Cytological and molecular investigations in Lens and Pisum
by Jason Gordon Walling

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Plant Sciences
Montana State University

Abstract:
Three separate investigations involving genetic characterization of the pea (Pisum sativum) and/or lentil (Lens culinaris) genome were performed.

The first experiment, evaluated the hypothesis of conserved gene synteny between lentil and pea species. Sixty-two segregating P. sativum anchor markers (STS, isozyme, RFLP, and morphological) and two lentil protein markers were tested within a lentil RIL population. Ten linkage groups were assembled that span over 300 centimorgans. The majority of the loci evaluated share linkage synteny with pea; however some notable genetic rearrangements were revealed. Supplementary to this study, data supported the presence of two reciprocal translocations that differ in each parent. In addition, approximately 25% displayed distorted segregation ratios.

The second experiment characterized and evaluated a lx Bacterial Artificial Chromosome (BAC) library constructed from P. sativum. The results suggested that approximately 1.0% of the clones contain chloroplast DNA sequence while approximately 0.35% would contain sequences homologous to the 45s rDNA array. The presence of lower copy sequences, such as gene family constituents and unique sequences were also evaluated. One clone, of the gene family small subunit of ribulose bisphosphate carboxylase (Rubisco), was identified from the library. One low copy gene, glutamine synthetase, was also identified within the BAC library. Inconsistent results obtained during the in-clone verification process are discussed.

The final experiment used the two partially characterized BACs obtained from the library screening to assess their use as in situ probes using Fluorescent In Situ Hybridization (FISH) on P. sativum chromosomes. rDNA probes as well as repetitive DNA probes were integrated into the study. The results verified the physical position of the NOR and 5s loci in both pea and lentil. Furthermore, a BAC containing the sequence for the Rubisco small subunit hybridized to three regions. However, the BAC containing the glutamine synthetase sequence did not hybridize to a specific location, possibly because the C0t-1 blocking did not provide sufficient inhibition to cross-hybridization. Results indicated that when using current protocols, positioning BACs on pea chromosomes is difficult, and improvements are needed before such an approach can be routinely integrated into an experimental design.

CYTOLOGICAL AND MOLECULAR INVESTIGATIONS IN *LENS* AND *PISUM*

by

Jason Gordon Walling

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

of

Master of Science

in

Plant Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2003

APPROVAL

of a thesis submitted by

Jason Gordon Walling

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 College of Graduate Studies.

Dr. Norman F. Weeden Norman F Weeden 4/18/03
(Signature) Date

Approved for the Department of Plant Sciences and Plant Pathology

Dr. Norman F. Weeden Norman F Weeden 4/18/03
(Signature) Date

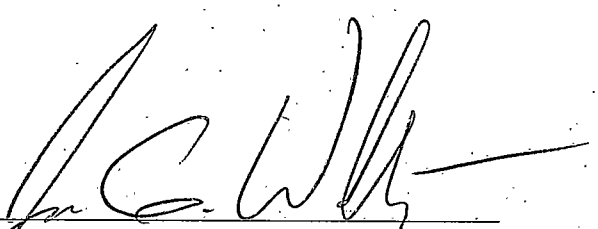
Approved for the College of Graduate Studies

Dr. Bruce R. McLeod Bruce R. McLeod 4-21-03
(Signature) Date

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.

Signature 

Date 04-18-2003

TABLE OF CONTENTS

TABLE OF CONTENTS-CONTINUED

TABLE OF CONTENTS-CONTINUED

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF FIGURES-continued

ABSTRACT

Three separate investigations involving genetic characterization of the pea (*Pisum sativum*) and/or lentil (*Lens culinaris*) genome were performed.

The first experiment, evaluated the hypothesis of conserved gene synteny between lentil and pea species. Sixty-two segregating *P. sativum* anchor markers (STS, isozyme, RFLP, and morphological) and two lentil protein markers were tested within a lentil RIL population. Ten linkage groups were assembled that span over 300 centimorgans. The majority of the loci evaluated share linkage synteny with pea; however some notable genetic rearrangements were revealed. Supplementary to this study, data supported the presence of two reciprocal translocations that differ in each parent. In addition, approximately 25% displayed distorted segregation ratios.

The second experiment characterized and evaluated a 1x Bacterial Artificial Chromosome (BAC) library constructed from *P. sativum*. The results suggested that approximately 1.0% of the clones contain chloroplast DNA sequence while approximately 0.35% would contain sequences homologous to the 45s rDNA array. The presence of lower copy sequences, such as gene family constituents and unique sequences were also evaluated. One clone, of the gene family small subunit of ribulose bisphosphate carboxylase (Rubisco), was identified from the library. One low copy gene, glutamine synthetase, was also identified within the BAC library. Inconsistent results obtained during the in-clone verification process are discussed.

The final experiment used the two partially characterized BACs obtained from the library screening to assess their use as *in situ* probes using Fluorescent *In Situ* Hybridization (FISH) on *P. sativum* chromosomes. rDNA probes as well as repetitive DNA probes were integrated into the study. The results verified the physical position of the NOR and 5s loci in both pea and lentil. Furthermore, a BAC containing the sequence for the Rubisco small subunit hybridized to three regions. However, the BAC containing the glutamine synthetase sequence did not hybridize to a specific location, possibly because the C_0t-1 blocking did not provide sufficient inhibition to cross-hybridization. Results indicated that when using current protocols, positioning BACs on pea chromosomes is difficult, and improvements are needed before such an approach can be routinely integrated into an experimental design.

CHAPTER 1

THESIS INTRODUCTION

Legumes

The legume family (Fabaceae) is the third largest family of flowering plants next to the orchid family (Orchidaceae) and the sunflower family (Asteraceae), respectively (Stern, 1997). The Fabaceae contains approximately 350 genera composed of 18,000 species of considerable genetic diversity that vary in both climatic and geographical distribution. Legumes are distributed throughout the world, from the tropics to arctic and alpine habitats.

There are several common characteristics that support the family of legumes. Most legumes possess the typical fruit morphology represented by a pod that encloses the seed(s), which when dry will split along two sutures facilitating seed dispersal. The word legume originates from the latin term *legumen* which literally means 'anything that can be gathered', presumabley referring to the grouping or gathering of seeds within the pod. The general flower morphology has been described as being either radial or bilateral (asymmetrical) (Stern, 1997). The asymmetrical flower, specifically the morphological features of the banner, keel, and wing-petals, are arguably the most recognizable features of the subfamily Papilionoideae. One common characteristic shared among most species of legumes is the symbiotic relationship shared with the bacteria *Rhizobium,* which facilitates the fixation of atmospheric N_2 to ammonia (NH_3) or other nitrogenous bases. The ability of legumes to fix nitrogen promotes their use in crop rotations and as a green manure (Campbell, 2002). The economic value represented by this family of legumes, is

surpassed only by that of the grasses (Poaceae) (Hymowitz, 1990). However, it has been argued that in terms of the number of species and various uses of legumes such as green manure, fuel, forage and feed crops, cover crops, chemicals and timber; members of the Fabaceae are probably more widely used by humans than any other family of plants (Hymowitz, 1990).

Within the Fabaceae, members of the subfamily Papilionoideae, specifically the tribes Vicieae (*Pisum, Lens, Vicia*) Phaseoleae (*Phaseolus, Vigna*), and Cicereae (*Cicer*) represent what have been referred to as the primary grain legumes (Hymowitz, 1990). The distinction between primary, secondary, and tertiary grain legumes is based on greater production numbers within the primary grain over constituents of the secondary and tertiary grain legumes (Hymowitz, 1990). Other commonly known members of the subfamily Papilioniodeae are the tribes lupin (Lupinus), clover (Trifolieae), and peanut (Arachis).

Members of the tribe Viceae such as *Pisum sativum* L. (field pea), *Lens culinaris* (lentil), *Vicia faba* (faba bean), and *Lathyrus sativus* (grass pea) are commonly labeled as cool season legumes, referring to the species preference of temperate/cool growth conditions. This tribe is thought to have originated along the Fertile Crescent. The domestication of the primary legumes was concurrent with the domestication of the cereal grains (Zohary et al., 1993). The phylogenetic relationship among these three tribes within the subfamily Papilionoideae is described in Figure 1.

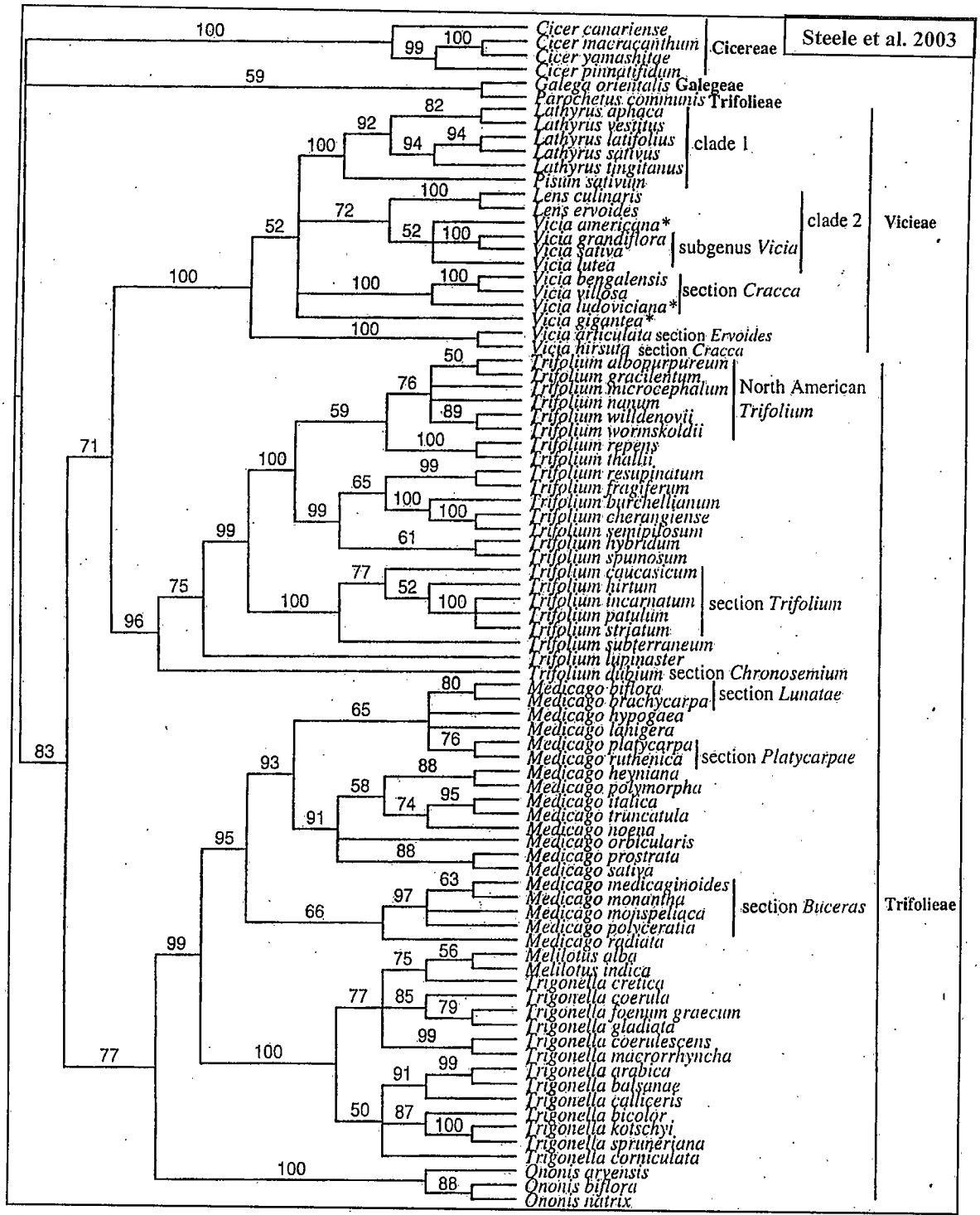


Figure 1. (Top) Phylogenetic relationship between members of the tribes Vicieae and Trifolieae based on plastid *MATK* sequence data (Steele et al. 2003).

Maize and beans were thought to have co-evolved in the New World, peas, lentils, and chickpeas co-evolved with wheat and barley in the Fertile Crescent, and rice co-evolved with soybeans in the Orient (Hymowitz, 1990).

Pisum (Field Pea)

Pisum sativum L. or garden/field pea is an annual cool season legume that prefers relatively humid conditions, ranging in temperature from 7 to 30°C (Muelhbauer, 1997). *P. sativum* is a self-pollinating diploid, with a haploid chromosome number of seven (n=7). Compared to most plant species, *Pisum sativum* has a large genome comprised of approximately 4,000Mbp (Megabasepairs, Mb) (Arumuganathan et al., 1991); consequently, the size of each of the seven chromosomes represented by the standard karyoptype is relatively large.

The morphology of the field pea has been described as bushy or climbing. The stems grow 30-150cm in length and are relatively weak. The leaves are alternate, pinnately compound and possess branched tendrils. Some cultivars have a mutation that converts all laminae to tendrils and are referred to as 'semileafless'. In addition to the characteristics of legume flower previously stated, the pea inflorescence is a raceme and the flowers range in color from white to purple. The fruit morphology has been described as pods that are curved or straight, swollen or compressed, and having 2-10 seeds which are dispersed by pod dehiscence along two sutures (Muehlbauer et al., 1983).

Peas are grown primarily for their fresh or dried seeds, tender pods and foliage. Peas are cultivated throughout the world's temperate regions and represent an important

export crop that comprises almost half the world trade in pulse crops. Most of the commercially produced peas in the United States are grown in Wisconsin, Minnesota, Washington, and Oregon (Muehlbauer, 1993). From 1993-1997 the number of harvested acres of pea in the U.S. was estimated to be 250,000 acres, yielding 450,000 tons of processed pea with a total value of approximately 140,000,000 dollars (Rhodes, 2002). However in 1990 it was estimated that the total acreage of peas in Canada was three times greater than that in the U.S. (Slinkard et al., 1990).

The elucidation of the genetics behind a few morphological features of *P. sativum* was concurrent with the establishment and initial development of genetics as a science. Specifically, the pioneering works of Gregor Mendel who chose the garden pea as the primary system in which to test his hypotheses on inheritance (Mendel, 1866). The seven contrasting pea morphological traits that Mendel analyzed in his famous study include seed shape, cotyledon color, flower color, pod shape, pod color, flower position, and plant stature. Mendel published his findings in scientific paper titled "Versuche über Pflanzen-Hybriden" or "Experiments on Plant Hybrids" in which he outlined his findings. Although the nomenclature came later, the principles of segregation and independent assortment that govern the science of genetics were based on Mendel's conclusions.

Lens (Lentil)

Lens spp. , like *Pisum,* are annual legumes that prefer temperate to warm season, although they require less humidity than pea. Lentil also shares the same haploid chromosome number and ploidy level as *P. sativum* (n=7) and also is primarily a self-

pollinating species. The genus *Lens* is comprised of five species that can be placed into two crossability groups (Ladizinsky et al., 1984). Group one contains *L. culinaris, L. orientalis,* and *L. odemensis* and group two contains *L. nigricans, L. ervoides*. Studies analyzing the variation of chloroplast DNA sequence among the five members revealed their phylogenetic relationship. As expected, considering the results from Ladizinsky's fertility studies, members of crossability group one had the closest sequence identity to each other. *L. culinaris* was most closely related to *L. orientalis* followed by *L. odemensis*. Furthermore, results from this experiment confirmed the hypothesis that *L. orientalis* represents the progenitor species to *L. culinaris*. The members of the second crossability group show the greatest amount of genetic variability at the plastid DNA level, with some results suggesting that *L. ervoides* may be closer to *L. culinaris* than *L. nigricans* (Muench et al., 1991).

The morphology of *L. culinaris* Medik. has been described as a bushy, semi-erect or erect with slender angular stems that reach a length of 18-36cm (Muehlbauer et al., 1985). The leaves are pinnate and alternate and there may be ten to sixteen leaflets attached to the rachis with simple tendrils extending from the top. The fruit consists of pods that are oval shaped containing 1-2 seeds. One to four flowers can form from each peduncle (2.5-5cm) and vary in color from white to purple. The inflorescence is a raceme. The seeds of the lentil come in a variety of shades of yellow, red or orange and have been described as lens shaped and small. The size of the seed varies in diameter from 6 to 9mm in the New World (macrosperma) species and 2 to 6mm in the Near East (microsperma) species (Muehlbauer et al., 1985).

Although to some extent lentil is used as a green manure crop, the seed is probably its greatest asset. The seed it typically eaten as dhal which is simply seed that has been peeled and split. Lentils are primarily grown in Turkey, India, Syria, Nepal, Canada, Ethiopia and Spain. In 1992 India lead the world in total lentil production with 750 tons produced followed by Turkey and Canada. In the United States the total amount of acreage sown with lentil in 1996 was estimated to be 60,000ha. The total production from 1984-1993 was approximately 65,000 tons (Muehlbauer, 1997).

The following three chapters will describe three separate investigations involving *Pisum* and *Lens*. Each experiment was designed to utilize the most current technology available to geneticists as well as to exploit and integrate information gained from past investigations within these systems. The first two experiments described in chapters 2 and 3 involve an assessment of the logistics for integrating a large insert library into *Pisum* genomics. The first experiment will evaluate the efficiency and accuracy of locating genes, via hybridization, within a newly developed 1x genome coverage, *Pisum sativum* Bacterial Artificial Chromosome (BAC) library (Coyne et al., 2000). The second experiment will use the BACs resulting from the first experiment to evaluate their potential use in subsequent pea cytology experiments. Fluorescence *In Situ* Hybridization (FISH) with fluorescently labeled BACs will be used to examine the hybridization characteristics of the large insert clones. The information acquired from these investigations will be valuable to future experiments directly involving the library as well as subsequent use of the BACs identified from the library.

The final experiment described in chapter 4 will examine the hypothesis of conserved gene synteny shared among members of the tribe Vicieae. A linkage map of *Lens* was generated primarily using a novel set of sequence tagged sites (STS) markers with clear positions on the *Pisum* linkage map (Brauner et al., 2002). The STS marker data was integrated with previously generated isozyme, protein and restriction fragment length polymorphism (RFLP) data (Weeden et al., 1992). Gaps between markers will be closed using RAPD (Random Amplified Polymorphic DNA) markers. The resulting map will be compared to the consensus map of *P. sativum* (Weeden et al., 1998) and areas of synteny will be revealed. Based on the results of this study, other members within the subfamily Papilionideae could be easily tested using this approach. Thus, the efficiency of generating linkage maps for each of the species will be discerned.

Literature Cited

Arumuganathan, K. E.D. Earle. (1991). Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter. 9:211-215.

Campbell N.A., J.B., Reece. (2002). Biology (6th ed). Benjamin Cummings, San Francisco, CA.

Brauner S., R.L. Murphy, J.G. Walling, J. Przyborowski, N.F. Weeden. (2002). STS Markers for Comparative Mapping in Legumes. J. Amer. Soc. Hort. Sci. 127 :616-622.

Coyne, C. K. Meksem, D.A. Lightfoot, K.E. Keller, R.R. Martin, M.T. McClendon, D.A. Inglis, E.W. Storlie, K.E. McPhee. (2000) Construction of bacterial artuficial chromosome library for pea (*Pisum sativum* L.) Pisum Genetics 32: 23-26.

Hymowitz, T., (1990). Grain legumes. In: J. Janick and J.E. Simon(eds.), Advances in new crops. Timber press, Portland Oregon.

Ladizinsky, G., D. Braun, D. Goshen, F.J. Muehlbauer (1984). The biological species of the genus *Lens* L. Bot. Gaz. 154:253-261.

Mendel. G. (1866). Versuche uber Pflazen-Hybriden. J. Heredity 42, 1951 p3-37. (Reprint).

Muehlbauer, F.J., R.W. Short, J.M. Kraft. (1983). Description and culture of dry peas. USDA-ARS Agricultural Reviews and Manuals, Western Series, No. 37. Western Region, Oakland, California. P.92.

Muehlbauer, F.J., J.I. Cubero, R.J. Summerfield. (1985). Lentil (*Lens culinaris* Medik.) P.266-311. In: R.J. Summerfield and E.H. Roberts (eds.) Grain Legume Crops. Collins 8 Grafton Street, London, UK.

Muehlbauer, F.J. (1993). Food and grain legumes. p.256-265. In: J. Janick and J.E. Simon (eds.), New Crops. Wiley, New York.

Muehlbauer, F.J. (1997). *Pisum sativum* L. NewCROP FactSHEET

Muehlbauer F.J., A. Tullu. (1997). *Lens culinaris* Medik. NewCROP FactSHEET.

Muench, D.G., A.E. Slinkard, G.J. Scoles. (1991) Determination of genetic variation and taxonomy in lentil (*Lens* Miller) species by chloroplast DNA polymorphism.

Rhodes D. (2002). U.S. Processing Pea Production Statistics (1992-1993). http://www.hort.purdue.edu/rhodcv/hort410/peas/pe00002.htm

Slinkard, A.E., R.S. Bhatty, B.N. Drew, R.A.A. Morrall. (1990) Dry pea and lentil as new crops in Saskatchewan: A case study. p.159-163. In: J. Janick and J.E. Simon (eds.), Advances in new crops. Timber Press, Portland OR.

Steele, K. P., and M. F. Wojciechowski. (2003). Phylogenetic systematics of tribes Trifolieae and Vicieae (Fabaceae). Advances in Legume Systematics. (in press).

Stern, K.R. (1997). Introductory Plant Biology. Wm.C.Brown publishers. Dubuque, IA.

Weeden, N.F., F.J. Muehlbauer, G. Ladizinsky. (1992) Extensive conservation of linkage relationships between pea and lentil genetic maps. J. Hered. 83:123-129.

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

Zohary, D., M. Hopf (1993). Domestication of plants in the old World – The origin and spread of wild cultivated plants in West Asia, Europe, and the Nile Valley. Clarendon Press, Oxford

CHAPTER 2

CHARACTERIZATION OF *PISUM* BACTERIAL ARTIFICIAL CHROMOSOME (BAC) LIBRARY

Literature Review

Introduction

Since the introduction of molecular cloning techniques in the late 1970s, DNA libraries have become a powerful means for accessing genetically important fragments of DNA. The segmentation and subsequent subcloning of genomes allows researchers to avoid some of the experimental limitations of manipulating complex genomes *in vitro*. By means of displaying sequence information in an ordered array such as in microtitre plates or on high-density filters, DNA libraries can facilitate the identification and cloning of DNA sequences that may otherwise remain unresolved. DNA libraries, of one form or another, have been constructed for nearly all agronomically and genetically important species and have played a crucial role in furthering the status of both structural and functional genomics in these species.

DNA Libraries

The extent to which DNA can be systematically analyzed has generally been governed by the recombinant cloning vector used in the cloning experiment. In particular, the size of the insert a vector can faithfully accept remains a limiting factor. Consequently, a variety of libraries have been developed that range in both size and complexity (Ausubel et al., 1999). Libraries employ the ability of genetically engineered

cloning vectors to accept fragments of DNA. These 'recombinant' vectors are subsequently transformed into a host cell to stabilize and propagate the vector. The components of a cloning vector such as a selectable marker, a sequence for controlling copy number, and multiple cloning sites allow the successful cloning of fragments to be accomplished with little effort. The simplest cloning vector, the plasmid, is used routinely in single event cloning experiments, such as the cloning of a PCR product. Although some plasmids have been shown to accept fragments up to 20kb, routine plasmid cloning is typically restricted to fragments that are between three and 10kb hindering their utility in constructing genomic libraries. Thus, plasmids generally are only used for subcloning arrays of larger fragments.

Bacteriophage lambda (λ) is a cloning vector that has the ability to package DNA into the phage particle, thus providing a stable means of maintaining insert integrity. Furthermore the high magnitude of phage multiplication provides access to large amounts of cloned DNA. Lambda cloning is also limited to relatively small fragments of DNA, with 25kb being the upper size limit. Thus, lambda vectors are typically used in small, targeted libraries. Cosmid cloning vectors are similar to plasmids but have the advantage of being to able hold larger inserts. This ability is gained from the addition of *cos* sites derived from the bacteriophage DNA. The *cos* sites facilitate proper packaging of the DNA clone by first ligating itself to the insert and creating circular DNA that is suitable to enter the capsids. Consequently cosmids can hold DNA fragments up to 45kb (Fairbanks et al., 1999).

The utility of cDNA libraries provides a means for restricting the DNA fragments examined to that of only actively transcribed genes. Copy DNA or cDNA libraries, made from RNA templates, are typically assembled using lambda or plasmid cloning vectors. mRNA are typically relatively short, thus, plasmid vectors are generally used (Ladish et al., 1999).

Of particular interest for this study are large insert libraries such as Yeast Artificial Chromosomes (YACs) and Bacterial Artificial Chromosomes (BACs). These libraries contain genomic fragments that can range from 0.1Mb to several Mb, and are essential for characterizing genomes as a whole (Wang et al., 1996). YACs have lost popularity among those using libraries due to their tendency to form chimeras, instability of clones, and difficulty in preparing DNA (Yu et al., 2000).

Bacterial Artificial Chromosome

A Bacterial Artificial Chromosome (BAC) can be defined as a genetically engineered *F* factor, isolated from *E. coli*, with the ability to hold fragments of DNA up to 300kb. BAC components consist of regulatory elements that confer selectable antibiotic resistance, regulate replication, and ensure low copy number within the host cell (Fairbanks, 1999). BAC DNA is typically introduced into the host cell using electroporation, a method that utilizes a brief but substantial electric pulse to open 'pores' in the cell membrane and permit insertion of large circular DNA elements into *E. coli* cells. The stability of BACs compared to YACs and the average insert size compared to cosmids, both demonstrate the superiority of BACs as vectors for constructing and

maintaining large insert libraries. Consequently, BAC libraries now exist in many significant plants species including but not limited to *Arabidopsis thaliana* (Wang et al., 1996), pearl millet (Allouis et al., 2001), potato (Song et al., 2000), sugarcane (Tomkins et al., 1999), *Medicago truncatula* (Nam et al., 1999), sorghum (Woo et al., 1994), barley (Yu et al., 2000) and pea (Coyne et al., 2000).

The maximum allowed insert size of BACs is 300kb. The large size promotes the logistics of cloning the total genetic composition of a species; however the task itself remains laborious. Briefly, total DNA is extracted in the case of pea, from seedlings that were grown in the dark for 3 days (Coyne et al., 2000). The total extract is then partially digested with six base pair restriction enzymes, and size selected twice, using pulsed field gel electrophoresis (PFGE). The size selected DNA is gel purified, ligated into the appropriate BAC cloning vector and electroporated into competent *E. coli* cells. The clones are then stored in 96 or 384 well microtitre plates (Nam et al., 1999; Coyne et al., 2000). The library can be screened using robot-arrayed high-density filters and DNA hybridization of radiolabeled probes or via PCR using pooled BACs (Nam et al., 1999).

An initial characterization of the BAC library with respect to plastid DNA contamination, average insert size, and the ability to locate specific sequences is necessary to assess the quality of the library. The evaluation of the average size of inserts within the library is typically accomplished by size fractionating random BAC clones using PFGE (Allouis et al., 2001; Tomkins et al., 2001; Song et al., 2000). The extent to which the library contains plastid DNA contamination can most easily be estimated by hybridization of radiolabeled plastid sequences to BAC array filters. Extra

nuclear DNA is typically not desired in a DNA library, thus it is important to identify clones containing such sequences. Certainly when dealing with a genome the size of pea, the more 'nuclear specific' the inserts of the library are, the fewer clones it would take to encompass a 1x coverage of the genome. For instance results from screening the potato BAC library (Song et al., 2000) indicated an estimated ~3% of the library contained BACs with organellar DNA inserts. When this number is applied to a library constructed from pea in which it is estimated that 250,000 clones would provide a 4-5x coverage approximately 7500 clones would contain cpDNA inserts. Thus, the importance of reducing the amount of contamination, or at minimum, acknowledging the presence of organellar DNA inserts is of importance to the project. The experiment demonstrates how the percent contamination can act as one indicator of the overall quality of the BAC library. It has been suggested that isolating tissue with relatively low amounts of chloroplasts can reduce the amount of chloroplast contamination. For example, isolating genomic DNA from source tissues that have either been light deprived, or are located within the innermost tissue of a particular plant, would decrease the amount of chloroplasts and thus cpDNA within the genomic DNA extract (Tomkins et al., 1999). Although plastid DNA contamination is generally undesirable, some, (1-3%) is expected (Yu et al., 2000).

Finally, the location of specific low copy sequences within the library can be determined in various ways. Anchored Sequence Tagged Sites (STS) and cDNA clones have been used to identify such sequences. Allouis et al. (2001) demonstrated the utility of using STS markers to locate clone homologs by choosing STS markers that are

genetically distributed throughout the pearl millet linkage map. Subsequent PCR amplification of a pooled BAC library with STS specific primers verified the quality of a pearl millet BAC library. Similarly, screening of pooled BACs allowed the identification of clones containing ethylene response genes in *Medicago* while establishing utility of the library (Nam et al., 1999).

BAC Applications

Although the previously described methods for confirming the utility of the library also can provide BAC insert data, such an approach is certainly not limited to just library characterization. The ability to locate BAC clones using molecular mapping markers, also known as positional cloning, has become an important step in both physical mapping and investigations of the sequences flanking the sequence of interest.

Resistance Gene Analogs (RGAs) provide a useful source of low copy sequences, appropriate for assessing condition of the library. Yu et al. (2000) used a set of RGA probes to screen BAC library filters of barley (*Hordeum vulgare*). One hundred twenty-one clones were identified that contained positive sequence homology. In a similar fashion, BACs containing sequences associated with cyst nematode resistance (CNR) in soybean (*Glycine max*) were identified from a soybean BAC library using RFLP-based markers tightly linked to the CNR locus (Danesh et al., 1998). Zwick et al. (1998) used six RFLP markers mapped to a linkage group in which *liguleless* is anchored, to screen a BAC library of *Sorghum bicolor*. Sixteen BACs were identified and physically mapped along chromosome 1 of *S. bicolor*. Results indicated that the physical map reflected the

linkage order established by RFLP mapping and also confirmed previous data indicating a conservation of the *liguleless* linkage group in rice and maize.

The large insert size in a BAC library allows the characterization of sequences flanking the original sequence identified. Screening flanking sequences using short anchor probes facilitates structural and evolutionary studies. Cregan et al. (1999) designed a targeted approach to identify simple sequence repeats (SSR) linked to previously established molecular markers in soybean. Two BACs identified by RFLP based markers conferring soybean cyst nematode resistance, were used to construct a subclone library. Hybridization screening of this subclone library with SSR probes identified 14 unique SSRs all localized within the genetically defined cyst nematode resistance locus.

BAC clones have been used to characterize complex areas of the genome. The arrangement and copy number of centromeric *Sau3*AI DNA fragments in sorghum was determined using subclones of BACs (Miller et al., 1998). Sequence analysis and *in situ* hybridization of the subclones indicated a complex arrangement of the *Sau3*AI fragments localized within the centromeres of sorghum. A method was devised to visually assess the complexity of the DNA found within the BAC, which could then be correlated with gene density within a particular portion of the genome. Using *in situ* hybridization of BAC subcloned DNA on isolated intact BAC clones, Jackson et al. (1999) was able to characterize areas within the BAC with respect to sequence composition and arrangement.

The informative DNA content of BACs when used in interspecies comparative studies can reveal mechanisms of gene or genome evolution. Areas of synteny between tomato and *Arabidopsis* were investigated by comparing tomato (*Lycopersicum esculentum)* BAC sequence, to *Arabidopsis* genomic sequence (Ku et al., 2000). Results obtained from comparing gene order and content, suggested that the driving factor influencing species divergence is likely to have been a series of large-scale genome duplications such as polyploidy rather than smaller rearrangements such as duplications and translocations. In a similar study, a BAC contig from *Arabidopsis* was used to propose a mechanism of genome expansion in *Brassica rapa* (Jackson et al., 2000). Rather than relying on sequence data, this experiment used a macroscale approach of arraying fluorescently labeled BAC contigs on extended DNA fibers of *Brassica*. The FISH signal was duplicated in the *Brassica* genome. However the orthologous loci retained the same length as seen in *Arabidopsis,* suggesting that large scale duplications, rather than an amassing of intergenic repeat sequences most likely contributed to the genome expansion.

Recently a BAC library was constructed of *Pisum sativum* (germplasm line PI 269818) genomic DNA with an estimated 1x genome coverage. As a prelude to the many applications this particular BAC library will contribute to, this study will assess the overall quality of the library. To test the utility of this BAC library, a series of experiments will be implemented to assess the quality of the BAC library. Initially, the amount of chloroplast DNA contamination will be estimated by quantifying the number of positive BAC clones identified by hybridization of chloroplast sequences to the

library. Secondly, the quality of the arrayed nuclear DNA clones will be investigated by hybridizing with characterized repetitive clones (25s rDNA). Finally, the ability to locate low copy (gene family) and unique sequences (single copy) within the library will be assessed using hybridization followed by a PCR verification, to find genetically anchored and characterized loci in the 1x *P. sativum* BAC library.

Materials and Methods

The BAC library, previously constructed by Coyne et al. (2000) of *Pisum sativum* DNA using germplasm line PI269818 of the Western Regional Plant Introduction (Pullman, WA.) had been prepared by cloning partially restricted DNA into vector pCLD04541 (V41). BIBAC vector V41 is a binary vector developed by Dr. J.D.G Jones and is suitable for maintaining large inserts in *E. coli* isolates with the additional capability of *Agrobacterium*-mediated transformation of large inserts into plants. Partially digested *Hind*III fragments were ligated into the cloning vectors, yielding an average insert size of 110kb. Consisting of approximately 50,000 clones, this library represents a 1x coverage of the total pea genome ($(5x10^4)(1.1x10^5)$= ~ 5000Mb). Robotics (Genomic Solutions) were employed to array each individual colony twice on a high density nylon membrane (Hybond N) (Coyne et al., 2000).

Dr. Clare Coyne, Washington State University- USDA NPGS, generously donated three sets of BAC filters to our program. Immediately upon receiving the BAC filters, they were stored at 4°C degrees until further use.

Probe Construction

Probes appropriate for identifying single copy and multi-copy DNA sequences in the *Pisum sativum* genomic BAC library were generated from genomic *Pisum* DNA (cv.'Sparkle') via PCR using previously characterized primers developed for genetic mapping in *Pisum* (see Table 1).

Table 1. Characteristics of Probes used to Screen BAC library

Gene Name	Accession Number	Gene Symbol	Product Length	Gene Organization
Chloroplast *trnL*-c & *trnL*-d	NA	*cpDNA*	510bp	High copy
25s rDNA	X52575	*25s*	400bp	High copy
Cytosolic glutamine sythetase	U28925	*Gsc*	1100bp	Single copy
Diminuto	D86494	*Dimin*	1200bp	Single copy
Ascorbate peroxidase	M93051	*Apx1*	1700bp	Single copy
Sucrose synthase	AJ012080	*Ssyn*	1850bp	Single copy
P393	AA430912	P393	516bp	Duplicate copies
P628	AA430910	P628	640bp	Single copy
T protein glycine decarboxylase complex	AJ222771	*Gdct*	1360bp	Single copy
PSL1 Lectin	M18160	*Lectin*	1050bp	*GF
Leghemoglobin	AB009844	*Lb*	1000bp	*GF
Ribulose 1,5-bisphosphate carboxylase small subunit	X00806.1	*RbcS*	610bp	*GF
Chlorophyll a/b-binding protein	M64619	*Cab*	660bp	*GF
Disease resistance response protein 49	U31669	*Drr49*	600bp	*GF
Ribosomal protein CL22	M60951	*Rpl22*	900bp	*GF
Farnesyltransferase	L08664	*Ftase*	900bp	*GF

***GF=Gene family member**

PCR products were immediately labeled using random priming and used for probe hybridization or cloned into plasmids for use at a later time. Genomic DNA extractions using the CTAB method followed previously described protocols (Torres et al., 1993): Freshly harvested meristem leaf tissue (about 1 cm^2) was placed into a small mortar. One milliliter of Extraction Buffer (50 ml 1M Tris-HCl, pH 7.5, 70 ml 5M NaCl, 40 ml. 0.5M EDTA, 340ml ddH_20, 1% mixed alkyltrimethylammonium bromide, and 0.4% β-mercaptoethanol) was layered on top of plant tissue. The tissue was ground with a pestal into a slurry. The slurry was incubated for 30 minutes at 65°C in the presence of 100μl of chloroform:isoamyl alcohol (24:1). The tube was to allowed cool to room temperature, filled to the top with additional 24:1 chloroform:isoamyl alcohol and vortexed to emulsify. The aqueous phase was separated from plant tissue and chloroform wash, using centrifugation in a fixed rotor centrifuge at 14*g* (Beckman Coulter Microfuge® 18 Centrifuge) for 5 minutes and carefully pipetted off the top. DNA was precipitated by adding 0.8-1.0 ml of ice cold 95% ETOH, which was initially layered on top of the buffer solution and incubated at room temperature for 10-20 minutes to allow precipitate to form at the solution interphase. Solutions were mixed by inversion to complete the precipitation of DNA. The precipitate was pelleted using a centrifugation of 10*g* for 5 minutes, and the extraction buffer and alcohol was carefully decanted. The pellet was washed once for ten minutes at room temperature with 1.0 ml of 0.2M sodium acetate in 75% ETOH and was washed a second time in 0.01M ammonium acetate in 75% ETOH for five minutes. DNA was dissolved in appropriate amount of TE (10mM

Tris, 1mM EDTA), typically ranging between 0.1 and 0.15 mls. One microliter of RNASE A was added to the samples to digest contaminating RNA.

PCR was performed in either an MJ Research® PTC-100 or PTC-200 "DNA Engine" thermocycler. Promega Taq or Platinum Taq (Invitrogen) polymerase was used in the PCR reaction for primer extension. Reaction preparation and solutions are as follows:

Reagent	µl/reaction	Stock Conc.	Reaction Conc.
ddH_2O	15.78		
$MgCl_2$ free buffer	2.5	10X	1X
$MgCl_2$	3.0	25mM	3mM
dNTP's	0.87	10mM each nucleotide	0.34mM each
Taq polymerase	0.1	5 units/µl	0.5units
Forward Primer	0.75	20µM	0.6µM
Reverse Primer	0.75	20µM	0.6µM
DNA	1.0		
	25.0 µl		

*Water was sterile

One microliter of template DNA was pipetted into individual microcentrifuge tubes. According to protocol above, a master mix was constructed by considering the number of reactions required and multiplying each specific addition by the number of reactions needed plus one or two extra reactions, to account for inherent pipetting errors. Twenty-four microliters of the master mix was aliquotted into each reaction tube, which already contained the one microliter of template DNA. If the PTC-100 was used, it was

necessary to top each tube with 1-2 drops of sterile mineral oil to reduce evaporation. The thermocycler was programmed to run according to the following protocol:

Cycling Parameters:

1-	94°C – 2.0 minutes	Initial Denaturation
2-	94°C – 1.0 minutes	Denaturation
3-	37-62°C – 2.0 minutes	Annealing
4-	72°C – 2.0 minutes	Extension
5-	35x to step #2	Cycling
6-	72°C – 8.0 minutes	Final Extension
7-	4°C – indefinitely	Holding Temperature

PCR products were separated on a 1% agarose gel to check the relative quality of amplification as well as size of amplification product. To isolate the PCR product from any unwanted contaminants such as unused primer and mis-primed products; the PCR product that corresponds to the expected size was extracted from the agarose gel and purified according to protocol supplied in "QIAquick Gel Extraction Kit" by Qiagen®. DNA was eluted from columns with minimal elution buffer.

Gel extract product was quantified by resolving the product next to "DNA Quantiladder" using agarose gel electrophoresis. Concentration was adjusted with TE to the standards required for the radiolabeling reaction.

Probe Labeling and Hybridization

Random priming reactions, using the Klenow fragment to fill gaps between primed random 10mers, were used to incorporate ^{32}P labeled dCTPs into the isolated

sequence. DECAprime™ II kits by Ambion® were used to facilitate the radiolabeling of between 1 and 3 probe sequences per reaction.

Upon completion of the labeling reaction, the product was cleaned using "Performa DTR Gel Filtration Systems" column by Edge Biosystems® according to manufacturer's protocol. This step is necessary to filter out unincorporated labeled nucleotides from previous reaction. Cherenkov counts were used to quantify the efficacy of labeling reaction using a Bioscan QC 400 xer scintillation counter.

Hybridization of probe to filters containing BAC clone arrays was carried out as follows. One or more filters was prehybridized in 20-30mls of "Prehyb solution" (5x SSC, 5x Denhardts (1g. Ficoll 400, 1g. PVP (Polyvinyl Pyrrolidone) , 1g. BSA (Bovine Serum Albumin), 1% SDS (Sodium Dodecyl Sulfate)), using a rotisserie style, hybridization oven, fitted for hybridization in tubes. The filter was prehybridized for 20-30 minutes during which time the probe and blocking DNA was prepared. Blocking DNA consisted of 20μl of sheared salmon sperm DNA (Eppendorf, 10mg/ml), that was denatured by boiling for 5 minutes, followed by a rapid cooling in an ice bath. Denatured salmon sperm DNA was added to the "Prehyb" within 5 min of the commencement of prehybridization.

Concurrent to prehybridizing and preparation of blocking, the radiolabeled probe was also denatured by incubating in boiling water for 10 minutes and likewise was rapidly cooled in an ice bath to avoid re-annealing. The entire amount of denatured probe was added directly to the prehybridization (now hybridization) solution and allowed to hybridize, overnight at either 37° (initially) or 55°C.

Stringency Washes and Autoradiography of BAC Filters

Following the hybridization reaction, the filters were subjected to a series of stringency washes. This step ensured that the filter was clean of hybridization solution and free of residual, non-specifically bound probe. The following is the protocol developed for utilizing stringency washes to prepare hybridized filters of the pea BAC library for autoradiography.

The hybridization solution was decanted into a proper radioisotope disposal container. The filter was rinsed of hybridization solution using a low stringency wash (LS=2xSSC, 0.1% SDS) for 10 minutes at 37°C, under slight agitation. Before administering an additional wash, the relative amount of isotope fixed to filter was estimated by scanning the filter with a Geiger counter. For clear resolution of signal it was empirically determined that the average intensity should be between 2000-3000cpm on the Geiger counter. This objective was met by applying additional washes; however, with each additional wash an increase in stringency and incubation temperature was used. For example, a more aggressive medium stringency (MS= 0.2xSSC/0.1%SDS) a 42°C incubation temperature was used. The MS wash was routinely followed by a high stringency (HS= 0.1xSSC/0.1%SDS) wash in which a 65°C incubation temperature was used. At times the filter needed only to be washed once at each stringency. However, on some occasions multiple washes at high stringency, in addition to the three routine washes were needed. Each wash was allowed to proceed for 15 minutes at which time a reading was taken with the Geiger counter. The number and stringency of washes used

was dependent on probe concentration and efficiency of isotope incorporation and thus had to be adjusted with each individual filter. A final 2x SSC rinse for 10 minutes at room temperature was used to remove excess SDS from washing steps. The filter was wrapped in one layer of cellophane and exposed to X-ray film within 24 hours of hybridization. It was beneficial to have a small amount of background, as it assisted in aligning the autoradiography to the filter map.

The filter was exposed to X-ray film (Kodak X-OMAT® AR Scientific Imaging Film) to visualize any hybridization signal. This exposure was carried out using a developing cassette supplemented with an intensifying screen and stored for 24-30 hours at -80°C. Subsequently, the cassette and film were allowed to thaw, and the film was developed according to manufacturer's instructions.

Hybridization Signal Scoring

The definition of legitimate clone was the presence of two probe signals within one square patch with the idea that each colony is duplicated within one patch. This system allows the scorer to differentiate between artifactual signal and a true hybridization. The two signals may be oriented in one of four ways: vertical, horizontal, diagonal right and diagonal left, with each arrangement representing the clone's plate of origin. Superimposing a map of the library over the developed film allowed the assignment of an identification number, based on the location and alignment of the signal, which is associated with the clone's origin within the library. Accordingly, This information was used to request the clone of interest from the library curator.

The library was initially screened for the presence of chloroplast DNA contamination. This was done using an annealing temperature of 37°C. *TrnL*-c and *trnL*-d primers, originally described by Taberlet et al. (1991), were used to generate the chloroplast DNA specific probes. The number of positive signals was counted and used to estimate the percent of clones with plastid inserts.

Clone Characterization and Sequence Analysis

A small aliquot of the freezer stock containing *E.coli* cloned BACs was streaked out on LB (Luria-Bertini Medium, premanufactured LB agar EM Science, EM Industries Inc. Gibbstown, NJ.) agar plates containing an appropriate selective agent. Using Vector PCLD045411 the selection was accomplished using ampicillin (20μg/ml) in the LB media.

Preliminary confirmation of the identity of the clone insert was accomplished by direct colony PCR. This method is similar to the described PCR reaction except that the template was obtained directly from the *E.coli* host, rather than purified DNA. The experiment was carried out by picking a small portion of an isolated colony, preferably a satellite, from the LB plate and adding the cells to a prepared PCR master mix that includes the primers suitable for amplifying the putative sequence from the insert. A duplicate PCR of each colony lift was performed using primers specific to sequences found on the vector as a positive check to confirm the polymerase chain reaction.

To further confirm the presence of the sequence of interest within the insert it was necessary to sequence a portion of the insert DNA- not including the rDNA and cpDNA

clones. Sequence template was obtained from the direct colony PCR reaction, which was either immediately sequenced or cloned into a vector and sequenced at a later time.

At times the sequencing reaction required a higher template yield than the direct colony PCR provided. This requirement was overcome by inserting the PCR fragment into a cloning vector suitable for sequencing. In the same manner as the previous gel excision, the PCR product was separated from the agarose gel and cloned into a vector. The cloning reaction, facilitated with TOPOCloneII™ by Invitrogen, ligates the PCR clone into a vector containing T3/T7 priming sites, thereby permitting subsequent sequencing reactions.

The sequencing reaction was carried out using ABI Prism BigDye Terminator v.2.0 100RR cycle sequencing kit (Applied Biosystems, Foster City, CA), to end sequence each PCR product. Sequences were generated in both forward and reverse directions using the T3 and T7 priming sites located on the cloning vector. Reaction conditions include mixing 0.5µl forward or reverse primer, 3.5µl PCR generated template and 2µl BigDye reaction mix. Reactions were performed on a Perkin Elmer GENEAMP PCR System 9600 with cycling parameters: 96°C for 10 seconds, 50°C for 5 seconds, and 60°C for 4 minutes. Products of sequencing reaction were cleaned and concentrated using a routine ethanol precipitation followed by re-dissolving in 0.8µl of Blue Dextran Dye. Sequencing products were applied to a paper comb and inserted into sequencing gel according to manufacturer's protocol. The sequencing gel was run on an ABI Prism 377 automated DNA Sequencer (Applied Biosystems, Foster City, CA). Chromatograph files were converted into forward and reverse sequences using Chromas™ version 1.45

software (Copyright Conor McCarthy, Griffith University, Southport, Queensland, Australia; available as freeware at http://www.technelysium.com.au/chromas.html). Sequences were edited at the 5' and 3' ends for low quality sequence using Bioedit Sequence Alignment Editor (copyright Tom Hall, North Carolina State University; available as freeware at http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Sequence analysis using BLAST (www.ncbi.nih.org/blast) verified the presence or absence of the putative gene sought within the BAC insert.