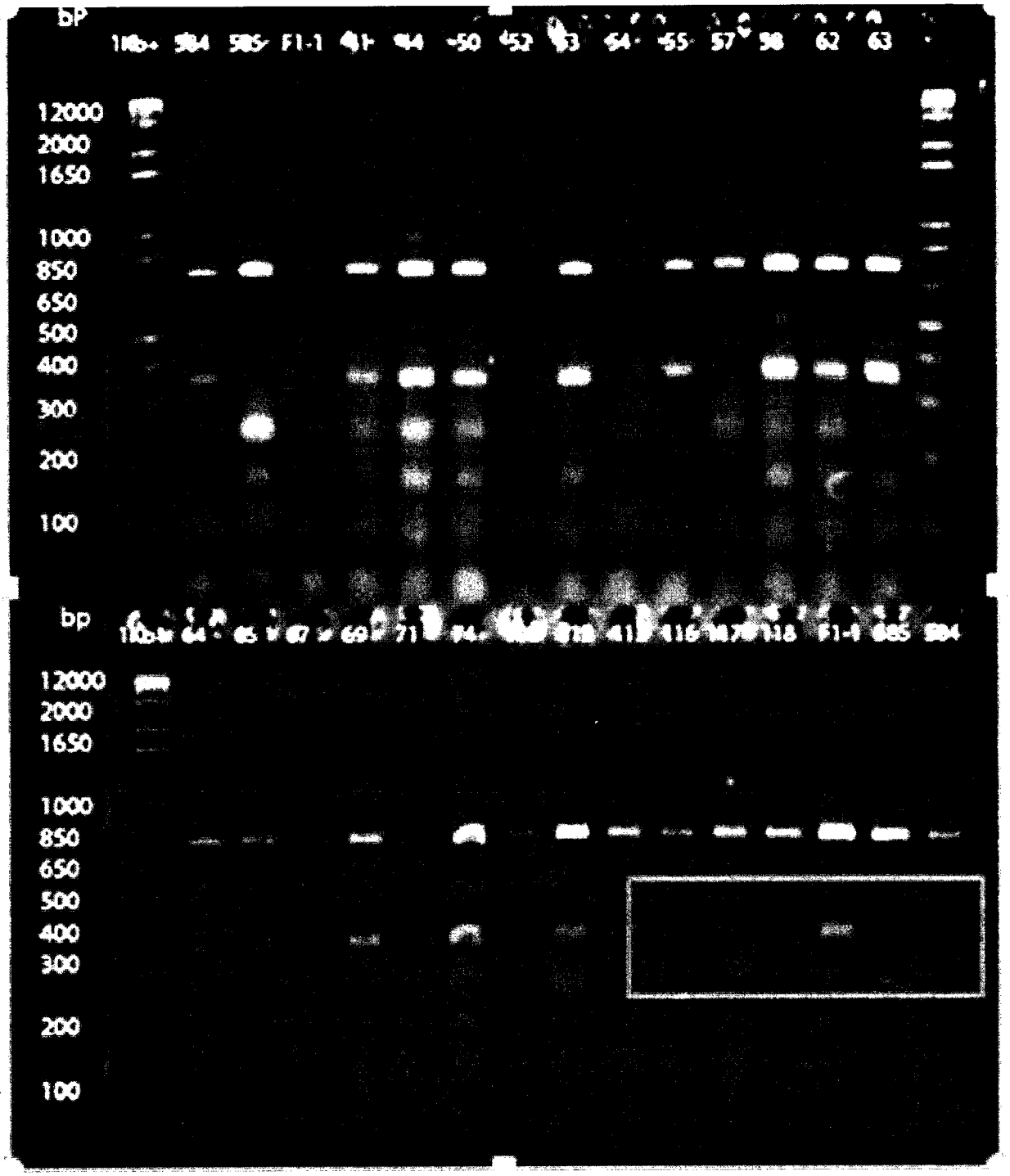


Figure 10: Gel electrophoresis of digested CMEN 584, CMEN 585, SAF1-1 and F2 generation-PCR products for I20 primer by *AluI* restriction enzyme in 2% agarose gel. Blue box was shown to observe the F2 generation digested fragment sizes compare to CMEN 585, CMEN 585 and SAF1-1.

The F2 generation plants were classified based on the banding patterns with respect to the 350 and 250 bp fragment sizes. Genotypes were inferred for CMEN 584 (350 bp band only = A^1A^1), CMEN 585 (250 bp band only = A^2A^2), and SAF1-1 (both 250 and 350 bp bands = A^1A^2) (Table 3).

Fragment Size (Kb)	F2 generation
250 bp	57
350 bp	53, 54, 55, 63, 65, 67, 69, 71, 116
250 bp and 350 bp	44, 41, 50, 58, 62, 64, 74, 108, 112, 113, 117, 118

Table 3: Segregation pattern of the I20 marker in the F2 generation mapping population.

These genotypes occurred in the F2 generation in frequencies of 9/22 (40.9%) A^1A^1; 12/22 (54.5%) A^1A^2; and 1/22 (4.5%) A^2A^2, respectively (Table 3). The observed ratio of 9:12:1 was subjected to a Chi Square Goodness-of-Fit test to the expected 1:2:1 ratio for normal Mendelian segregation of codominant alleles at a single locus. The Chi Square value was 11.72, with a p value of 0.05. Based on the, highly significant difference, the null hypothesis for normal segregation was rejected. This deviation from the expected monogenic 1:2:1 ratio is evidently a case of segregation distortion.

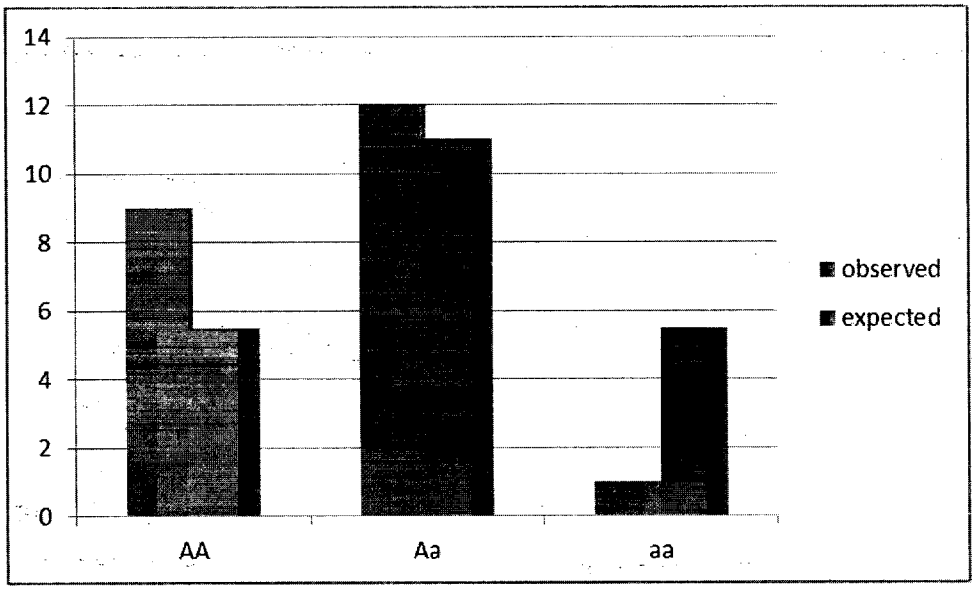


Figure 11: Comparison of observed and expected CMEN 585, CMEN 584, andSAF1-1 in the F2 generation.

To genotype the J13 marker, the PCR products generated by primer pair J13 from the two parents, SAF1-1 and the F2 generation digestion with restriction enzyme *Hae*III of produced the results shown in figure 12. The undigested PCR product sizes were 1650bp. Upon digestion, a band of size 1500 bp plus one or two weak smaller bands were observed in CMEN 584, while bands of sizes 1500 and 1550 bp and some weak smaller bands were observed in CMEN 585. Thus, parent 585 appeared to be heterozygous at the J13 locus. However, the F2 generation was monomorphic for this marker: all plants had the 1600 bp band. This result may be explained by the fact that the SAFS-1 F1 plant had the genotype AA, and thus inherited the A allele from both parents and was not heterozygous (Figure 12).

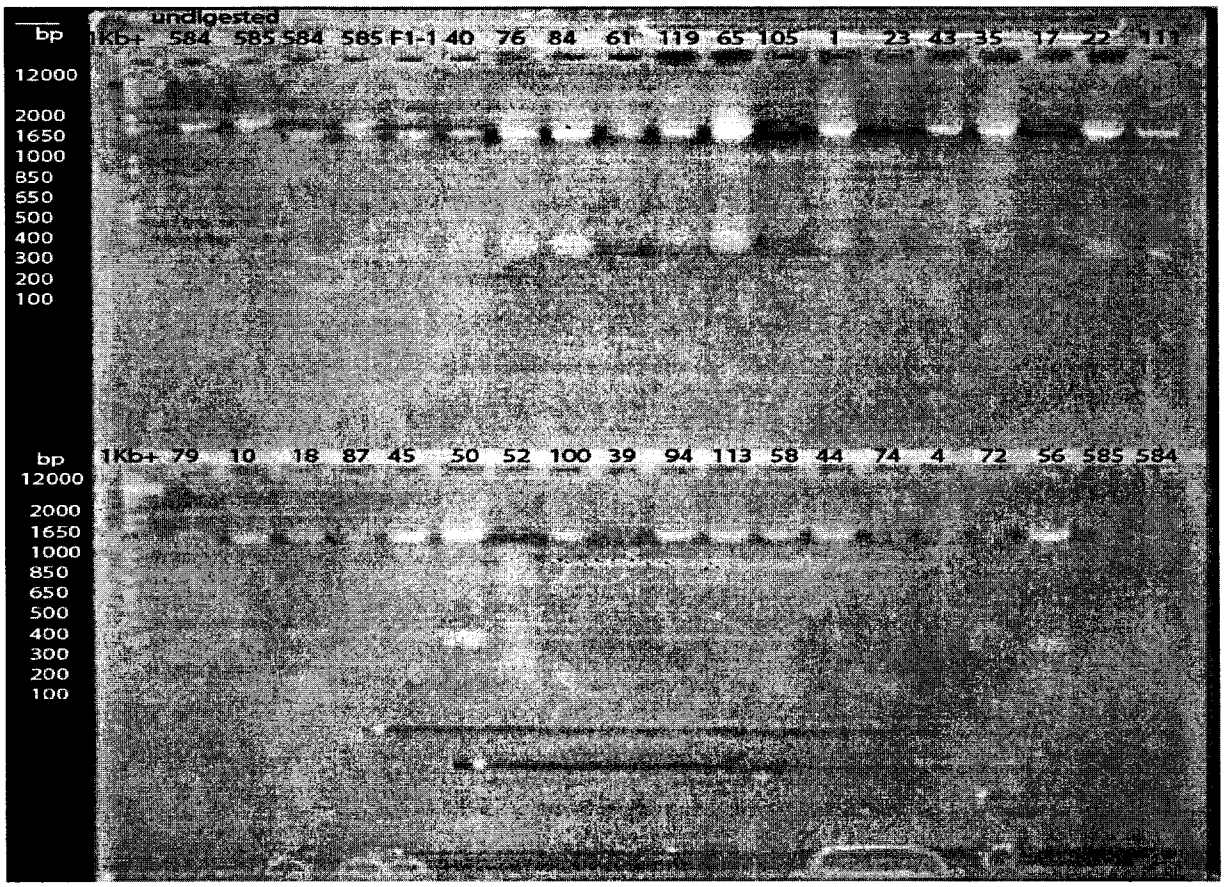


Figure12: Gel electrophoresis of undigested CMEN 584, CMEN 585, and digested CMEN 584, CMEN 585, SAF1-1 and F2 generation of J13-PCR product by *HaeIII* in 2% Agarose gel.

One set of primers (F03) was designed from transposons to show why transposons weren't useful in this research. The undigested and digested F03-PCR products of CMEN 584, 585, SAF1-1 and F2 generation individuals were electrophoresis on the gel. (Figure 13)

The undigested PCR products of CMEN 584 and CMEN 585 (Figure 10), SAF1-1 and F2 generation plants (result not shown) all had an estimated fragment size of about 1400bp. The F03 primer pair reveals polymorphisms detected via restriction digestion with the *RsaI* restriction enzyme but polymorphism wasn't consistent. The F03-PCR product of CMEN 584 digested with *RsaI* had 2 bands with 850 and 1200 bp. Digested fragment sizes of CMEN

585 and F2 generation: 43 had 3 bands with estimated bands of 850, 1000 and 1200 bp. Digested fragment sizes of SAF1-1 and F2 generation. 61, 17, 56, 105 and 4 had 2 bands with estimated sizes of 850 and 1000 bp. F2 generation: 72 had 1 band with the estimated fragment size of about 850bp (Figure 13).

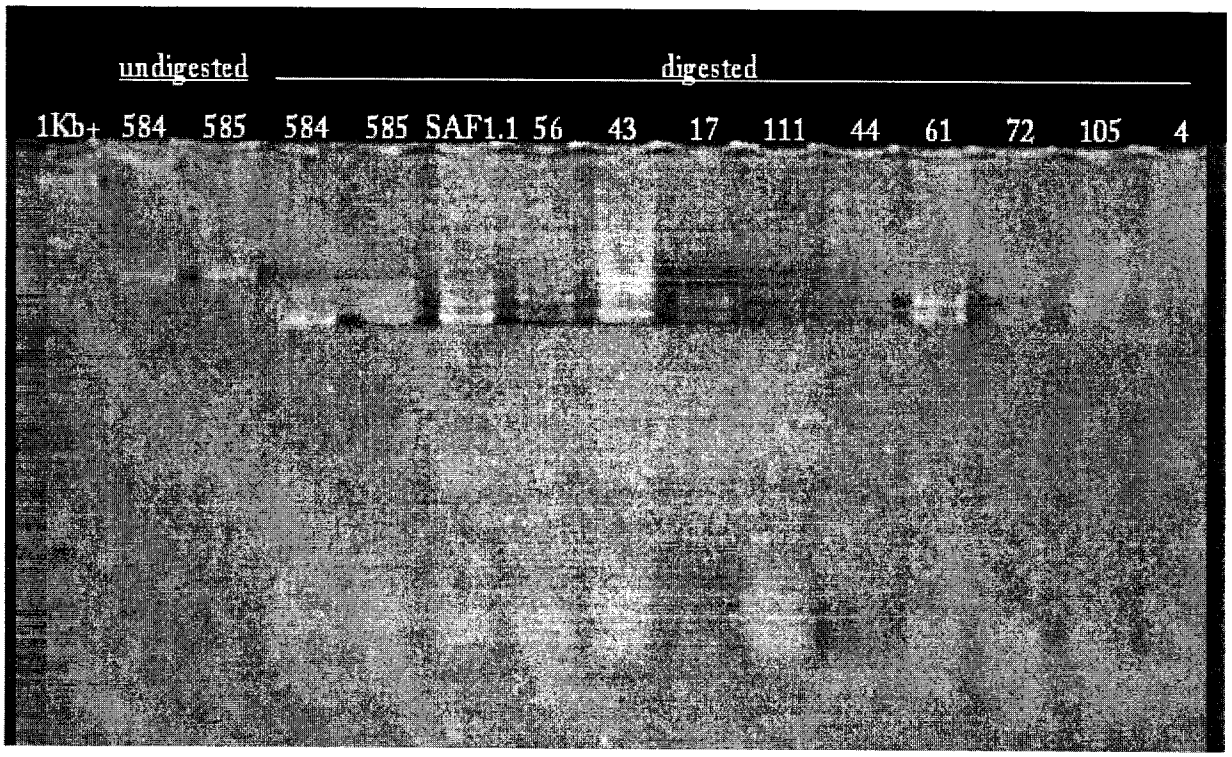


Figure 13: Gel electrophoresis of undigested CMEN 584, CMEN 585, and digested CMEN 584, CMEN 585, SAF1-1 and F2 generation (56, 43, 17, 111, 44, 61, 72, 105, 4) of PCR products for F03 primer by RsaI in 2% Agarose gel.

Designed primers also were used to amplify and test two other germplasm accessions classified most resistant and susceptible to Verticillium wilt (5). CMEN 17 was defined consistently resistant to Verticillium wilt, while CMEN 516 as highly susceptible (5). There was no polymorphisms evidence in any of the accession (Figure 14).

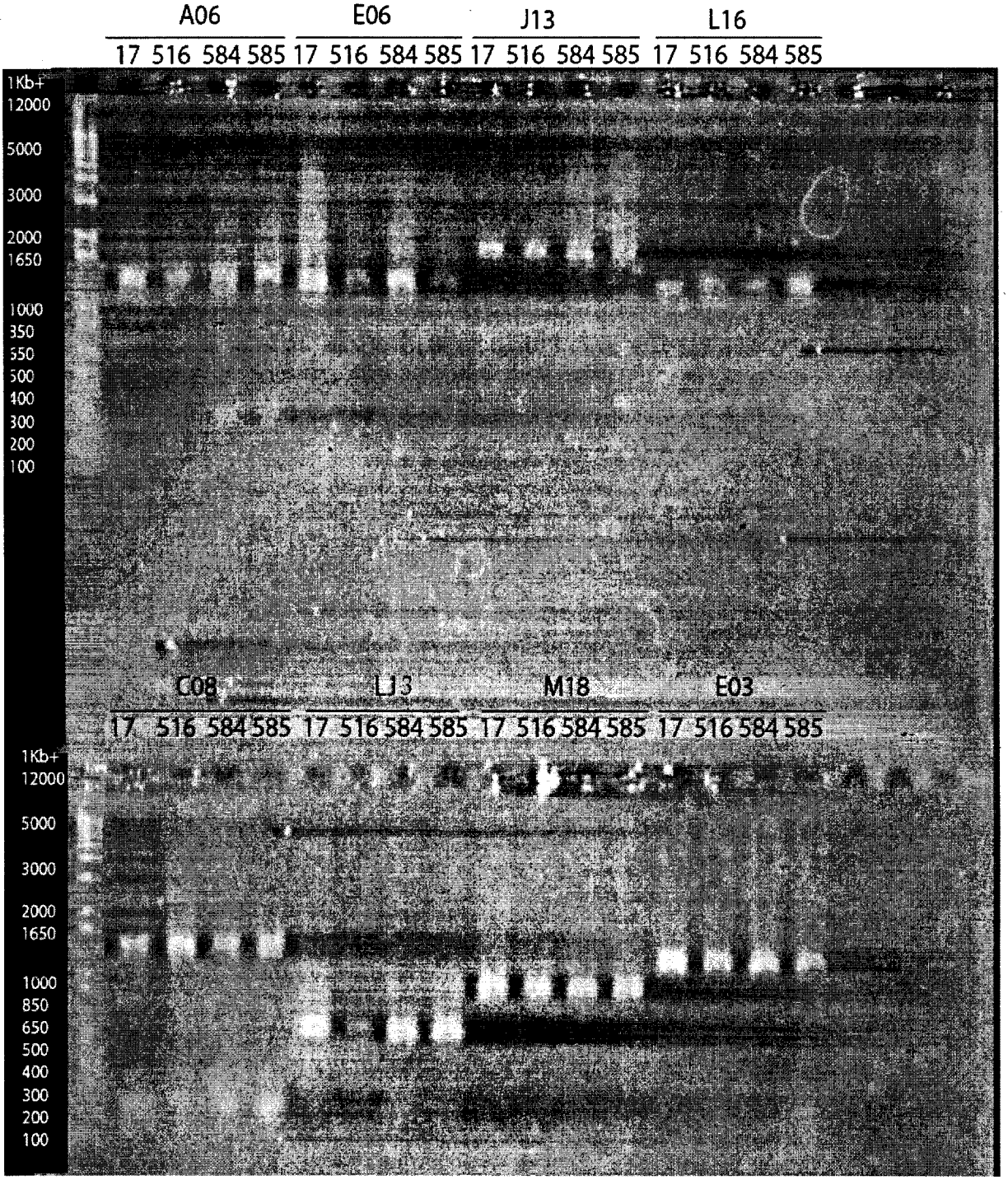


Figure 14: Gel electrophoresis of CMEN 584, CMEN 585, CMEN 17, and CMEN 516 PCR products with the A06, E06, J13, L16, C08, LJ3, M18, and E03 designed primers from the genomic library in 2% Agarose gel.

PCR products from CMEN 17, CMEN 516, CMEN 584, and CMEN 585 were cleaned up and sent to the UNH genomic center for sequencing. Some

sequences were better than others. The best primer PCR product sequence (M18) was chosen to find polymorphisms in these accessions (CMEN 17, CMEN 516, CMEN 584, and CMEN 585) (Figure 15). M18 primer PCR product alignment of four parents revealed a few polymorphisms (SNPs) between different parents. Comparing parents from South Africa and Europe showed that Polymorphisms (here: SNPs) between two parents from one origin were fewer than two parents from different origin (Figure 15).

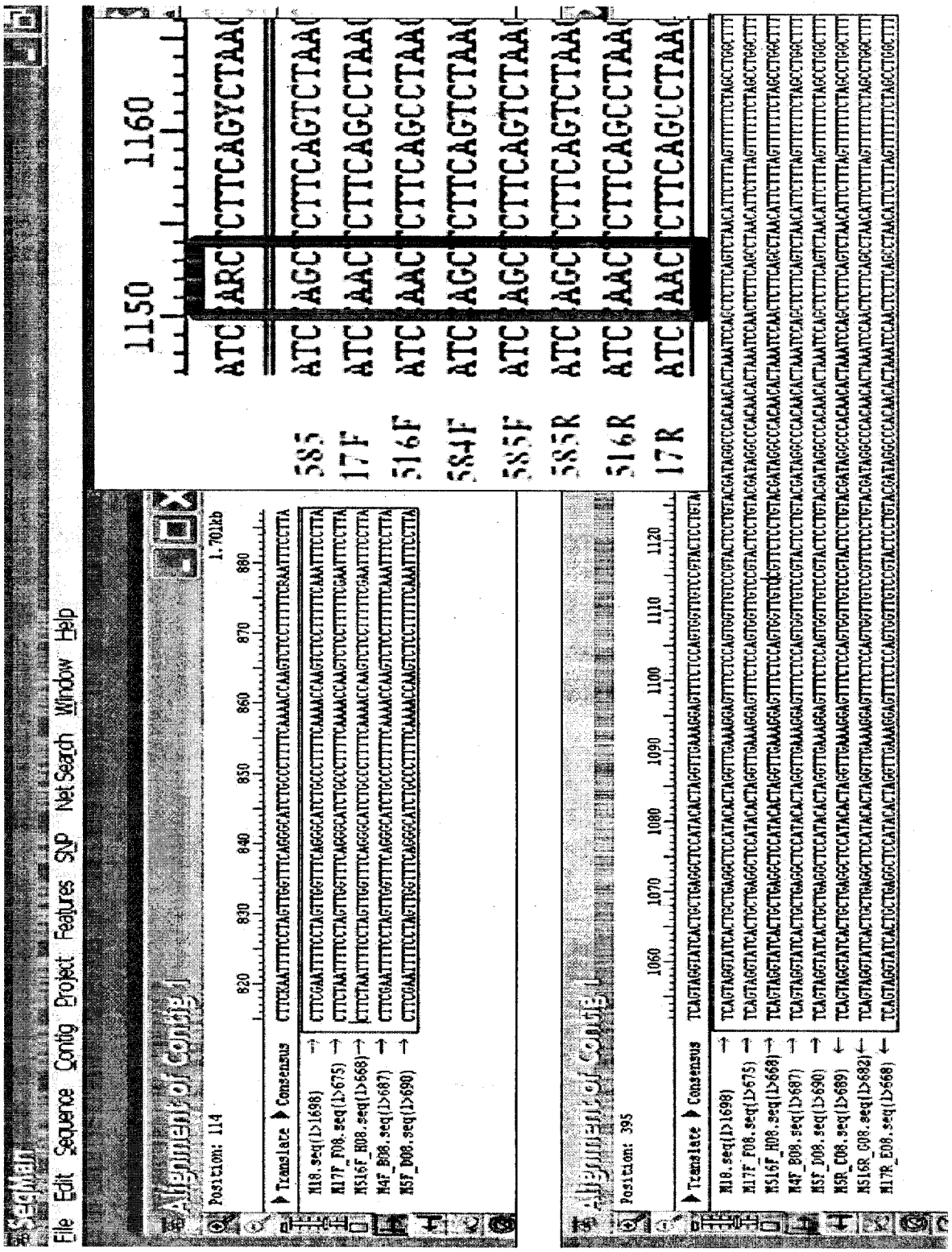


Figure 15: Sequence alignment of CMEN 584, CMEN 585, CMEN 17, and CMEN 516 PCR product with the M18 designed primer from the genomic library. Red box at the top reveals the SNPs between accessions from South Africa and accessions from Europe.

The LJ3 primer pair, which targets the limonene synthase gene, was used as a control with other primers on the gel (Figure 16). (20) This gene is involved with essential oil production in mint. The PCR product of this primer didn't show any polymorphism between CMEN 17, CMEN 516, CMEN 584, and CMEN 585. PCR product of this primer was sent for sequencing to the UNH genome center. Sequences of parents from different origins didn't align as well as parents from the same origin (Figure 16).

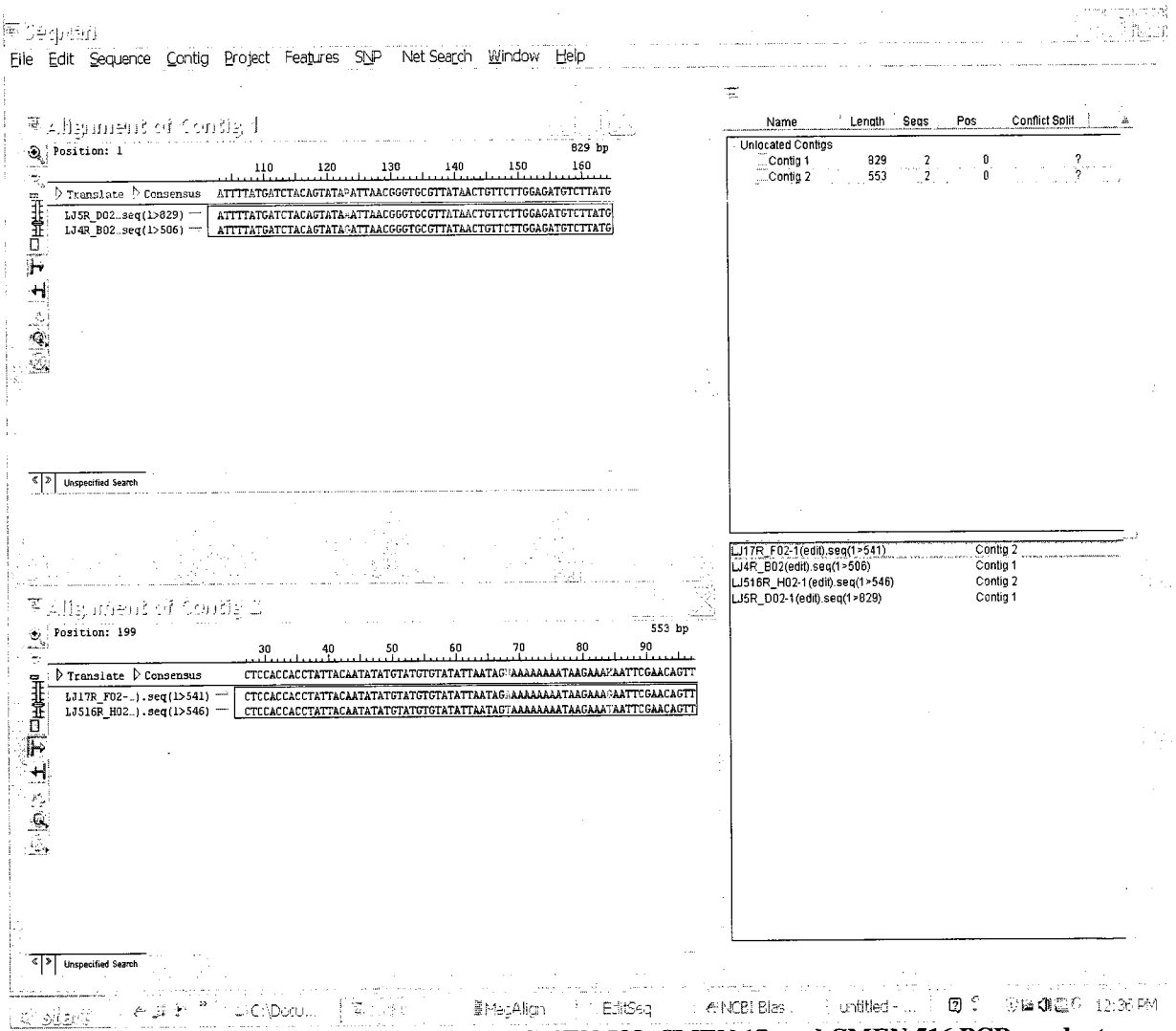


Figure 16: Sequence alignment of CMEN 584, CMEN 585, CMEN 17, and CMEN 516 PCR product with the LJ3 designed primer from the genomic library.

Obtained sequences were compared with the existing LJ3 cDNA (20) to find an intron position of the parent sequences. An intron was obtained is shown in Figure 17. The intron part begins after the highlighted sequence (Figure 17).

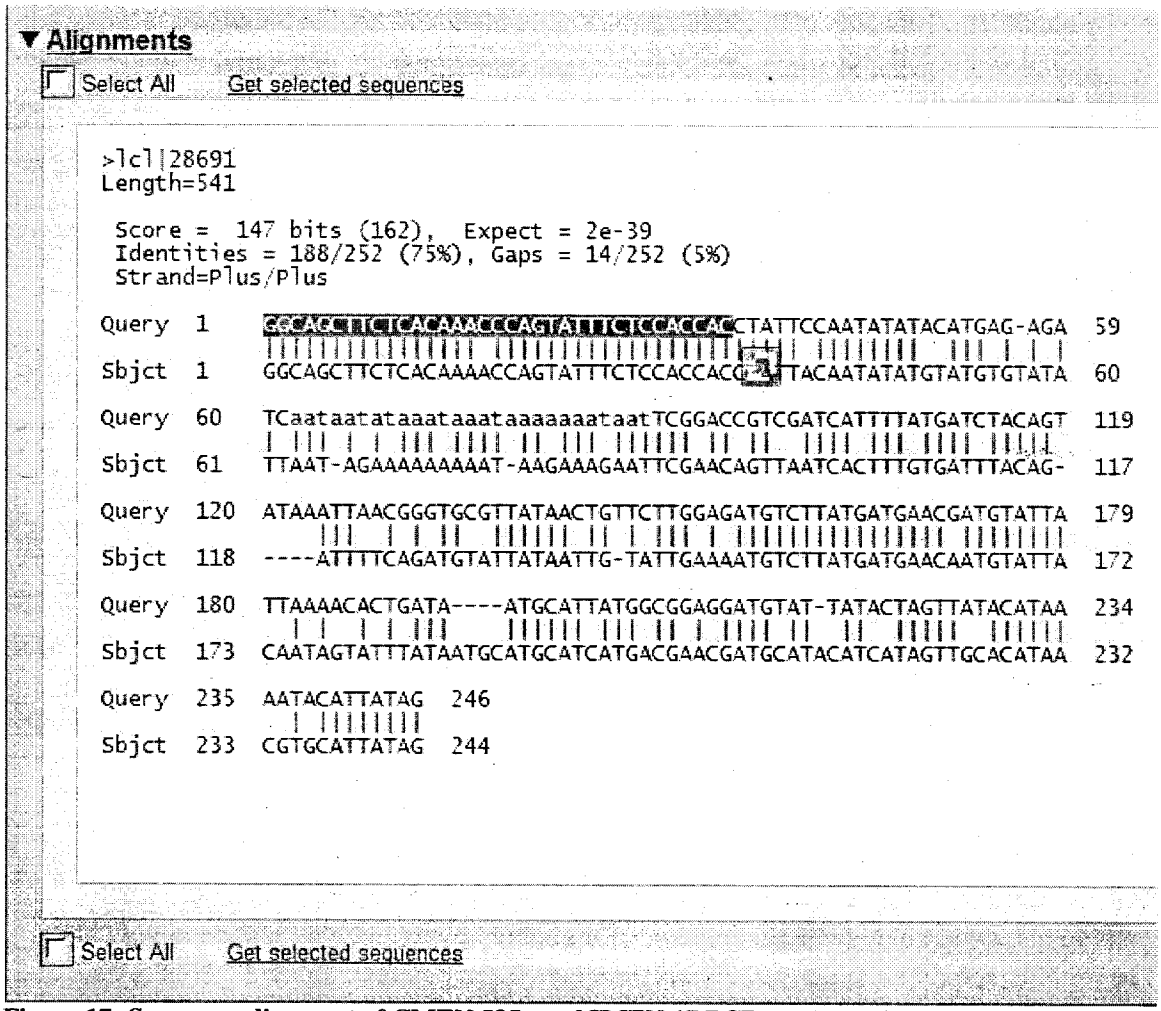


Figure 17: Sequence alignment of CMEN 585 , andCMEN 17 PCR product with the LJ3 primer. Quary: LJ3 –CMEN 17 and subject: LJ3 –CMEN 585.

DISCUSSION

With the goal of assessing the usefulness of an existing F2 generation (CMEN 585 x CMEN 584) population as a linkage mapping population in *Mentha longifolia*, a small genomic library of CMEN 585 was initially constructed, and 279 clones were sequenced and annotated. Using this resource of new DNA sequences from mint, 19 PCR primer pairs were designed and used to assess genetic diversity among four accessions of *M. longifolia*. Some useful molecular markers were developed, despite the many technical difficulties encountered during this process. Sufficient data was generated to show that the CMEN 585 x CMEN 584 mapping population was not sufficiently polymorphic to allow map construction using the CAPS marker approach. However, as discussed below, useful genomic resources for further research on the genetic basis of verticillium wilt resistance in *M. longifolia* were developed, and a promising future direction for this research is proposed.

Genomic libraries are typically larger than the one constructed in this research, containing 1000's or 10,000's of colonies. However, for the purpose of this research, a few hundred colonies were sufficient. The applied method of making a genomic library in this study is adequate to make a larger genomic library. A 2 ul aliquot of DNA end-repaired genomic DNA was enough to produce 3 petri plates with an average of 93 colonies each. An additional 24 ul of DNA end-repaired genomic DNA is retained frozen: enough to expand the available library ten-fold if the need arises in future investigations. Also, a greater yield of

library colonies might be obtainable by use of electrocompetent *E. coli* instead of chemically competent *E. coli*. Electroporation using electrocompetent *E. coli* is considered a more efficient transformation method than the heat shock method using chemically competent cells as applied here (24).

Sequences from the CMEN 585 genomic library were aligned to matching sequences from the Viridiplantae database of GenBank Sequences by using Blastx and tBlastx algorithms.

These homology searches indicated that the obtained genomic sequences compriseded 215,127 bp DNA from the nuclear genome, 44,694 bp DNA from chloroplast the chloroplast genome (cpDNA), and 11,669 bp DNA from the mitochondria mitochondrial genome (mtDNA). Thus observed ratio of cloned sequences from nucleus, chloroplast, and mitochondria was about 18:4:1 (Appendix A).

The estimated size of the *M. longifolia* nuclear genome is about 450 Mb (5), which is about 3 x 10^3 times the size of a typical chloroplast genome. The range size of cpDNAs in angiosperms and land plants can be between 120-160 kb (21). However, a typical plant cell (e.g., in the palisade layer of a leaf) might contain as many as 50 chloroplasts, and each chloroplast contains multiple copies of its genome (22). Thus, it is not surprising that nuclear genome sequences outnumbered chloroplast genome sequences in my library by a ratio of only about 4.6 to 1.

Similarly, the mitochondria number within plant cells ranges from a few hundred to thousands (18, 21). In comparison to the chloroplast genome, the

size of the mitochondrial genome is quite variable (200- 2400 kb) (22, 23). The size of the mint mitochondrial genome is not known. However, since the mtDNA sequences were in the lowest proportion in my library, it is reasonable to speculate that the mint mitochondrial genome is unlikely to be unusually large.

The best matches from the Blast alignments (i.e., the best E-values) were associated with genes from the chloroplast and mitochondria due to their highly conserved genomes. These genomes are usually transmitted uniparentally (26), and thus are not subject to heterozygosity and recombination. Unfortunately, the presumed uniparental inheritance of chloroplast and mitochondria genomes, in which their genes are not transmitted through the pollen, makes them useless as a basis for developing the segregating molecular markers needed to make a genetic linkage map. In the CMEN 585 x CMEN584 population for which I had hoped to make a linkage map, Verticillium wilt resistance was segregating (5), and therefore was presumed to be determined by nuclear genes. So there was no need to find molecular markers from the chloroplast or mitochondrial genomes.

Approximately 37% of obtained genomic sequences were transposon-related. The amount of retrotransposons differs greatly among plant genomes. In broad bean (*Vicia faba*) plants with a chromosome number of a 2n=12, 10% of the genome is retroposable elements, while in strawberry (*Fragaria* spp.) with 2n = 56 chromosome number, 15% is retroposons, and maize (*Zea mays* L. ssp) with 2n = 20 chromosome number is close to 90% of genome is retroposable elements (28, 29, 30).

Transposons weren't considered a good choice for designing primers in my study. Retrotransposons exist as large, multicopy familes, so primers based on them may produce multiple bands that are hard to interpret as markers, as illustrated by the results with primer pair F03 (Figure 13).

The last and best option for marker development was protein-encoding nuclear genes that were not transposon-related, and are inherited bi-parentally. These comprised about 12% of the sequences. The exon positions of such genes are likely to be very stable, and these genes are not expected to accumulate as many mutations as introns or transposons. This stability is one of their positive points which made them best for designing primers that had a high probability of working on all the germplasm accessions to be assayed. The protein-encoding nuclear genes with the best Blast search E-values were given the most attention and the primer apirs such as I20 or J13 were designed from them.

To acquire molecular markers the CAPs method was applied. Surprisingly, the CAPs method had some limitations. Parents CMEN 584 and CMEN 585 were different morphologically, in oil composition and Verticillium wilt resistance (5), and so it had been anticipated that they would differ sufficiently at the DNA sequence level to facilitate the use of the CAPS method. When DNA sequences from these two mapping parents were amplified using 11 of the primer pairs (D07, M18, I04, A06, J13, E06, C04, L16, M06, N08, and N07), no electrophoretic band mobility polymorphisms were detected (Figure 5), thus providing no immediate evidence of indel (insertion/deletion) polymorphisms

between CMEN 585 and CMEN 584. Even when the PCR products were digested with several different restriction enzymes to produce multiple digestion fragments, electrophoretic band mobility polymorphisms were detected in only a few of the digests, as for instance with primer pairs J13 and I20. Even the identified polymorphisms did not always segregate in the F2 population, possibly because they were heterozygous in one or both of the parents, and alternate alleles were not always transmitted to the F1 plant (SAF1) used to make the F2 generation, as evidenced by the results with primer pair J13 (Figures 7, 8).

For making a genetic linkage map, the number of markers needed in different plants depends on their chromosome number. In mint, with 12 pairs of chromosomes, and at least 5 markers needed per each pair of chromosome, a total of 60 markers is considered the minimum needed. Unfortunately, application of the CAPS method using arbitrarily chosen restriction enzymes wasn't sensitive enough to detect enough polymorphisms between CMEN 585 and CMEN 594 to provide the number of molecular markers needed to make a genetic linkage map.

When selected PCR products from CMEN 585 and CMEN 584 were sequenced and compared, some sequence polymorphisms such as SNPs and small indels were observed, but only a few of these resulted in changes in restriction sites or the sizes of restriction fragments that could be detected as CAPS markers. This shows that the CAPS method was not sensitive enough to detect most of variations that did exist between the two mapping parents. Thus, it was concluded that genotyping methods such as direct sequencing or direct detection of SNPs are required for future studies with this germplasm.

Accessions CMEN 585, CMEN 584, CMEN 17, and CMEN 516 had been evaluated previously based on their geographic origin, oil composition, and *Verticillium* wilt resistance and suitability (5). In this research, the aim was check molecular marker polymorphisms with respect to their aspects of morphology, and especially verticillium resistance, which Vining defined in CMEN 585 and CMEN 17 as consistently resistant to Verticillium wilt, while CMEN 584 and CMEN 516 were highly susceptible (5).

CMEN 585 and CMEN 584 were from South Africa, while CMEN 17 and CMEN 516 were from Europe. Comparison betwee the accessions from South Africa or between those from Europe didn't reveal sufficient variation for marker development using the CAPS method. However, much greater variation existed between the South Africa and Europe accessions. The variation between the South African and European accessions is indicative of genomic divergence between accessions from distant geographical origins. The polymorphism between the European and South African accessions is sufficient to recommend a plan for developing a new mapping population by making a segregating population from a Europe and South Africa cross.

In conclusion, the genomic resources that were developed in this study, including a small genomic library of *M. longifolia* resistant variety CMEN 585, sequences of 279 genomic clones which have been deposited in GenBank under accession numbers HR308784- HR309062 and 19 pairs of PCR primers for protein-encoding nuclear genes, help to advance the genetic study of the mint diploid model species, *Mentha longifolia* (Appendix A and B). The results of this

study provide a basis for outlining a strategy for future research directions in this system. Specifically, I recommend crossing between accessions CMEN 516 and CMEN 585 as the next step, with the aim of creating a new population that is segregating for Verticillium wilt resistance/susceptibility. CMEN 516 as a highly susceptible is from Europe and CMEN 585 as a highly resistance is from South Africa. The CMEN 585 genomic library and designed primers which have been made in this research are useful to find a polymorphism between these two accessions as parents, and their F1 and F2 generation population.

Given that at least 60 markers are needed to make a genetic map, more primers will be needed to find enough markers. More primers can be obtained by making bigger genomic library, 2) CAPS method will be easier and faster by sequencing to define a restriction sites, 3) CAPS method can be replaced with other genotyping methods such as Restriction Fragment Length Polymorphism (RFLP) or SNPs.

REFERENCES

1. Cantino, P.D., and Sanders, R.W. 1986. Subfamilial classification of Labiateae. Syst. Bot. 11: 163-185.

2. Harley, R.M., and Brighton, C. A. 1977. Chromosome numbers in the genus *Mentha* L. J. Linn. Soc. Bot. 74: 71-96.

3. *Reinhart, H.M.,* and *Hill, G.D.* 1991. *Agricultural plants,* Cambridge University Press. 3: 285-287.

4. Tepe, Ş., and Ellialtıoğlu, Ş. 2002. Obtaining polyploidy mint (*Mentha longifolia* L.) plants with in vitro colchicine treatment. AKDENİZ ÜNİVERSİTESİ ZİRAAT FAKÜLTESİ DERGİSİ. 15(2): 63-69.

5. Vining, K.J. 2007. Studies of Verticillium wilt and characterization of candidate verticillium wilt resisrance genes in the mint species *Mentha longifolia* (L.) Huds. University of New Hampshire.

6. Bunsawat, J., Elliott, N.E., Hertweck, K.L., Sproles, E., and Alice, L.A. 2004. Phylogenetics of *Mentha* (Lamiaceae): evidence from chloroplast DNA sequences. Systematic Botany 29(4):959-964.

7. USDA, plant profile, *Mentha arvensis,* <http://plants.usda.gov/java/profile?symbol=MEAR4>

8. Bowers, J., and Klingeman, P. 2005. Diagnoses and management of Verticillium wilt, a nursery and landscape plant pathogen. Tennessee Greentime: 20- 23.

9. Weiner, M.A. 1990. Weiner's herbal: the guide to herb medicine. 2nd, Quantum Books, Mill Valley, CA.284.

10. Semagn, K., Bjørnstad, Å. and Ndjiondjop, M. N. 2006. An overview of molecular marker methods for plants. African Journal of Biotechnology. 5 (25): 2540-2568.

11. Kunihisa, M., Fukino, N., and Matsumoto, S. 2003. Development of cleavage amplified polymorphic sequence (CAPS) markers for identification of strawberry cultivars. Euphytica 134: 209–215.

12. Orcheski, B.B., and Davis, T.M. 2010. An enhanced method for sequence walking and paralog mining: TOPO® Vector-Ligation PCR. BMC Research Notes. 3:61.

13. Barone, A. 2003. Molecular marker-assisted selection for resistance to pathogens in tomato. A paper presented during the FAO international workshop on "Marker assisted selection: A fast track to increase genetic gain in plant and animal breeding". 17-18 October 2003, Turin, Italy.

14. Diwan, N., Fluhr, R., Eshed, Y., Zamir, D., and Tanksley, S.D. 1999. Mapping of *Ve* in tomato: a gene conferring resistance to the broad-spectrum pathogen, Verticillium dahliae race 1. Theoretical and Applied Genetics. 98(2): 315–319.

15. Query result for CMEN 17, <http://www.ars-grin.gov/cgi-in/npgs/acc/search.pl?accid=%20PI+557755>

16. Query result for CMEN 584, <http://www.ars-grin.gov/cgi-in/npgs/acc/search.pl?accid=%20PI+557769>

17. Query result for CMEN 585, <http://www.ars-grin.gov/cgi-in/npgs/acc/search.pl?accid=%20PI+557767>

18. Plant genome and structure, <http://www.ndsu.edu/pubweb/~mcclean/plsc731/genome/genome7.htm>

19. Lacey M.L., Stephens C.T., Green R.J., and York Jr. A. Mint production in the Midwestern United States. NCR publication number 155.

20. Jones A.J. 1997. Characterization of *4S-limonene synthase*: a gene involved with essential oil production in the mint family. University of New Hampshire.

21. Palmer J.D. 1985. Chloroplast DNA and Molecular Phylogeny. BioEssays 2(6): 263-267.

22. Ward B.L., Anderson R.S., Bendich A.J. 1981. The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). Cell 25: 793-803.

23. Palmer J.D. 1985a. Evolution of chloroplast and mitochondrial DNA in plants and algae. In: MacIntrye RJ (ed) Monographs in evolutionary biology: molecular evolutionary genetics. Plenum. New York. 131-240.

24. Wul, N., Matand, K., Kebede, B., Acquaah, G., and Williams, S. 2010. Enhancing DNA electrotransformation efficiency in *Escherichia coli* DH10B electrocompetent cells, Electronic Journal of Biotechnology ISSN: 0717-3458

25. Bennetzen, J.L. 1996. The contribution of retroelements to plant genome organization, function and evolution. Trends in Microbiology 4: 347-353.

26. Birky, C.W. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: Mechanisms and evolution (organelles/non-Mendelian inheritance/maternal inheritance. Proceedings of the National Academy of Sciences USA: 92: 11331-11338

27. Weir, B.S., 2001. Bioprospecting environmental metagenomes - BTech Honours report. Bioprospecting Environmental Genomes. BTech Honours Thesis and the University of Auckland.

28. Rédei, G.P. 1999. Genetics manual: current theory, concepts, terms. World Scientific Publishing Co Pte, Ltd.

29. Sleper, D.A., and Poehlman J.M. 2006. Breeding field crops. Fifth Edition Blackwell Publishing professional, Iowa.

30. Wikipedia, Vicia faba, <http://en.wikipedia.org/wiki/Vicia_faba>

APPENDICES

APPENDIX A: Result of blastx and tblastx algorithms of CMEN 585 genomic library.

Colony Number	Few Rev Contig	Length	Chloroplast Mitoch Nucleous	Top Blast match	Genome Number	E-value
A02	F	804	N	transposase	gb|ABD32507.1|	2.00E-09
A04	C	877	N	hypothetical protein	ref|XP_001690467.1	4.1
A06	C	1380	N	hypothetical protein	ref|XP_002283442.1|	1.00E-30
A08	F	775	N	hypothetical protein	ref|XP_002537594.1|	3.2
A08	R	883	N	hypothetical protein	emb|CAN72461.1|	4.00E-69
A10	F	832	M	NADH ubiquinone oxidoreductase subunit 5	gb|AAB97305.1|	8.00E-30
A10	R	812	N	hypothetical protein	ref|XP_002535064.1|	3.40E-02
A12	C	636	N	kinase	ref|XP_002527895.1|	9.00E-09
A18	R	878	N	hypothetical protein	ref|XP_002452397.1|	0.044
A20	C	1554	N	hypothetical protein	ref|XP_002280371.1|	1.40E+00
A22	R	872	N	hypothetical protein	ref|XP_002526925.1	1.40E+00
A24	R	438	N	MADS transcription factor	emb|CAI47596.1|	7.00E+00
B01	F	831	C	anthranilate synthase	ref|XP_002316223.1	7.00E-22
B01	R	772	C	anthranilate synthase alpha subunit	emb|CAC29060.1|	7.00E-08
B02	F	596	C	RNA polymerase beta I subunit	gb|ADA69917.1|	0.28
B02	R	789	N	RNA polymerase beta I subunit	ref|YP_567068.1	9.00E-138
B04	R	840	N	RAB6-interacting protein, putative	ref|XP_002523956.1|	2.00E-10
B04	R	568	N	RAB6-interacting protein, putative	ref|XP_002523956.1|	3.00E-23
B05	C	1421	N	hypothetical protein	ref|XP_002269501.1	1.00E-16
B06	F	835	N	hypothetical protein	ref|XP_002268850.1|	4.00E-33
B07	C	1536	N	hypothetical protein	ref|XP_002267479.1|	8.00E-47
B09	F	737	N	hypothetical protein	ref|XP_002488942.1	2.00E-07
B09	R	877	N	putative gag-pol polyprotein	gb|AAL76004.1|	5.00E-71
B10	F	929	C	protein kinase atmrk1	ref|XP_002517090.1	7.00E-64
B10	R	804	N	protein kinase atmrk1	ref|XP_002517090.1	5.00E-23
B11	C	1242	N	RNA polymerase beta II subunit	gb|ADA69916.1|	2.00E-113
B12	F	795	N	lycopene beta-cyclase	gb|ACJ66628.1|	9.20E+00
B13	R	876	N	replication protein A 70b	dbj|BAC77529.1	0.02
B14	R	886	N	predicted protein	ref|XP_001783991.1	3.70E-01
B15	R	917	N	putative plasmalemma Na+/H+ antiporter	emb|CAG30524.1|	1.00E-31
B16	F	869	N	unnamed protein product	emb|CBI26990.1	4.10E+00
B16	R	897	N	hypothetical protein	emb|CAN80692.1|	0.015
B17	C	1001	N	hypothetical protein	gb|EAY81992.1|	7.00E+00
B18	R	885	N	maturase	gb|AAL77621.1|	5.40E+00

B19	C	1615	N	gag-pol polyprotein	gb|AAO73529.1|	2.00E-30
B20	C	371	C	Ycf2	gb|ADA69968.1|	4.00E-48
B21	R	853	N	retrotransposon protein	gb|ABG22406.1|	4.40E-02
B22	R	913	N	polyprotein	gb|ABG22120.1|	1.00E-55
B23	R	865	N	unknown protein	ref|NP_194735.2|	6.30E-01
B24	R	750	N	unknown protein product	emb|CBI25960.1	2.20E-01
C04	R	729	N	hypothetical protein	ref|XP_002265846.1|	3.00E-58
C06	F	880	M	diphosphoinositol polyphosphate phosphohy	ref|XP_002513032.1|	2.00E-04
C06	R	917	N	amino acid transporter	ref|XP_002526910.1|	1.00E-39
C08	C	1414	N	predicted protein	ref|XP_002330863.1|	9.00E-21
C10	F	794	M	NADH ubiquinone oxidoreductase subunit 5	gb|AAB97305.1	8.00E-30
C10	R	727	N	hypothetical protein	ref|XP_002535064.1	3.40E-02
C14	F	887	N	hypothetical protein	emb|CAN78493.1	3.00E-07
C14	R	898	N	gag-pol polyprotein	gb|AAR13317.1|	4.00E-08
C16	C	1392	N	maturase K	gb|AAW64933.1|	0.00E+00
C20	R	920	C	cytochrome b559 beta chain	gb|AAS46065.1|	2.00E-15
C22	R	836	N	alpha-gliadin	gb|ACY71754.1|	1.10E+00
C24	F	963	N	putative transposase	gb|AAO38443.1|	3.00E-35
C24	R	964	N	Zinc finger, RING-type	ref|XP_002277399.1|	7.00E-29
D02	F	891	N	hypothetical protein	emb|CAN75466.1	3.7
D02	R	821	N	putative wall-associated serine/threonine kin	dbj|BAD19239.1|	0.089
D03	C	1290	N	GmMYB29A2	dbj|BAA81733.2|	8.00E-24
D05	C	1382	C	envelope membrane protein	ref|NP_683815.1|	0.34
D07	C	1480	N	putative gag-pol polyprotein	gb|AAO37957.1|	3.00E-117
D08	F	863	N	retrotransposon protein	gb|ABA97860.1|	3.00E-39
D08	R	867	N	putative retroelement pol polyprotein	gb|AAD20433.1|	9.00E-18
D12	F	782	N	predicted protein	ref|XP_001774633.1|	8.30E+00
D13	R	821	N	EIF3A_TOBAC RecName: Full=Eukaryotic translation	sp|Q40554.1|	5.00E-64
D16	F	759	N	predicted protein	ref|XP_002304020.1|	8.30E+00
D16	R	956	N	maturase K	gb|ACJ13361.1|	2.80E+00
D19	R	853	N	sugar transporter	ref|NP_197997.1|	1.00E-35
D20	R	334	N	putative copia-like polyprotein	gb|AAK84483.1|	2.00E-28
D21	R	864	N	putative copia-like otein	gb|AAK84483.1|	4.00E-72
D22	F	659	C	hypothetical protein	emb|CAN60947.1|	7.00E-41
D22	R	877	N	putative copia-like polyprotein	gb|AAK84483.1|	6.00E-75
D23	F	891	N	F14N23.4	gb|AAD32866.1|	5.70E-01
D23	R	836	N	sulfate transporter	ref|XP_002517550.1|	6.30E+00

E02	F	775	N	MPAA2_AMBAR RecName: Full=Pollen allergen Amb a 2	sp|P27762.1|	5.10E+00
E02	R	879	N	putative AP endonuclease/reverse tra	gb|AAM82604.1|	2.00E-34
E04	C	1467	C	envelope membrane protein	ref|NP_683815.1|	2.70E-01
E06	C	1449	N	lipid binding protein	ref|XP_002512201.1|	6.00E-28
E08	F	965	N	binding	ref|NP_974047.1|	4.00E-27
E08	R	860	N	binding	ref|NP_974047.1|	2.00E-48
E12	C	480	N	unconventional myosin	gb|AAB71528.1|	3.00E-07
E16	C	805	N	putative somatic embryogenesis receptor kinas	emb|CAJ77499.1|	0.78
E18	C	1584	N	transposon protein	gb|ABF96044.1|	3.00E-07
E20	F	227	N	Zinc knuckle containing protein	gb|ABD65024.1|	4.00E-04
E20	R	819	N	putative copia-like polyprotein	gb|AAK84483.1|	2.00E-85
E24	F	860	N	transposase	gb|AAL69349.1|	6.60E+00
E24	R	940	N	Zinc finger, RING-type	gb|ABD32420.1|	6.00E-28
F01	F	699	N	hypothetical protein SORBIDRAFT	ref|XP_002440609.1|	6.60E+00
F01	R	867	N	retrotransposon protein	gb|ABA95977.1|	3.00E+00
F03	C	1545	N	transposon protein	gb|AAU89193.1|	9.00E-56
F04	C	1592	N	f-box family protein	ref|XP_002319379.1|	5.00E-06
F05	F	825	C/M	NADH dehydrogenase 2-like ORF 260	prf||1211235DE	3.00E-85
F05	R	830	C/M	NADH dehydrogenase subunit 2	ref|YP_001671727.1|	4.00E-60
F07	F	822	N	gag-pol polyprotein	gb|AAR13317.1|	7.00E-37
F07	R	835	N	hypothetical protein	dbj|BAD03833.1|	6.60E+00
F08	F	933	N	RNA polymerase subunit, RPB5; RNA polymerase R	gb|ABN07995.1|	8.00E-11
F09	F	704	N	RNA polymerase beta chain	gb|ABD93758.1|	9.00E-56
F09	R	779	N	RNA polymerase beta chain	gb|ABD93758.1|	2.00E-86
F10	C	816	N	Early nodulin 20 precursor	ref|XP_002515714.1|	1.7
F12	F	854	N	cytochrome P450 like_TBP	dbj|BAA10929.1|	1.00E-60
F12	R	743	N	nucleoporin	gb|ABF99044.1|	3.00E+00
F13	F	822	N	hypothetical protein	emb|CAN65852.1|	3.90E+00
F13	R	934	N	hypothetical protein	gb|EEH51255.1|	1.70E+00
F14	C	959	N	hypothetical protein	ref|XP_002443155.1	3.00E+00
F15	F	787	N	LysM-domain containing receptor-like kinase	gb|ABN50364.2|	2.30E+00
F15	R	920	N	helicase-like protein	dbj|BAB02793.1|	4.00E-69
F16	R	849	C	NADH dehydrogenase subunit F	gb|ABR12227.1|	3.00E+00
F17	F	883	N	hypothetical protein	ref|XP_002277484.1	2.30E+00
F18	F	814	N	hypothetical protein	ref|XP_002264078.1	8.60E+00
F20	R	574	N	hypothetical protein	emb|CBI30815.1	8.60E+00

F22	F	879	N	Putative gag-pol polyprotein	Putative gag-pol p	1.00E-03
F22	R	813	M	retrotransposon protein	gb|ABG66113.1|	8.00E-12
F24	C	371	C	Ycf2	ref|YP_635682.1|	7.00E-48
G02	C	1298	C	photosystem II protein I	ref|YP_002720022.1|	4.00E-07
G06	F	802	N	ring finger containing protein	ref|XP_002524891.1|	0.12
G06	R	861	N	RING-H2 finger protein RHG1a-like	dbj|BAD82497.1|	2.00E-27
G10	F	709	M	NADH dehydrogenase subunit 5	ref|YP_173349.1|	2.00E-20
G10	R	754	N	hypothetical protein	ref|XP_002535064.1|	0.018
G12	F	932	N	hypothetical protein	dbj|BAC84788.1|	8.3
G12	R	856	C	AF314010_1 PSII D1 protein	gb|AAG60671.1|	2.00E-44
G14	C	1268	N	putative gag-pol polyprotein	gb|ACS74199.1|	6.00E-71
G16	C	806	N	putative histone deacetylase complex, SIN	ref|XP_001783401.1|	2.90E+00
G18	C	1449	C	cytochrome b559	prf||1011228A	2.00E-58
G20	R	826	N	hypothetical protein	gb|EEH51255.1|	3.70E+00
G22	F	876	N	ring finger containing protein	ref|XP_002524891.1|	1.20E-01
G22	R	872	N	RING-H2 finger protein RHG1a-like	dbj|BAD82497.1|	2.00E-27
G24	C	913	N	protein disulfide isomerase	ref|XP_002512495.1|	5.00E-23
H01	R	822	N	retrotransposon protein	gb|ABA95977.1|	3.31E+01
H03	C	792	N	leucine-rich repeat-containing protein 2	ref|XP_002520787.1|	6.00E-26
H04	F	926	N	ATP binding protein	ref|XP_002526444.1|	2.00E-13
H04	R	966	C/M	70 kDa heat shock protein	gb|ACT65562.1|	6.00E-08
H05	C	1312	N	CPK related protein kinase 7	ref|XP_002310736.1|	4.30E+00
H06	R	711	N	F-box domain containing protein	ref|NP_001159080.1|	8.00E-04
H07	F	807	C	pentatricopeptide (PPR) repeat-containing pr	ref|NP_172286.1|	8.00E-31
H07	R	909	N	leucine-rich repeat resistance protein-like pr	gb|AAK70805.1|	1.00E-16
H10	C	658	C	ribosomal protein L14	gb|ADA69961.1|	5.00E-62
H11	C	708	C	hypothetical chloroplast RF1	ref|YP_001936563.1|	2.00E-13
H12	F	943	N	Hyp-rich glycoprotein	prf||1814452D	0.1
H12	R	889	C	AF314010_1 PSII D1 protein	gb|AAG60671.1|	3.00E-50
H13	R	887	N	leucine-rich repeat family protein	ref|NP_190638.1|	7.30E+00
H14	C	1058	N	f-box family protein	ref|XP_002308787.1|	3.00E-05
H15	C	636	N	kinase	ref|XP_002527895.1|	1.00E-08
H17	F	810	N	Na+/H+ antiporter	gb|ABF60872.1|	5.00E-37
H17	R	821	N	putative plasmalemma Na+/H+ antiporter	emb|CAG30524.1|	1.00E-31
H18	F	806	N	alpha-1,2-fucosidase	dbj|BAF85832.1|	3.00E-01
H18	R	810	N	polyprotein	gb|ABG22120.1|	7.00E-51

H19	C	985	N	WRKY56; transcription factor	ref|NP_176583.1|	4.00E-13
H21	R	889	N	retrotransposon protein	gb|ABG22406.1|	4.60E-02
H22	C	371	C	Ycf2	gb|ADA69968.1|	4.00E-48
H23	R	766	N	Early nodulin 20 precursor	ref|XP_002515714.1|	1.90E+00
H24	C	803	N	Early nodulin 20 precursor	ref|XP_002515714.1|	1.90E+00
I04	C	1564	N	zinc finger protein, putative	ref|XP_002526412.1|	9.00E-80
I08	F	949	N	hypoxanthine phosphoribosyltransferase	gb|ACF74328.1|	0.011
I08	R	921	N	hypoxanthine-guanine phosphoribosyltransf	ref|XP_002527093.1|	1.00E-11
I12	F	805	N	maturase K	emb|CAE45223.1|	9.5
I12	R	960	C/M	box ATP-dependent RNA helicase, puta	ref|XP_002514133.1|	2.00E-41
I14	F	847	N	hypothetical protein OsJ_20744	gb|EEE65407.1|	3.30E+00
I16	R	840	N	putative plasmalemma Na+/H+ antiporter	emb|CAG30524.1|	9.00E-32
I18	F	864	N	Na+/H+ antiporter	gb|ABF60872.1|	5.00E-37
I18	R	864	N	putative plasmalemma Na+/H+ antiporter	emb|CAG30524.1|	1.00E-31
I20	F	888	N	Na+/H+ antiporter	gb|ABF60872.1|	1.00E-36
I20	R	869	N	putative plasmalemma Na+/H+ antiporter	emb|CAG30524.1|	4.00E-27
I22	F	922	N	maturase K	gb|ABI17989.1|	5.6
I22	R	869	N	hypothetical protein	ref|XP_002280260.1|	0.002
I24	C	914	N	unknown	gb|ACJ84687.1|	6.00E-26
J01	C	994	N	unnamed protein product	emb|CBI40531.1|	7.00E-04
J02	F	946	N	RNA polymerase beta subunit	gb|ADA69918.1|	9.00E-118
J02	R	858	C	ORF111 [Pinus koraiensis]	ref|YP_001152093.1|	2.00E-10
J03	C	1513	N	thiosulfate sulfurtransferase, putative, expre	gb|ABA99272.1|	1.00E-10
J05	C	1360	N	CPK related protein kinase 7	ref|XP_002310736.1|	5.70E+00
J06	R	907	N	unknown protein	ref|NP_188591.2|	8.90E-01
J09	F	892	N	putative protein	emb|CAB82966.1|	1.00E-15
J09	R	884	N	hypothetical protein	emb|CAN82333.1|	1.00E-08
J10	C	1503	N	stress associated endoplasmic reticulum p	ref|XP_002514902.1|	3.00E-16
J13	R	785	N	glycosyltransferase	dbj|BAG80547.1|	5.00E-91
J14	C	1077	N	f-box family protein	ref|XP_002308787.1|	5.00E-05
J15	C	708	C	Ycf1	gb|ADA69984.1|	1.00E-17
J16	C	245	C/M	NADH dehydrogenase subunit 7	gb|ACL26691.1|	1.00E-39
J19	F	822	N	unknown protein	ref|NP_001117613.1|	5.70E+00
J19	R	704	N	predicted protein	ref|XP_001765196.1|	6.00E-15
J21	F	876	N	ATP binding protein, putative	ref|XP_002526444.1|	2.00E-13
J21	R	902	C/M	70 kDa heat shock protein	gb|ACT65562.1|	7.00E-08

J24	F	861	N	hypothetical protein	gb|EEC78103.1|	3.40E+00
K02	C	1061	N	maturase	emb|CAQ57610.1|	8.90E-01
K04	C	1112	N	predicted protein	gb|EEH57282.1|	1.50E+00
K06	R	573	N	retrotransposon protein	gb|ABA95859.1|	9.00E-03
K08	F	937	N	bifunctional protein	gb|ACG27941.1|	2.00E-09
K08	R	827	N	hypoxanthine-guanine phosphoribosyltransf	ref|XP_002527093.1|	8.00E-05
K10	F	904	N	integrase	gb|ABG37653.1|	2.00E-14
K10	R	901	N	gag-pol polyprotein	gb|AAR13317.1|	3.00E-06
K12	C	1395	N	maturase K	gb|AAW64976.1|	3.00E-177
K14	F	832	N	binding	ref|NP_974047.1|	3.00E-26
K14	R	800	N	binding	ref|NP_974047.1|	2.00E-34
K16	C	879	N	kinase, putative	ref|XP_002518336.1|	8.00E-78
K18	F	895	N	Na+/H+ antiporter	gb|ABF60872.1|	6.00E-37
K18	R	858	N	putative plasmalemma Na+/H+ antiporter	emb|CAG30524.1|	2.00E-31
K24	C	1327	N	unnamed protein product	emb|CBI24364.1|	5.00E-05
L01	C	1212	C	ATP-dependent clp protease, putative	ref|XP_002512853.1|	3.00E-18
L02	R	582	N	unnamed protein product	emb|CBI24364.1|	6.00E-05
L03	F	363	N	hypothetical protein	emb|CAN83374.1|	5.00E-06
L03	R	798	N	hypothetical protein	emb|CAN66234.1|	4.00E-01
L04	F	918	N	unnamed protein product	emb|CBI30553.1|	2.60E+00
L04	R	962	N	prenyl-dependent CAAX protease, putative	ref|XP_002517936.1|	1.00E-04
L05	C	1361	N	CPK related protein kinase 7	ref|XP_002310736.1|	7.6
L07	F	896	C/M	translation initiation factor IF1	sp|P69041.1|	2.00E-32
L07	R	867	C	RNA polymerase alpha subunit	gb|ADA69957.1|	1.00E-128
L09	C	191	C	ndhD	emb|CAP62548.1|	2.00E-14
L12	R	897	N	ATP-dependent RNA helicase, putative	ref|XP_002510773.1|	1.00E-07
L13	C	1121	N	hypothetical protein	gb|AAT73668.1|	7.60E+00
L14	R	822	N	hypothetical protein	dbj|BAC98570.1|	5.80E+00
L16	C	1397	C	putative protein	gb|ABZ89183.1|	2.00E-57
L17	C	814	N	hypothetical protein	gb|EEE54370.1|	0.71
L18	R	853	C	NADH dehydrogenase subunit F	gb|ABR12227.1|	5.8
L19	C	1216	N	hypothetical protein	gb|ABD65088.1|	2.00E-05
L20	C	244	C/M	NADH dehydrogenase subunit 7	gb|ACL26691.1|	9.00E-41
L23	F	797	N	predicted protein	ref|XP_001772858.1|	0.9
L23	R	811	N	unknown protein	ref|NP_194735.2|	1.2
L24	C	908	N	unknown	gb|ACJ84687.1|	1.00E-25

M02	C	1322	C	photosystem II protein I	ref|YP_002720022.1|	1.00E-07
M04	F	931	N	hypothetical protein	ref|XP_002515777.1|	2.2
M04	R	897	N	prenyl-dependent CAAX protease	ref|XP_002517936.1|	1.00E-04
M06	R	923	C/M	DEAD box RNA helicase	gb|AAN74635.1|	5.00E-77
M08	C	939	N	hypothetical protein	ref|XP_002269252.1|	8.5
M10	C	636	N	kinase, putative	ref|XP_002527895.1|	2.00E-08
M14	C	1411	N	maturase K	gb|AAW64933.1|	0.00E+00
M16	C	1266	N	transporter, putative, expressed	gb|ABG22349.1|	3.00E-04
M18	C	1688	N	hypothetical protein	emb|CAN70566.1|	7.00E-73
M20	C	1200	N	hypothetical protein NitoCp007	ref|YP_398843.1|	6.00E-10
M24	F	876	N	putative protein	emb|CAB82966.1|	1.00E-15
M24	R	937	N	Retrotransposon gag protein	gb|ABD63142.1|	6.00E-44
N01	C	1310	C	photosystem II protein I	ref|YP_002720022.1|	2.00E-08
N02	C	1300	C	photosystem II protein I	gb|AAZ04048.1|	5.00E-05
N03	C	1468	N	Disease resistance response protein, puta...	ref|XP_002509891.1|	3.00E-17
N05	F	891	N	putative protein	emb|CAB82966.1|	1.00E-15
N05	R	862	N	Retrotransposon gag protein	gb|ABD63142.1|	2.00E-30
N06	F	762	N	GTP binding protein, putative	ref|XP_002514929.1|	1.00E-15
N06	R	876	N	emb1688 (embryo defective 1688); GTP binding	ref|NP_176911.2|	6.00E-22
N07	R	850	C/M	DEAD box RNA helicase	gb|AAN74635.1|	3.00E-74
N08	F	920	N	retrotransposon protein, putative, Ty3-gypsy s	gb|ABA95229.1|	1.00E-66
N08	R	903	N	putative retroelement pol polyprotein	gb|AAD20433.1|	6.00E-19
N09	F	671	N	hypothetical protein	emb|CAN70032.1|	5
N09	R	866	N	cytochrome P450 like_TBP	dbj|BAA10929.1|	5.00E-74
N10	F	927	N	integrase	gb|ABM55238.1|	2.00E-66
N11	C	1103	C	Ycf1 [Olea europaea]	gb|ADA69984.1|	6.00E-116
N12	F	412	N	conserved hypothetical protein	ref|XP_002516088.1|	1.30E+00
N15	R	913	N	glycosyltransferase	dbj|BAG80542.1|	2.00E-108
N17	F	875	C	copper ion binding protein, putative	ref|XP_002528251.1|	3.00E-20
N18	R	877	N	NADH dehydrogenase subunit F	gb|ABR12227.1|	6.5
N19	R	962	N	Retrotransposon gag protein	gb|ABD63142.1|	7.00E-05
N21	C	1282	C	hypothetical protein	emb|CAJ32479.1|	3.00E-11
N24	F	797	N	hypothetical protein	gb|EEC78103.1|	5.00E+00
N24	R	872	N	hypothetical protein	ref|YP_635779.1|	3.80E+00
O02	F	913	C	photosystem I P700 chlorophyll a apoprotein	ref|NP_084672.2|	2.00E-159
O02	R	875	C	photosystem I P700 apoprotein A1	gb|ADA69926.1|	1.00E-86

O04	F	939	N	hypothetical protein	ref|XP_002515777.1|	1.3
O04	R	858	N	integrase	gb|ABG37653.1|	5.00E-14
O06	R	903	N	gag-pol polyprotein	gb|AAR13317.1|	3.00E-06
O08	C	245	N	Os03g0152400	dbj|BAF10906.2|	5.10E+00
O10	F	888	N	putative reverse transcriptase	gb|AAC32926.1|	0.35
O12	F	905	N	binding	ref|NP_974047.1|	5.00E-27
O12	R	862	N	binding	ref|NP_974047.1|	9.00E-29
O14	C	1407	N	maturase K	gb|AAW64933.1|	0.00E+00
O16	F	828	N	galactose-binding lectin	gb|AAL09163.1|	5.1
O16	R	709	C	receptor kinase, putative	ref|XP_002535240.1|	6.00E-81
O18	F	879	N	sorbitol transporter	gb|AAT06053.2|	0.78
O18	R	845	N	predicted protein	ref|XP_001762478.1|	6.60E+00
O20	C	416	N	hypothetical protein OsI_34071	gb|EEC67196.1|	5.00E-06
O22	C	1640	N	hypothetical protein	ref|XP_002283430.1|	3.00E-17
O24	C	1299	C	photosystem II protein I	gb|AAZ04048.1|	3.00E-05
P02	F	526	N	hypothetical protein	ref|XP_002467429.1|	1.00E+00
P02	R	822	N	cytochrome P450 like_TBP	dbj|BAA10929.1|	3.00E-65
P03	C	1218	C	ATP-dependent clp protease, putative	ref|XP_002512853.1|	2.00E-18
P04	F	879	N	hypothetical protein	ref|XP_002268850.1|	4.00E-33
P05	C	1608	N	hypothetical protein	ref|XP_002267479.1|	3.00E-50
P07	C	1303	C	Ycf1	gb|ADA69984.1|	4.00E-147
P08	F	771	N	predicted protein	ref|XP_001693172.1|	5.1
P09	F	754	N	putative protein	emb|CAB82966.1|	6.00E-12
P09	R	859	N	hypothetical protein	emb|CAN82333.1|	1.00E-08
P10	R	879	N	putative retroelement pol polyprotein	gb|AAD26943.1|	0.002
P12	C	1204	N	hypothetical protein	emb|CAN72835.1|	7.00E-07
P13	F	883	N	hypothetical protein	ref|XP_002277484.1|	3.00E+00
P14	F	904	N	similar to retrotransposon pro...	ref|XP_002271693.1|	1.00E-62
P15	C	904	C	ATP-dependent protease subunit	gb|ADA69950.1|	4.00E-06
P16	C	371	C	Ycf2	gb|ADA69968.1|	5.00E-48
P17	F	848	N	hypothetical protein	|XP_002282425.1|	5.10E+00
P18	F	896	N	alpha-1,2-fucosidase	dbj|BAF85832.1|	0.054
P18	R	840	N	polyprotein	gb|ABG22120.1|	1.00E-54
P19	C	811	N	ubiquitin-protein ligase/ zinc ion binding pro	gb|ACG44198.1|	3.9
P20	F	882	N	alpha-1,2-fucosidase	dbj|BAF85832.1|	0.032
P20	R	926	N	polyprotein	gb|ABG22120.1|	1.00E-54
P23	R	853	N	retrotransposon protein, putative, Ty3-gypsy s	gb|ABA97145.1|	7.00E-34
P24	C	1104	N	gag-pol polyprotein	gb|AAO73525.1|	6.00E-25

APPENDIX B: The list of GSS assigned Genbank Accession Numbers of 585 genomic library sequences

dbGSS_Id	User_Id	GenBank_Accn
30163542	A02_Forward_804	HR308784
30163543	A04_Contig_877	HR308785
30163544	A06_Contig_1380	HR308786
30163545	A08_Forward_755	HR308787
30163546	A08_Reverse_883	HR308788
30163547	A10_Forward_832	HR308789
30163548	A10_Reverse_812	HR308790
30163549	A12_Contig_636	HR308791
30163550	A18_Reverse_878	HR308792
30163551	A20_Contig_1554	HR308793
30163552	A22_Reverse_872	HR308794
30163553	A24_Reverse_438	HR308795
30163554	B01_Forward_831	HR308796
30163555	B01_Reverse_772	HR308797
30163556	B02_Forward_596	HR308798
30163557	B02_Reverse_789	HR308799
30163558	B04_Forward_840	HR308800
30163559	B04_Reverse_568	HR308801
30163560	B05_Contig_1421	HR308802
30163561	B06_Forward_835	HR308803
30163562	B07_Contig_1536	HR308804
30163563	B09_Forward_737	HR308805
30163564	B09_Reverse_877	HR308806
30163565	B10_Forward_929	HR308807
30163566	B10_Reverse_804	HR308808
30163567	B11_Contig_1242	HR308809
30163568	B12_Forward_795	HR308810
30163569	B13_Reverse_876	HR308811
30163570	B14_Reverse_886	HR308812
30163571	B15_Reverse_917	HR308813
30163572	B16_Forward_869	HR308814
30163573	B16_Reverse_897	HR308815
30163574	B17_Contig_1001	HR308816
30163575	B18_Reverse_885	HR308817
30163576	B19_Contig_1615	HR308818
30163577	B20_Contig_371	HR308819

30163578	B21_Reverse_853	HR308820
30163579	B22_R4_913	HR308821
30163580	B24_Reverse_750	HR308822
30163581	C04_Reverse_729	HR308823
30163582	C06_Forward_880	HR308824
30163583	C06_Reverse_917	HR308825
30163584	C08_Contig_1414	HR308826
30163585	C10_Forward_794	HR308827
30163586	C10_Reverse_727	HR308828
30163587	C14_Forward_887	HR308829
30163588	C14_Reverse_898	HR308830
30163589	C16_Contig_1392	HR308831
30163590	C20_Reverse_920	HR308832
30163591	C22_Reverse_836	HR308833
30163592	C24_Forward_963	HR308834
30163593	C24_Reverse_964	HR308835
30163594	D02_Forward_891	HR308836
30163595	D02_Reverse_821	HR308837
30163596	D03_Contig_1290	HR308838
30163597	D05_Contig_1382	HR308839
30163598	D07_Contig_1480	HR308840
30163599	D08_Forward_863	HR308841
30163600	D08_Reverse_867	HR308842
30163601	D12_Forward_782	HR308843
30163602	D13_Reverse_821	HR308844
30163603	D16_Forward_759	HR308845
30163604	D16_Reverse_956	HR308846
30163605	D19_Reverse_853	HR308847
30163606	E02_Forward_775	HR308848
30163607	E02_Reverse_879	HR308849
30163608	E04_Contig_1467	HR308850
30163609	E06_Contig_1449	HR308851
30163610	E08_Forward_965	HR308852
30163611	E08_Reverse_860	HR308853
30163612	E12_Contig_480	HR308854
30163613	E16_Contig_805	HR308855
30163614	E18_Contig_1584	HR308856
30163615	E20_Forward_227	HR308857
30163616	E20_Reverse_819	HR308858
30163617	E24_Forward_860	HR308859
30163618	E24_Reverse_940	HR308860

30163619	F01_Forward_699	HR308861
30163620	F01_Reverse_867	HR308862
30163621	F03_Contig_1545	HR308863
30163622	F04_Contig_1592	HR308864
30163623	F05_Forward_825	HR308865
30163624	F05_Reverse_830	HR308866
30163625	F07_Forward_822	HR308867
30163626	F07_Reverse_835	HR308868
30163627	F08_Forward_933	HR308869
30163628	F09_Forward_704	HR308870
30163629	F09_Reverse_779	HR308871
30163630	F10_Contig_816	HR308872
30163631	F12_Forward_854	HR308873
30163632	F12_Reverse_743	HR308874
30163633	F13_Forward_822	HR308875
30163634	F13_Reverse_934	HR308876
30163635	F14_Contig_959	HR308877
30163636	F15_Forward_787	HR308878
30163637	F15_Reverse_920	HR308879
30163638	F16_Reverse_849	HR308880
30163639	F17_Forward_883	HR308881
30163640	F18_Forward_814	HR308882
30163641	F20_Reverse_574	HR308883
30163642	F21_Contig_361	HR308884
30163643	F22_Forward_879	HR308885
30163644	F22_Reverse_813	HR308886
30163645	F24_Contig_371	HR308887
30163646	G02_Contig_1298	HR308888
30163647	G06_Forward_802	HR308889
30163648	G06_Reverse_861	HR308890
30163649	G10_Forward_709	HR308891
30163650	G10_Reverse_754	HR308892
30163651	G12_Forward_932	HR308893
30163652	G12_Reverse_856	HR308894
30163653	G14_Contig_1268	HR308895
30163654	G16_Contig_806	HR308896
30163655	G18_Contig_1449	HR308897
30163656	G20_Reverse_826	HR308898
30163657	G22_Forward_876	HR308899
30163658	G22_Reverse_872	HR308900
30163659	G24_Contig_913	HR308901

30163660	H01_Reverse_822	HR308902
30163661	H03_Contig_792	HR308903
30163662	H04_Forward_926	HR308904
30163663	H04_Reverse_966	HR308905
30163664	H05_Contig_1312	HR308906
30163665	H06_Reverse_711	HR308907
30163666	H07_Forward_807	HR308908
30163667	H07_Reverse_909	HR308909
30163668	H10_Contig	HR308910
30163669	H11_Contig	HR308911
30163670	H12_Forward_943	HR308912
30163671	H12_Reverse_889	HR308913
30163672	H13_Reverse_887	HR308914
30163673	H14_Contig	HR308915
30163674	H15_Contig	HR308916
30163675	H17_Forward_810	HR308917
30163676	H17_Reverse_821	HR308918
30163677	H18_Forward_806	HR308919
30163678	H18_Reverse_810	HR308920
30163679	H19_Contig	HR308921
30163680	H20_Contig	HR308922
30163681	H21_Reverse_889	HR308923
30163682	H22_Contig	HR308924
30163683	H23_Reverse_766	HR308925
30163684	H24_Contig	HR308926
30163685	I12_Forward_805	HR308927
30163686	I12_Reverse_960	HR308928
30163687	I14_Forward_847	HR308929
30163688	I16_Reverse_840	HR308930
30163689	I18_Forward_864	HR308931
30163690	I18_Reverse_864	HR308932
30163691	I20_Forward_888	HR308933
30163692	I20_Reverse_869	HR308934
30163693	I22_Forward_922	HR308935
30163694	I22_Reverse_869	HR308936
30163695	J01_Contig	HR308937
30163696	J02_Forward_946	HR308938
30163697	J02_Reverse_858	HR308939
30163698	J03_Contig	HR308940
30163699	J05_Contig	HR308941
30163700	J06_Reverse_907	HR308942

30163701	J09_Forward_892	HR308943
30163702	J09_Reverse_884	HR308944
30163703	J10_Contig	HR308945
30163704	J13_Forward_836	HR308946
30163705	J13_Reverse_785	HR308947
30163706	J14_Contig	HR308948
30163707	J15_Contig	HR308949
30163708	J16_Contig	HR308950
30163709	J19_Forward_822	HR308951
30163710	J19_Reverse_704	HR308952
30163711	J20_Forward_911	HR308953
30163712	K02_Contig	HR308954
30163713	K04_Contig	HR308955
30163714	K06_Reverse_573	HR308956
30163715	K08_Forward_937	HR308957
30163716	K08_Reverse_827	HR308958
30163717	K10_Forward_904	HR308959
30163718	K10_Reverse_901	HR308960
30163719	K12_Contig	HR308961
30163720	K14_Forward_832	HR308962
30163721	K14_Reverse_800	HR308963
30163722	K16_Contig_879	HR308964
30163723	K18_Forward_895	HR308965
30163724	K18_Reverse_858	HR308966
30163725	K24_Contig	HR308967
30163726	L01_Contig	HR308968
30163727	L02_Reverse_582	HR308969
30163728	L03_Forward_363	HR308970
30163729	L03_Reverse_798	HR308971
30163730	L04_Forward_918	HR308972
30163731	L04_Reverse_962	HR308973
30163732	L05_Contig	HR308974
30163733	L07_Forward_896	HR308975
30163734	L07_Reverse_867	HR308976
30163735	L09_Contig	HR308977
30163736	L12_Reverse_897	HR308978
30163737	L13_Contig	HR308979
30163738	L14_Reverse_822	HR308980
30163739	L16_Contig	HR308981
30163740	L17_Contig	HR308982
30163741	L18_Reverse_853	HR308983

30163742	L19_Contig	HR308984
30163743	L20_Contig	HR308985
30163744	L23_Forward_797	HR308986
30163745	L23_Reverse_811	HR308987
30163746	L24_Contig	HR308988
30163747	M02_Contig	HR308989
30163748	M04_Forward_931	HR308990
30163749	M04_Reverse_897	HR308991
30163750	M08_Contig	HR308992
30163751	M10_Contig	HR308993
30163752	M14_Contig	HR308994
30163753	M16_Contig	HR308995
30163754	M18_Contig	HR308996
30163755	M20_Contig	HR308997
30163756	M24_Forward_876	HR308998
30163757	M24_Reverse_937	HR308999
30163758	N01_Contig	HR309000
30163759	N02_Contig	HR309001
30163760	N03_Contig	HR309002
30163761	N05_Forward_891	HR309003
30163762	N05_Reverse_862	HR309004
30163763	N06_Forward_762	HR309005
30163764	N06_Reverse_876	HR309006
30163765	N07_Reverse_850	HR309007
30163766	N08_Forward_920	HR309008
30163767	N08_Reverse_903	HR309009
30163768	N09_Forward_671	HR309010
30163769	N09_Reverse_866	HR309011
30163770	N10_Forward_927	HR309012
30163771	N11_Contig	HR309013
30163772	N12_Forward_412	HR309014
30163773	N15_Reverse_913	HR309015
30163774	N17_Forward_875	HR309016
30163775	N18_Reverse_877	HR309017
30163776	N19_Forward_538	HR309018
30163777	N19_Reverse_962	HR309019
30163778	N21_Contig	HR309020
30163779	N24_Forward_797	HR309021
30163780	N24_Reverse_872	HR309022
30163781	O02_Forward_913	HR309023
30163782	O02_Reverse_875	HR309024

30163783	O04_Forward_939	HR309025
30163784	O04_Reverse_858	HR309026
30163785	O06_Reverse_903	HR309027
30163786	O08_Contig	HR309028
30163787	O10_Forward_888	HR309029
30163788	O12_Forward_905	HR309030
30163789	O12_Reverse_862	HR309031
30163790	O14_Contig	HR309032
30163791	O16_Forward_828	HR309033
30163792	O16_Reverse_709	HR309034
30163793	O18_Forward_879	HR309035
30163794	O18_Reverse_845	HR309036
30163795	O20_Contig	HR309037
30163796	O22_Contig	HR309038
30163797	O24_Contig	HR309039
30163798	P02_Forward_526	HR309040
30163799	P02_Reverse_822	HR309041
30163800	P03_Contig	HR309042
30163801	P04_Forward_879	HR309043
30163802	P05_Contig	HR309044
30163803	P07_Contig	HR309045
30163804	P08_Forward_771	HR309046
30163805	P09_Forward_754	HR309047
30163806	P09_Reverse_859	HR309048
30163807	P10_Reverse_879	HR309049
30163808	P12_Contig	HR309050
30163809	P13_Forward_883	HR309051
30163810	P14_Forward_904	HR309052
30163811	P15_Contig	HR309053
30163812	P16_Contig	HR309054
30163813	P17_Forward_848	HR309055
30163814	P18_Forward_896	HR309056
30163815	P18_Reverse_840	HR309057
30163816	P19_Contig	HR309058
30163817	P20_Forward_882	HR309059
30163818	P20_Reverse_926	HR309060
30163819	P23_Reverse_853	HR309061
30163820	P24_Contig	HR309062