DEVELOPMENT OF A MOLECULAR MARKER TO TRACK APA G40199 INTROGRESSION IN COMMON BEAN FOR BRUCHID RESISTANCE

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Lucy Ione Mazaheri

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Program:
Genomics and Bioinformatics

July 2018

Fargo, North Dakota

North Dakota State University
Graduate School

Title

DEVELOPMENT OF A MOLECULAR MARKER TO TRACK APA G40199 INTROGRESSION IN COMMON BEAN FOR BRUCHID RESISTANCE

By

Lucy Mazaheri

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Phillip McClean
Chair

Dr. Robert Brueggeman

Dr. Justin Faris

Dr. Juan Osorno

Approved:

07/12/2018
Date

Phillip McClean
Department Chair

ABSTRACT

In common bean (*Phaseolus vulgaris*), the main seed storage pests are the bruchid beetles. Damage done to the seed by the larvae has a large impact on seed quality and yield. Arcelin (ARC), phytohaemagglutinin (PHA), and α-amylase inhibitor (α-AI) are linked seed storage proteins that form the APA locus on chromosome Pv04 and are associated with resistance. A major breeding objective is to introduce bruchid resistance into common bean from a resistant tepary genotype, G40199, by introgressing the resistant APA locus into susceptible common bean backgrounds. Here we developed a molecular marker that tracks the introgression. A set of PCR primers to the α-amylase inhibitor locus amplified a DNA fragment that showed a 45 base pair insertion in the middle of a lectin Leg_b domain. This enhanced locus characterization and insertion/deletion marker may preclude the need for bruchid resistance screening early in the breeding.

ACKNOWLEDGEMENTS

Support was provided by the Feed the Future Innovation Lab for Collaborative Research on Grain Legumes by the United States Agency for International Development (USAID) as part of Feed the Future, the U.S. Government's global hunger and food security initiative, under the terms of Cooperative Agreement No. EDH-A-00-07-00005-00. This work could not have been done in conjunction without the efforts of Dr. James Beaver and Dr. Kelvin Kamfwa. Dr. Beaver and Dr.Kamfwa were responsible for the development and scoring of the recombinant inbred line population that I used in this study.

I would also like to thank my committee members Dr. Juan Osorno, Dr. Robert Brueggeman and Dr. Justin Faris for setting aside time in your busy schedules to help me better my research and prepare my thesis for publication. I would like to extend a heartfelt thank you to Dr. McClean for being a consistent source of guidance throughout my time here at North Dakota State University. Dr. McClean's depth of knowledge has been invaluable to me as I navigated this journey from student to member of the scientific community. I am also so grateful for the wonderful lab environment that Dr. McClean has fostered and all of the people who were part of the team. Thank you to Rian Lee for guidance in my day to day struggles and for always being supportive and kind when I felt like I was in over my head.

Finally, I must acknowledge my support system for helping me along the way. I am forever indebted to my parents and sister for their unwavering support and unconditional love. As for my husband Mitchell, my gratitude, like my love, is boundless and this accomplishment would not have been possible without you.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

INTRODUCTION

Common bean (*Phaseolus vulgaris L.*) is one of the most important grain legumes in the developing world. Approximately 14 million hectares of arable land produces 11 million metric tons of common bean worldwide (Singh 1999; FAO 2005). In southern Africa, beans are the second most important source of dietary protein (Cardona, 1989). Additionally, this inexpensive crop has a positive mix of protein, complex carbohydrates, fiber, iron, folic acid concentrations that make beans a nearly perfect food (Pachicho, 1989). Damage done by post-harvest predators is one of the biggest problems facing both small and large-scale dry bean farmers. The bean weevils (*Acanthoscelides obtectus)* and the Mexican bean weevil (*Zabrotes subfasciatus)* are the most notable pests. These bruchids attack seeds in the field prior to harvest as well as in storage. In parts of Africa and Latin America, 30-35% of stored seeds suffer damage and/or contamination due to bruchid damage (Nahimana, 1992; Uebersax, 1996). The process currently used to identify resistance is time and labor intensive. Adult beetles are given the opportunity to lay eggs on seeds and in 30-60 days the number of exit holes from emerging adults can be scored (Singh, 1985).

Resistance *to Z. subfasciatus* was originally found in wild Mexican tepary bean accessions, and a major breeding effort was started to breed the resistance into susceptible *P. vulgaris* genotypes. Only recently has a source of strong resistance to *A. obtectus* been discovered in a tepary genotype G40199 (Kusolwa, 2007). Resistance appears to be associated with a group of linked seed storage proteins, arcelin (ARC), phytohaemagglutinin (PHA), and α-amylase inhibitor (α-AI) encoded by the APA locus on chromosome Pv04. Arcelin, originally thought to be the main resistance factor, was isolated and analyzed *in vitro*, and it was shown to not be solely responsible for resistance (Goosens, 2000). It appears a combination of arcelin, α-

amylase inhibitor, and other closely linked genetic factors have the greatest effect on resistance. As the interactions of these proteins became better understood, developing a molecular marker to track the introgression of the members of the APA locus became a goal in order to more quickly identify resistance in populations without having to infest and measure the susceptibility in every individual.

LITERATURE REVIEW

Dry Bean

Common bean (2n = 2x = 22) is the most important grain legume and accounts for more than 29 million metric tons of seed world-wide (McClean and Raatz, 2017). Beans are legumes that belong to the Fabaceae family which is the third-largest family of flowering plants. These are vitally important to agriculture and the environment. Legumes provide a substantial fraction of all nutritional protein and reduce the need for exogenous fertilizers due to their capacity for symbiotic nitrogen fixation (Kumar, Srivastava, Singh and Pratap, 2014). After the Poaceae grass family, this family represents the second most important family of crop plants. Grain legumes account for 27% of world crop production and provide 33% of the dietary protein consumed by humans while pasture and forage legumes provide a vital part of animal feed (Smýkal et al., 2015).

Legumes are comprised of several evolutionary lineages derived from a common ancestor ~54 million years ago (Wojciechowski, 2003). The Papilionoids form the largest clade, dating nearly to the origin of legumes and contain most cultivated species (Lavin, Herendeen, & Wojciechowski, 2005). Currently, based on morphological characters, three major groups are recognized as Papilionoid subfamilies: mimosoid legumes, Mimosoideae (four tribes and 3,270 species); papilionoid legumes, Papilionoideae (or family Fabaceae/Papilionaceae with 28 tribes and 13,800 species); and caesalpinioid legumes, Caesalpinioideae (or family Caesalpiniaceae with four tribes and 2,250 species) (Lewis, Schrire, Mackinder, & Lock, 2005).

Beans, along with soybeans (*Glycine max*) and cowpeas (*Vigna unguiculata*) belong to the most economically important tribe of the legume family, Phaseoleae. There are two gene pools of bean (Andean and Mesoamerican) that are thought to have diverged from a common

ancestor more than 100,000 years ago (Mamidi et al., 2012). The wild gene pools, found in South America and present-day Mexico, resulted in two unique parallel evolutionary lineages within the same species.

The domestication of legumes included changes in plant architecture, seed gigantism, transition to self-pollination, reduced seed dispersal, and loss of seed dormancy (Hammer, 1984). An increase in the seed size of domesticates compared to their wild relatives is suggested to be related to greater planting depth in agricultural systems, with larger seeds producing more vigorous seedlings (Abbo et al., 2011). At the same time, early farmers may have selected for a higher proportion of starch, oil, and protein.

Grain legumes were domesticated in conjunction with the domestication of grasses (Shahal Abbo, Lev-Yadun, & Gopher, 2012). However, more legumes were domesticated overall, and as a result the Fabaceae family contains the largest number of domesticated species. Pea, faba bean, lentil, pea, and chickpea are some of the world's oldest domesticated crops and were common in Fertile Crescent of Mesopotamian agriculture (Smýkal et al., 2015). Common bean in the Americas probably has the longest history as a domesticated species, originating in parallel in two separate centers of domestication: 1) the Andean mountains of South America, giving rise to the Andean gene pool, and 2) the Central American highlands and lowlands, giving rise to the Mesoamerican (Middle American) gene pool (Blair, Díaz, Buendía, & Duque, 2009).

Market Classes

There are many characteristics that distinguish beans the Andean and Middle American gene pools. Leaflet size and shape, leaf hairiness, length of internodes, pod beak origin, days to maturity, and growth habit are just a few of these characteristics (Singh, 1992). The Middle American gene pool encompasses the Mesoamerican, Durango, and Jalisco races. Middle

American races are generally small to medium seeded and range in seed shape from cylindrical, kidney, oval, to rhombohedric. Races are further divided into commercial market classes; most commonly pinto, great northern, small red, navy, small white and black in the U. S.

The Andean gene pool contains races Nueva Granada, Peru and Chile (Singh, 1992). Some of the more obvious characteristics that differentiate these races from the Middle American gene pool include medium (25-40 g per 100 seed) to large seeds (>40 g per 100 seed) and oval, round and kidney shaped seeds. Nueva Granada is the most common Andean race in the U.S and includes light red kidney, dark red kidney, white kidney and cranberry beans. The importance of the Andean beans in parts of Africa, Europe and South America has sparked a large increase in research in genetic improvement for many dry bean market classes.

Dissemination of the Domesticated Lines

It was originally thought that the center of origin of common bean was in the western hills of the Andes in northern Peru and Ecuador, and the wild bean was dispersed north into Colombia, Central America, and Mexico and south into Peru, Bolivia, and Argentina (Kami, Velásquez, Debouck, & Gepts, 1995). Recently, AFLP (Rossi et al., 2009) and simple sequence repeats (Kwak & Gepts, 2009) analyses of wild and domesticated *P. vulgaris* accessions, suggested, there was a severe bottleneck in the Andean populations prior to domestication, and that there was a Mesoamerican origin of the common bean. The center of origin of common bean is now thought to be in Mexico which coincides with the domestication maize (*Zea mays*) and squash (*Cucurbita* spp.) (Zizumbo-Villarreal, 2010).

Pinto and great northern beans seem to have migrated from central Mexico into the central region of the US and were moved northward by the Native Americans into Canada. Elsewhere, navy and black beans were carried from the Caribbean to the east coast of the US and

northward into the Great Lakes region. Kidney beans, domesticated from the wild Andean gene pool, were moved along the slave trade route to Europe and on to eastern Africa (Kelly, 2010).

Grown in mild climates with a mean temperature of 15-25°C, beans are considered a short-day crop and can survive in environments up to 3000 m above sea level (White and Laing, 1989). Beans are often intercropped either at the same time or in relay with maize which acts as a support for climbing varieties. Beans provide additional income and serve as a nutritional backup if the maize crop grows poorly. The crop possesses variability for photoperiod sensitivity, growth habit, and disease resistance in the different market classes. This contributes to diversity in production systems in a variety of climates spanning North America, Latin America, Europe, Africa and Asia (Singh, 1992; Woolley et al., 1991).

Bruchids

Bruchids belong to the order coleopteran in the family bruchidae. The word bruchid is a general term for beetles, such as the bean weevil (*Acanthoscelides obtectus* Say.) and the Mexican bean weevil (*Zabrotes subfasciatus* Boh.). These are the two most common and devastating seed storage pests of common bean (Kornegay & Cardona, 1991). Both bruchids feed on other species of Phaseolus including the wild tepary (*P.acutifolius* A. Gray), runner (*P. coccineus* L.), and lima (*P. lunatus* L.) beans. The larvae are most economically devastating to the common bean, and damage done to these seeds by the bruchid larvae can cut the marketable of the beans by 7-40% (Slumpa and Ampofo, 1991).

The *Z. subfasciatus* originated in Central America and was described in the early 19th century. It is a small black beetle that ranges in size from 2.0-3.0 mm in length and 1.3-1.5 mm in width. It is tropicopolitan and favors warm temperatures. Other closely related hosts include *Glycine max, Phaseolus acutifolius, Phaseolus coccineus, Vigna angularis, Vigna mungo,* and

Vicia faba. Commercial distribution of dried beans has allowed this seed storage pest to infiltrate South America, Africa and parts of India (Southgate, 1964). *Z. subfasciatus* females oviposit only on seeds that have fallen from the pod or are still inside a dehisced pod. This makes infestation detection before the prepupal stage challenging. Once the pupal stage has commenced, the larva has compromised the inner surface of the integument with a hole to facilitate emergence of the adult beetle (Pierre and Pimbert 1981).

Z. subfasciatus only infests stored beans, and not those in the field. Some control strategies are only applicable on a small scale. For example, farmers can place their seeds into storage and add dust or ashes to fill up the spaces between the seeds making it more difficult for the females to lay their eggs. Tumbling the seeds inside the storage containers and frequent drying in the sun helps mitigate infestation by creating unfavorable conditions for the pest (Quentin et al. 1991). The use of pesticide controls is available, but they come at a financial and environmental cost that is often times too great for small farmers. Phosphine fumigants and pyrethroid compounds are extremely toxic and can have a negative impact on the environment and effect the health of applicator and consumer (Agrofit 20011). Studies in developing countries have found an incident of acute pesticide poisoning in 35 out of every 100,000 workers every year (Henao, 2002). In search of an alternative form of control, research moved towards botanical pesticides and found neem (*Azadirachta indica* A. Jus), eucalyptus (*Eucalyptus citriodora* Hook), lemongrass (*Cymbopogon citratus* Stapf.) and camellia (*Camellia oleifera*) essential oils all repelled *Z. subfasciatus* weevils from laying eggs on *P. vulgaris* (Oliveira 1999 and França 2012). Another method of controlling infestation uses diatomaceous earth (DE). DE is a soft sedimentary rock that is easily crumbled into powder. It consists of fossilized remains of a hard-shelled algae called diatoms, from which it gets its name. Due to the chemical properties

of DE, the powder absorbs lipids on the exoskeletons of insects causing desiccation and death (Golob, 1997). DE has no known toxic effect on animals or humans and is in turn a good alternative as a safe pest control.

A. obtectus Say, commonly known as the bean weevil, originated in Central America and is specialized on *P. vulgaris* L. (Delgado-Salinas, 1999). It is now found worldwide due to commercial trade, distribution, and its ability to reproduce in slightly cooler and higher altitude climates. It is characterized as a small beetle, 2-5 mm in length, are light to dark brown with a red posterior border. Adult females lay eggs in groups of 2-20 in or on the seed pods and can lay up to 200 eggs. The first instar larvae emerge within 30-45 days. A few days after that the larvae molt into 2nd instar larvae and spend the next 3-4 weeks consuming seeds inside the pod. The larvae then enter the holes they formed in the seed and pupate. The entire life cycle of a single generation takes approximately 100-110 days and when in favorable climates (22-29°C) multiple generations can reproduce one after another. Due to the neotropical nature of the insect, any temperature higher or lower than 22-29°C can drastically decrease average fecundity of the laid eggs.

A. obtectus infestation control is more complex because it starts in the field and continues in storage. In an effort to combat infestation from the field, Schmale and colleagues (2002) researched the use of the parasitoid *Horismenus ashmeadii* (Dalla Torre) as a biological control. In the field the parasitoid would feed on the *A. obtectus* larvae and reduce infestation. However, the effectiveness did not extend into storage. There are a few insecticides for *Z. subfasciatus* that are effective for the control of *A. obtectus* as well. DE, malathion, aluminum phosphide, magnesium phosphide, gaseous phosphine, carbon dioxide, and methyl bromine are the most common. However most of these methods either have widespread insect resistance or require

well-sealed and climate-controlled granaries. This poses a problem for small subsistence and shareholder farmers who may not have these types of facilities or agents.

Logistic issues, like these, highlight the need for research to develop a genetic resistance to these pests. It is known that certain seed storage proteins are toxic to these bruchids and confer a natural resistance in some wild Mexican varieties (Osborn, 1988). Utilizing these proteins and introgressing them into breeding lines would reduce the use of these insecticides and would provide farmers with limited storage capacity a means to protect their yield.

APA Locus and Bruchid Resistance

One defense mechanism the Fabaceae family employs against insects, birds, or mammals is to accumulate antinutritional seed storage proteins. Specifically, in *P. vulgaris*, the seed storage proteins include phaseolin, lectins, phytohaemagglutinins, and lectin-like proteins (LLP). Lectin-like proteins are associated with plant defense in *Phaseolus* (Chrispeels and Raikhel, 1991). Lectins reversibly bind to carbohydrate molecules (Peumans and Van Damme, 1995). This aids in defense by interfering with the carbohydrate metabolism of an organism that has ingested the plant. Phytohemagglutinin is a major lectin found in common bean and is homologous to arcelins and α-amylase inhibitors, all of which are involved in bruchid resistance (Chrispeels and Raikhel, 1991).

Evolutionarily, the APA (Arcelin/Phytohemagglutinin/α-Amylase inhibitor) locus found in *Phaseolus* species is a result of an ancestral paralogous duplication event that produced the true lectin and the lectin-related genes. Eventually, these lectin-related genes evolved through a second duplication event into the α-AI and ARC, which is only found in some wild Mexican varieties (Lioi et al., 2003). This locus, which encodes all three proteins is located on chromosome Pv04 (previously B4) in the *P. vulgaris* genome (Nodari et al., 1993) and is

estimated to be 500 kbp in size (Galasso, 2005; Sparvoli, 2008; Gepts, 2008). These genes have no introns and are between 732-825 base pairs. These are homologous with high (45-85%) amino acid sequence identity. Varying degrees of resistance to pests have been attributed to the APA locus, however the molecular effect that confers resistance is still not known. A study by Kami and colleagues developed a figure from a BAC clone that helps visualize how the locus is thought to look (Fig. 1). The figure shows six purple arrows that denote APA genes. Interestingly the reference genome only has three gene models (Phvul.004G158000, Phvul004G158100, and Phvul004G158200) that are associated with the APA locus. This highlights how many more genes could be involved in the resistance and just not accounted for in the reference as of now.

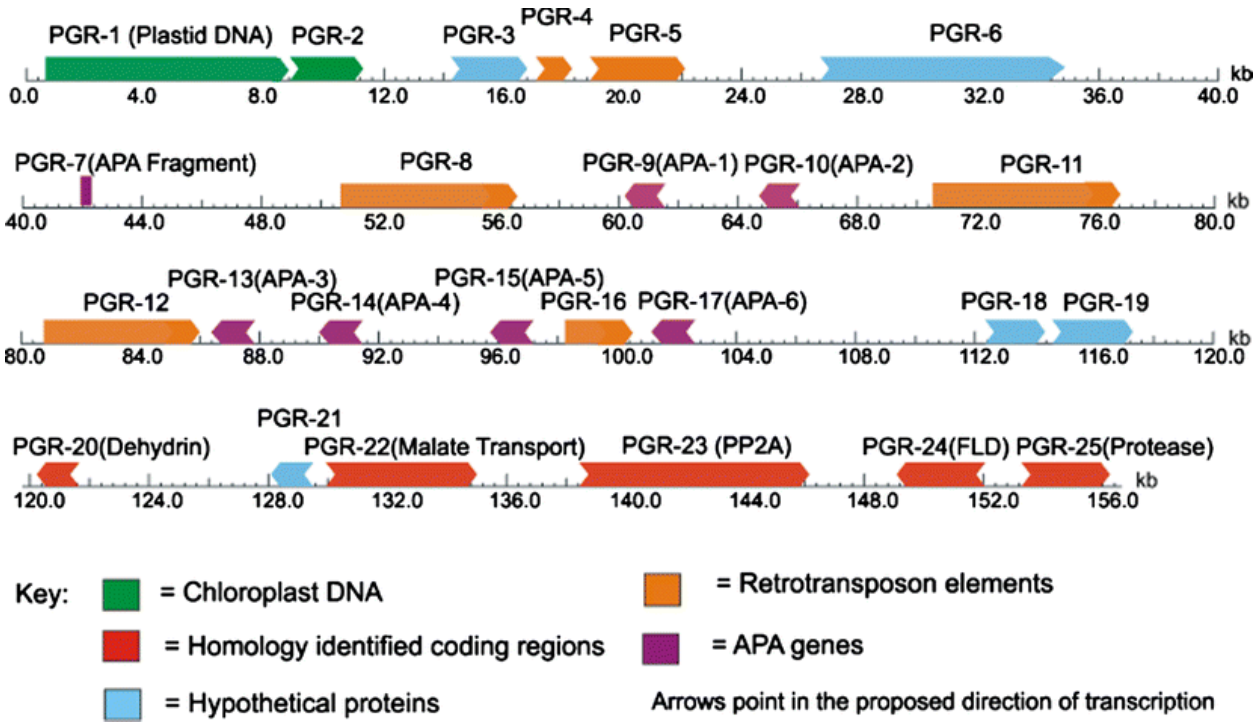

Figure 1. Schematic of the distribution of putative genes identified in P. vulgaris G02771 BAC clone 71F18. (From Kami et. al. 2006, Fig 3)

Arcelin (ARC) is a lectin related protein that may aid in resistance to bruchids. Although the specific mechanism is unknown it is thought that this protein is difficult to digest and may contribute to a change in gut structure and epithelial cells (Bardocz et al., 1995; Minney et al., 1990). Arcelin was originally found in wild common bean accessions from Mexico (Osborn et al., 1988a and 1988b). There are eight known variants that are differentiated based on amino acid sequence and electrophoretic polypeptide comparison: Arc-1, Arc-2, Arc-3 and Arc-4 (Osborn et al., 1986; 1988b; Hartweck et al., 1991), Arc-5 (Lioli, 1989; Goossens et al., 1994), Arc-6 (Santino et al., 1991), Arc-7 (Acosta-Gallegos et al. 1998) and Arc-8 (Zaugg, 2013). Resistance from all of these variants is effective on *Z. subfasciatus*, but Arc-4 is the only variant to show resistance for both species. In regards to differences in nomenclature in previous publications, variants in this document are denoted by Arc-#

Osborn and colleagues initially identified Arc-1, Arc-2, Arc-3 and Arc-4 by electrophoretic differences after screening 106 wild bean accessions that contained a novel protein band that had not previously been reported (1986). The source for Arc-1 is PI325690 or G12882. More research was conducted to find the differences in these initial arcelin variants. The source for Arc-2 was G12866, and through the comparison of cDNA clone sequences it was found to be 99.3% homologous with Arc-1 (John, 1990). Arc-5 (Lioli, 1989) was further characterized and revealed it was made up of three polypeptides (Arc-5a, Arc-5b and Arc-5c). Using G02271 as the source, two cDNA clones, Arc-5-I and Arc-5-II, were isolated and found to have 98.3% similarity at the nucleotide level (Goossens, 1994). Arc-4-I and Arc-4-II were discovered in 1994 from G12949 as the source. Arc4-II was later found to be identical to Arc-3-II (Lioi, 2003). Arc-6 was later found to be highly similar to Arc-1 and Arc-2 based on cDNA sequence from G11051 by the same group. Arc-7 was discovered in 1998 from wild accession

G24584, and like Arc-5 made up of two polypeptides (Acosta-Gallegos, 1998). Lastly, in 2012 Arc-8 was discovered in an accession named QUES and was found to be closely related to Arc-4 which are the only variants resistant to both types of bruchid beetle.

The phytohemagglutinin (PHA) family consists of two major polypeptides, PHA-E and PHA-L. Both of these isoforms cause agglutination but have different targets. PHA-E acts on red blood cells and PHA-L acts on white blood cells. The amino acid sequence identity of these tandemly linked genes is 82% similar (Rudiger and Gabius, 2001). PHA can bind to carbohydrates in the insect's digestive system and prevents the absorption of nutrients across the intestinal wall (Sakthivelkumar, 2013). A study Pusztai and colleagues demonstrated the antinutritional effect of PHA on rats when they were fed purified PHA or bean seeds (1979 &1983).

α-Amylase inhibitor (α-AI) is the third lectin-like proteins that makes up the APA locus. This protein inhibits the α-amylase activity in some insect species. The α-amylase activity aids in starch breakdown, and interrupting that process deprives the insect of carbohydrates and retards larval development. There are multiple allelic variants of α-AI that combined with other proteins have varying levels of resistance to bruchid attack (Ishimoto and Chrispeels, 1996; Morton et al., 2000). Genotypes with α-AI-2 are more resistant compared to α-AI-1, however studies have shown that the resistance mechanism is complicated, and high expression of α-AI-2 is not enough to confer resistance on its own (Nishizawa et al., 2007). α-AI variants have been found in *P. acutifolius* that are successful in protecting beans from *Z. subfasciatus* due to their similarity to α-AI-2 (Yamada et al., 2001). The complexity of this region and the possibility that there may be protein interactions that are not well understood yet makes it difficult to utilize this locus correctly.

Breeding and Hybridization

One way to cultivate bruchid resistance would be to incorporate proteins with known resistance properties into popular common bean genotypes. This was done originally with a backcross-breeding scheme with the arcelin gene (Osborn et al. 1988b; Cardona et al. 1990; Kornegay and Cardona, 1991; Kornegay et al. 1993; Misangu, 1997; Hartweck et al. 1991). Arl-1 was the first variant introgressed into cultivated beans, and Arl-2, Arl-3 and Arl-4 were soon to follow (Osborn et al. 1998b; Romero-Andreas et al. 1986; Cardona et al. 1990; Hartweck et al. 1991; Kornegay et al. 1993). In order to increase accumulation of arcelin in the seed, a phaseolin null allele was introduced to the lines with variants 1, 2 and 4 of arcelin. The increased arcelin activity of these 'Sanilac' mutant (SMARC) lines have been highly successful and imparting resistance to *Z. subfasciatus* but are still ineffective against *A. obectus* (Cardona et al. 1990; Kornegay and Cardona 1991; Kornegay et al. 1993; Hartweck et al. 1997; Acosta-Gallego et al. 1998; Paes et al. 2000; Sales et al. 2000).

With the ability to isolate and compare the resistance of different lines with a variety of combinations of seed storage proteins, it is probable that other factors are involved. Harnessing the ability to transfer the whole locus from one resistant line to a susceptible line could be a way to preserve resistance without *a priori* understanding of the exact mechanism. Transferring genes from *P. acutifolius* to *P. vulgaris* is a difficult process due to genetic and reproductive barriers. Interspecific hybrids from these two species present a challenge due to a small number of offspring with low fertility; abnormal chromosome recombination; need for embryo rescue; and high linkage drag from the tepary parent (Haghighi et al. 1988; Mejia-Jiménez et al. 1994; Rabakaorihanta et al. 1980). Despite the challenges, it is possible. Singh and Muñoz (1999) successfully transferred common bean blight (*Xanthomonas campestris* pv. phaseoli) resistance

from tepary to common bean. Modifications to the congruity backcrossing method (CBC) resulted in the double congruity backcross (DCBC) method which resulted in increased hybrid fertility and gene transfer (Mejia- Jiménez et al. 1994). As the backcrossing system for APA locus becomes more reliable and introgression from tepary G40199 is more successful, a method to track that introgression is becoming greater. Transferring and tracking the APA locus from G40199 to other common bean cultivars will be an important step towards introducing resistance to both *Z. subfasciatus* and *A. obectus.*

In 2007 P.M. Kusolwa and J.R. Myers discovered the utility of G40199 as it relates to its high level of *A. obectus* resistance. Following a F_1 interspecific hybrid between G40199 and *P. vulgaris*, subsequent backcrosses with ICA Pijao were made to recover fertility and the development of a F_2 interspecific population from G40199 and Brown Tepary. The population showed a single dominant gene coded for a 33 kDa seed storage protein in G40199. Subsequent lines developed from these lines show significant bruchid resistance and the 33 kDa protein was linked to co-expression of arcelin-like and PHA protein subunits.

The main objective of this study is to develop a molecular marker to track the introgression of the APA locus from *Phaseolus acutifolius* into *Phaseolus vulgaris* in order to utilize its resistance to seed storage pests. A secondary objective as to better characterize the locus itself through sequence analysis and phylogenetic mapping.

MATERIALS AND METHODS

Plant Material

Four bruchid resistant and four bruchid susceptible genotypes were used in this research (Table 1). *P. acutifolius* genotype, G40199, is the source of bruchid resistance. The four resistant genotypes are: PR1012-29-3-1A and PR1012-3-3A, medium sized red beans; and PR1464-1, and PR1464-6, black beans. The four susceptible genotypes are: XRAV-40-4 and Rojo, used as parents to develop the new resistant lines; and Verano and Badillo, susceptible checks. G40199 was originally crossed with ICA-Pijao, then backcrossed twice. Using a similar scheme, Rojo and SMARC-2-PN-1 were crossed and backcrossed three times. The progeny of those two crossing schemes were crossed to develop the PR1012-29-3 line. The individuals PR1012-29-3-1A and PR1012-29-3-3A were selected from that line and used as the resistant parents in a cross with XRAV-40-4 to develop the PR1464-1 and PR1464-6 genotypes. Table 1 shows the pedigrees of these lines below.

Table 1. Resistant and susceptible genotypes used to test marker

Genotype	Resistant/Susceptible	Pedigree	Size/Color
PR1464-1	Resistant	PR1012-29-3/XRAV-40-4	Black
PR1464-6	Resistant	PR1012-29-3/XRAV-40-4	Black
PR1012-29-3-1A	Resistant Parent	Rojo#3/SMARC-2-PN-1//ICA-Pijao#2/G40199(Tepary)	Medium/Red
PR1012-29-3-3A	Resistant Parent	Rojo#3/SMARC-2-PN-1//ICA-Pijao#2/G40199(Tepary)	Medium/Red
XRAV-40-4	Susceptible Parent		Black
Rojo	Susceptible Parent		Red
Verano	Susceptible Check		White
Badillo	Susceptible Check		LRK
G40199	Resistant Source		Tepary

Genomic DNA Preparation and PCR Analysis

Seeds from resistant and susceptible genotypes were planted in the greenhouse and grown to the first trifoliate stage. One to two young trifoliate leaves (approximately .1 g of tissue) of each genotype were collected from multiple plants and stored in the -80 C freezer. The samples were ground into powder using liquid nitrogen with a plastic pestle. Genomic DNA was extracted using the IBI Plant Genomic DNA Mini Kit IB47231 as per the kits protocol and quantified using a Nanodrop (www.ibisci.com). Each individual DNA sample was brought to a concentration of 20 ng/ul using MilliQ water. These samples were stored in the -20 C freezer until further use. Using an online sequence alignment program, MultAlin (http://multalin.toulouse.inra.fr/multalin/), sequences from Zaugg et. al. 2013 were obtained from GeneBank and aligned to show significant differences. Those GenBank IDs are as follows: Arc-1 (P19329), Arc-2 (P19330), Arc-4-I (CAD29134), Arc-4-II (CAD58679), Arc-4-III (CAD58657), Arc-5a (CAA90585), Arc-5b (CAA85405), Arc-5c (AAF23725), Arc-6 (CAA04960), Arc-7 (CAD28677), α-AI-1 (AAA33769), α-AI-2 (CAD28676), ARL-4 (CAD28840), ARL-4-I (JQ675761), PHA-E (CAD28837), PHA-L (CAD28838), Lec-4B17 (CAD29133). We were then able to develop internal primers that were specific to the insertion/deletion difference now called INS45 (see results Fig. 1). The sequence of the internal primer, named, α-AI-1 is: (5' – CCCGTCCAGCCCGAATCCAAAGG- 3'and 5'-CATCCCAAGGCACGCTTTTGATATCG-3'). DNA was amplified using a standard PCR protocol across all eight genotypes. The cycles were as follows: 3 mins at 95 C for initial denaturation, 30 sec at 95 C, 30 sec at 55 C, 1 min at 72 C for 45 cycles, followed by 10 min at 72◦C for one cycle. Each 20 µl PCR reaction consisted of 1.2 µl of DNA at 20 ng/µl, 2 µl of 10x PCR buffer that contained 15 µM $MgCl_2$, 1 µl of 10 µM total dNTPs, 1 µl of each forward and reverse primers at 5 µM, and 1-2 units of Taq-

polymerase. The PCR product was visualized on a 3% (w/v) agarose gel stained with ethidium bromide, and band sizes were estimated by comparison to the 1 KB Plus DNA Ladder (Invitrogen Cat. No. 10787-026) (www.thermofisher.com).

Marker Development

The α-AI-1 PCR product showed a single band that was approximately ~275 base pairs (bp) in length for the resistant genotypes and a slightly shorter product (~225 bp) that was visible on the susceptible genotypes. The PCR products for the resistant and susceptible genotypes were visualized on a 1.5% agarose gel, and individual bands of the resistant and susceptible genotypes were excised. The gel fragments were melted and purified with a Promega Wizard SV Gel and PCR Clean-up System (Cat. No. A9281) following the kit's instructions (www.promega.com). The products were sent for Sanger sequencing to Eton Biosciences (http://www.etonbio.com/). The sequences were analyzed and edited using the Staden package (Bonfield, 1995).

To gain further understanding of this locus additional primer sets were developed. Three sets of primers were designed to capture full gene sequences for alpha-amylase inhibitor, phytohaemagglutinin, and arcelin from BAC clone 71F18 (GenBank, DQ323045; Kami, 2006). The primers were designed to the start and stop sites of the AAI, ARC and PHY genes.

Table 2. Oligonucleotide sequence specific primers

Target gene	Primer Sequence
PHY-gene	Forward: 5'-TCAACGAAACCAACCTTATCCTCCAACG- 3'
	Reverse: 5'-GGTCCTGAATTGTTGGGAACGTCGAT-3'
ARC-gene	Forward: 5'-TCCACCAAGGCGATCCCAAACTTA-3'
	Reverse: 5'-ATTGTTATTGCTCCTAGAGCGGATACTGAATG-3'
α-AI-gene	Forward: 5'-ATGGCTTCCTCCAAGTTCTTCACTGTCCT-3'
	Reverse: 5'-CTAGAGGATGTTGTTGAGGAGGATGTTGGAACGTTGAGATG-3'
α-AI-1	Forward: 5'-CCCGTCCAGCCCGAATCCAAAGG- 3'
	Reverse: 5'-CATCCCAAGGCACGCTTTTGATATCG- 3'

These new primers were used to amplify DNA using the same protocol as previously stated across the same eight genotypes and run on an agarose gel. Bands were visible on six of the eight genotypes for the AAI-gene primer (four resistant and two susceptible). All eight genotypes amplified a product with the PHY-gene primer set. The amplified products were isolated from the agarose gel and purified using the Wizard SV Gel and PCR Clean-up System (Cat. No. A9281) following the kit instructions (www.promega.com). Samples were then sent for sequencing to Eton Biosciences (http://www.etonbio.com/).

The α-AI-1primers (Table 2) were then used on a population of recombinant inbred lines (RILs) to test the marker's effectiveness by comparing genotype to the resistance field scoring. The RILs were developed by Dr. Kelvin Kamfwa by crossing AO1012-29-3-3A, as the resistant source, and Badillo, as the susceptible parent. AO-1012-29-3-3A (Reg. No. GP-299, PI 675563) was derived from a population developed by Dr. Paul Kusolwa and Dr. James Myers at Oregon State University. The University of Puerto Rico bean breeding program received BC3 F4 lines from the cross 'Rojo'*3///SMARC-2-PN-1//'ICA Pijao'*2/G40199. These parents were used to

make the recombinant inbred line of 215 individuals. Seeds from this population were grown in the greenhouse and grown to the first trifoliate stage. Genomic DNA was prepared as previously stated, and the DNA was amplified using the α-AI-1 primers under standard PCR protocol. The PCR products were visualized on a 3% (w/v) agarose gel with the same conditions and compared to a 1 kb molecular weight standard.

Scoring of susceptibility in each individual was done by placing recently emerged adult pairs in containers with 10 well dried seeds. The adults were allowed to mate and then removed from the containers. The seeds are then monitored for visible egg development and checked for perforations made by the new emerging adults at 30, 45, and 60 days after infestation. The score is made up of two numbers, the first is a scale to measure the total number of perforations on the seeds, 1 = 0 perforations, 2 = 1-5 perforations, 3 = 6-10 perforations, 4 = 11-15 perforations, and 5 = >15 perforations. The second number is the proportion of seeds with perforations out of the total infested/evaluated seeds. For example, if a score was (2, 3/10), out of the 10 seeds evaluated there were 1-5 perforations on a seed, and 3 seeds were perforated.

Cloning and Phylogeny of APA PCR Products

Because the amplification products for the four resistant genotypes using the PHY-gene, ARC-gene, and α-AI-gene prime pairs were not monomorphic, the product from using the PHY-gene, ARC-gene primers were purified. The purified products were cloned using a Promega pGEM-T vector system (Cat. No. A3600) (http://www.promega.com/). Using X-gal blue/white screening, five individual white colonies per PCR reaction were grown in liquid culture for approximately 24 hrs. 1.5 mL of the cell suspension was centrifuged into a pellet and resuspended as per the directions in the Promega Wizard Plus SV Minipreps DNA Purification System (Cat. No. A1330) (http://www.promega.com/). The DNA from the individual clones

were sequenced as previously stated. To compare the amplification products using the primer pairs designed to amplify fragments the APA locus and not another region of the genome, the clones with a translated sequence that contained a lectin leg_B domain (PF00139) (www.pfam.xfam.org), were selected from the collection of proteins predicted from the annotation of the common bean genome v2.0 assembly (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Pvulgaris). The sequences of the cloned fragments from the α-AI and PHY primer pairs, and the ARC sequences from the previous primer sets were added as well. A MUSCLE (Edgar, 2004) alignment of all lectin leg_B domain containing protein sequences and predicted proteins of all the amplification products was performed in MEGA v.7 (Kumar et al. 2015). The alignment was then used to develop a neighbor-joining tree with a 1000 bootstrap value using MEGA v.7 with the default parameters.

RESULTS

INS45 Marker Screening

Amplification with the α-AI-1 PCR internal primers produced products varied in size between the resistant and susceptible genotypes (Fig. 2). The gel shows that lanes 1-4 are slightly larger fragments than in lanes 5-9. Based on the phenotypic data in Table 1, the resistant genotypes correspond with the larger fragment (~275 bp) and the susceptible genotypes correspond with the smaller (~225 bp) bands. These fragments were sequenced, and sequence data revealed an in-frame 45 base pair insertion in the resistant genotypes that was absent in the susceptible genotypes. (Fig. 3). Those individuals with the presence of the 45-base pair insertion will be referred to as containing the INS45 band.

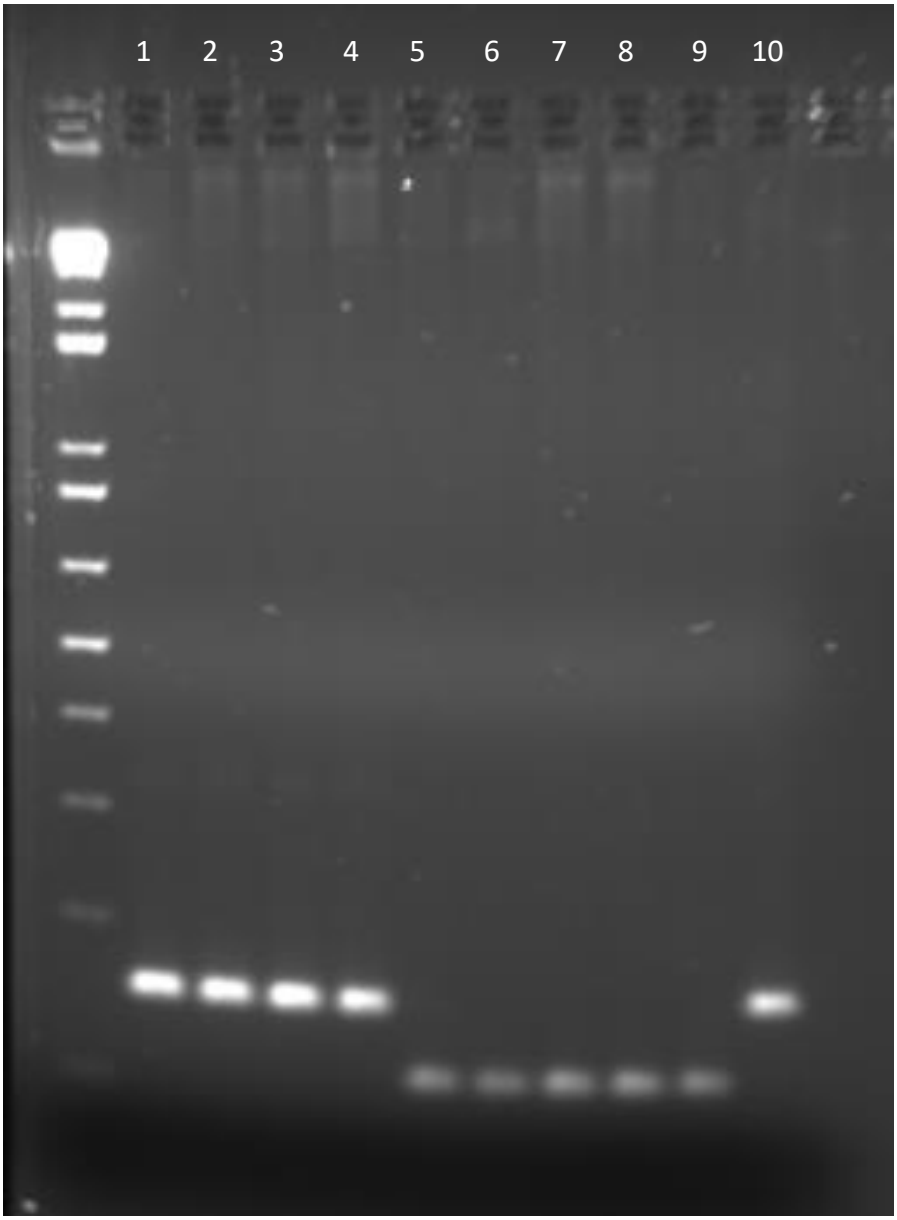

Figure 2. Differences in PCR products generated using the α-AI-1 internal primers with resistant (1-4) and susceptible (5-9) genotypes. Lane 1, PR1464-1; lane 2, PR1464-6; lane 3, PR1012-29-3-1A; lane 4, PR1012-29-3-3A; lane 5, XRAV-40-4; lane 6, Rojo; lane 7, Verano; lane 8, Badillo; lane 9, G19833; lane 10, G40144

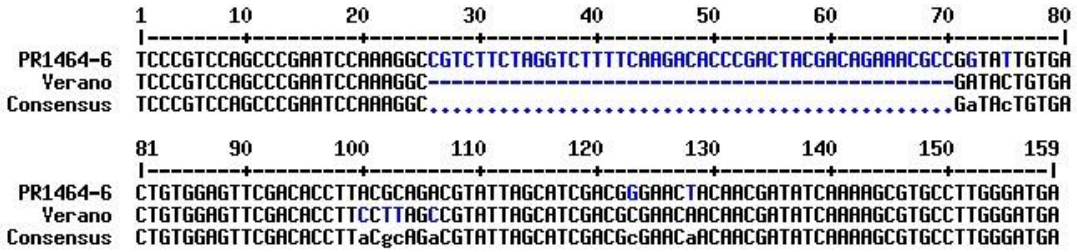

Figure 3. Sequence obtained from the amplicons using α-AI-1 internal primers showing the 45 bp insertion in the resistant PR1464-6 compared to the susceptible Verano

After amplifying members of the SA recombinant inbred lines with the α-AI-1 primers, the samples were scored relative to band size and compared to the field resistance severity score. Of the 215 members of the population, data was available for 206 individuals. As shown in Table 3, five individuals were scored as field resistant and had the INS45 band that is associated with resistance. It also shows that 159 individuals that were scored as susceptible and did not have the INS45 band as expected. Among the phenotypically resistant, six lacked the INS45 insertion, and 20 susceptible individuals contained the INS45 insertion. Also, 14 individuals contained both bands. The observed number of individuals with the resistance is much lower than the expected 1:1 ratio. This segregation distortion could be accounted for by the fact that this RIL was developed to track introgression from one species to another. A contingency chi-square test was performed to determine if the presence of the INS45 marker phenotype was associated with resistance. The p-value for this test was p=0.0011 suggesting that the presence of INS45 insertion marker was associated with resistance.

Table 3. Number of individuals from the SA population that had the INS45 band or did not have the INS45, compared to the original resistance scoring method.

	Resistant	**Susceptible**
+ INS45	5	20
- INS45	6	159
Both	3	11

Sequence Analysis of Leg_b Domain Containing Sequence

The α-AI-gene primers were used to amplify DNA of eight test genotypes. These PCR products were slightly different in size. The resistant genotypes produced a product of ~750 bp (Fig. 4, lanes 1-4). However, among the susceptible genotypes, amplification products from XRAV-40-4 and Verano were present and larger than expected, while no product was generated from susceptible genotypes Rojo and Badillo DNA (Figure 4, lanes 6 and 8, respectively).

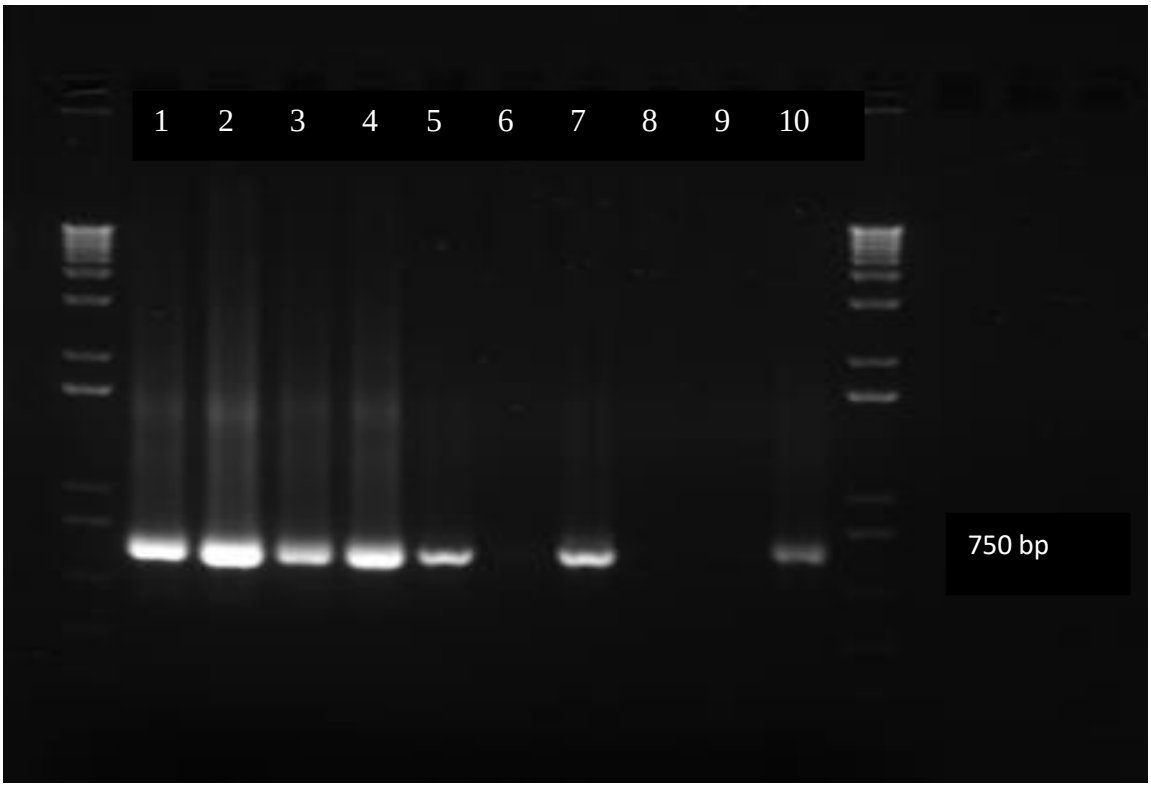

Figure 4. Differences in resistant (1-4) and susceptible (5 and 7) bands with α-AI-gene primer. Lane 1, PR1464-1; lane 2, PR1464-6; lane 3, PR1012-29-3-1A; lane 4, PR1012-29-3-3A; lane 5, XRAV-40-4; lane 6, Rojo; lane 7, Verano; lane 8, Badillo; lane 9, G19833; lane 10, G40199

When working with a multigene family of closely related individual gene models it is important to understand that the primers might amplify sequences across related families. The ARC and PHY primer pairs were designed to isolate gene specific differences, yet at the

nucleotide sequence level the amplification products among different genotypes were highly similar. A large collection of cloned fragments was sequenced among resistant and susceptible genotypes as well as the *P. acutifolius* donor of the bruchid resistance. After aligning these sequences, it became clear that these clones formed subgroups (see Fig 6). These subgroups consisted of sequences from G40199, the crosses, and parents, and within each subfamily there was a high level of identity. A full inventory of how many clones for each primer and genotype that was sequences can be found in appendix Table A1. From this point forward in the study any clones that no insertion/deletion differences and <5 nucleotide substitution differences to another clone were considered as one sequence.

Phylogenetics

Due to the high similarity of cloned sequences associated with genes in the Leg_b multigene family, multiple neighbor-joining trees were developed to better understand the relationship between these clones and the other ~60 Leg_b proteins in the reference common bean genome. The first tree was constructed with amino acid sequences to investigate the relationship of all Leg_b gene models annotated from the reference *P. vulgaris* genome (Fig 5). The three gene models for the α-AI, ARC and PHY genes and another Leg_b domain are clustered in a single node of the tree (bootstrap equals 93; see red nodes; Phvul.004G158000, Phvul.004G158100, Phvul.004G158200, Phvul.004G158300). The cluster is uniquely separated from the other Leg_b domain proteins suggesting a unique evolutionary relationship.

An additional neighbor-joining tree was constructed using all the clone sequences from the PHY, ARC, and AAI primers from the resistant and susceptible genotypes (Fig 6). In this tree, due to the high bootstrap values, the relationship of primer specific products can be observed. The three red node clusters make up a PHY/AAI #1 subgroup. The two blue node

clusters differentiate the PHY/AAI #2 subgroup. The green nodes (bootstrap equals 99) shows the ARC #1 subgroup, and (bootstrap equals 84) define the ARC #2 subgroup. Specificity of the ARC primer pair is strong because the tree shows all but one ARC product is found in the ARC clusters. An introgression is expected to bring a block of genes in that are linked (for example the APA locus) into introgression lines. We see evidence of that since the four PHY primer products from G40199 also clustered with the ARC #1 and #2 groups.

The high number of individual clones that grouped together at nodes with large bootstrap values allowed for further analysis into their similarity at the nucleotide level. When analyzing sequence of all clones regardless of primer origin, nine different groups were recognized (Table 4). Every sequence was analyzed as part of an all-by-all BLAST to compare similarities. Only sequences with e-values of 0 were considered to be part of the same group. Using these groups, the many individual sequences that make up Fig 6 were consolidated even further. Clones that had no insertion/deletion differences, had less than 5 substitution differences, and were from the same amplification product (clones 1-5) were considered to represent a single individual (Table 6). Representative clones from each group were chosen for further tree development. Figure 7 shows the similarities between the representative clones from each group compared to one another using a color code to denote high and low consensus.

Using the two ARC groups denoted in Table 4, a neighbor-joining tree was developed with published ARC sequence from *P.vulgaris, P. acutifolius*, and the two ARC group representatives identified in this research, ARC_G_B3 and ARC_G_T1. This allowed the determination of the relationship of these cloned ARC sequences with the other published ARC sequences (Figure 8). This tree also supports that an arcelin introgression took place in the

introgression lines since ARC_G_T1, a representative of ARC subgroup #2, is nearly identical to an arcelin amino acid sequence previously characterized in *P. acutifolius*.

Thirteen clones were not part of any group due to insertion/deletions and were analyzed as individuals along with the parsed clone representatives from Table 4 in a neighbor-joining tree (Figure 9). Black circles were used to denote the representative sequences. To get a larger picture of how the clones are related to the Leg_b domains, a neighbor-joining tree was constructed that merges the information found in Figures 4 and 7 (Figure 10). The nodes with bootstrap values of 99 and 98 (see triangles) show the clusters of clones and the relationship to the Leg_b domain models Phvul.004G158000, Phvul.004G158100, Phvul.004G158200 and Phvul.004G158300.1. This tree determines that these cloned sequences were actually present in the APA locus on chromosome 4 and were not the product of amplification of other Leg_b subdomain genes elsewhere in the genome. Figure 10 compares the previously characterized sections of the APA locus from BAC clone 71F18 (APA-1, APA-2, APA-3, APA-4, APA-5 and APA-6) (Kami, 2006), along with the ARC, PHY and α-AI whole gene primer clones from this project.

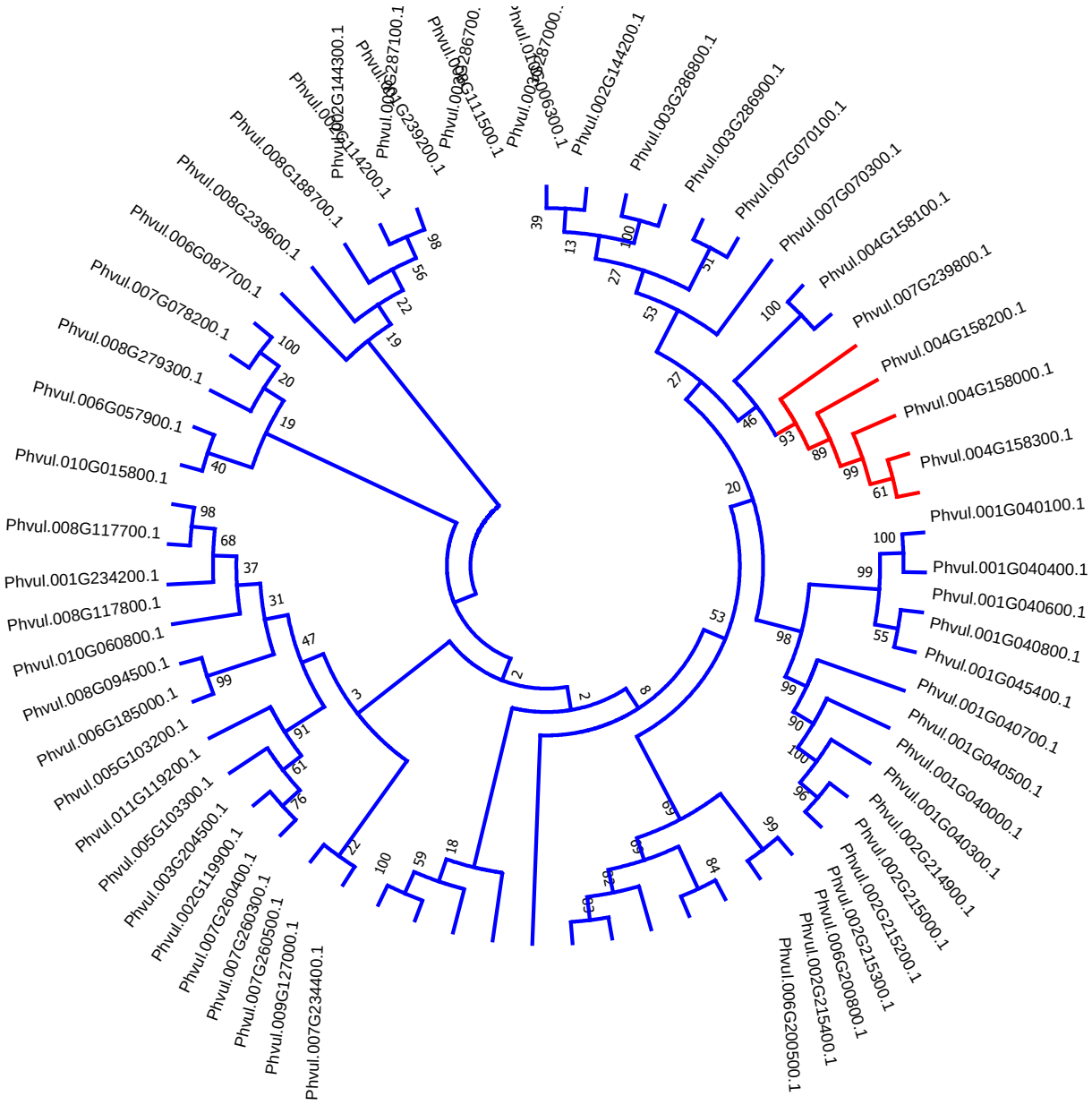

Figure 5. Neighbor joining tree comparing Leg_b amino acid sequences across the *Phaseolus vulgaris* genome. The red nodes indicate where the four Pv04 gene models separate from the other Leg_b domains across the genome which are blue

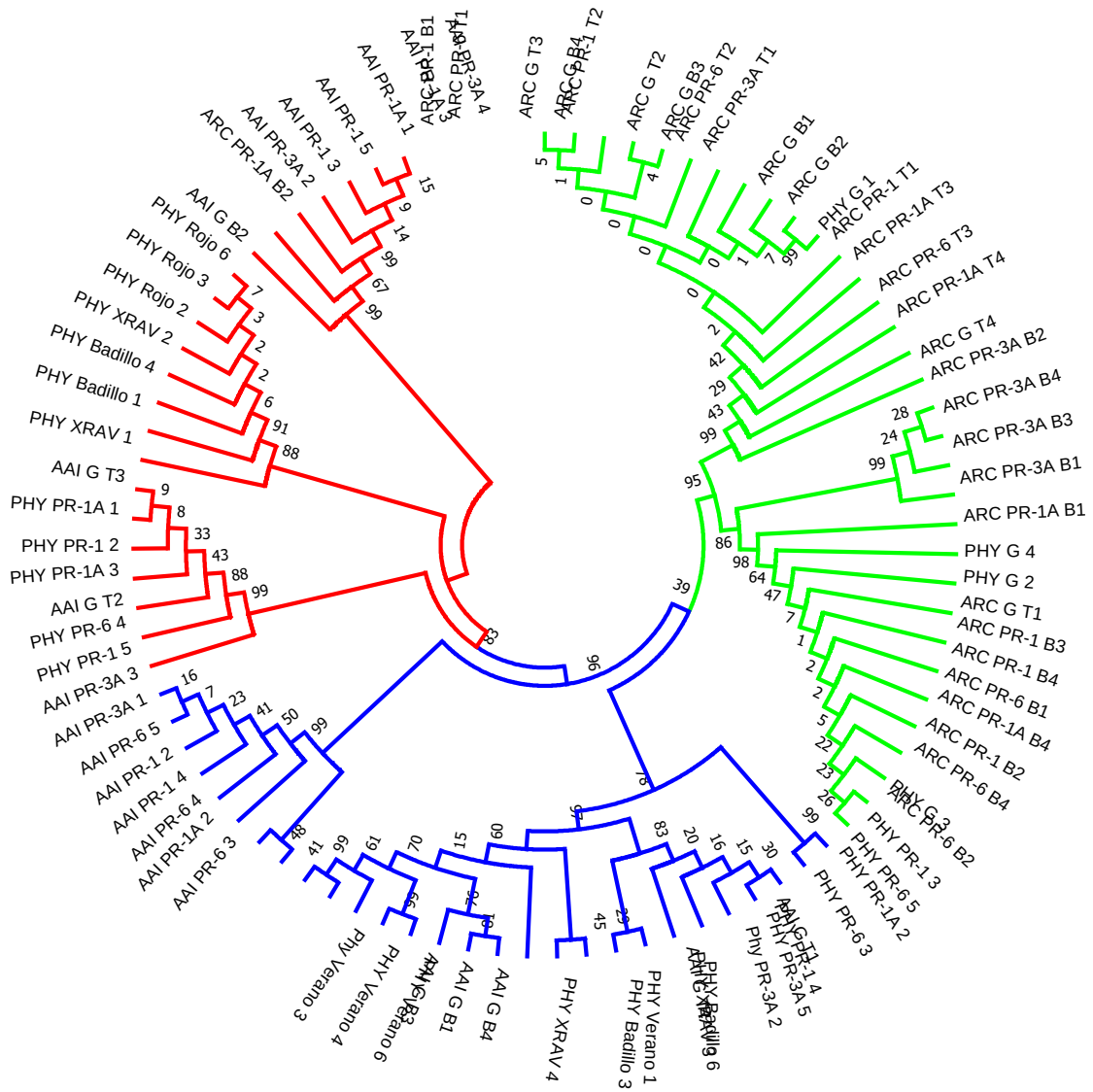

Figure 6. Neighbor-joining tree containing all cloned amino acid sequences from the resistant genotypes (PR1464-1 and PR1464-6), resistant parents (PR1012-29-3-1A and PR1012-29-3-3A), susceptible parents (XRAV-40-4 and Rojo), and susceptible checks (Verano and Badillo) and the source G40199 amplified with ARC, PHY and AAI primers. B or T indicates bottom or top from a double banded sample. 1-5 indicates the clone number. PHY/AAI #1 subgroup nodes are red, PHY/AAI #2 subgroup nodes are blue, and ARC subgroup nodes are green.

Table 4. Clones parsed into groups based on differences in the nucleotide sequence. The representative individual clone for each group is listed.

Group Members	Differences in Sequence	Group Representative
ARC_G_T2 ARC_G_B3 PHY_G_1 ARC_PR-1A_T2 ARC_PR-3A_T1 ARC_PR-6_T1 ARC_PR-1_T1	2 nt substitution between members	ARC_G_B3
ARC_G_T1 PHY_G_2 ARC_PR-1A_B1 ARC_PR-6_B1 ARC_PR-1_B2	5 nt substitution between members	ARC_G_T1
AAI_PR-3A_1 AAI_PR-1_2 AAI_PR-6_5	2 nt substitution between members	AAI_PR-1_2
AAI_G_B2 PHY_Rojo_2 PHY_Badillo_4 PHY_XRAV_2 PHY_Rojo_6	20 nt substitution between members	AAI_G_B2
AAI_PR-3A_4 AAI_PR-1A_1 AAI_PR-1_3	1 nt substitution between members	AAI_PR-1_3
AAI_G_B4 PHY_XRAV_4 PHY_Badillo_3	18 nt substitution between members	AAI_G_B4
AAI_G_T2 PHY_PR-1A_3 PHY_PR-6_4	6 nt substitution between members	AAI_G_T2
AAI_G_T1 PHY_PR-1A_2 PHY_PR-6_3 PHY_PR-1_4	18 nt substitution between members	AAI_G_T1
PHY_PR-1_3 PHY_PR-6_5	5 nt substitution between members	PHY_PR-6_5

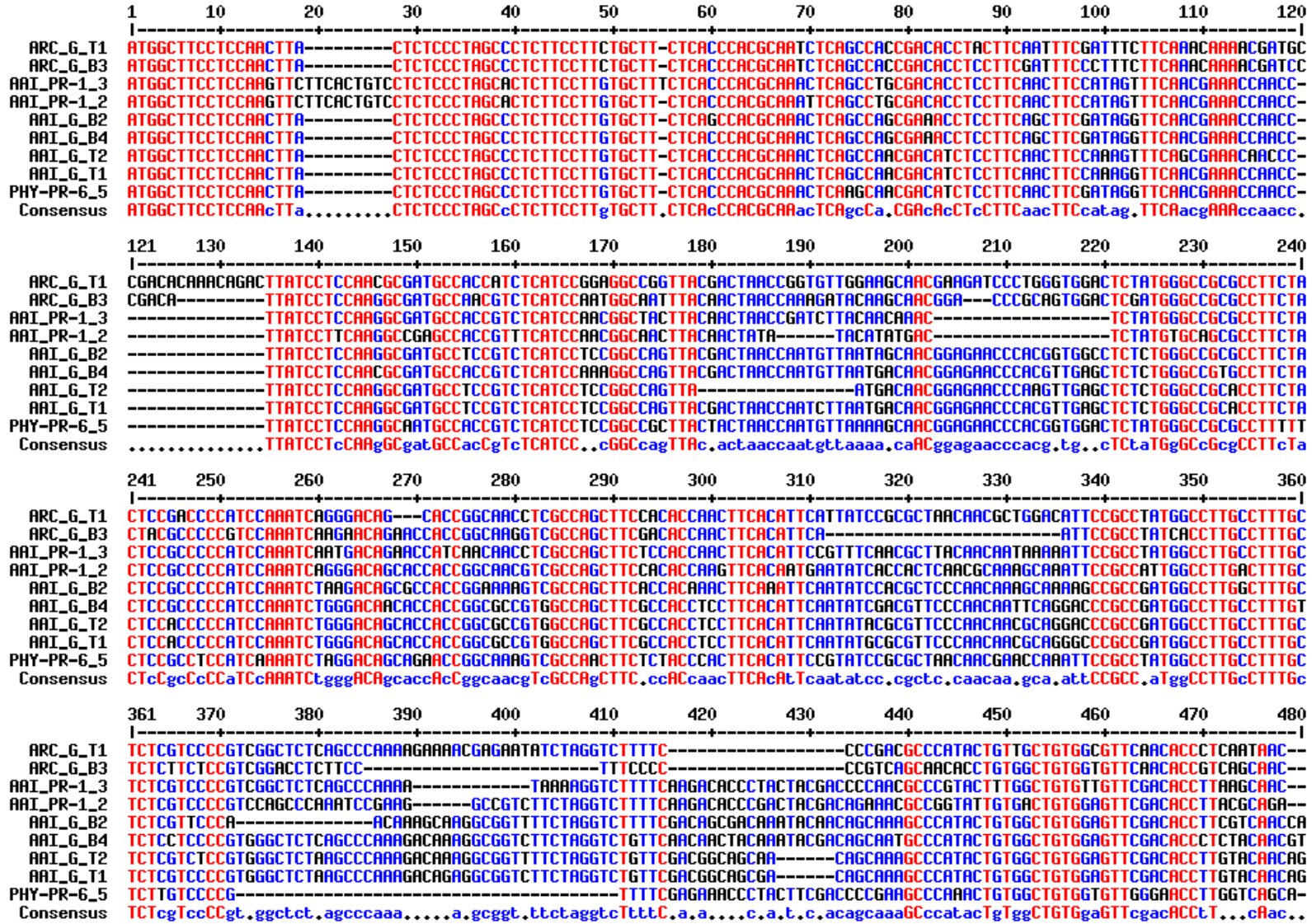

Figure 7. Alignment of nucleotide sequence from each of the representative clones. High consensus color is red, low consensus color is blue and neutral consensus is black.

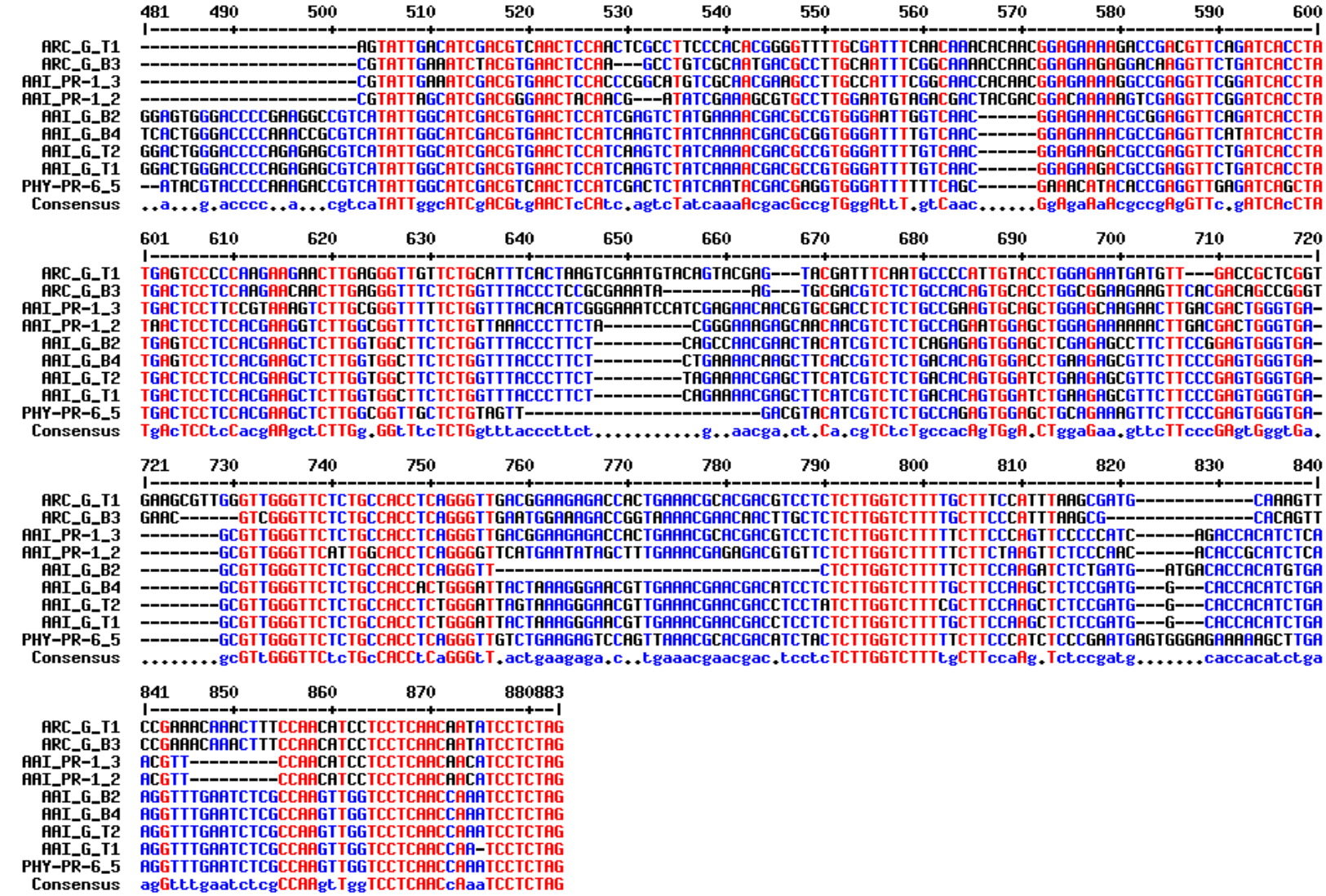

Figure 7. Alignment of nucleotide sequence from each of the representative clones (continued). High consensus color is red, low consensus color is blue and neutral consensus is black.

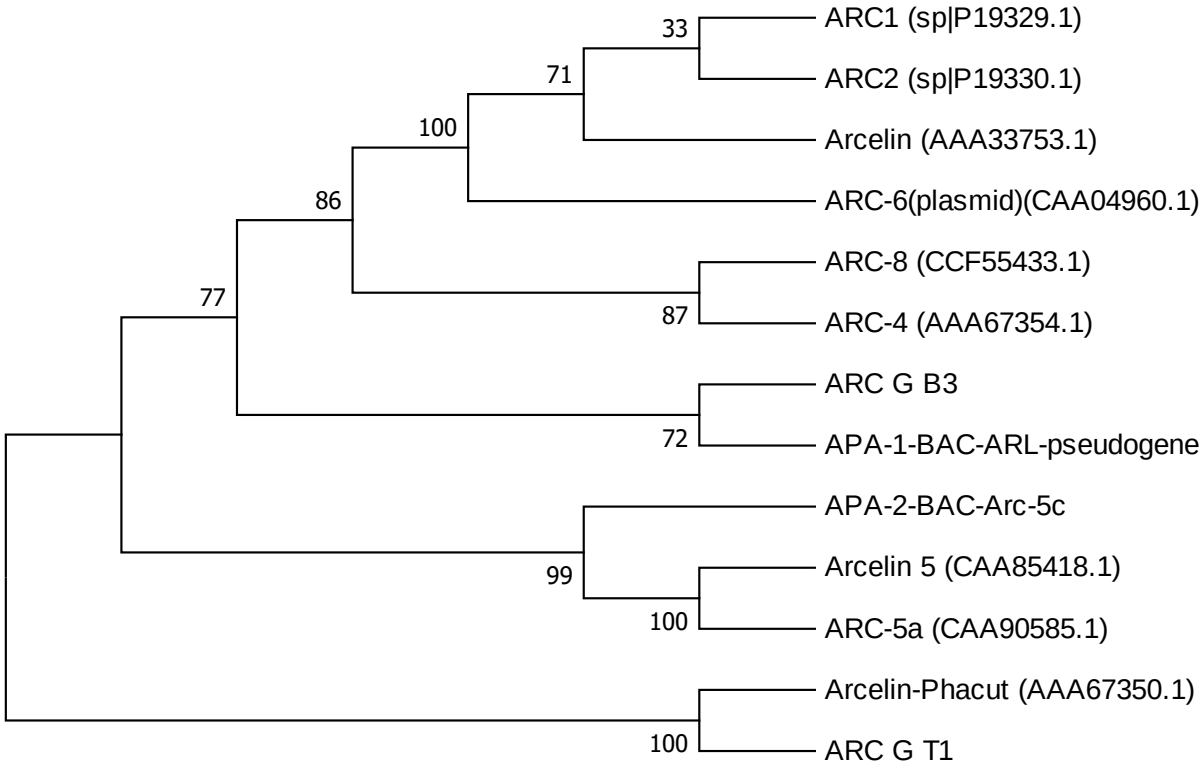

Figure 8. Neighbor-joining tree with the two ARC representative clones, APA 1, APA 2, and previously characterized ARC amino acid sequence from *P. vulgaris* and *P. acutifolius.*

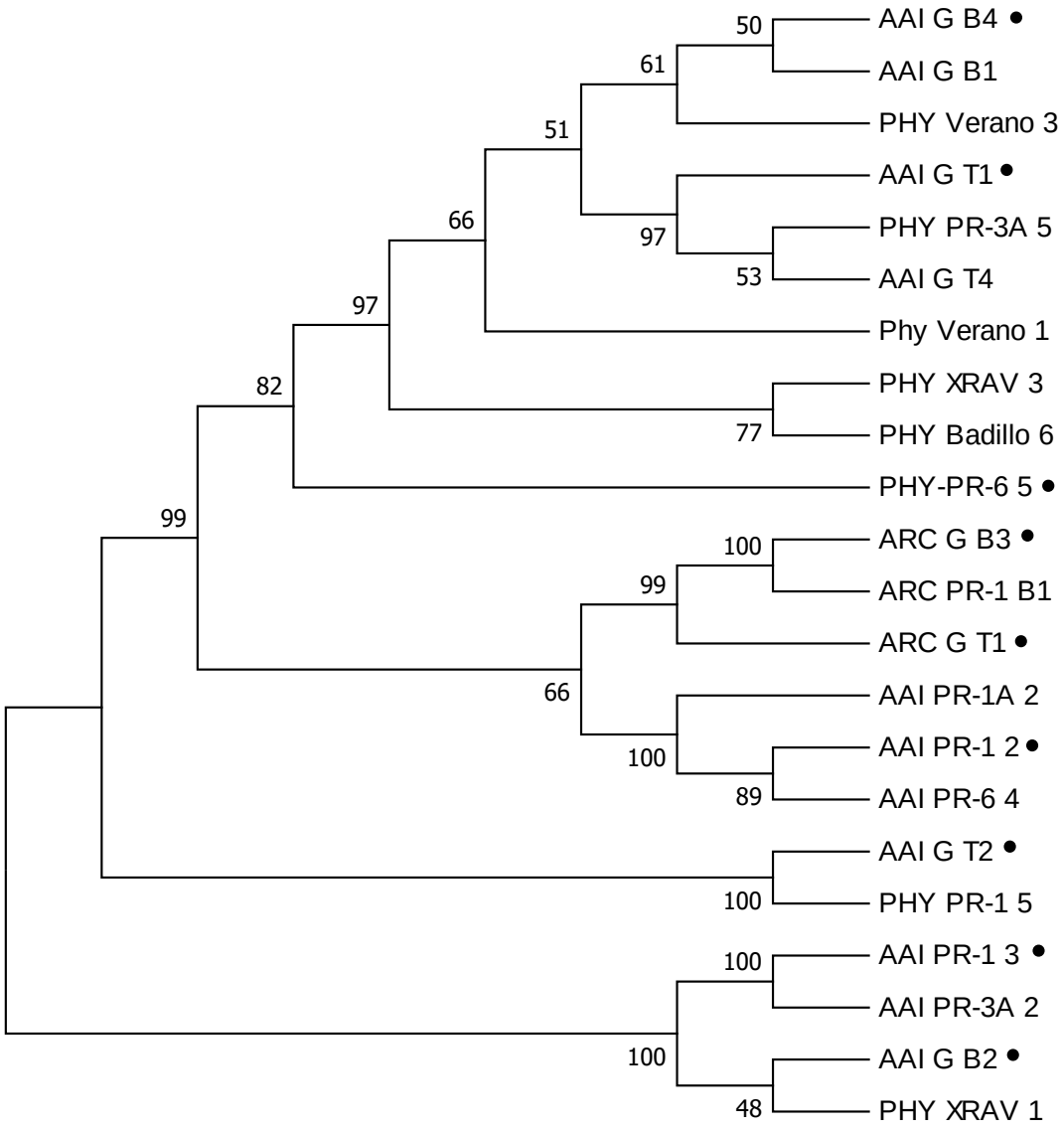

Figure 9. Neighbor joining tree comparing the amino acid sequence from the ARC, PHY and AAI group representatives with the individual clones that do not fit into groups

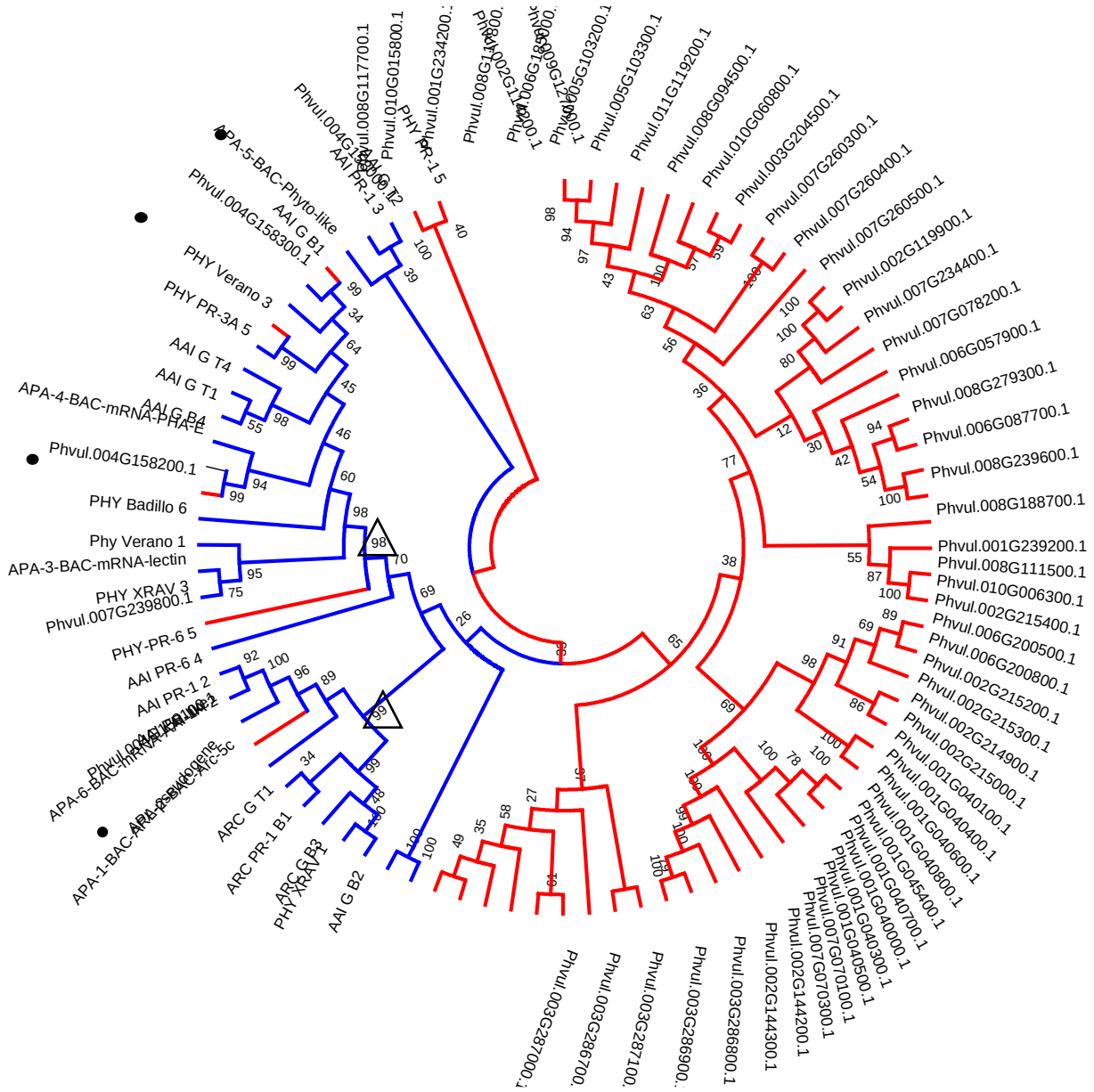

Figure 10. Neighbor-joining tree comparing the amino acid sequence of the representative ARC, PHY and AAI group members with the Leg_b domain models and the previously characterized APA 1-6 sequences.

DISCUSSION

INS45 Marker Durability

As previously mentioned, the development of a molecular marker to track the introgression of this locus could dramatically decrease the time and labor involved in screening for bruchid resistance. Having an easily identifiable marker could allow breeders to screen hundreds of individuals in a matter of days, as opposed to 8 weeks which is typical in the traditional method (Singh, 1985). The α-AI-1 INS45 indel marker discovered here is important, not only because it distinguishes the resistant and susceptible individuals, but also because it gives more insight into what proteins and domains are involved. As previously stated, the Pfam database showed that this indel occurs in the middle of the Lectin_legB (PF00139) domain. With more than 70 lectins in the common bean reference genome, this is the largest lectin family and is a member of the superfamily, concanavalin, which includes a multitude of carbohydrate binding domains (Sharon, 1990). This is important because it brings together the specific genetic information we know about the marker and the information about the lectins that may be interfering with the carbohydrate metabolism in bruchid beetles (Chrispeels and Raikhel, 1991).

Using the SA recombinant inbred line population, we were able to test the marker and analyze its usefulness as a tool to track the introgression of the APA locus from the tepary bean G40199 source. The marker itself has an easy to identify size difference that allows for repeatable and consistent scoring across many different individuals. As previously mentioned, when the population was screened against this marker it correctly identified resistant genotypes in 5 individuals, and correctly identified susceptible genotypes in 159 individuals. However, due to multiple effects in this APA locus there were instances where the marker did not correspond to the phenotypic score. This could be partly be the result of an intense scoring system for

resistance that is too rigid for this marker screening. Some members of the population showed no signs of infestation (1,0/10) for up to 60 days in the first replication and then were deemed susceptible when one or two seeds were affected 60 days into the second replication. This screening system is important information for breeding purposes but does not lend itself well to the complexities of this complex locus, especially since the specific mode of resistance is not well understood yet.

While this screen of 206 individuals in the SA population has not shown the INS45 marker to be a perfect predictor of the introgression of the *P. acutifolius* source of bruchid resistance, it has shown it could be a valuable tool in the screening process. Selecting on the INS45 band across a whole population would result in low type II error and would pick out many of the individuals that are susceptible. Additionally, the remainder of the population, and individuals with both bands could be screened for damage using the original scoring system on a smaller scale.

Sequence Analysis of Leg_b Domain Containing Sequence

The presence of the PCR product for only XRAV-40-4 and Verano was an unexpected but ultimately important part of this study. Discovering that Verano, the susceptible check, with no part in the breeding scheme had the same amplification as XRAV-40-4 is what prompted the cloning effort. After looking into the background of Verano it was discovered that VAX6, another tepary genotype was present in its pedigree (Beaver, 2008). It is possible that the presence of the unexpected PCR product could be an artifact of having another tepary genotype in the background

After aligning and comparing the Leg_b domain containing clone sequences it became apparent that there is a highly similar and trackable presence of ARC sequences from source, to

parent, to progeny in the ARC_G_B3 and ARC_G_T1 groups as evidenced by the fact that the APA locus from the *P. acutifolius* source G40199 was indeed introgressed into these resistant genotypes. Interspecific crosses in which genetic blocks are inherited from one species to another can be difficult and tracking whether the important loci is introgressed can be even more difficult to track. For the work here, having clear proof that the important Leg_b subdomain sequences from the APA locus were indeed introduced is crucial to developing resistant lines, even when the mechanism for resistance is unclear.

Phylogeny

Looking at the Leg_b gene model tree it is evident that there are many gene models with this domain on almost every chromosome. In Figure 5 the gene models that are important to this Pv04 Leg_b domain (Phvul.004G158000.1, Phvul.004G158100.1, Phvul.004G158200.1, and Phvul.004G158300.1) separate at a node with a 93-bootstrap value from gene models on chromosome Pv07. Alternatively, in Figure 6 there is strong differentiation between subgroups that are made up of clones ARC and PHY/AAI clones. Figure 8 uses previously characterized ARC sequences from *P. vulgaris* and *P. acutifolius* to support that the introgression itself was successful. Figure 9 allows for a consolidation of data and the ability to look across primer groups and the chance to put less weight in the name of the primer and analyze its relationship with other sequences. Using the ARC_G_T1 group from Table 4 and the relationship it has with the *P. acutifolius* arcelin sequence in Figure 8 it shows a clear introgression from the G40199 tepary source that can be tracked into the newly developed lines. Alternatively, groups like AAI_PR-1_2 and AAI_PR-1_3 that had no G40199 sequences represented could come from the *P. vulgaris* unique sequences relative to *P. acutifolius*. In other words, those clones sequence are from the susceptible parents and not from the tepary source.

In Figure 10 when both of these trees are analyzed together the big picture was revealed. Merging the clone representatives with the Leg_b domain sequences verifies that the clone sequences are highly related to APA Leg_b protein on Pv04, not to other regions of the genome. In this figure clusters become more apparent, for example, with a bootstrap value of 96 AAI and ARC sequence show relation to characterized APA sequences from BAC clones described by Kami et al. (2006) and Leg_b domain Phvul.004G158100.1. On the adjacent node a cluster of PHY clones show a close relationship to the previously characterized PHA-E and lectin domain sequences Phvul.004G158200.1 and Phvul.004G158300.1. It proved difficult to develop gene specific primers that could clearly distinguish sequences because of the extreme similarity and variants of these genes. It remains unclear why the AAI_G_T1, AAI_G_T4, AAI_G_B1 and AAI_G_B4 sequences show such similarity to the PHY clone clusters in Figure 10, and why we did not see as a clear of relationship between the clones from the resistant cross genotypes and the susceptible parents.

Up until this point using the three Phvul gene models from the reference genome has been the best model available for how the locus was arranged. Now, taking into consideration this phylogenetic evidence, it could be more beneficial to look at the model in Figure 1 (Kami et al, 2006). During the development and annotation of reference genomes sequences that are highly similar or repetitive are often left out. In this case the BAC clone, which denotes the APA genes in purple, gives a much more accurate representation of how many genes could be involved in this locus. A major challenge of this project was getting rooted in nomenclature. Much of the previous research was so focused on isolating genes, getting them into specific lines, and trying to discover resistant variants, that the nomenclature surrounding this locus became too specific for what was actually known. Perhaps with this new phylogenetic

information the use of new gene models the language around this research can start to reflect the complexity and big picture approaches needed to understand this locus. With more research a better understanding of the locus will provide more insight into the resistance mechanism and advance the development of more specific markers.

CONCLUSION

The ability to track the introgression of this APA locus into common bean is the key to being able to understand this source of bruchid resistance. Tracking the locus is essential to knowing what existing lines to use and what crosses to make to develop new lines. Use of a marker to allow breeders to select lines with resistance would be helpful for small shareholder farmers who can't afford to lose a substantial portion of their crop because their infrastructure to store seed is not ideal (CIAT, 1986). Post-harvest losses require global attention in the fight for food security. Sub-Saharan Africa is estimated to lose $4 billion USD annually due to post harvest losses, and even a 1% advancement in loss prevention would bring back $40 million dollars to the people in the most food insecure areas in the world (World Bank, 2010).

Resistance to these seed storage pests with the presence of the APA locus from wild tepary bean genotype G40199 resistance can be strong. The use of interspecific crosses between *P. acutifolius* and *P. vulgaris* has allowed the introgression of this locus into dry bean cultivar. The availability of a marker that tracks the introgression with accuracy allows breeders to screen large populations for this locus without having to grow every individual to seed and test bruchid resistance. While the data from the INS45 marker hasn't shown it to be a definitive screen for resistance, it can be used as an effective tool in reducing the number of lines that need to be evaluated with the laborious screening procedure. This enables breeders to develop a larger number of early generation lines that contain the G40199 source of bruchid resistance, screen these larger populations with the INS45 marker, and only phenotype a reduced number of lines.

In an effort to better characterize the locus, the sequence and phylogeny work in this study details the relationship among members of the complex Leg_B family. This survey has allowed for increased understanding about the relationships between the sequences that make up

subgroups and the gene models by comparing them on the sequence level. At the same time, it also opens up a discussion about how well this data fits the current gene models based on the reference genome. There is much more research to be done but the work done in this study has shown that given a recombinant inbred line made from one of these interspecific crosses, this INS45 marker could help track the introgression in the population. Furthermore, additional research into allele optimization for the numerous variants of ARC, PHY and α-AI primers will quicken the development of lines with the targeted variant introgression.

REFERENCES

Abbo, S., S. Lev-Yadun, and A. Gopher. 2012. Plant Domestication and Crop Evolution in the Near East: On Events and Processes. Critical Reviews in Plant Sciences, 31(3): 241–257.

Abbo, S., Y.T. Mesghenna, and H. Van Oss. 2011. Interspecific hybridization in wild Cicer sp. Plant Breeding = Zeitschrift Fur Pflanzenzuchtung. 130(2): 150–155.

Acosta-Gallegos, J.A., C. Quintero, J. Vargas, O. Toro, J. Tohme, C. Cardona. 1998. A new variant of arcelin in wild common bean, *Phaseolus vulgaris* L., from southern Mexico. Genet. Res. Crop. Evol. 45:235-242.

Bardocz, S., G. Grant, S. W. Ewen, T. J. Duguid, D. S. Brown, K. Englyst, and A. Pusztai. 1995. "Reversible Effect of Phytohaemagglutinin on the Growth and Metabolism of Rat Gastrointestinal Tract." Gut 37 (3). gut.bmj.com: 353–60.

Beaver, J. S., T. G. Porch and M. Zapata. 2008. Registration of 'Verano' White Bean. J. Plant. Reg. 2:187-189. doi:10.3198/jpr2008.02.0110crc

Blair, M.W., L.M. Díaz, H.F. Buendía, and M.C. Duque. 2009. Genetic diversity, seed size associations and population structure of a core collection of common beans (Phaseolus vulgaris L.). TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik, 119(6): 955–972.

Bonfield, J.K., Smith, K.F. and R. Staden. 1995. A new DNA sequence assembly program. Nucleic Acids Res. 23: 4992-4999.

Broughton, W.J., G. Hernandez, M. Blair, S. Beebe, P. Gepts, and J. Vanderleyden. 2003. Beans (Phaseolus Spp.) --Model Food Legumes. Plant and Soil 252 (1). Springer: 55–128.

Cardona, C. 1989. Insects and Other Invertebrate Bean Pests in Latin America. Bean Production Problems in the Tropics. Cali, Colombia: CIAT[Links]. sidalc.net. http://www.sidalc.net/cgibin/wxis.exe/?IsisScript=CATALCO.xis&method=post&formato=2&cantidad=1&expresion=mfn=060064.

Cardona, C., J. Kornegay, C.E. Posso, F. Morales, H. Ramirez. 1990. Comparative value of four arcelin variants in the development of dry bean lines resistant to the Mexican bean weevil Entomol. Exp. Appl. 56: 197-206.

Chrispeels, M.J., N.V. Raikhel. 1991. Lectins, lectin genes, and their role in plant defense. Plant Cell 3:1-9.

CIAT, Nov. 1986. Principal insects attacking dry stored beans, 2nd edition

Crowson, R. A. 1981. The Biology of the Coleoptera, Academic Press, London, New York, Toronto, Sydney, San Fransisco.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 32(5):1792-1797.

FAO Statistics online http://faostat.fao.org

Goossens A., R. Geremia, G. Bauw, M. Van Montagu, G. Angenon. 1994. Isolation andcharacterization of arcelin-5 proteins and cDNAs. Eur. J. Biochem. 225: 787-795.

Goossens, A., C. Quintero, W. Dillen, R. De Rycke, J.F Valor, J. De Clercq, and G. Angenon. 2000. Analysis of bruchid resistance in the wild common bean accession G02771: no evidence for insecticidal activity of arcelin 5. Journal of Experimental Botany. 51(348): 1229-1236.

Haghighi, Y., and P.D Ascher. 1988. Fertile intermediate hybrids between *P. vulgaris* and *Pacutifolius* from congruity backcrossing. Sex. Plant Reprod. 1: 51-58.

Hammer, K. 1984. Das Domestikationssyndrom. Die Kulturpflanze: Berichte Und Mitteilungen Aus Dem Institut Fur Kulturpflanzenforschung Der Deutschen Akademie Der Wissenschaften Zu Berlin in Gatersleben Krs. Aschersleben, 32(1): 11–34.

Hartweck, L.M., and T.C. Osborn. 1997. Altering protein composition by genetically removing phaseolin from common bean seeds containing arcelin or phytohaemagglutinin. Theor. Appl. Genet. 95: 1012-1017.

Hartweck, L.M., R.D. Voegelzang, and T.C. Osborn. 1991. Characterization and comparison of arcelin seed protein variants from common bean. Plant Physiol. 97: 204-211.

Henao S., M.P. Arbelaez. 2002. Epidemiological situation of acute pesticide poisoning in the Central American Isthmus, 1992-2000. Pan American Health Organization (PAHO) PLAGSALUD. Epidemiol Bull 2002; 23: 5-9.

Huignard, J., and J.C. Biemont. 1981. Reproductive Polymorphism of Populations of Acanthoscelides Obtectus from Different Colombian Ecosystems. In the Ecology of Bruchids Attacking Legumes (Pulses). Series Entomologica. Springer Netherlands. 149-164.

Ishimoto, M., T. Sato, M.J. Chrispeels, and K. Kitamura. 1996. Bruchid resistance of transgenic adzuki bean expressing seed α-amylase inhibitor of common bean. Entomol. Exp. Appl. 79: 309-315.

John, M., C.M. Long. 1990. Sequence analysis of arcelin 2, a lectin-like plant protein. Elsevier. 86:171-176.

Kami, J., Poncet, V., Geffroy, V., & Gepts, P. (2006). Development of four phylogenetically-arrayed BAC libraries and sequence of the APA locus in Phaseolus vulgaris. *Theoretical and applied genetics*, *112*(6), 987-998.

Kelly, J.D. 2010. The story of bean breeding, white paper prepared for bean CAP & PBG works on the topic of dry bean production and breeding research in the USA. Michigan State University. 1–29.

Kornegay, J., C. Cardona, A. van Schoonhoven, O. Voysest, and Others. 1991. Breeding for Insect Resistance in Beans. Common Beans: Research for Crop Improvement. CAB International. 619–48.

Kornegay, J., C. Cardona, and C.E. Posso. 1993. Inheritance of resistance to Mexican bean weevil in common bean, determined by bioassay and biochemical tests. Crop Sci. 33: 589-594.

Kornegay, J.L., and C. Cardona. 1991. Inheritance of resistance to Acanthoscelides obtectus in a wild common bean accession crossed to commercial bean cultivars. Euphytica 52: 101-111.

Kumar, S., G. Stecher and K. Tamura. 2015. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. Molecular Biology and Evolution (submitted).

Kusolwa, P.M. and Myers, J.R. 2007. Identification of expressed lectin-like protein profiles from tepary bean G40199 and interspecific hybrids and phylogenetic relationship to other bean lectins. Breeding for Bruchid Resistance in Common Bean (Phaseolus vulgaris L.): Interspecific Introgression of Lectin-like Seed Proteins from Tepary Bean (P. acutifolius A. Gray), Genetic Control and Bruchid Resistance Characterization., p.74.

Kwak, M., and P. Gepts. 2009. Structure of genetic diversity in the two major gene pools of common bean (Phaseolus vulgaris L., Fabaceae). TAG. Theoretical and Applied Genetics. 118(5): 979–992.

Lavin, M., P.S. Herendeen, and M.F. Wojciechowski. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. Systematic Biology, 54(4): 575–594.

Leroi, B. 1981. Feeding, Longevity and Reproduction of Adults of Acanthoscelides Obtectus Say in Laboratory Conditions. In the Ecology of Bruchids Attacking Legumes (Pulses). Series Entomologica. Springer Netherlands. 101-111.

Lewis, G.P., B. Schrire, B. Mackinder, and M. Lock. 2005. Legumes of the World. Royal Botanic Gardens Kew.

Lioi, L., and R. Bollini. 1989. Identification of a new arcelin variant in wild bean seeds. Annu. Rep. Bean Improv. Coop. 32: 28.

Lioi, L., F. Sparvoli, I. Galasso, C. Lanave, and R. Bollini. 2003. Lectin-Related Resistance Factors against Bruchids Evolved through a Number of Duplication Events. TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik 107 (5). Springer: 814–22.

Mamidi, S., M. Rossi, S.M. Moghaddam, et al. 2013. Demographic factors shaped diversity in the two gene pools of wild common bean Phaseolus vulgaris L. Heredity. 110(3):267-276. doi:10.1038/hdy.2012.82.

McClean, P. E., & Raatz, B. (2017). Common Bean Genomes: Mining New Knowledge of a Major Societal Crop. In *The Common Bean Genome* (pp. 129-145). Springer, Cham.

Mejia-Jiménez, A., C. Muñoz, H.J. Jacobsen, W.M. Roca, and S.P. Singh. 1994. Interspecific hybridization between common and tepary beans: Increased hybrid embryo growth, fertility, and efficiency of hybridization through recurrent and congruity backcrossing. Theor. Appl. Genet. 88: 324-331.

Minney, B., P. Hugo, M.R. Angharad, P. Dobie, J. Dendy, C. Cardona, and J.A. Gatehouse. 1990. Biochemical Bases of Seed Resistance to Zabrotes Subfasciatus (bean Weevil) in Phaseolus Vulgaris (common Bean); A Mechanism for Arcelin Toxicity. Journal of Insect Physiology 36 (10). Elsevier: 757–67.

Misangu, N.R. 1997. Distribution of bean bruchid species in the major bean growing regions of Tanzania and breeding beans for resistance to *Acanthoscelides obtectus* (Say) and *Zabrotes subfasciatus* (Boh). Ph.D. thesis pp 37-170.

Morton, R.L., H.E. Schroeder, K.S. Bateman, M.J. Chrispeels, E. Armstrong, T.J.V Higgins. 2000. Bean α-amylase inhibitor 1 in transgenic peas (*Pisum sativum*) provides complete protection from pea weevil (*Bruchus pisorum*) under field conditions. Proc. Natl. Acad. Sci. USA 97: 3820-3825.

Nahimana, M. 1992. Highlights of bruchid research in the Great Lakes Region. Proceedings of 3rd Southern Africa Development Community/Centro Internacional de Agricultura Tropical (SADC/CIAT) Bean Research Workshop. CIAT African Workshop Series No. 27. pp. 153-163.

Nishizawa, K., M. Teraishi, S. Utsumi, and M. Ishimoto. 2007. Assessment of the importance of α-amylase inhibitor-2 in bruchid resistance of wild common bean. Theor. Appl. Genet. 114:755-764.

Nodari, R. O., S. M. Tsail, R. L. Gilbertson, and P. Gepts. 1993. Towards an Integrated Linkage Map of Common Bean 2. Development of an RFLP-Based Linkage Map. TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik 85 (5). Springer: 513–20.

Osborn, T.C., D. Alexander, S.S.M. Sun, C. Cardona, and F.A. Bliss. 1988b. Insecticidal activity and lectin homology of arcelin seed protein. Science 240: 207-210.

Osborn, T.C., T. Blake, P. Gepts, and F.A Bliss. 1986. Bean arcelin 2. Genetic variation, inheritance and linkage relationships of a novel seed protein of *Phaseolus vulgaris* L. Theor. Appl. Genet. 71: 847-855.

Osborn, T.C., M. Burrow, and F.A. Bliss. 1988a. Purification and characterization of arcelin seed protein from common bean. Plant Physiol. 86: 399-405.

Pachico, D. 1989. Trends in world common bean production. Bean Production Problems in the Tropics. 10:1.

Paes, N.S., I.R. Gerhardt, M.V. Coutinho, M. Yokoyama, E. Santana, N. Harris, M.J. Chrispeels, and M.F. Grossi de Sá. 2000. The effect of arcelin-1 on the structure of the midgut of bruchid larvae and immunolocalization of the arcelin protein. J. Insect Physiol. 46:393-402.

Peumans, W.J., and J.M. Van Damme. 1995. Lectins as plant defense proteins - update on plant defense proteins. Plant Physiol. 109: 347-352.

Pfaffenberger, G.S. 1985. Description, Differentiation, and Biology of the Four Larval Instars of Acanthoscelides Obtectus (Say) (Coleoptera: Bruchidae). The Coleopterists' Bulletin 39 (3). The Coleopterists Society: 239–56.

PROMEGA www.promega.com

Quentin, M.E., J.L. Spencer, and J.R. Miller. 1991. Bean tumbling as a control measure forcommon bean weevil, *Acanthoscelides obtectus*. Entomol. Exper. Appl. 60: 105-109.

Rabakoarihanta, A., C.T. Shii, M.C. Mok, D.W.S. Mok. 1980. Meiosis and fertility recovery of interspecific hybrids between *Phaseolus vulgaris* L. and *P. acutifoluis* A. Gray. Theor. Appl. Genet. 57: 59-64.

Romero-Andreas, J., B.S. Yandell, and F.A. Bliss. 1986. Bean arcelin 1. Inheritance of a novel seed protein of *Phaseolus vulgaris* L. and its effect on seed composition. Theor. Appl. Genet. 72:123-128.

Rossi, M., E. Bitocchi, E. Bellucci, L. Nanni, D. Rau, G. Attene, and R. Papa. 2009. Linkage disequilibrium and population structure in wild and domesticated populations of Phaseolus vulgaris L. Evolutionary Applications, 2(4): 504–522.

Rudiger, H., H.J. Gabius. 2001. Review – Plant lectins: Occurrence, biochemistry, functions and applications. Glycoconjugate J. 18: 589-613.

Sales, M.P., I.R. Gerhardt, M.F. Grossi de Sá, J. Xavier-Filho. 2000. Do legume storage proteins play a role in defending seeds against bruchids? Plant Physiol. 124: 515-522.

Sakthivelkumar, S., M.I. Jesse, V. Veeramani, P. Ramaraj, K. Kathiravan, M. Arumugam, and S. Janarthanan. 2013. Diversity and Analysis of Sequences Encoded by Arcelin Genes from Indian Wild Pulses Resistant to Bruchids. Process Biochemistry 48 (11). Elsevier: 1697–1705.

Santino, A., B. Valsasina, L. Lioi, A. Vitale, R. Bollini. 1991. Bean (*Phaseolus vulgaris* L.) seed lectins: a novel electrophoretic variant of arcelin. Plant Physiol (Life Sci Adv) 10:7-11.

Schmale, I., F.L. Wackers, C. Cardona, S. Dorn. 2002. Field Infestation of *Phaseolus vulgaris* by *Acanthoscelides obtectus* (Coleoptera: Bruchidae), Parasitoid Abundance, and Consequences for Stoage Pest Control. Environ. Entomol. 31: 859-863.

Schmutz, J., P.E. McClean, S. Mamidi, G.A. Wu, S.B. Cannon, J. Grimwood, S.A. Jackson. 2014. A reference genome for common bean and genome-wide analysis of dual domestications. Nature Genetics, 46(7): 707–713.

Singh, B. B., S. R. Singh, and O. Adjadi. 1985. Bruchid Resistance in Cowpea1. Crop Sci. 25:736-739. doi:10.2135/cropsci1985.0011183X002500050004x

Sharon, N., H. Lis. 1990. Legume lectins--a large family of homologous proteins. FASEB J. 4 (14): 3198–208. PMID 2227211

Singh, S.P., C.G. Muñoz. 1999. Resistance to common bacterial blight among *Phaseolus* species and common bean improvement Crop Sci. 39:80-89.

Singh, S.P. 1992. Common Bean Improvement in the Tropics. Plant Breeding Reviews 10. books.google.com: 199–269.

Slumpa, S., and J. Ampofo. 1991. Evaluation of Different Methods for the Control of Bean Bruchids (Acanthoscelides Obtectus). Annual Report of the Bean Improvement Cooperative. Bean Improvement Cooperative. agris.fao.org. http://agris.fao.org/agris-search/search.do?recordID=US9151351.

Smýkal, P., C.J. Coyne, M.J. Ambrose, N. Maxted, H. Schaefer, M.W. Blair, … R.K. Varshney. 2015. Legume Crops Phylogeny and Genetic Diversity for Science and Breeding. Critical Reviews in Plant Sciences, 34(1-3), 43–104.Allen, D. J., ed. 5-7 October 1992. Mbabane, Swaziland

Uebersax, M.A., K. Jai-Neung, and C. Yong-Soo. 1996. Packaging and handling systems for dry edible beans. Michigan Dry Bean Digest, 20(2): 5-13

White, J.W., and D.R. Laing. 1989. Photoperiod Response of Flowering in Diverse Genotypes of Common Bean (Phaseolus Vulgaris). Field Crops Research 22 (2). Elsevier: 113–28.

Wojciechowski, M.F. 2003. Reconstructing the phylogeny of legumes (Leguminosae): An early 21st century perspective. Pages 5-35 in Advances in legume systematics (B. Klitgaard and A. Bruneau, eds.). Royal Botanic Garden, Kew, United Kingdom

Woolley, J., et al. 1991. The agronomy of intercropping with beans. Common beans: research for crop improvement. 707-735.

World Bank, 2010. Missing food, The case of postharvest grain losses in Sub Saharan Africa. Washington DC. The World Bank.

Yamada, T., K. Hattroi, M. Ishimoto. 2001. Purification and characterization of two a-amylase inhibitors from seeds of tepary bean (*Phaseolus acutifolius* A. Gray). Phytochemistry 58: 59-66.

Zaugg, I., C. Magni, D. Panzeri, M. Gloria Daminati, R. Bollini, B. Benrey, S. Bacher, and F. Sparvoli. 2013. QUES, a New Phaseolus Vulgaris Genotype Resistant to Common Bean Weevils, Contains the Arcelin-8 Allele Coding for New Lectin-Related Variants. TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik 126 (3). Springer: 647–61.

. Zizumbo-Villarreal, D. & Colunga-GarcíaMarín, P. 2010. Origin of agriculture and plant domestication in West Mesoamerica. Genet. Resour. Crop Evol. 57: 813–825.

APPENDIX. CLONE INVENTORIES FOR ARC, PHY AND α-AI CLONES

SEQUENCED FROM THE NUMBER OF PRODUCTS

Number of ARC Primer Clones

	One Band	Two Bands	
Genotype	One	Top	Bottom
G40199	x	3	4
PR1464-1	x	2	4
PR1464-6	x	3	3
PR1012-29-3-1A	x	2	3
PR1012-29-3-3A	x	1	4

Number of PHY Primer Clones

	One Band	Two Bands	
Genotype	One	Top	Bottom
G41099	4	x	x
Rojo	3	x	x
Badillo	4	x	x
XRAV-40-4	4	x	x
Verano	4	x	x
PR1464-1	4	x	x
PR1464-6	3	x	x
PR1012-29-3-1A	3	x	x
PR1012-29-3-3A	2	x	x

Number of α-AI Gene Primer Clones

	One Band	Two Bands	
Genotype	One	Top	Bottom
G41099	x	4	4
PR1464-1	4	x	x
PR1464-6	3	x	x
PR1012-29-3-1A	3	x	x
PR1012-29-3-3A	4	x	x