IDENTIFYING REGIONS OF CONSERVED SYNTENY BETWEEN PEA (*PISUM SPP.*), LENTIL (*LENS SPP.*), AND BEAN (*PHASEOLUS SPP.*)

by

Matthew Durwin Moffet

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Plant Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

August 2006

© COPYRIGHT

by

Matthew Durwin Moffet

2006

All Rights Reserved

APPROVAL

of a thesis submitted by

Matthew Durwin Moffet

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Division of Graduate Education.

Dr. Norman Weeden

Approved for the Department of Plant Sciences and Plant Pathology

Dr. John Sherwood

Approved for the Division of Graduate Education

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.

Matthew Durwin Moffet

August 2006

ACKNOWLEDGEMENTS

I would like to acknowledge and thank the following individuals: my past and present academic advisors Dr. Dan Bergey, Dr. John Sherwood, and Dr. Norm Weeden; my graduate committee Dr. Ed Dratz, Dr. Luther Talbert, and Dr. Norm Weeden; my supporting family; Amber Moffet (wife), Louis Moffet (father), Kim Wallinder Moffet (mother), Judith Wallinder (grandmother), and Dr. P. Gepts of U. C Davis (provided important data making my research possible).

TABLE OF CONTENTS

1. INTRODUCTION 1

Introduction 1
Background 4
Hypothesis 9

2. DESIGNING AND SCREENING UNIVERSAL GENE-SPECIFIC MARKERS 10

Introduction 10
Materials and Methods. 10
Primer Design: Gene Selection 10
Primer Design: Compiling DNA Sequence for Analysis. 12
Primer Design: Alignment of Multiple DNA Sequences 13
Primer Design: Identifying Uniqueness of Primer Sequences 15
Primer Screen: PCR Amplification 18
Primer Screen: Electrophoresis 19
Primer Screen: Restriction Digests 19
Results & Discussion 20

3. *PHEOPHORBIDE A MONOOXYGENASE* (*PAO1*) IS LOCATED ON LG VII NEAR *AMY* IN PEA AND LENTIL. 30

Introduction 30
Materials and Methods 32
Amplification of *PsPao* 32
Amplification of Other Gene Fragments 33
Sequencing 33
Populations Used for Mapping *PsPao1* 33
Detection of Polymorphism 34
Estimation of Recombination Frequencies 34
Results 34
Discussion 38
Conclusion 41

4. MAPPING GENE-SPECIFIC MARKERS IN MULTIPLE LEGUME SPECIES 43

Introduction 43
Materials and Methods 44
CAPS Screening for Polymorphism 44
Plant Genetic Material for *Pisum* Linkage Mapping 44
Plant Genetic Material for *Lens* Linkage Mapping 46
Plant Genetic Material for *Phaseolus* Linkage Mapping 46
Results 53

TABLE OF CONTENTS-CONTINUED

Segregation Data for Universal Gene-Specific Primer Pairs 54
Working Universal Gene-Specific Primer Pairs 61
Non-Working Universal Gene-Specific Primer Pairs 66
Discussion 67

5. IDENTIFICATION AND ANALYSIS OF SYNTENY 69

Introduction 69
Materials and Methods 70
Regions of Suspected Conserved Synteny 70
Regions of Conserved Synteny Between *Pisum* and *Lens*.....74
Regions of Conserved Synteny Between *Pisum* and *Lens* to *Phaseolus* 75
Results 75
Conserved Linkage Between Pea, Lentil, and Bean 75
Comparison of *Pisum* LG III and *Phaseolus* Linkage (see Figure 5.4) ...78
Comparison of *Pisum* LG V and *Phaseolus* Linkage (see Figure 5.5) 80
Comparison of *Pisum* LG VII and *Phaseolus* Linkage (see Figure 5.6)...82
Discussion 85

REFERENCES CITED 88

APPENDICES 95

APPENDIX A: LIST OF ALL PRIMER PAIRS SCREENED BY PCR 96
APPENDIX B: PERMISSION LETTER 99

LIST OF TABLES

Table Page

2.1 List of Restriction Enzymes 20

2.2 List of Primer Sequences for Universal Markers 24

2.3 Designed Universal Primer Pairs: Monomorphic 27
Table describes the genes that resulted in monomorphic fragments in bean and were not mapped.

2.4 Designed Universal Primer Pairs: Polymorphic 28

2.5 Designed Universal Primer Pairs: Only Polymorphic in Bean 29

3.1 Primer Combinations- Expected vs. Actual Fragment Size in 2% SB agarose gels 35

3.2 Joint segregation analysis between the loci *Amy* and *Pao* in four RIL populations 37

Pair-wise Comparison of Marker Segregation in *Phaseolus* (BAT93 x JaloEEP558) 54

4.1 *Enolase* Linkage Segregation Data in *Phaseolus vulgaris* 54

4.2 *Paal* Linkage Segregation Data in *Phaseolus vulgaris* 55

4.3 *TufM* Linkage Segregation Data in *Phaseolus vulgaris* 55

4.4 *Paal, Enolase,* and *TufM* Linkage Segregation Data in *Phaseolus vulgaris* 55

4.5 *RCA* Linkage Segregation Data in *Phaseolus vulgaris* 56

4.6 *RCPME* Linkage Segregation Data in *Phaseolus vulgaris* 56

4.7 *GRF3* Linkage Segregation Data in *Phaseolus vulgaris* 56

4.8 *Apyrase* Linkage Segregation Data in *Phaseolus vulgaris* 57

4.9 *Tkp* Linkage Segregation Data in *Phaseolus vulgaris* 57

4.10 *UNK21* and *ALDH7* Linkage Segregation Data in *Phaseolus vulgaris* 58

LIST OF TABLES-CONTINUED

Table Page

4.11 *FK* Linkage Segregation Data in *Phaseolus vulgaris* 58

4.12 *GPT* Linkage Segregation Data in *Phaseolus vulgaris* 58

4.13 *Pao* Linkage Segregation Data in *Phaseolus vulgaris*.... 59

4.14 *CRS* Linkage Segregation Data in *Phaseolus vulgaris* 59

4.15 *TubA1* Linkage Segregation Data in *Phaseolus vulgaris* 59

4.16 *Aat* Linkage Segregation Data in *Phaseolus vulgaris* 60

4.17 *PGIp* Linkage Segregation Data in *Phaseolus vulgaris* 60

4.18 *bFRUCT* Linkage Segregation Data in *Phaseolus vulgaris* 60

4.19 *Cipor* Linkage Segregation Data in *Phaseolus vulgaris* 61

4.20 *Apx1* Linkage Segregation Data in *Phaseolus vulgaris* 61

4.21 Mapped Universal Primers Pairs Summary Table 62

LIST OF FIGURES

Figure Page

1.1 Detailed phylogeny of legumes based on *trnK/matK* (Hu *et al.,* 2000) 7

3.1 Basic pathway of chlorophyll degradation 31
(**Substrates --> (**abbreviations: RCC, red chlorophyll catabolite; pFCC, primary fluorescent chlorophyll catabolite; NCC, non-fluorescent chlorophyll catabolite)
(**Enzymes --> (**abbreviations: CHL, chlorophyllase; PAO, pheophorbide *a* oxygenase; RCCR, red chlorophyll catabolite reductase; *, *[↓pH]* and multiple enzymes)

3.2 End sequences of the Pao-F2/R3 fragment showing >94% DNA sequence identity between WL808 and MN313) 35
PsPao-F2 DNA sequence alignment of *P. abyssincum* (WL808) and *P. sativum* (MN313). Identities = 379/401 (94.5%).
PsPaoWL808-F2 Obtained DNA sequence
PsPaoWL808-R3 Reverse complement of obtained DNA sequence

3.3 2% Agarose Gel of *Pao* 37
(A)The PCR amplification of JI1794(Lane 1), SLOW(Lane 2), JI1794 x Slow progeny(Lane 3-6) showing 2-JI1794 and 2-Slow genotypes.(B)Abyss. x Sparkle w/CfoI, 2-Abyss., 4-Sparkle, and 1 heterozygous genotype.

3.4 *PsPao* gene model based on the gene splicing of *AtPao* 37
Primer and obtained sequence locations within the *PsPao* gene

4.1 *Pisum* Linkage Map (M.S.U linkage map of Ji1794 x Slow) 45

4.2 Integrated Linkage Map of LG 1 47

4.3 Integrated Linkage Map of LG 2 48

4.4 Integrated Linkage Map of LG 3 48

4.5 Integrated Linkage Map of LG 4 49

4.6 Integrated Linkage Map of LG 5 49

4.7 Integrated Linkage Map of LG 6 50

4.8 Integrated Linkage Map of LG 7 50

LIST OF FIGURES-CONTINUED

Figure Page

4.9 Integrated Linkage Map of LG 8 51

4.10 Integrated Linkage Map of LG 9 51

4.11 Integrated Linkage Map of LG 10 52

4.12 Integrated Linkage Map of LG 11 52

4.13 *Phaseolus* Linkage Map (Davis Bat x Jalo linkage map with M.S.U markers) ...66

5.1 The *Pisum* LG7 lower arm region 70
Identification of orthologous loci in pea and thale cress by NCBI Mapviewer, genes showing conserved synteny between pea and lentil were queried to find proper genes (some nomenclature differences).

5.2 Microsyntenic Comparison by BLAST: The RST1 Gene is Conserved 72

5.3 Consensus Linkage Map of *Pisum sativum* and *Lens* 74

5.4 Comparison of *Pisum* Linkage Group III to Bean Linkage Groups 79

5.5 Comparison of *Pisum* Linkage Group V to Bean Linkage Groups 81

5.6 Comparison of *Pisum* Linkage Group VII to Bean Linkage Groups 82

5.7 Comparison of Conserved Synteny between *Arabidopsis, Pisum, Lens, Medicago* and *Phaseolus* 86
The bars and relative distances and/or orientation from top to bottom is not drawn to scale. The figure is an illustration depicting relative gene order.

GLOSSARY AND NOMENCLATURE

Ancestry- Individuals sharing a common ancestor at any point of time.

bp- Base Pair. The number of nucleotide monomers in a DNA sequence.

cM- Centi Morgan. Defined as 1cM equals 1 recombination per 100 F2 individuals or 1 recombination per 50 recombinant inbred lines.

CAPS- Cleaved Amplified Polymorphic Sequence. The PCR product of two or more individuals is cut with a restriction endonuclease exposing a difference in DNA sequence between individuals (polymorphism). As DNA sequences diverge through time the recognition sites for restriction endonucleases are either lost or gained creating polymorphisms.

Conservation of synteny- Regions of two genomes from different species that show similar synteny.

EPIC-PCR- Exon-primed intron-crossing PCR. This is one acronym that closely describes the method of PCR used in this genomic study. DNA primers are designed to the most highly conserved DNA sequence in a gene, usually the exon ends near highly divergent introns. Size of designed DNA fragment should be inversely proportional to the amount of polymorphism present in the populations being tested.

Homologous- DNA sequence diverged from a common ancestor that still maintains enough similarity to be identified as similar. Paralogous and orthologous are delineations of homology.

Kb- Kilobase. 1,000 monomer nucleotide units, see bp.

Linkage- Using a segregating population to identify a preponderance of parental genotypes between two or more genetic markers. Used in direct reference of data.

Mb- Megabase. 1 million monomer nucleotide units, see bp.

Macrosynteny- Generally means linkage of two different DNA markers (loci) located in the same proximity or location on the same chromosome in two species. Used within this thesis as markers found on the same linkage groups/chromosomes, either rearranged or not, usually meaning 50cM / Mb or less.

Microsynteny- Meaning two different DNA markers are located in close proximity on a chromosome in two species. Used within this thesis as markers found on the same BAC, either rearranged or not, usually meaning 0.5cM / Mb or less.

GLOSSARY AND NOMENCLATURE-CONTINUED

Orthologous- DNA sequence diverged from a common ancestor that still maintains enough similarity to perform the same type of function in two distinct species. Most likely diverged from a speciation event.

Paralogous- DNA sequence diverged from a common ancestor that still maintains enough similarity to be identified as close in DNA sequence, but does not perform the same function. Most likely diverged from a duplication event.

Synteny- Loci located on the same linkage group, implying chromosome. Also means marker co-linearity. Syntenic genes stated as "genes thought to reside on the same chromosome because of their behavior in somatic cell hybridization experiments." (King *et al.*, 1997)

ABSTRACT

The identification of conserved synteny in legumes can facilitate many different types of gene discovery. Techniques like marker assisted selection and the candidate gene approach can benefit greatly by identifying conserved synteny and genes located within those regions. Both *Pisum* and *Phaseolus* have detailed linkage maps, but a limited number of markers have been located in both species. In the present study I mapped 21 genes in *Phaseolus vulgaris*, 16 of which had already been located on the *Lens* and *Pisum sativum* linkage maps, the markers were used to look for conserved synteny between *Pisum, Lens* and *Phaseolus*. In particular, I was able to target marker/gene-rich regions of pea linkage groups V and VII, as well as pea linkage group III, with *Pisum* STS markers and universally designed gene-specific markers already located on the pea linkage map. About 50% of the tested genes amplified an appropriate sized fragment in *Phaseolus*, but less than 40 % of the gene-specific markers showed polymorphism by cleaved amplified polymorphic sequence (CAPS) analysis in bean. The data reveals little evidence for extensive gene order conservation, and even some closely linked (<5cM) loci in *Pisum* are not linked in *Phaseolus vulgaris.* The only example of conserved synteny was approximately 15cM on pea and lentil linkage group V and bean linkage group 1. *Paal*, *Enolase*, and *TufM* were first identified in the syntenic area and allowed identification of *fin/det*, one of several *TFL* genes already mapped in bean, as another orthologous loci between pea and bean. Finding conservation of synteny with *Paal*, identified *Paal3* and *TFL1* genes as linked loci in *Arabidopsis thaliana* on linkage group 5. The pea locus *Paal1,2* is then speculated to be a tandem duplication of a *Paal3* homolog in a ancient common ancestor and probably occurring after the speciation of *Pisum*.

CHAPTER 1

INTRODUCTION

Introduction

When it exists, the conservation of gene order among species can be a valuable asset for gene mapping and gene discovery within taxonomic groups. Assignment of homologous markers into linkage maps will allow comparisons of gene order between species. Using linkage data to extrapolate conserved synteny among several legume species, especially *Medicago truncatula* (barrel medic) and the three species we are currently working with (pea, lentil, bean) will expedite the field of genomics in the Fabaceae plant family. Known as comparative genomics, the approach utilizes known gene sequences to locate specific genes on the linkage maps of multiple species. Described conserved synteny between *Medicago, Pisum*, *Lens, and Phaseolus* comparative genomic analyses will aid in a more complete comparative analysis of legumes.

The study presented here describes the development of universal DNA markers to be used in a comparison of gene order between pea and bean. The universal primer pairs were designed to be complementary to highly conserved DNA sequences in multiple species for mapping genes. Mapping of the same genes in pea, lentil and common bean indicate the conservation of the primer sequence and should have general utility for comparative mapping in Fabaceae and other plant families.

Several similar approaches have been used in pea and other crops. Gilpin *et al.* (1997) sequenced cDNAs obtained from pea etiolated seedlings and designed primers to those sequences. The authors mapped a number of these on the pea consensus map or linkage maps developed in other crosses using a CAPS approach. Schneider *et al.* (1999) obtained gene sequences from GenBank (www.ncbi.nlm.nih.gov) to develop primers for and map these genes in *Beta vulgaris*, again primarily by means of CAPS analysis. In these two cases, the size of the PCR product was usually under 1000 bp, requiring a high degree of DNA polymorphism between the two parents in order to obtain a CAPS polymorphism within the two PCR products. Intron targeting by PCR has been a focus of multiple gene mapping labs to target conserved regions of DNA (genes) by polymerase chain reaction (PCR), while still finding sufficient polymorphism in the PCR product for mapping (intron divergence> exon divergence). Another previous research project on the sequence divergence of >300 human spliceosome genes strengthens the belief that intron divergence is greater than exon divergence. Different PCR products that contained more introns were found to be more polymorphic than genes with more exons (Stephens and Schneider, 1992). Exon-primed intron-crossing PCR (EPIC-PCR) has been described in rice to develop genome-wide markers and is similar to the method used herein and has become the current method for isolating polymorphism in genes by targeted primer design (Palumbi *et al.*, 1995; Wang *et al.*, 2005).

Other research groups have used restriction fragment length polymorphisms (RFLP) probes or cDNA derived probes, but some probes hybridize to multiple fragments slightly limiting the number of useful, cross-species markers (Ahn and Tanksley, 1992; Tanksley *et al.,* 1992; Devos *et al.*, 1995). Microsatellites and many

genomic DNA probes are non-coding DNA, also causing hybridization of repetitive DNA and random DNA sequence. Only rarely could non-coding DNA be homologous in distant comparative analyses, except in small introns. The usefulness of microarray technology has been best suited to expression analyses then comparative genomics. ESTs (expressed sequenced tags) generate a more stringent comparative analysis than genomic sequence. Microsatellites can produce comparative markers in close genetic comparisons, generally meaning within the genus, where previous comparisons of *Triticum timopheevii* and *T. aestivum* showed 78% of the markers amplified from the A genome in each and <50% amplified in the secondary genomes (Salina *et al.*, 2005). Previous usage of microsatellites in comparing *Medicago* to pea shows that less than one-third will amplify in PCR (Wang *et al.*, 2004). Similarly, within a tribe, microsatellites designed for pea genomic sequence usually do not amplify the homologue in lentil. (Weeden, unpublished).

Previous mapping projects showed high levels of macrosynteny between pea and lentil (Weeden *et al.*, 1992; Walling and Weeden, 2003; Walling, 2003; Moffet and Weeden, 2005; Weeden *et al.*, 2005). The usage of universal primer pairs for specific genes will be used to expand the comparative study of pea and lentil to include bean. The developed universal markers used in this described study have shown successful mapping of genes in pea, lentil and common bean by gene-specific intron-targeted PCR and incorporates gene-specific intron-targeted PCR markers used in the pea-lentil comparison described above and continuing in our lab (Walling, 2003).

Here we present primer sequences and relevant data for 16 genes in legumes that have been successfully mapped in previously described pea, lentil and common bean

linkage maps using restriction endonucleases for identifying polymorphism within the parents and progeny of certain established mapping populations. The background that follows will support the findings of this research and also show discrepancies on the amount of conserved synteny in the legumes.

Background

The usage of genomics in comparing species has grown with the expansion of DNA sequencing. Comparative genomic studies have been performed in the major plant families; Solanaceae, Brassicaceae, Poaceae and Fabaceae. Published research indicates that many plant species share marker co-linearity/synteny; where, gene order on the chromosome is conserved between two species. Comparative genomic studies on conserved synteny use DNA marker information obtained in previous research to find homologous markers between different species.

Extrapolation of conserved synteny, synteny among multiple species from different genera, has been done on the more intensively studied plant genomes. Comparative genomic studies in the family Solanaceae, using tomato, potato, and pepper as examples, show high amounts of conserved synteny. Tomato and potato show several inversions, but pepper has an estimated six times the chromosomal rearrangements of tomato or potato mainly attributable to inversions (Prince *et al.*, 1993). Based on previous taxonomic work, it was anticipated and found that within the Solanaceae, pepper is less syntenic to tomato than is potato (Livingstone *et al.*, 1999). Marker design in the major Solanaceae comparisons was based on RFLP sequence used in creation of the tomato-potato comparative linkage map (Tanksley *et al.*, 1992). The choosen markers

were found to be very useful, by hybridizing to the same number of bands and/or mapping to the same loci in many cases for potato and pepper.

The comparative genomic studies in the family Brassicaceae have been comparisons of *Brassica napus* (rapeseed), and *Arabidopsis thaliana* (thale cress) the only plant genome to be nearly completed in sequenced (Lan *et al.*, 2000). The comparison within the Brassiceae tribe revealed extremely high conservation of synteny with only a few small inversions (Lydiate *et al.*, 1993). Indications of synteny are found in many locations throughout the plant family with a significantly higher occurrence of breakpoints occurring in the more evolutionary distant genomic comparisons (Lydiate *et al.*, 1993; Prince *et al.*, 1995). Marker design was based on genomic sequence from thale cress and rice to make comparisons across the family. The cDNA-derived RFLP markers were found to be very useful, considering many hybridized to the same number of bands (within the Brassicaceae family) and mapped to the expected loci in other species.

The conservation of synteny among the cereals or grain crops has been of economic interest, so there have been comparisons between some distantly related species in the family Poaceae. The realized value to breeders and geneticists of a grasses genomic database has prompted the formation of Gramene (http://www.gramene.org) to collate genomic information on grasses (Jaiswal *et al.*, 2006). The small genome size of rice resulted in this species being chosen as the model grass species genome. Comparative analyses between the grass species will centralize to the rice genome (Gale *et al.*, 1996; Moore *et al.*, 1995; Devos *et al.*, 2000). *Oryza sativa* (rice) has even been shown to have conservation of synteny, to some extent with *Arabidopsis thaliana* annotated coding sequences (CDS) used to identify 60 syntenic regions involving 4–22

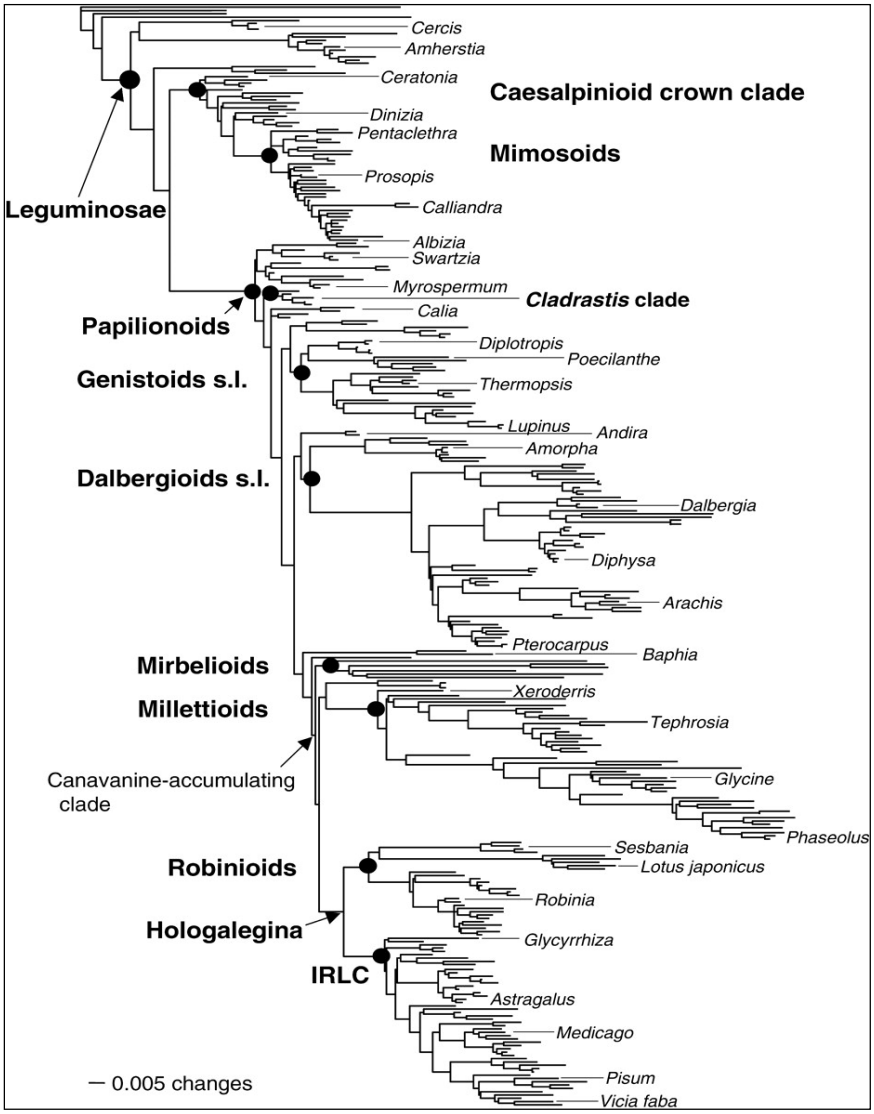
rice orthologs covering ≤3.2 cM (Salse *et al.*, 2002). S. Ahn and S. Tanksley (1992) found 32 conserved linkage segments, containing two or more loci that could be identified between the rice and maize genomes, from 5 cM to 85 cM in length. Another comparison of rice and *Brachypodium distachyon used BAC landing and fiber FISH to find synteny between each of the 10 linkage groups and regions from rice (R. Hasterok et al., 2006).* A comparison of rye and wheat showed as many as 13 chromosomal rearrangements (Devos *et al.*, 1992). Hulbert *et al.* (1990) identified regions of conserved synteny in maize and sorghum, while other research groups have added to the comparison (Whitkus *et al.*, 1992; Berhan *et al.*, 1993; Binelli *et al.*, 1993; Chittenden *et al.*, 1994; Pereira *et al.*, 1994). Recently sorghum has been compared to rice and found that (Kim *et al.*, 2005). The evaluation of genome size evolution in the genus *Sorghum* has proposed no model for the expansion of genomes in the grass family, meaning conserved synteny may be present at higher levels than in other plant families (Price *et al.*, 2005). Comparative genomic studies in the family Poaceae (ex. rice, maize, sorghum, barley, and millet) have shown many areas of conserved synteny, but comparisons of wheat and rye has revealed over a dozen chromosomal rearrangements, comparable to the amount of divergence in the Brassicaceae, and more than what has been found in the other plant families (Paterson *et al.*, 1996; Paterson *et al.*, 2005).

A Bayesian tree of 330 legume matK sequences based on the strict consensus tree of 12 equally parsimonious trees from the *trnK*/*matK* data set making a very diverse comparison of legume taxonomy (Hu *et al.*, 2000). "Sequence divergence values …....vary from 0.008 substitution per site (between two *Wisteria* spp.) to 0.184 substitution per site (between pea and cowpea)." Using estimations of genome similarity

can be useful, but deceptive, only known gene order will have direct effects on plant improvement by comparative genomics. Figure 1.1 details phylogeny in the legumes as shown by previous research (Hu *et al.*, 2000).

Initial comparative genomic studies in the family Fabaceae compared the linkage maps of important diploid crops (pea, lentil, bean, chickpea, beans). Figure 1.1 is intended to give a general view of the Fabaceae plant family and show the relative evolutionary distance between pea, lentil, and bean.

Figure 1.1 Detailed phylogeny of legumes based on *trnK*/*matK* (Hu *et al.*, 2000).



More recently legume genomics has been focused around *Medicago truncatula* and *Lotus japonicus*, because they are the sequenced model plant(s) in Fabaceae. Kalo *et al.* (2004) made a comparison of *Medicago sativa* and pea finding instances where synteny was conserved over large regions of the linkage groups. Weeden *et al.* (1992) used RFLP probes and isozymes to show synteny between pea and lentil, but found no synteny to bean (Personal communication: N.F Weeden). With limited DNA sequence, consensus primer design was shown to aid comparative analyses between pea and lentil by amplifying genes that had not been sequenced from that species (Weeden *et al.*, 1993).

The comparison of bean to cowpea has shown higher levels of synteny conservation than other comparisons, as would be expected for sister genera (Vallejos *et al.*, 1992; Boutin *et al.*, 1995). Boutin *et al.* (1995) found an average of 37cM for areas showing conservation of synteny, including one region approaching 100cM. The level of genome conservation between *Phaseolus* and *Vigna* is comparable to the observed synteny within *Vigna* (*V. radiata* and *V. unguiculata*), while a substantial number of linkage blocks of up to 50 cM in length were conserved among all three genera when *Glycine* was included using 173 RFLP probes (Menancio-Hautea *et al.*, 1993; Boutin *et al.*, 1994).

Choi *et al.* (2004) made a comparative map of six legume species, finding many markers associated to similar chromosomes, but many linkage groups have only two comparative markers making it difficult to say we are observing conserved synteny. A major lack of marker density can be a limiting factor in proper identification of synteny, along with a lack in marker density. One region with many markers was pea linkage group 3 and *Medicago* linkage group 3 showing >50cM macrosynteny, but when

compared to *Vigna* only several close markers showed conserved synteny. Zhu *et al.* (2003) was able to show conservation of synteny between *Medicago truncatula* and *Arabidopsis thaliana* showing small continuous regions and gene duplications that confounded the problem of identifying orthologs and regions truly syntenic. Zhu *et al.* (2005) used EST sequences to design primers for eight legumes and found many of the same linkage group correlations between species as Choi *et al.* (2004) found. A confirmation between studies that soybean compared to common bean is more highly rearranged than common bean to cowpea as expected. This break in synteny suggests any other species from a different tribe, when compared to *Phaseoleae*, will definitely show major rearrangements.

More information will eventually lead to an answer for the question about conservation of synteny between eudicot and monocot plant species, and to some extent we can already answer that question by the end of this thesis (Devos *et al.*, 1999).

Hypothesis

Previous literature states close relationships of gene order between *Medicago* and *Pisum* and this distantly conserved synteny has then been extrapolated to many of the species in the Fabaceae family. Due to the large size and considerable diversity of the Fabaceae, highly conserved macrosynteny will be primarily found within closely related tribes. The pea and lentil genomes have significant conserved macrosynteny, and if bean does not show conserved macrosynteny, then we can expect lineages within Papilionoideae to show macrosynteny only within closely similar ancestries.

CHAPTER 2

DESIGNING AND SCREENING UNIVERSAL GENE-SPECIFIC MARKERS

Introduction

Described in this chapter is the design of homologous DNA markers. The intent of the research is to map these markers in pea, lentil, and bean. The research goal is to increase the number of DNA markers that can be mapped in multiple legume species. Universal primer design was explored to increase the probability that a primer pair will amplify the homologous fragment in another species, such as that found in Chapter 3 where the *Pao* gene was used as an example in universal primer design.

Materials and Methods

Primer Design: Gene Selection

Initiation of a genomic study requires generating as much data as possible to increase the probability of finding valuable comparative points. Decisions must be made on the method of data acquisition to get the highest quality data in the most efficient manner. The *Arabidopsis* genome is estimated at ~25, 000 genes, giving an over abundance of possible loci to target, thus creating a need for an efficient method in data mining for gene selection as reported by the *Arabidopsis* Genome Initiative (A.G.I., 2000). When targeting genes we must consider how many genes have available DNA sequence from multiple species. The probability of acquiring possible loci is greatly diminished when only using highly sequenced genes. Legume species have limited DNA

sequence data due to delayed funding compared to other plant species. Currently several legume species are finally developing into substantial model plant genomes. Many of the genes we screened with BLAST had a high similarity match in *Medicago truncatula* and frequently matched a *Lotus japonicus* sequence.

The choice of certain genes for this type of homologous gene-targeted PCR is based on available gene sequences. Criteria used in gene selection included: genes sequenced in many species, genes shown to be single or low copy in many genomes, and genes involved in well known biochemical pathways. Universal priming sites were identified at the ends of exons, and the fragment was selected to span the maximum distance of 1 Kb to 3 Kb of DNA sequence. An optimum location for priming sites would be located at the end of the translated region (expected to be the most highly conserved and gene-specific sequences).

Selection of only the most conserved genes between species helped to insure the targeted gene amplified. The selection implied specific gene products are necessary for life and as a consequence the DNA sequence may be more conserved than in genes considered less vital. A later focus was on PCR fragments larger than 1500 bp to increase the efficiency of the possible genes in the screening process. Genes involved in concurrent research interests were used as a starting point, due to the amount of sequence knowledge on specific genes (*Pao, Paal, Hsp70*). The *Pao* gene is used as an example of universal primer design and is described in detail as Chapter 3. Genes involved in gene families where DNA sequence is highly similar were mostly excluded, to prevent identifying non-orthologous genes. The only member of a multiple gene family that was investigated by this research was *ALDH7,* its PCR reaction produced two DNA

fragments of different restriction enzyme recognition sites and same length, so only one polymorphism was chosen to map in bean.

A majority of the genes screened in this study could be considered vital genes, defined here as genes necessary for life. We can expect DNA differences in vital genes between species, but also a higher level of sequence conservation within species. As time and natural selection persists vital genes are more likely to be monomorphic, due to the need for a functional gene product for support of life. The goal in gene selection and primer design was to compromise between the level of DNA conservation and the amount of DNA polymorphism.

Primer Design: Compiling DNA Sequence for Analysis

Primer sequences were generally designed from genomic and mRNA sequences available in the GenBank database (http://www. ncbi.nlm.nih.gov), TIGR (http://www.tigr.org) and PlantGDB (http://www.plantgdb.org). When a specific gene had been chosen the gene name was queried to find an initial DNA sequence through NCBI Entrez (http://ncbi.nlm.nih.gov/entrez) or genbank acession numbers in the interesting journal article detailing the gene of interest (GOI). DNA and RNA sequences where found by BLASTn in Biology Workbench and compiled into the Biology Workbench (http://www.sdsc.org) for sequence alignment. In general, *Medicago* and *Arabidopsis* sequences were used to design all the universal primers, but determined the conserved nature of this sequence by comparing the entire gene sequence with those of homologous genes in other legumes and other available plant sequences. In most cases, the size of the PCR fragment expected in pea could be predicted from the *Arabidopsis* or

Medicago genomic sequence, some amplified fragments had larger introns than expected. In the case of *Pao* the same region of the gene in pea and lentil have undergone intron enlargement giving two alleles found in our PCR and CAPS screening, CAPS also identified other alleles. Priming sites were then designed to generate a fragment >1000bp and span multiple introns. The major experimental advantage for amplifying a fragment larger than 1000 bp is to increase the probability of finding polymorphism. In several cases we were able to design primers that gave much larger PCR fragments using less conserved regions of the genes, but such efforts to increase the chance of polymorphism did not prove successful as shown by the Results and Discussion section in this chapter.

Many genes were eliminated after the compiling of nucleotide sequences in the Biology Workbench, due to the lack of conserved DNA sequence (not highly conserved genes) or lack of sufficient sequence to determine what areas are truly conserved (not highly characterized). Estimation of nucleotide sequence by protein sequence or known motifs can amplify known genes, but success and validity without sequencing is highly uncertain. Using the all the available DNA sequences will ensure PCR amplification in multiple species as shown in the results.

Primer Design: Alignment of Multiple DNA Sequences

Alignment of the multiple BLAST acquired DNA sequences was the most crucial step of success in homologous PCR amplification. When available DNA and RNA sequences from *Medicago* and *Arabidopsis*, plus RNA sequence from *Lotus* and tomato could be found, PCR amplification success was markedly increased.

Several genes were eliminated after the alignment of nucleotide sequences in the Biology Workbench, due to the lack of conserved DNA sequence. This further prompted the elimination of several groups of primers that at first were considered as potential sources of markers; UNISTS markers and *Phaseolus* EST sequences. All primer sequences were screened on the database containing the *Phaseolus*, and many other plants, EST sequences known as PlantGDB (http://www.plantgdb.org/PlantGDB-cgi/blast/PlantGDBblast) by BLAST and no sequences were found using very lenient parameters. UNISTS markers (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unists) are being used to integrate STS markers into genetic maps by sequencing the amplified fragments to find polymorphism. The UNISTS database was first queried for markers containing pea, bean, and *Medicago* sequences. UNISTS markers average <500bp and it was first believed that larger DNA fragments for universal markers could be developed and amplified by finding over-lapping regions with the UNISTS sequence and genomic DNA sequence or RNA sequence.

Nucleotide sequences compiled in the Biology Workbench and aligned by ClustalW used anywhere from two nucleotide sequences to ten or more nucleotide sequences. In certain cases different nucleotide sequences were used in designing different primer pairs to test in the different parental backgrounds. In most cases the entire portion of the DNA sequence could not be used in the alignment, because it would skew and misalign the multi-species alignment. Every combination of the multiple sequences would be aligned in ClustalW to determine the most highly conserved sequences and regions of sequence that may need to be trimmed only to the area conserved and of interest for primer design. The ClustalW alignment was then imported

into the Texshade module (TeXshade version 1.4. http://homepages.uni-tuebingen.de/beitz/txe.html) for easy viewing. Texshade shows sequence name, nucleotide numbering, and colored shading for each nucleotide sequence. Many gene sequences were eliminated due to a prediction of a small-amplified fragment size (<800 bp). Once two conserved regions had been identified, and estimated to be at least 800 bp apart using *Medicago* or *Arabidopsis* genomic DNA sequence, then an exact primer location was determined by identifying the uniqueness of the DNA sequence found within the two conserved regions.

Primer Design: Identifying Uniqueness of Primer Sequences

There are several major considerations important in the design of universal primers from a multi-species alignment; exon-intron junctions, locations of motifs, and uniqueness of the sequence. The method described is a modified version of EPIC-PCR, because here we try to target and identify the most highly conserved exon next to the least conserved intron, thus maximizing the probability of PCR success in multiple species and find a CAPS polymorphism.

Once an alignment is created, the naked eye can distinguish areas highly conserved with the same or similar gene ontological categories, by using BLAST with the Spidey program at NCBI (http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/) to identify the transcript splice sites/exon-intron junctions we will know where a primer can and can't be designed. The priming sequence needs to be located within an exon to limit sequence divergence between the 5' and 3' ends of the primer (meaning the whole primer sequence is conserved and not just the ends of the primer). Additionally, primer locations

near the exon-intron junction are not used if one or more species have a shifted location of the intron/exon junction. The regions of DNA sequence found within introns and the ends of exons with changes in the location of exon-intron junctions signify a higher rate of sequence divergence. Natural selection is based on survival of the fittest, so once an individual has an evolved form of a gene changes within the coding sequence are more likely to create a less fit version of the gene. Therefore, the consequence of changing coding sequence (exons), will be limited or more conserved then within introns, which are non-coding regions of DNA. Primer design is usually close to a highly conserved intron-exon junction making it important to verify the 3' end will be as close as possible to the actual DNA sequence for high fidelity during PCR. Design was the most crucial aspect of the universal primer design process in preventing PCR mis-priming.

Locations of motifs can also dictate the identification of highly conserved regions of DNA. Identifying genes containing the same motif are not useful for identifying unique homologous DNA sequences within unique single-copy genes, unless using them to avoid designing primers that may hybridize to that particular sequence. By targeting a motif in primer design we may have cross-hybridization with any other genes also having the same motif, which could be hundreds or thousands of loci being amplified. To prevent mis-priming, motifs were located inside the aligned sequence and avoided as areas for universal primer design; except their proximity that can be conserved like exon-intron junctions. Commonly, unique DNA sequences can be found adjacent to important motifs controlling gene function, so these can be areas of conserved DNA sequence that are gene-specific.

Regions identified as being the most useful for universal primer design, unique and highly conserved nucleotide sequence, appear to be the DNA sequences adjacent to a major motif, adjacent to the exon-intron junctions, and the location interior of the 5' end. Mutations earlier in gene coding sequences are more likely to have an effect on the overall function of the gene product, so there is a tendency to see more conservation of nucleotide sequence at the beginning of a gene than end. When coding sequence is changed near the 5'end a frame-shift may cause the protein to be completely un-functional. Changes near the 5' end will affect initial protein folding after translation and can change the localization signal at the beginning of the 5' end, which is so complex we can still not predict every proteins signal.

Verification that the primer is unique proved necessary to prevent cross-hybridization of non-complementary DNA during the annealing set of PCR. To accomplish an effective hybridization test of the primer the BLAST module was performed on the primer sequences with less stringent conditions then pre-set BLASTn parameters. Conducting BLASTn with the pre-set parameters on a 20-mer, only 19 nucleotide matches out of 20 total will not be filtered out or 16 consecutive nucleotide matches out of 20 will not be filtered out. To eliminate all possibilities of non-complementary primers that might hinder PCR productivity requires checking hairpin structures, self-annealing or the other primer (primer-dimers), and primers that contain high similarity to other genes at the important 3' end. Hairpin structures and primer-dimers were the only deviations allowed, because only the most severe and not allowed hybridization will cause amplification of a non-complementary DNA sequence. If ~8 nucleotides at the 3' end of a primer anneal to DNA then PCR is very likely to occur, so

nucleotide matches with non-complementary DNA near the 3' end are important to avoid in primer design. To obtain sequences fitting a more appropriate level of stringency in the BLASTn parameters two alterations were chosen for search criteria, the expected value and word count value. The expected value for a 20-mer should be changed from the preset value of 1 x 10^{-4} to 10 to allow BLAST hits of overall less nucleotide matches to not be filtered out, since more sequences will be shown you may also need to increase the number of displayed BLAST hits to find non-complementary sequences. The word count should be lowered from 11 to 7, to identify DNA sequences that match 7 bp in a row versus 11, especially the last 8-10 bp of the primer ends should be avoided.

Primer Screen: PCR Amplification

PCR amplification of the universally designed primers was first done by a general touchdown PCR (TD-PCR), which, worked for the vast majority of the primers with *Taq* polymerase (Promega Corp.; Madison, WI) in the screening of pea, lentil, and bean parents. The general TD-PCR was 10 cycles with an annealing temperature of 60ºC and decreasing 0.5 ºC each cycle, initial denaturation was 2 minutes and the subsequent denaturing was 20 seconds long, extension was 2 minutes, with an additional 40 cycles at 56 ºC annealing temperature to end the PCR protocol robustly.

PCR fragments below 2000 bp in size were amplified using *Taq* polymerase (Promega). For larger fragments or multiple PCR fragments from mis-priming Platinum *Taq* polymerase (Invitrogen Corp.; Carlsbad, CA) was used to increase the fidelity. PCR reactions consisted of 25ul reactions with 3.5ul being used for an agarose gel to check PCR success of different primer pair combinations.

Primer Screen: Electrophoresis

All polymorphisms were resolved on 2% agarose gels with ethidium bromide (1.5ul/100ml buffer) stain in the gels before pouring. We successfully used 1x sodium-borate (SB) agarose gels @ 100-120v for all gel electrophoresis performed. To make 4L of 5x SB 8gr. of NaOH crystals were dissolved in 4L ddH_2O, then titrated to pH 8.5 with boric acid.

Primer Screen: Restriction Digests

When a specific primer pair amplified an appropriately sized product in bean, comparable to pea, lentil and expected size based on the gene model, then the parents would go through the CAPS (cleaved amplified polymorphic sequence) screening process, where each of the parents involved in our mapping populations were amplified and screened.

The screening process involved finding a CAPS polymorphism between parents by screening the available restriction enzymes described in Table 2.1. PCR of the *Phaseolus* population for each specific universal primer pair was performed in 96 well PCR plates before the digestion. The restriction enzymes are listed from top to bottom in a rank order. The restriction enzymes at the top of the list have a 4-base recognition site where as we go down the list towards the 6-base recognition sites. The enzymes at the top of the list appeared more likely to cut the PCR product, but that does not mean they were the most polymorphic enzymes either. Enzymes with a low cost are placed near the top of the list to be cost-effective.

Table 2.1 Restriction Enzyme List.

RANK of USE		RECOGNITION SEQ.	INCUBATION TEMP.	SUPPLIER	BUFFER
Hinf 1		GANTC	37	NEB	1
		CTNAG		PROMEGA	B
Rsa 1		GTAC	37	PROMEGA	C
		CATG			
Hae 3		GGCC	37	PROMEGA	C
		CCGG			
Alu 1		AGCT	37	NEB	2
		TCGA			
Nde 2	(same as Mbo1)	GATC	37	PROMEGA	D
		CTAG			
Cfo 1		GCGC	37	PROMEGA	B
		CGCG			
HpyCH4 4		ACGT	37	NEB	1
		TGCA			
Bfa 1		CTAG	37	NEB	4
		GATC			
Hsp92		CATG	37	PROMEGA	K/H
		GTAC			
Mse 1	(same as Tru9)	TTAA	37	NEB	2+BSA
		AATT			
Msp 1		CCGG	37	PROMEGA	B
		GGCC			
Taq 1		TCGA	65	PROMEGA	E/MC
		AGCT			
Tsp509 1		AATT	65	NEB	1
		TTAA			
Aci 1		CCGC	37	NEB	3
		GGCG			
BstU 1		CGCG	60	NEB	2
		GCGC			
Dde 1		CTNAG	37	PROMEGA	D
		GANTC			
Tsp45 1		GT(C/G)AC	65	NEB	1
		CA(G/C)TG			
Nci 1		CCSGG	37	PROMEGA	B
		GGSCC			
BstO 1		CCWGG	60	PROMEGA	C
		GGWCC			
Sin 1	(same as Sau96)	GGWCC	37	PROMEGA	C
	(same as Ava 2)	CCWGG			
Bsp1286	G(G/A/T)GC(C/A/T)C	GNGCNC	37	PROMEGA	A
real rec.seq->	C(C/T/A)CG(G/T/A)G	CNCGNG			
Xho 2		(A/G)GATC(T/C)	37	PROMEGA	C
		(T/C)CTAG(A/G)			
EcoR 1		GAATTC	37	NEB	EcoR1 Buffer
		CTTAAG			
Afl 3		AC(PU)(PY)GT	37	NEB	3
		TG(PY)(PU)CA			
Hind 3		AAGCTT	37	PROMEGA	E
		TTCGAA			
BamH 1		GGATCC	37	PROMEGA	E
		CCTAGG			
EcoR5		GATATC	37	PROMEGA	D
		CTATAG			
Dra 1		TTTAAA	37	PROMEGA	B
		AAATTT			
NEB- New England Biolabs. Ipswich, MA					

Results and Discussion

The first group of genes to be considered in designing universal primers were the sequenced RFLPs and ESTs of *Phaseolus vulgaris*. The small number of sequences

combined with a lack of identified genes quickly prompted elimination of this category. Two genes were targeted from sequenced RFLPs, replication protein –subunit 32 kilodaltons (*RPA32)*, mevalonate kinase (*MevlK*), and pyrroline-5-carboxylate synthase (P5CS); but only the *Phaseolus vulgaris* and one or two other sequences could be found, between the sequences a good alignment was provided. The primer pairs used in amplification were RPA32_F (CGGCGGCGGATTCATGTC), RPA32_R (TCAAARTTTGTCCRGGCCTG), MevlK_F (TCCRCGTTTCTCTGGYTRTAC), MevlK_R1 (CCTCCAGCTCCWGTTARTTTGGA), P5CS_F1 (GTTCCAGMGGCTTCAGG), and P5CS_R1 (TGTRCTTGCRTTGTGRAAHACAGC). The other primer for MevlK also worked, MevlK_R (CATKGACWGGAGCAGACCT). Multiple primer pairs were designed with a maximum of 5 degeneracies, meaning different combinations of 1 to 4 different nucleotides at specified base pair locations. Screening the parents with the designed primer pairs with the parents, and some cases the mapping populations. All three primer sets produced multiple bands in pea and lentil with a single band in bean for *MevlK* and multiple bands for *RPA32* and *P5CS*. Indicating the sequence from the *Phaseolus* ESTs (*MevlK*, *RPA32, P5CS)* was probably more specific to bean then pea and lentil and could not be used in designing universal primer pairs for legumes (Appendix A). Generally speaking the available EST sequences failed to identify any well-characterized genes screened in this research. Each primer pair was screened in PlantGDB for each gene attempted, and never produced a significant match to any of the other genes in the primer design pipeline.

The second group of primers to be tested was the utilization of UNISTS markers, by developing an additional primer to extend the small (<500 bp) UNISTS fragments into

large (>1,500 bp) fragments. Greater than 50 markers were found by query and several prospects, with sequences available (>10) were found. Several markers, (tRALS, SHMT, and RNAH) were aligned with an available genomic sequence and a universal priming location was found. The marker tRALS was found to be too variable outside the UNISTS region, and was not pursued. Primers were developed from two UNISTS based markers, SHMT and RNAH. They were amplified and put through a parental screen with a CAPS polymorphism found in pea and lentil, but not bean. SHMT produced a band ~1 Kb in bean, while RNAH produced a fragment ~800 bp. Both were monomorphic in bean by the CAPS analysis. SHMT and RNAH are two markers shown by Choi et al. (2004) to disrupt conservation of synteny in the legumes. The problem in this research with developing UNISTS markers into larger fragments is their low level of characterization, meaning there is not enough DNA sequence available for those genes to develop primers amplifying fragments of large size. Further work on the group of UNISTS markers was terminated, because finding conserved sequences with a sufficient distance to include a polymorphism was too difficult and time consuming for the low potential success rate. In certain cases (*bfruct, GAPA, ALDH7*) STS markers produced larger fragments then initially expected, a positive benefit only seen in the STS designed markers.

The initial screening of *Phaseolus* sequences and UNISTS markers provided a very limited number of appropriate universal markers. The intention further research was designing new markers to be universal for amplification of homologous fragments in legume species, as described in the following paragraphs.

The third group of DNA markers screened was the STS primer pairs used in mapping *Pisum sativum* (Weeden *et al.*, 1998). Using multiple primer pairs originally developed for pea, the lentil and bean parents were screened for amplification of homologous fragments, based on fragment size and restriction fragment sizes observable by CAPS analysis. Many of the genes mapped in this comparative study came from the *Pisum* STS category, either by the initial screen or the redesign of an original primer pair in favor of a universally designed primer pair, examples of the latter including *Apx1*, *Enolase*, *Rcpme*, *Aat*, and *TubA1*. All other primer pairs were based on new gene regions found to be conserved among species, even though, they were previously mapped in *Pisum*. Regardless of the fact that they were already mapped genes in pea, the new primer pairs had to be verified in pea to ensure the proper locus was being amplified and co-segregated with the original marker.

When new primer pairs were made of already mapped genes in pea, the first step was to test ten samples of the already genotyped pea lines by CAPS analysis. The design of universal primers entails making larger PCR fragments, so it was important to ensure the original polymorphism will be visible in the larger fragment. All the universally designed primers showed the original polymorphism mapped in pea (data not shown).

The last group of potential homologous markers to be screened included universal primer pairs designed to genes not previously mapped in pea, lentil, or bean. This last marker category accounts for the largest number of markers developed in this study, implicating how crucial primer design is and how bean and pea are divergent even at the level of highly conserved coding regions.

Table 2.2 List of Primer Sequences for Universal Markers.

Designed Universal Primer Pairs: Monomorphic

Primer pair	Primer sequence (5' to 3')
AC3_F	GCGATAATGGAACCGGAATGG
AC3_R	TTCCTGTGTACTATTGATGG
Acl_F3	CATCCCACACAATGAAGAGT
Acl_R2	TCTAATCCATGAATGAAAGA
Alat_F2	CTGACCTTCCCTCGCCAG
Alat_R2	CCTTGCTCGCACCATCAGT
cMDH_F4*	CATCACTCCCCTAGCAATC
cMDH_R4*	GATGTCAMCCATGCAACTGT
Crabs Claw_F	TCTGCAACACYTYCTTGCG
Crabs Claw_R	CTGAAAGCTTCWCGRTGWGG
Fbpp_F	CTTACTCTCCTTCACGTCT
Fbpp_R	CTTTTCAACCTTCTCCACCT
GAPA_F	TYCTGAATTCTCTGGYCTCCG
GAPA_R	CGAATTTCTGRTCAACRACYTTG
GAPA_R1	TGGCAACAATGTCAGCCAAATC
Gsn1_F2	CTCTCACTCTCTATTGCT
Gsn1_R2	ATGGAGCAATAGAGAGTG
Gsp_F2	CTCTCACTCTCTATTGCT
Gsp_R2	ATGGAGCAATAGAGAGTG
HsfA_F	CTTCGTTCGTCAACTCAACACC
HsfA_R	AAAAACCCTTCATTCGCAAAT
LASY_F	ACCCTGATGAGCCCACCA
LASY_R	GCAGCCCGATCTGAGTC
MNSOD_F	GAACATGCCTACTACTTACAG
MNSOD_R	CTCTCTTTCTCATATACTTCAC
NPAC_F2	TGGCTCCAGGTCCAGTTRTTGA
NPAC_R	CGATGGCTGGAACAGAAC
NUP133_F1	CTTGATGTGCAGGTAGATGAT
NUP133_R1	GCACCTCTGGTTGTTGTCC
GmPEPCN_F	CCAGCAGCTATACCCTTCATTGT
GmPEPCN_R	ATGGATGAGATGGCTGTCATTGC
PI4p_F	CTGGAGAGCAGACCCATTCAAT
PI4p_F2	TAGGTGCTGATATTCAATTTGG
PI4p_R	GCGAAAGCCTCCAATCCAC
PPE_F3	GGCTGGTTGTGATTGGATTCA
PPE_R3	GGRGTACCAGGGTTTAGGA
RCC_F1*	TCACACAACTCCTTTAACGAAA
RCC_R1*	TTATGCTAAATTCCTTGATGAAGC
RBRP_F	ATCATTCTGTGTTGCTTCTACGG
RBRP_R	TGAAAGGGCTCTGATAAGCATG
RPL35A_F*	GTGCCCCAAATAAGCTCTC
RPL35A_R*	GTGAGGCGTCTTCTAATAGC
SAHH_F	GATTGGGGTCCTGGTGGTG
SAHH_R	AAGAGCAGCAACCTTCTCATCAA
SBE1_F	GCYCCICGIATMTAYGARGC
SBE1_R	RTTGCCYTCYCTIGGRAARTC

Designed Universal Primer Pairs: Polymorphic

Primer pair	Primer sequence (5' to 3')
Apx1_F2	CCGATTACCAGAAGGCCATTG
Apx1_R1	CAAGAGGGCGGAATACAG
Pao_INT-F	GCAGGAGGGAAGGATAGATGAG
Pao_R3	GTGAAMGTGAGSRTTGTGTACTG
Tkp_F	GGIGAYGGITGYCARATGGA
Tkp_R	GAIGCRAGRTCIGCRCTIC
RCA_F2	AGCTTRTTGTTCACATCACCAAG
RCA_R2	TGWACTCTCTTCACATTCTCTTG
TuFM_F2	GATGGACCWATGCCTCAAAC
TuFM_R2	ACCCTCATCCTTTGTGAGTAC
Paal_F	GAGGTGAARCGYATGGTGG
Paal_R	GGTTTYCCTTGCATAACTTCAGC
ENOL_F2	TTCCATCAAGGCCCGTCAGA
ENOL_R2	TTGCACCAACCCCATTCATT
Apy_F2	GCCTGCWATCWCTTCCTCCC
Apy_R2	GCCTARAGGCCATGCRGCTTCC
CRS_F	GGCRTCTGGGATGTGTGCTA
CRS_R	TACACGGAGATGCAGCGTTTTC
Rcpme_F	GTTCAAAACTGTGGCTGAGGCT
Rcpme_R	GTGTTCTGGTTTGGGTCTTCTC
FK_F4	GTTCGGTCACATGCTCGCC
FK_R2	GCACCAACAAAYGAATCACC
Gpt_F3	GCTGAAGCTCCTAAGACTGATC
Gpt_R2	TGCGGCTCCYAGAGCATTGAC
Cipor_F2	GCTACTGCTAAGGCWTTGGCKG
Cipor_R2	CCWGATTTTGTTAGGCTTGGATCAC
Aatp_F	ATGGCTCCTCCAGATCCAA
Aatp_R	ATCTGCAAGGTATTCACAT
BFruct_F2	ACTTTCAACCTCTCAARAATTGG
BFruct_R2	CCCACACAATRTTYCCCCA
TubA1_F	GGAAACGCCTGCTGGGAG
TubA1_R	AACAGTTGGAGGCTGATAAT
ALDH7_F1	GGWTGGAATGCTTGCATTGC
ALDH7_R	TTGTRCAMCGYTGACCAG
GRF3_F1	CTCAGTGGCARGAACTTGAAC
GRF3_R1	GGCTTTCTTGAACGGTTGCG
Hsp70_F1	CGGNATYGATCTCGGNACSAC
Hsp70_R1	CATCTTNGCRATGATNGGRTTGC
PGIp_F2leg	AAGGCCTTTCGTGAAGCAGG
PGIp_R2leg	CCATCTGTAGCCCARTACCAA
UNK21_F	TCGCCTCCATGTCCACCTC
UNK21_R	CGGCCTTGCTAAATCAGTCAG

Table 2.2 List of Primer Sequences for Universal Markers-Continued.

Monomorphic cont..

Primer pair	Primer sequence (5' to 3')	Primer pair	Primer sequence (5' to 3')
SEP3_F2	GTTATGAGTCTCTACAACGCTCTC	SYM29_F	CAAAGAGATGAAAAGTATCAC
SEP3_R2	TCAGCACTCAGATTCAGTTGG	SYM29_R	TCGCCAGTGAGAGATTATC
SGR_F	ACCACAACTCACAAGTCACTTC	Thiop_F1	AGATGGAGAAGTCACCACAC
SGR_R	CTTCCAACACCTTCTCAACTC	Thiop_R1	CCATGAGTTCTTCAYTATCC
SHMT_F	ACCACAACTCACAAGTCACTTC	TIC55_F*	GTGGGGACGAGAGAAAGATGGRR
SHMT_R	CTTCCAACACCTTCTCAACTC	TIC55_R*	AGGAAGCTGATAGTCCTGCAACA
STV3_F	ATGGRTGGRWGGAATATGC	VR_F1	GAGTGTGTGTAACTGGAGGTACAGG
STV3_R	TCKCATGGGRGARTGRATT	VR_R1	ATGCCTAATGCGCCATCGACGGTTC

* Multiple bands in the PCR reaction made identification of the proper band during CAPS analysis impossible by only using EtBr staining or an improper sized band.

Due to the growing amount of publicly available DNA sequence, primer design will be more influential on comparative studies as genetic sequencing and genomics come of age. Table 2.2 shown above gives the primer sequences designed to amplify homologous fragments in pea, lentil, and bean.

These markers were designed to be universal based on the fact that they were designed to all available DNA sequence. Universal means the primer pairs will work in multiple species, some available sequence aligned poorly and was discounted, therefore the universal primers are universal based on the sequences used to design them. The universal marker for *Pao*, for example, was designed with cyanobacteria, maize, rice, tomato, potato, pine, cotton, and barrel medic DNA sequence, so it should be more universal than other markers with less amounts of DNA sequence and from a smaller screen of species. It also means that the forward primer may be more universal then the reverse and vice versa, based on the quality of DNA sequence at that specific site. The universal gene-specific marker *Paal* was designed with ~5 dicot species and two monocot species, so the universal primers ability to work in monocots will be less than in dicots. Universal gene-specific primer pairs were designed to the most conserved regions, so we

can assume that many of the primers have a possibility to work in distantly related species.

All of the chosen genes that passed the compiling step and the alignment step and also showed two conserved regions >1,500 bp, worked as universal gene-specific markers, except *GAPA, UNK21, ALDH7, HSP70*. The first primer pair for *GAPA* (GAPA_F/R) amplified a fragment 1.8 Kb and produced no polymorphisms when screened by 27 enzymes. New primers (GAPA_F1/R1) were designed to amplify a fragment 2.5 Kb and again *GAPA* was still monomorphic. CELI nuclease, proprietary name is surveyor nuclease, (Transgenomic Inc.; Omaha, NE) was used to see if a polymorphism could be identified, still no polymorphism was visible using the agarose gel detection kit from the company. CELI identifies and cleaves dsDNA at mis-matches when ssDNA is hybridized to its homologous ssDNA counterpart. CELI recognizes single base pair mis-matches up to large insertions and deletions. Primers designed specifically from a *Medicago* sequence would usually amplify a similar sized fragment in pea and lentil, but not in bean. GAPA was designed from many sequences, so primer sets for a large fragment could be designed that worked equally as well in all three species. GAPA was mapped in pea and lentil, but in bean it may require sequencing to identify a polymorphism.

Sequencing may have allowed transformation of some non-working and/or monomorphic primer pairs into working and polymorphic ones, by identifying SNPs (single nucleotide polymorphisms). Sequencing also might have helped in developing new primers by 5' or 3' RACE (rapid amplification of cDNA ends), allowing extension of known sequence in production of longer PCR fragments.

Table 2.3 shown below lists the genes found to be monomorphic by CAPS analysis. Many of the listed genes amplify well and are good candidates for further screening and possible use in future comparative analyses.

Table 2.3 Designed Universal Primer Pairs: Monomorphic.
-Table describes the genes that resulted in monomorphic fragments in bean and were not mapped.

Primer pair	Gene Product	PCR Fragment Size			PCR
		Pea	Lentil	Bean	
AC3	*Actin 3*	~1.3 Kb	~1.3 Kb	~1.3 Kb	G
Acl_F3/R2	*ATP: Citrate lyase*	~0.9 Kb	~0.9 Kb	~1.0 Kb	VG
Alat_F2/R2	*Alanine aminotransferase*	~0.8 Kb	~0.8 Kb	~1.6 Kb	VG
cMDH_F4/R4	*Cytoplasmic malate dehydrogenase*	~1.6 Kb	~1.6 Kb	~0.8 & 0.2 Kb	VG
Crabs Claw_F/R	*myb tanscription factor*	~1.2 Kb	*na*	~1.2 Kb	F
Fbpp_F/R	*Fructose bisphosphatase*	~0.7 Kb	~0.7 Kb	~0.7 Kb	G
GAPA_F/R	*Glyceraldehyde-3-phosphate DH (subA)*	~1.3 Kb	~1.3 Kb	~1.4 Kb	VG
GAPA_F/R1	*Glyceraldehyde-3-phosphate DH (subA)*	~2.0 Kb	~2.0 Kb	~2.5 Kb	VG
Gsn1_F/R	*Glutamine synthetase (cytosolic)*	~0.6 Kb	~0.6 Kb	~0.6 Kb	VG
Gsp_F2/R2	*Glutamine synthetase (plastidic)*	~0.8 Kb	~0.8 Kb	~0.8 Kb	VG
HsfA_F/R	*Heat shock factor A*	~0.1 & 0.5 Kb	~0.1 & 0.5 Kb	~0.3 & 0.5 Kb	VG
LASY_F/R	*Lipoic acid synthase*	~0.9 Kb	~0.7 & 0.9 Kb	~0.7 & 0.9 Kb	F
MNSOD_F/R	*Manganese superoxide dismutase*	~1.2 Kb	~1.2 Kb	~1.2 Kb	G
NPAC_F/R	*Putative actin*	~0.2 & 0.5 Kb	~0.2 Kb	~0.2 Kb	G
NUP133_F/R	*Nucleoporin 133*	~1.0 Kb	~1.0 Kb	~1.0 Kb	VG
PEPCN_F/R	*Phosphenolpyruvate carboxylase*	~1.1 & 1.3 Kb	~1.1 & 1.3 Kb	~1.1 & 1.3 Kb	VG
PI4p_F1/R1	*TGA-type basic leucine zipper protein*	~0.4 Kb	~0.45 Kb	~0.9 Kb	G
PI4p_F2/R1	*TGA-type basic leucine zipper protein*	~1.4 & 0.4 Kb	~1.4 Kb	~1.8 Kb	G
PPE_F3/R3	*Ribulose-5-phosphate-3-epimerase*	~0.7 Kb	~0.8 Kb	~1.3 Kb	VG

Reproducibility- VG= Very Good, G= Good, F= Fair. Listed in rank; VG= bright band and F= noticeable band.

Table 2.3 Designed Universal Primer Pairs: Monomorphic-Continued.

Primer pair	Gene Product	PCR Fragment Size			PCR
		Pea	Lentil	Bean	
RBRP_F/R	*Retinoblastoma-related protein*	~0.8 Kb	~0.8 Kb	~1.0 & 1.4 Kb	F
RCC_F/R	*Red chlorophyll catabolite reductase*	<1.0 Kb	<1.0 Kb	<1.0 Kb	F
RPL35A_F/R	*Ribosomal protein L35 subunit A*	~0.4 Kb	~0.2 & 0.3 Kb	~0.2 & 0.3 Kb	F
SAHH_F/R	*S-Adenosyl Homocysteinase*	~1.1 Kb	~1.1 Kb	~1.1 & 1.9 Kb	VG
SBE_F2/R2	*Starch branching enzyme*	~1.2 Kb	~1.2 Kb	~1.2 Kb	VG
SEP3_F2/R2	*SEPELA-like homologue*	~0.7 Kb	~0.7 Kb	~0.7 Kb	F
SGR_F/R	*Stay-green factor*	~0.6 & 1.0 Kb	~0.4 Kb	~0.6 Kb	F
SHMT_F/R	*Serine hydroxymethyltransferase*	~1.2 Kb	~1.2 Kb	~1.2 Kb	VG
STV3_F/R	*Sialyl transferase*	~1.4 Kb	~1.4 Kb	~1.4 Kb	VG
SYM29_F/R	*Serine-threonine protein kinase*	~0.4 Kb	~0.5 Kb	~0.5 Kb	F
Thiop_F1/R	*Thiol protease*	~1.2 Kb	~1.2 Kb	~1.2 Kb	VG
TIC55_F/R	*55 kD chloroplast translocation protein*	~0.5 Kb	~0.6 Kb	~0.6 Kb	F
VR_F1/R1	*Vestitone reductase*	~0.6 Kb	~1.0 Kb	~0.9 Kb	F

Reproducibility- VG= Very Good, G= Good, F= Fair. Listed in rank; VG= bright band and F= noticeable band.

Table 2.4 Designed Universal Primer Pairs: Polymorphic.

Primer pair	Gene Product	PCR Fragment Size			PCR
		Pea	Lentil	Bean	
Apx1_F2/R1	*Ascorbate peroxidase*	~1.8 Kb	~1.6 Kb	~1.6 Kb	VG
Pao_INT-F/R3	*Pheophorbide a monoxygenase*	1.2 & 1.3 Kb	~1.1 Kb	~1.1 Kb	VG
Tkp_F/R	*Transketolase (plastid-specific)*	~1.1 Kb	~1.1 Kb	~1.25 Kb	F
RCA_F2/R2	*Ribulose bisphosphate activase*	~1.0 Kb	~1.0 Kb	~1.9 Kb	G
TuFM_F2/R2	*Mitochondrial elongation factor*	~2.0Kb	~2.0Kb	>2.5Kb	VG
Paal_F/R	*Phenylalanine ammonia lyase*	~1.0 Kb	~0.9 Kb	~1.3 Kb	VG
ENOL_F2/R2	*Enolase*	~1.2 Kb	~1.2 Kb	~1.3 Kb	VG
Apy_F2/R2	*Apyrase S-Type*	~0.6 & 1.2 Kb	~0.6 & 1.2 Kb	~1.4 Kb	G
CRS_F/R	*Cystathionine-gamma-synthase precursor*	~1.3 Kb	~1.3 Kb	~1.5 Kb	G
Rcpme_F/R	*Pectin methylesterase*	~0.6 & 0.7 Kb	~0.6 Kb	~0.65 Kb	G
FK_F4/R2	*Fructokinase*	~1.1 Kb	~1.2 Kb	~2.0Kb	G
Gpt_F3/R2	*Glucose 6-phosphate translocator*	~1.8 Kb	~1.8 Kb	~2Kb	G
Cipor_F2/R2	*Protochlorophyllide reductase*	~1.1 Kb	~1.1 Kb	~1.2 Kb	VG
Aat_F/R	*Aspartate aminotransferase (cytosolic)*	~0.5 & 1.2 Kb	~0.5 & 1.2 Kb	~0.6 & 1.4 Kb	G
BFruct_F2/R2	*Beta-Fructose(invertase)*	~1.1 Kb	~1.1 Kb	~0.6 & 1.1 Kb	VG
TubA1_F/R	*Alpha tubulin*	~1.4 Kb	~1.4 Kb	~2.5 Kb	G

PCR Reproducibility- VG= Very Good, G= Good, F= Fair. Listed in rank; VG= bright band and F= noticeable band.

Table 2.4 describes the genes that resulted in polymorphic fragments and were mapped in pea, lentil and bean.

The genes listed below in Table 2.5 are only mapped in bean. The genes were only mapped in bean, because they were either monomorphic in pea by a CAPS screen or they have not gone through the full CAPS screening process.

Table 2.5 Designed Universal Primer Pairs: Only Polymorphic in Bean.

Primer pair	Gene Product	PCR Fragment Size			PCR
		Pea	Lentil	Bean	
ALDH7_F1/R	*Aldehyde dehydrogenase (family 7)*	~0.8 Kb	~1.0 Kb	~1.8 Kb	VG
GRF3_F1/R1	*Growth regulating factor 3*	~0.4 Kb	~0.4 Kb	~0.9 Kb	VG
Hsp70_F1/R1	*Heat Shock Protein 70 kD subunit*	~2.1 & 2.3 Kb	~2.1 & 2.3 Kb	~2.3 & 2.5 Kb	F
PGIp_F/R	*Phosphoglucose isomerase*	~1.5 Kb	*na*	~1.0 & 1.5 Kb	G
UNK21_F/R	*Unknown protein 21*	~0.7 & 1.4 Kb	~0.5 Kb	~1.1 Kb	G

Reproducibility- VG= Very Good, G= Good, F= Fair. Listed in rank; VG= bright band and F= noticeable band.

The primer names given above may not coincide with names in other tables. The names provided in these tables (Table 2.2, 2.3, 2.4, 2.5) are the names we originally assigned to the primer pairs. The citable primer names are given in a table at the end of chapter 4 (Table 4.21 or Appendix A).

CHAPTER 3

PHEOPHORBIDE A MONOOXYGENASE (*PAO*) IS LOCATED ON LG VII NEAR *AMY* IN PEA AND LENTIL

Introduction

The gene coding pheophorbide a monooxygenase (PAO) was first identified by a light dependent lesion mimic in *Zea mays*, called the lethal leaf spot mutation (*lls1*) (J. Gray *et al.*, 1997). The leaves of this mutant stay green during dark incubation and accumulate pheide *a,* causing a light-dependent lesion mimic phenotype. The *Pao* DNA sequence codes for a Rieske iron–sulfur cluster and conserved C-terminal motif in both plant and cyanobacteria that have sequenced *Pao* (J. Gray *et al.*, 1997). It has also been shown to be a member of a small non-heme oxygenase family that includes chlorophyll *a* oxygenase (CAO), choline monooxygenase (CMO), a 55 kD inner chloroplast membrane translocation protein (Tic55), and a 52 kD translocation protein for protochlorophyllide oxidoreductase A in chloroplasts (Ptc52) (J. Gray *et al.*, 2002). Homologs of *Lls1* (*Pao1*) have been subsequently identified in *Lycopersicon esculentum, Oryza sativa, Arabidopsis thaliana, Medicago truncatula, Pineas tuneada, Solanum tuberosum, Vigna unguiculata,* and other plant species by different research projects (J. Gray *et al.*, 2004; M. Wang *et al.*, 2004). The *A. thaliana* homolog, *accelerated cell death* (*Acd1,* At3g44880, Gi: 42565541), displayed less than 80% sequence identity with the *Zea mays* gene and not universally accepted as the *Pao1* ortholog until the latter gene was used to complement an *Acd1* null mutant in *Arabidopsis* (M. Wang et al., 2004). Additional articles have

been published confirming and restating the *Lls1* product is PAO, which catalyzes the third step in chlorophyll degradation (Fig. 3.1) (J. Gray *et al.*, 2004; S. Hortensteiner *et al.*, 1999; A. Pruzunska *et al.*, 2003; A. Pruzunska *et al.*, 2005; F. Vicentini *et al.*, 1995).

CHL Mg Dechelatase PAO RCCR *
chlorophyll a ---> chlorophyllide a----> pheophorbide a ---> RCC ---> pFCC ---> NCC

Figure 3.1 Basic pathway of chlorophyll degradation.

(**Substrates --> (**abbreviations: RCC, red chlorophyll catabolite; pFCC, primary fluorescent chlorophyll catabolite; NCC, non-fluorescent chlorophyll catabolite)

(**Enzymes --> (**abbreviations: Chl, chlorophyllase; Pao, pheophorbide *a* oxygenase; RCCR, red chlorophyll catabolite reductase; *, *[↓pH]* and multiple enzymes)

The 'stay-green' phenotype of the *lls1* mutant lead some researchers to propose *Pao* as a candidate for Mendel's *i* gene in pea, particularly after it was demonstrated that PAO activity was greatly reduced in green cotyledons (H. Thomas *et al.*, 1996). This study has led to the acceptance by many that the genetic basis of the *i* mutant is known (B. Buchanan *et al.*, 2000; H. Thomas *et al.*, 1996, H. Thomas *et al.*, 2000). Here we present evidence that the *Pao* gene in pea is located on LG VII, not on LG I and, therefore, cannot be synonymous with the genetic defect generating the green/yellow cotyledon polymorphism.

The stay-green phenotype has recently been found by collaboration of IGER and many researchers. We have since confirmed that the stay-green locus is the same locus as Mendel's *i* gene and will not be discussed in this thesis as what the gene product

codes. Since this chapter was already published during the time of this research and the publication verifying Mendel's *i* gene is still in the early phases of journal article submission at the time of the thesis deadline, so the primer sequence and discussion will be omitted. This paper was published in the on-going efforts to isolate both the *SGR* and *i* genes.

Materials and Methods

Amplification of *PsPao*:

Homologous mRNA sequences were obtained using a tomato EST (AF321984 [gi: 16973464]) for BLAST at NCBI; identifying *Arabidopsis thaliana* (Atg43880 [gi: 42565541]), maize (U77346 [gi: 1935910]), *Medicago truncatula* (TC100809) and *Vigna unguliata* (Gi: 30088963) for comparison of highly conserved regions and positions of introns. The gene structure (Fig. 3.4) and expected fragment sizes of the primer sequences in *Pisum, Lens, and Phaseolus* (Table 3.1) is presented below. Sequencing of the original DNA fragment amplified by the Pao-F2/R3 primer pair gave the sequences presented in Fig 3.2. To prevent mis-priming at lower annealing temperatures a touchdown protocol was used with each primer pair for sequencing. CAPS analysis was done with a standard PCR protocol (all primer combinations worked with a 58º C annealing temperature and a 2 minute extension) using *Taq* polymerase (Promega).

The two *PsPao* fragments were cloned using a pCR2.1 kit (Invitrogen) according to the manufacturer's recommended procedures. The two clones PsPao-WL808 and PsPao-MN313 were partially sequenced from both ends, and internal primers were

designed to permit sequencing of the remaining genomic sequence (final sequence not available).

Amplification of Other Gene Fragments:

PCR procedures used for IVDH, b-amylase, Sym29 and RAPDs (random amplified polymorphic DNA) were all done with Promega *Taq* polymerase. IVDH, b-amylase, and Sym29 were done with a standard PCR protocol using 60º C annealing temperature, a 2 minute extension and 40 cycles. The RAPDs were done with a 37º C annealing temperature, 2 minute extension with 35 cycles.

Sequencing:

Sequencing was carried out by the CORE facility at WSU-Pullman. Purified sequence and primers were sent to the sequencing lab for dye incorporation and sequencing. Complete sequence is not yet available, but sufficient sequence was obtained to verify the clones (WL808, an *P. sativum* ssp *abyssinicum*; MN313, a *P. sativum* ssp. *sativum*) identity near the 5' and 3' ends (Figure 3.2).

Populations Used in Mapping *PsPao1*:

The amplicon generated by Pao_F2/R3 was initially mapped in the JI1794 x 'Slow' RIL population that formed the basis of the *Pisum sativum* consensus map (N.F. Weeden *et al.*, 1998). The map position was confirmed in two other pea RIL populations. The first of these is a [*P. sativum* ssp. *sativum* x *P. s.* ssp. *abyssincum*] *x P. s.* ssp. *sativum* BC_1F_5 population consisting of 62 lines. The second is an F_2-derived F_5 population from the cross PI 220174 x A578-239. PI 220174 was collected in

Afghanistan and possesses a phenotype (R, M, Le, small seeds) typical of many landraces in that region. A578-239 is a multiple-marker line developed by and obtained from G.A. Marx. A final confirmation of the position of *Pao* was performed in the F_8 RIL population derived from the cross *Lens culinaris* 7 x *L. ervoides* 32 (S. Brauner *et al.*, 2002, Y. Tadmor *et al.*, 1987, Walling and Weeden, 2003, N.F. Weeden *et al.*, 1992).

Detection of Polymorphism:

Polymorphism in DNA fragments was detected based on original fragment size, when different, or Cleaved Amplified Polymorphic Sequences (CAPS) analysis. Fragments were separated on 2% agarose gels. The β-amylase polymorphism was also confirmed in the pea populations by isozyme analysis on a tris-citrate/lithium borate system (Weeden *et al.*, 1992) using young leaf extracts.

Estimation of Recombination Frequencies:

Recombination frequencies between pairs of loci in F_2-derived RILs were obtained by the equation $r = R/(2-2R)$ (B. Burr and F.A Burr, 1988). For small recombination frequencies ($r<0.1$) this equation provides a good approximation for *r* in the backcross RIL population as well.

Results

The fragment generated by primers Pao-F2 (5'GGAGAGATCATTGGTAYC CAGTTTC3') and Pao-R3 (5'GTGAAMGTGAGSRTTGTGTACTG3') was approximately the same size as predicted from the *Arabidopsis* gene model (Table 3.1). Table 3.1 shows the observed fragment size for *Pao* in pea, lentil, and bean. Sequencing

of the original DNA fragment amplified by the Pao-F2/R3 primer pair showed high sequence similarity between the two lines and shows five possible SNPs, as presented in Fig. 3.2. Two additional primers were designed to the verified nucleotide sequence and were used in mapping *Pao*. Sequences for the additional primers are Pao_INT-F (5'GCAGGAGGGAAGGATAGATGAG3') and Pao_INT-R (5'CTATCNCTTCTTCCTGTGACC3').

Table 3.1 Primer Combinations- Expected vs. Actual Fragment Size in 2% SB agarose gels.

	Exp. Pea Size	***Act. Pea Size***	**Lens Size**	**Phaseolus Size**
Pao-F/R	1550bp	~1500bp	Did not try.	Did not try.
Pao_INT-F/INT-R	700bp	>700bp	900bp	Faint band.
Pao-F2/INT_R*	1100bp	~1100bp*	1300bp*	Faint band.
Pao_INT-F/R3*	1300bp	~1300bp/1200bp*	1050bp	1150bp
Pao-F2/R3^{+}	1616bp	~1600/1700bp^{+}	Faint band.	Faint band.

**Primers used to map Pao.* $^{+}$*Primer pairs used in sequencing Pao.* *SB- Sodium-borate buffer.*

Figure 3.2 End sequences of the Pao-F2/R3 fragment showing >94% DNA sequence identity between WL808 and MN313

PsPao-F2 DNA sequence alignment of *P. sativum abyssincum* (WL808) and *P. sativum sativum* (MN313).
Identities = 379/401 (94.5%)

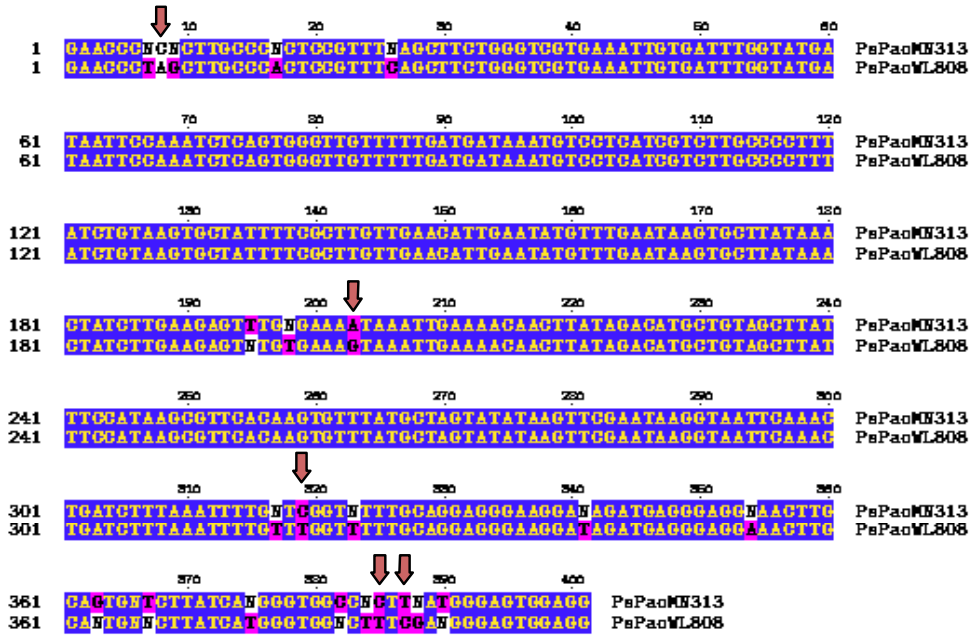


Figure 3.2 Continued

PsPaoWL808-F2 Obtained DNA sequence →
GAACCCTAGCTTGCCCACTCCGTTTCAGCTTCTGGGTCGTGAAATTGTGATTT
GGTATGATAATTCCAAATCTCAGTGGGTTGTTTTTGATGATAAATGTCCTCAT
CGTCTTGCCCCTTTATCTGTAAGTGCTATTTTCGCTTGTTGAACATTGAATATG
TTTGAATAAGTGCTTATAAACTATCTTGAAGAGTNTGTGAAAGTAAATTGAA
AACAACTTATAGACATGCTGTAGCTTATTTCCATAAGCGTTCACAAGTGTTTA
TGCTAGTATATAAGTTCGAATAAGGTAATTCAAACTGATCTTTAAATTTTGTT
TGGTTTTTGCAGGAGGGAAGGATAGATGAGGGAGGAAACTTGCANTGNNCTT
ATCATGGGTGGNCTTTCGANGGGAGTGGAGG

PsPaoWL808-R3 Reverse complement of obtained DNA sequence →
TCCCCATGGCACCCGGTAAGACTCGCTCCATTGTTTGCAGTGCTCGATAATTT
CTTCCAGTTCAACAGTGCCTGGGCCTGCCTGGTGGCAAGTAAGATTCGATTCC
GTTCTTTTTGATAAAACCACCCAATGTTCCATTAGGCTTATATATGTCATTTCT
TGTTCATGTTTCAGGTGGTTCCTAGATGGTACGAACATTGGACTTCAAACAAA
GTATATGATGGAGACATGATTGTCCTTCAAGGTCAAGAAAAAGTGTTTCTCG
CACAAACCGAGC

The size of the *Pao* fragment amplified by Pao_INT-F/R3 or F2/R3 differed between JI 1794 and Slow, allowing the segregation pattern to be observed directly on the 2% SB agarose gel without further restriction (Fig 3.3A). Similarly, a size difference was observed in the direct products amplified in the two *Lens* parents when the same primer pairs were used. In the other two populations CAPS was used to reveal a polymorphism, with Pao-F2/INT_R being used in the Abyssincium backcross RIL with *CfoI* (Fig 3.3B) and Pao_INT-F/R3 in the Afghanistan RIL cut with *HinfI*. In both pea and lentil a size polymorphism could be viewed when the second half or two-thirds of the gene was amplified, when the first half of the gene was amplified no size polymorphism was observed, so pea and lentil have two apparent alleles both differing in the second half of the gene by ~100 bp. Segregation of the various markers investigated in the analysis did not differ significantly from the expected 1:1 or 3:1 ratios expected in the F_2-derived

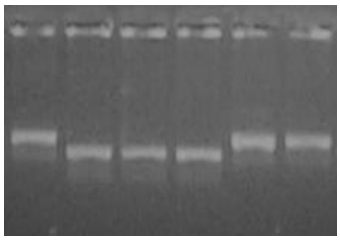
or BC-derived RILs, respectively. Joint segregation analysis between *Amy* and *Pao* in each of the pea populations indicted a recombination value between 3 and 7 cM (Table 3.2). A comparable recombination value was obtained between the two loci in *Lens*. *Pao* could be further fine-mapped in the JI1794 x Slow population using the dense marker map available for this set of RILs. The segregation data unambiguously placed *Pao* on LG VII between *Amy* and *Sym29* (Fig. 3.3).

Table 3.2 Joint segregation analysis between the loci *Amy* and *Pao* in four RIL populations.

	Number of lines with designated genotype										
Population	P_1P_1	P_1H	P_1P_2	HP_1	HH	HP_2	P_2P_1	P_2H	P_2P_2	N	r
JI1794 x Slow	19	0	3	0	0	0	0	0	27	49	3.3
BC1F5	11	0	3	0	0	0	2	0	46	62	4.4
Afghanistan	8	0	0	1	0	1	2	0	12	24	7.1
L.c. 7 x L.e. 32	22	0	2	0	0	0	0	0	22	46	2.3

Figure 3.3 2% Agarose Gel of *Pao.*

(A)The PCR amplification of JI1794(Lane 1), SLOW(Lane 2), JI1794 x Slow progeny(Lane 3-6) showing 2-JI1794 and 2-Slow genotypes.(B)Abyss. x Sparkle w/CfoI, 2-Abyss., 4-Spark., and 1 heterozygous genotype.

(A) 

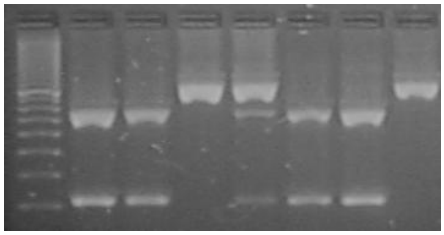
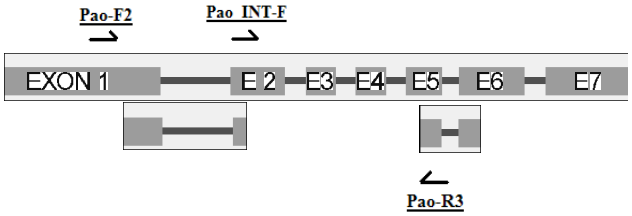
(B) 

Figure 3.4 *PsPao* gene model based on the gene splicing of *AtPao.*

PsPao DNA sequence →



Discussion

The nature of the "stay-green" phenotype in pea implies "green-cotyledon" is the same mechanism, indicating a breakdown in the chlorophyll degradation pathway before *Pao*. Chlorophyll is a green pigment where pheophorbide *a* is a light-green product, the phenotype is "stay-green", because chlorophyll levels have not changed in the chloroplasts and lack of plastoglobuli in the gerontoplast like WT pea plants (H. Thomas, 1977). We then believe *I* is a gene involved in the chlorophyllase step of chlorophyll degradation or the magnesium dechelatase step altering the pathway and resulting in only the first step of chlorophyll degradation or an entirely different mechanism that indirectly effects the degradation of chlorophyll and the subsequent disruption in the chloroplast, very possibly an effector of PAO or an earlier step.

A mutation at the *Pao* locus could provide an explanation for the green cotyledon phenotype of pea because of the enzyme's participation in the short pathway of chlorophyll degradation. The candidacy of *Pao* as the true identity of the *i* gene was further enhanced by the research of Thomas showing PAO activity was significantly reduced in pea lines with green cotyledons (H. Thomas *et al.*, 1996). However, our mapping data here shows *Pao* located on LGVII of pea, not at the *I* locus on LG I. We are forced to conclude that *I* is either a factor controlling transcription of *Pao* or a factor influencing another enzyme in the degradation pathway, probably an upstream enzyme like chlorophyllase. Protein localization and sequence studies have shown PAO is targeted to the chloroplast by a transit peptide of six or more serine residues, meaning chlorophyll degradation is occuring in the chloroplast, so the other enzymes in

chlorophyll degradation should be localized in the chloroplast. The identification of the *I* gene will, therefore, not be based solely on the localization of a gene product. The identification of several database sequences further verifies the homology of *Acd1* and the multiple species-specific forms of *Pao* that researchers have shown to be *Lls1/Pao1*. PAO reacts with oxygen, probably ferredoxin, and is possibly phosphorylated; many situations could arise that shift the kinetics of PAO and its degradation of pheophorbide *a*. Whether or not, the over accumulation of pheide *a* or free radical oxygen triggered and/or caused the lesions, is still left to be completely understood. Previous research showed *Pao* expression has been positively correlated with senescence and was subjected to post-translational regulation (A. Pruzinska *et al.*, 2003).

CAPs analysis of *Pao* in *Lens culinaris*, and *Phaseolus vulgaris* (unpublished) showed synteny to close markers in *Lens*, but not in *Phaseolus*. Using the http://medicago.org/genome database and the location of Pao1 on the Ji1794 x Slow linkage map we compared the genomic sequence. A comparison between a *Medicago* BAC clone (mt2-11o04) and an *Arabidopsis* BAC clone F28D10 found there is enough sequence similarity that mth2-11o4 contains the *Pao* gene. Genomic sequence of *MtPao*, matched the identified *MtPao*/*Lls1* cDNA with >99% sequence identity, but different intron lengths then *PsPao* and *AtPao*. The basic *PsPao* gene model is based on the *Arabidopsis* splicing sites and some alterations using *Medicago* with the higher sequence similiarity is presented above showing the region where *PsPao* sequence was obtained (Fig. 3.4).

The synteny in this region of the genome appears to be conserved in *Lens* and at least partially conserved in *Medicago truncatula*. *Medicago* BACs containing *Pao*

(mth2-11o4), *Apx* (mth2-27j21), *IVDH* (mth2-65c4), *SKDH* (BAC not found), *GS1* (BAC not found) and *Sym29* (mth11-o4) all are on the lower arm of LG 4 as presented in www.medicago.org. The relative gene order for these genes is also maintained in *Arabidopsis* and *Pisum*. Sequence analysis by BLAST has shown other Reiske Fe-S proteins possibly a TIC55 homolog on MtLG5 (mte1-51b2) and ~350bp of *AtPao* matching *AtCL9*, a gene tandem to *AtPao* shown to be a chloroplast ribosomal protein with high sequence identity to *ZmCL10*. In *Phaseolus vulgaris* stay-green is associated with the retention of thylakoid proteins like light-harvesting chlorophyll-binding protein of photosystem II, the 33 kDa polypeptide of the oxygen-evolving complex, cytochrome f and the psaF protein of photosystem I; while showing no unusual persistence of chlorophyll biosynthesis enzymes (A. Bachman *et al.*, 1994). Synteny has not been demonstrated in another species with respect to pea LG I aiding the identification of the *I* gene. One intriguing aspect is the location of chlorophyllase subunits close to *Arabidopsis* chalcone synthase (*CHS*), and the *I* gene maps near one of the two chalcone synthases in pea, but the *CHS* in *Arabidopsis* appears to be syntenic to the second *CHS* cluster on *Pisum* LG III.

The genes involved in chlorophyll degradation could be considered candidate genes (CG) for a green cotyledon phenotype. These CG's have similiar protein structures or functions and could interact or alter PAO kinetics by changing the PAO substrate or making substrate unavailable for PAO, as demonstrated by the lack of PAO translational control in Pružinská *et al.* 2005 and *Pao* not contributing to "stay-green" in Arabidopsis (R. Tanaka *et al.*, 2003).

Related to Thomas 1997 where accumulation of pheophorbide a and chlorophyllide was interpretated as a change in PAO activity in pea, common bean has no-change in the quantity of chlorophyll degradation products, stay-green in bean is only associated with the retention of thylakoid proteins, indicating regulation of PAO activity does not cause the stay-green phenotype in *Phaseolus* (A. Bachman *et al.*, 1994). Yang et al. 2004 used a PAO antibody for detection of *PsPao* in etiolated seedlings, also indicating PAO is constitutively expressed at low levels, but can cause a stay-green phenotype in *Arabidopsis* if the gene is knocked-out (A. Pruzinska *et al.*, 2005). The mapping of PsPao and northern blot analysis of LePao (Moffet M., Bergey D. unpublished) proves that PAO is continuatively expressed and not controlling the green cotyledon phenotype in pea, but regulated in a wound response pathway and up regulated in senescence.

Conclusion

Using primers that worked in the tomato northern blots we amplified the *Pao* gene in multiple lines to confirm the presence of *PsPao*. A multi-species sequence comparison and design of universally conserved primers allowed sequencing in *Pisum sativum* ssp. *abyssinicum (WL808) and P. sativum* ssp. *sativum (MN313)*. Screening by restriction enzymes/(CAPS) analysis showed two alleles in our collection of *Pisum* parental lines and the reliability of the primers to work in multiple species (tomato, pea, lentil and common bean). *PsPao* is relatively polymorphic even within species and is a single locus that makes it useful in PCR-based map generation and marker assisted selection (MAS) in breeding projects. Finding only one locus in the pea genome should

dictate the nomenclature, so we propose *Pao1* should be referred to as *Pao*. The primers used in this study should require little altercation (Pao-F2, Pao_INT-F, and Pao-R3) to work reliably in many other species for amplifying *Pao*, due to their conservation and robustness between legumes.

In relation to the "green cotyledon" phenotype, we must find other controlling factor(s) of PAO, in the pathway of its synthesis or the path leading to PAO's interaction with its appropriate substrate(s). We conclude the *I* gene is probably connected to chlorophyll degradation, but has been demonstrated by mapping to not be *PsPao* and thus probably a transcription factor affecting chlorophyllase or changes in the thylakoid architecture.

CHAPTER 4

MAPPING GENE-SPECIFIC MARKERS IN MULTIPLE LEGUME SPECIES

Introduction

The mapping of genes in *Phaseolus vulgaris* was the rate-limiting step in universal marker development. The mapping populations for pea and lentil were more polymorphic than the bean mapping population, and the success rate of working primer pairs was better in pea subsequently lentil and bean. A summary table (Table 4.22) at the end of the chapter will show that the success rate was based on all the markers working in pea and lentil then bean. Many of the screened markers worked in one or two species, but finding a marker that both amplified a clean product and was polymorphic was difficult. The 16 genes mapped in all three mapping populations were previously mapped in pea and lentil by myself or J. Walling prior to my starting of the masters research program, so details of mapping those genes in pea and lentil was excluded. Presented in this chapter is the information on the pea, lentil, and bean mapping populations and the methods used to acquire linkage data in the *Phaseolus vulgaris* RIL mapping population BAT93 x JaloEEP558 maintained and provided by Dr. P. Gepts of U. C Davis along with segregation data.

Quikmap EXCEL macro allows the simultaneous comparison of linkage across the entire genome as one marker is tested at a time, thus allowing the easy identification of strong or weak linkage for every marker in the genome. The markers were compared by hand using Quikmap, because data in the excel spreadsheet allows the user to scan the

genome looking for linkages based on the number of matching genotypes, meaning to scan the genome with a naked eye to find the strongest matches.

Materials and Methods

CAPS Screening for Polymorphism

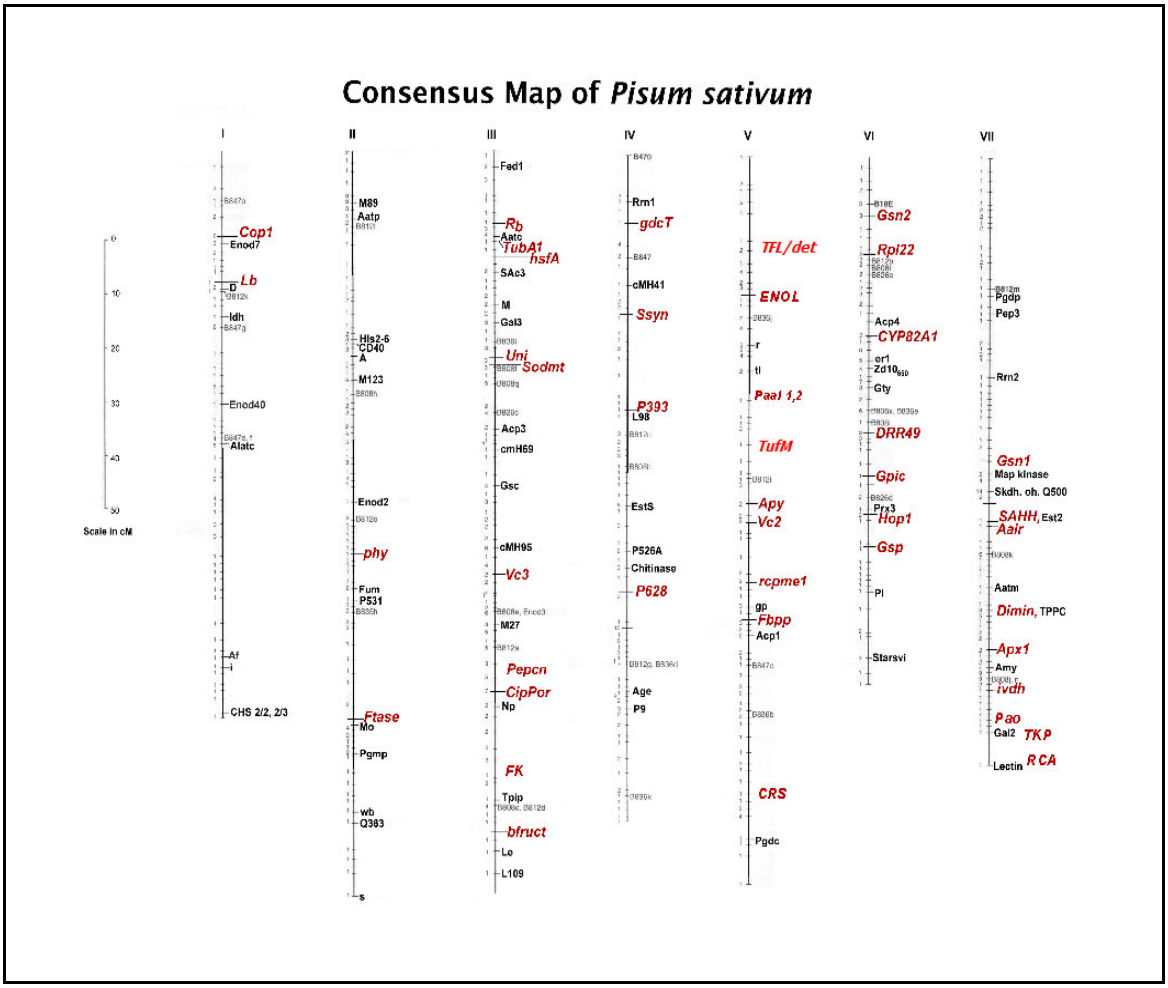
Restriction analysis generally was limited to endonucleases with 4-base recognition sequences, giving a total of ~25 unique restriction enzymes used in the polymorphism screening process. In certain cases PCR worked better for one species than another, so the screening process might not be concurrent between the different parental lines. If a PCR fragment was not polymorphic between the two parents of any mapping population, after all 25 restriction enzyme digests, the primer pair screening was terminated to permit time for other marker advancement that would work in pea, lentil and bean. The CAPS screen (partially described in chapter 2) utilized the 21ul left after the 4ul PCR check gel and was dependent on the intensity of the PCR amplification some or all the PCR sample would be digested. A check gel is 4ul of PCR reaction mixture loaded into a thin 1-2% agarose gel to verify the PCR amplification worked and to what amount. The amount of PCR product would determine if the remaining PCR sample could be subdivided into four 5.25ul digests, three 7ul digests, two 10.5ul digests, or one 21ul digest.

Plant Genetic Material for *Pisum Linkage Mapping*

Three populations of recombinant inbred lines (RILs) were used for mapping the amplified sequences. The 50 lines of JI1794 x Slow F2-derived F12 RILs constitute the

mapping population used to generate the *Pisum* consensus map shown as Figure 4.1 below (Weeden *et al.*, 1998). The population has considerable polymorphism and is segregating for many morphological characters. *Pisum sativum* is a diploid (2*n* = 14), a large genome (~4,800 Mb), and has many phenotypic markers, making it an interesting model for genome research. The mapping population consists of ~50 morphological markers, 15 isozyme markers, > 1000 RAPD's, >200 RFLP markers, 50-60 AFLP markers and ~180 CAPS markers.

Figure 4.1 Pisum Linkage Map (M.S.U linkage map of Ji1794 x Slow)



Plant Genetic Material for *Lens* Linkage Mapping

The lentil population used for mapping *Lens culinaris* 7 x *L. ervoides* 32 F7-RIL consisted of 46 lines and already has genetic maps developed (Tadmor *et al.,* 1987; Walling and Weeden, 2003; Weeden *et al.*, 1992; Brauner *et al.,* 2002). The lentil genome has a size of ~4,100 Mb and a chromosome number (2*n* = 14). Currently there are >100 CAPS markers, RAPDs, isozymes, and RFLPs mapped in the *L.c.* 7 x *L.e.* 32 population. Mapping of homologous gene-specific primers was previously done by a former graduate student, Jason Walling. Walling (2003) mapped ~ 20 genes in the *L.c. 7* x *L.e.* 32 population creating the M.S.U lentil linkage map (Walling and Weeden, 2003). Since that time I have mapped in *Lens* ~20 more genes (Weeden *et al.*, 2005). The gene-specific markers on the current M.S.U. lentil linkage map share in their conservation of synteny almost every marker mapped with the M.S.U. pea linkage map, therefore, the focus has been mapping the genes in bean.

Plant Genetic Material for *Phaseolus Linkage Mapping*

The common bean genome is diploid (2*n* =22) and approximately 590Mb. Its small genome size makes it an interesting system to explore conservation of synteny with large and evolutionarily distant genomes, like pea and lentil. The primary RIL mapping population for common bean (*Phaseolus vulgaris*) is BAT93 x JaloEEP558 developed by P. Gepts and is representative of the *Phaseolus* consensus linkage map shown as Figures 4.3-4.11 (Freyre *et al.*, 1998). Once a polymorphism was observed between the parents, the RIL population was scored and the resulting segregation pattern compared to all previously mapped markers using the Quikmap EXCEL macro (Weeden and Barnard,

1994). The segregation data for already mapped markers was provided by P. Gepts via personal communication and by his website (http://www.agronomy.ucdavis.edu/gepts/bjril7.htm).

The BAT93 x JaloEEP558 population is a wide cross from the two separate gene pools of Mesoamerican and Andean beans. Multiple primer pairs were developed to amplify fragments larger than the fragment originally mapped in *Pisum*. In common bean there are ~70 RFLPs and ~50 AFLPs including isozymes, SCARs, and SSRs. The integrated map consists of several RFLPs per linkage group mapped in multiple maps/crosses at different labs (Davis, Florida and Paris RFLP and RAPD maps combined) of bean was presented at the Plant Genome conference III (Freyre *et al.*, 1995).

Figure 4.2 *Integrated Linkage Map of LG 1*

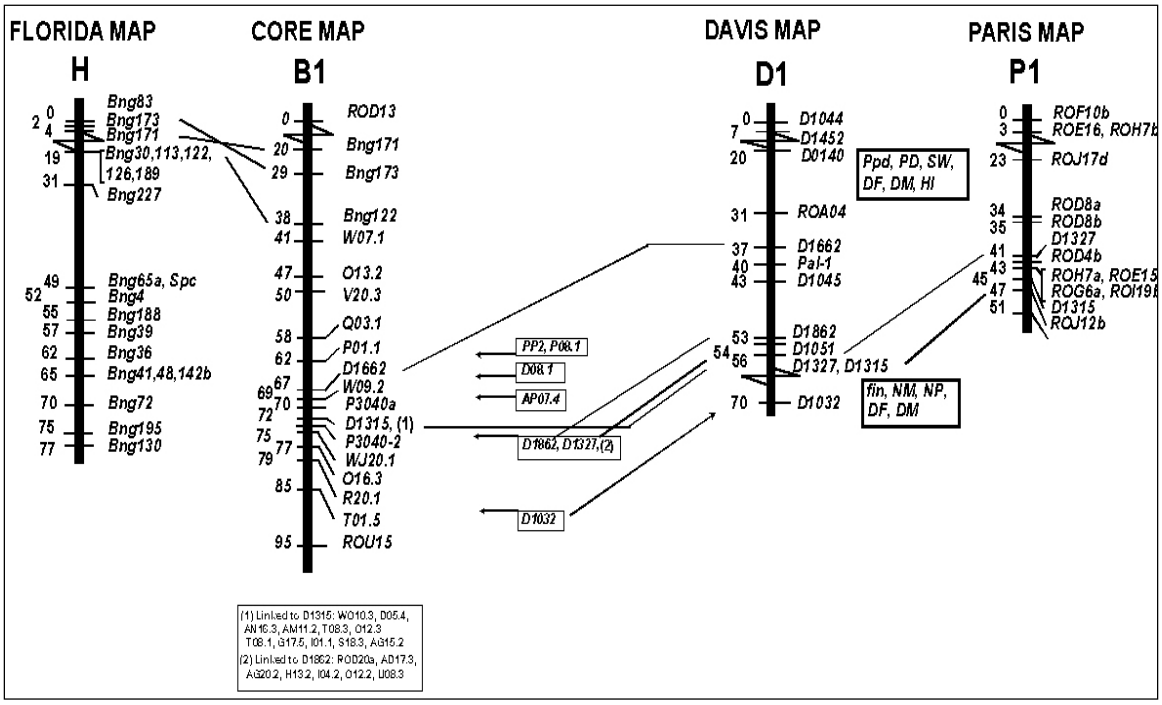


Figure 4.3 *Integrated Linkage Map of LG 2*

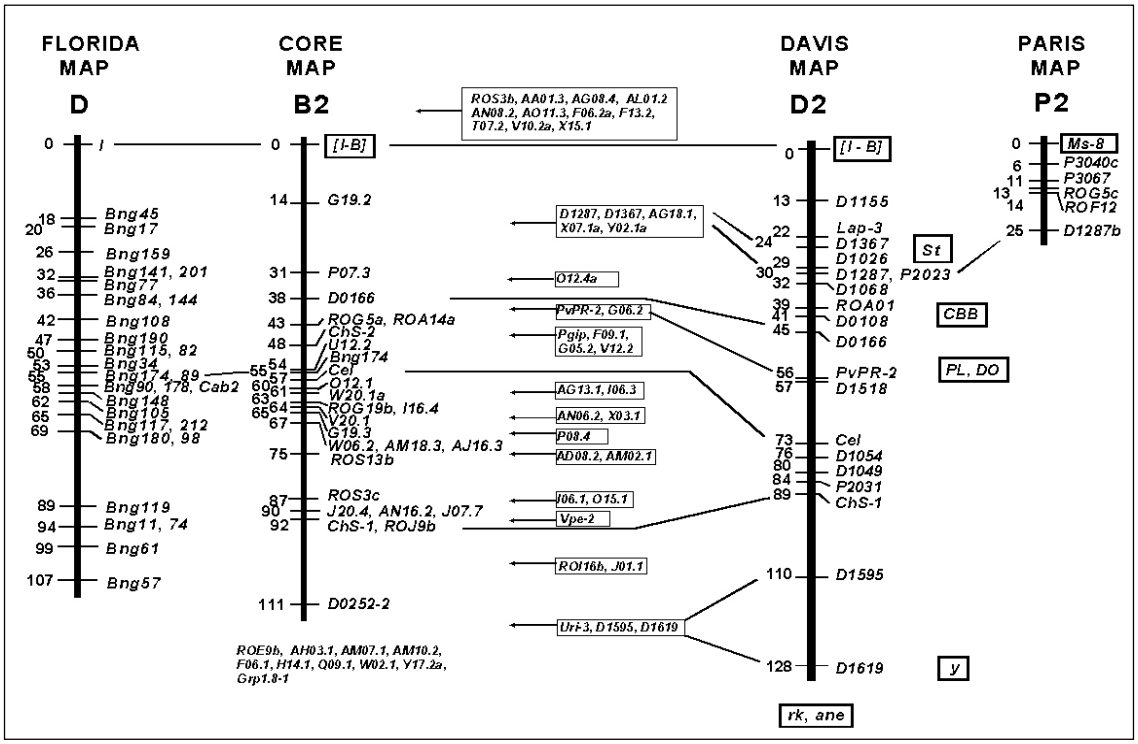


Figure 4.4 *Integrated Linkage Map of LG 3*

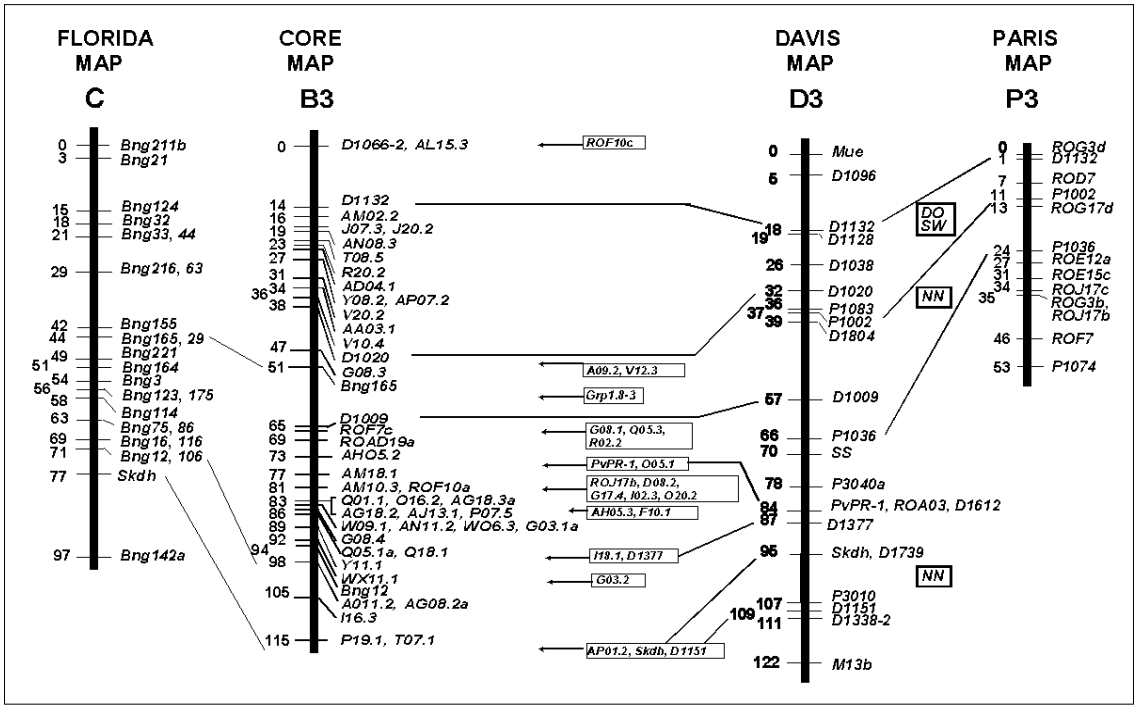


Figure 4.5 *Integrated Linkage Map of LG 4*

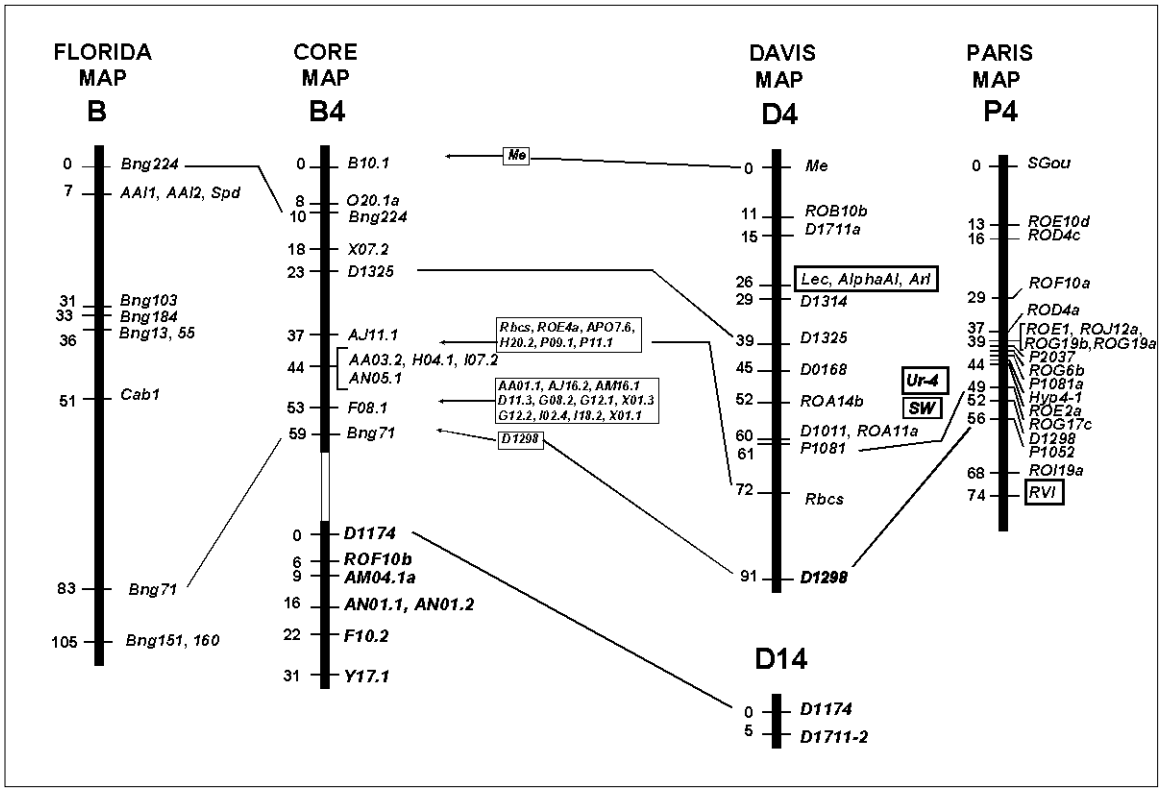


Figure 4.6 *Integrated Linkage Map of LG 5*

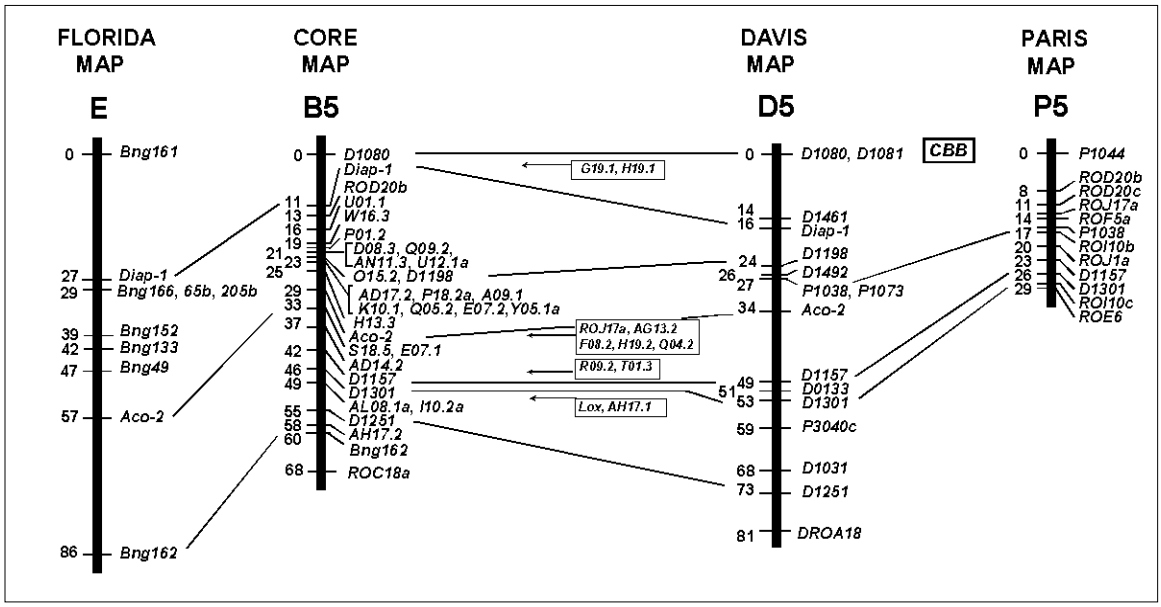


Figure 4.7 *Integrated Linkage Map of LG 6*

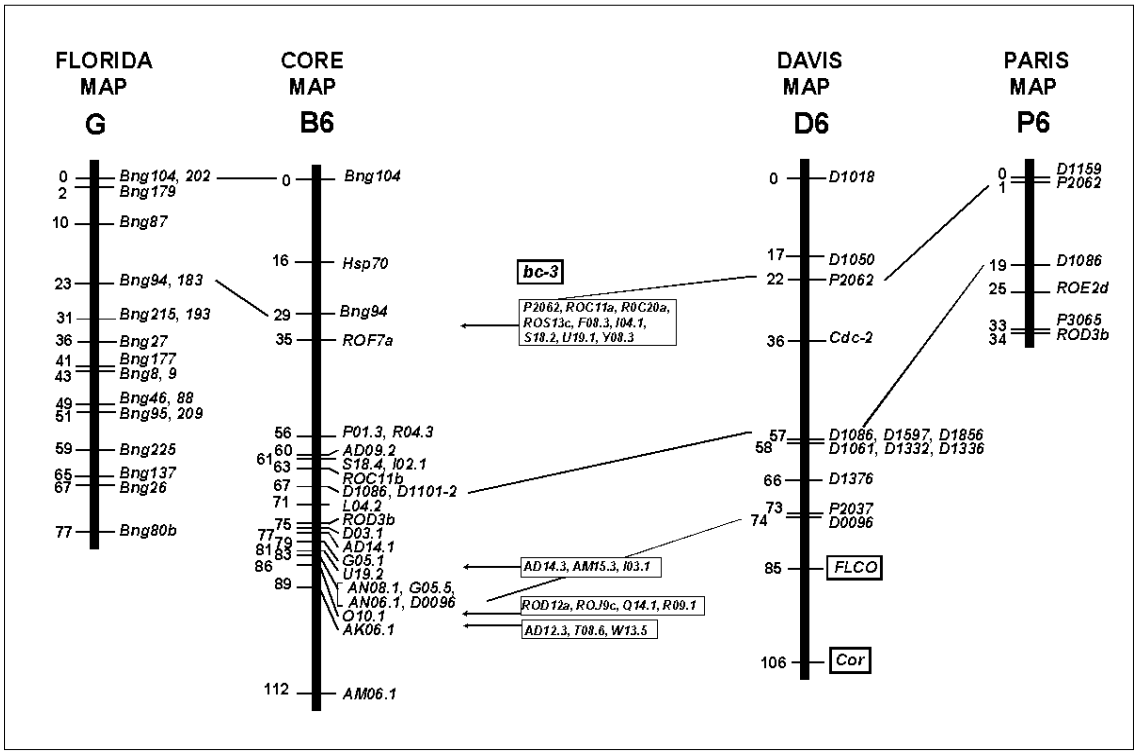


Figure 4.8 *Integrated Linkage Map of LG 7*

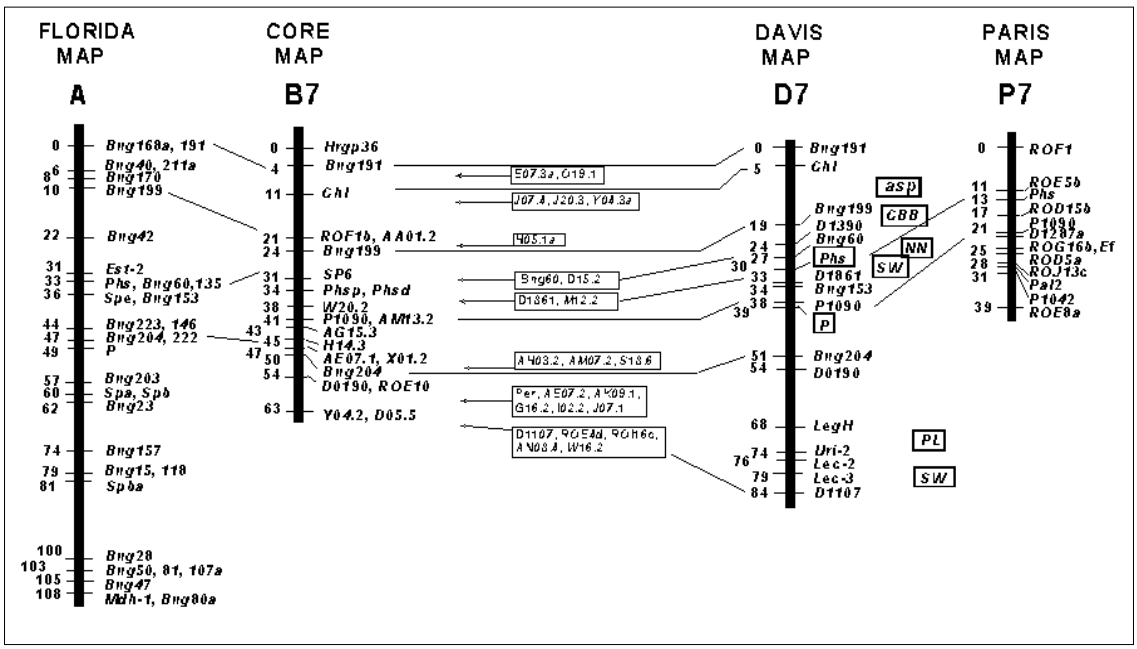


Figure 4.9 *Integrated Linkage Map of LG 8*

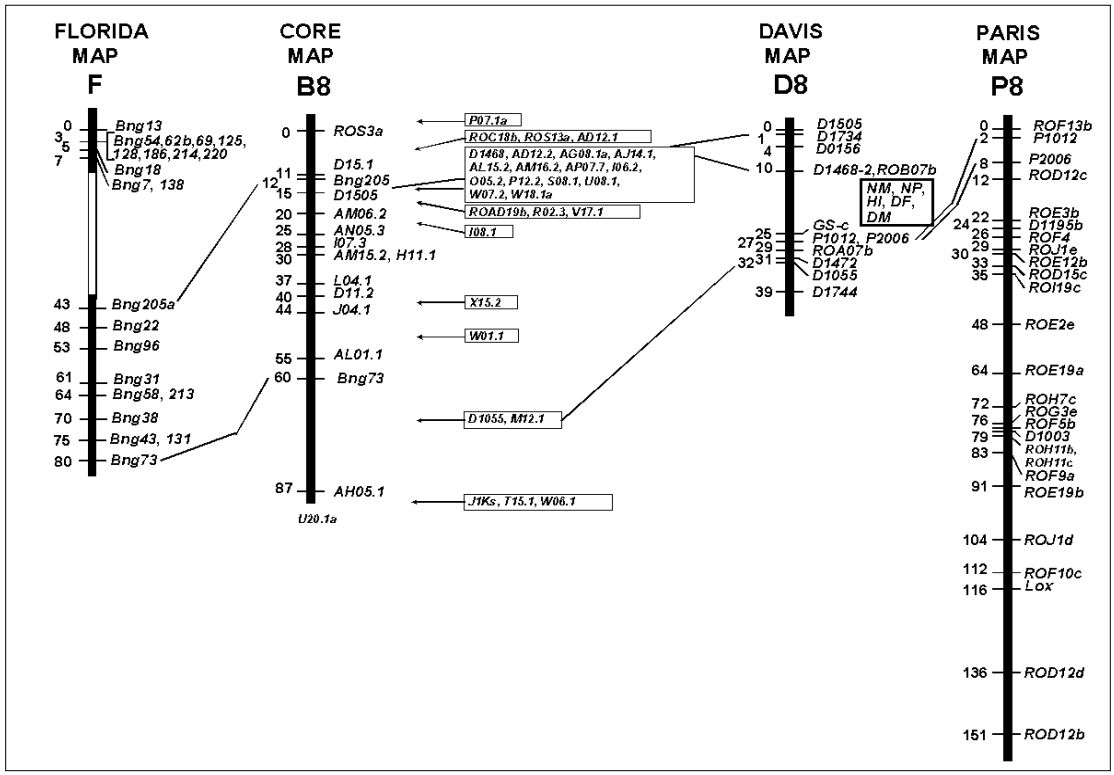


Figure 4.10 *Integrated Linkage Map of LG 9*

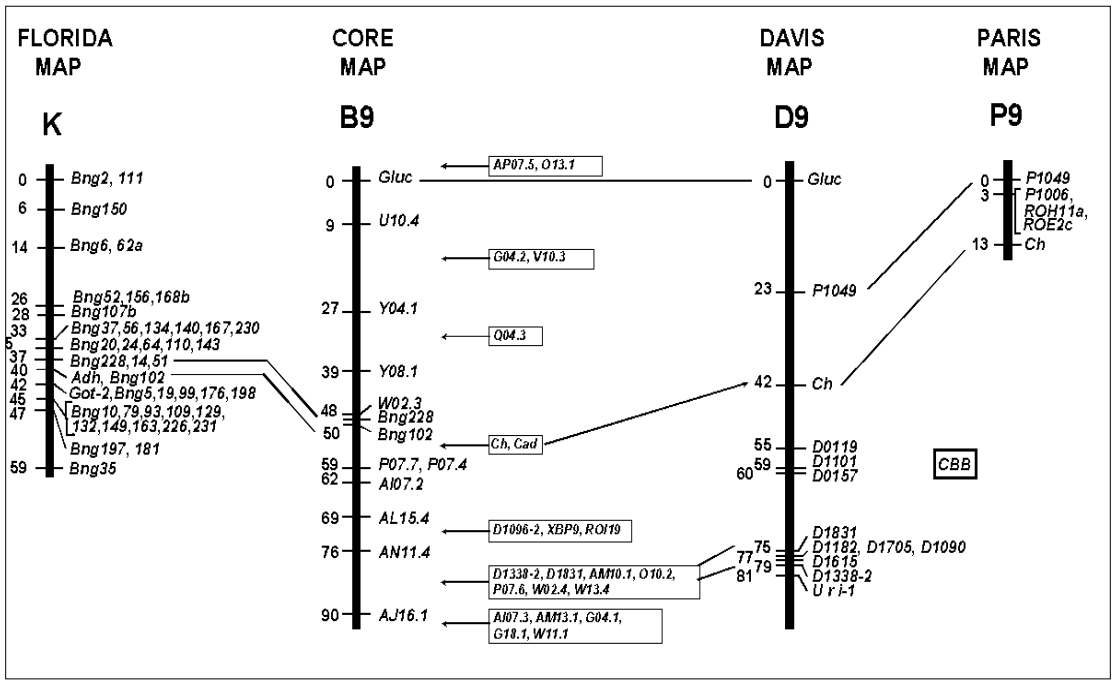


Figure 4.11 *Integrated Linkage Map of LG 10*

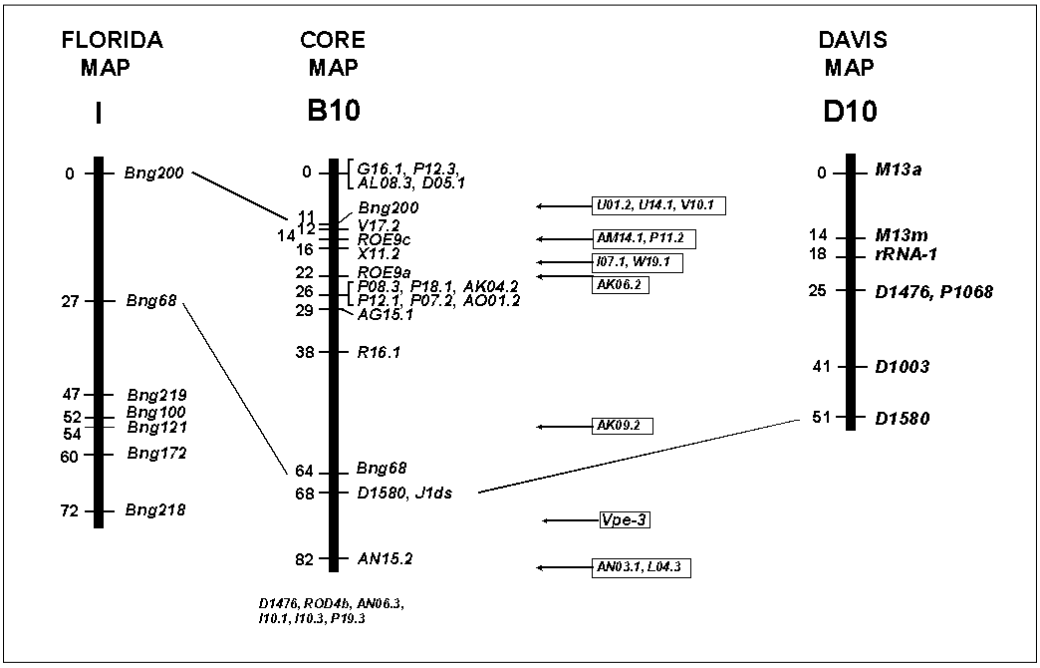
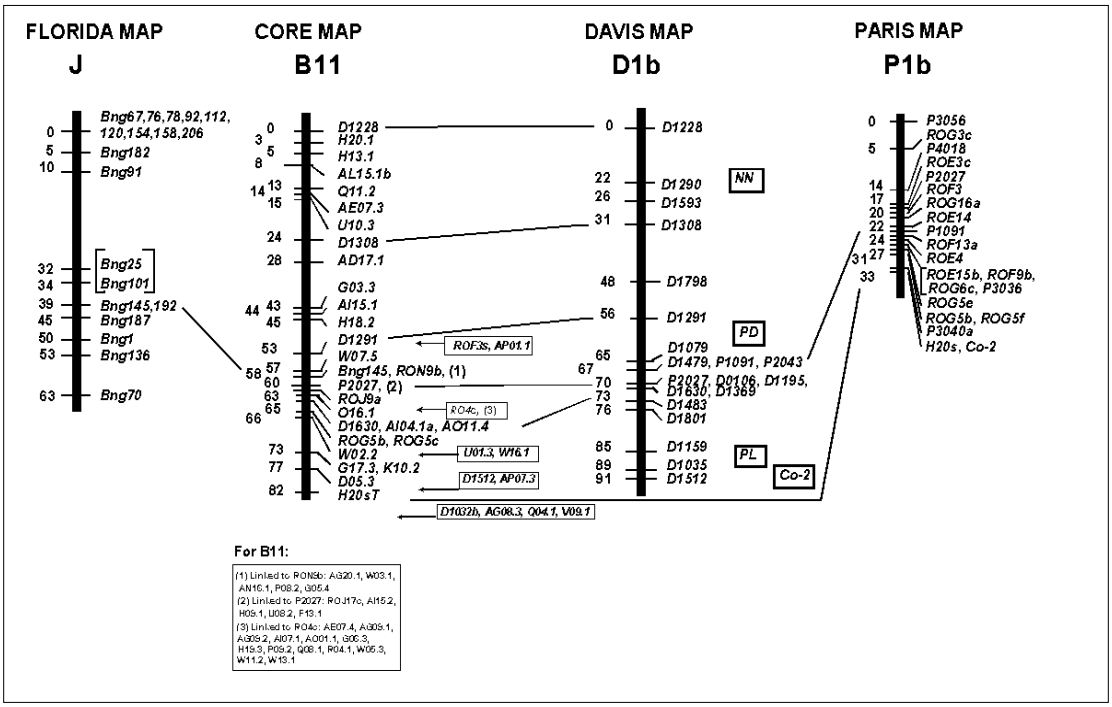


Figure 4.12 *Integrated Linkage Map of LG 11*



Results

Of the 100 genes for which 124 primer pairs were constructed, 53 gave fragments of the appropriate size in at least one of the species examined and similar size bands in the other two species (Appendix A), while 16 (16%) mapped in all three species. The 100 genes screened by PCR in bean consisted of 20 (20%) polymorphic, 37 (37%) monomorphic, and 43 (43%) genes not amplifying an appropriate fragment in the PCR reaction. The amplified products, polymorphic and monomorphic, give a PCR success rate of 57/100 genes or 57% of the genes appeared to amplify the proper gene in bean. Overall, the success of finding polymorphism in the universal markers was 16/53 or 30% of the genes that amplified also showed CAPS polymorphism in all three species. There was an obvious correlation between the number of DNA sequences used in primer design and the success of the designed primer pairs; more was always better. The limiting factor was the lack of CAPS polymorphisms between the bean parents. The majority of the 100 genes are polymorphic in pea and lentil. As a result of the monomorphic data the suggestion should be made that the bean population has roughly 1/3 the sequence divergence within conserved genes, as compared to the pea and lentil mapping populations.

Each of the 16 universal markers for which a CAPS polymorphism was identified in bean (Chapter 2) were located on the U.C. Davis bean linkage map (Freyre *et al.*, 1998). In general the position of the genes are unambiguous based on the segregation patterns provided by P. Gepts and those acquired by the CAPS analysis in the bean mapping population. Quikmap analysis revealed a high correlation between the

segregation pattern of each universal marker and a unique/specific region on the U.C. Davis linkage map with two exceptions. The two exceptions, *UNK21* and *ALDH7*, segregated independently of all regions except the end of linkage group 4 (Table 4.10). The two markers showed the strongest linkage to this area and based on double recombinants the markers were mapped to a locus beyond the most distal marker on the U.C. Davis linkage map. The joint segregation analyses below show the pair-wise comparisons of all the mapped markers and their flanking RFLP or isozyme markers (Tables 4.1 to 4.20). AFLPs may occur much closer to the mapped marker, but only a sub-set of the bean population was run with the AFLP primer combinations, so more segregation data is available for the RFLP markers. RFLP markers are denoted by a capital "D" where isozyme markers are given by just the name. The universal markers that were mapped in pea, lentil, and bean are in **bold** type and underlined. The tables below show the amount of similar genotypes (AA, AB, BA, BB) between the markers (M1-M2, M2-M3, M1-M3 for most cases). The heterozygous genotypes (HA, AH, HH, BH, HB) were ignored in the segregation analyses, because they occurred at a very low frequency. M# refers to Marker #, which has the name of the marker listed above.

Segregation Data for the Universal Gene-Specific Primer Pairs

Phaseolus Linkage Group 1 Markers:

Table 4.1 *Enolase* Linkage Data

		D1862	***Enolase***	D1662				
		Marker 1	Marker 2	Marker 3				
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	7	2	2	21	32	7.1	23.6	<.001
M2-M3	11	0	0	23	34	0.0	38.2	<.001
M1-M3	12	4	4	37	57	8.2	39.3	<.001

A=BAT B=JALO

Table 4.2 *Paal* Linkage Data

	D1862 Marker 1	***Paal*** Marker 2		D1662 Marker 3				
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	8	1	3	19	31	7.4	20.7	<.001
M2-M3	11	2	0	21	34	3.1	29.5	<.001
M1-M3	13	4	4	36	57	8.2	37.6	<.001

A=BAT B=JALO

Table 4.3 *TufM* Linkage Data

	D1862 Marker 1	***TufM*** Marker 2		D1662 Marker 3				
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	8	1	3	20	32	7.1	22.3	<.001
M2-M3	10	2	1	22	35	4.7	27.9	<.001
M1-M3	13	4	4	37	58	8.0	39.2	<.001

A=BAT B=JALO

Table 4.4 *Paal, Enolase, and TufM* Linkage Data

	D1862 Marker 1	***TufM*** Marker 2	***Enolase*** Marker 3	***Paal*** Marker 4	D1662 Marker 5			
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	8	1	2	21	32	5.2	26.2	<.001
M1-M3	7	2	2	21	32	7.1	23.6	<.001
M1-M4	8	1	3	18	30	7.7	19.2	<.001
M1-M5	13	4	4	37	58	8.0	39.2	<.001
M2-M3	11	2	1	26	40	4.1	34.4	<.001
M2-M4	12	1	2	24	39	4.2	31.4	<.001
M2-M5	10	2	1	22	35	4.7	27.9	<.001
M3-M4	12	0	2	25	39	2.7	35.8	<.001
M3-M5	11	0	0	24	35	0.0	39.8	<.001
M4-M5	11	2	0	21	34	3.1	29.5	<.001

A=BAT B=JALO

Table 4.5 *Rca* Linkage Data

	D1862 Marker 1	***Rca*** Marker 2		D1662 Marker 3				
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	8	2	4	20	34	10.7	17.6	<.001
M2-M3	11	2	4	18	35	10.3	15.7	<.01
M1-M3	14	4	14	31	63	20.0	14.2	<.01

A=BAT B=JALO

Phaseolus Linkage Group 2 Markers:

Table 4.6 *Rcpme* Linkage Data

	J9b Marker 1	***Rcpme*** Marker 2		CHS 1 Marker 3				
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	28	1	2	15	46	3.5	38.3	<.001
M2-M3	27	2	1	12	42	3.8	36.0	<.001
M1-M3	38	2	2	26	68	3.1	54.8	<.001

A=BAT B=JALO

Table 4.7 *Grf3* Linkage Data

	J9b Marker 1	***Grf3*** Marker 2		CHS 1 Marker 3				
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	24	1	2	12	39	4.2	31.4	<.001
M2-M3	23	2	1	9	35	4.7	29.4	<.001
M1-M3	38	2	2	26	68	3.1	54.8	<.001

A=BAT B=JALO

Phaseolus Linkage Group 3 Marker:

Table 4.8 *Apyrase* Linkage Data

		D1009 Marker 1	***Apyrase*** Marker 2	Bng165 Marker 3				
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	9	2	6	13	30	18.2	6.0	≤.1
M2-M3	14	1	2	13	30	5.6	19.0	<.001
M1-M3	25	5	11	29	70	14.8	18.2	<.001

A=BAT B=JALO

The data for *Tkp* (shown below) was incomplete so both of the observed regions showing linkage are given in the table below, which includes linkage groups 3 and 7 in bean. The two regions appear to have similar segregation patterns for this area of the genome as linkage is shown between M1-M4. Confusion in the marker order could also contribute to the observed results, because the strongest linkage appears to be D1151 to D0190.

Table 4.9 *Tkp* Linkage Data

	LG3 D1151 Marker 1	LG3 Skdh Marker 2	***Tkp*** Marker 3	LG7 D0190 Marker 4	LG7 DBng204 Marker 5			
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	21	9	17	19	66	32.5	0	*na*
M1-M3	6	0	2	7	15	7.7	8.4	<.05
M1-M4	19	14	20	17	70	47.2	0	*na*
M1-M5	20	11	24	15	70	50	0	*na*
M2-M3	5	3	2	5	15	25	0.1	*na*
M2-M4	23	13	12	14	62	33.8	0	*na*
M2-M5	24	12	14	5	55	44.8	0	*na*
M3-M4	6	2	2	5	15	18.2	2.3	<.1
M3-M5	6	2	2	6	16	16.7	3.0	<.1
M4-M5	36	1	8	24	69	7.5	40.0	<.001

A=BAT B=JALO

Phaseolus Linkage Group 4 Markers:

Table 4.10 *Unk21* and *ALDH7* Linkage Data

	Bng224	Me	***Unk21***	***ALDH7***				
	Marker 1	Marker 2	Marker 3	Marker 4				
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	29	13	8	19	69	21.9	6.4	<.1
M1-M3	14	13	5	10	42	37.5	0	*na*
M1-M4	13	12	9	8	42	50.0	0	*na*
M2-M3	16	5	5	17	43	15.2	10.0	<.05
M2-M4	14	7	8	15	44	25.9	0	*na*
M3-M4	17	4	5	18	44	12.9	13.5	<.01

A=BAT B=JALO

Phaseolus Linkage Group 5 Marker:

Table 4.11 *Fk* Linkage Data

	D1251	***Fk***	D1301					
	Marker 1	Marker 2	Marker 3					
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	16	3	2	16	37	7.8	19.1	<.001
M2-M3	18	0	2	16	36	2.9	28.7	<.001
M1-M3	32	6	7	27	72	11.0	27.4	<.001

A=BAT B=JALO

Phaseolus Linkage Group 6 Markers:

Table 4.12 *Gpt* Linkage Data

	D0096	***Gpt***	ROD3b					
	Marker 1	Marker 2	Marker 3					
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	20	0	1	13	34	1.5	31.6	<.001
M2-M3	18	4	2	15	39	9.1	18.2	<.001
M1-M3	32	6	2	25	65	7.0	37.2	<.001

A=BAT B=JALO

Table 4.13 *Pao* Linkage Data

	D0096	***Pao***	ROD3b					
	Marker 1	Marker 2	Marker 3					
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	18	1	0	14	33	1.6	29.6	<.001
M2-M3	17	3	1	16	37	6.1	22.5	<.001
M1-M3	32	6	2	25	65	7.0	37.2	<.001

A=BAT B=JALO

Phaseolus Linkage Group 8 Markers:

Table 4.14 *Crs* Linkage Data*

	RF10c	***Crs***	J1k(scar)					
	Marker 1	Marker 2	Marker 3					
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	10	10	4	5	29	46.7	0	<1.0
M2-M3	18	0	2	19	39	2.7	31.5	<.001
M1-M3	22	20	10	13	65	42.9	0	<1.0

A=BAT B=JALO

*Linkage group 8 data contains <12 suitable markers for this study in the entire linkage group.

Table 4.15 *TubA1* Linkage Data

	D1468	***TubA1***	Bng205					
	Marker 1	Marker 2	Marker 3					
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	16	0	3	18	37	4.4	26.3	<.001
M2-M3	21	1	0	17	39	1.3	35.5	<.001
M1-M3	35	2	5	28	70	5.6	45.1	<.001

A=BAT B=JALO

Phaseolus Linkage Group 9 Markers:

Table 4.16 *Aatc* Linkage Data

	D1096-2	***Aatc***	Bng102					
	Marker 1	Marker 2	Marker 3					
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	14	3	2	18	37	7.8	19.5	<.001
M2-M3	15	4	4	15	38	13.3	11.1	<.05
M1-M3	21	9	9	25	64	19.6	7.4	<.1

A=BAT B=JALO

Table 4.17 *Pgip* Linkage Data

	D1096-2	***Pgip***	Bng102					
	Marker 1	Marker 2	Marker 3					
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	14	6	6	17	43	19.4	5.3	<.1
M2-M3	20	2	1	22	45	3.6	33.7	<.001
M1-M3	21	9	9	25	64	19.6	7.4	<.1

A=BAT B=JALO

Phaseolus Linkage Group 10 Markers:

Table 4.18 β*Fruct* Linkage Data

	D1580	**β*Fruct***	Bng68					
	Marker 1	Marker 2	Marker 3					
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	20	4	5	12	41	14.1	12.5	<.01
M2-M3	21	5	4	13	43	13.2	14.2	<.01
M1-M3	29	3	1	25	58	3.7	43.2	<.001

A=BAT B=JALO

Phaseolus Linkage Group 11 Markers:

Table 4.19 *Cipor* Linkage Data

	RON9b Marker 1	***Cipor*** Marker 2		D1291 Marker 3				
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	17	1	0	18	36	1.4	32.2	<.001
M2-M3	15	2	3	17	37	7.8	19.2	<.001
M1-M3	24	5	5	34	68	8.6	33.9	<.001

A=BAT B=JALO

Table 4.20 *Apx1* Linkage Data

	D1291 Marker 1	***Apx1*** Marker 2		RON9b Marker 3				
	AA	AB	BA	BB	TOTAL	*cM*	X^2	*p*
M1-M2	15	3	2	17	37	7.8	19.2	<.001
M2-M3	16	1	2	17	36	4.5	24.8	<.001
M1-M3	24	5	6	33	68	9.6	30.6	<.001

A=BAT B=JALO

The chi-square value was calculated as Total X^2 - X^2 due to segregation distortion.

Working Universal Gene-Specific Primer Pairs

The results for the universal primer pairs are provided below. Table 4.21 shows the universal primer pairs mapped in *Pisum*, *Lens*, and *Phaseolus*. Table 4.21 shows the genes that exhibited polymorphism between the two bean parents. A fragment size is shown for the amplified fragment in bean with all sizes shown are rough estimates and not exact. A method to detect sequence polymorphism is now needed. If a size polymorphism is present CAPS is not needed. The polymorphisms per number of restriction enzymes screened until a polymorphism was found is given as # Poly. R.E.

The linkage groups for pea, lentil, and bean are listed as LG species. The reproducibility of the universal primer pair to amplify the desired fragment is given at the far right.

Table 4.21 Universal Primer Pairs Mapped in *Pisum*, *Lens*, and *Phaseolus*

		Gene Product	Fragment Size	Mapping method	Differentiation	# Poly. R.E.	LG Pea	LG Lens	LG Bean	Reproducibility
1	Apx1_F2/R1	*Ascorbate peroxidase*	~1.4 Kb	CAPS	*DdeI*	2 of 3	VII	IV	11	Very Good
2	Pao_INT-F/R3	*Pheophorbide a monoxygenase*	~1.3 Kb	Size/CAPS	*HpychIV 4*	2 of 8	VII	IV	6	Very Good
3	Tkp_F/R	*Transketolase (plastid-specific)*	~1.25 Kb	CAPS	*HpychIV 4*	1 of 12	VII	IV	3	Fair
4	RCA_F2/R2	*Ribulose bisphosphate activase*	~1.9 Kb	CAPS	*Ava II*	1 of 27	VII	IV	1	Good-Fair
5	TuFM_F2/R2	*Mitochondrial elongation factor*	>2.5Kb	CAPS	*Hinf I*	1 of 2	V	V	1	Very Good
6	PaaI_F/R	*Phenylalanine ammonia lyase*	~1.6 Kb	CAPS	*Taq I*	1 of 15	V	V	1	Very Good
7	ENOL_F2/R2	*Enolase*	~1.2 Kb	CAPS	*RSA I*	*1 of 5*	V	V	1	Very Good
8	Apy_F2/R2	*Apyrase S-Type*	~1.4 Kb	CAPS	*Alu I*	1 of 27	V	V	3	Good
9	CRS_F/R	*Cystathionine-gamma-synthase precursor*	*~1.5 Kb*	CAPS	*Hae III*	*1 of 3*	V	V	8	Good
10	Rcpme_F/R	*Pectin methylesterase*	*~650bp*	Size	*Jalo=slow*	*na*	V	V	2	Good
11	FK_F4/R2	*Fructokinase*	*>2.1Kb*	CAPS	*Hinf I*	*3 of 12*	III	III	5	Good
12	Gpt_F3/R2	*Glucose 6-phosphate translocator*	~2Kb	CAPS	*Hinf I*	1 of 3	III	III	6	Good
13	Cipor_F2/R2	*Protochlorophyllide reductase*	~1.2 Kb	CAPS	*Hae III*	2 of 4	III	VI	11	Very Good
14	Aat_F/R	*Aspartate aminotransferase (cytosolic)*	~1.5 Kb	CAPS	*HpychIV4*	1 of 10	III	II or VII	9	Good
15	BFruct_F2/R2	*Beta-Fructose(invertase)*	*~1.1 Kb*	CAPS & (+/-)	*Alu I*	*4 of 4*	III	III	10	Very Good
16	TubA1_F/R	*Alpha tubulin*	*~2.5 Kb*	(+/-)	*Jalo=+*	*na*	III	III	8	Good
17	ALDH7_F1/R	*Aldehyde dehydrogenase (family 7)*	*~1.75 Kb*	CAPS	*BstO1*	*1 of 25*	?	?	4	Very Good
18	GRF3_F1/R1	*Growth regulating factor 3*	*~880bp*	CAPS	*HpychIV 3*	*2 of 27*	Monomorphic	Faint	2	Very Good
19	Hsp70_F1/R1	*Heat Shock Protein 70kD subunit*	*<2.5 Kb*	(+/-)	*Jalo= -*	*na*	V or VII	Faint	*3 & 8	Fair-Poor
20	PGIp_F/R	*Phosphoglucose isomerase*	*~1.6 Kb*	CAPS	*Hinf I*	*1 of 10*	?	?	9	Good
21	UNK21_F/R	*Unknown protein 21*	*~1Kb*	CAPS	*RSA I*	*1 of 3*	Faint	?	4	Good

* = An HSP70 already mapped to LG6 in bean.

Markers 17-21 are primer pairs only found to be polymorphic in bean by the CAPS analysis.

The universal primer pair for *Apx1* was based on many plant sequences. The *Apx1* marker showed linkage to already mapped RFLP markers, D1291 and 1308, on bean LG 11. *Apx1* showed strong linkage to D1291, but weak linkage to D1308. The segregation pattern dictates that D1308 is the nearest RFLP marker to that side of *Apx1*. *Cipor* also mapped to this linkage group and is not syntenic in pea or lentil with *Apx1*. *Cipor* showed linkage to a similar marker (D1308) as *Apx1*, and also RON9b. Multiple *Apx* genes can be found in a single plant genome, so this could be an alternate *Apx* and not the ortholog of *Apx1*.

The universal primer pair for *bFruct*, an invertase, was based on many plant sequences. *bFruct* showed linkage to markers D1580 and Bng 68 on LG 10 and was the only universal marker that mapped in LG 10.

The universal primer pair for *PGIp* and *Aat* was based only on legume sequences. *PGIp* and *Aat* showed linkage to D1096-2 and Bng 102 markers on LG 9, yet *PGIp* has been mapped to LG 2. The linkage data suggests that *Aat* is closer to D1096-2 and *PGIp* is closer to Bng 102. The universal primer pair for *PGIp*, an enzyme that catalyzes the reversible isomerization of glucose-6-phosphate and fructose-6-phosphate and *Aat*, Aspartate aminotransferase (cytosolic), amplified efficiently.

The universal primer pair for *CRS* was based on many plant sequences. The universal primer pair for *TubA1*, the α-subunit in tubulin, was a pair of primers based on many plant sequences. RF10c and J1k (scar) on LG 8 showed linkage to *CRS*. *TubA1* showed linkage to two different markers on LG 8, D1468 and Bng 205.

The universal primer pair for *Pao* was based on many plant sequences. *Pao* showed strong linkage to D0096 on LG 6 and to the flanking marker ROD3b. Multiple *Pao* genes cannot be found in a single plant genome that we have tested, so if this is not *Pao* it could be a member of the same gene family (ex. *PTC52, Cao, or RCC* described in Ch.2). Extensive testing of the *Pao* gene makes it highly unlikely that the designed primers match a different gene. The universal primer pair for *Gpt* was based on two pair of plant sequences. *Gpt* showed strong linkage to D0096 on LG 6 and to the flanking marker ROD3b. *Gpt* mapped tightly to this linkage group and is not syntenic in pea or lentil compared to *Pao*.

The universal primer pair for *FK*, a fructokinase, was based on multiple legume plant sequences. *FK* has multiple primer pairs listed in Appendix A that span multiple regions and the majority of the gene, all appear to amplify well and are polymorphic in our tested species. D1251 and D1301 showed tight linkage to *FK*, the only universal marker mapped to LG 5.

The universal primer pair for *Tkp*, a transketolase (plastid-specific), was based on several plant sequences. *Tkp* showed strong linkage to D1291, but weak linkage to D1308 (described above). The universal primer pair for *Apyrase* was based on many plant sequences and mapped strongly to LG 3.

The universal primer pair for *Rcpme* was also based on many plant sequences. The universal primer pair for *GRF3*, a transcription factor affecting plant growth, mapped on linkage group 2 and was a pair of primers based on one plant sequence from *Arabidopsis*.

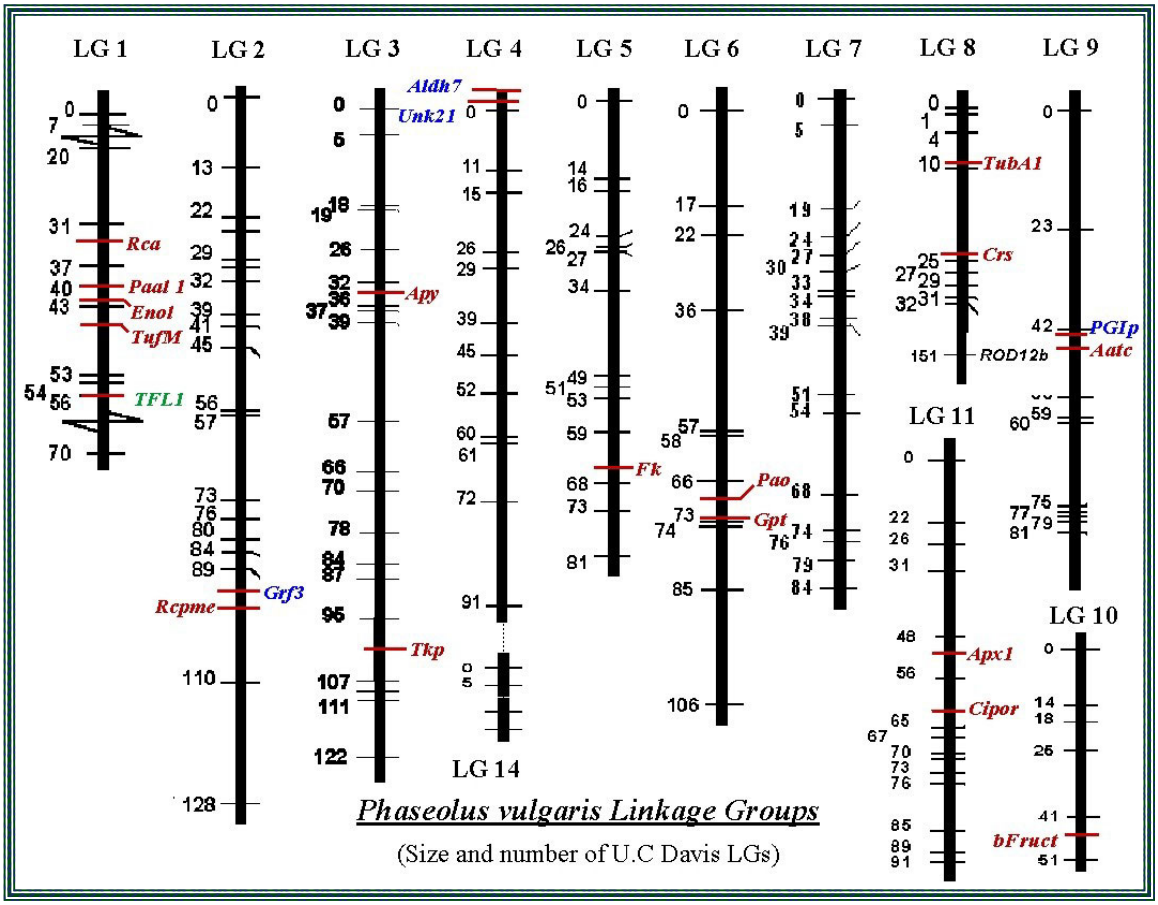
The universal primer pair for *RCA* was based on many plant sequences and mapped to LG 1. *Paal, TufM, and Enolase* also showed stronger linkage to D1291, but weak linkage to D1308. The universal primer pair for *Paal* was based on many plant sequences (*Paal 1-3*) like the universal primer pair for *TufM* and *Enolase*. The *Enolase* marker showed linkage to *Pal* already mapped and to two RFLP markers, D1862 and 1662. *TufM* also mapped to this linkage group and appears to be syntenic in pea or lentil with other markers, the observed conservation of synteny will be discussed more in Chapter 5. Multiple *Pal/Paal* genes can be found in a single plant genome, so this could be an alternate *Pal* and not the ortholog of *Paal1,2*.

Some of the universal primer pairs did not produce PCR products in all three species. This paragraph describes some genes added to the *Phaseolus* linkage map, but not to the pea and lentil maps. *ALDH7*, *UNK21*, *PGIp*, *GRF3* and *Hsp70* were only mapped in the *Phaseolus* mapping population. Limited time has prevented mapping of *ALDH7*, *PGIp*, and *GRF3* that have not exhibited polymorphism in the several CAPS screens for pea or lentil parents. *UNK21* was designed to an unknown coding sequence found in *Medicago* and amplifies small non-specific bands, all other universal markers used three or more genomic DNA or cDNA sequences. *Hsp70* has been mapped in pea and lentil, but this primer pair maps to a different location than published in the bean map. *Hsp70* is mapped in pea, but appears to have two separate copies in the genome. The analysis of this research did not determine which of the two genes was mapped, so it was included in this category. The gene *PGIp* is also included, because the location of the gene using this set of primer pairs has not been verified in pea as PGIp (plastidic) and not PGIc (cytoplasmic), eventhough one of the primers was designed to be PGIp by Dr. Weeden by using several legume sequences. The universal primer pair for *ALDH7*, an enzyme in the 7th family of aldehyde dehydrogenases, was based on many plant sequences and mapped near *UNK21*. The map positions of *UNK21* and *ALDH7* both appear to be outside the boundaries of bean linkage group 4. One marker for an aldehyde dehydrogenase, *ALDH7*, mapped to an area of high interest, due to linkage only with this region in the genome. *ALDH7* and *UNK21* mapped to the end of bean linkage group 4 extending the length of the linkage group past *Me* another 20cM.

Under a small sample size of recombinant inbred lines RILs (~40) there are slight linkages, but based on the segregation patterns and possible double recombinants

between the genes; they can only fit at this location and are unlikely to fall into a gap in the linkage map. Distances shown in Figure 4.13 below are relative and not absolute.

Figure 4.13: BAT x JALO bean map (M.S.U *Phaseolus vulgaris* Map)



Non-Working Universal Gene-Specific Primer Pairs

The Appendix A shows a list of possible primer pairs that either produced a polymorphism, strong PCR amplification and were monomorphic, multiple bands, or very weakly amplified PCR products, meaning the further development of these as universal markers could be done with more time.

Discussion

The research data presented in this paper shows 21 working gene-specific markers that nearly doubles the number of known genes in common bean. The research shows the robustness of the U.C. Davis linkage map, but also the technique used to achieve universal gene-specific markers containing the highest amount of polymorphism within the bean genome.

Greater than 90% of the 21 mapped markers could be visualized after a restriction digest with 5-8ul of the PCR reaction mix, due to the high amount of PCR product.

The CAPS screen, not only identifies polymorphism between parents, but also the verification of homology. The amplification of homologous fragments can be shown by the banding pattern for the different restriction enzymes with the PCR product. In many cases the banding patterns were similar and size polymorphisms could be identified to suspected fragments, while certain restriction enzymes might cut pea and/or lentil once and in bean the restriction site might be lost, or vice versa. The presented evidence is not as rigorous as DNA sequencing.

The case I concentrated on using, previously mapped pea STS markers or widely investigated genes, made it possible to identify specific genes that show conserved synteny. DNA markers that map to non-conserved syntenic regions could be non-orthologous sequences, representing possible duplicated genes or genes of related ontology, but the markers that do show conserved synteny are very likely to be the orthologous sequence. Mapping of the universal primer pairs in more legume species,

coupled with sequencing, will verify the exact genetic composition of each marker and will still provide very efficient markers within a different species.

CHAPTER 5

IDENTIFICATION AND ANALYSIS OF SYNTENY

Introduction

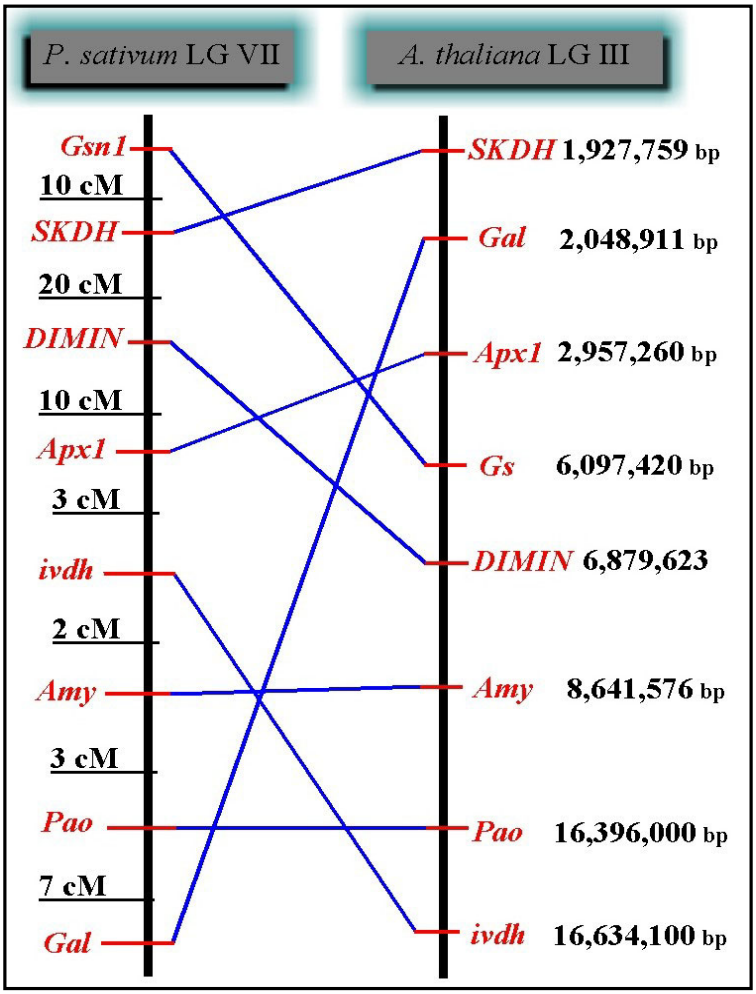
Identification of conserved synteny has obvious benefits in understanding the genetics of legumes. Identification of synteny is revealed by the mapping of orthologous loci between individuals and the orientation of those loci. Commonly the mapping of homologous markers can refer to microsynteny (an example is loci found on a corresponding BAC) or macrosynteny (an example is loci found on corresponding chromosomes). Depending on the level of marker saturation and sequence availability studies could show a mix of macrosynteny and microsynteny based on what information is found. In certain comparisons BAC sequencing projects facilitate microsyntenic studies; where species without large scale sequencing projects can only look for macrosynteny or the hybridization to BAC libraries. Markers dispersed across the chromosomes and the genome will lead to a macrosyntenic study. A lack of suitable in-house markers prevented a genome wide comparison between pea and bean, or *Medicago* and bean, so all recently in-house designed pea markers were screened in bean. The results of the screened STS markers was described in chapter 4, where this chapter will conclude upon the gene order of the mapped genes. Chapter 5 will further examine the conservation of synteny between pea & bean and extrapolate these results to other legumes and the genomic model species *Arabidopsis*.

Materials and Methods

Regions of Conserved Suspected Synteny

An early initiative in this study was to take previously known synteny in other species to estimate regions that might show conservation of synteny. *Pao* and *Paal* were the first genes to be mapped in all three species. As a preliminary step to mapping, DNA markers expected to have synteny conservation between *Pisum* and *Arabidopsis* were explored to identify possible homologous markers between *Pisum* and *Phaseolus*.

Figure 5.1 The *Pisum* LG7 lower arm region.
Identification of possible homologous loci in pea and thale cress by NCBI Mapviewer, genes showing conserved synteny between pea and lentil were queried to find proper genes (some nomenclature differences).



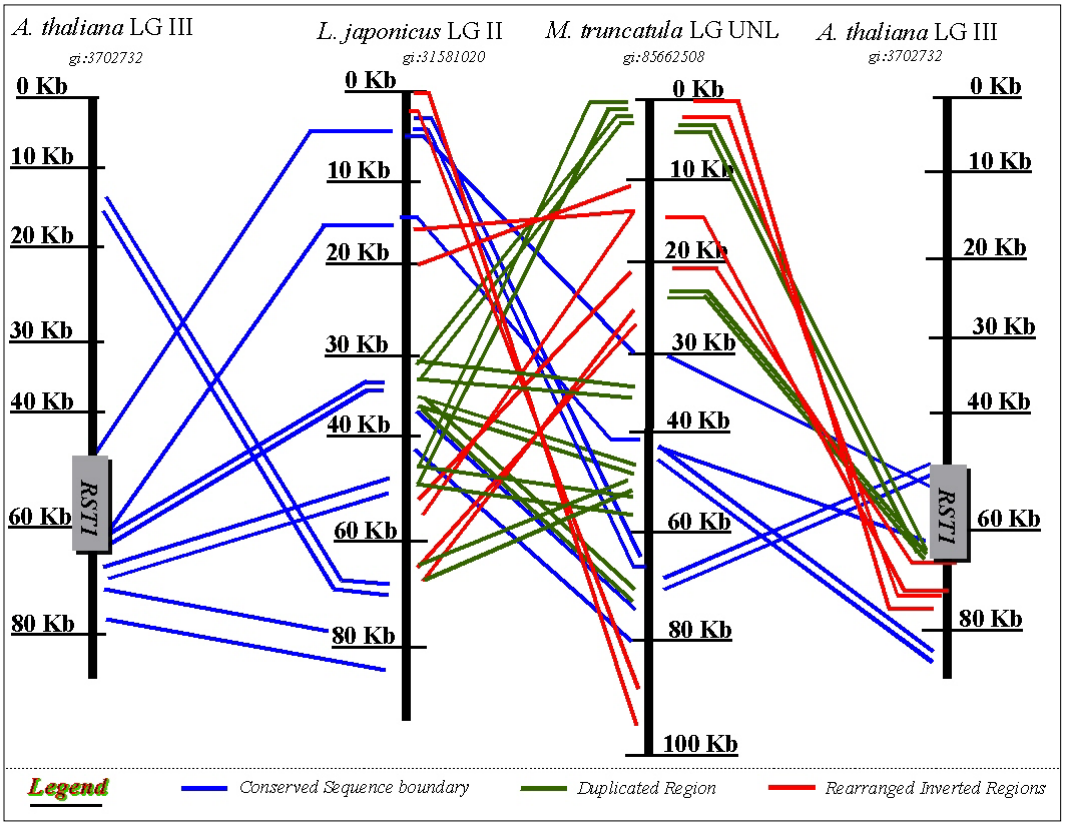
Conservation of synteny between *Pisum* and *Arabidopsis* has been implied, but not to the continuous size identified and shown in Figure 5.1. *Pao* showed linkage to several genes on pea LG VII and Arabidopsis LG III, correlating with pea (Figure 5.1) and was the basis for starting at this point. In *Pisum* we can only report recombination distances whereas in *Arabidopsis* we can report the approximate distance on the chromosome. The comparison above involves half of pea LG VII and nearly all of *Arabidopsis* linkage group 3.

Other cases were explored, such as identifying microsynteny in *Medicago* and *Lotus* and then screening those regions in pea, lentil, and bean for conservation of synteny. Regretfully, this method was not possible until later in the research. Early in the research the candidate BACs were executed through a lenient BLAST and only rarely contained multiple well-described genes to target with the CAPS approach. The method of microsyntenic comparisons between *Medicago* and *Lotus* will be accelerated with the completion of the respective BAC sequencing projects and then all analysis will be by BLASTing BACs for potential homologs.

Below is a figure describing a case of microsynteny with a newly found gene in *Arabidopsis*, called the resurrection gene (*RST1*). It was of interest due to its large size (~10Kb) and lack of identification in any other species to date (Chen *et al.*, 2005). To identify microsynteny by DNA sequence the BLAST utility, BL2SEQ (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi) was used with less stringent parameters then the standard presets. The size of the *RST* gene appears to have influenced the amount of synteny conservation in the *Medicago* and *Lotus* BACs. The two entire BAC sequences were imported into BL2SEQ and parameters were set with

BLASTn: match +1, mismatch –1, open gap penalty 2, extension gap penalty 2, expect 100, word size 7. Pair-wise alignments retrieved were screened by eye for sequences showing >60% similarity for a continuous 60 bp or more, based on the amount and percentage of similarity in the comparison, we can speculate that this region is actually homologous and many identified alignments exceed the basic requirements to label them as microsyntenic. The map location for this region is known in *Arabidopsis* and *Lotus*, but not at this time for *Medicago*. Current research on the conservation of synteny between *Lotus japonicus* and *Medicago truncatula*, speculated to be quite high in similarity, will identify the location of many genes. When a locus is known in one species we can estimate its location in other legume species (Cannon *et al.*, 2006).

Figure 5.2 Microsyntenic Comparison by BLAST: The RST1 Gene is Conserved.



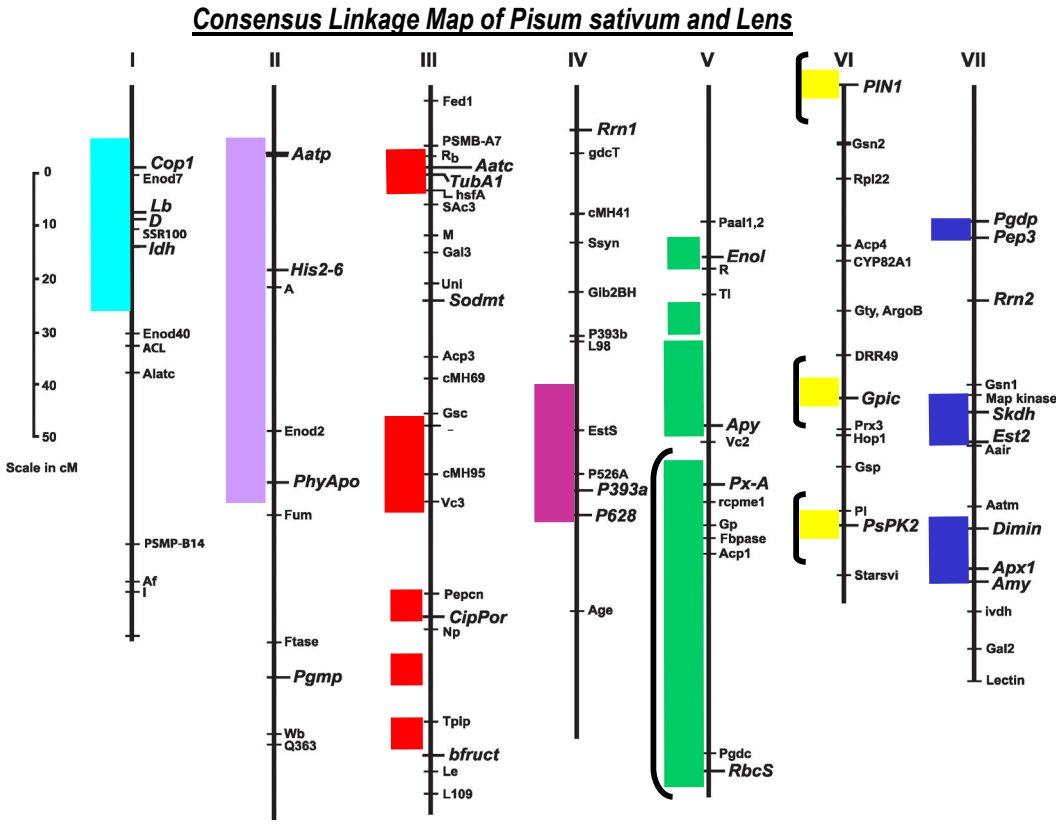
The primer screen of RST1 showed multiple bands with one band at the approximate expected size in the pea and lentil PCR reactions, not in bean. The fragment generated by primers RST_F (5' TRTTCARGAAATGRRACTC 3') and RST_R (5' TCAATRAGCCACTGTCT 3') was approximately the same size as predicted from the *Arabidopsis* gene model, ~2 Kb. This analysis was done very late in the research and new universal primers have yet to be designed. The lines shown in the figure below depict the 5' and 3' boundaries of similar DNA sequences as described above.

Dependent on the stringency of the BLAST parameters we can only identify sections of significantly similar DNA and classify it as homologous. If we look at the probability of acquiring a specific DNA sequence, as a function of the length, we should conclude the necessary requirements to be >50 or 100 bp to be identified as likely syntenic. This is based on the probability of acquiring a specific set of 50 random nucleotides within a genome. Since each nucleotide has a 25% or ¼ probability of randomly occurring, then 50 nucleotides has the theoretical probability of $(¼)^{50}$ or $7.88860905 \times 10^{-31}$. Meaning within the bean genome (~588 Mb) we assume a specific DNA sequence will randomly occur once in the genome if the probability is 5.88×10^{-8}. The size of the pea genome (~4100 Mb) means a specific DNA sequence will randomly occur once in the genome if the probability is 4.10×10^{-9} ($¼^{14}$ nucleotides = 3.7252903×10^{-9}). Identification depends on the similarity of the functional coding in the DNA sequence and its position to other syntenic sequences. We are only comparing conserved sequence in the described manner, so we can speculate on observed homology.

Regions of Conserved Synteny Between *Pisum* and *Lens*

In chapter one we described previous research on the pea-lentil comparison. Previous research has shown ~50% conserved synteny between pea and lentil as shown in the figure below.

Figure 5.3 Consensus Linkage Map of *Pisum sativum* and *Lens* (Walling, 2003)



After J. Walling finished his work I have mapped many markers in the *Lens culinaris* 7x *Lens ervoides* 32 population, but only the primers designed to be universal have been further mapped in bean. *Lens* linkage groups correlating to pea linkage groups have been extended on LG I, II, IV, and V gaps in the linkage groups were closed on III, V, and VII. By mapping additional gene-specific markers in *Lens* the estimate of synteny

conservation is now approaching 70%. The *Pisum-Lens* comparative map contains <100 DNA and morphological markers that could be homologous.

Regions of Conserved Synteny Between *Pisum* and *Lens* to *Phaseolus*

The process to determine the conservation of synteny between pea/lentil and bean was a visual process of aligning an illustration of one linkage group to another and looking for the same genes to occur on both linkage groups. Of course, the locations of the genes must already be known. Development of other visual methods such as dot plots (diagonal line comparison) and CMAP (computer module) can display the relationships of gene order, just as sketches or manual illustrations/figures. Illustrations below (Figures 5.4, 5.5, 5.6) will show the linkages in *Phaseolus vulgaris* associated with genes already mapped in certain *Pisum* linkage groups.

The available STS markers from pea and the universally designed primers were first tested from linkage groups (LG) VII, then LG III, and LG V. Primer pairs from the other linkage groups were tested and either found to be monomorphic or the PCR did not amplify a single band of the appropriate size.

Results

Conserved Linkage Between Pea, Lentil, and Bean.

The results of mapping 21 gene-specific primer pairs in *Phaseolus vulgaris* resulted in 16 markers mapped in all three papilionoid species (pea, lentil, and bean). An additional benefit is that many of these markers are already mapped in other species and will benefit future comparisons on gene order in the tree of life. PCR amplification of 43

genes with >98% giving single bands produced 21 polymorphic fragments and 22 monomorphic fragments. PCR screening occurred on 136 primer pairs and 100 genes, many of which came from the labs supply of previously designed primers. Mapping of the 21 polymorphic genes in bean out of 43 amplified is a 49% success rate, but only 21.0% of the total genes tested from pea and lentil. Comparing the total number of universally designed primers to the number that mapped, 19 out of the 21 mapped genes that produced a good PCR fragment, thus equating to 19 mapped out of ~30 designed to multiple species. Some primer pairs in pea were designed using a different method and could be considered universally designed as the process described within. Availability in the growing number of DNA sequences has increased the number working universal markers in this study. Some markers were designed to one or two sequences and were not included in the calculations for universally designed primers.

The 21 mapped markers will eventually be mapped in other species and can be useful in future comparative studies. Three markers on the bean map can also be included in our comparison *Pal1, SKDH, TFL*. *Pal1* (*Paal1,2* in pea) was one of the first markers tested, because it was already on the bean and pea maps it was used for verification of the universal primer pair mapping a homologous gene in pea/lentil and bean (Freyre *et al.*, 1998; Kuznetsova *et al.*, 2005). SKDH is an isozyme marker previously mapped in all three of our mapping populations. We know there is conservation of synteny between *Arabidopsis*, pea, and lentil in regards to markers flanking SKDH and expected bean to show conservation of synteny (Figure 5.1). Even after the present study we still lack evidence in bean for synteny with SKDH. The only

possible exception is the weak linkage to *Tkp*, which has incomplete data (poor amplification) and located >20 cM away on the linkage map.

Synteny was found on pea LG V consisting of *TFL* (*fin* & *det*), *Enolase*, *Paal*, and *TufM* markers between pea/lentil and bean spanning ~10-20 cM. It was only recently that *fin* in bean was described as a *TFL* gene controlling determinacy (Kwak *et al.*, 2006). *Paal* does show synteny conservation with *TFL/det* in pea and *Arabidopsis*, but breakpoints occur with the other genes in pea LG V and in the end of *Medicago* LG V. Examining the location of these genes in *Medicago* and *Arabidopsis* we can see breakpoints in the conservation of synteny, markers are too distant for accuracy, but breakpoints are at least placed between two specific loci in the linkage group. Previous research says that *Pisum* linkage group V and *Medicago* 7 are highly conserved (Kalo *et al.*, 2004), but my results cannot confirm these or other previous results (Choi *et al.*, 2004). The linkage analysis performed with the markers on *Pisum* linkage group V shows they cannot be conserved on the described linkage group end in *Phaseolus vulgaris*. These areas could be targeted by further microsyntenic studies to elucidate a more accurate determination of the breakpoints and the regions truly conserved from the common progenitor ancestor.

The following sections show by illustrations the locations and associations of genes mapped in pea and bean. Figures 5.4, 5.5, and 5.6 all have the pea linkage groups on the left side and used as the species to be compared against with bean, so each figure has only one pea linkage group and one to many linkage groups for bean. When an approximate distance could be obtained for markers showing synteny, then it is given in cM parallel to the depicted linkage group. In many cases there will be marker names on

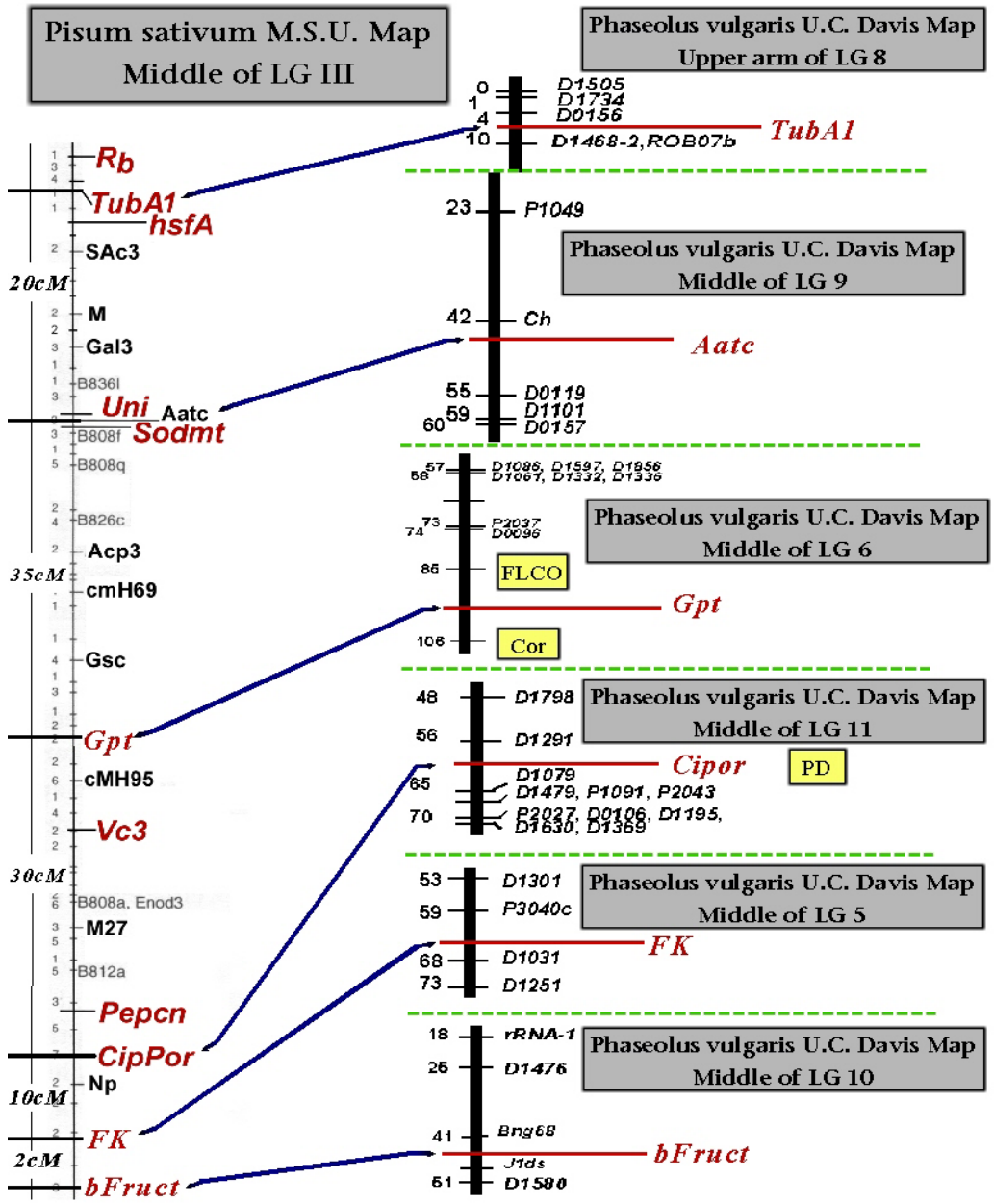
the maps that we have not discussed here, they are RFLP, RAPD, and STS markers used in construction of the individual maps and also used in the linkage analysis. The *Phaseolus vulgaris* linkage groups are on the right side of the figure and only show specific segments of linkage groups. Within the bean linkage groups RFLP and SCAR markers are listed in black, the morphological, disease and QTL markers are shown in yellow boxes. The universal markers from this study are indicated by long blue lines connecting the pea and bean linkage groups. It is also important to note that all the universal markers are mapped in pea, lentil, bean, and in some cases other legumes and plant species.

Comparison of *Pisum* Linkage Group III and *Phaseolus* Linkage (see Figure 5.4)

Six gene-specific markers were evenly spread across three-fourths of pea linkage group III and all produced single bright bands, except *bFruct* where the JaloEEP558 allele produced a secondary band that co-segregated with the CAPS segregation pattern. Two of the markers (*TubA1* and *Aatc*) were found by screening STS primer pairs mapped in pea. Three markers (*Gpt*, *FK*, and *bFruct*) were designed by rearchers in the lab and were based on *Arabidopsis*, *Medicago* and a *Lotus* or soybean DNA sequence or less (Weeden *et al.*, 1998). One universal primer pair (*Cipor*) was designed using six DNA sequences when the previous primer pair from pea showed poor reproducibility.

Each of the *Pisum* linkage group III markers were mapped in *Phaseolus* and they all mapped to different bean linkage groups (5, 6, 8, 9, 10, 11). *TubA1* at the top of PsLG III shows strong linkage to markers in PvLG 8. *Aat, Gpt, Cipor*, and *FK* also showed strong linkage to different linkage groups and produced a bright single PCR product.

Figure 5.4 Comparison of *Pisum* Linkage Group III to Bean Linkage Groups.



Even the two DNA markers *FK* and *bFruct* that were only 2cM in pea, were not linked by going to the smaller genome of bean, a probable result due to multiple *FK* and *bFruct* genes in the genome found by query of NCBI entrez and *Arabidopsis* map viewer. Gene

families will probably be the most difficult aspect of creating a comparative genomic analysis on legumes.

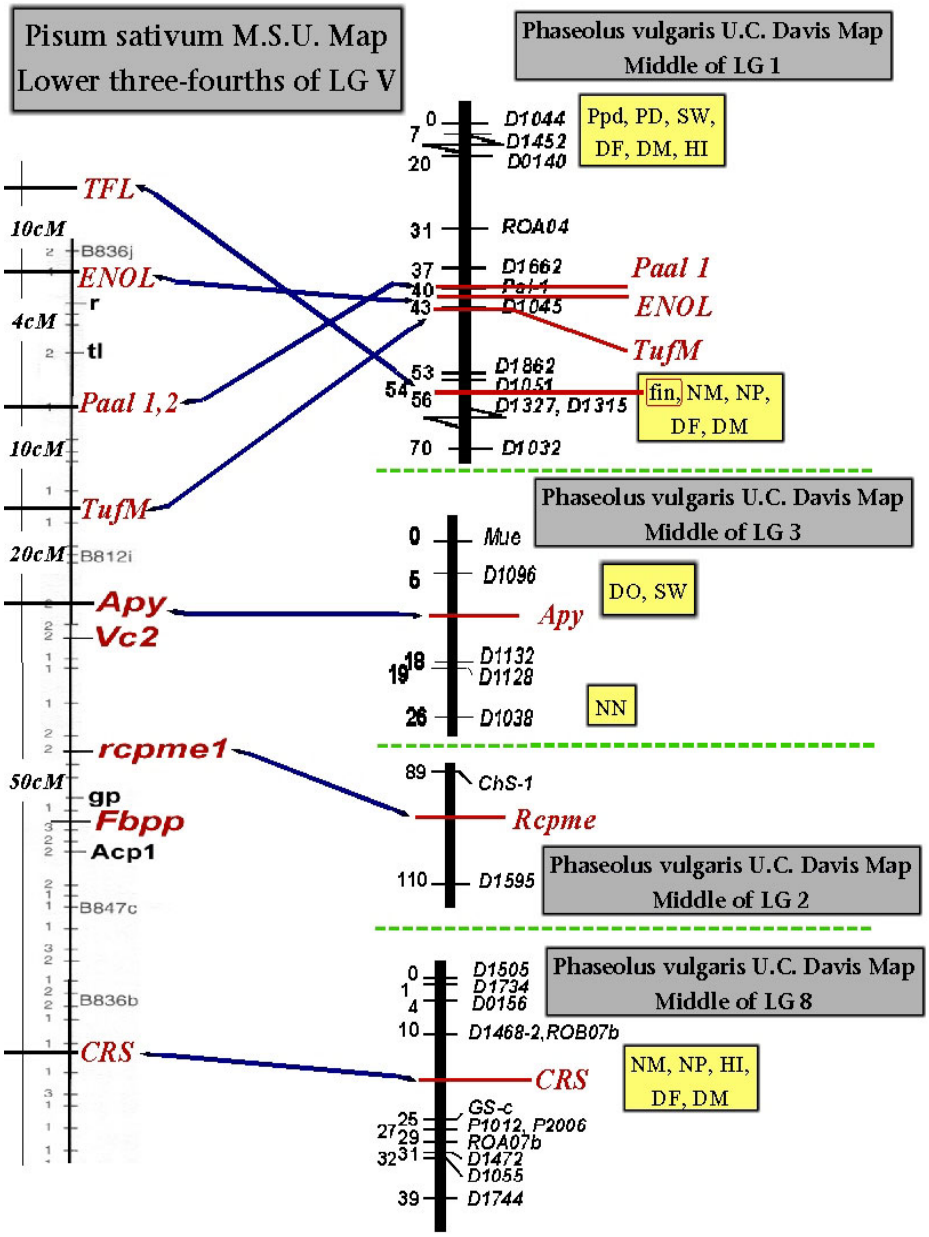
Comparison of *Pisum* Linkage Group V and *Phaseolus* Linkage (see Figure 5.5)

The universal gene-specific markers showing synteny grouped at the top half of pea linkage group V. An additional two markers mapped to other linkage groups. *CRS* was located at the very bottom of pea linkage group V, *Apy* in the middle, were genes previously mapped by Dr. Weeden. When the pea linkage group V markers were mapped in bean three mapped to bean linkage group 1 as a cluster, the same as pea, and 2 universal markers mapped to two different linkage groups. A lack of more than two markers limits the approximation of a syntenic break-point, more markers are needed in the vicinity, *Apy* only places the break-point within 20 cM with no other verification.

TFL (fin) appears orthologous to a *TFL(det)* in pea and *TFL1* in *Arabidopsis* and is un-identified as syntenic, including the exclusion of this region in the *Medicago-Arabidopsis* comparison (Zhu *et al.*, 2003). *Apy* and *Tkp* showed weak linkage to bean linkage groups 8 and showed strong linkage to linkage group 3, while producing a bright single PCR product.

Relating genetic distance in pea to bean does not correlate based on only one region of synteny, rearrangements in the gene order, rearrangements in the individual species, overall lack of genome-wide linkage data, and the huge difference in genome size between pea and bean.

Figure 5.5 Comparison of *Pisum* Linkage Group V to Bean Linkage Groups.

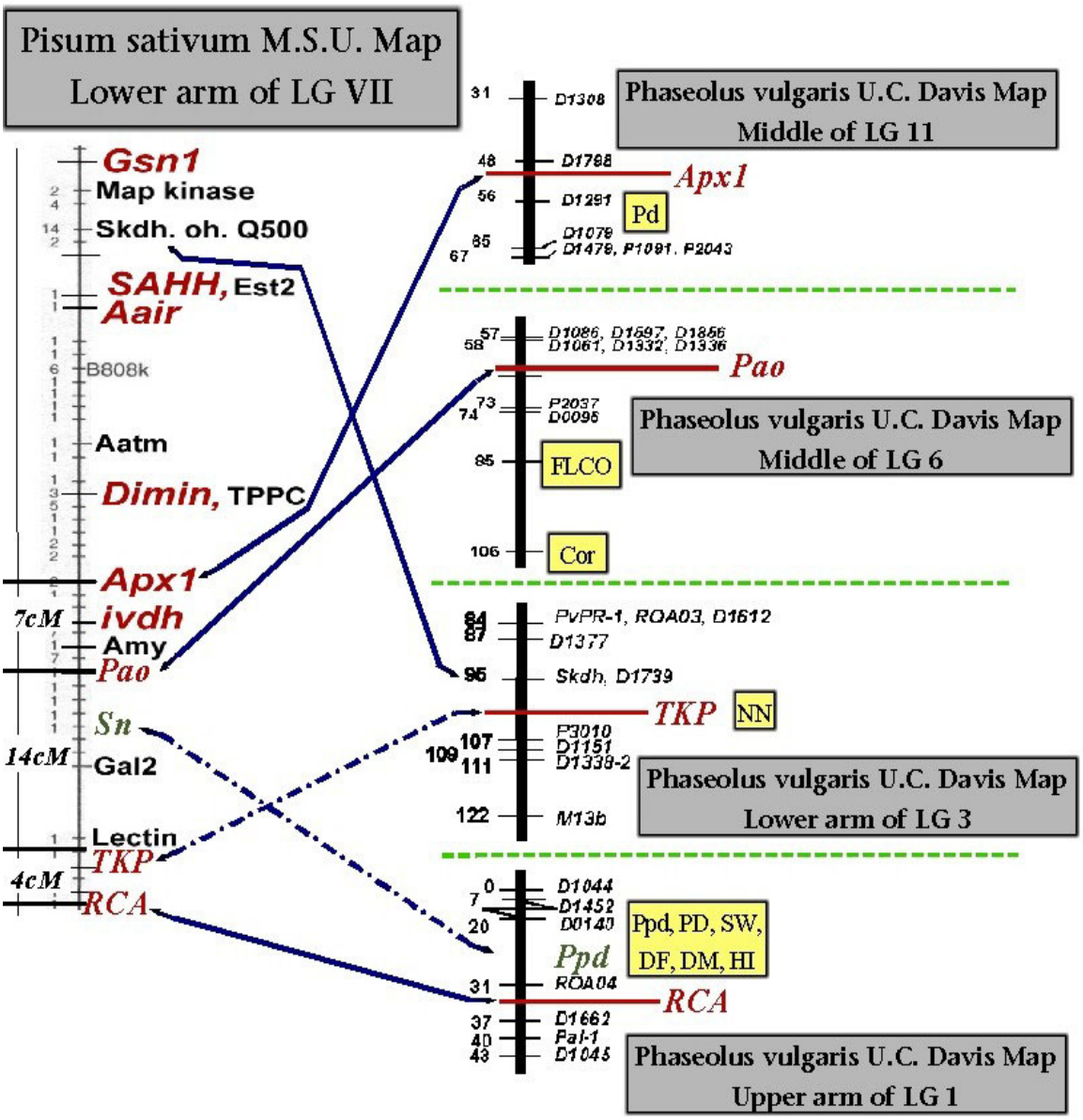


Through personal correspondence at the Plant and Animal Genome Conference (XIV) with Myounghai Kwak of the P. Gepts lab at U.C. Davis, did I become aware they had mapped multiple *TFL* genes in bean. One of the TFL genes had co-segregated with *fin* (determinancy gene in bean) and only later noticed *det* (determinancy in pea) showed linkage to the same markers already showing conservation of synteny.

Comparison of *Pisum* Linkage Group VII and *Phaseolus* Linkage (see Figure 5.6)

Gene-specific markers were grouped on pea linkage group VII, including *SKDH* an isozyme already mapped, and STS markers. Universal markers produced very weak bands and all other markers produced bright monomorphic bands or no proper band.

Figure 5.6 Comparison of *Pisum* Linkage Group VII to Bean Linkage Groups.



When the PsLG VII markers were mapped in bean they all mapped to different *Phaseolus* linkage groups, like the majority of the other markers. *SKDH* synteny could not be efficiently tested without other close markers. *Apx1*, *Pao*, and *RCA* also showed strong linkage and produced a bright single PCR product. The same correlation was tested, based on the belief that short genetic distance increases the chance of finding synteny, but no synteny was found with this group of markers, except *Tkp* and *SKDH*. *Tkp* produced very poor data so this linkage is not considered to be real until further data verifies this conclusion.

Ppd is the major photoperiod gene in bean, where in pea *Sn* is the major photoperiod gene. The strong linkage data of *RCA* and *Sn*/*Ppd* mean that the major photoperiod genes in pea and bean might be orthologous to one another.

The primer sequences that successfully amplified a product of the expected length in pea were given in Tables 2.3 and 2.4 (Monomorphic and Polymorphic). In almost every case the initial primers worked well; however, in several cases (e.g. *GAPA, Apx, Pao*) several sets of primers were tested before an appropriate combination giving the best amplification was obtained. Linkage data was provided by P. Gepts of U.C. Davis for incorporation of our new 16 markers by CAPS analysis.

The method for scoring genotypes, the expected fragment size, the restriction enzyme used and the linkage group on which the locus was placed can be found in Table 4.21. Table 4.21 gives a good comprehensive view of the major results for analysis of synteny. Chapter 2 displays the segregation of a polymorphism in pea and bean for the sequence encoding pheophorbide a oxygenase (*Pao*) and is used as an example for the universal primers designed. In the case of *Pao*, the genomic sequence in the pea and

lentil parents contained a size polymorphism of equal size, whereas *Phaseolus vulgaris Pao* needed a CAPS polymorphism for segregation analysis. The restriction digest screen confirms that the bean PCR product is more than likely *Phaseolus vulgaris Pao*.

We found that >50% of the gene primer combinations tried, previously working in pea and lentil, would produce a band in bean. In contrast, multiple primers pairs we previously designed to different regions of the *Cop*1 gene sequence, and all gave satisfactory amplification in pea and lentil, but gave no amplification in bean. *Enolase* amplified with the second primer pair like many other primer pairs that mapped in bean (Figure 5.5). In the Table 4.21 we only report the primer pair that generated a fragment in pea, lentil, and bean while also being polymorphic and robust. Chapter 5 describes the major results of how the genes in Table 4.21 were located within the bean genome.

The identity of *Phaseolus*-amplified products was confirmed by restricting the fragment with as many endonucleases needed (27 total) to find polymorphism and comparing the size of the resulting subfragments with those obtained from *Pisum* and *Lens*. *GAPA* and *PI4P* were also screened with Surveyor nuclease from Transgenomic Inc. and were still monomorphic.

Among the 16 homologous genes, many are spread across the genome in bean. *TKP* and *RCA, Apx1* and *Pao, Paal and Enolase* are all linked by <10cM in pea and lentil, but only *Paal* and *Enolase* were conserved in this <10cM range of bean. The only marker >10cM that was syntenic was *TufM*, next to *Paal* and *Enolase,* the only region of conserved synteny that contained more than two conserved markers. These results clearly establish one area of linkage conservation between pea, lentil, and bean. There are most likely other regions, but this is the only one we have multiple markers in the

region and weak linkages were excluded. The presented data does not support the hypothesis that synteny in the legumes is highly conserved as described in other research (Choi *et al.*, 2004, Kalo *et al.*, 2004).

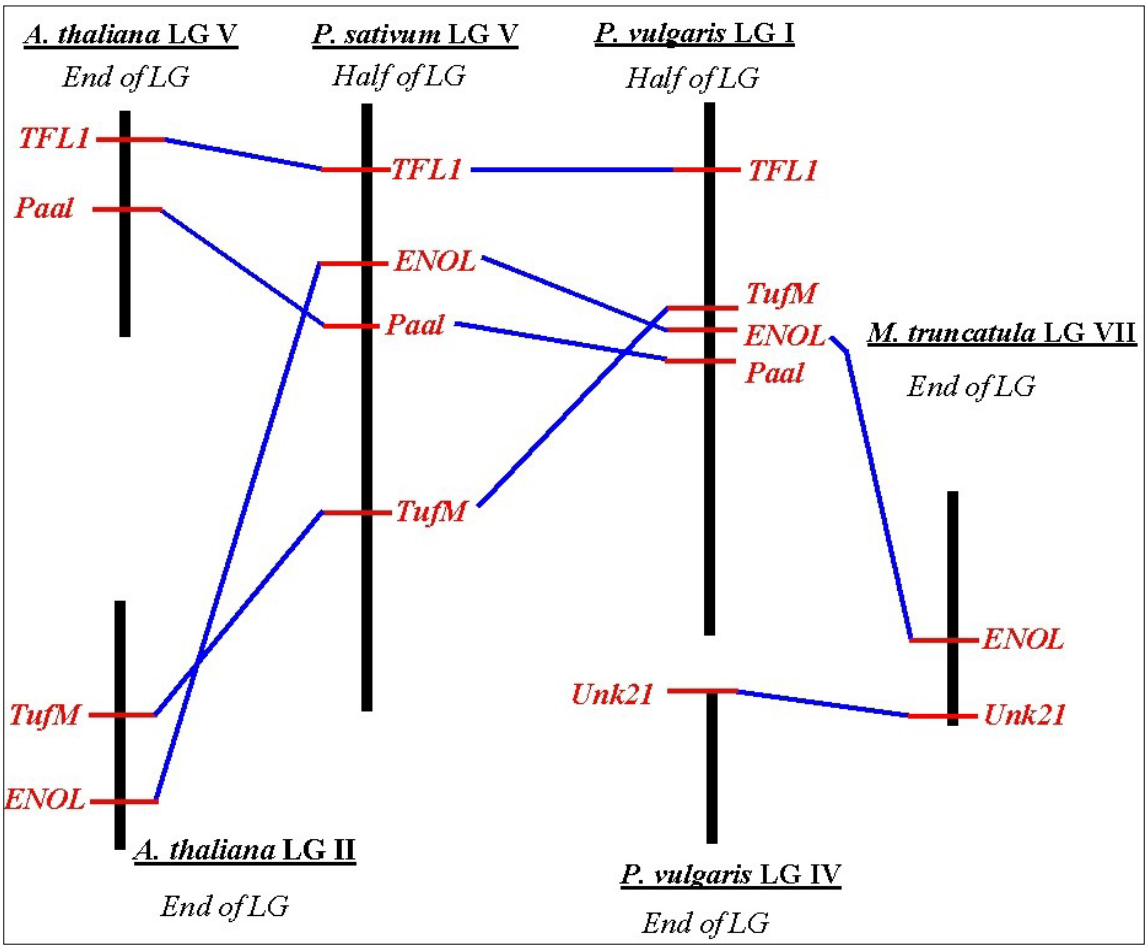
Discussion

The approach we have used to locate genes on the pea, lentil and common bean linkage maps appears to be a method applicable in most species. The 16 genes listed in the Table 4.21 represent less than one-fourth of the genes we have attempted to map using a CAPS analysis. Our success rate in lentil has been greater than 90% when a sequence from pea is available. In common bean success rate is closer to 25% screening primers developed for pea, where primers designed to be universal in legumes and other species are ~50%. Extrapolating from soybean or *Arabidopsis* sequences alone has proven more challenging, with a success rate of much less than 50%. Conservation of gene sequence appears to be greater between *Medicago truncatula* and other legumes than soybean with pea and lentil.

The growing number of gene sequences available; indicate that this approach will provide a sufficient number of markers for comparative analyses in *Pisum* and *Phaseolus* macrosynteny. It has also become apparent that the greater number of sequences used in the primer design means an increased chance of PCR success. By acquiring new gene sequences from the sequencing projects of *Arabidopsis thaliana*, *Medicago truncatula*, *Lotus japonicus*, *Glycine max*, and *Phaseolus vulgaris* the design of universal primers is possible and becoming more feasible than without the sequence information from legumes.

Figure 5.7 Comparison of Conserved Synteny between *Arabidopsis, Pisum, Lens, Medicago* and *Phaseolus.*

*The bars and relative distances and/or orientation from top to bottom is not drawn to scale. The figure is an illustration depicting relative gene order.



Many STS mapping projects are not appropriate for comparative genomics, primers are not designed to unique highly conserved regions or end up designed to motifs, it has become apparently crucial for comparative analyses to use the most appropriate primer possible. Comparative mapping requires a priming sequence that is unique only to the single-specific loci being amplified. Many markers presented here

went through multiple CAPS and primer set screening before finding a universal primer combination.

With our limited number of 20 new polymorphic genes in *Phaseolus* we found only one area that appears to be syntenic between pea and bean, *Pisum* LG V with *Phaseolus* LG I. The one unexpected result is the close proximity of syntenic markers in bean (~14 cM), while in pea they are slightly more spread out ~20 cM, which is a smaller expansion of the syntenic region then expected based on different genome sizes (Figure 5.7). It brings to question, if gene order is really not conserved or non-homologous sequences were amplified due to nucleotide changes in the highly conserved regions. The highly involved method of primer design and CAPS screening showed primer specificity. Universal primer design also showed *Paal1,2* was a homolog of *PAL3* in *Arabidopsis*. We conclude there is conserved macrosyntenic areas between pea and bean by comparing 19 loci and finding one 10-20cM region showing conservation of synteny, but other regions of conserved synteny appear to be limited in their occurrence compared to more closely related species.

REFERENCES CITED

A.G.I. 2000. Analysis of the genome of the flowering plant *Arabidopsis thaliana*. Nature. 408: 796-815.

Ahn, S. and S.D.Tanksley. 1993. Comparative Linkage Maps of the Rice and Maize Genomes. Proceedings of the National Academy of Sciences 90: 7980-7984.

Bachman, A., Fernandez, L.J., Ginsburg, S., Thomas, H., Bouwkamp, J.C., and T. Solomos. 1994. Stay green genotypes of *Phaseolus vulgaris L*: chloroplast proteins and chlorophyll catabolites during foliar senescence. New Phytology. 126: 593-600.

Bennetzen, J.L., 2000. Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell 12: 1021-1029.

Berhan, A.M., Hulbert, S.H., Butler, L.G., and J.L. Bennetzen. 1993. Structure and evolution of the genomes of *Sorghum bicolor* and *Zea mays*. Theoretical Applied Genetics 86: 598-604.

Binelli, G., Gianfrancesci, L., Pè, M.E., Taramino, G., Busso, C., Stenhouse, J., and E. Ottaviano. 1993. Similarity of maize and sorghum genomes as revealed by maize RFLP probes. Theoretical Applied Genetics 84: 10-16.

Boutin, S.R., Young, N.D., Olson, T.C., Yu, Z.H., Shoemaker, R.C., and C.E. Vallejos. 1995. Genome conservation among three legume genera detected with DNA markers. Genome. 38: 928-937.

Boutin, S.R., Olson, T., Yu, Z.H., Young, N.D., Shoemaker, R.C., and C.E. Vallejos. 1994. Substantial Genome Conservation Among Legume Genera Detected With DNA Markers. Plant and Animal Genome Conference XIV. San Diego, California.

Brauner, S., Murphy, R.L., Walling, J.G., Przyborowski, J., and N.F. Weeden. 2002. STS Markers for Comparative Mapping in Legumes. Journal of American Society for Horticultural Science. 127: 616-622.

Buchanan, B., *Biochemistry and Molecular Biology of Plants*. John Wiley and Sons. 2000. 1057.

Burr, B., Burr, F.A., Thompson, K.H., Albertson, M.C., and C.W. Stuber. 1988. Gene Mapping with Recombinant Inbreds in Maize. Genetics 118: 519-526.

Cannon, S.B., Rombauts, S., Sterck, L., Mudge, J., Wang, X., Sato, S., Vasdewani, J., Cheung, F., Oldroyd, G., Cook, D., Yves, V., and B.A. Roe. 2006. A Sequence-Based Comparison Of The Medicago And Lotus Genomes. Plant and Animal Genome Conference XIV. San Diego, California.

Chen, X., Goodwin, S.M, Liu, X., Chen, X., Bressan, R.A., and M.A. Jenks. 2005. Mutation of the RESURRECTION1 Locus of *Arabidopsis* Reveals an Association of Cuticular Wax with Embryo Development. Plant Physiology. 139: 909-919.

Chittenden, L.M., Schertz, K.F., Lin, Y.R., Wing, R.A., and A.H. Paterson. 1994. A detailed RFLP map of Sorghum bicolor x S. propinquum suitable for high-density mapping suggests ancestral duplication of chromosomes or chromosomal segments. Theoretical Applied Genetics 87: 925-933.

Choi, H., Kim, D.J., Zhu, H., Mun, J.H., Beak, J.M., Roe, B., Ellis, N., Young, N.D., Doyle, J., Kiss, G.B., and D.R. Cook. 2004. Estimating genome conservation between crop and model legume species. PNAS 101: 15289-15294.

Devos, K.M., Atkinson, M.D., Chinoy, C.N., Liu, C.J., and M.D. Gale. 1992. Chromosomal rearrangements in rye genome relative to that of wheat. Theoretical Applied Genetics 85: 673-680.

Devos, K.M., Bryan, G.J., Collins, A.J., Stephenson, P., and M.D. Gale. 1995. Application of two microsatellite sequences in wheat storage proteins as molecular markers. Theoretical Applied Genetics 90: 247-252.

Devos, K.M., Beales,J., Nagamura, Y., and T. Sasaki. 1999. *Arabidopsis*-rice: Will colinearity allow gene prediction across the eudicot-monocot divide? Genome Research 9: 825–829.

Devos, K.M. and M.D. Gale. 2000. Genome relationships: the grass model in current research. Plant Cell 12: 636–646.

Freyre, R., Skroch, P.W., Geffroy, V., Adam-Blondon, A.F., Shirmohamadali, A., Johnson, W.C., Llaca, V., Nodari, R.O., Pereira, P.A., Tsai, S.M., Tohme, J., Dron, M., Nienhuis, J., Vallejos, C.E., and P. Gepts. 1998. Towards an integrated linkage map of common bean. Theoretical Applied Genetics. 97: 847-856.

Freyre, R., Skroch, P.W., Adam-Blondon, A., Shirmohamadali, A., Johnson, W.C., Nodari, R.O., Koinange, E.M.K., Sevignac, M., Bannerot, H., Nienhuis, J., and P. Gepts. 1995. Integrated Linkage Map of Common Bean (*Phaseolus vulgaris L.*). Plant and Animal Genome Conference III. San Diego, California.

Gale, M.D., Devos, K.M., and G. Moore. 1996. Rice as the pivotal genome in the new era of grass comparative genetics. [IRRI] International Rice Research Institute. Rice

Genetics III. Proceedings of the Third International Rice Genetics Symposium, 16-20 Oct. (1995)

Gepts, P. 1998. Genome mapping resources in common bean: towards a unified linkage map in recombinant inbred population BAT93 x Jalo EEP558. Annual Report of the Bean Improvement Cooperative. 41: 91-92.

Gilpin, B.J., McCallum, J.A., Frew, T.J., and G.M. Timmerman-Vaughan. 1997. A linkage map of the pea (*Pisum sativum L.*) genome containing cloned sequences of known function and expressed sequence tags (ESTs). Theoretical Applied Genetics. 95: 1289-1299.

Grant, D., Cregan, P., and R.C. Shoemaker. 2000. Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. Proceedings National Academy of Sciences USA. 97: 4168-4173.

Gray, J., Close, P.S., Briggs, S.P., and G.S Johal. 1997. A novel suppressor of cell death in plants encoded by the Lls1 gene of maize. Cell. 89: 25–31.

Gray, J., Janick-Buckner, D., Buckner, B., Close, P.S., and G.S Johal. 2002. Light-Dependent Death of Maize *lls1* Cells Is Mediated by Mature Chloroplasts. Plant Physiology. 130: 1–14.

Gray, J., Wardzala, E., Yang, M. Reinbothe, S., Haller, S., and F. Pauli. 2004. A small family of LLS1-related non- heme oxygenases in plants with an origin amongst oxygenic photosynthesizers. Plant Molecular Biology. 54: 39-54.

Grusak, M.A., Li, C.M, Moffet, M.D., and Weeden, N., 2004. Map position of the *FRO1* locus in *Pisum sativum.* Pisum Genetics 36: 6-8.

Hasterok, R., Marasek, A., Donnison, I., Armstead, I., Thomas, A., King, I., Wolny, E., Idziak, D., Draper, J., and Glyn Jenkins. 2006. Alignment of the genomes of *Brachypodium distachyon* and temperate cereals and grasses using BAC landing with fluorescent *in situ* hybridization. Genetics. 10.1534

Hortensteiner, S., 1999. Chlorophyll breakdown in higher plants and algae. Cellular Molecular Life Sciences. 56: 330–347.

Jaiswal, P., Ni, J., Yap, I., Ware, D., Spooner, W., Youens-Clark, K., Ren, L., Liang, C., Zhao, W., Ratnapu, K., Faga, B., Canaran, P., Fogleman, M., Hebbard, C., Avraham, S., Schmidt, S., Casstevens, T.M., Buckler, E.S., Stein, L., and S. McCouch. 2006. Gramene: a bird's eye view of cereal genomes. Nucleic Acids Research. 34: D717-D723.

Kaló, P., Endre, G., Zimanyi, L., Csanadi, G., and G.B. Kiss. 2000. Construction of an improved linkage map of diploid alfalfa (*Medicago sativa*). The molecular genetic linkage map of the model legume *Medicago*. Theoretical Applied Genetics. 100: 641-657.

Kaló, P., Seres, A., Taylor, S.A., Jakab, J., Kevei, Z., Kereszt, A., Endre, G., Ellis, T. H. N., and G. B. Kiss. 2004. Comparative mapping between *Medicago sativa* and *Pisum sativum*. Molecular Genetics and Genomics. 272: 235 - 246

Kim, J.S., Islam-Faridi, M. N., Klein, P. E., Stelly, D. M., Price, H. J., Klein, R. R., and J. E. Mullet. 2005. Comprehensive Molecular Cytogenetic Analysis of Sorghum Genome Architecture: Distribution of Euchromatin, Heterochromatin, Genes and Recombination in Comparison to Rice**.** Genetics. 171(4): 1963 - 1976.

King, R.C., and W.D Stansfield. *A Dictionary of Genetics 5th Edition*. Oxford University Press. 1997. 332.

Kuznetsova, E.V., Tzyganov, V.E., Pinaev, A.G., Borisov A.Y., Tikhonovich, I.A., and M.D. Moffet. 2005. Fine mapping of the gene Crt of *Pisum sativum L.* controlling root morphogenesis and being involved in interactions with symbiotic microorganisms. Pisum Genetics. 37: 42-43.

Kwak, M., Kami, J.A., and P. Gepts. 2006. Identification Of The Determinacy Gene (*Fin*) And Its Evolution During Domestication In Common Bean (*Phaseolus vulgaris* L.). Plant and Animal Genome Conference XIV. San Diego, California.

Lan, T.H., Delmonte, T.A., Reischmann, K.P., Hyman, J., Kowalski, S., McFerson, J., Kresovich, S., and A.H. Paterson. 2000. EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. Genome Research. 10: 776-788.

Livingstone, K.D., Lackney, V.K., Blauth, J.R., Van Wijk, R., and M. K. Jahn. 1999. Genome Mapping in *Capsicum* and the Evolution of Genome Structure in the *Solanaceae*. Genetics. 152: 1183-1202.

Lydiate, D., Sharpe, A., Lagercrantz, U., and I. Parkin. 1993. Mapping the Brassica genome. Outlook on Agriculture. 2: 85-92.

Moffet, M.D., and N.F. Weeden. 2005. *Pheophorbide a monooxygenase* (*Pao1*) is located on LG VII near *Amy* in *pea* and *lentil*. Pisum Genetics. 37: 24-29.

Moffet, M.D., and N.F. Weeden. 2006. Investigation Of Synteny Conservation Between *Pisum* And *Phaseolus*. Plant and Animal Genome Conference XIV. San Diego, California.

Moore, G., Foote, T., Helentjaris, T., Devos, K.M., Kurata, N., M.D. Gale. 1995. Was there a single ancestral cereal chromosome? Trends in Genetics. 11: 81-82.

Paterson, A.H., Lan, T.H., Reischmann, K.P., Chang, C., Lin, Y.R., Liu, S.C., Burow, M.D., Kowalski, S.P., Katsar, C.S., DelMonte,T.A., Feldmann, K.A., Schertz, K.F., and J.F. Wendel. 1996. Toward a unified genetic map of higher plants, transcending the monocot–dicot divergence. Nature Genetics. 14: 380–382.

Paterson, A.H., Bowers, J.E., Burow, M.D., Draye, X., Elsik, C.G., Jiang, C.X., Katsar, C.S., Lan, T.H., Lin, Y.R., Ming, R., and R.J. Wright. 2000. Comparative Genomics of Plant Chromosomes. Plant Cell. 12(9): 1523-1540.

Palumbi, S. R. 1995, Nucleic acids II: the polymerase chain reaction, In Hillis, D. and Moritz, C. (Eds.). *Molecular Systematics* 2nd edn Sunderland, MA Sinauer Associates Inc. pp. 205–247.

Pereira, M.G., Lee, M., Bramel-Cox, P., Woodman, W., Doebley, J., and R. Whitkus. 1994. Construction of an RFLP map in sorghum and comparative mapping in maize. Genome. 37: 236-243.

Price, H., Dillon, S., Hodnett, G., Rooney, W., Ross, L., and J. Johnston. 2005. Genome Evolution in the Genus Sorghum (*Poaceae*). Annuals of Botany. 95: 219-227.

Prince, J., Pochard, P.E., and S.D. Tanksley. 1993. Construction of a molecular linkage map of pepper, and a comparison of synteny with tomato. Genome. 36: 404-417.

Pružinská, A., Tanner, G., Anders, I., Roca, M., and S. Hörtensteiner. 2003. Chlorophyll breakdown: Pheophorbide *a* oxygenase is a Rieske-type iron–sulfur protein, encoded by the *accelerated cell death 1* gene. Proceedings National Academy of Sciences USA. 100: 15259–15264.

Pružinská, A., Tanner, G., Aubry, S., Anders, I., Moser, S., Müller, T., Ongania, K., Kräutler, B., Youn, J., Liljegren, S.J., and S. Hörtensteiner. 2005. Chlorophyll breakdown in senescent *Arabidopsis* leaves. Characterization of chlorophyll catabolites and of chlorophyll catabolic enzymes involved in the degreening reaction. Plant Physiology. 139: 52-63.

Salina, E.A., Leonova, I.N., Efremova, T.T., and Marion S. Röder. 2006. Wheat genome structure: translocations during the course of polyploidization. Functional & Integrative Genomics. 6: 71-80.

Salse, J., Piégu, B., Cooke, R., and Michel Delseny. 2002 . Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project. Nucleic Acids Research. 30: 2316-2328.

Schneider, K., Borchardt, D.C., Schäfer-Pregl, R., Nagl, N., Glass, C., Jeppsson, A., Gebhardt, C., and F. Salamini. 1999. PCR-based cloning and segregation analysis of functional gene homologues in *Beta vulgaris*. Molecular Genome Genetics. 262: 515-524.

Stephens, R.M. and T.D. Schneider. 1992. Features of Spliceosome evolution and function inferred from an analysis of information at human splice sites. Journal of Molecular Biology. 228: 1124-1136.

Tadmor, Y., Zamir, D., and G. Ladizinsky. 1987. Genetic mapping of an ancient translocation in the genus *Lens.* Theoretical Applied Genetics. 73: 883-892.

Tanaka, R., Hirashima, M., Satoh, S., and A. Tanaka. 2003. The Arabidopsis-accelerated cell death gene ACD1 is involved in oxygenation of pheophorbide a: inhibition of the pheophorbide a oxygenase activity does not lead to the "stay-green" phenotype in *Arabidopsis*. Plant Cell Physiology. 44: 1266-1274.

Tanksley, S.D., Ganal, M W., Prince, J.P., de Vicente, M.C., Bonierbale, M.W., Broun, P., Fulton, T.M., Giovannoni, J.J., Grandillo, S., Martin, G.B., Messeguer, R., Miller, J.C., Miller, L., Paterson, A.H., Pineda, O., Roder, M. S.,Wing, R. A., Wu, W., and N.D. Young. 1992. High Density Molecular Linkage Maps of the Tomato and Potato Genomes. Genetics. 132: 1141-1160.

TeXshade version 1.4, by Eric Beitz (eric.beitz@uni-tuebingen.de). http://homepages.uni-tuebingen.de/beitz/txe.html

Thomas, H.. 1977. Ultrastructure, polypeptide composition and photochemical activity of chloroplasts during foliar senescence of a non-yellowing mutant genotype of *Festuca pratensis* Huds. Planta. 137: 53–60.

Thomas, H., Schellenberg, M., Vicentini, F., and P. Matile. 1996. Gregor Mendel's green and yellow pea seeds. Botanica Acta. 109:3–4.

Thomas, H., and C.J. Howarth. 2000. Five ways to stay green. Journal of Experimental Botany. 51: 329-337.

Vallejos, C.E., Sakiyama, N.S., and C.D. Chase. 1992. A Molecular Marker-Based Linkage Map of *Phaseolus vulgaris L.*. Genetics. 131: 733-740.

Vicentini, F., Hörtensteiner, S., Schellenberg, M., Thomas, H. and P. Matile. 1995. Chlorophyll breakdown in senescent leaves: Identification of the biochemical lesion on a stay-green genotype of *Festuca pratensis*. New Phytol. 129: 247–252.

Walling, J., Masters Thesis. Montana State University. Bozeman, Montana. 2003.

Walling, J. and N.F. Weeden. 2003. Comparative Mapping Between Pisum and Lens Using Genetically Anchored STS Markers Reveals Areas of Conserved Gene Synteny. Plant and Animal Genome Conference XI. San Diego, California.

Wang, X., Zhao, X., Zhu, J., and W. Wu. 2005. Genome-wide Investigation of Intron Length Polymorphisms and Their Potential as Molecular Markers in Rice (*Oryza sativa* L.). DNA Research 12(6): 417-427

Weeden, N.F., Muehlbauer, F.J., and G. Ladizinsky. 1992. Extensive Conservation of Linkage Relationships Between Pea and Lentil Genetic Maps. Journal of Heredity 83: 123–129.

Weeden, N.F., and J. Barnard. 1994. N.Y. Agricultural Research Station. Geneva Computer Center Technical Report. 20: 2.

Weeden, N.F., Ellis, T.H.N., Timmerman-Vaughan, G.M., Święcicki, W.K., Rozov, S.M. and V.A. Berdnikov. 1998. A consensus linkage map for *Pisum sativum*. Pisum Genetics 30: 1-4.

Weeden, N.F. and M.D. Moffet. 2002. Identification of genes affecting root mass and root/shoot ratio in a JI 1794 x 'Slow' RIL population. Pisum Genetics. 34: 28-31.

Weeden, N.F, Walling, J.G., and M.D. Moffet. 2005. Conservation of Synteny Between Pea and Lentil With Extrapolation to *Medicago, Lotus* and *Arabidopsis*. Plant and Animal Genome Conference XIII. San Diego, California.

Weeden, N.F., Moffet, M.D, Hance, S.T., and K. E. McPhee. 2005. The Location and Action of Genes Influencing Root Growth in *Pisum Sativum*. Plant and Animal Genome Conference XIII. San Diego, California.

Weeden, N.F. and D.A. DeMason. 1999. *Hop1*, Encoding a homeodomain-containing transcription factor, is located near *Prx3* on linkage group VI. Pisum Genetics. 31: 37-38.

Whitkus, R., Doebley, J., and M. Lee. 1992. Comparative genome mapping of sorghum and maize. Genetics. 132: 119-130.

Yang, M., Wardzala, E., Johal, G.S., and J. Gray. 2004. The Wound-Inducible *Lls1* Gene from Maize is an Orthologue of the *Arabidopsis Acd1* Gene, and the LLS1 Protein is Present in Non-Photosynthetic Tissues. Plant Mol. Biol. 54: 175-191.

Zhu, H., Kim, D., Baek, J., Choi, H., Ellis, L.C., Küester, H., McCombie, W.R., Peng, H., and D.R. Cook. 2003. Syntenic Relationships between *Medicago truncatula* and *Arabidopsis* Reveal Extensive Divergence of Genome Organization. Plant Physiology. Vol. 131: 1018-1026.

APPENDICES

APPENDIX A

LIST OF ALL PRIMER PAIRS SCREENED BY PCR

List of All Primer Pairs Screened by PCR (continued on next page)

- Dark lines in graph do not denote any characterization.

* Monomorphic PCR products not screened with all 27 restriction enzymes.

Primer Pair	Forward Primer Sequence	Reverse Primer Sequence	Poly-morphic	Mono-morphic	PCR Non-Specific or Weak	No Bands
1,3 GLU_F1/R1	TATGGCTTCTTTCTTTGCTAG	ATCACTCTTAAGAGAAACAG				X
AAIR_F/R	TTYAARGGIATYAAGCARATIGGIG	CGICCIGCYAAIACAACRCT			X	
AAIR_F1/R1	CACACATCCCGTTTCGCTC	GAATGCCCCTTCTTCCTCAAG			X	
AAIR_F/R1	TTYAARGGIATYAAGCARATIGGIG	GAATGCCCCTTCTTCCTCAAG				X
AAIR_F1/R	CACACATCCCGTTTCGCTC	CGICCIGCYAAIACAACRCT				X
AATc_F/R	TGTTGCTGATGGAGGCGAAT	RATCCTCCCATCAGATGTCA				X
AATp_F/R	ATGGCTCCTCCAGATCCAA	ATCTGCAAGGTATTCACAT	X		X	
AC3_F/R	GCGATAATGGAACCGGAATGG	TTCCTGTGTACTATTGATGG		X		
ACL_F3/R2	CATCCCACACAATGAAGAGT	TCTAATCCATGAATGAAAGA		X		
ALAT_F2/R2	CTGACCTTCCCTCGCCAG	CCTTGCTCGCACCATCAGT		X		
ALDH7_F/R	GATCAATTATMCCWTCAGAAC	TTGTRCAMCGYTGACCAG		X*		
ALDH7_F/R1	GATCAATTATMCCWTCAGAAC	TARCCRAARTTWATTCCTTG		X*		
ALDH7_F1/R1	GGWTGGAATGCTTGCATTGC	TARCCRAARTTWATTCCTTG		X*		
ALDH7_F1/R	GGWTGGAATGCTTGCATTGC	TTGTRCAMCGYTGACCAG	X			
ANS_F/R	GGSTAYGGAAGYAAGYTTGC	CTYTTGTACTTNGCATTG		X*		
ANT_F2/R2	CTCATGGGWGGTGTTTCAGC	AGGATGTTYGCACCYGCWCC				X
APX_F2/R1	CCGATTACCAGAAGGCCATTG	CAAGAGGGCGGAATACAG	X			
APX_F1/R1	ATCGCTGAGAAGAAATGC	CAAGAGGGCGGAATACAG				X
APY_F2/R2	GCCTGCWATCWCTTCCTCCC	GCCTARAGGCCATGCRGCTTCC	X		X	
BAMY_F/R	TATGTTCCAGTTTATGTBATG	TTGTCATATCACTGATTCAG				X
BAMY_F/R2	TATGTTCCAGTTTATGTBATG	ACTTCAATGTTTTCTCTCCA				X
BEBT_F3/R3	GCTTCGTTTCCAGATACCAG	GAACCCAGCTCGTGTCACG				X
BFruct_F2/R2	ACTTTCAACCTCTCAARAATTGG	CCCACACAATRTTYCCCCA	X			
CAO_F/R2	TGGTAYCCMGTTGCTTTC	CAGACATGGAGTTGRTGRTG			X	
CAO_F2/R2	ATGRTCTGGATYTGGCCYGG	CAGACATGGAGTTGRTGRTG			X	
CCR_F3/R3	GTAACACATAATAAATGGAG	GGTGTAGTGTTCACTTC				X
ClPOR_F/R	ACTGCTAAGGCTTTGGCTGA	AGATTTTGTTAGGCTTGGATCACT				X
ClPOR_F2/R2	GCTACTGCTAAGGCWTTGGCKG	CCWGATTTTGTTAGGCTTGGATCAC	X			
cMDH_F4/MDH_R3	CATCACTCCCCTAGCAATC	CTCTTCGGCTGTCAAGTCCA			X	
Crabs Claw_F/R	TCTGCAACACYTYCTTGCG	CTGAAAGCTTCWCGRTGWGG		X*		
CRS_F/R	GGCRTCTGGGATGTGTGCTA	TACACGGAGATGCAGCGTTTTC	X			
CT16_F/R	ACGTGGCATTGGTGGAC	AACCCATCTAGCCCAAGTT			X	
CT16_F2/R2	CTTCTTCTATTTTCTTCGCACC	CATAACACCACTGTACCCTTC			X	
CT16_F2/R	CTTCTTCTATTTTCTTCGCACC	AACCCATCTAGCCCAAGTT			X	
DCOP_F2/R2	CCTCCCGCAGCAAATCCT	AGGCTGAAGAACCACGTCAG				X
DIMIN_F/R	AAGAGGAAGAAGATTTGGGT	TCAAAGCCTGCTTCTGGGTA				X
DOSSGA18_F/R	CCGATTTCTTCGTCGTGTCC	CAACTCCAGGCAGAACCAAACA			X	
EGL_F2/R2	TTTTCAAGGTCAGAGGTCAGG	TCAGCAGCAACATCAGAACC				X
EGL_F2/R3	TTTTCAAGGTCAGAGGTCAGG	TTGAAGAGGGTGAGTTAGGCA				X
ENOD12_F/R	AAGTGGTCAAAGATGATAAGAC	GCTTTAGAAATGGATGTTATGTTC				X
ENOD40_F1/R1	CCTCTAAACCAATCCATTAT	TAACAGTTACCTACCTTAC				X
ENOL_F2/R2	TTCCATCAAGGCCCGTCAGA	TTGCACCAACCCCATTCATT	X			
Fbpp_F/R	CTTACTCTCCTTCACGTCT	CTTTTCAACCTTCTCCACCT		X		
FDGOGAT_F1/R1	CCTTGAAAGCAATGAGATG	TAGGATATGGTGCAAGG			X	
FK_F2/R2	GTSTCCCTYGCAGAAGC	GCACCAACAAAYGAATCACC			X	X
FK_F3/R2	AGATCAGCACACTTGAAGGC	GCACCAACAAAYGAATCACC	X			
FK_F3/R3	AGATCAGCACACTTGAAGGC	GCCTTCAAGTGTGCTGATC				X
FK_F4/R3	GTTCGGTCACATGCTCGCC	GCCTTCAAGTGTGCTGATC	X			
FK_F4/R2	GTTCGGTCACATGCTCGCC	GCACCAACAAAYGAATCACC	X			
FRO_F/R	AGAACATGGCCCTAAAGTGTACGTAA	CCCTTAGAGAAGTTGAGTTCAACGG				X
FTase_F/R	GGTGGATATGCTGGGGGA	AGAATATATTGCTGTAAGGC				X
GAPA_F/R	TYCTGAATTCTCTGGYCTCCG	CGAATTTCTGRTCAACRACYTTG		X		
GAPA_F/R1	TYCTGAATTCTCTGGYCTCCG	TGGCAACAATGTCAGCCAAATC		X		
GIB2BH_F1/R1	CTTGTGAAGATTTTGGATTCT	ACTTTTGAACCTCCCATTAGTC				X
GIB2BH_F2/R1	TCTCCTCACAACCAATCAAGA	ACTTTTGAACCTCCCATTAGTC				X
GLO_F/R	GCRTCIGCWGCIGGIAC	AGYTGKCGIGCICCRTG			X	
GLUPRX_F/R	GCAAGYAACTTCTGGATRTCCTT	GCAAGYAACTTCTGGATRTCCTT			X	
GmPEPCN_F/R	CCAGCAGCTATACCCTTCATTGT	ATGGATGAGATGGCTGTCATTGC		X	X	
GPT_F3/R2	GCTGAAGCTCCTAAGACTGATC	TGCGGCTCCYAGAGCATTGAC	X			
GRF3_F1/R1	CTCAGTGGCARGAACTTGAAC	GGCTTTCTTGAACGGTTGCG	X			
GRF3_F2/R1	GCAGTAGTAACGGAGGAGGT	GGCTTTCTTGAACGGTTGCG		X		

GSc_F/R	CCTCTCCATTGATGCCAC	CCTCTCCATTGATGCCAC				X
Gsn1_F2/R2	TGGCCGTAGATTGATTGTTG	ATTGCTCCCACTTTCCATGA		X	X	
Gsp_F3/R2	CTCTCACTCTCTATTGCT	ATGGAGCAATAGAGAGTG		X	X	
HEXK_F/R	AACTCAATCTCCATCCCCA	AACTCAATCTCCATCCCCA				X
HsfA_F/R	CTTCGTTCGTCAACTCAACACC	AAAAACCCTTCATTCGCAAAT		X*	X	
HOP_F1/R1	CTTACACTAACCCTTGTTCC	CATCACCACCCATTCTCACT				X
Hsp70_F1/R1	CGGNATYGATCTCGGNACSAC	CATCTTNGCRATGATNGGRTTGC	X		X	
IDH_F/R	CCAGAAAATCAAAGTCGCCA	TAAYWCGGGTCATTTCATCTC				X
IVDH_F1/R1	CACTTCCTTCCTCTTTGACGAC	AACCTGCGTTGCTCTTTCAG				X
IVDH_F1/R2	CACTTCCTTCCTCTTTGACGAC	TGAGCAACGGTCCCATTAGT				X
LASY_F/R	CCGAAACCTAAATGGATGAAGGAATC	CACAGCACTCTGAAGCTCTTC			X	
LASY_F2/R2	ACCCTGATGAGCCCACCA	GCAGCCCGATCTGAGTC				X
LEGA_F1/R1	CATGGCTAAGCTTCTTGCAC	GCTCTGTTCTCAGACTCACG		X*		
LEGK_F1/R1	GCTTCTCCAATTCACCACTTCACC	GACTGAGGGTGAACCAGAGGGCC		X*		
LEP_F/R	ATGAACACAACATCATC	TTAGGAGCCAAAGTAGT			X	
LHY_F/R	CAAAGRGAACGATGGAC	KGCYTCYTTCTCCAACT			X	
LHY_F1/R1	AAGTTGGAGAARGARGC	GAGGAGGYGAAAAGCTTT			X	
LIP_F3/R3	CATCACTGGATCAGATTCAC	GCTTATCTGCTGCTTGCC				X
MEVLK_F/R	TCCRCGTTTCTCTGGYTRTAC	CATKGACWGGAGCAGACCT	X			
MEVLK_F/R1	TCCRCGTTTCTCTGGYTRTAC	CCTCCAGCTCCWGTTARTTTGGA	X			
MNSOD_F/R	GAACATGCCTACTACTTACAG	CTCTCTTTCTCATATACTTCAC				X
NPAC_F/R	TCAAAGCAAAGCAGAAGYGAG	CGATGGCTGGAACAGAAC		X*	X	
NPAC_F2/R	TGGCTCCAGGTCCAGTTRTTGA	CGATGGCTGGAACAGAAC		X*	X	
NUP133_F1/R1	CTTGATGTGCAGGTAGATGAT	GCACCTCTGGTTGTTGTCC		X*	X	
P393_F/R	CTGGTTGGTCCTTCCTT	AAAGAGTGACAAGAACC				X
P5CS_F1/R	GTTCCAGAMGGCTTCAGG	ACTGGACCHCGAGCATGRATGG			X	
P5CS_F/R	TCWGGTGCTGTTGGYCTTGG	ACTGGACCHCGAGCATGRATGG			X	
P5CS_F/R1	TCWGGTGCTGTTGGYCTTGG	TGTRCTTGCRTTGTGRAAHACAGC			X	
P5CS_F1/R1	GTTCCAGAMGGCTTCAGG	TGTRCTTGCRTTGTGRAAHACAGC			X	
P628_F/R	ATGTGGTCTCAACTGAC	AGCTATTGCAAACATGAC				X
Paal1,2_F/R	GAGGTGAARCGYATGGTGG	GGTTTYCCTTGCATAACTTCAGC	X		X	
PAAL2_F/PAAL_R	CAATAACATCAAAGTGAGTGAC	GGTTTYCCTTGCATAACTTCAGC		X*		
Pao_INT-F/INT-R	GCAGGAGGGAAGGATAGATGAG	CTATCNCTTCTTCCTGTGACC		X*		
Pao_INT-F/R3	GCAGGAGGGAAGGATAGATGAG	GTGAAMGTGAGSRTTGTGTACTG	X			
Pao_F2/R3	GGAGAGATCATTGGTAYCCAGTTTC	GTGAAMGTGAGSRTTGTGTACTG		X*		
PGlp_F2leg/R2leg	AAGGCCTTTCGTGAAGCAGG	CCATCTGTAGCCCARTACCAA	X			
PGLM_F2/R2	CRTTACCTTGTTTCTCCMCC	TCTCYGGKGGAGAAACAAGG				X
PHYAPO_F1/R1	CCTATTTTGGCAGTGGATGTTG	ACTTCCGACTGATGTGTAGGCT		X		
PIN1_F/R	ATGATGGGTGGTGGAAGAAA	GCTTGAATAAGTGTTTGGGTTT				X
PI4P_F/R	CTGGAGAGCAGACCCATTCAAT	GCGAAAGCCTCCAATCCAC		X		
PI4P_F2/R	TAGGTGCTGATATTCAATTTGG	GCGAAAGCCTCCAATCCAC		X		
POSSNP_F/R	CTTGCAGCACGGACCATAAGG	GGTTKAGGATAACAAGTSGACATTCC			X	
PPE_F3/R3	GGCTGGTTGTGATTGGATTCA	GGRGTACCAGGGTTTAGGA		X		
Q363_F/R	TCGAGACTTTAAACCCG	GTCTTTGTCTTTAGGTTG				X
RBRP_F/R	ATCATTCTGTGTTGCTTCTACGG	TGAAAGGGCTCTGATAAGCATG		X	X	
RCA_F2/R2	AGCTTRTTGTTCACATCACCAAG	TGWACTCTCTTCACATTCTCTTG	X		X	
RCC_F1/R1	TCACACAACTCCTTTAACGAAA	TTATGCTAAATTCCTTGATGAAGC		X*		
Rcpme_F/R	GTTCAAAACTGTGGCTGAGGCT	GTGTTCTGGTTTGGGTCTTCTC	X		X	
RPL15_F1/R1	GATTTGGATGCACTTCTTGAC	GGTTCTGGCTTCTCACCA			X	
RPL15_F1/R2	GATTTGGATGCACTTCTTGAC	CCTTCAARATAAACTGCAAAGTAC				X
RPL22_F/R	CTCTCTCTTTAGCCATTAAC	CTTCCTTGTCAGACTCATC				X
RPL24_F/R	TGCCGATTCAGTGGTCMAAG	TTCTTSGCTTTCTTCTCATCC				X
RPL35A_F1/R1	GTGCCCCAAATAAGCTCTC	GTGAGGCGTCTTCTAATAGC		X	X	
RPA32KDA_F/R	CGGCGGCGGATTCATGTC	TCAAARTTTGTCCRGGCCTG			X	
RST1_F/R	TRTTCARGAAATGRRACTC	TCAATRAGCCACTGTCT			X	
SAHH_F1/R1	GATTGGGGTCCTGGTGGTG	AAGAGCAGCAACCTTCTCATCAA		X	X	
SBE_F2/R2	GCYCCICGIATMTAYGARGC	RTTGCCYTCYCTIGGRAARTC		X		
SEP3_F2/R2	GTTATGAGTCTCTACAACGCTCTC	TCAGCACTCAGATTCAGTTGG		X	X	
SGR_F/R	ACCACAACTCACAAGTCACTTC	CTTCCAACACCTTCTCAACTC		X		
SHMT_F/R	ACCACAACTCACAAGTCACTTC	CTTCCAACACCTTCTCAACTC		X		
STV3_F/R	ATGGRTGGRWGGAATATGC	TCKCATGGGRGARTGRATT		X*		
STV3_F/R2	ATGGRTGGRWGGAATATGC	GATTTCTCTGTGTGTCAGAC			X	
Sym29_F/R	CAAAGAGATGAAAAGTATCAC	TCGCCAGTGAGAGATTATC			X	
t-CAH_F1/R1	ATCTCAACCACCGCAACCTC	TTGTCCTCCTTTCTCAGAAGTGTC			X	
THIOP_F/R	AGATGGAGAAGTCACCACAC	CCATGAGTTCTTCAYTATCC		X		
TIC55_F/R	GTGGGGACGAGAGAAAGATGGRR	AGGAAGCTGATAGTCCTGCAACA			X	
TKP_F/R	GGIGAYGGITGYCARATGGA	GAIGCRAGRTCIGCRCTIC	X		X	
TPI_F/R	GGIGAYTCIGCIATGGARGA	CRTCCARIGCIGCCATRCA				X
TPI_F2/R2	TTCGTCGGYGGYAACTGGAA	CCWGTWCCAATRGCCCARAC			X	
TubA1_F/R	GGAAACGCCTGCTGGGAG	AACAGTTGGAGGCTGATAAT	X		X	
TUFM_F2/R2	GATGGACCWATGCCTCAAAC	ACCCTCATCCTTTGTGAGTAC	X			
UNK21_F/R	TCGCCTCCATGTCCACCTC	CGGCCTTGCTAAATCAGTCAG	X			
VR_F1/R1	GAGTGTGTGTAACTGGAGGTACAGG	ATGCCTAATGCGCCATCGACGGTTC		X		

APPENDIX B

PERMISSION LETTER

MONTANA
STATE UNIVERSITY

4-17-2006

Dr. Norm Weeden
Pisum Genetics Editor
MSU-Bozeman

Dear Dr. Norm Weeden:
The presented letter confirms our recent communication on permission to use and print the journal article in *Pisum Genetics* 2005 Volume 37 p.23-28. I am completing a masters of science program with thesis at Montana State University Bozeman entitled "IDENTIFYING REGIONS OF SYNTENY BETWEEN PEA (*PISUM SPP.*), LENTIL (*LENS SPP.*), AND BEAN (PHASEOLUS *SPP.*)." I would like your permission to reprint in my dissertation excerpts from the described journal article. The excerpts to be reproduced are the journal article is its entirety. The requested permission extends to any future revisions and editions of my dissertation, including non-exclusive world rights in all languages, and to the prospective publication of my masters thesis by UMI. These rights will in no way restrict republication of the material in any other form by you or by others authorized by you. Signing of this letter will also confirm that *Pisum Genetics* owns the copyright to the above-described material. If these arrangements meet with your approval, please sign this letter where indicated below and return it to me in the enclosed return envelope.
Thank you very much.

Sincerely,
Matt Moffet
M.S. Graduate student
Research Assistant
BIOL 102 TA

PERMISSION GRANTED FOR THE
USE REQUESTED ABOVE:

_____
Dr. Norman F. Weeden

Date: _____