Title

Developing a molecular linkage map for and understanding the biochemical mechanisms and underlying genetic architecture of biotic stress resistance in lima bean (Phaseolus lunatus L.)

Author

Zullo, Stephanie Smolenski

Publication Date

2021

Peer reviewed|Thesis/dissertation

Developing a molecular linkage map for and understanding the biochemical mechanisms and underlying genetic architecture of biotic stress resistance in lima bean (*Phaseolus lunatus* L.)

by

STEPHANIE SMOLENSKI ZULLO
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Horticulture & Agronomy

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Paul Gepts, Chair

_____

Daniel Kliebenstein

_____

Allen Van Deynze

Committee in Charge

2021

Abstract

Lima bean (*Phaseolus lunatus* L.) is a leguminous crop grown in the Americas, Europe, Africa, and Asia by subsistence farmers and commercial farmers alike. Lima beans are the most significant dry bean crop grown in California, and due to heavy biotic stress, there is a demand for the breeding of more productive and resilient cultivars. Lima bean has lagged in the development of genetic resources to advance the understanding of the population structure, domestication, and divergence from other legume species, hindering the development of improved lima bean cultivars. This dissertation establishes a high-quality molecular linkage map for lima bean and evaluates the role of two putative biochemical mechanisms of biotic stress resistance to gain information on the underlying genetics of and enhance breeding efforts for biotic stress resistance in lima bean.

The first chapter of this dissertation focuses on the development of a genetic map based on a biparental recombinant inbred line population of lima beans, aligned to both the lima bean and common bean reference genomes. Alignment of the genetic map to the new lima bean reference genome enabled the characterization of chromosome morphology, calculation of recombination rates within and across chromosomes, and definitions of the euchromatic and pericentromeric regions across chromosomes for lima bean. Alignment of the genetic map to the common bean reference genome highlighted the variations in chromosome morphology between lima bean and common bean and identified previously unknown and confirmed suspected genomic rearrangements, while confirming the strong synteny between the two closely related species.

The second chapter of this dissertation focuses on the identification of significant QTL for key agronomic and domestication traits in the UC 92 – UC Haskell biparental recombinant inbred line population of lima beans. QTLs of domestication traits in lima bean are identified in this population and their putative orthologs in common bean. A major QTL underlying determinacy was identified on chromosome 1 and collocates with major QTLs for flowering time and inflorescence position. The genetic interactions and transgressive segregation and directionality of desirable traits for breeding are reviewed and their correlations with other traits and potential impacts towards breeding advancements are analyzed.

The third chapter analyzed the variation in volatile cyanide production in different reproductive tissues of lima beans across environments subject to different levels of biotic stress as a putative biochemical mechanism of stress resistance in lima beans. Using a panel of 12 cultivars adapted to California, the variation in volatile cyanide production was evaluated across different time intervals and following repeated injury in the floral bud, immature pod, mature pod, and fresh seed tissue. There was a significant decline in volatile cyanide production from the immature (floral bud and immature pod) to the mature (mature pod and fresh seed) tissues. Studying the volatile cyanide production in the biparental recombinant inbred line population across environments subject to a range of biotic stress levels due to insecticide applications, demonstrated an increase in volatile cyanide production in plots that were subject to biotic stress. Using this biparental RIL population, a major QTL underlying volatile cyanide production in the floral bud and immature pod tissue was identified on chromosome 5, and with two and three other minor QTLs, respectively, explain over 97.5% of the phenotypic variation for these traits and maintained high heritability across environments.

The fourth chapter analyzed the variation in polygalacturonase inhibition in the floral bud tissue of lima beans across environments subject to different levels of biotic stress as a putative biochemical mechanism of stress resistance in lima beans. The variation in crude plant extract protein concentration and the inhibition of polygalacturonase sourced from *Lygus hesperus* and *Aspergillus niger* was evaluated across environments subject to different biotic stress due to insecticide applications; these biochemical traits were evaluated in a panel of 12 California-adapted cultivars and a biparental recombinant inbred line population. There was a significant ($p < 0.05$) and consistent difference between the insecticide applications on all three biochemical traits studied, with higher levels observed in the treatments that controlled for biotic stress with insecticide applications. Major QTLs were identified for both crude plant extract protein concentration and polygalacturonase inhibition of *A. niger* on chromosome 1. While all three biochemical traits had low levels of heritability, they had positive correlations with yield across the different environments subject to biotic stress.

Collectively this dissertation provides a high-quality genetic map as a key genetic and genomic resource for lima bean research, and insight into the synteny and rearrangements between lima bean and its closest relative, common bean. This genetic map combined with analysis of two putative mechanisms for biotic stress resistance allows QTL mapping of biotic stress resistance traits, which can aid in the development and advancement of lima bean cultivars with improved biotic stress resistance in the California region.

Dedication

To my children, Peter and Madeline, who joined me during my graduate journey and make every day brighter. To my husband, Peter, for his encouragement and support. To my parents and grandparents who emphasized education and perseverance to achieve one's goals in life.

Acknowledgments

I would like to thank my major professor, Paul Gepts, for the guidance, support, and wisdom he shared with me throughout my graduate journey at UC Davis.

I would like to thank members of the Gepts Lab throughout the years, including Antonia Palkovic, Kimberly Gibson, Zachary Dashner, Andrea Ariani, Travis Parker, Jorge Berny, Lorena Lopes De Sousa, Jessica Delfini, Talissa Floriani, Yolanda Guzman, Angelina Bahena and those who have helped me through various parts of my research work in the lab, in the field and in the greenhouses, and for all the stimulating discussions on bean and plant genetics.

I would like to thank Emilio Laca for the incredible amount of statistical wisdom that I have gained in his company, and for his intellectual and personal encouragement throughout my graduate years. I would like to thank Michael Kovach for introducing me to the wonderful world of plant breeding and for being a constant source of personal, professional and intellectual encouragement.

I would like to thank those within the Horticulture & Agronomy Graduate Group and Plant Science Department who have aided me in my graduate journey and research work, including Lisa Brown, my academic advisor, Andy Walker, members of my dissertation committee, Allen Van Deynze and Dan Kliebenstein, as well as John Labavitch, and members of my qualifying exam committee, Jeff Mitchell, Tom Gradziel, Mike Miller and Dan Runcie.

This research could not have been completed without the financial support of the California Dry Bean Board, the NIFA USDA OREI grant #2015-51300-24157, the Horticulture & Agronomy Graduate Groups Jastro Research Awards, and the Monsanto STEM Fellowship.

Table of Contents

Introduction

Lima beans (*Phaseolus lunatus* L.) are grown throughout the world as a succulent or dry grain legume in temperate and subtropical regions. Lima beans are an herbaceous annual dicot crop (Figure 1) harvested for their highly nutritious seeds, which contain around 20% protein and over 50% carbohydrates. They are a beneficial nitrogen-fixing rotator crop for growers (B. Holland et al., 1991; D. B. Jones et al., 1922). Lima beans are primarily self-fertilizing, but with some outcrossing due to pollinators, and are maintained as open pollinated inbred landraces and cultivars. Lima beans are notable for their crescent shaped pods, which undoubtedly guided their scientific name, *Phaseolus lunatus* L. (Bailey, 1923; Freytag & Debouck, 2002; Gutiérrez Salgado, Gepts, & Debouck, 1995).

Figure 1. Lima bean prints in the records of the Museo de Ciencias Naturales, Universidad Central de Venezuela, Caracas, Venezuela

Lima bean background & domestication

Lima bean is a member of the Fabaceae family (or Leguminosae), the third largest family of angiosperms containing approximately 750 genera and 19,500 species, and the *Phaseolus* genus, the most important leguminous genus for human consumption, globally containing 70-80 species (Christenhusz & Byng, 2016; Freytag & Debouck, 2002; P. Gepts, 2001). Lima bean is one of the five domesticated species belonging to this genus; the other domesticated *Phaseolus* species include common bean (*Phaseolus vulgaris* L.), runner bean (*Phaseolus coccineus* L.),

tepary bean (*Phaseolus acutifolius* A. Gray) and year bean (*Phaseolus dumosus* MacFady), in order of descending agricultural significance (P. Gepts, 2001). The *Phaseolus* genus was initially defined and broadly (mis)characterized by Linnaeus in 1753, and underwent multiple rounds of revisions in the centuries that followed, notably by de Candolle who declared the Americas as the origin for common and lima beans (de Candolle, 1885; Daniel G. Debouck, 1999).

Lima bean has a broad ecological adaptation across its center of diversity, ranging in both latitude and altitude, from central Mexico to northern Argentina and from sea level to 2,700 meter highlands (Paul Gepts, 2014a; Gutiérrez Salgado et al., 1995; Motta-Aldana et al., 2010). They are thought to have been cultivated in the Neotropics since the origin of agriculture some 10,000 years ago, and have had two distinct domestications supported by morphological, biochemical, paleobotanical and molecular evidence (Caicedo et al., 1999; Fofana et al. , 1999; Paul Gepts, 1988; Gutiérrez Salgado et al., 1995; Kaplan, 1965; Kaplan & Kaplan, 1988; Lioi et al., 1999; Motta-Aldana et al., 2010).

Lima bean is rare among agronomic crops but not alone in the *Phaseolus* genus in having dual domestications; common bean and lima bean each had two separate domestication events, one of Mesoamerican and another of Andean origin, which gave rise to two distinct intraspecific gene pools for each species (Mackie, 1943). The dual domestications of lima bean (Figure 2) and common bean are thought to contribute to the broader geographical span and broader genetic base of these two species compared to the other three domesticated species of *Phaseolus*, which only had a single domestication event (Baudoin et al. 2004; Daniel G. Debouck, 1999). The Andean gene pool contains the large flat lima bean seeds, known as Big

Lima varieties, and are believed to have been domesticated in the mid-altitude western valleys between Ecuador and Peru around 4000 BP (Andueza-Noh et al., 2013; Gutiérrez Salgado et al., 1995; Mackie, 1943; Motta-Aldana et al., 2010; Serrano-Serrano et al., 2010). The Mesoamerican gene pool consists of the round, small and flat, oval-shaped seeds, known as Potato and Sieva varieties, respectively, and are believed to have been domesticated in western central Mexico around 1400 BP (Andueza-Noh et al., 2013; Gutiérrez Salgado et al., 1995; Mackie, 1943; Motta-Aldana et al., 2010; Serrano-Serrano et al., 2010). The Mesoamerican and Andean gene pools of lima bean are believed to have diverged approximately 0.5 MY BP (Serrano-Serrano et al., 2010). Each has two geographically and genetically distinct groups emerge from additional domestications within their gene pools, referred to as MI, MII, AI, and AII. The AI gene pool is dispersed throughout the Andean highlands of Peru and Ecuador, and more recent genomic research proposed a genetically distinct AII gene pool distributed around central Colombia (Baudet, 1977; Chacón-Sánchez & Martínez-Castillo, 2017; Mackie, 1943). The MI gene pool is dispersed around the center of domestication occurring in western central Mexico, while the MII gene pool is dispersed around the region of eastern Mexico, Guatemala, and Costa Rica, as far east as the Yucatan Peninsula and the Caribbean and as far south as northeastern Colombia (Andueza-Noh et al., 2013; Serrano-Serrano et al. ,

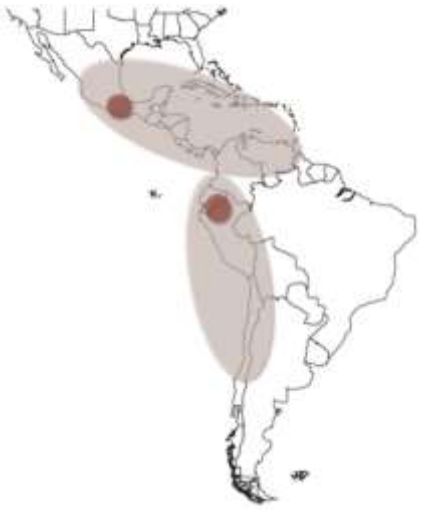

Figure 2. Domestication areas and centers of diversity of lima bean. Darker color denotes the centers of domestication for the Mesoamerican and Andean gene pools. Lighter color denotes centers of diversity for the Mesoamerican and Andean domesticated gene pools.

2012). The two Mesoamerican gene pools differ not only in their geographical regions of distribution, but in the environmental adaptations as well. The MI gene pool dispersed throughout tropical dry forests of the Mexican Pacific coastal plains at an average altitude of 448 m asl, and the MII gene pool dispersed in the more humid lowlands of the Mexican gulf at an average altitude of 543 m asl (Serrano-Serrano et al., 2012).

The MI gene pool in western central Mexico coincides with proposed regions of domestication for two other *Phaseolus* species, common and tepary bean, in addition to other important Mesoamerican crops, maize and squash (Kwak & Gepts, 2009; Kwak, Kami, & Gepts, 2009; Matsuoka et al., 2002; Nanni et al., 2011; Sanjur et al., 2002). The evidence of these multiple domestications suggests this region may be key for the origin of agriculture and the development of human civilization in Mesoamerica (Piperno et al., 2007; Ranere et al., 2009; Zizumbo-Villarreal & Colunga-GarcíaMarín, 2010).

During the domestication process in lima bean, morphological and physiological traits were subject to selection similar to other domesticated crop species (Ornduff & Harlan, 1977). Lima beans underwent convergent evolution, with Andean and Mesoamerican gene pools undergoing selection of similar traits during domestication, including loss of seed or pod shattering, loss of photoperiod sensitivity, loss of seed dormancy, and the adaption towards a prostrate growth habit (Andueza-Noh et al., 2013; Ornduff & Harlan, 1977).

The dual domestications and conspecific gene pools putatively minimize the loss of genetic diversity subject to a species under a single domestication event, and lima bean is consistent with common bean in the reduction in genetic diversity in their Mesoamerican gene

pools (Andueza-Noh et al., 2015; Chacón-Sánchez & Martínez-Castillo, 2017; Rendón-Anaya et al., 2017; Schmutz et al., 2014). However, the reduction in genetic diversity for lima bean varied among gene pools, due to deviances in founder effects or intensity of selection for domestication traits. Strong selection intensity for key traits may have reduced genetic diversity, including the selection against antinutritional compounds in seeds, including cyanogens or lectins. Antinutritional compounds may be an effective trait against pests and herbivory, but are detrimental to human consumption and would require a high selection intensity during the domestication of lima beans (Chacón-Sánchez & Martínez-Castillo, 2017; Debouck, 1996).

While selection of plants with unique desirable domestication traits may lead to a reduction in genetic diversity around loci of interest, domestication of lima bean led to increased genetic and morphological diversity in its harvested seeds, both towards an increase in the size, a wider variety of seed shapes, and the selection for novel seed colors and patterns. One of the remarkable features of domesticated lima beans is the vast array of colors and patterns in their seeds and the size and shape variation. Domesticated seeds are uniquely distinct from their wild counterparts due to this variation in the seed coat color, pattern, and sizes; the wild seeds commonly being much smaller with a dull, brown speckled pattern.

Phylogenetic classification

Throughout the 20th century and into the early 21st century, dozens of new species of *Phaseolus* have been discovered and continue to be reevaluated to determine if they are synonymous with preexisting species or are in fact distinct species. Initially, lima beans were subdivided into botanical varieties with the domesticated germplasm belonging to var. *lunatus*

and the wild germplasm belonging to var. *silvester*, however, this formal division between wild and domesticated germplasm has been abandoned due to the absence of barriers to gene flow between them (Baudet, 1977; Delgado-Salinas, 1985; Gutiérrez Salgado et al., 1995; Maréchal & Mascherpa, 1978). Phylogenetic relationships among species and subspecies is continually clarified by combining botanical and molecular evidence and is combined in Figure 3 below; hybridization studies reveal the gene flow among species and their wild allies.

Debouck classification	Delgado-Salinas classification
Paniculati section	Lunatus group: Sect. *Paniculati* , Subsect. *Volubili* Pro Parte
P. albinervus	*P. augusti*
P. augusti (= P. bolivianus)	*P. bolivianus*
P. jaliscanus	*P. lignosus*
P. juquilensis	*P. lunatus*
P. lignosus	*P. mollis*
P. longiplacentifer	*P. pachyrrhizoides*
P. lunatus	*P. viridis (= P. longiplacentifer)*
P. maculatifolius	*Polystachios group: Sects. Coriacei and Paniculati, Subsects. Volubili and Lignosi*
P. marechalii	
P. mollis	
P. nodosus	*P. albinervus*
P. pachyrrhizoides	*P. jaliscanus (= P. scrobiculatifolius)*
P. polystachios	*P. juquilensis (= P. acinaciformis)*
P. rotundatus	*P. maculatifolius*
P. salicifolius	*P. maculatus (= P. venosus)*
P. scrobiculatifolius	*P. marechalii*
P. sinuatus	*P. nodosus*
P. smilacifolius	*P. novoleonensis*
P. sonorensis	*P. polystachios*
P. viridis	*P. reticulatus*
P. xolocotzii	*P. ritensis*
Coriacei section	*P. rotundatus*
P. maculatus	*P. salicifolius*
P. novoleonensis	*P. sinuatus*
P. reticulatus	*P. smilacifolius*
P. ritensis	*P. sonorensis*
P. venosus	*P. xolocotzii*

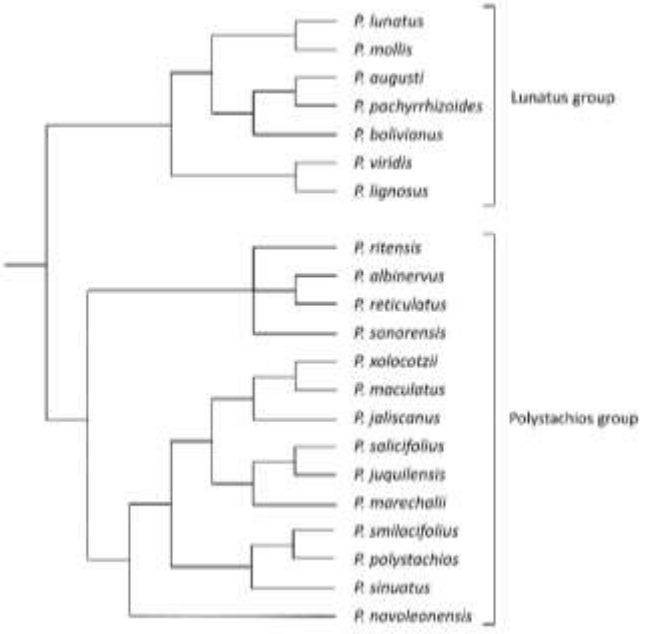

Figure 3. Classification of lima bean wild allies according to Debouck's botanical and Delgado-Salinas' molecular classification, and a phylogenic tree of the Lunatus and Polystachios groups (excluding *P. maculatifolius, P. nodosus, and P. rotundatus*).

Current phylogenetic studies have classified all *Phaseolus* species within the Phaseolus crown clade, and have subdivided the genus into Clade A and Clade B. Clade B contains all of the domesticated *Phaseolus* species and a much broader geographical and ecological adaptation than the former clade (Delgado-Salinas et al., 2006). Within Clade B, lima bean is the most phylogenetically distant of the domesticated species, belonging to a separate

phylogenetic group/section than the one shared by the other four domesticated *Phaseolus* species (Baudoin et al., 2004). Delgado-Salinas has classified the B clade based on molecular evidence into five groups: Filiformis, Vulgaris, Leptostachyus, Lunatus and Polystachios, whereas Debouck has classified the latter clade based on botanical evidence into six sections: *Acutifolii*, *Coriacei*, *Flacati*, *Paniculati*, *Phaseoli* and *Rugosi* (Debouck, 2016; Delgado-Salinas et al., 2006). The Lunatus group is believed to have been formed around 1 MYBP, and includes the wild species *P. augusti* Harms, *P.bolivianus* Piper*, P. lignosus* Britton*, P. mollis* Hook*, P. pachyrrhizoides* Harms*,* and *P. viridis* Piper[1] (Baudoin et al., 2004; Delgado-Salinas et al., 2006; P. Gepts, 2001). According to Debouck's classification in Figure 3, the *Paniculati* section includes the wild species *P. albinervus* Freytag & Debouck, *P. augusti* Harms, *P. jaliscanus* Piper, *P. juquilensis* A.Delgado, *P. lignosus* Britton, *P. longiplacentifer* Freytag, *P. lunatus* L.[1], *P. maculatifolius* Freytag & Debouck, *P. marechalii* A.Delgado, *P. mollis* Hook, *P. nodosus* Freytag & Debouck, *P. pachyrrhizoides* Harms, *P. polystachios* L. Britton, Sterns & Poggenb.[2], *P. rotundatus* Freytag & Debouck*, *P. salicifolius* Piper, *P. scrobiculatifolius* Freytag, *P. sinuatus* Nutt, *P. smilacifolius* Pollard, *P. sonorensis* Standl, *P. viridis* Piper, and *P. xolocotzii A*. Delgado (Debouck, 2016).

All species included in Debouck's classification of the *Paniculati* and *Acutifolii* sections overlap with Delgado-Salinas' classification of the Lunatus and Polystachios groups, with Delgado-Salinas subclassifying species that Debouck considers distinct, i.e., *P. longiplacentifer,*

[1] *P. longiplacentifer may be a subspecies of P. viridis* (Delgado-Salinas et al., 2006)
[2] *P. polystachios/polystachyus* contains three subspecies *P. polystachios*, *P. sinuatus* Nutt, and *P. smilacifolius* Pollard (Baudoin et al., 2004)
*denotes that there are no germplasm accessions and <3 herbarium samples collected for these species (Ramírez-Villegas, Khoury, Jarvis, Gabriel Debouck, & Guarino, 2010)

P. scrobiculatifolius, P. venosus. The Serrano-Serrano et al. (2010) molecular classification supports the organization of the Polystachios and Lunatus groups created in Delgado-Salinas's work, with the only exceptions of *P. lignosus* and *P. xolocotzii* being classified in the Polystachios and Lunatus groups, respectively. This botanical and molecular classification provides evidence of the close phylogenetic relationships of lima bean and its two dozen closest wild allies.

Gene flow and hybridization

The closest phylogenetic relatives of lima bean are three wild allies of Andean origin commonly included in the secondary gene pool: *P. augusti* Harms, *P. bolivianus* Piper, and *P. pachyrrhizoides* Harms, supporting the phylogeographical and phylogenetic belief that lima bean is of Andean origin[3] (Maquet et al., 1999; Serrano-Serrano et al., 2010). There is some debate if these three Andean allies are distinct species or a spectrum of a single species with geographical variants. Currently, Debouck and the CIAT germplasm repository classify *P. bolivianus* as a subspecies of *P. augusti* (Baudoin et al., 2004; Maquet et al., 1999). *P. augusti* is the most broadly dispersed of these three allies found in the highlands from Ecuador to Argentina and ranging from 1400 to 2980 masl [4600 to 9777 ft asl], *P. bolivianus* is found throughout Peru and Bolivia, and *P. pachyrrhizoides* is found on the highlands of the western side of Peru ranging from 1950 to 3080 masl [6400 to 10105 ft asl]. The next closest set of relatives to lima bean include *P. mollis*, which is endemic to the Galapagos Islands and believed to have South American origins, followed by the non-South American allies, *P. viridis* of Mexican

[3] *P. rosei* was identified as a wild form of lima bean of Andean origin

origin and *P. lignosus,* which is endemic to the Bermuda archipelago. These latter three allies are yet to be determined as members of the secondary or tertiary gene pool of lima bean.

Lima bean is unable to naturally hybridize with any of its domesticated *Phaseolus* relatives, and no natural interspecific hybrids have been reported between any of lima beans wild allies either (Baudoin et al., 2004; Debouck, 2016). Common bean is the closest domesticated relative to lima bean and attempts to hybridize the species through embryo rescue have resulted in high lethality rates and complete sterility in a single intermediate plant

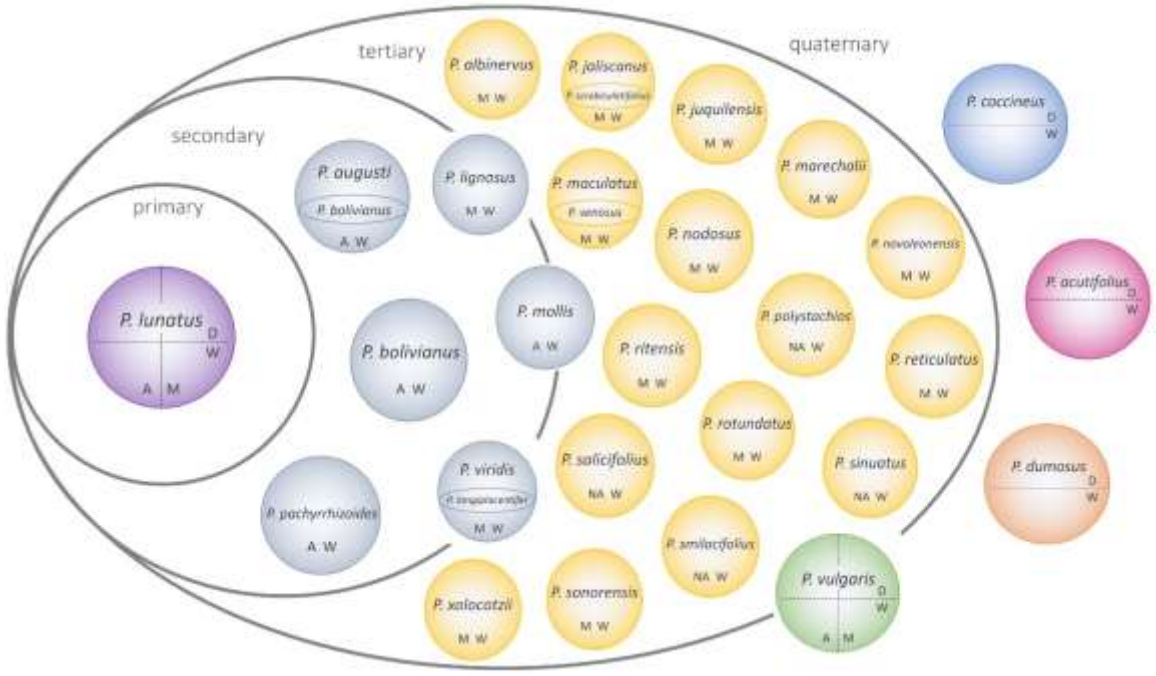

Figure 4. Summary of primary, secondary, tertiary, and quaternary gene pools of lima bean. ' A' denotes Andean origin, 'M' denotes Mesoamerican origin, 'NA' denotes North American origin, 'D' denotes domesticated, and 'W' denotes wild.

that survived to flowering (Kuboyama et al., 1991). Successful gene transfer has occurred, overcoming post-fertilization barriers, between lima bean and several closely related wild

species: *P. jaliscanus*, *P. maculatus*[4], *P. marechalii*, *P. ritensis*, *P. salicifolius*, *P. sinuatus*, and *P. smilacifolius*, and these are believed to be several of the species within the tertiary gene pool of lima bean (Baudoin & Katanga, 1990). Other species believed to be in the tertiary gene pool of lima bean include *P. albinervus, P. juquilensis, P. nodosus, P. novoleonensis, P. polystachios, P. reticulatus, P. rotundatus, P. sonorensis, and P. xolocotzii.* The theoretical classification of species into the different gene pools of lima bean are shown in Figure 4 (Baudoin & Katanga, 1990; Baudoin et al., 2004; Debouck, 2016; Kuboyama et al., 1991).

Lima bean cultural significance from the ancient to modern world

The distribution of wild and domesticated lima beans throughout central and south America stretches over 10,000 kilometers and overlaps with regions associated with the pre-Mayan and Incan ancient cultures and civilizations (Baudoin, 1988). Lima beans offer an important source of nutrition, fiber and protein for human consumption, and the significance of this crop in early civilizations is apparent in the recovered samples and artifacts of Meso- and South America. Some of the oldest archaeobotanical remains of domesticated lima beans were found in the coastal valleys of Peru and the Peruvian Andes around 5600 AD and 3500 AD,

Figure 5. Moche illustration of warrior beans ("pallares guerreros") (Jackson, 2009)

[4] *P. maculatus contains the subspecies P. ritensis*

respectively (Kaplan & Lynch, 1999). Linguistic references to beans in Mesoamerica dates back to a proto-Mayan language around 3400 BP (Brown, 2006; Paul Gepts, 2014a).

Numerous illustrations and ceramics featuring bean iconography circa 100-800 AD have been recovered in the Moche and Chicama valleys along the northern coast of Peru as artifacts from the Moche culture, highlighting their reverence for lima beans (Figure 5 and Figure 6) (Melka, 2010; Ryser, 1998). Evidence suggests that lima bean's role as a subsistence crop evolved to greater cultural significance, particularly in anthropomorphic illustrations of the bean warriors (Figure 5) (Jackson, 2009), and the development of a proto-record system utilizing lima beans to mark different patterns with the full meaning still unknown (Ryser, 1998). The Moche culture was succeeded by the Chimu civilization around 900 AD, and later conquered by the Incan Empire around 1470 AD before the arrival of Spanish conquistadors during the 1500s, and throughout the lima bean persisted as an important staple crop (Ryser, 1998).

Figure 6. Moche ceramic featuring lima bean iconography (Museo Larco, Lima, Peru)

Lima beans were dispersed throughout the Americas over thousands of years, and then spread rapidly across the world following the arrival of European explorers in the Americas. Spanish conquistadors were introduced to lima beans by the Incans and spread this legume throughout Europe and Asia, and eventually to Africa through the slave trade (Mackie, 1943; Nwokolo, 1996). Lima bean's popularity is marked by its appearance in common recipes, including the Greek dish *gigandes plaki*, and the Balkan/Serbian dish *prebranac* and the incorporation as a staple in stir-fry recipes throughout China and Asia.

Lima bean has been referenced sporadically throughout historical and cultural texts, which allows us to determine its level of significance at various times. Native Americans relied on the cultivation of the Three Sisters (corn, squash, and beans) and cooked these together to make the popular dish succotash. This recipe was shared with colonial settlers and led to the adopted propagation of lima beans by colonial settlers and later by several of the founding fathers. George Washington grew lima beans at his Mount Vernon garden and wrote of sending lima beans in a correspondence in 1793 (Burnett & Fitzpatrick, 1934; Manca, 2012). Thomas Jefferson cultivated lima beans in his Monticello garden and his daughter, Martha, includes lima beans as an ingredient in her recipe for okra soup (Hatch, 2010). The lima beans cultivated by the Cherokees and other tribes along the Atlantic coasts, from Florida to Virginia, resembled those grown in the American Southwest by the Hopi tribes, reflecting Mesoamerican origins (Mackie, 1943). Lima beans found in the Caribbean and West Indies are believed to have been transported in canoes by the Carib tribe of the Yucatan peninsula and are also consistent with Mesoamerican origins, but are distinct from those found in the Southeastern US (Mackie, 1943).

Lima beans were introduced, or likely reintroduced, to California when a shipment from South America docked in Santa Barbara in the 1870s. The common name in English traces its origins to this time when the novel bean was referred to by the Peruvian city the shipment originated from- the Lima (LEE-muh) bean or the lima (LIE-muh) bean (Dubroff, 2014). Lima beans reached peak production with 150,000 acres planted in 1914 when 75% of the world's production was from Ventura County alone (Dubroff, 2014; Lazicki et al., 2016). Lima bean production in the early 20th century was primarily along the coast, but the introduction of

irrigation in California in the 1960s and 1970s shifted coastal production towards higher value crops, including strawberries and avocadoes, and lima bean production shifted inland to the Sacramento and San Joaquin Valleys.

To preserve the genetic diversity of beans and improve the crop, the International Center for Tropical Agriculture (CIAT) was created in 1967. It is based in Cali, Colombia, and it housed the largest and most diverse collection of beans globally. CIAT is a part of the CGIAR (Consultative Group on International Agricultural Research), which focuses on food security-related research (Figure 7). To-date, CIAT maintains nearly 38,000 *Phaseolus* accessions containing 44 taxa from 112 countries of origin (www.ciat.cgiar.org, 2020). CIAT is a vital resource for the genetic preservation and research and advancement of bean species; its collection of lima bean includes some 3,305 accessions across cultivated, wild, escape, weedy and hybrid lima bean statuses. Other germplasm collections of lima bean include the USDA National Plant Germplasm System located at the Western Plant Introduction Station in Pullman, Washington containing 6,065 accessions of lima beans, and EMBRAPA CENARGEN in Brasilia (www.genesys-pgr.org, 2020). The number of accessions for lima bean and its wild relatives, as well as their gene pool and classifications, are included in Table 1.

Table 1. Summary of lima bean and its wild relatives' classifications, gene pools, origins and germplasm accessions.

Genus species	Delgado-Salinas classification	Debouck classification	Gene Pool	Origin
P. lunatus L.	Lunatus Group	Paniculati Section	Primary	Andean-Mesoamerican
P. augusti Harms	Lunatus Group	Paniculati Section	Secondary	Andean
P. bolivianus Piper	Lunatus Group	Paniculati Section	Secondary	Andean
P. pachyrrhizoides Harms	Lunatus Group	Paniculati Section	Secondary	Andean
P. mollis Hook	Lunatus Group	Paniculati Section	Tertiary	Galapagos Islands-Andean
P. viridis Piper	Lunatus Group	Paniculati Section	Tertiary	Mesoamerican
P. viridis Piper *subsp. longiplacentifer* Freytag & Debouck	Lunatus Group	Paniculati Section	Tertiary	Mesoamerican
P. lignosus Britton	Lunatus Group	Paniculati Section	Tertiary	Bermudan Islands
P. albinervus Freytag & Debouck	Polystachios Group	Paniculati Section	Quarternary	Mesoamerican
P. jaliscanus Piper	Polystachios Group	Paniculati Section	Quarternary	Mesoamerican
P. jaliscanus Piper *subsp. scrobiculatifolius* Freytag	Polystachios Group	Paniculati Section	Quarternary	Mesoamerican
P. juquilensis A.Delgado	Polystachios Group	Paniculati Section	Quarternary	Mesoamerican
P. juquilensis A.Delgado *subsp. acinaciformis* Freytag & Debouck	Polystachios Group	Paniculati Section	Quarternary	Mesoamerican
P. maculatifolius Freytag & Debouck	Polystachios Group	Paniculati Section	Quarternary	Mesoamerican
P. maculatus Scheele	Polystachios Group	Coriacei Section	Quarternary	Mesoamerican
P. maculatus Scheele *subsp. ritensis* (M.E. Jones) Freytag	Polystachios Group	Coriacei Section	Quarternary	Mesoamerican
P. maculatus Scheele *subsp. venosus* Piper	Polystachios Group	Coriacei Section	Quarternary	Mesoamerican
P. marechalii A.Delgado	Polystachios Group	Paniculati Section	Quarternary	Mesoamerican
P. nodosus Freytag & Debouck	Polystachios Group	Paniculati Section	Quarternary	Mesoamerican
P. novoleonensis Debouck	Polystachios Group	Coriacei Section	Quarternary	Mesoamerican
P. polystachios L. Britton et al.	Polystachios Group	Paniculati Section	Quarternary	North America
P. polystachios L. Britton et al. *subsp. polystachios*	Polystachios Group	Paniculati Section	Quarternary	North America
P. polystachios (L.) Britton et al. *subsp sinuatus* Nutt	Polystachios Group	Paniculati Section	Quarternary	North America
P. polystachios L. Britton et al. *subsp. smilacifolius* Pollard	Polystachios Group	Paniculati Section	Quarternary	North America
P. reticulatus Freytag & Debouck	Polystachios Group	Coriacei Section	Quarternary	Mesoamerican
P. rotundatus Freytag & Debouck	Polystachios Group	Paniculati Section	Quarternary	Mesoamerican
P. salicifolius Piper	Polystachios Group	Paniculati Section	Quarternary	Mesoamerican
P. sonorensis Standley	Polystachios Group	Paniculati Section	Quarternary	Mesoamerican
P. xolocotzii A.Delgado	Polystachios Group	Paniculati Section	Quarternary	Mesoamerican

**P. bolivianus* categorized within *P. augusti* in the GENESYS and CIAT gene banks.

Lima bean production in California

Today, domesticated lima bean, including landraces and cultivars, retains morphological diversity for traits subject to consumer preference that distinguish their current market classes for production. Most notably is the distinction between large and baby lima beans. Early classification of lima bean designated the two seed classes as separate species, *P. lunatus* containing the small, round or flat *sieva* beans and the large, flat beans belonging to a separate species, *P. limensis* (Bailey, 1923; Gutiérrez Salgado et al., 1995; Yarnell, 1965). However, today the size of lima beans denotes the two subclasses of the *P. lunatus* species for the smaller, round 'baby' limas and the larger, flat 'large' limas. The secondary distinction between lima beans, which is more important for farmers than consumers, is the indeterminate vine and the determinate bush growth habit. The vigorous, climbing or prostrate growth habit of the indeterminate lima bean necessitates a different field management approach from its upright and compact counterpart in the determinant growth habit, and, thus, they are not often found intermixed within a field.

Lima bean production in the United States is divided among succulent bean production in the mid-Atlantic region and dry bean production in California. Lima beans account for nearly half of the total dry bean production in California, with Stanislaus, San Joaquin and Sutter counties leading lima bean production over the past decade (Figure 7A) (USDA NASS, 2019). Dry lima bean production in the US is valued around $30 million dollars annually over the past

decade, accounting for 60-80% of total global production, with California as the sole producer of dry lima bean production nationally (Long et al., 2014; USDA ERS, 2004; USDA NASS, 2019).

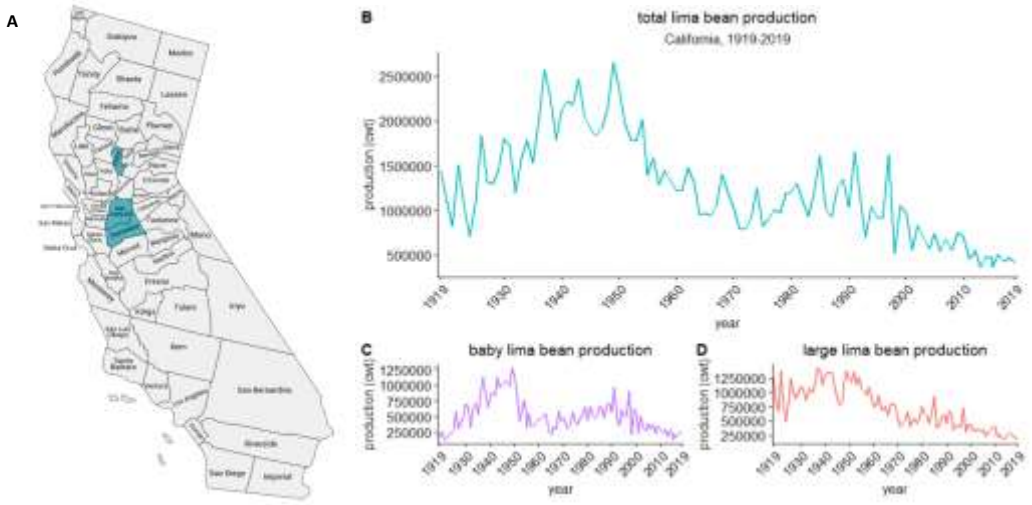

Figure 7. A. Location of the top three Counties for lima bean production B. Total lima bean production in California for the past 100 years C. Total baby lima bean production in California for the past 100 years D. Total large lima bean production in California for the past 100 years.

Lima bean production has been in decline since the mid-twentieth century for both large and baby lima bean production (Figure 7B-D) (USDA NASS, 2019). This production decline may be attributed in part to competition from higher-value irrigated crops, but in recent years lima bean yields have also suffered from significant pathogen and pest pressure. Bean fields are commonly furrow irrigated; however, the high clay content of many fields in the Central Valley may result in increased waterlogging, soil pathogens, and root disease if overirrigated. Particularly significant soilborne diseases include seedling root rot caused by *Rhizoctonia solani* and *Pythium* fungi, *Fusarium* root rot, *Sclerotinia* white mold, and *Macrophomina* charcoal rot at maturity (Long et al., 2014). Drip and sprinkler irrigation systems mitigate overirrigation resulting from furrow irrigation, however, these alternatives may increase production costs,

and sprinkler irrigation may increase the likelihood of foliar pathogens, such as bacterial blights (Long et al., 2014).

The most significant factor impacting lima bean production in California in recent year is pest pressure. Pest pressure is particularly significant in organic lima bean production, which is limited in access to effective pesticides to deter pests. In the California Central Valley, common pests affecting lima bean production include lygus bugs, nematodes, aphids, and spider mites, which each vary in their mobility, seasonal peaks, and targeted organ tissue for lima beans.

Lygus hesperus Knight

The Western tarnished plant bug, *Lygus hesperus* Knight, is the most severe contributor to agronomic losses of lima bean in the Western US and can result in a 70% yield reduction (Dohle, 2017; Long et al., 2014). *L. hesperus* is a pest to other valuable agronomic crops in California, including strawberry, alfalfa seed, and cotton, and has a relative in the tarnished plant bug, *Lygus lineolaris,* which impacts cotton, carrot seed, and other crops in the eastern US (Scott, 1977; Young, 1986). Both species belong to the Hemiptera order, the Heteroptera suborder, and Miridae family and, with other species, colloquially fall within the classification of mirid bugs, which are easily identified by a distinct triangular segment at the base of their wings, and true bugs, which are piercing and sucking insects.

L. hesperus can be found from southern Canada to northern Mexico, hosted on hundreds of agronomic and native plants, is highly mobile and is able to inflict significant damage to crop yields even at low population levels (Long et al., 2014; Mueller et al., 2003). Due to the high mobility of adult lygus insects, chemical methods are limited in their

effectiveness on adult population control. Thus, there is a demand for varieties that confer resistance to lygus damage. Increasing popularity of organic production in crops, particularly strawberry, has led to trap cropping with the lygus preferred host, alfalfa, and multiple rounds of tractor-mounted vacuuming to control lygus populations in the field (Swezey et al., 2007). Previous research has concluded that resistance to lygus in lima beans is heritable, however, the mode of inheritance and mechanisms of resistance are unknown (Dohle, 2017).

Lygus sp. are piercing and sucking insects that use stylets to penetrate plant tissue, deposit salivary enzymes for extra-oral digestion, and siphon degraded plant tissue for consumption, a process known as lacerate-and-flush (Miles & Taylor 1994). Lygus saliva initiates the process of tissue degradation in the plant cell walls using pectinase and polygalacturonase digestive enzymes (Strong 1970). These enzymes work effectively to degrade cell wall molecules and enable pests to actively digest the plant tissue of many economically important crops, including lima bean. During flowering season, lygus insects target the floral bud tissue of lima bean with this lacerate-and-flush method of feeding resulting in floral and pod abortion and a reduction of grain yield. When these lygus insects feed on the developing pods later in the season, it can result in grains with noticeable dimples and blemishes that reduce the quality of the beans harvested and lead to reduction in harvestable yield or a lower price per weight for the grower.

Figure 8. Photo of adult Lygus bug on lima bean floral buds.

Plant defense mechanisms

Plants have evolved sophisticated defense mechanisms to protect against herbivore damage, including morphological traits and biochemical compounds. Resistance can be expressed throughout the plant or locally in vulnerable tissues at different developmental stages. Defense mechanisms to herbivory may either be constitutively expressed by the plant or induced upon herbivory. Constitutive resistance is expressed solely by the underlying genetics of the plant and not in response to the herbivory by the insect (Rechcigl & Rechcigl, 2000). Induced resistance is the expression of resistant traits in the plant following environmental stimulation caused by herbivory (Rechcigl & Rechcigl, 2000). Inducibility as a form of phenotypic plasticity, is particularly important for the judicious allocation of plant resources towards defensive traits, and in the absence of herbivory, these limited resources could be allocated towards fitness and reproduction (Holeski et al., 2010).

Resistance, both constitutive or induced, can be categorized as either antixenosis, antibiosis, or tolerance, depending on the interaction between the plant and the herbivore (Edwards & Singh, 2006; Elger et al., 2009; Garber, 1951; Kogan & Ortman, 1978). Antixenosis, also known as non-preference, occurs when the plant deters insects from initial contact, feeding or oviposition damage by repelling the herbivore in their sensory perceptions of the plant, resulting in adverse effects on the insect's behavior but no effect on the insects survival (Garber, 1951; Rechcigl & Rechcigl, 2000). Mechanisms of antixenosis may include the release of volatile organic compounds, the presence of glandular and nonglandular trichomes, trichome density, and the presence of surface waxes on the plant that deter the initiation of herbivore damage. Antibiosis occurs when the insect actively feeds on the plant tissue and there is an

adverse effect on the fitness and mortality of the insect (Garber, 1951; Rechcigl & Rechcigl, 2000). The ingestion of detrimental compounds may include cyanogens, digestive inhibitors such as high lignin, inhibiting proteins, and silica tissues, and allelochemicals. Tolerance is the third category of resistance which is the ability of a plant to regenerate damaged tissues and maintain fitness following uninhibited insect damage (Garber, 1951; Rechcigl & Rechcigl, 2000). This method of resistance does not result in any adverse impact on the insect's fitness or survival, and solely focuses on the plants ability to recover following herbivory.

One potential mechanism of antibiosis resistance that is of interest for lima bean resistance to lygus includes the constitutive storage of cyanogenic glucosides in the cell which is metabolized to produce volatile hydrogen cyanide upon herbivory. The production of cyanide involves the reaction of the cyanogenic glucoside substrate, stored within the vacuole, and the β-glucosidase enzyme, located in the apoplast, which react to produce toxic hydrogen cyanide in response to plant tissue damage (Frehner & Conn, 1987; Miller & Conn, 1980). The specific substrate of this reaction in lima bean is linamarin and appropriately the associated enzyme is linamarase. In the absence of herbivory, the enzyme and the substrate are separated by the cell membrane and vacuolar membrane, which prevent the production of hydrogen cyanide unless those membranes are disrupted. It is only after herbivore damage ruptures both membranes that the linamarin substrate and linamarase enzyme react producing volatile hydrogen cyanide as the product of interest, and acetone and glucose as byproducts (Figure 9). The repellent effect of cyanogenesis on plant herbivores has been observed in lima bean with the pest *Schistocerca gregaria* Forskal, as well as in white clover and cassava, and may play a role in

antixenosis and/or antibiosis in lima bean's response to lygus pressure (Ballhorn, Lieberei, & Ganzhorn, 2005; Miller & Conn, 1980).

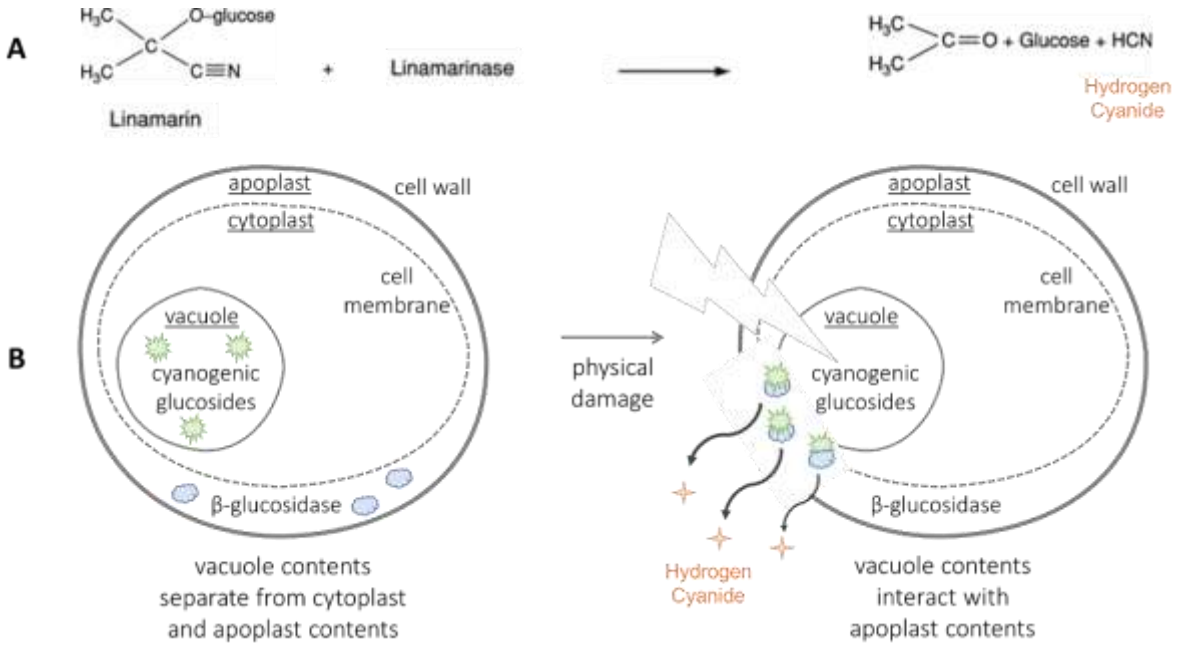

Figure 9. A. The chemical reaction of linamarin (cyanogenic glucoside substrate) with linamarase (B-glucosidase enzyme) to produce an acetone, glucose, and hydrogen cyanide B. The cellular disruption following herbivory that allows the substrate and enzyme to react.

A second potential mechanism of resistance to herbivore damage includes the production of polygalacturonase inhibiting proteins (PGIPs) that bind to and mitigate enzymatic damage caused by the release of digestive enzymes, polygalacturonases, upon lygus feeding (Bergmann et al., 1994; Celorio-Mancera et al., 2008; Shackel et al., 2005). PGIP production and other plant defenses are induced by the oligogalacturonides that result from polygalacturonase digestion of the cell wall (Figure 10) (Karr & Albersheim 1970, Aziz et al. 2004). These polygalacturonases are produced in the saliva of lygus (Allen & Mertens, 2008; Celorio-Mancera et al., 2008; Celorio-Mancera et al., 2009). Genes coding for PGIPs have been identified in wheat, rice, chickpea, and, of specific interest, four genes have been characterized in common

bean that cluster in a 50 kb locus (Kalunke et al., 2015). These PGIPs may play a role in inhibiting the action of pathogen and possibly insect polygalacturonases (Agüero et al., 2005; Powell et al., 2000).

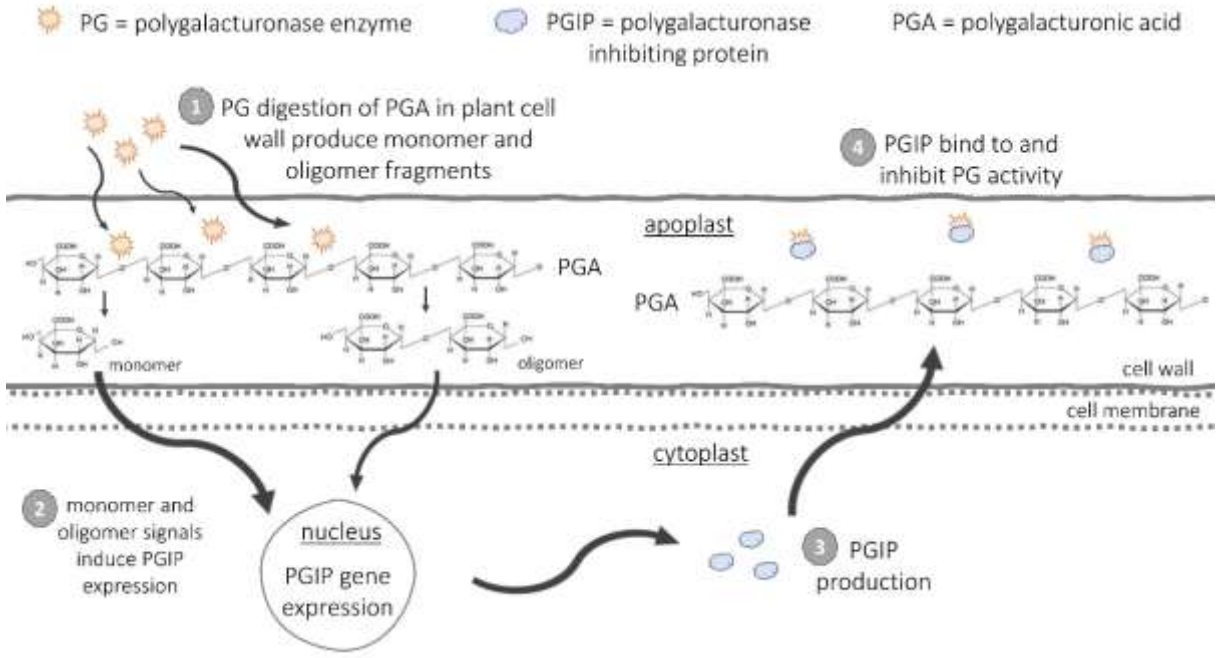

Figure 10. Polygalacturonase inhibiting proteins (PGIP) cellular response to polygalacturonase (PG) enzymatic digestion of polygalacturonic acid (PGA) polymers in the cell wall.

Dissertation objectives

Lygus herbivory is currently the largest threat to lima bean production in California and understanding different mechanisms and traits that confer resistance will enhance lima bean production in the primary growing region of the California Central Valley. The vastness of lima beans adaptation to drier and hotter environments, especially compared to the close relative common bean, may increase its breeding potential for biotic and abiotic stresses, including tolerance to lygus pressure and other undetermined stresses that may arise with climate change (Gutiérrez-Salgado et al 1995; Mackie 1943). Maintaining genetic variation in plant

populations and phenotypic plasticity is critical to adapting to dynamic herbivore pressure and identifying effective mechanisms of resistance.

The primary objectives of this dissertation are to (1) develop a molecular linkage map for lima bean that will (a) establish the relationship between genetic distance and physical position of loci within the UC 92 – UC Haskell recombinant inbred line population and (b) highlight chromosomal rearrangements between the common bean and lima bean reference genomes revealed by the two tandem alignments of the genetic maps, (2) evaluate the role of cyanogenesis as a putative mechanism of lygus resistance in the reproductive tissues of lima beans, (3) evaluate the role of polygalacturonase-inhibiting proteins as a putative mechanism of lygus resistance in the reproductive tissues of lima beans, and (4) evaluate the role of agronomic and lygus -resistance traits in yield production under different environmental conditions and map relevant trait QTL.

References

Agüero, C. B., Uratsu, S. L., Greve, C., Powell, A. L. T., Labavitch, J. M., Meredith, C. P., & Dandekar, A. M. (2005). Evaluation of tolerance to Pierce's disease and Botrytis in transgenic plants of Vitis vinifera L. expressing the pear PGIP gene. *Molecular Plant Pathology*, *6*(1), 43–51. https://doi.org/10.1111/J.1364-3703.2004.00262.X

Allen, M. L., & Mertens, J. A. (2008). Molecular cloning and expression of three polygalacturonase cDNAs from the tarnished plant bug, Lygus lineolaris. *Journal of Insect Science*, *8*. https://doi.org/10.1673/031.008.2701

Andueza-Noh, R. H., Martínez-Castillo, J., & Chacón-Sánchez, M. I. (2015). Domestication of small-seeded lima bean (Phaseolus lunatus L.) landraces in Mesoamerica: evidence from microsatellite markers. *Genetica*, *143*(6), 657–669. https://doi.org/10.1007/s10709-015-9863-0

Andueza-Noh, R. H., Serrano-Serrano, M. L., Chacón Sánchez, M. I., Sanchéz del Pino, I., Camacho-Pérez, L., Coello-Coello, J., … Martínez-Castillo, J. (2013). Multiple domestications of the Mesoamerican gene pool of lima bean (Phaseolus lunatus L.): Evidence from chloroplast DNA sequences. *Genetic Resources and Crop Evolution*, *60*(3), 1069–1086. https://doi.org/10.1007/s10722-012-9904-9

Bailey, L. H. (1923). Article IV. Various Cultigens, and Transfers in Nomenclature. *Gentes Herbarum*, *1*(3), 120-.

Ballhorn, D. J., Lieberei, R., & Ganzhorn, J. U. (2005). Plant cyanogenesis of Phaseolus lunatus and its relevance for herbivore-plant interaction: The importance of quantitative data. *Journal of Chemical Ecology*, *31*(7), 1445–1473. https://doi.org/10.1007/s10886-005-5791-2

Baudet, J. C. (1977). taxonomic status of the cultivated types of lima bean (Phaseolus lunatus L.). *Trop Grain Legume Bull*.

Baudoin, J. P. (1988). Genetic Resources, Domestication and Evolution of Lima Bean, Phaseolus lunatus (pp. 393–407). Springer, Dordrecht. https://doi.org/10.1007/978-94-009-2786-5_16

Baudoin, J. P., & Katanga, K. (1990). Phyletic relationships within the genus Phaseolus on basis of pollen morpho-logy and experimental hybridization. *Bean Improvement Cooperative*, *Annual Rep*, 117–118.

Baudoin, J. P., Rocha, O., Degreef, J., Maquet, A., & Guarino, L. (2004). *Ecogeography, demography, diversity and conservation of Phaseolus lunatus L ... - Google Books*. *Systematic and Ecogeographic Studies on Crop Genepools*. Retrieved from https://books.google.com/books?id=OyAFXM82EyEC&pg=PA3&lpg=PA3&dq=phaseolus+augusti+pachyrrhizoides+bolivianus&source=bl&ots=G4xHn7zaxz&sig=ACfU3U0QPCXwWmo_vxrXu-yMhjpBDFMRPQ&hl=en&sa=X&ved=2ahUKEwjEw-

aAyvroAhXJJjQIHR7zD0UQ6AEwAXoECAwQAQ#v=onepage&q&f=fals

Bergmann, C. W., Ito, Y., Singer, D., Albersheim, P., Darvill, A. G., Benhamou, N., … De Lorenzo, G. (1994). Polygalacturonase-inhibiting protein accumulates in Phaseolus vulgaris L. in response to wounding, elicitors and fungal infection. *The Plant Journal*, *5*(5), 625–634. https://doi.org/10.1111/j.1365-313X.1994.00625.x

Brown, C. H. (2006). Prehistoric Chronology of the Common Bean in the New World: The Linguistic Evidence. *American Anthropologist*, *108*(3), 507–516. https://doi.org/10.1525/aa.2006.108.3.507

Burnett, E. C., & Fitzpatrick, J. C. (1934). The Writings of George Washington from the Original Manuscript Sources, 1745-1799. *The American Historical Review*, *39*(3), 535. https://doi.org/10.2307/1840172

Caicedo, A. L., Gaitán, E., Duque, M. C., Toro Chica, O., Debouck, D. G., & Tohme, J. (1999). AFLP fingerprinting of Phaseolus lunatus L. and Related wild species from South America. *Crop Science*, *39*(5), 1497–1507. https://doi.org/10.2135/cropsci1999.3951497x

Candolle, E. de. (1885). Origin of Cultivated Plants. *Origin of Cultivated Plants.*

Celorio-Mancera, M. de la P., Allen, M. L., Powell, A. L., Ahmadi, H., Salemi, M. R., Phinney, B. S., … Labavitch, J. M. (2008). Polygalacturonase causes lygus-like damage on plants: Cloning and identification of western tarnished plant bug (Lygus hesperus) polygalacturonases secreted during feeding. *Arthropod-Plant Interactions*, *2*(4), 215–225. https://doi.org/10.1007/s11829-008-9050-7

Celorio-Mancera, M. de la P., Greve, L. C., Teuber, L. R., & Labavitch, J. M. (2009). Identification of endo- and exo-polygalacturonase activity in Lygus hesperus (Knight) salivary glands. *Archives of Insect Biochemistry and Physiology*, *70*(2), 122–135. https://doi.org/10.1002/arch.20282

Chacón-Sánchez, M. I., & Martínez-Castillo, J. (2017). Testing domestication scenarios of Lima bean (Phaseolus lunatus L.) in mesoamerica: Insights from genome-wide genetic markers. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.01551

Christenhusz, M. J. M., & Byng, J. W. (2016). The number of known plants species in the world and its annual increase. *Phytotaxa*, *261*(3), 201–217. Retrieved from https://www.biotaxa.org/Phytotaxa/article/view/phytotaxa.261.3.1/20598

Debouck, D. G. (1996). Colombian Common and Lima Beans: Views on their Origin and Evolutionary Significance. *Corpoica Ciencia y Tecnología Agropecuaria*, *1*(1), 7. https://doi.org/10.21930/rcta.vol1_num1_art:146

Debouck, D. G. (1999). Diversity in Phaseolus Species in Relation to the Common Bean (pp. 25–52). Springer, Dordrecht. https://doi.org/10.1007/978-94-015-9211-6_2

Debouck, D. G. (2016). *Beans: seven crops to breed, seventy-seven genetic resources to use sustainably*. Retrieved from

https://cgspace.cgiar.org/bitstream/handle/10568/82959/Conferencia Villahermosa 27 Septiembre 2016.pdf?sequence=1

Delgado-Salinas, A. O. (1985). Systematics of the genus Phaseolus (Leguminosae) in North and Central America. *Ph.D. Thesis. University of Texas, Austin*. Retrieved from https://elibrary.ru/item.asp?id=7465320

Delgado-Salinas, A. O., Bibler, R., & Lavin, M. (2006). *Phylogeny of the Genus Phaseolus (Leguminosae): A Recent Diversification in an Ancient Landscape*. *Source: Systematic Botany* (Vol. 31).

Dohle, S. (2017). *Development of Resources for Lima Bean (Phaseolus lunatus) Breeding and Genetics Research*. University of California, Davis.

Dubroff, H. (2014). Spilling the beans on Ventura County's lima legacy. *Pacific Coast Business Times*. Retrieved from www.pacbiztimes.com/2014/05/02/spilling-thebeans-on-ventura-countys-lima-legacy/

Edwards, O., & Singh, K. B. (2006). Resistance to insect pests: What do legumes have to offer? *Euphytica*, *147*(1–2), 273–285. https://doi.org/10.1007/s10681-006-3608-1

Elger, A., Lemoine, D. G., Fenner, M., & Hanley, M. E. (2009). Plant ontogeny and chemical defence: Older seedlings are better defended. *Oikos*, *118*(5), 767–773. https://doi.org/10.1111/j.1600-0706.2009.17206.x

Fofana, B., Baudoin, J. P., Vekemans, X., Debouck, D. G., & Du Jardin, P. (1999). Molecular evidence for an andean origin and a secondary gene pool for the Lima bean (Phaseolus lunatus L.) using chloroplast DNA. *Theoretical and Applied Genetics*, *98*(2), 202–212. https://doi.org/10.1007/s001220051059

Frehner, M., & Conn, E. E. (1987). The Linamarin β-Glucosidase in Costa Rican Wild Lima Beans (Phaseolus lunatus L.) Is Apoplastic . *Plant Physiology*, *84*(4), 1296–1300. https://doi.org/10.1104/pp.84.4.1296

Freytag, G. F., & Debouck, D. G. (2002). Taxonomy, Distribution, and Ecology of the Genus Phaseolus (Leguminosae–Papilionoideae) in North America, Mexico, and Central America. *Botanical Research Institute of Texas, Forth Worth*. https://doi.org/10.1663/0013-0001(2003)057[0421:bredfa]2.0.co;2

Garber, R. J. (1951). Insect Resistance in Crop Plants. *Agronomy Journal*, *43*(12), 638–638. https://doi.org/10.2134/agronj1951.00021962004300120016x

Gepts, P. (2001). Phaseolus vulgaris (Beans). In *Encyclopedia of Genetics* (pp. 1444–1445). Elsevier. https://doi.org/10.1006/rwgn.2001.1749

Gepts, Paul. (1988). *Genetic Resources of Phaseolus Beans*. *Genetic resources of Phaseolus beans* (Vol. 6). https://doi.org/10.1007/978-94-009-2786-5

Gepts, Paul. (2014). Beans: Origins and Development. In *Encyclopedia of Global Archaeology* (pp. 822–827). Springer New York. https://doi.org/10.1007/978-1-4419-0465-2_2169

Gutiérrez Salgado, A., Gepts, P., & Debouck, D. G. (1995). Evidence for two gene pools of the Lima bean, Phaseolus lunatus L., in the Americas. *Genetic Resources and Crop Evolution*, *42*(1), 15–28. https://doi.org/10.1007/BF02310680

Hatch, P. J. (2010). Thomas Jefferson's Legacy in Gardening and Food | Thomas Jefferson's Monticello. Retrieved July 2, 2020, from https://www.monticello.org/house-gardens/center-for-historic-plants/twinleaf-journal-online/thomas-jefferson-s-legacy-in-gardening-and-food/

Holeski, L. M., Chase-Alone, R., & Kelly, J. K. (2010). The genetics of phenotypic plasticity in plant defense: Trichome production in mimulus guttatus. *American Naturalist*, *175*(4), 391–400. https://doi.org/10.1086/651300

Holland, B., Widdowson, E. M., Unwin, I. D., & Buss, D. H. (1991). *Vegetables, Herbs and Spices: Fifth Supplement to McCance and Widdowson's The Composition of Foods*. *Vegetables, Herbs and Spices: Fifth Supplement to McCance and Widdowson's The Composition of Foods*. Retrieved from https://books.google.com/books?hl=en&lr=&id=-M1khtZ31BwC&oi=fnd&pg=PA7&dq=holland+lima+beans+1991+unwin&ots=TcSyx1myu_&sig=itr-6lk4sqMWKzALCk2EeHpOkjM

Jackson, M. (2009). Moche art and visual culture in ancient Peru. *Choice Reviews Online*, *47*(01), 47-0360-47–0360. https://doi.org/10.5860/choice.47-0360

Jones, D. B., Gersdorff, C. E. F., Johns, C. O., & Finks, A. J. (1922). The proteins of the lima bean, Phaseolus lunatus. *Journal of Biological Chemistry*, *53*(2), 231–240. https://doi.org/10.1016/s0021-9258(18)85778-7

Kalunke, R. M., Tundo, S., Benedetti, M., Cervone, F., De Lorenzo, G., & D'Ovidio, R. (2015). An update on polygalacturonase-inhibiting protein (PGIP), aleucine-rich repeat protein that protects crop plants against pathogens. *Frontiers in Plant Science*, *6*(MAR), 1–17. https://doi.org/10.3389/fpls.2015.00146

Kaplan, L. (1965). Archeology and domestication in American Phaseolus (Beans). *Economic Botany*, *19*(4), 358–368. https://doi.org/10.1007/BF02904806

Kaplan, L., & Kaplan, L. N. (1988). Phaseolus in Archaeology (pp. 125–142). Springer, Dordrecht. https://doi.org/10.1007/978-94-009-2786-5_7

Kaplan, L., & Lynch, T. F. (1999). Phaseolus (Fabaceae) in archaeology: AMS radiocarbon dates and their significance for pre-colombian agriculture. *Economic Botany*, *53*(3), 261–272. https://doi.org/10.1007/BF02866636

Kogan, M., & Ortman, E. F. (1978). Antixenosis–A New Term Proposed to Define Painter's "Nonpreference" Modality of Resistance. *Bulletin of the Entomological Society of America*, *24*(2), 175–176. https://doi.org/10.1093/BESA/24.2.175

Kuboyama, T., Shintaku, Y., & Takeda, G. (1991). Hybrid plant of Phaseolus vulgaris L. and P. lunatus L. obtained by means of embryo rescue and confirmed by restriction endonuclease analysis of rDNA. *Euphytica*, *54*(2), 177–182. https://doi.org/10.1007/BF00039605

Kwak, M., & Gepts, P. (2009). Structure of genetic diversity in the two major gene pools of common bean (Phaseolus vulgaris L., Fabaceae). *Theoretical and Applied Genetics*, *118*(5), 979–992. https://doi.org/10.1007/s00122-008-0955-4

Kwak, M., Kami, J. A., & Gepts, P. (2009). The putative mesoamerican domestication center of Phaseolus vulgaris is located in the Lerma-Santiago basin of Mexico. *Crop Science*, *49*(2), 554–563. https://doi.org/10.2135/cropsci2008.07.0421

Lazicki, P., Geisseler, D., & Horwath, W. R. (2016). Dry Bean Production in California. *University of California, Davis*.

Lioi, L., Sparvoli, F., & Bollini, R. (1999). Variation and genomic polymorphism of lectin-related proteins in Lima bean (Phaseolus lunatus L.) seeds. *Genetic Resources and Crop Evolution*, *46*(2), 175–182. https://doi.org/10.1023/A:1008630330008

Long, R., Temple, S., Meyer, R., Schwankl, L., Godfrey, L., Canevari, M., & Roberts, P. (2014). Lima Bean Production in California. *Lima Bean Production in California*, (May). https://doi.org/10.3733/ucanr.8505

Mackie, W. W. (1943). Origin, dispersal, and variability of the lima bean, Phaseolus lunatus. *Hilgardia*, *15*(1), 1–29. https://doi.org/10.3733/hilg.v15n01p001

Manca, J. (2012). *George Washington's eye: Landscape, architecture, and design at Mount Vernon*. *George Washington's Eye: Landscape, Architecture, and Design at Mount Vernon*. https://doi.org/10.1093/jahist/jat018

Maquet, A., Vekemans, X., & Baudoin, J.-P. (1999). *Phylogenetic study on wild allies of Lima bean, Phaseolus lunatus (Fabaceae), and implications on its origin*. *Plant Syst. Evol* (Vol. 218).

Maréchal, R., & Mascherpa, J. (1978). Etude taxonomique d'une groupe d'espèces des genres Phaseolus et Vigna (Papilionaceae) sur la base de données morphologiques et polliniques. *Boissiera*, *28*, 1–273.

Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez, J. G., Buckler, E., & Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(9), 6080–6084. https://doi.org/10.1073/pnas.052125199

Melka, T. S. (2010). The moche lima beans recording system, revisited. *Folklore*, *45*, 89–136. https://doi.org/10.7592/FEJF2010.45.melka

Miller, J. M., & Conn, E. E. (1980). Metabolism of Hydrogen Cyanide by Higher Plants. *Plant Physiology*, *65*(6), 1199–1202. https://doi.org/10.1104/pp.65.6.1199

Motta-Aldana, J. R., Serrano-Serrano, M. L., Hernández-Torres, J., Castillo-Villamizar, G., Debouck, D. G., & Chacón, M. I. S. (2010). Multiple origins of Lima bean landraces in the Americas: Evidence from chloroplast and nuclear DNA polymorphisms. *Crop Science*, *50*(5), 1773–1787. https://doi.org/10.2135/cropsci2009.12.0706

Mueller, S. C., Summers, C. G., & Goodell, P. B. (2003). A field key to the most common Lygus species found in agricultural fields of the central San Joaquin Valley of California. *University of California Division of Agriculture and Natural Resources*, 1–12. Retrieved from http://cottoninfo.ucdavis.edu/files/133229.pdf

Nanni, L., Bitocchi, E., Bellucci, E., Rossi, M., Rau, D., Attene, G., … Papa, R. (2011). Nucleotide diversity of a genomic sequence similar to SHATTERPROOF (PvSHP1) in domesticated and wild common bean (Phaseolus vulgaris L.). *Theoretical and Applied Genetics*, *123*(8), 1341–1357. https://doi.org/10.1007/s00122-011-1671-z

Nwokolo, E. (1996). Lima bean (Phaseolus lunatus L.). In *Food and Feed from Legumes and Oilseeds* (pp. 144–158). Springer US. https://doi.org/10.1007/978-1-4613-0433-3_15

Ornduff, R., & Harlan, J. R. (1977). Crops and Man. *Systematic Botany*, *2*(3), 227. https://doi.org/10.2307/2418272

Piperno, D. R., Moreno, J. E., Iriarte, J., Holst, I., Lachniet, M., Jones, J. G., … Castanzo, R. (2007). Late Pleistocene and Holocene environmental history of the Iguala Valley, Central Balsas Watershed of Mexico. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(29), 11874–11881. https://doi.org/10.1073/pnas.0703442104

Powell, A. L. T., Van Kan, J., Ten Have, A., Visser, J., Greve, L. C., Bennett, A. B., & Labavitch, J. M. (2000). Transgenic expression of pear PGIP in tomato limits fungal colonization. *Molecular Plant-Microbe Interactions*, *13*(9), 942–950. https://doi.org/10.1094/MPMI.2000.13.9.942

Ramírez-Villegas, J., Khoury, C., Jarvis, A., Gabriel Debouck, D., & Guarino, L. (2010). A Gap Analysis Methodology for Collecting Crop Genepools: A Case Study with Phaseolus Beans. *PLoS ONE*, *5*(10). https://doi.org/10.1371/journal.pone.0013497

Ranere, A. J., Piperno, D. R., Holst, I., Dickau, R., & Iriarte, J. (2009). The cultural and chronological context of early Holocene maize and squash domestication in the central balsas river Valley, Mexico. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(13), 5014–5018. https://doi.org/10.1073/pnas.0812590106

Rechcigl, J. E., & Rechcigl, N. A. (2000). Biological and biotechnological control of insect pests. *Choice Reviews Online*, *37*(08), 37-4486-37–4486. https://doi.org/10.5860/choice.37-4486

Rendón-Anaya, M., Montero-Vargas, J. M., Saburido-Álvarez, S., Vlasova, A., Capella-Gutierrez, S., Ordaz-Ortiz, J. J., … Herrera-Estrella, A. (2017). Genomic history of the origin and domestication of common bean unveils its closest sister species. *Genome Biology*, *18*(1), 1–17. https://doi.org/10.1186/s13059-017-1190-6

Ryser, G. L. (1998). *Beans: Prehistoric Indicators of Social Relations and Organization in the Moche Valley, Peru*.

Sanjur, O. I., Piperno, D. R., Andres, T. C., & Wessel-Beaver, L. (2002). Phylogenetic relationships among domesticated and wild species of Cucurbita (Cucurbitaceae) inferred from a mitochondrial gene: Implications for crop plant evolution and areas of origin. *Proceedings*

of the National Academy of Sciences of the United States of America, *99*(1), 535–540. https://doi.org/10.1073/pnas.012577299

Schmutz, J., Jackson, S. A., Wang, M., Wing, R. A., Gao, D., Zhang, M., … Cannon, S. B. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics*, *46*(7), 707–713. https://doi.org/10.1038/ng.3008

Scott, D. R. (1977). An annotated list of host plants for Lygus hesperus Knight. *Bull. Entomol. Soc. Am*, *23*, 19–22.

Serrano-Serrano, M. L., Andueza-Noh, R. H., Martínez-Castillo, J., Debouck, D. G., & Chacón S, M. I. (2012). Evolution and Domestication of Lima Bean in Mexico: Evidence from Ribosomal DNA. *Crop Science*, *52*(4), 1698–1712. https://doi.org/10.2135/cropsci2011.12.0642

Serrano-Serrano, M. L., Hernández-Torres, J., Castillo-Villamizar, G., Debouck, D. G., & Chacón Sánchez, M. I. (2010). Gene pools in wild Lima bean (Phaseolus lunatus L.) from the Americas: Evidences for an Andean origin and past migrations. *Molecular Phylogenetics and Evolution*, *54*(1), 76–87. https://doi.org/10.1016/j.ympev.2009.08.028

Shackel, K. A., Celorio-Mancera, M. de la P., Ahmadi, H., Greve, L. C., Teuber, L. R., Backus, E. A., & Labavitch, J. M. (2005). Micro-injection of lygus salivary gland proteins to simulate feeding damage in alfalfa and cotton flowers. *Archives of Insect Biochemistry and Physiology*, *58*(2), 69–83. https://doi.org/10.1002/arch.20033

Swezey, S. L., Nieto, D. J., & Bryer, J. A. (2007). Control of Western Tarnished Plant Bug Lygus hesperus Knight (Hemiptera: Miridae) in California Organic Strawberries Using Alfalfa Trap Crops and Tractor-Mounted Vacuums. *Environmental Entomology*, *36*(6), 1457–1465. https://doi.org/10.1093/EE/36.6.1457

USDA ERS. (2004). *Economic research service*. *Choice Reviews Online* (Vol. 42). https://doi.org/10.5860/choice.42-2206

USDA NASS. (2019). *USDA/NASS QuickStats*. Retrieved from https://quickstats.nass.usda.gov/results/9979DE70-2420-3A8E-A718-4D3E4C830346

www.ciat.cgiar.org. (2020). CIAT.

www.genesys-pgr.org. (2020). Retrieved January 7, 2021, from https://www.genesys-pgr.org/

Yarnell, S. H. (1965). Cytogenetics of the vegetable crops IV. Legumes (continued). *The Botanical Review*, *31*(3), 247–330. https://doi.org/10.1007/BF02859129

Young, O. P. (1986). Host Plants of the Tarnished Plant Bug, Lygus lineolaris (Heteroptera: Miridae). *Annals of the Entomological Society of America*, *79*(4), 747–762. https://doi.org/10.1093/aesa/79.4.747

Zizumbo-Villarreal, D., & Colunga-GarcíaMarín, P. (2010). Origin of agriculture and plant domestication in West Mesoamerica. *Genetic Resources and Crop Evolution*, *57*(6), 813–825. https://doi.org/10.1007/s10722-009-9521-4

Chapter 1: Developing the first molecular linkage map for lima bean, *Phaseolus lunatus*

Abstract

The absence of genomic resources for lima bean, *Phaseolus lunatus*, research has led to decades of reliance on genomic resources from its closest domesticated relative, common bean, *Phaseolus vulgaris.* However, advances in sequencing technology and the reduction in cost on a per sample and per nucleotide basis have allowed the development of a high-density molecular linkage map of lima bean. Using the genotyping-by-sequencing (GBS) method, a molecular linkage map of lima bean was developed for a recombinant inbred line (RIL) population (n = 234) resulting from the cross of two contrasting California cultivars, UC Haskell and UC 92. In the absence of a lima bean reference genome, this linkage map was initially aligned to the common bean reference genome and, upon later collaborative research, was aligned to the first whole-genome sequence for lima bean. The comparison of genetic distances of markers in this recombinant lima bean population with the physical positions of these markers on the common bean reference genome allowed the detection of genomic areas of inter-species genetic divergence, chromosomal re-arrangements, and recombination rate variation between the two closely related species. This linkage map also allowed the genetic mapping of key agronomic traits in this recombinant inbred line population and was used in the assembly of the first high-quality reference genome sequence of lima bean, included in Phytozome. It also offers the potential to improve current breeding methods and varietal improvement in lima beans.

Introduction

Lima bean is an important warm-season grain legume that is a vital source of protein and nutrition for consumers and a beneficial nitrogen-fixing crop for growers globally (B. Holland et al., 1991; D. B. Jones et al., 1922). Lima bean and common bean are the two most agronomically and economically significant species within the *Phaseolus* genus, both coveted for their highly nutritious seeds that can be eaten fresh or dried, and stored for longer periods of time (Delgado-Salinas et al., 2006; Freytag & Debouck, 2002; Paul Gepts, 2014b). The *Phaseolus* genus contains 70-80 characterized species, including five that have been domesticated in distinct environments of Central and South America: *Phaseolus acutifolius* A. Gray (tepary bean), *Phaseolus coccineus* L. (runner bean), *Phaseolus dumosus* Macfady (year bean), *Phaseolus lunatus* L. (lima bean), and *Phaseolus vulgaris* L. (common bean) (Delgado-Salinas et al., 2006; Freytag & Debouck, 2002; Paul Gepts, 2014b).

Lima bean is rare among agronomic crops, but not alone in the *Phaseolus* genus, in having dual domestications that it shares with its closest domesticated relative, common bean. Common bean and lima bean each had two separate domestication events, one of Mesoamerican and another of Andean origin, which gave rise to two distinct gene pools for each species (Mackie, 1943). Lima bean is categorized into two main gene pools: Andean and Mesoamerican, and recent genetic research has subdivided these main gene pools into two Andean (AI and AII) and two Mesoamerican (MI and MII) gene pools (Chacón-Sánchez & Martínez-Castillo, 2017). The dual domestications and conspecific gene pools putatively minimize the loss of genetic diversity subject to a species under a single domestication event, and lima bean is consistent with common bean in the reduction in genetic diversity in their

Mesoamerican gene pools (Andueza-Noh et al., 2015; Chacón-Sánchez & Martínez-Castillo, 2017; Schmutz et al., 2014). Lima bean also has an expansive distribution range with a broader ecological and climactic range than any of the other domesticated *Phaseolus* species, offering potential traits for adaptation to abiotic and biotic stresses which may be necessary with the threat of climate change (Delgado-Salinas & Gama-López, 2015).

Lima bean and common bean are both autogamous diploid species with 2n = 2x = 22 chromosomes with high levels of homozygosity throughout their genomes. Common bean has a haploid genome size of ~587 Mb. Due to the previous absence of a reference genome for lima bean and a high degree of synteny between the two species, lima bean genetic research deferred to the common bean reference genome (Cícero Almeida & Pedrosa-Harand, 2013; Bonifácio, Fonsêca, Almeida, dos Santos, & Pedrosa-Harand, 2012; Mercado-Ruaro & Delgado-Salinas, 1998; Schmutz et al., 2014). The recent development of a reference genome for lima bean has produced 512 contigs assembling into a 542 Mbp reference genome, and provides a more accurate identification of relevant genomic regions and loci (Garcia et al., 2020). The high macro-collinearity between the two species has been characterized in cytogenetic mapping, and has highlighted three chromosomes with significant chromosomal rearrangments between the two species. These three rearrangments may contribute to the high barrier to gene flow between these two domesticated relatives. The development of the lima bean reference genome would have been difficult without the initial development of a genetic map to identify the cytogenetic differences between the two species reference genomes, and this chapter will present greater details of the development of this genetic map aligned to both reference genomes. These new genetic and genomic resources will ultimately allow for more accurate

QTL mapping and advances in lima bean breeding and development to improve production and quality traits.

The objectives of this study were to (1) develop the first molecular linkage map for lima bean, which establishes the relationship between genetic distance and physical position of loci within this recombinant inbred population, (2) identify recombination rate variation within and across chromosomes and unique regions of conservation with common bean, and (3) highlight chromosomal rearrangements between the common bean and lima bean reference genomes revealed by the two tandem alignments of the genetic maps.

Material and Methods

Mapping population

A biparental recombinant inbred line (RIL) population was developed by a previous graduate student in the Gepts lab (Dohle, 2017) from reciprocal crosses of two polymorphic California Central Valley-adapted cultivars, UC 92 and UC Haskell, belonging to the two domestication pools of lima bean. UC Haskell is a small-seeded, indeterminate vine-type cultivar of Mesoamerican origin and UC 92 is a large-seeded, determinate bush-type cultivar of Andean origin; the development of the RIL population and their distinctive traits for seed size and plant growth habit are shown in Figure 1.1. A total of 238 recombinant inbred lines were established in the F_8 generation following single seed descent with an estimated 99% homozygosity across all lines in this generation (Fehr, 1987). The resulting RIL population demonstrated segregation for many agronomic traits, including germination rate, flowering

time, inflorescence position, plant height, plant habit, pod position, pod density, yield, and biotic stress tolerance (Dohle, 2017).

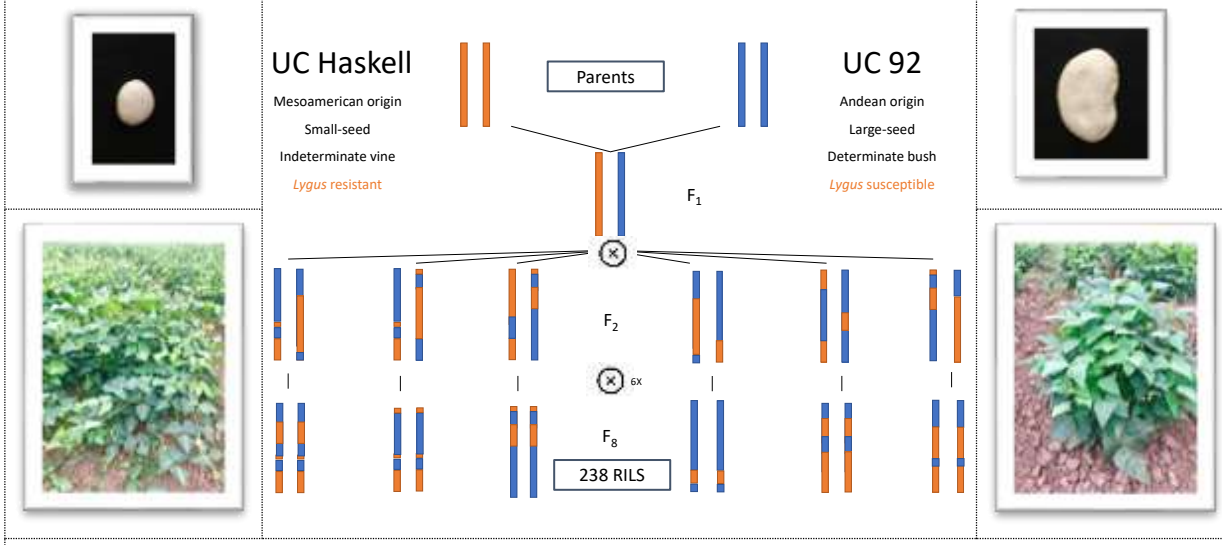

Figure 1.1 Development of the UC 92 x UC Haskell RIL population and the distinctive traits for seed size and plant growth habit.

DNA extraction, GBS digestion, ligation, GBS sequencing

Leaf tissue for DNA extraction was sampled from two-week-old seedlings of 238 RILs in the F_8 generation and the two inbred parents. Leaf tissue samples were collected into 96-well plates from plants grown in a greenhouse, immediately put on ice, and lyophilized for 24 hours. DNA was extracted using an adapted DNA extraction protocol for lima and common bean (Ariani, Berny Mier y Teran, & Gepts, 2016). Leaf tissue was lyophilized for 24 hours and macerated using ceramic beads and a bead-beater. To each well, 600 µl DNA extraction buffer was added, and the plates were incubated for 30 minutes at 65°C. The plates were cooled for 15 minutes at 4°C. To each well, 300 µl of 6M ammonium acetate was added and the plates were incubated for 15 minutes at 4°C. The plates were then centrifuged for 15 minutes at 5,000 *g* and 600 µl of supernatant was transferred to new plates containing 360 µl of isopropanol in

each well. The plates were mixed vigorously and allowed to incubate for 5 minutes at room temperature. The plates were centrifuged again for 15 minutes at 5,000g and the supernatant was discarded. The DNA pellet was washed and dislodged with 400 µl of cold 100% ethanol in each well. The plates were centrifuged again for 15 minutes at 5,000g, the supernatant was discarded, and the remaining pellets were put into the lyophilizer for 5 minutes to dry. The dried pellets were resuspended in 200 µl TE buffer and 0.4 µl RNase A (100 ηg/µl) and allowed to incubate for 5 minutes at room temperature and then 30 minutes at 37°C.

The concentration and quality of DNA was confirmed for each sample by verifying that the 260nm/280nm absorbance ratios were above 1.8 using a NanoDrop Lite (Thermo Fisher Scientific). DNA was quantified for each sample using the Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific), and 100 ηg of DNA from each sample was transferred to a PCR plate. Two genomic libraries were prepared, each containing 144 unique barcode identifiers for the *Cvi*AII restriction enzyme, and a total of 240 unique genetic lines were sequenced (238 recombinant inbred lines and the 2 parental lines). GBS barcode libraries and adapters for the *Cvi*AII restriction enzyme (New England Biolabs) were prepared using a protocol adapted for common bean (Ariani et al., 2016). The restriction enzyme, *Cvi*AII, CutSmart buffer (New England Biolabs) and unique barcode identifiers for each well were added to the plates, spun down, and run on a PCR machine for 2 hours at 25°C. T4 (10x) ligation buffer was added to the wells and the samples were run on the thermal cycler for 1 hour at 22°C, followed by 30 minutes at 65°C. Seven µl from each sample within a genomic library was added to a petri dish, mixed, and transferred to Eppendorf tubes. Binding buffer and isopropanol were added to the tubes, mixed, and 800 µl was transferred to a GeneJET (Thermo Fisher Scientific) purification

column, centrifuged for 60 seconds, and eluted with water. DNA was quantified for each library on the QUBIT dsDNA HS Assay Kit (Thermo Fisher Scientific/Invitrogen) prior to GBS sequencing. The libraries were sequenced using the SR100 protocol on two lanes of an Illumina HiSeq flowcell at the University of California, Davis, Genome Center.

Sequence preprocessing, genome(s) alignment and SNP calling

Sequence data was demultiplexed using the GBSprep Python script (https://github.com/aariani/GBSprep) to remove the CATG remnant site from the restriction enzyme and the common adaptor sequence, in addition to quality trimming the sequence reads. The demultiplex data was processed with the Burrows-Wheeler Aligner (http://bio-bwa.sourceforge.net/) and aligned to the *Phaseolus vulgaris* G19833 v2.0 reference genome, available through the Phytozome website (https://phytozome.jgi.doe.gov/pz/portal.html), and later to the whole-genome sequence of *Phaseolus lunatus*, created by collaborators in Colombia (Garcia et al., 2021). The *Phaseolus vulgaris* reference genome was sequenced from an inbred landrace accession G19833 from the Andean gene pool (-6.27° S. Lat.; -77.75° W. Long.; 1860.0 masl Alt.) (Schmutz et al., 2014). The *Phaseolus lunatus* reference genome was sequenced from an inbred landrace accession of Mesoamerican origin (9.55° N. Lat.; -75.32° W. Long.; 100.0 masl Alt.) from the MI gene pool, G27455 (Garcia et al., 2021). Following alignment, SNP variants were called between the two lima bean parents of the RIL population using Next Generation Sequencing Experience Platform (https://sourceforge.net/p/ngsep/wiki/Home/). Filtering of biparental calls against both genomes was made using VCFtools (https://vcftools.github.io/man_latest.html) to identify SNPs with minimum depth set to 3, minimum genotype quality of 30, and maximum missing

data of 0.99. Filtering of RIL calls was made with a minimum depth set to one, minimum genotype quality of 30, and maximum missing data set to 0.5, which produced data sets including both parental genotypes, 238 RILs and 63,291 SNP markers for the *Phaseolus lunatus* reference genome alignment and 68,205 SNP markers for the *Phaseolus vulgaris* reference genome alignment.

Genetic map development (Phaseolus lunatus alignment)

The linkage map aligned to the *Phaseolus lunatus* reference genome was created from 93 RILs and 10,498 polymorphic SNP markers using the ASMap and R/qtl packages in R statistical software program (Broman, 2009; R Core Team, 2013; J. Taylor & Butler, 2017). Markers with less than 20% missing genotypes were used for map construction, and individuals with more than 50% missing genotypes were removed from map construction and verified for less than 95% matching genotype pairs with other individuals. More details about the number of missing genotypes per marker and individual, as well as the proportion of matching genotype pairs among individuals and the parental allelic distribution are shown in Figure 1.2.

Linkage groups were constructed using the 'mstmap' function based on recombination frequencies of the SNP markers using the 'Kosambi' mapping function and a *p*-value of $1e^{-7}$. Linkage groups were merged when originating from the same chromosome and recombination frequencies were recalculated by chromosome with lower *p*-values. After initial construction, chromosome 1 and 9 were each split into two linkage groups each, and markers on chromosome 8 and 11, and chromosome 5 and 10 were strongly associated on the same linkage groups. After the establishment of the 11 distinct linkage groups, the linkage groups were re-named according to their chromosomal alignment on the *Phaseolus lunatus* reference

genome which is consistent with the chromosomal numbering of the *Phaseolus vulgaris* reference genome. Genotype errors that result in tight double cross-over events were filtered out and the order of markers within each linkage group was established based on recombination frequencies. Individuals with excessive recombination rates were removed, after expected recombination rates were calculated from a Poisson distribution with a mean recombination rate of 44, determined by the number of chromosome arms multiplied by the meiotic events in fixed regions. Markers with high double crossover rates and significant genotype error rates with LOD greater than 3 were removed.

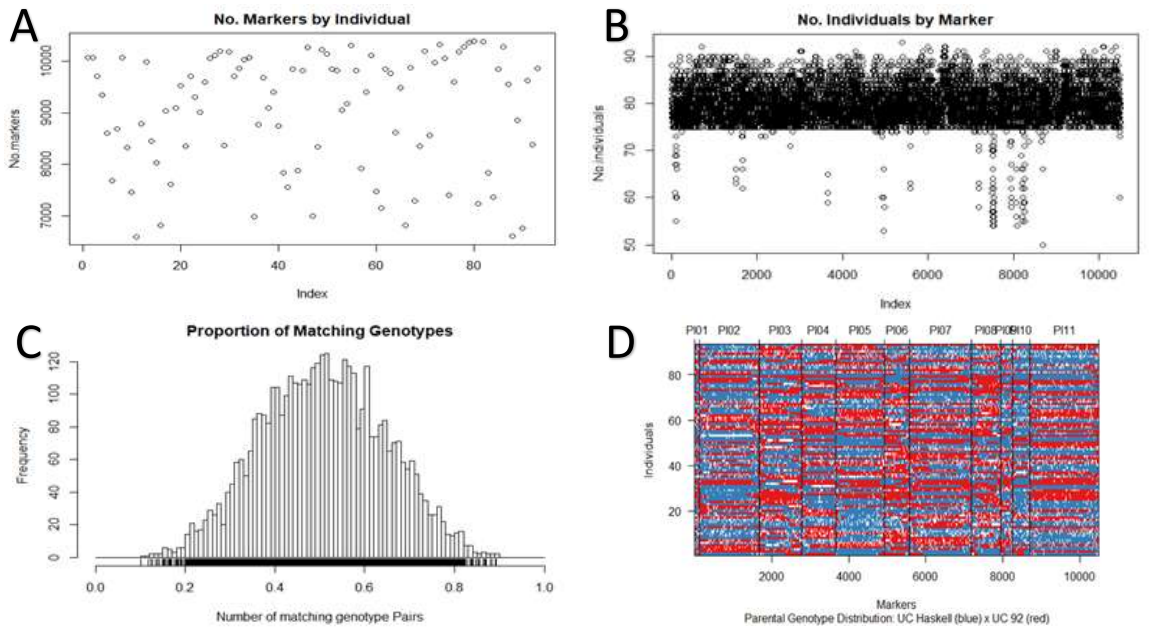

Figure 1.2 Quality control figures of the alignment to the *Phaseolus lunatus* reference genome*:*
A. Number of markers per individual B. Number of individuals per marker C. Distribution frequency of matching genotype pairs for indivudals D. Parental genotype distribution by individuals x markers across linkage groups.

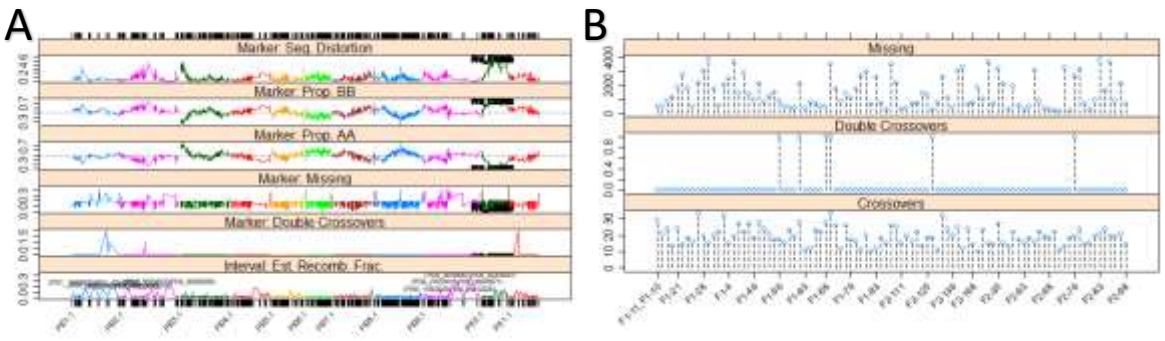

Figure 1.3 Quality control figures of the alignment to the *Phaseolus lunatus* reference genome*:* A. Marker profile by chromosome for segregation distortion, marker proportion for BB (UC Haskell) and AA (UC 92) genotypes, missing markers, marker double crossovers and estimated recombination frequency B. Individual profile for missing markers, double crossovers and single crossovers.

Recombination rates across the genetic map were calculated using the 'MareyMap' function using the 'sliding window' interpolation method every 200 kbp in a 1 Mbp sliding window (Rezvoy et al., 2007). Pericentromeric regions of the chromosomes were defined when recombination rates exceeded 2 Mbp/cM for a given locus, consistent with the standards of pericentromeric definition in the first published common bean reference genome (Schmutz et al., 2014). Recombination rates were calculated separately by chromosome for the pericentromeric regions, within the euchromatic regions of each short arm, within the euchromatic regions of each long arm, and then combined to calculate the collective recombination rate for the euchromatic regions of the chromosome. The recombination rate for each chromosome was calculated from the genetic distance and the physical span of the first and last markers, and the genome-wide recombination rate was calculated by combining the genetic distances and physical span of markers included in the linkage map for all chromosomes.

Genetic map development (Phaseolus vulgaris alignment)

The linkage map aligned to the *Phaseolus vulgaris* reference genome was created from 94 RILs and 11,579 SNP markers using the ASMap package in R statistical software program (Broman, 2009; R Core Team, 2013; J. Taylor & Butler, 2017). Markers with less than 20% missing genotypes were used for map construction, and individuals with more than 50% missing genotypes were removed from map construction and verified for less than 95% matching genotypes with other individuals. More details about the number of missing genotypes per marker and individual, as well as the proportion of matching genotype pairs among individuals and the parental allelic distribution are shown in Figure 1.4.

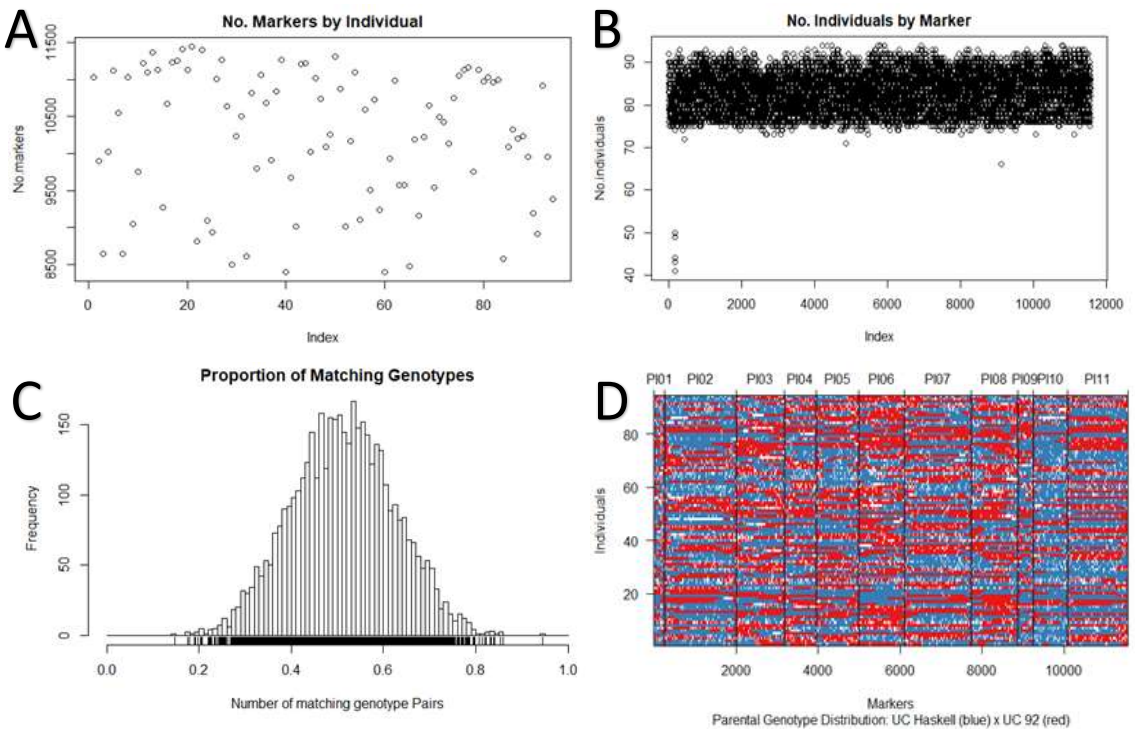

Figure 1.4 Quality control figures of the alignment to the *Phaseolus vulgaris* reference genome: A. Number of markers per individual B. Number of individuals per marker C. Distribution frequency of matching genotype pairs for individuals D. Parental genotype distribution by individuals x markers across linkage groups.

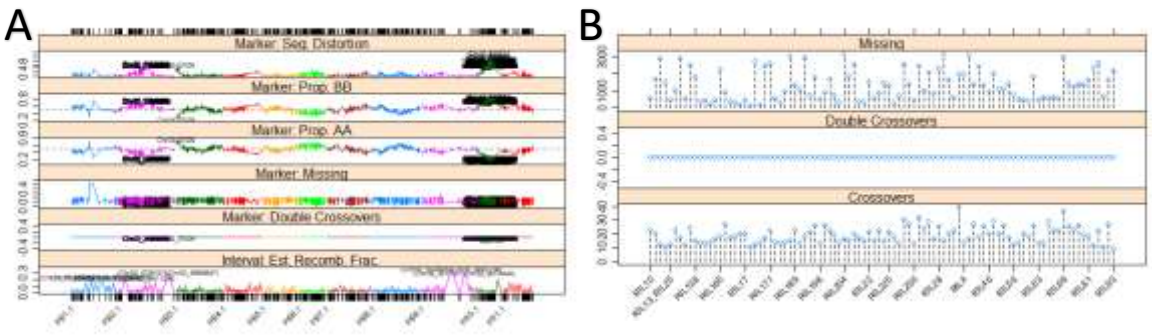

Figure 1.5 Quality control figures of the alignment to the *Phaseolus vulgaris* reference genome: A. Marker profile by chromosome for segregation distortion, marker proportion for BB (UC Haskell) and AA (UC 92) genotypes, missing markers, marker double crossovers and estimated recombination frequency B. Individual profile for missing markers, double crossovers and single crossovers.

Linkage groups were constructed using the 'mstmap' function based on recombination frequencies of the SNP markers using the 'Kosambi' mapping function and a *p*-value of $1e^{-8}$. Linkage groups were merged when originating from the same chromosome and recombination frequencies were recalculated by chromosome with lower *p*-values. After initial construction, the 11 established linkage groups were re-named according to their chromosomal alignment on the *Phaseolus vulgaris* reference genome. Chromosome 1 contained three distinct linkage groups, and chromosomes 2 and 9 each contained two linkage groups, and linkages were associated between chromosome 2, 3,4, 7 and 10. After the establishment of the 11 distinct linkage groups, the linkage groups were re-named according to their chromosomal alignment on the *Phaseolus vulgaris* reference genome. Genotype errors that result in tight double cross-over events were filtered out and the order of markers within each linkage group was established based on recombination frequencies. Individuals with excessive recombination rates were removed, after expected recombination rates were calculated from a Poisson distribution with a mean recombination rate of 44, determined by the number of chromosome

arms multiplied by the meiotic events in fixed regions. Markers with high double crossover rates and significant genotype error rates with LOD greater than 3 were removed.

Recombination rates across the genetic map were calculated using the 'MareyMap' function using the 'sliding window' interpolation method every 200 kbp in a 1 Mbp sliding window (Rezvoy et al., 2007). Pericentromeric regions of the chromosomes were defined when recombination rates exceeded 2 Mbp/cM for a given locus, consistent with the standards of pericentromeric definition in the common bean reference genome (Schmutz et al., 2014). Recombination rates were calculated by chromosome for the pericentromeric regions, and within the euchromatic regions of the short arm, within the euchromatic regions of the long arm, and then combined to calculate the collective recombination rate for the euchromatic regions of the chromosome. The recombination rate for each chromosome was calculated from the genetic distance and the physical span of the first and last markers, and the genome-wide recombination rate was calculated by combining the genetic distances and physical span of markers included in the linkage map for all chromosomes.

Results

Alignment to the Phaseolus lunatus reference genome

A linkage map aligned to the *Phaseolus lunatus reference genome* was developed for the UC 92 – UC Haskell RIL population with 10,497 SNP markers collapsing into 522 non-recombinant loci with an estimated genetic length of 1064 cM. The marker density across genetic distances of the linkage map and physical positions of the reference genome are shown in Figure 1.6. Linkage groups were established for each of the 11 chromosomes. The number of markers, genetic distances, and physical coverage for each chromosome are summarized in Table 1.1, and the complete genetic distance and physical positions are in Appendix 1.1.

Table 1.1 Marker coverage, genetic length, and physical coverage across the chromosomes for the UC 92 – UC Haskell RIL population genetic map aligned to the *Phaseolus lunatus* reference genome.

Chr	No Markers	Loci	First Marker (bp)	Last Marker (bp)	Physical Coverage (Mbp)	Genetic Length (cM)	Max Dist Between Loci (cM)	Avg Dist Between Loci (cM)	Avg Dist Between Loci (kbp)
Pl01	119	27	591,375	47,588,051	47.0	113.3	20.5	4.2	1740.6
Pl02	1554	60	720,985	51,607,776	50.9	136.5	24.4	2.3	848.1
Pl03	1102	68	2,903	45,042,368	45.0	116.9	7.8	1.7	662.3
Pl04	887	45	66,315	49,299,606	49.2	91.4	12.2	2.0	1094.1
Pl05	1263	43	24,078	38,089,793	38.1	74.7	7.5	1.7	885.2
Pl06	661	46	2,763	36,649,143	36.6	60.3	3.5	1.3	796.7
Pl07	1604	56	1,778	47,820,557	47.8	98.4	8.2	1.8	853.9
Pl08	749	73	35,793	57,107,478	57.1	109.8	6.7	1.5	781.8
Pl09	310	43	2,050	40,853,806	40.9	133.7	32.8	3.1	950.0
Pl10	453	26	53,329	53,972,946	53.9	62.0	11.4	2.4	2073.8
Pl11	1795	35	30,471	48,067,523	48.0	66.7	9	1.9	1372.5
Total	10497	522			514.6	1063.8	144.0		
Mean	954.3	47.5	139,258	46,918,095	46.8	96.7	13.1	2.2	1096.3

This linkage map had an average genetic and physical spacing between loci of 2.04 cM and 0.99 Mbp across the genome, respectively. Genetic gaps larger than 20 cM were observed on three linkage groups: Pl01, Pl02 and Pl09, with 20.5, 24.4 and 32.8 cM gaps, respectively. Marker coverage varied across and within linkage groups with the densest marker coverage observed in the pericentromeric regions of Pl02, Pl05, Pl07 and Pl11, and the sparsest marker

coverage observed in the pericentromeric regions of Pl01 and Pl09 and can be observed in Figure 1.8.

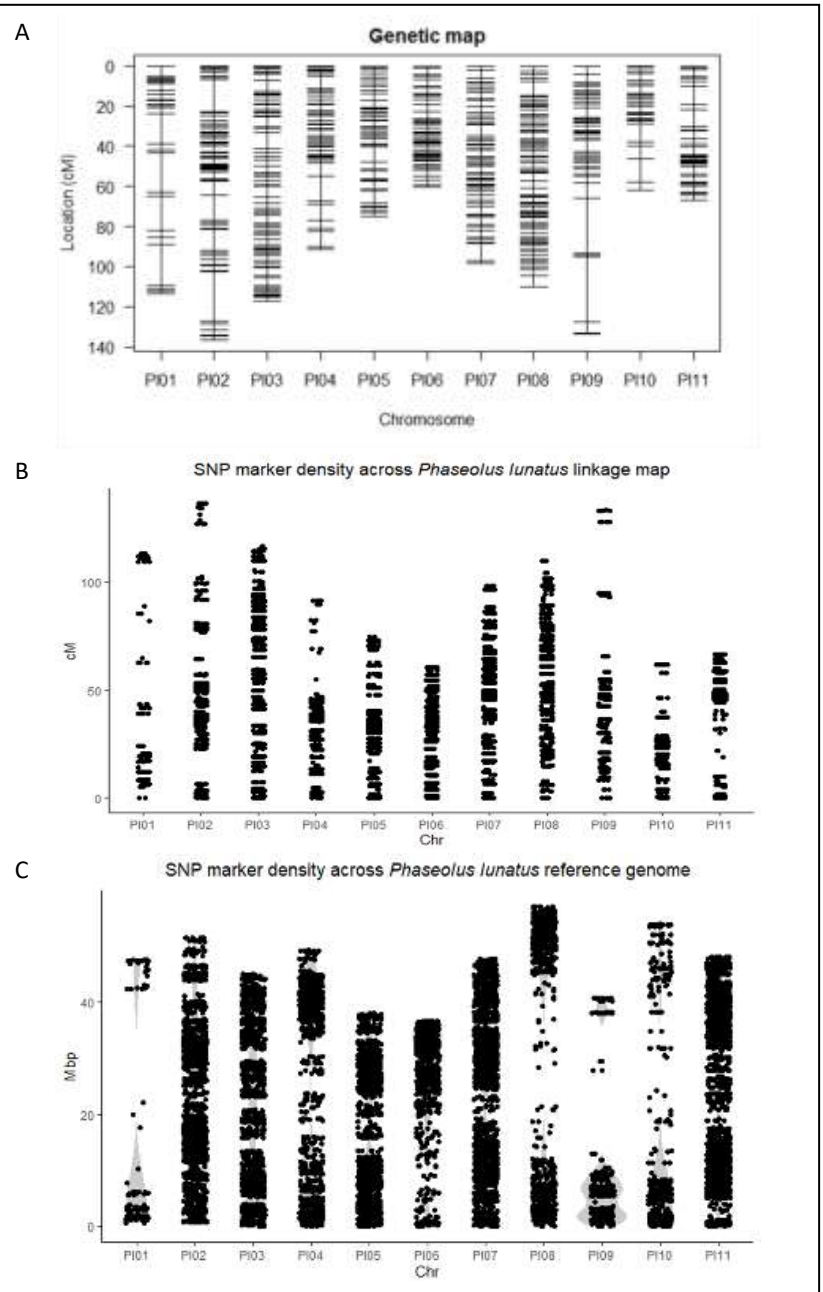

Figure 1.6 Genetic map and marker coverage aligned to the Phaseolus lunatus reference genome: A. Genetic map for the UC 92 - UC Haskell RIL population B. SNP marker density across the genetic map. C. SNP marker density across the reference genome.

Table 1.2 Estimated recombination rates across chromosomes and chromosomal regions and boundaries for pericentromeric regions for the *Phaseolus lunatus* alignment.
*denotes a limited estimation due to minimal marker coverage in this region.

Chr	Chromosome (cM/Mbp)	Euchromatic Arms (cM/Mbp)	Euchromatic Short Arm (cM/Mbp)	Euchromatic Long Arm (cM/Mbp)	Pericentromere (cM/Mbp)	Start of Pericentromere (Mbp)	End of Pericentromere (Mbp)	Pericentromere Length (Mbp)
Pl01	2.41	8.62	7.52	9.81	0.10	7.720	22.11*	14.38*
Pl02	2.68	5.43	5.29	5.51	0.16	9.984	36.417	26.43
Pl03	2.60	3.83	5.23	3.45	0.15	6.030	23.711	17.68
Pl04	1.86	7.08	5.43	9.66	0.09	7.927	44.546	36.62
Pl05	1.96	4.63	4.00	5.35	0.15	8.325	30.802	22.48
Pl06	1.65	3.42	--	3.42	0.20	0.003	19.761	19.76
Pl07	2.06	4.78	5.50	4.21	0.27	8.422	37.151	28.73
Pl08	1.92	4.29	4.48	4.11	0.23	14.141	45.172	31.03
Pl09	3.27	6.24	6.61	5.98	0.56	8.180	29.51*	21.33*
Pl10	1.15	5.43	4.15	6.93	0.10	5.766	49.100	43.33
Pl11	1.39	6.04	8.38	3.60	0.09	5.414	42.758	37.34
Total								299.12
Mean	2.09	5.44	5.66	5.64	0.19			27.19

Recombination rate varied within linkage groups, with the lowest rates of recombination in the centromeric and pericentromeric regions and the highest rates towards the telomeric ends. Estimated recombination rates for each chromosome and the euchromatic and pericentromeric regions are included in Table 1.2 with the limits of the pericentromeric regions for each chromosome, and the estimated recombination frequencies and LOD scores between markers for each linkage group is included in a heatmap in Figure 1.7.

The Pl03 linkage group had the highest average recombination rate, with recombination events occurring every 662 kbp. The Pl10 linkage group had the lowest average recombination rate, with recombination events occurring on average every 2,074 kbp, which may be influenced by the high degree of segregation distortion observed in the pericentromeric region towards the UC Haskell haplotype. The largest spans of the pericentromeric regions were on the Pl04, Pl10 and Pl11 linkage groups, and the shortest spans were on Pl03 and Pl06. The Pl01 and Pl09 linkage groups had particularly sparse marker coverage across the pericentromeric

regions, which likely reduced the accuracy of the recombination rates in these regions and the definition of the pericentromeric regions for these linkage groups.

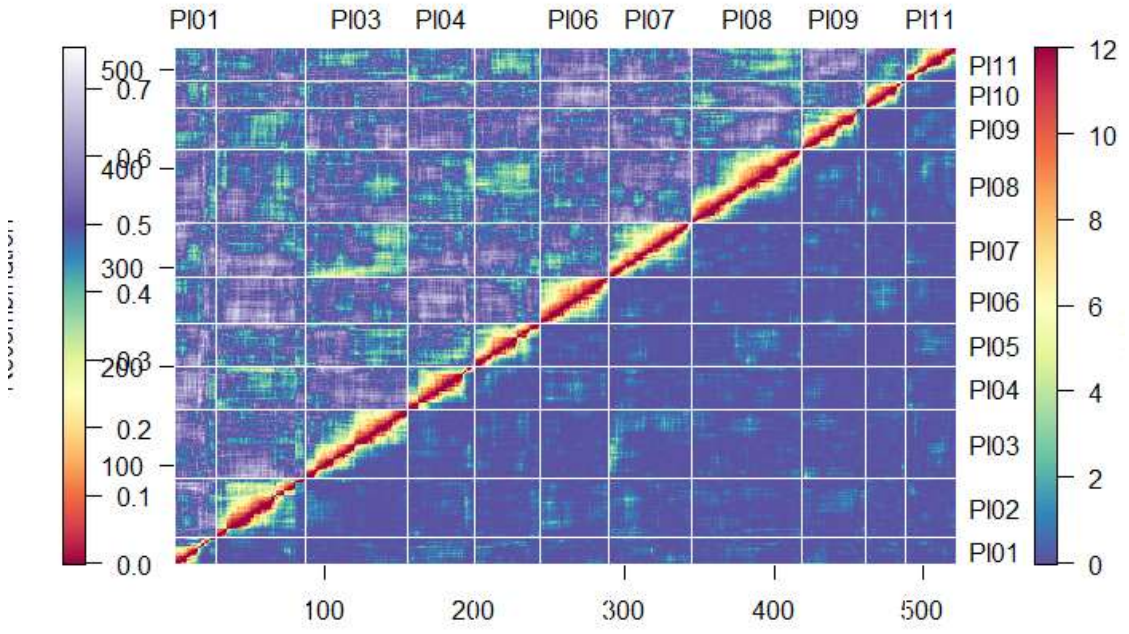

Figure 1.7 Pairwise recombination frequencies and LOD scores between loci on linkage groups aligned to the *Phaseolus lunatus* reference genome.

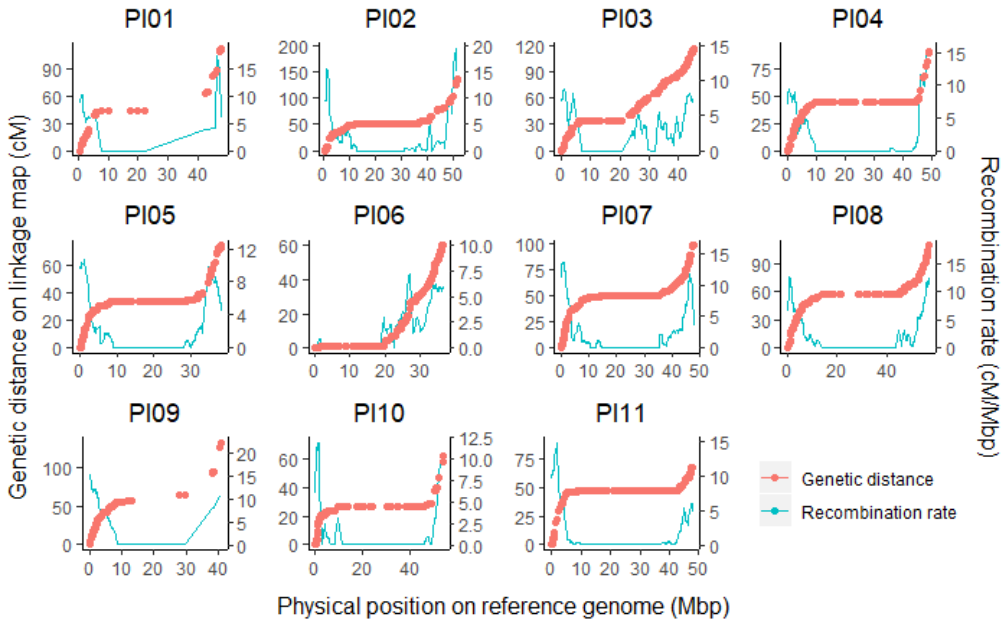

Figure 1.8 Genetic distance (cM) and recombination rate (cM/Mbp) by physical position (Mbp) for the UC 92 – UC Haskell RIL population aligned to the *Phaseolus lunatus* reference genome.

Alignment to the Phaseolus vulgaris reference genome

A linkage map aligned to the *Phaseolus vulgaris* reference genome was developed for the UC 92 – UC Haskell RIL population with 11,560 SNP markers collapsing into 429 non-recombinant loci with an estimated genetic length of 1085 cM. The marker density across genetic distances of the linkage map and physical positions of the reference genome are shown in Figure 1.9. Linkage groups were established for each of the 11 chromosomes. The number of markers, genetic distances, and physical coverage for each chromosome are summarized in Table 1.3.

Table 1.3 Marker coverage, genetic length and physical coverage across the chromosomes for the UC 92 – UC Haskell RIL population genetic map aligned to the *Phaseolus vulgaris* reference genome.

Chr	No Markers	Loci	First Marker (bp)	Last Marker (bp)	Physical Coverage (Mbp)	Genetic Length (cM)	Max Dist Between Loci (cM)	Avg Dist Between Loci (cM)	Avg Dist Between Loci (kbp)
Pl01	260	24	14,879	51,433,871	51.4	111.3	16.7	4.6	2142.5
Pl02	1743	43	12,353	49,668,150	49.7	137.5	30.3	3.2	1154.8
Pl03	1171	48	12,844	53,422,222	53.4	112.2	6.7	2.3	1112.7
Pl04	780	36	21,321	47,926,300	47.9	94.2	9.6	2.6	1330.7
Pl05	1035	44	31,868	40,909,946	40.9	82.8	7.4	1.9	929.0
Pl06	1126	47	31,462	31,216,289	31.2	62.4	3.4	1.3	663.5
Pl07	1624	49	1,023	40,036,525	40.0	108.4	10.4	2.2	817.1
Pl08	1140	54	4,193	63,047,315	63.0	115.6	6.0	2.1	1167.5
Pl09	385	29	378,670	38,238,091	37.9	128.5	36.3	4.4	1305.5
Pl10	832	38	36,326	44,268,009	44.2	61.3	14.8	1.6	1164.0
Pl11	1483	29	7,594	53,572,347	53.6	71.6	10.4	2.5	1847.1
Total	11579	441			513.2	1085.8	152.0		
Mean	1052.6	40.1	50,230	46,703,551	85.5	98.7	13.8	2.6	1239.5

This linkage map had an average genetic and physical spacing between loci of 2.46 cM and 1.16 Mbp across the genome, respectively. Genetic gaps larger than 20 cM were observed on two linkage groups: Pl02 and Pl09, with 30.3 and 36.3 cM gaps, respectively. Marker coverage varied across and within linkage groups with the densest marker coverage observed in the pericentromeric regions of Pl02, Pl05, Pl07 and Pl11, and the sparsest marker coverage observed in the pericentromeric regions of Pl01, Pl08 and Pl09.

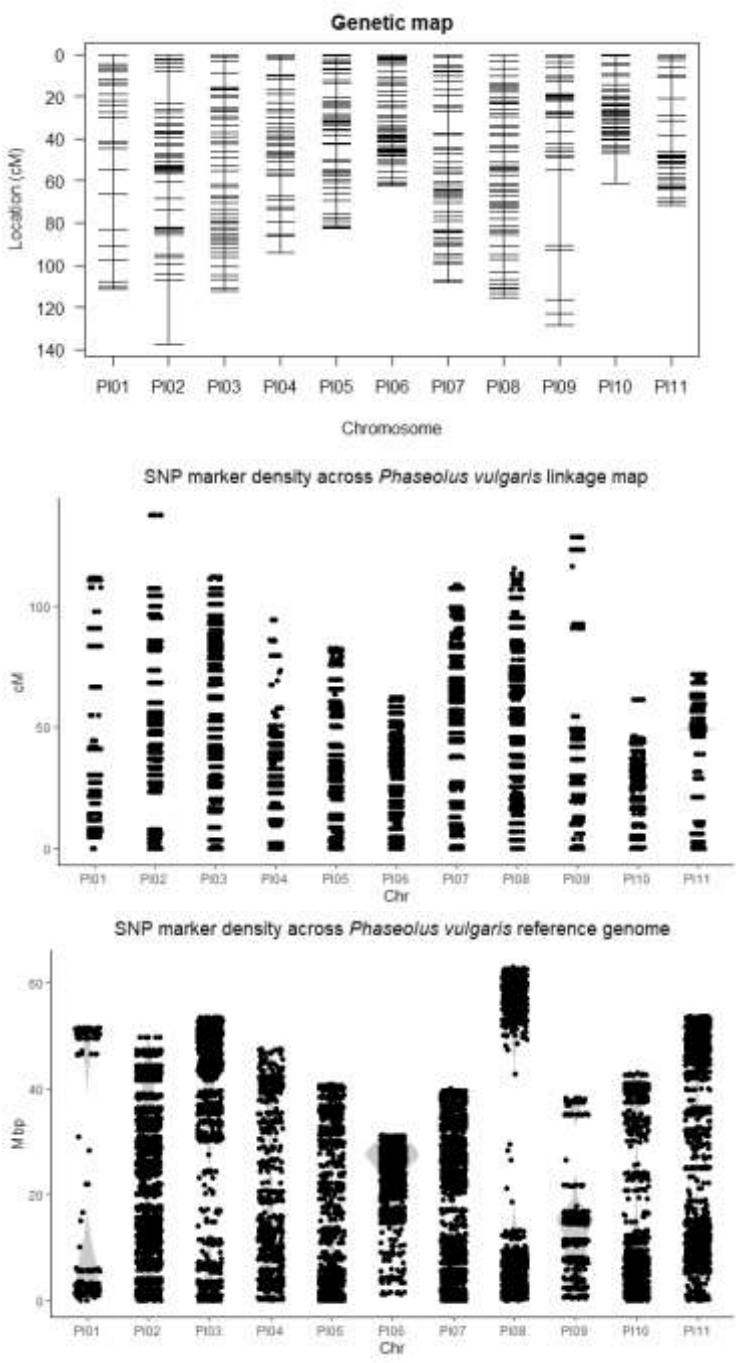

Figure 1.9 Genetic map and marker coverage aligned to the *Phaseolus vulgaris* reference genome*:* A. Genetic map for the UC 92 - UC Haskell RIL population B. SNP marker density across the genetic map. C. SNP marker density across the reference genome.

Table 1.4 Estimated recombination rates across chromosomes and chromosomal regions and boundaries for pericentromeric regions for the *Phaseolus vulgaris* alignment.
*denotes a limited estimation due to minimal marker coverage.

Chr	Chromosome (cM/Mbp)	Euchromatic Arms (cM/Mbp)	Euchromatic Short Arm (cM/Mbp)	Euchromatic Long Arm (cM/Mbp)	Pericentromere (cM/Mbp)	Start of Pericentromere (Mbp)	End of Pericentromere (Mbp)	Pericentromere Length (Mbp)
Pl01	2.18	8.19	7.55	8.91	0.51	6.102	30.91*	24.81
Pl02	2.77	5.39	5.30	5.44	0.20	0.97-0.03/5.82-22.11	/32.31-34.43	18.40
Pl03	2.10	3.22	3.27	3.20	0.35	10.455	31.074	20.62
Pl04	1.98	5.16	3.62	9.10	0.09	12.886	42.397	29.51
Pl05	1.99	5.72	4.66	6.83	0.14	6.845	34.131	27.29
Pl06	2.09	3.88	--	3.88	0.15	1.281	15.504	14.22
Pl07	2.72	4.37	5.43	3.58	0.20	9.88-24.36/	/34.17-32.23	16.41
Pl08	1.84	4.95	5.74	4.28	0.24	10.099	50.646	40.55
Pl09	3.40	4.19	3.34	4.79	0.66	16.67-26.46/*	/6.74-10.23*	13.28*
Pl10	1.82	4.38	3.73	5.52	0.14	0-0.86/9.50	38.146	29.50
Pl11	1.32	5.29	7.46	3.19	0.08	6.361	47.013	40.65
Total								275.24
Mean	2.14	4.98	5.01	5.34	0.25			25.02

Recombination rates varied within linkage groups, with the lowest rates of recombination in the centromeric and pericentromeric regions and the highest rates towards the telomeric ends. Estimated recombination rates for each chromosome and the euchromatic and pericentromeric regions are included in Table 1.4 with the limits of the pericentromeric regions for each chromosome. The estimated recombination frequencies and LOD scores between markers for each linkage group is included in a heatmap in Figure 1.10.

The Pl09 and Pl02 linkage groups had the greatest average recombination rates, with recombination events occurring every 294 and 361 kbp, respectively. The Pl11 and Pl10 linkage groups had the lowest average recombination rates, with recombination events occurring on average every 758 and 549 kbp, respectively. The largest spans of the pericentromeric regions were on the Pl04, Pl10 and Pl11 linkage groups, and the shortest spans were on Pl03 and Pl06 (Table 1.4). The Pl01 and Pl09 linkage groups had particularly sparse marker coverage across the pericentromeric regions, which likely reduced the accuracy of the recombination rates in these regions and the definition of the pericentromeric regions for these linkage groups and

can be observed in Figure 1.11.

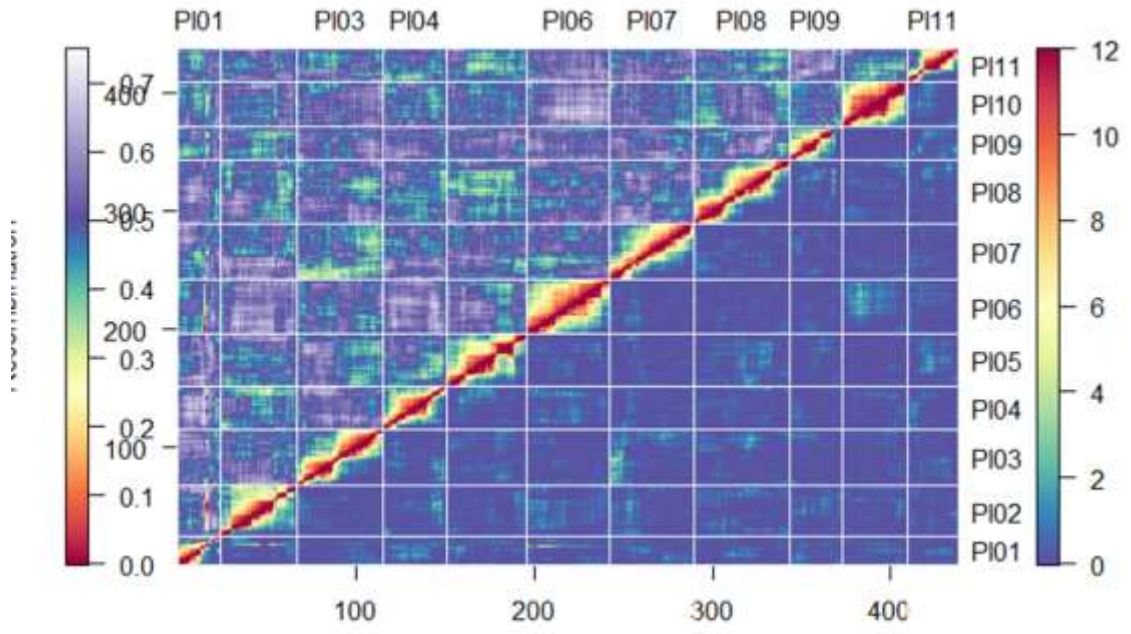

Figure 1.10 Pairwise recombination frequencies and LOD scores between loci on linkage groups aligned to the *Phaseolus vulgaris* reference genome.

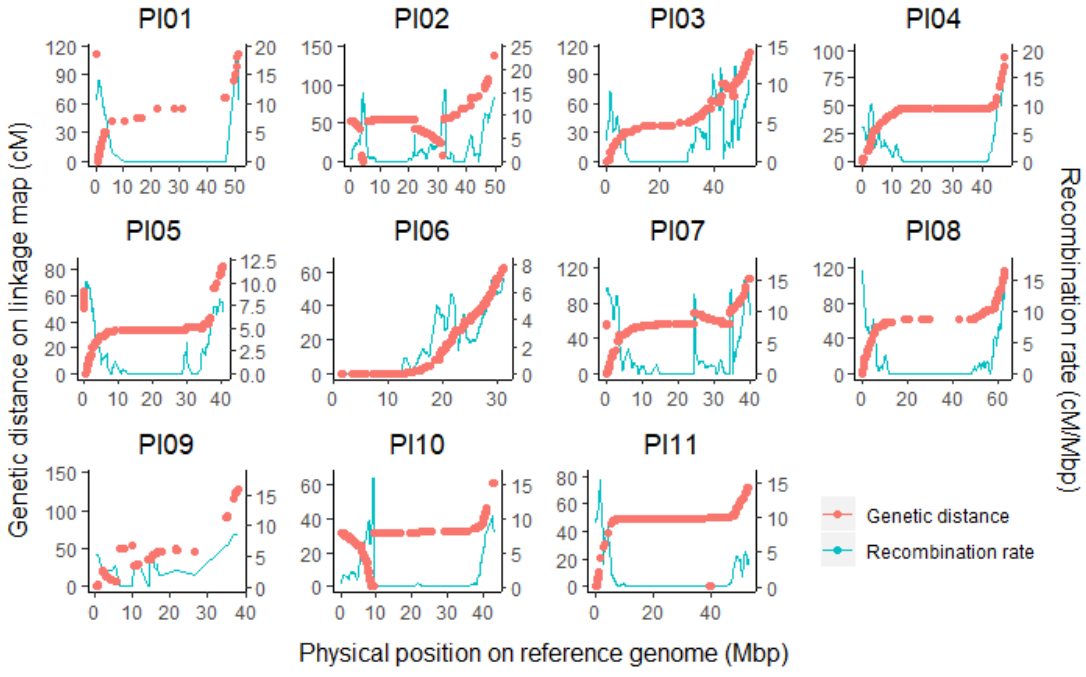

Figure 1.11 Genetic distance (cM) and recombination rate (cM/Mbp) by physical position (Mbp) for the UC 92 – UC Haskell RIL population aligned to the *Phaseolus vulgaris* reference genome.

Throughout the 11 chromosomes (2x = 2n =22) shared between common bean and lima bean, all detected structural rearrangements including inversions, translocations, insertions and duplications are detailed in Table 1.5; this includes nine rearrangements affecting regions larger than 2 Mbp between the common bean genome and this lima bean recombinant inbred line population. The most substantial structural events identified in this study are a complete inversion of the short arm of chromosome Pl10 and a large 10 Mbp intra-chromosomal translocation of the pericentromeric region of Pv02 to the short arm of chromosome Pl02. Additional interspecific rearrangements include the inversion of the majority of the short arm of Pl02 (including the translocation region), two inversions on the long arm of chromosome Pl03, a 10Mbp inversion within the long arm of chromosome Pl07, a complex translocation within an inversion on the short arm of chromosome Pl09, a nearly 3 Mb insertion in the pericentromeric region of Pl07, and a nearly 800 kb insertion on the long arm of Pl05. A 167 kb duplication from the long arm of chromosome Pv07 was detected at the beginning of the short arm of chromosome Pl11.

Discussion

Genetic map development

The comparison of the two genetic maps confirms the collinearity and strong degree of synteny between the common bean and lima bean genomes reported previously (Cícero Almeida & Pedrosa-Harand, 2013; Bonifácio et al., 2012). Two genetic maps were developed based on marker positions in the two reference genomes, which do not have accurate collinearity established between the two genomes resulting in slight deviations of the established maps. The two genetic maps were developed using the same quality control

parameters and highlight the differences between the alignments to the reference genomes of the two species. The common bean and lima bean reference genomes were both sequenced from domesticated landraces; however, they differ in the domesticated gene pools they originate from, the environmental and ecological characteristics of their collection locations, and other key botanical traits that could influence read alignment with this specific RIL population. Differences in read alignments attributed to the differences in the gene pools of the sample genomes sequenced could influence the markers selected for contribution to the genetic map, particularly for significant deletion regions or rearrangements in either of the reference genomes. The interspecific structural rearrangements are highlighted in the alignment of the lima bean genetic map to the common bean reference genome, whereas there were no intraspecific structural rearrangements identified in the alignment of the lima bean genetic map to the lima bean sample genome. This suggests minimal intraspecific genetic distance between the lima bean sample genome and the UC 92 - UC Haskell biparental RIL population, in spite of the divergence of the two parents of the RIL population, which belong to the Andean and Mesoamerican gene pools, and the landrace sequenced belonging to the Mesoamerican gene pool.

The pericentromeric region of chromosome 1 for the UC 92 – UC Haskell genetic map has the lowest marker coverage both within lima bean and between lima bean and common bean. This large genetic gap in marker coverage across this pericentromeric region may influence the estimated recombination rate for this region and the accuracy of the estimated range of the pericentromeric region.

Sequence data was obtained across this region for both parents, suggesting there is no major deletion in this pericentromeric region on Pl01; however, the read depth declined in this region for both parents and across the recombinant inbred lines. This region contains the lima bean homolog of the Arabidopsis *TFL1* gene (*PlTFL1*) at 41 Mbp of Pl01 and the common bean homolog (*PvTFL1y*) at 45 Mbp of Pv01. These homologs control the expression of the determinate growth habit, as well as loci involved in flowering time and stress induced enzymatic inhibition and protein production (Garcia et al., 2021; Pañeda et al., 2008; Repinski, Kwak, & Gepts, 2012). These traits may be highly conserved in domesticated landraces and cultivars for both species and result in the conservation of the pericentromeric region of Pv01 and Pl01.

Across the lima bean genome, the average genetic distance between loci was 2.2 cM and the range across chromosomes varied between 1.3 – 4.2 cM. The genetic map contained few gaps greater than 10 cM across the genome, which generally occurred across pericentromeric regions of chromosomes. The physical coverage of this recombinant inbred line populations genetic map to the lima bean reference genome is high, estimated at 95%. The marker coverage of this genetic map across the lima bean genome and the population size of this recombinant inbred line population offer enough statistical power for accurate QTL mapping and estimations of QTL effects for this population (H. Li et al., 2010; Piepho, 2000).

Chromosome morphology

Overall, there is strong karyotype stability between these two domesticated *Phaseolus* species, with all chromosomes exhibiting primarily metacentric-submetacentric morphology excluding the acrocentric nature of chromosome 6 for both species and the submetacentric-

subtelocentric nature of chromosome 3 in lima bean. Lima bean exhibits greater metacentric morphology among its chromosomes compared to common bean, excluding chromosomes 6 and chromosome 3 (Bonifácio et al., 2012). Slight variations exist within both species in the size and classification of the structure of chromosome 6; previous research has characterized chromosome 6 as being metacentric in the common bean line 'Saxa' and the lima bean line 'Vermelhinha', whereas it is characterized as being acrocentric in the common bean line 'BAT93' and in the UC 92 - UC Haskell lima bean population (Bonifácio et al., 2012; Fonsêca et al., 2010; Moscone, Klein, Lambrou, Fuchs, & Schweizer, 1999). Chromosome 6 is consistent with being classified as acrocentric and having the smallest physical length in the common bean lines 'BAT93' and G19833 and the lima bean line G27455. Chromosome 6 also has the smallest genetic distance in the Stampede – Red Hawk common bean RIL population and the UC 92 – UC Haskell lima bean RIL population (Fonsêca et al., 2010; Garcia et al., 2021; Schmutz et al., 2014). The long arm of chromosomes 3 in lima bean has the longest physical length of any chromosome arms resulting in the classification of this chromosome as submetacentric bordering subtelocentric in morphology, varying from common beans metacentric-submetacentric characterization of this chromosome (Cícero Almeida & Pedrosa-Harand, 2013; Bonifácio et al., 2012; Fonsêca et al., 2010). The long arms of chromosomes 3 and 6 have the longest physical length and also include regions within those arms that have mild declines in recombination, which may be considered recombination "cold-spots" in these euchromatic regions.

Recombination rate variation

Recombinant rate estimations for lima bean show a reduced level in recombination across the pericentromeric regions and an increase when approaching the subtelomeric chromosomal ends. Chromosome 10 has the lowest recombination rate calculated across its entire chromosome, particularly in the pericentromeric region. This decline in recombination rate on chromosome 10 in this RIL population may be influenced by the high degree of segregation distortion observed in the pericentromeric region towards the UC Haskell haplotype, which can be seen in Figures 1.3 and 1.5. This deviation from the expected chi-square ratio of 1:1 has been observed in other genetic maps of related species. While the underlying reason(s) for this segregation distortion is still unknown, there may be strong conservation of loci of interest in this chromosomal region.

Lima bean is consistent with common bean in having extended suppression of recombination around the centromeres which results in large pericentromeric regions (Pedrosa-Harand, Kami, Gepts, Geffroy, & Schweizer, 2009). Lima bean has between 30-80% of the physical length of their chromosomes occupied by pericentromeric regions, but much smaller fractions of their genetic distances, 2-9%, occupied by pericentromeric regions across their chromosomes. Chromosomes 4, 10 and 11 have the largest physical spans (36.63 Mbp, 43.33 Mbp, 37.34 Mbp, respectively) and the largest percentages (74%, 80% and 78%, respectively) of their physical length occupied by pericentromeric regions, and the lowest recombination rates across their pericentromeric regions (0.09 cM/Mbp, 0.10 cM/Mbp, 0.09 cM/Mbp). Chromosome 3 is consistent with prior cytogenetic mapping in lima bean in having the shortest pericentromeric length and the smallest percentage of the physical chromosome

(Bonifácio et al., 2012), and additionally has the smallest genetic distance and smallest percentage of the genetic distance of the chromosome occupied by pericentromeric regions[5].

Plant species with smaller genomes, including *Arabidopsis* and rice, maintain uniform recombination across their chromosomes with limited suppression extending beyond a few Mbp of their centromeres, whereas species with larger genomes tend to have larger regions of recombination suppression around their centromeres, including wheat, potato and tomato (Cheng et al., 2001; Gill et al., 1996; Iovene et al., 2008; Schmidt et al., 1995; Tang et al., 2008). Common bean and lima bean are unique among these other species in having relatively small genomes (~550 Mbp) and larger portions of their chromosomes occupied by recombination suppressing pericentromeric regions (Garcia et al., 2021; Pedrosa-Harand et al., 2009; Schmutz et al., 2014).

Due to chromosomal rearrangements on the genetic map alignment to the common bean reference genome, convergent regions leading up to chromosomal rearrangements were omitted from the recombination rate calculations and may influence the calculated recombination rates in these regions and across the chromosomes. Overall, the recombination rates and estimated parameters of the pericentromeric regions are more accurately estimated on the genetic map alignment to the lima bean reference due to the absence of observed chromosomal rearrangements.

[5] Excluding the limited estimate of the pericentromeric region of chromosome 1, given the lack of markers across this region on chromosome 1.

Interspecific structural rearrangements

Structural rearrangements were identified throughout the 11 chromosomes shared between common bean and lima bean, including nine rearrangements affecting regions larger than 2 Mbp, in this recombinant inbred line population. Deletions in the lima bean reference genome are more difficult to confirm and distinguish from gaps in marker coverage. Some of these rearrangements could contribute to previously observed barriers to reproductive gene flow and hybridization between lima bean and common bean (Kuboyama et al., 1991; Leonard, Stephens, & Summers, 1987). Two of the largest and most significant structural events are the complete inversion of the short arm of chromosome Pl10 and a large 10 Mbp intra-chromosomal translocation of the pericentromeric region of Pv02 to the short arm of chromosome Pl02. These two rearrangements are confirmed and provide clarity to previous work suggesting pericentric inversions in these chromosomes based on fluorescence *in situ* hybridization (FISH) assays (C. Almeida & Pedrosa-Harand, 2011; Bonifácio et al., 2012). The duplication of a region on the long arm of chromosome Pv07 detected on the short arm of chromosome Pl11 confirms previous molecular work suggesting there is a duplication on the common bean genome on chromosomes Pv07 and Pv11 that is only observed on chromosome Pl11 in lima bean (Bonifácio et al., 2012).

Conclusion

This work presents the first molecular linkage map for lima bean. The development of this linkage map benefited from the alignment of markers against the novel lima bean reference genome and the pre-existing reference genome of lima bean's closest domesticated relative, common bean (Garcia et al., 2021; Schmutz et al., 2014). This work provides insight into the structural rearrangements between the two *Phaseolus* species and allows for the identification of significant agronomic loci in lima bean. The establishment of the relationship between the genetic distance and physical position of loci in the lima bean genome is critical for the mapping of QTL and advancement of breeding efforts for high heritability traits for this crop. Further research should continue to develop different lima bean populations to understand any intraspecific rearrangements between populations, and particularly between different gene pools or domestication statuses. Developing more populations of lima bean can improve the accuracy of mapping genetic loci for different populations and traits of interest, which will continue to be important for broadening the genetic diversity incorporated into breeding programs for crop improvement.

References

Almeida, C., & Pedrosa-Harand, A. (2011). Contrasting rDNA Evolution in Lima Bean (Phaseolus lunatus L.) and Common Bean (P. vulgaris L., Fabaceae). *Cytogenetic and Genome Research*, *132*(3), 212–217. https://doi.org/10.1159/000321677

Almeida, Cícero, & Pedrosa-Harand, A. (2013). High macro-collinearity between lima bean (Phaseolus lunatus L.) and the common bean (P. vulgaris L.) as revealed by comparative cytogenetic mapping. *Theoretical and Applied Genetics*, *126*(7), 1909–1916. https://doi.org/10.1007/s00122-013-2106-9

Andueza-Noh, R. H., Martínez-Castillo, J., & Chacón-Sánchez, M. I. (2015). Domestication of small-seeded lima bean (Phaseolus lunatus L.) landraces in Mesoamerica: evidence from microsatellite markers. *Genetica*, *143*(6), 657–669. https://doi.org/10.1007/s10709-015-9863-0

Ariani, A., Berny Mier y Teran, J. C., & Gepts, P. (2016). Genome-wide identification of SNPs and copy number variation in common bean (Phaseolus vulgaris L.) using genotyping-by-sequencing (GBS). *Molecular Breeding*, *36*(7), 1–11. https://doi.org/10.1007/s11032-016-0512-9

Bonifácio, E. M., Fonsêca, A., Almeida, C., dos Santos, K. G. B., & Pedrosa-Harand, A. (2012). Comparative cytogenetic mapping between the lima bean (Phaseolus lunatus L.) and the common bean (P. vulgaris L.). *Theoretical and Applied Genetics*, *124*(8), 1513–1520. https://doi.org/10.1007/s00122-012-1806-x

Broman, K. W. (2009). *Statistics for Biology and Health A Guide to QTL Mapping with R / qtl*. https://doi.org/10.1007/978-0-387-92125-9

Chacón-Sánchez, M. I., & Martínez-Castillo, J. (2017). Testing domestication scenarios of Lima bean (Phaseolus lunatus L.) in mesoamerica: Insights from genome-wide genetic markers. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.01551

Cheng, Z., Presting, G. G., Buell, C. R., Wing, R. A., & Jiang, J. (2001). High-Resolution Pachytene Chromosome Mapping of Bacterial Artificial Chromosomes Anchored by Genetic Markers Reveals the Centromere Location and the Distribution of Genetic Recombination Along Chromosome 10 of Rice. *Genetics Soc America*. Retrieved from www.genome.

Delgado-Salinas, A. O., Bibler, R., & Lavin, M. (2006). *Phylogeny of the Genus Phaseolus (Leguminosae): A Recent Diversification in an Ancient Landscape*. *Source: Systematic Botany* (Vol. 31).

Delgado-Salinas, A. O., & Gama-López, S. (2015). Diversity and distribution of wild beans in Mexico. *Revista Digital Universitaria*, *16*(2), 1607–6079. Retrieved from http://ru.tic.unam.mx/handle/123456789/2292

Dohle, S. (2017). *Development of Resources for Lima Bean (Phaseolus lunatus) Breeding and Genetics Research*. University of California, Davis.

Fehr, W. (1987). *Principles of Cultivar Development: Crop Species*. *Agronomy Books*. Retrieved from https://lib.dr.iastate.edu/agron_books/2

Fonsêca, A., Ferreira, J., Dos Santos, T. R. B., Mosiolek, M., Bellucci, E., Kami, J., … Pedrosa-Harand, A. (2010). Cytogenetic map of common bean (Phaseolus vulgaris L.). *Chromosome Research*, *18*(4), 487–502. https://doi.org/10.1007/s10577-010-9129-8

Freytag, G. F., & Debouck, D. G. (2002). Taxonomy, Distribution, and Ecology of the Genus Phaseolus (Leguminosae–Papilionoideae) in North America, Mexico, and Central America. *Botanical Research Institute of Texas, Forth Worth*. https://doi.org/10.1663/0013-0001(2003)057[0421:bredfa]2.0.co;2

Garcia, T., Duitama, J., Zullo, S. S., Gil, J., Ariani, A., Dohle, S., … Chacón-Sánchez, M. I. (2021). Comprehensive genomic resources related to domestication and crop improvement traits in Lima bean. *Nature Communications*, *12*(1), 1–17. https://doi.org/10.1038/s41467-021-20921-1

Gepts, P. (2014). The contribution of genetic and genomic approaches to plant domestication studies. *Current Opinion in Plant Biology*, *18*(1), 51–59. https://doi.org/10.1016/j.pbi.2014.02.001

Gill, K. S., Gill, B. S., Endot, T. R., & Boyko, E. V. (1996). *Identification and High-Density Mapping of Gene-Rich Regions in Chromosome Group 5 of Wheat*. *Genetics* (Vol. 143). Retrieved from https://www.genetics.org/content/143/2/1001.short

Holland, B., Widdowson, E. M., Unwin, I. D., & Buss, D. H. (1991). *Vegetables, Herbs and Spices: Fifth Supplement to McCance and Widdowson's The Composition of Foods*. *Vegetables, Herbs and Spices: Fifth Supplement to McCance and Widdowson's The Composition of Foods*. Retrieved from https://books.google.com/books?hl=en&lr=&id=-M1khtZ31BwC&oi=fnd&pg=PA7&dq=holland+lima+beans+1991+unwin&ots=TcSyx1myu_&sig=itr-6lk4sqMWKzALCk2EeHpOkjM

Iovene, M., Wielgus, S. M., Simon, P. W., Buell, C. R., & Jiang, J. (2008). Chromatin Structure and Physical Mapping of Chromosome 6 of Potato and Comparative Analyses With Tomato. *Genetics Soc America*. https://doi.org/10.1534/genetics.108.093179

Jones, D. B., Gersdorff, C. E. F., Johns, C. O., & Finks, A. J. (1922). The proteins of the lima bean, Phaseolus lunatus. *Journal of Biological Chemistry*, *53*(2), 231–240. https://doi.org/10.1016/s0021-9258(18)85778-7

Kuboyama, T., Shintaku, Y., & Takeda, G. (1991). Hybrid plant of Phaseolus vulgaris L. and P. lunatus L. obtained by means of embryo rescue and confirmed by restriction endonuclease analysis of rDNA. *Euphytica*, *54*(2), 177–182. https://doi.org/10.1007/BF00039605

Leonard, M. F., Stephens, L. C., & Summers, W. L. (1987). Effect of maternal genotype on development of Phaseolus vulgaris L. ×P. lunatus L. interspecific hybrid embryos. *Euphytica*, *36*(1), 327–332. https://doi.org/10.1007/BF00730679

Li, H., Hearne, S., Bänziger, M., Li, Z., & Wang, J. (2010). Statistical properties of QTL linkage

mapping in biparental genetic populations. *Heredity*, *105*(3), 257–267. https://doi.org/10.1038/hdy.2010.56

Mackie, W. W. (1943). Origin, dispersal, and variability of the lima bean, Phaseolus lunatus. *Hilgardia*, *15*(1), 1–29. https://doi.org/10.3733/hilg.v15n01p001

Mercado-Ruaro, P., & Delgado-Salinas, A. (1998). Karyotypic studies on species of Phaseolus (Fabaceae: Phaseolinae). *American Journal of Botany*, *85*(1), 1–9. https://doi.org/10.2307/2446547

Moscone, E. A., Klein, F., Lambrou, M., Fuchs, J., & Schweizer, D. (1999). Quantitative karyotyping and dual-color FISH mapping of 5S and 18S-25S rDNA probes in the cultivated Phaseolus species (Leguminosae). *Genome*, *42*(6), 1224–1233. https://doi.org/10.1139/g99-070

Pañeda, A., Rodríguez-Suárez, C., Campa, A., Juan, ·, Ferreira, J., & Ramón Giraldez, ·. (2008). Molecular markers linked to the fin gene controlling determinate growth habit in common bean. *Euphytica*, *162*, 241–248. https://doi.org/10.1007/s10681-007-9596-y

Pedrosa-Harand, A., Kami, J., Gepts, P., Geffroy, V., & Schweizer, D. (2009). Cytogenetic mapping of common bean chromosomes reveals a less compartmentalized small-genome plant species. *Chromosome Research*, *17*(3), 405–417. https://doi.org/10.1007/s10577-009-9031-4

Piepho, H. P. (2000). Optimal marker density for interval mapping in a backcross population. *Heredity*, *84*(4), 437–440. https://doi.org/10.1046/j.1365-2540.2000.00678.x

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundatrion for Statistical Computing. Retrieved from http://www.r-project.org/

Repinski, S. L., Kwak, M., & Gepts, P. (2012). The common bean growth habit gene PvTFL1y is a functional homolog of Arabidopsis TFL1. *Theoretical and Applied Genetics*, *124*(8), 1539–1547. https://doi.org/10.1007/s00122-012-1808-8

Rezvoy, C., Charif, D., Gueguen, L., & Marais, G. A. B. (2007). MareyMap: an R-based tool with graphical interface for estimating recombination rates. *Bioinformatics*, *23*(16), 2188–2189. https://doi.org/10.1093/bioinformatics/btm315

Schmidt, R., West, J., Love, K., Lenehan, Z., … C. L.-, & 1995, U. (1995). Physical map and organization of Arabidopsis thaliana chromosome 4. *Science.Sciencemag.Org*. Retrieved from https://science.sciencemag.org/content/270/5235/480.abstract

Schmutz, J., Jackson, S. A., Wang, M., Wing, R. A., Gao, D., Zhang, M., … Cannon, S. B. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics*, *46*(7), 707–713. https://doi.org/10.1038/ng.3008

Tang, X., Szinay, D., Lang, C., Ramanna, M. S., Van Der Vossen, E. A. G., Datema, E., … Bai, Y. (2008). Cross-species bacterial artificial chromosome-fluorescence in-situ hybridization painting of the tomato and potato chromosome 6 reveals undescribed chromosomal

rearrangements. *Genetics*. https://doi.org/10.1534/genetics.108.093211

Taylor, J., & Butler, D. (2017). R package ASMap: Efficient genetic linkage map construction and diagnosis. *Journal of Statistical Software*, *79*(Stam 1993). https://doi.org/10.18637/jss.v079.i06

Chapter 2: QTL mapping for key agronomic and domestication traits in lima bean

Abstract

Lima beans, *Phaseolus lunatus,* has been considered an orphan crop due to its relative lack of research and varietal improvement, and lower production acreage compared to other legume crops, such as common bean, *Phaseolus vulgaris*. The lack of genome resources available for lima bean has hindered the understanding of important traits in this crop and has so far relied upon the assumption of synteny between lima bean and common bean. The development of a recombinant inbred line population, a genetic map, and an annotated reference genome for lima bean in recent years has enabled for the first time QTL mapping of key agronomic traits to propel breeding advancements of lima bean. Agriculturally relevant QTL were identified for plant habit, seed weight, seed color, flowering time, floral bud position, and crown pod density.

Introduction

Lima bean (*Phaseolus lunatus*) is a source of protein, fiber, and nutrients for people throughout the world; an important nitrogen-fixing crop for many growers; and the most important grain legume in California. Dry lima bean production in the US is valued around $30 million dollars annually over the past decade, accounting for 60-80% of total global production, with California as the sole producer of dry lima bean production nationally (Long et al., 2014; USDA ERS, 2004; USDA NASS, 2019). The University of California Davis lima bean breeding program is continually developing improved cultivars of large and baby lima beans. To enhance breeding efforts, advanced breeding techniques need to be deployed to accelerate the rate of

genetic gain towards lima bean cultivars with enhanced biotic and abiotic stress resistance, and improved yield and quality.

Until recently, there has been limited genetic research on lima bean, which has led to decades of reliance on genetic information from its close domesticated relative, common bean (*Phaseolus vulgaris*). Several advances over the past decade have led to a cascade of improvements in lima bean genetic and genomic resources, including the development of the first biparental recombinant inbred line population of lima beans at UC Davis, the subsequent development of a high quality molecular linkage map, and the assembly of the first whole-sequence genome publicly available with annotations (Dohle, 2017; Garcia et al., 2021). These advances in genetic tools provide the infrastructure to identify the chromosomal locations of key agronomic and domestication traits in lima beans segregating in this biparental lima bean population and enhance our understanding of their genetic function by identifying putative candidate genes.

The parental polymorphisms between UC 92 and UC Haskell lead to segregations of multiple agronomic and domestication traits that can be mapped throughout the biparental recombinant inbred line population resulting from their hybridization. Depending on the complexity of inheritance and epistatic interactions of these polymorphic traits, some transgressive segregation can be expected within the population. Key chromosomal regions underlying important physiological, domestication or agronomic traits may often be syntenic between lima and common beans on the same chromosomal region. In common bean, loci for determinacy and maturity on chromosome Pv01 may offer insight into putative homologs in lima bean (Kwak et al., 2012). QTL mapping of this population may identify regions containing

homologs in related *Phaseolus* and legume species and identify candidate genes for important agronomic and domestication traits.

The objectives of this study are to map QTL for segregating traits identified in the biparental recombinant inbred line population underlying key agronomic and domestication traits, to calculate the QTL effects for these traits and identify putative candidate genes and homologs for these traits in related legume species. Identifying the location of causative QTLs within lima bean will enhance the knowledge of gene function and the establish the relationship between collocating QTLs to accelerate breeding efforts. The utilization of molecular markers derived from these QTL analyses can support a more efficient method of breeding by incorporating marker-assisted-selection in this species.

Materials and Methods

Plant material and experimental design

A biparental recombinant inbred line (RIL) population was developed at the University of California, Davis (38.542790° N. Lat.; 121.763049° W. Long.) from reciprocal crosses of two contrasting California Central Valley-adapted cultivars, UC 92 and UC Haskell, resulting from distinct domestication gene pools of Lima bean. UC Haskell is a small-seeded, vine-type cultivar of Mesoamerican domestication origin with lygus-resistance and UC 92 is a large-seeded, bush-type cultivar of Andean origin with lygus-susceptibility.

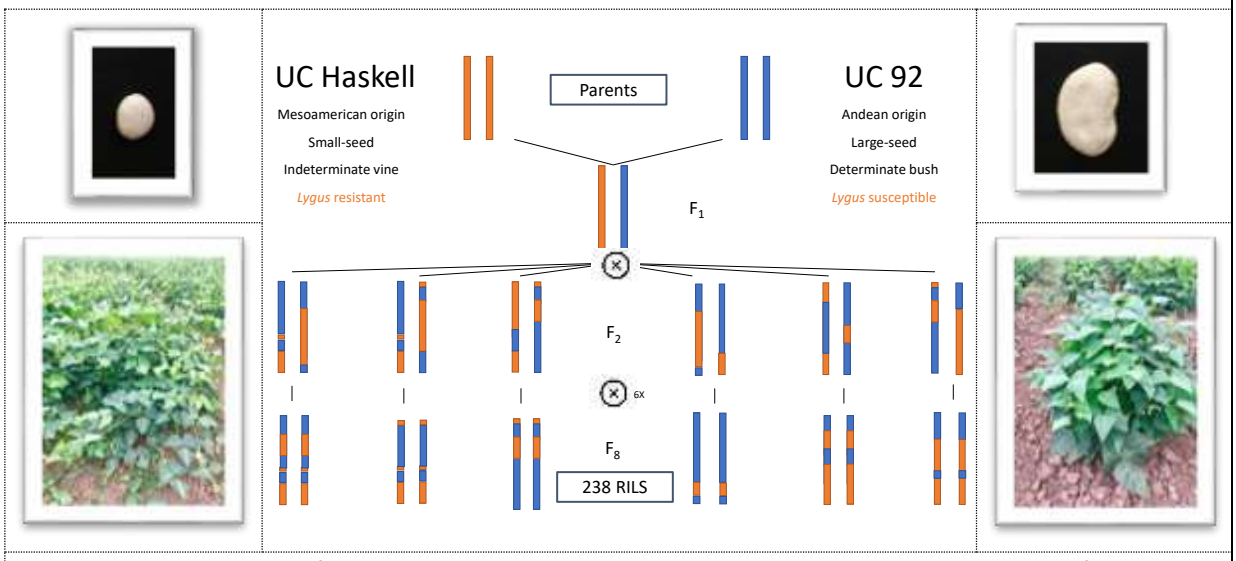

Figure 2.1 Development of the UC 92 x UC Haskell RIL population and the distinctive traits for seed size and plant growth habit.

During the summer of 2017, 238 recombinant inbred lines and the two inbred parental lines were planted as single plants into 8" azalea pots with standard UC Davis potting soil and grown within a greenhouse at the University of California, Davis. Leaf tissue was sampled from two-week-old seedlings of 238 RILs in the F_8 generation and the two inbred parental lines; following extraction of their DNA, these lines were sequenced using GBS as described in the Materials and Methods of Chapter 1. Seeds were hand-harvested from these individual sequenced plants and propagated by single-seed-descent over two additional greenhouse cycles; 180 RILs with sufficient seed were selected for field trials.

A field experiment of 180 recombinant inbred lines were planted in a randomized, augmented balanced incomplete block, split-plot design at two locations in May of 2018. Each plot was a 4-foot [1.22 m] double-row planting on 60-inch [1.52 m] beds with a 4-inch [10.16 cm] between-plant spacing of 24 seeds. Due to the differences in management of different plant habits, bush RILs and vine RILs were separated into distinct experiments. The vine

experiment had 110 RILs with a between-plot spacing of 6-feet [1.83 m], four checks (Calico Cat, Hopi12, RIL39, RIL72), that were replicated twice, and two controls [UC Haskell (lygus-resistant) and Cariblanco (lygus-susceptible)] that were replicated in each of three balanced incomplete blocks. To prevent cross-contamination of vine RILs across beds, a double-row bed of black-seeded bush-type common beans, *P. vulgaris*, were planted between the beds of vine RILs. The bush experiment had 61 RILs with a between-plot spacing of 3-feet [0.91 m], three checks (Dixie Speckled, Henderson, Jackson Wonder) that were replicated twice, and two controls [UC 92 (lygus-susceptible) and UC Beija Flor (lygus-resistant)], which were replicated four times in each of the three balanced incomplete blocks.

The split plot experimental design was created for the application of pesticide treatments to mitigate lygus damage on one plot and no pesticide treatments targeting lygus insects on the other plot. The pesticide treatment plots received applications of Warrior II (3.75 oz/acre) and Mustang (4 oz/acre) when more than three lygus bugs were observed on more than half of the control plots. The no-pesticide treatment main plots did not receive either of these pesticide treatments that would target lygus insects throughout the season.

Phenotypic data collection and statistical analysis

Phenotypic characterization for plant habit, hundred-seed weight and seed color in the QTL mapping study was performed on single plants grown within a greenhouse during the summer of 2017 at the University of California, Davis (38.542790° N. Lat.; 121.763049° W. Long.). These individual plants were evaluated for plant habit at maturity and classified as either ‘determinate’ or ‘indeterminate’. Seeds were hand-harvested from individual plants and

weights were collected a hundred-seed-count basis. If a recombinant inbred line did not have one-hundred seed, a minimum of twenty-five seeds were counted, weighed and the hundred-seed weight was calculated. While UC 92 and UC Haskell both have white seed coat colors, progeny within this population segregated for a mild green hue on their seed coat and was classified as having either 'color' or 'no color'.

Phenotypic data for days to first flower, inflorescence position, plant height, crown pod density, and yield were collected from the split-plot experiment grown at two locations in Davis, California, in 2018, as described in detail above. Days to first flower was recorded by subtracting the flower date (when a majority of plants in the plot had at least one floral bud open) from the planting date. Inflorescence position was recorded as a majority of the inflorescences in a plot being: below the canopy, below and at the canopy, at the canopy, at and above the canopy, above the canopy and far above the canopy on a rating scale of 0-5, respectively. Plant height was measured in centimeters using a meter stick that measured the height of the average plant in the plot from the soil surface at maturity. Crown pod density was recorded at harvest as the location and dispersal of pods throughout the plant in a majority of the plants in a plot as: no pods, very few pods, few pods, some pods, many pods, and many pods densely around the crown on a rating scale of 0-5, respectively. Yield data was recorded as the weight in grams of good quality seed harvested from the plot and normalized across the number of plants in each plot.

An ANOVA of a generalized linear mixed model, including genotype and treatments and their interaction as fixed effect factors and location and blocks as random effect factors, was performed for these traits using the 'stats' and 'lme4' packages in R statistical software (Bates

et al., 2015; R Core Team, 2013). The residuals of the models were evaluated to meet the assumptions of normality and homogeneity to perform analysis of variance for these data using the 'shapiro.test' and 'leveneTest' functions in R statistical software, respectively (Fox & Weisbery, 2019; R Core Team, 2013). The experimental design assured that all plots analyzed met the assumption of independence to perform an analysis of variance. The least squares means for the genotypes were calculated using the 'lsmeans' package, and the correlations between traits were calculated using the 'corrplot' package (Lenth, 2016; Wei et al., 2017).

QTL mapping

QTL mapping for plant habit, days to first flower, inflorescence position, plant height, crown pod density, yield, hundred-seed weight, and seed coat color was performed in the UC 92 - UC Haskell RIL population using the R/qtl package (Broman, 2009). A genome-wide scan for single QTLs was performed using the 'scanone' function with the extended Haley-Knott regression method. For traits that were controlled by a single QTL, composite interval mapping was performed using the 'cim' function with the extended Haley-Knott regression method, the Kosambi mapping function, and 1000 permutations to identify the position and LOD score of QTLs above the 95% significance threshold. For traits that contained multiple significant QTL, the 'makeqtl' and 'fitqtl' functions were used to identify the optimal multiple QTL model, and to calculate LOD scores, percent of phenotypic variation explained by the QTL and the QTL effects. Major QTLs are defined as those that account for > 10% of the phenotypic variation explained, and minor QTLs are defined as those that account for < 10% of the phenotypic variation explained (Collard et al., 2005).

Results

Quantitative trait loci (QTL) were mapped in the UC 92 - UC Haskell biparental population for agronomic and quality traits. Thirteen QTL were identified in the biparental population, and the LOD scores, QTL effects and population distribution are summarized in Table 2.1 below.

Table 2.1 QTL peaks, LOD scores, percent of phenotypic variation explained, the QTL effect and the parental direction of the positive allele.

Trait	QTL Peak	LOD	% PVE	QTL Effect	QTL Effect (+) Allele
Determinacy	Pl01_42432553	78.60	78.71	--	--
Flowering Time	Pl01_42432553	9.80	29.9	5.3 days	UC Haskell
Inflorescence Position	Pl01_42448854	10.6	22.71	--	--
	Pl11_33694066	3.58	7.00	--	--
Plant Height	Pl04_42709541	3.68	12.49	6.19 cm	UC Haskell
Pod Set Density	Pl08_49414866	4.05	11.47	--	--
Yield	Pl03_39354044	4.07	8.5	165.19 grams	UC Haskell
Hundred Seed Weight	Pl10_39674709	8.32	12.5	6.2 grams	UC 92
	Pl03_26637359	4.30	6.20	3.7 grams	UC 92
	Pl09_6453659	3.66	5.2	3.4 grams	UC 92
	Pl04_2884774	3.28	4.7	3.4 grams	UC 92
Seed Coat Color	Pl07_46426599	11.60	20.3	--	--
	Pl07_45067785	4.50	8.5	--	--

One major QTL for determinacy was identified on the long arm of chromosome Pl01, explaining 78% of the phenotypic variation. The peak LOD score for determinacy was located at the first marker on the long arm, after a nearly 20 cM and 20 Mbp gap in the pericentromeric region of chromosome Pl01. A major QTL was also found on chromosome Pl01 for flowering

time at the same peak marker as the QTL for determinacy. The QTL for flowering time explained only 30% of the phenotypic variation, which suggests that other genes influence flowering time in this population. For flowering time, transgressive segregation was observed for both earlier and later progenies than the UC 92 and UC Haskell parents, respectively.

A major QTL for inflorescence position also collocated to the same region of chromosome Pl01 as the QTLs for determinacy and flowering time and explained 23% of the phenotypic variation associated with this trait. A minor QTL for inflorescence position was also found on the long arm of chromosome Pl011 and explained an additional 7% of the phenotypic variation for the trait. The two QTLs for this trait are additive in nature, with the UC 92 haplotypes for both QTLs increasing the position of the inflorescence above the canopy. Inflorescence position was only transgressive towards phenotypes that had inflorescences that extended far above the canopy, a phenotype that was primarily observed in determinate recombinant inbred lines and transgressed beyond the UC 92 phenotype.

A single QTL was identified for plant height on the long arm of chromosome Pl04, which explained 12.5% of the phenotypic variation in the population, and transgressive segregation was observed towards taller and shorter progenies within the population. The QTLs and associated information for determinacy, flowering time, inflorescence position and plant height are observed in Figure 2.2 below.

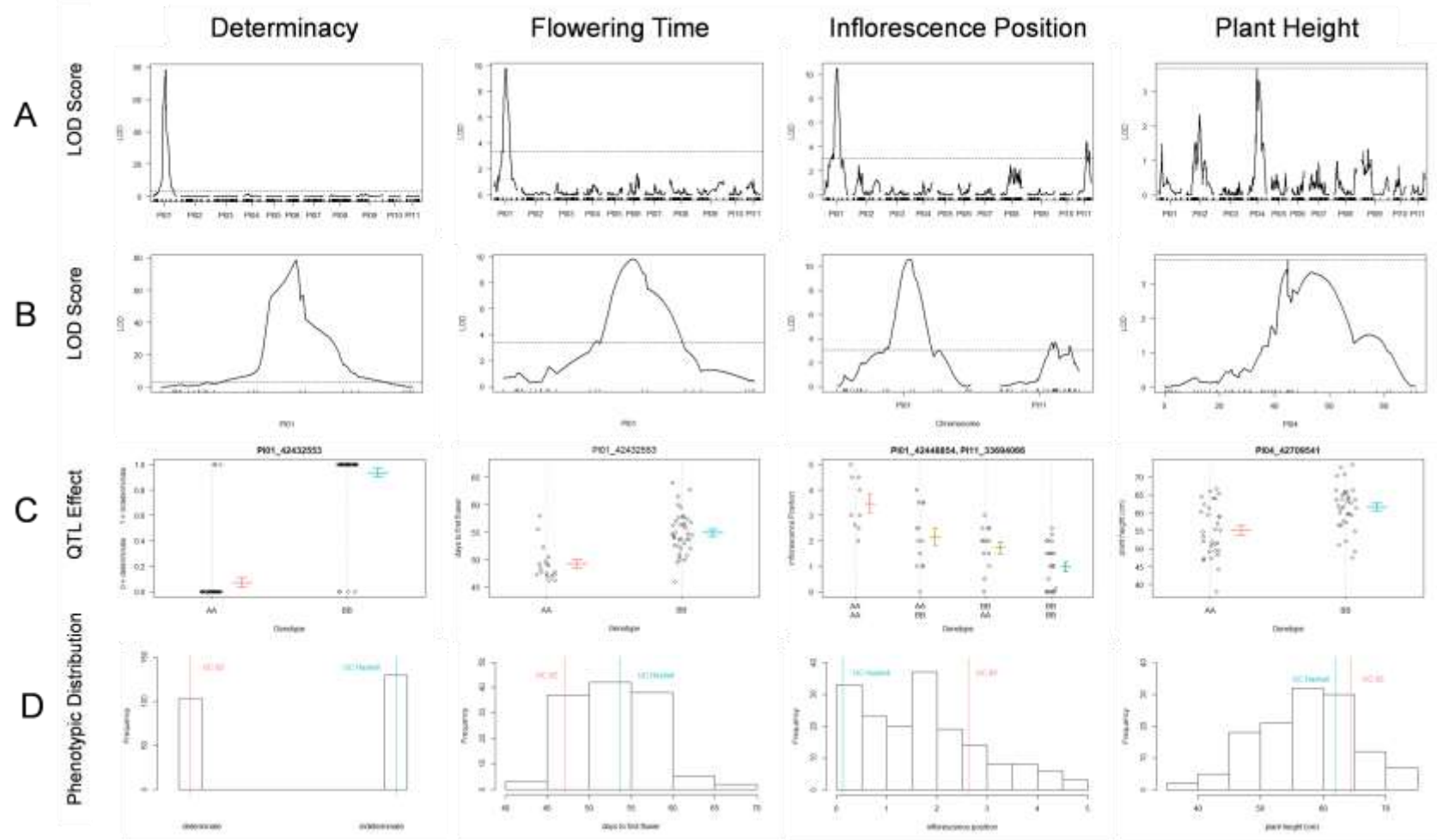

Figure 2.2 QTL mapping results in the UC 92 – UC Haskell RIL population A. LOD scores across all chromosomes, B. LOD scores on peak chromosomes, C. QTL effects with parental haplotype combinations and D. the phenotypic distribution in the RIL population for determinacy, flowering time, inflorescence position and plant height.

A single QTL was identified for pod set density on the long arm of chromosome Pl08, which explained 11.5% of the phenotypic variation observed in the population. Transgressive segregation was observed for progenies that had fewer and less dense pod distribution at maturity. A single QTL was identified for yield on the long arm of chromosome Pl03, which explained 8.5% of the phenotypic variation. This trait also had transgressive segregation observed only in progenies that had lower yields than the lesser yielding parent, UC 92.

Four additive QTL were identified for hundred seed weight, collectively explaining 28% of the phenotypic variation, including one major QTL on the long arm of chromosome Pl10 that explained 12.5% of the phenotypic variation. Transgressive segregation was only observed for seed weights below the small-seeded parent UC Haskell, and not for larger seed weights.

Two QTL for seed coat color were found on the long arm of chromosome Pl07 that explained 29% of the phenotypic variation associated with this trait. The expression of seed coat color in this population is due to the epistatic interaction of these two QTLs or complementary gene action since the light green hue on the seed coat is only observed in progeny and not observed in either parent. The QTLs and associated information for pod set density, yield, hundred seed weight and seed coat color are observed in Figure 2.3 below.

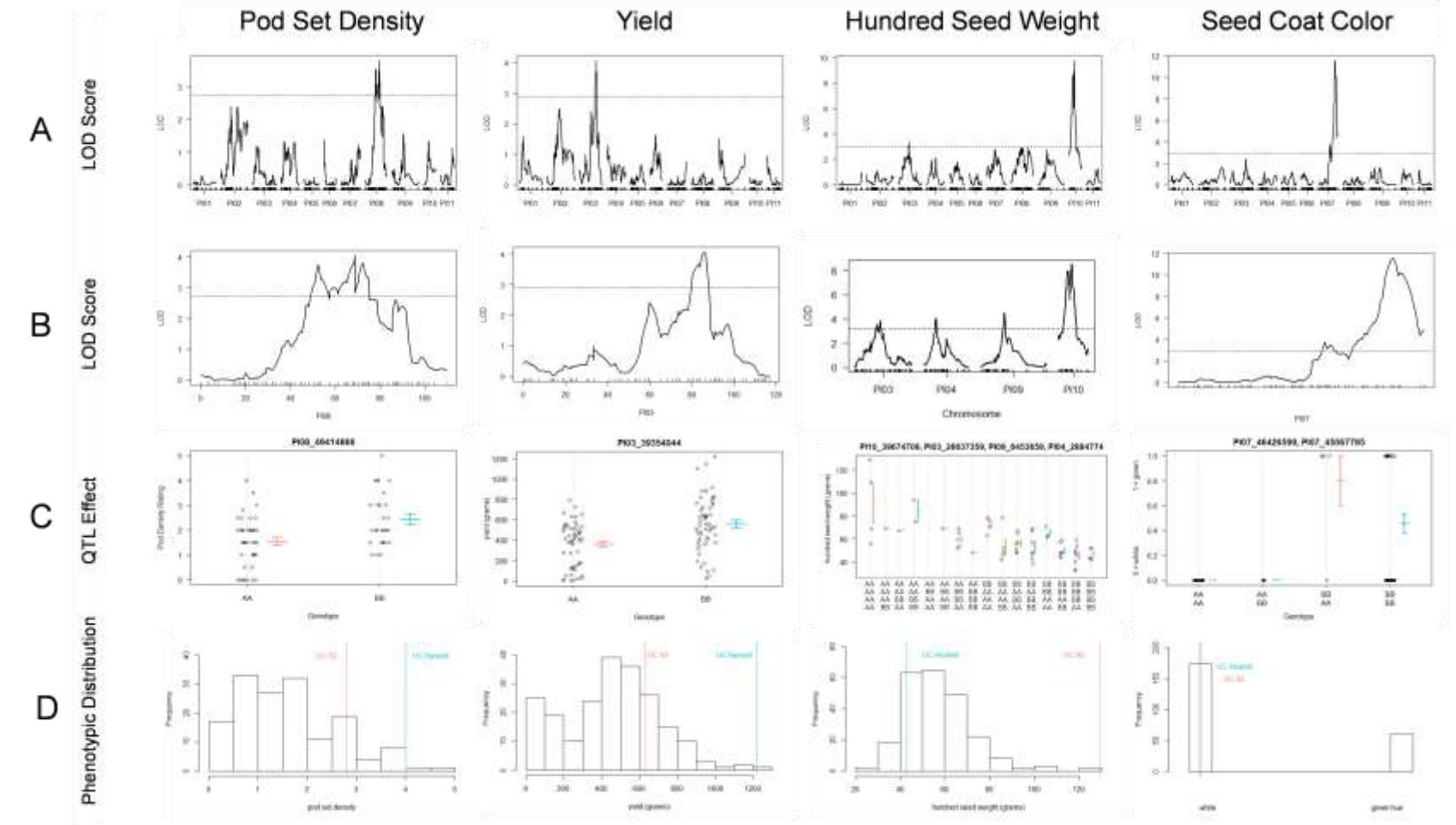

Figure 2.3 QTL mapping results in the UC 92 – UC Haskell RIL population A. LOD scores across all chromosomes, B. LOD scores on peak chromosomes, C. QTL effects with parental haplotype combinations and D. the phenotypic distribution in the RIL population for pod set density, yield, hundred seed weight and seed coat color

The correlations between all eight of these traits were evaluated, and results are shown in Figure 2.4 below. Yield had positive correlations of 0.34 and 0.45 with pod set density and plant height, respectively, at the $p < 0.001$ significance levels. Days to first flower had a significant correlation of 0.71 with determinacy at the $p < 0.001$ significance level, where there was a strong correlation among determinate habits and early flowering times. Days to first flower also had a significant correlation of 0.45 with inflorescence position at the $p < 0.001$ significance level, where there was a strong correlation among inflorescence positions above the canopy and early flowering times. There was also a negative correlation of 0.31 between pod set density and seed weight, at the $p < 0.001$ significance level. Plant height also had positive correlations of 0.20 and 0.18 with pod set density and the indeterminate habit, respectively, which were both significant at the $p < 0.05$ significance levels.

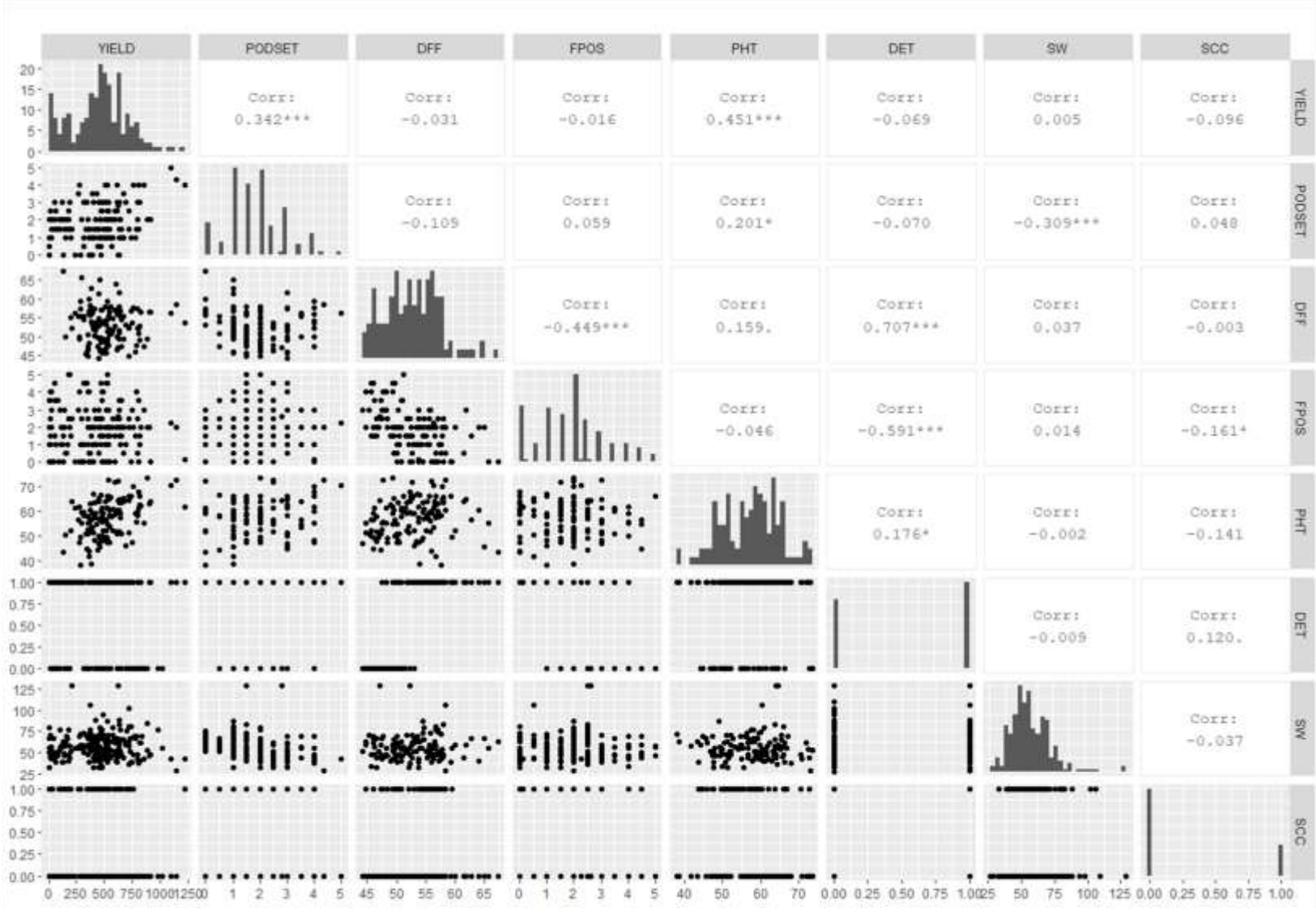

Figure 2.4 Correlations between yield, pod set density, flowering date, inflorescence position, plant height, determinacy, seed weight and seed coat color in the UC 92 – UC Haskell recombinant inbred line population.

Discussion

QTL mapping of the UC 92 - UC Haskell population offers not only the ability to more rapidly advance key traits in a breeding population for future varietal release, but provides insight on the drivers of domestication traits through QTL mapping. A likely causative gene for this QTL is *PlTFL1* at 41 Mbp on chromosome Pl01. This gene is an ortholog of the Arabidopsis gene TFL1 and the common bean determinacy gene *PvTFL1y*. *PvTFL1y* was previously mapped at 45Mbp of chromosome Pv01 (Campa et al., 2018; Repinski et al., 2012). As in common bean, the presence of the *PlTFL1* gene is associated with a QTL for both determinacy and flowering time: determinacy can cause earlier flowering by converting the terminal meristems from a vegetative state into reproductive ones (Koinange, Singh, & Gepts, 1996). In this experiment, the determinacy trait was contributed by the Andean parent, UC 92, which is consistent with findings in common bean, where the determinacy trait is generally also of Andean origin (Repinski et al., 2012).

The QTL for determinacy collocated to the same region of chromosome Pl01 as the QTLs for flowering time and inflorescence position; this observation suggests that these traits may be influenced by a single gene or by a cluster of collocating genes that cosegregate within this population. This is also supported by the strong correlation among determinate habits within the progeny and earlier flowering times, both traits that were under heavy selection for during the domestication process of beans (Cober & Tanner, 1995; Kwak et al., 2012). The findings of these collocating QTLs on chromosome Pl01 for these three traits is consistent with prior work conducted on this lima bean population in earlier generation evaluations (Dohle, 2017).

Unfortunately, the pericentromeric region of chromosome Pl01 that these three traits map to has sparse marker coverage and limits the ability to fine-map QTLs in this region.

Multiple traits scored in the UC 92 – UC Haskell population showed transgressive segregation to varying extents. However, seed weight only showed transgressive segregation for seed weights below the small-seeded parent UC Haskell, and not for the larger seed weights. This observation is consistent with prior results showing a shift towards smaller-seeded segregants, observed in common bean (Blair et al., 2006; Checa & Blair, 2012). This observation highlights the difficulty in developing large-seeded improved progenies in populations arising from selfing only. Possible solutions are the development of population based on backcrosses to a large-seeded parent. An alternative is suggested by the phenotypic distribution for HSW, discriminated by allelic combinations. The presence of four marker alleles of the large-seeded parent (UC 92) for four seed weight QTLs corresponds to the heaviest seeds and vice-versa for the individuals with the three marker alleles of the small-seeded parent (UC Haskell). Thus, indirect selection using these markers in early generations may shift breeding populations towards larger seeds.

Determinacy and seed size are two traits that are strongly associated with the domestication syndrome in beans, and both of these traits are polymorphic in the parents and segregating in the recombinant inbred line progeny and underlying QTLs have been identified that are consistent with QTLs found in common bean (Blair et al., 2006; Repinski et al., 2012). A QTL in seed weight was identified on chromosome Pl07 (at the Pl07_43161814 marker), which was significant at the $p < 0.10$ significance level but omitted in the results since it was not significant at the $p < 0.05$ significance level. This location on chromosome Pl07 corresponds to

an important locus for domestication, the phaseolin locus, located on chromosome Pl07 in common bean, which is associated with seed weight and may be important for future lima bean research efforts (Blair et al., 2006; P. Gepts, Osborn, Rashka, & Bliss, 1986; Paul Gepts, 1988b).

The location of two QTLs on chromosome Pl07 for seed coat color is consistent with the *P* locus in common bean on linkage group 7, which controls the expression of color in beans (Caldas & Blair, 2009; Mcclean et al., 2018). The green hue on the seed coat was not observed in either parent, but the epistatic interactions of two distinct QTL on chromosome Pl07 result in the expression of this phenotype in a fraction of the population. Increased seed coat color in beans is associated with higher tannins, phenolics, and nutritional value, and may be a desirable trait for breeding lima beans in the future (Caldas & Blair, 2009).

The strong correlations between pod set density and plant height with yield provide insight into traits and loci that can be selected for to increase the agronomic productivity of lima bean lines from this biparental population. More vigorous plants and an increase in pod set density around the crown of the plant may offer more photosynthate reserves for pod filling or perhaps greater insulation from the abiotic and/or biotic stress of the California Central Valley that results in increases in yield. Since significant QTLs were identified for these traits, there is potential for accelerated molecular breeding towards these traits to increase agronomic productivity.

Conclusion

The identification of these 13 significant QTLs for domestication and agronomic traits in this biparental lima bean population provide opportunities to enhance breeding efforts for these traits in lima beans, both by increasing productivity and nutritional content for growers and consumers alike. These results reveal the genetic complexity of these traits and the interactions among QTLs that may assist in future breeding efforts for desirable phenotypes. By identifying local markers to these significant causative loci, marker-assisted-selection may be deployed to improve breeding efforts towards desirable traits. This QTL mapping study also identified homologous traits among lima bean and common bean and offer insight into the homology between these *Phaseolus* species at the genetic level, especially for traits involved in the domestication process.

References

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Blair, M. W., Iriarte, G., & Beebe, S. (2006). QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean x wild common bean (Phaseolus vulgaris L.) cross. *Theoretical and Applied Genetics*, *112*(6), 1149–1163. https://doi.org/10.1007/s00122-006-0217-2

Broman, K. W. (2009). *Statistics for Biology and Health A Guide to QTL Mapping with R / qtl*. https://doi.org/10.1007/978-0-387-92125-9

Caldas, G. V., & Blair, M. W. (2009). Inheritance of seed condensed tannins and their relationship with seed-coat color and pattern genes in common bean (Phaseolus vulgaris L.). *Theoretical and Applied Genetics*, *119*(1), 131–142. https://doi.org/10.1007/s00122-009-1023-4

Campa, A., Murube, E., & Ferreira, J. J. (2018). Genetic Diversity, Population Structure, and Linkage Disequilibrium in a Spanish Common Bean Diversity Panel Revealed through Genotyping-by-Sequencing. *Genes*. https://doi.org/10.3390/genes9110518

Checa, O. E., & Blair, M. W. (2012). Inheritance of yield-related traits in climbing beans (Phaseolus vulgaris L.). *Crop Science*, *52*(5), 1998–2013. https://doi.org/10.2135/cropsci2011.07.0368

Cober, E. R., & Tanner, J. W. (1995). Performance of related indeterminate and tall determinate soybean lines in short-season areas. *Crop Science*, *35*(2), 361–364. https://doi.org/10.2135/cropsci1995.0011183X003500020011x

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005, January). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*. https://doi.org/10.1007/s10681-005-1681-5

Dohle, S. (2017). *Development of Resources for Lima Bean (Phaseolus lunatus) Breeding and Genetics Research*. University of California, Davis.

Fox, J., & Weisbery, S. (2019). *An R Companion to Applied Regression - John Fox, Sanford Weisberg - Google Books*. *Sage Publications*. Retrieved from https://books.google.com/books?hl=en&lr=&id=uPNrDwAAQBAJ&oi=fnd&pg=PP1&dq=fox+weisberg+car&ots=MwF64I2A61&sig=PZXVVsfUb19NftmcF9vV1uuahhU#v=onepage&q=fox weisberg car&f=false

Garcia, T., Duitama, J., Zullo, S. S., Gil, J., Ariani, A., Dohle, S., … Chacón-Sánchez, M. I. (2021). Comprehensive genomic resources related to domestication and crop improvement traits in Lima bean. *Nature Communications*, *12*(1), 1–17. https://doi.org/10.1038/s41467-021-20921-1

Gepts, P., Osborn, T. C., Rashka, K., & Bliss, F. A. (1986). Phaseolin-protein Variability in Wild Forms and Landraces of the Common Bean (Phaseolus vulgaris): Evidence for Multiple Centers of Domestication. *Economic Botany*, *40*(4), 451–468. https://doi.org/10.1007/BF02859659

Gepts, Paul. (1988). *Genetic Resources of Phaseolus Beans*. *Genetic resources of Phaseolus beans*. https://doi.org/10.1007/978-94-009-2786-5_11

Koinange, E. M. K., Singh, S. P., & Gepts, P. (1996). Genetic Control of the Domestication Syndrome in Common Bean. *Crop Science*, *36*(4), 1037–1045. https://doi.org/10.2135/cropsci1996.0011183X003600040037x

Kwak, M., Toro, O., Debouck, D. G., & Gepts, P. (2012). Multiple origins of the determinate growth habit in domesticated common bean (Phaseolus vulgaris). *Annals of Botany*, *110*(8), 1573–1580. https://doi.org/10.1093/aob/mcs207

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*. https://doi.org/10.18637/jss.v069.i01

Long, R., Temple, S., Meyer, R., Schwankl, L., Godfrey, L., Canevari, M., & Roberts, P. (2014). Lima Bean Production in California. *Lima Bean Production in California*, (May). https://doi.org/10.3733/ucanr.8505

Mcclean, P. E., Bett, K. E., Stonehouse, R., Lee, R., Pflieger, S., Moghaddam, S. M., … Mamidi, S. (2018). White seed color in common bean (Phaseolus vulgaris) results from convergent evolution in the P (pigment) gene. *New Phytologist*, *219*(3), 1112–1123. https://doi.org/10.1111/nph.15259

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundatrion for Statistical Computing. Retrieved from http://www.r-project.org/

Repinski, S. L., Kwak, M., & Gepts, P. (2012). The common bean growth habit gene PvTFL1y is a functional homolog of Arabidopsis TFL1. *Theoretical and Applied Genetics*, *124*(8), 1539–1547. https://doi.org/10.1007/s00122-012-1808-8

USDA ERS. (2004). *Economic research service*. *Choice Reviews Online* (Vol. 42). https://doi.org/10.5860/choice.42-2206

USDA NASS. (2019). *USDA/NASS QuickStats*. Retrieved from https://quickstats.nass.usda.gov/results/9979DE70-2420-3A8E-A718-4D3E4C830346

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., & Zemla, J. (2017). R package “corrplot”: Visualization of a Correlation Matrix. *Statistician*, *56*, 316–324. Retrieved from https://peerj.com/articles/9945/Supplemental_Data_S10.pdf

Chapter 3: Cyanogenesis as a putative mechanism of resistance to lygus herbivory in lima bean

Abstract

Cyanogenesis is a constitutive defense mechanism that releases volatile hydrogen cyanide from damaged plant tissue, and is believed to deter herbivory in lima bean, *Phaseolus lunatus*. Cyanogenesis has been reported in various other crops, including cassava, white clover and almonds as a mechanism conferring herbivore resistance. The Western tarnished plant bug, *Lygus hesperus* Knight, is a widespread pest impacting lima bean production in the California Central Valley. This study evaluates the volatile cyanide expression in the reproductive tissues of lima bean in a recombinant inbred line population to understand if cyanogenesis is an effective mechanism of herbivore resistance to lygus. It also evaluates the heritability of this biotic stress resistance trait through QTL mapping using a novel high-throughput quantitative analysis of volatile cyanide expression.

Introduction

Significance of lima bean and lygus herbivore damage

Lima bean, *Phaseolus lunatus,* is the most important grain legume in California, a vital source of protein and nutrition for people throughout the world, and an important nitrogen-fixing crop for many growers. Lima bean accounts for nearly half of the total dry bean production in California, with Stanislaus, San Joaquin and Sutter counties leading lima bean production over the past decade (USDA ERS, 2004; USDA NASS, 2019). Dry lima bean production in the US is valued around $30 million dollars annually over the past decade,

accounting for 60-80% of total global production, with California as the sole producer of dry lima bean production nationally (Long et al., 2014; USDA ERS, 2004; USDA NASS, 2019).

The most significant factor impacting lima bean production in California in recent year is pest pressure. Pest pressure is particularly significant on organic lima bean production, which is limited in the effectiveness of applications to deter pests. In the California Central Valley, the Western tarnished plant bug, *Lygus hesperus* Knight, is the most severe contributor to agronomic losses of lima bean and can result in a 70% yield reduction (Dohle, 2017; Long et al., 2014). Lygus is a pest to other valuable agronomic crops throughout the Western US and California, including strawberry, alfalfa seed and cotton, and has a problematic relative in the tarnished plant bug, *Lygus lineolaris,* which impacts cotton, carrot seed, and other crops in the Eastern US (Scott, 1977; Young, 1986). These species can be found from southern Canada to northern Mexico, host on hundreds of agronomic and native plants, are highly mobile, and are able to inflict significant damage to crop yields with low population levels (Long et al., 2014).

Due to the high mobility of adult lygus insects, chemical methods are limited in their effectiveness on adult population control and damage, and there is a demand for varieties that confer resistance to lygus damage. Increasing popularity of organic production in other crops, particularly strawberry, has led to trap cropping with lygus' preferred host, alfalfa, and continuous rounds of tractor-mounted vacuuming to control lygus populations in the field (Swezey et al., 2007). Previous research has concluded that resistance to lygus in lima beans is heritable, however the mechanisms for resistance and mode of inheritance need to be further researched (Dohle, 2017).

Development of defense mechanisms in plants

The development of biochemical mechanisms of defense to herbivory exists in thousands of species across the plant kingdom (Poulton, 1990). These defense mechanisms involve the deployment of different chemical compounds prior to, during and/or following herbivory, and include volatile organic compounds, including terpenoids and phenols, glucosinolates, and cyanogenic glucosides (Ballhorn et al., 2009; Halkier & Du, 1997; Walling, 2000). Several important agricultural crops produce cyanogenic glucosides that volatilize into hydrogen cyanide as part of a defense mechanism; these include almonds, apples, apricots, cherries, cassava, white clover, flax, and lima bean (Conn, 1969; D. A. Jones, 1998; Poulton, 1990; Vetter, 2000). This two-component chemical defense mechanism involves the reaction of the cyanogenic glucoside substrate stored within the vacuole and the β-glucosidase enzyme stored in the apoplast, which produces toxic hydrogen cyanide in response to plant tissue damage (Frehner & Conn, 1987; Miller & Conn, 1980).

Different plant species utilize different cyanogenic glucosides as primary precursors for cyanogenesis, including amygdalin in almonds and prunasin in cherries and apricots (Conn, 1969; Frehner et al., 1990; Poulton, 1990). The primary substrate of this reaction in lima beans is linamarin and appropriately the associated enzyme is linamarase; the secondary substrate is lotaustralin, which is structurally identical to linamarin, with an

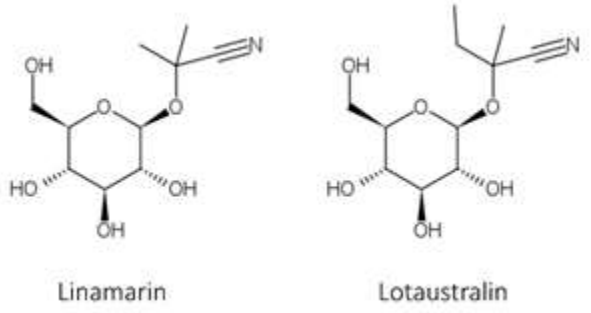

Figure 3.1 Linamarin and lotaustralin cyanogenic glucoside precursors to cyanide production in lima bean tissue.

additional methyl group (Figure 3.1) (Shlichta et al., 2014). In the absence of herbivory, cyanogenic glucosides are constitutively stored in the vacuole of the plant cells and do not react to produce the toxic defense compound. It is only after physical damage ruptures the vacuole and cell membranes that the linamarin substrate and linamarase enzyme can react and produce volatile hydrogen cyanide as the defense compound (Figure 3.2). Tissue damage, including herbivory and freezing temperatures, creates an environment for the cyanogenic glucoside and the β-glucosidase enzyme to react and produce volatile hydrogen cyanide. The repellent effect of cyanogenesis on plant herbivores has been observed in lima bean with the pest *Schistocerca gregaria* Forskal, and in white clover and cassava; it may play a role in antixenosis and/or antibiosis in lima bean in response to lygus (Ballhorn et al., 2005; Miller & Conn, 1980).

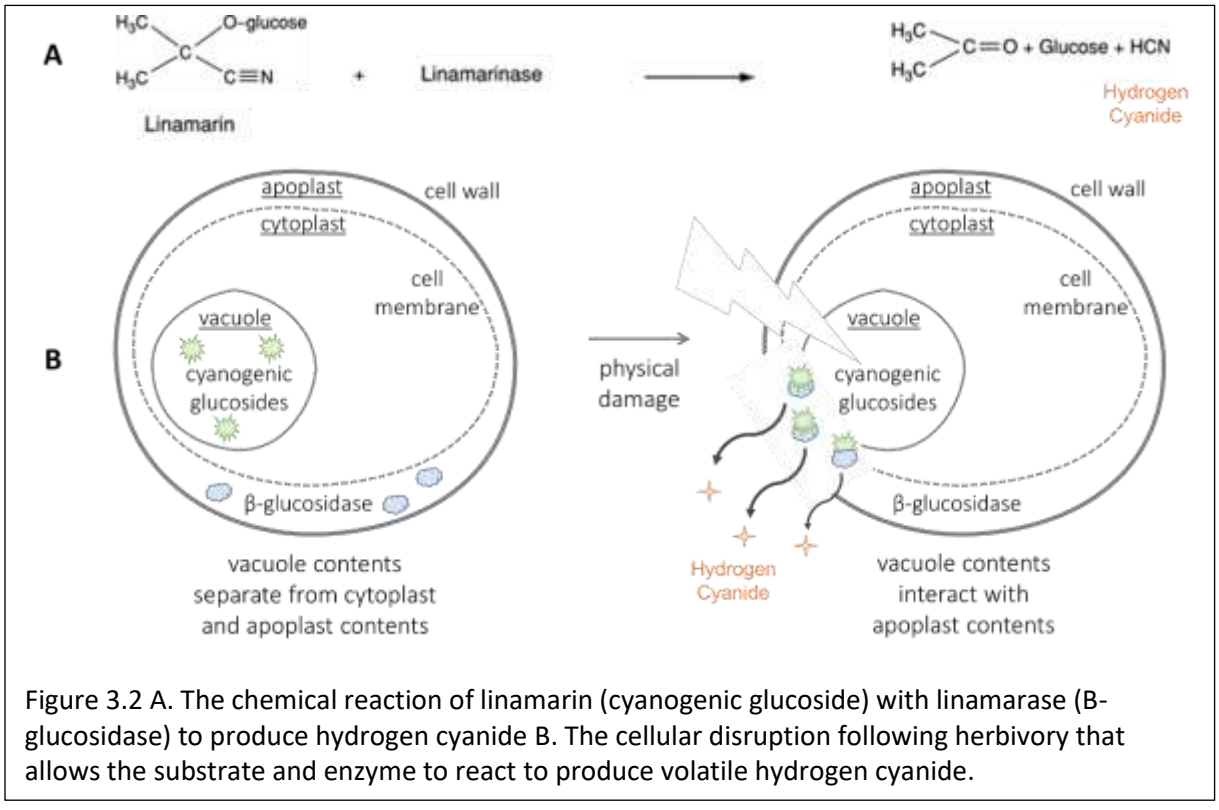

Figure 3.2 A. The chemical reaction of linamarin (cyanogenic glucoside) with linamarase (B-glucosidase) to produce hydrogen cyanide B. The cellular disruption following herbivory that allows the substrate and enzyme to react to produce volatile hydrogen cyanide.

Wild accessions of lima beans have been recorded having elevated levels of cyanide production upon tissue damage (Ballhorn et al., 2009). During the domestication process the production of hydrogen cyanide in the seed was selected against to reduce toxicity in favor of human consumption; however, it is unknown if the reduction in cyanogenic compounds was selected against in other plant tissues.

This research focuses on the existence of volatile hydrogen cyanide by reproductive tissues of lima bean as a putative mechanism of resistance to lygus damage that impacts yield productivity. Multiple studies were conducted to evaluate the differences in volatile cyanide production among different reproductive tissues of lima beans, across different environments, across different insecticidal treatments and assuming different biotic stress, following repeated injury, across different time intervals, across different genotypes, across different diversity statuses, and with different allelic variants.

The California-adapted cultivar study conducted in 2017, 2018, 2019 and 2020 evaluated the differences in cyanide production in different reproductive tissues (floral bud, immature pod, mature pod tissue and fresh bean tissue) among 12 California-adapted cultivars in field locations with two different insecticidal treatments over two years, and then tested in the greenhouse in three different years. The cyanide production was analyzed across different time intervals (0-30 minutes, 30-60 minutes, 60-90 minutes) and following repeated injury (primary, secondary, tertiary) on the same tissue samples. The objectives of this preliminary study were to identify if there are significant differences in volatile cyanide production (1) among the different reproductive tissue types, (2) across different time intervals, (3) following

repeated injury, (4) between different field and greenhouse environments and (5) among the different genotypes.

The status diversity study conducted in the fall of 2019 evaluated the differences in cyanide production in different reproductive tissues across different intervals for a diverse collection of accessions of different status levels (cultivated, escape, landrace, and wild). These accessions were obtained from the CIAT (Centro Internacional de Agricultura Tropical, Cali, Colombia) germplasm repository with the objective to identify differences in reproductive volatile cyanide production among the different status levels.

Materials and Methods

California cultivar study

An experiment of 12 California-adapted lima bean cultivars was grown with two replicates sampled during the summers of 2017 and 2018 on conventionally managed fields with two adjacent blocks receiving different insecticidal treatments in Davis, CA. The different insecticide applications were the application of Warrior II (3.75 oz/acre) and Mustang (4 oz/acre) when more than three lygus bugs were observed to mitigate lygus damage on one field plot, and no insecticide treatments targeting lygus insects on the other field plot. Each row or plot was a double row planting on 60-inch [1.52 m] beds with a 4-inch [10.16 cm] between-plant spacing. The 12 cultivars were also grown in single-plant 8-inch [20.32 cm] azalea pots in greenhouses during the summer of 2018, the spring of 2019 and the spring of 2020 in Davis, CA. Tissue was collected for floral buds and immature pods at all locations and years, tissue was collected for mature pods in the field in the summer of 2017 and the greenhouse in the fall of 2019, and fresh bean tissue was collected in the greenhouse in the spring of 2020. The cultivars

included in the study were Calico Cat (G26451), Dixie Speckled, Henderson Bush, Hopi 12 (G25623), Jackson Wonder, RIL 39, RIL 72, UC 92, UC Beija Flor, UC Cariblanco N, UC Haskell and UC Lee.

Status diversity study

Accessions of wild, escape, and landrace lima beans were collected from CIAT based on their diversity for seed size, seed color, origin, altitude, and habit and include accessions from both the Andean and Mesoamerican gene pools. These accessions were grown with UC 92 and UC Haskell as checks. Seeds were planted in a greenhouse in Davis, California, in the late fall of 2018 and floral bud tissue, immature pod, and mature pod tissue were collected in the early months of 2019 for analysis on the production of volatile cyanide in these reproductive tissues. The accessions included in the study are listed in Appendix 3.1.

Recombinant inbred line study

A recombinant inbred line (RIL) population was created from a cross of a lygus-resistant, small-seeded vine-type California cultivar, UC Haskell, and a lygus-susceptible, large-seeded bush-type California cultivar, UC 92 (Dohle, 2017). An experiment of 180 RILs was planted in a randomized augmented balanced incomplete block split-plot design at two locations in May of 2018. Each plot was a 4-foot [1.22 m] double-row planting on 60-inch [1.52 m] beds with a 4-inch [10.16 cm] between-plant spacing of 24 seeds. Due to the differences in management of different plant habits, bush RILs and vine RILs were separated into distinct experiments. The vine experiment had 110 RILs with a between-plot spacing of 6-feet [1.83 m], four checks (Calico Cat, Hopi12, RIL39, and RIL72), that were replicated twice, and two controls [UC Haskell

(lygus-resistant) and Cariblanco (lygus-susceptible)], which were replicated in each of three balanced incomplete blocks. To prevent cross-contamination of vine RILs across beds, a double-row bed of black-seeded bush-type common beans, *Phaseolus vulgaris*, were planted between the beds of vine RILs. The bush experiment had 61 RILs with a between-plot spacing of 3-feet [0.91 m], three checks (Dixie Speckled, Henderson, and Jackson Wonder) that were replicated twice, and two controls [UC 92 (lygus-susceptible) and UC Beija Flor (lygus-resistant)] that were replicated four times in each of the three balanced incomplete blocks. Tissue samples were collected from floral bud and immature pod tissue from this experiment in 2018.

The split-plot experimental design was created for the application of insecticide treatments to mitigate lygus damage on one plot and no insecticide treatments targeting lygus insects on the other plot. The plots that were controlled with insecticide treatments received applications of Warrior II (3.75 oz/acre) and Mustang (4 oz/acre) when more than three lygus bugs were observed on more than half of the control plots. The plots that did not receive pesticide treatments did not receive either of these pesticide treatments that would target lygus insects throughout the season.

High-throughput cyanide quantification

A high-throughput method of cyanide quantification was developed from an adapted protocol, which provided automated quantification of colorimetric data imagery collected from captured volatile cyanide samples (Schneider et al., 2012; Takos et al., 2010). Tissue samples were collected in triplicated subsamples for each plot and organized onto 96-well plates. Tissue samples were kept on dry ice and frozen at -80°C for a minimum of 24 hours. Hydrogen cyanide

production was collected at three-time points (30, 60 and 90 minutes) for all collected samples in 2017, 2018, and 2019.

The preliminary 2017 study on California-adapted cultivars included additional steps for preliminary testing and analysis. To observe the relationship between cyanide expression and tissue weight, the weights of all tissue samples were collected prior to freezing the samples at -80°C. The trait data was measured as nM HCN*30 min^{-1}*mg fresh plant $tissue^{-1}$, however the data had significant deviations from the assumptions of normality and homoscedasticity. No correlation was observed between nM HCN*30 min^{-1} and mg fresh plant tissue and no further analysis was conducted. Weight measurement of sample tissue was deemed unnecessary, and volatile cyanide production was measured as nM HCN*30 min^{-1}.

Feigl-Anger paper was prepared fresh for each reaction by separately dissolving 0.25 g of copper ethylacetylacetate in 25 ml of chloroform and 0.25 g of tetra base 4,4'-methylenebis (N,N-dimethylaniline) in 25 ml chloroform in separate glass beakers, thoroughly mixing each solution with a glass wand for 60 seconds, and then combining the solutions with additional mixing with a glass wand for 60 seconds. The combined solution was then poured into a glass 150 mm PYREX Petri dish (Fisher Scientific, 08-747F), and Whatman 3MM chromatography paper (GE, 3030-6185) cut to 8x12cm, was wetted in the impregnated solution for a few seconds on each side and allowed to dry under a hood. Tissue samples were removed from the -80°C freezer and allowed to thaw at room temperature for 30 minutes. The prepared Feigl-Anger paper was placed over each 96-well plate and a weighted flat surface was put on top to seal each well of the plate for 30-minute intervals across a total of 90 minutes. To observe the effects of repeated injury within tissue samples, tissue samples were returned to the -80°C

freezer for at least 24 hours and tested across three 30-minute intervals (30, 60 and 90 minutes) for a second and third time.

Standards of hydrogen cyanide were created with concentrations of 1000, 750, 500, 250, 100, 50, 25 and 0 nM KCN in 1M NaOH solution. 25 µl of 4 M H_2SO_4 were added to the 60 µl standards and exposed to Feigl-Anger paper for 30 minutes (Takos et al., 2010). A standard regression curve was calculated between the standard HCN concentrations and the blue absorbance intensity calculated as -log (Mean/255); nanomolar concentrations were calculated for each sample as a nM/30 minute rate across the intervals measured. To improve the prediction accuracy of the regression curve across lower and higher molar concentrations, two regression curves were developed for concentration conversions (Figure 3.3).

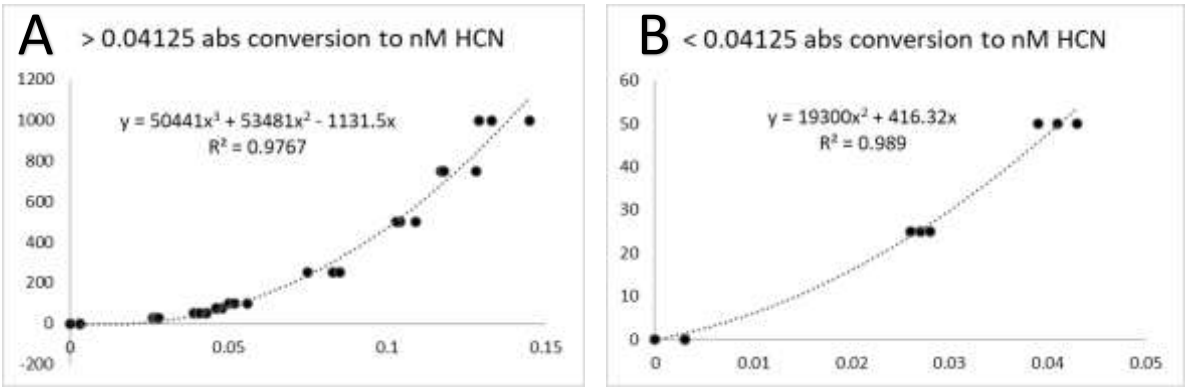

Figure 3.3 A. The standard regression curve of absorbance to nM HCN concentration for absorbance values above 0.04125 B. The standard regression curve of absorbance to nm HCN concentration for absorbance values below 0.04125.

High throughput automated data imagery analysis and quantification

The Feigl-Anger paper was immediately scanned onto a computer using 300 DPI image resolution, and quantification of the blue absorbance intensity was calculated using ImageJ

software (https://imagej.nih.gov/) and the 'readplate2' plugin (https://imagej.nih.gov/ij/plugins/readplate/index.html) (Schneider et al., 2012). For each 96-well plate and time interval, the scanned image underwent visual quality control in ImageJ to verify each well on the 96-well plate was properly aligned for accurate quantification of the blue light intensity for each well. Data output for each well was the blue light absorbance calculated as -log (Mean/255). After correcting the absorbance values for the blank across each 96-well plate, the absorbance value was converted to nanomolar concentrations of volatile hydrogen cyanide. This conversion was calculated for each sample as a nM/30 minute rate across the intervals measured based on the regression curve established from the standards of hydrogen cyanide.

Statistical analysis

An analysis of variance was performed to determine significant differences among variables depending on the experiment. The different volatile cyanide traits analyzed include the different tissue types (floral bud, immature pod, mature pod, and fresh seed), injury (primary, secondary, and tertiary) and time intervals (0 - 30 min, 30 - 60 min and 60 - 90 min for the floral bud, immature pod and mature pod tissue, and 0 - 240 min for the fresh seed tissue at physiological maturity). Independent variables include genotype (cultivar or QTL haplotypes or status), location, and treatment as fixed effects and block and year as random effect factors. Conversion of the absorbance measurements to nM led to strong positive skews of the data for all of the time points and tissues measured. Therefore, statistical analyses were conducted on

the absorbance measurements that produced normal distributions and met the assumptions necessary to conduct analyses of variances for these traits.

For the recombinant inbred line population, the volatile cyanide absorbance data for floral bud and immature pod produced bimodal distributions which violated the assumptions of normality to perform an analysis of variance for these traits. Following preliminary QTL mapping of these traits, significant QTL peaks were incorporated into the linear models for these traits and produced residuals that met the assumptions of normality and homogeneity to perform analysis of variance for these data. The experimental design assured that all plots analyzed met the assumption of independence to perform an analysis of variance. Linear models and an analysis of variance were performed for the RIL population incorporating the significant QTL peaks on chromosomes Pl05, Pl08, and Pl10 and chromosomes Pl05 and Pl10 for floral bud and immature pod traits, respectively. The best fitting linear model for floral bud cyanogenesis included genotype, location, and treatment as fixed effect factors.

The additive mixed effect models were created using the 'lme4' package and the ANOVA was performed with the 'stats' package in R statistical software (Bates et al., 2015; R Core Team, 2013). Prior to results being reported, the assumptions of independence, normality and homoscedasticity were tested by running the Shapiro-Wilk test of normality and the Levene's test for homogeneity of variance using the 'stats' and 'car' packages in R statistical software, respectively (Fox & Weisbery, 2019; R Core Team, 2013). Phenotypic data that deviated from the assumptions were set to missing. The adjusted R^2 incorporating the fixed and random effects of the linear models and the coefficient of variation of the linear models were calculated using the 'piecewiseSEM' and 'sjstats' packages in R statistical software, respectively (Lüdecke,

2020; Nakagawa & Schielzeth, 2013). The least significant differences between fixed effect factors were calculated using the 'lsmeans' package, the broad sense heritability of each trait was calculated following the heritability estimator equation across multiple environments and treatments, and the correlations between traits were calculated using the 'corrplot' package in R statistical software (J. B. Holland et al., 2003; Lenth, 2016; Wei et al., 2017).

QTL mapping

QTL mapping was conducted on the UC 92 – UC Haskell RIL population using the R/qtl package in R statistical software (Broman, 2009). A genome-wide scan for single QTLs was performed using the 'cim' function with the extended Haley-Knott regression method, the Kosambi mapping function, and 1000 permutations to identify the position and LOD score of QTLs above the 95% significance threshold. Since multiple significant QTL were identified, the 'makeqtl' and 'fitqtl' functions were used to identify the optimal multiple QTL model, and to calculate LOD scores, percent of phenotypic variation explained by the QTL and QTL effects. Major QTLs are defined as those that account for > 10% of the phenotypic variation explained, and minor QTLs are defined as those that account for < 10% of the phenotypic variation explained (Collard et al., 2005).

Results

Differences in reproductive tissue types, time intervals, and repeated injury in California-adapted cultivars

Repeated injury of California-adapted cultivars across all time intervals

Significant differences were detected among the 12 California-adapted cultivars of lima bean, among different reproductive tissues, among repeated injuries, among time intervals,

Table 3.1 Descriptive statistics for volatile cyanide production of 12 California cultivars across multiple injuries, different tissue types, time intervals, treatments and locations.

Factor	F values	Factor	Level	Means nM volatile HCN
Geno	63.50***	Tissue	floral bud	110.83b
Tissue	1317.80***		immature pod	183.33c
Trt	52.80***		mature pod	6.21a
Loc	53.48***		fresh seed[1]	0.05a
Time	12.22***			
Injury	77.75***	Time	0-30 min	74.1b
Geno*Tissue	20.84***		30-60 min	105.1b
Geno*Trt	7.36***		60-90 min	108.7b
Geno*Loc	10.87***		0-240 min[1]	0.05a
Tissue*Loc	103.23***			
Trt*Loc	5.05**	Injury	1	159.9b
Tissue*Time	43.14***		2	64.7a
Tissue*Injury	49.83***		3	63.3a
Time*Injury	12.10***			
Geno*Tissue*Trt	5.56***			
Geno*Tissue*Loc	10.72***	Tissue*Injury	floral bud 1	187.86c
Geno*Trt*Loc	9.64***		floral bud 2	62.18ab
Tissue*Trt*Loc	8.60***		floral bud 3	29.92a
Tissue*Time*Injury	4.34***		immature pod 1	248.62d
Geno*Tissue*Trt*Loc	1.51.		immature pod 2	181.11bcd
			immature pod 3	201.82bcd
CV (%)	20.22		mature pod 1	7.12a
R^2 (%)	87.4		mature pod 2	3.16a
H^2 (%)	3.35		mature pod 3	10.40a

'***' denotes that $p < 0.001$, '**' < 0.01, '*' < 0.05, '.' < 0.10, 'ns' > 0.10
'abcd' letters represent significantly different adjusted lsmeans groups at the $p < 0.05$ significance level

among the applications of insecticide treatments, among field and greenhouse locations, and in almost all of the genotype by environment interactions evaluated for the production of volatile cyanide. Tissue was the most significant factor followed by injury number, genotype, location, treatment and time interval, which were all significant at the $p < 0.001$ significance level. The descriptive statistics and volatile cyanide production across all genotypes, tissues, injuries, time intervals, treatments and locations are included in Table 3.1 and Figure 3.4.

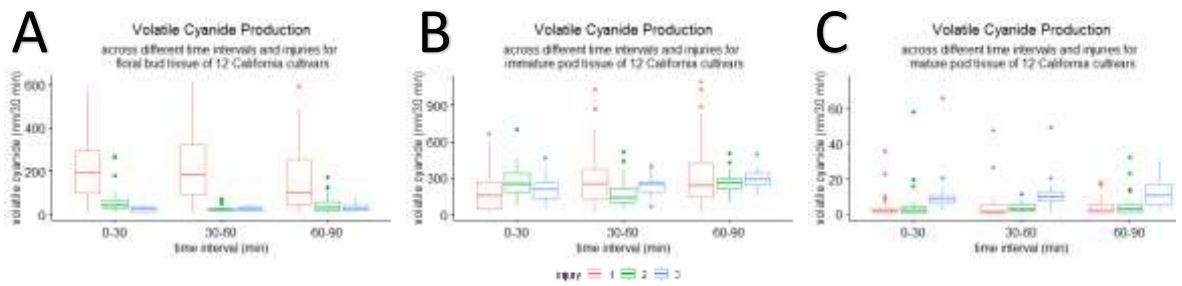

Figure 3.4 Volatile cyanide production across the 0-30, 30-60 and 60-90 minute time intervals following primary, secondary and tertiary injury for A. floral bud B. immature pod and C. mature pod tissue.

Comparing the least squares means for volatile cyanide production among reproductive tissues, immature pod tissue had the greatest production of volatile cyanide and was significantly greater than floral bud tissue, which was significantly greater than mature pod tissue and fresh seed tissue at the $p < 0.05$ significance level. The latter two tissue types were not significantly different from each other at the $p < 0.05$ significance level. Fresh seed produced the lowest levels of volatile cyanide across all varieties and required an extended time interval for measurement before differences were observed among genotypes. The least squares means for reproductive tissues are reported in Table 3.1. Comparing the least squares means for volatile cyanide production for repeated injuries, cyanide production following the

primary tissue injury was significantly greater than that after the secondary and tertiary injuries, the latter injuries not being significantly different from each other at the $p < 0.05$ significance level. There was no significant difference in volatile cyanide production across the three consecutive 30 minute time intervals tested for the floral bud, immature pod and mature pod tissue across different injuries; however, the 240 minute time interval required for the fresh seed tissue was significantly lower and the only significantly different time interval at the $p < 0.05$ significance level. Floral bud tissue had significant differences in volatile cyanide production across all three injuries at the $p < 0.05$ significance level, with a decline in volatile cyanide production following repeated injury. The production of volatile cyanide in the immature pod and mature pod tissues did not significantly differ across the different injury numbers at the $p < 0.05$ significance level. The least squares means for the interaction between tissue and injury are reported in Table 3.1.

There were significant positive correlations in the production of volatile cyanide between the floral bud and immature pod tissues (r = 0.222, $p < 0.01$) and between the immature pod and mature pod tissues (r = 0.264, $p < 0.001$). There was a significant negative correlation between the floral bud and mature pod tissues of -0.272 ($p < 0.001$). These correlations and the distribution of the volatile cyanide production in the different tissues can be observed in Figure 3.5.

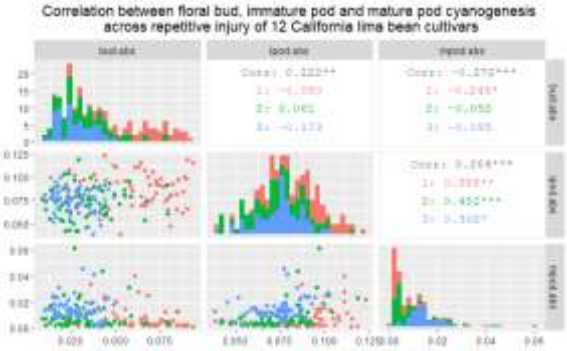

Figure 3.5 Correlations between volatile cyanide production (absorbance) in the floral bud, immature pod and mature pod tissues, and across repeated injuries. '***' denotes that p < 0.001, '**' < 0.01, '*' < 0.05

Primary injury of California-adapted cultivars across all time intervals

Another analysis of variance was conducted for the primary injury alone of the 12 California adapted cultivars and detected significant differences among reproductive tissues, among genotypes, among applications of insecticide treatments, among field and greenhouse locations, and in almost all of the genotype by environment interactions evaluated. Reproductive tissue was the most significant factor followed by genotype, treatment, the

Table 3.2 Descriptive statistics for volatile cyanide production of 12 California cultivars across different tissue types, time intervals, treatments and locations for primary injury.

Factor	F values	Factor	Level	Means nM volatile HCN
Geno	64.49***	Tissue	floral bud	190.81b
Tissue	530.02***		immature pod	250.05c
Trt	60.32***		mature pod	2.68a
Loc	33.15***		fresh seed[1]	0.05a
Time	22.40***			
Geno*Tissue	15.58***	Time	0-30 min	118b
Geno*Trt	5.48***		30-60 min	155bc
Tissue*Trt	8.30***		60-90 min	170c
Geno*Loc	4.22***		0-240 min[1]	0.05a
Tissue*Loc	20.77***			
Trt*Loc	3.32*			
Tissue*Time	39.17***	Tissue*Time	floral bud 0-30 min	202.97cd
Geno*Tissue*Trt	3.31***		floral bud 30-60 min	218.37de
Geno*Tissue*Loc	6.28***		floral bud 60-90 min	150.75bc
Geno*Trt*Loc	6.72***		immature pod 0-30 min	179.07cd
Tissue*Trt*Loc	3.08*		immature pod 30-60 min	266.91ef
Geno*Tissue*Trt*Loc	1.00ns		immature pod 60-90 min	296.86f
			mature pod 0-30 min	4.68a
CV (%)	19.31		mature pod 30-60 min	5.48a
R^2 (%)	83.09		mature pod 60-90 min	4.38a
H^2 (%)	7.77		fresh seed 0-240 min[1]	0.05a

'***' denotes that $p < 0.001$, '**' < 0.01, '*' < 0.05, '.' < 0.10, 'ns' > 0.10
'abcdef' letters represent significantly different adjusted lsmeans groups at the $p < 0.05$ significance level

interaction between tissue and time interval, location and time interval, which were all significant at the $p < 0.001$ significance level. The descriptive statistics of this analysis of variance for the primary injury across all time intervals, tissue types, genotypes, treatments and locations are included in Table 3.2.

Comparing the least squares means for volatile cyanide production in the primary injury alone among reproductive tissues, immature pod tissue had the greatest production of volatile cyanide and was significantly greater than floral bud tissue, which was significantly greater than mature pod tissue and fresh seed tissue at the $p < 0.05$ significance level. The latter two tissue types were not significantly different from each other at the $p < 0.05$ significance level. There were significant differences among the interactions of tissue type and time interval, and several overlapping interactions among floral bud and immature pod tissues time intervals that are not significantly different from each other at the $p < 0.05$ significance level. Floral bud and mature pod tissue had peak cyanide production at the 30-60 minute interval, but immature pod tissue had peak cyanide production at the 60-90 minute interval. The least squares means for

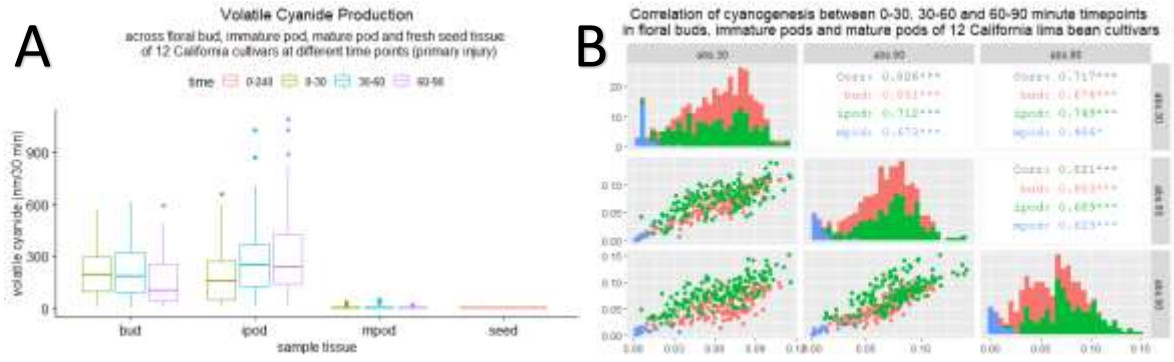

Figure 3.6 A. Volatile cyanide production for floral bud, immature pod, mature pod and fresh seed tissue across the 0-30, 30-60 and 60-90 minute time intervals following primary injury. B. Correlation of volatile cyanide production (absorbance) between the different time intervals for floral bud, immature pod and mature pod tissue. '***' denotes that p < 0.001

reproductive tissues are reported in Table 3.2 and the volatile cyanide production across reproductive tissues can be observed in Figure 3.6A.

There were significant correlations at the $p < 0.001$ significance level in the volatile cyanide production between the three time intervals both across all tissue types and within the floral bud, immature pod and mature pod tissue types. The correlation between the 0-30 and 30-60, the 30-60 and 60-90, and the 0-30 and 60-90 time intervals for volatile cyanide production was 0.826, 0.821 and 0.717, respectively. These correlations and the associated correlations within each type of reproductive tissue are shown in Figure 3.6B.

Primary injury for the initial 30-minute time interval of floral bud tissue

An analysis of variance was conducted for the volatile cyanide production of the primary injury and initial 0-30 minute time interval for the floral bud of the 12 California adapted cultivars. Significant differences were identified among genotypes, among the applications of

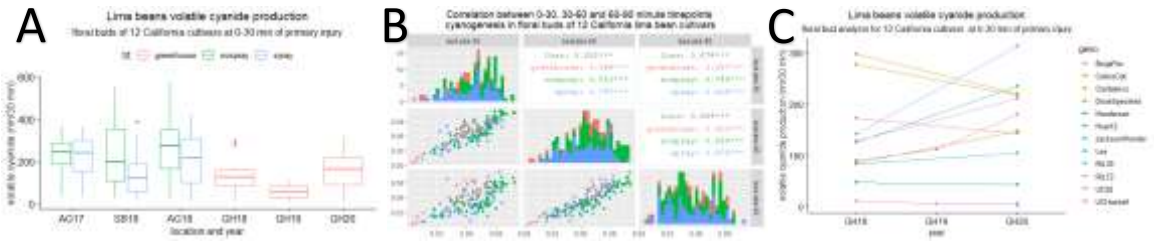

Figure 3.7 Volatile cyanide production for floral bud tissue following primary injury at the 0-30 min time interval A. Across different locations and treatments B. Correlation of volatile cyanide production (absorbance) between different time intervals across different treatments. '***' denotes that p < 0.001 C. Across different cultivars and greenhouse locations.

insecticide treatments, and among field and greenhouse locations at the $p < 0.001$ significance level, and in all of the genotype by environment interactions tested. The descriptive statistics of

these analyses of variance for floral bud for volatile cyanide production following the primary injury during the 0-30 minute time interval are included in Table 3.3, and the volatile cyanide production is observed in Figure 3.7A.

Table 3.3 Descriptive statistics for volatile cyanide production in the floral buds of 12 California cultivars across different treatments and locations for the primary injury and 0-30 time interval.

	Trait	Floral Bud HCN (primary 0-30 min)				
	units	nM volatile HCN				
	Loc	all	GH	2017-FD	2018-FD1	2018-FD2
F values	Geno	25.97***	17.72***	28.34***	6.40***	9.55***
	Trt	35.30***	-	5.43*	6.00*	10.02**
	Loc	11.04***	1.15ns	-	-	-
	Geno*Trt	3.61***	-	-	-	-
	Geno*Loc	2.21**	-	-	-	-
	Geno*Trt*Loc	1.79.	-	-	-	-
Geno means	Beija Flor	175.3bcd	153.40abc	238.3bcd	140.0abc	218.8bc
	Calico Cat	317.2e	254.95c	173.0abc	431.8d	434.3e
	Cariblanco	220.3cde	243.12bc	290.3cd	154.1abc	310.8cde
	Dixie Speckled	185.9bcde	112.00abc	212.9bcd	137.8abc	298.0bcde
	Henderson	48.7a	40.59ab	161.4abc	43.8a	41.8a
	Hopi12	175.3abcd	176.13abc	320.8d	97.7ab	220.2abcd
	Jackson Wonder	117.3abc	89.75abc	257.6d	98.1ab	120.5ab
	Lee	68.7abcd	17.82ab	101.6ab	-	-
	RIL 39	292.3de	223.37bc	367.6d	236.9abcd	442.0cde
	RIL 72	278.9de	165.20abc	377.2d	316.6bcd	307.0bcde
	UC 92	110.0ab	7.58a	34.9a	155.8abc	135.0ab
	UC Haskell	265.4de	126.86abc	328.2d	283.8cd	349.2de
Trt means	No spray	248b	-	253a	226b	296b
	Spray	185a	-	225a	155a	228a
	Greenhouse	131a	-	-	-	-
Loc means	2017-FD	233.8bc	-	-	-	-
	2018-FD1	169.7ab	-	-	-	-
	2018-FD2	242.6c	-	-	-	-
	2018-GH	120.2a	119a	-	-	-
	2019-GH	56.4abc	123a	-	-	-
	2020-GH	151.9abc	152a	-	-	-
CV (%)		10.14	9.56	3.76	16.9	16.2
R^2		0.90	0.95	0.97	0.70	0.68
H^2		0.56	0.85	0.79	0.44	0.44

'***' denotes that p < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.10, 'ns' > 0.10
'abcde' letters represent significantly different adjusted lsmeans groups at the p < 0.05 significance level

There were significant correlations among the three different time intervals for the primary injury of both the floral bud at the $p < 0.001$ significance level; these correlations were strongest across the different greenhouse environments and are shown in Figure 3.7B. There was a general conservation of genotype rank for volatile cyanide production of the 12 cultivars is displayed across the three different greenhouse environments (Figure 3.7C).

Primary injury for the initial 30-minute time interval of immature pod tissue

Another analysis of variance was conducted for the volatile cyanide production of the primary injury and initial 0-30 minute time interval for the immature pod tissue of the 12 California adapted cultivars. Significant differences were identified among genotypes, among the applications of insecticide treatments, and among field and greenhouse locations at the $p < 0.001$ significance level, and in all of the genotype by environment interactions tested. The descriptive statistics of these analyses of variance for the immature pod tissue for volatile cyanide production following the primary injury during the 0-30 minute time interval are included in Table 3.4, and the volatile cyanide production is observed in Figure 3.8A.

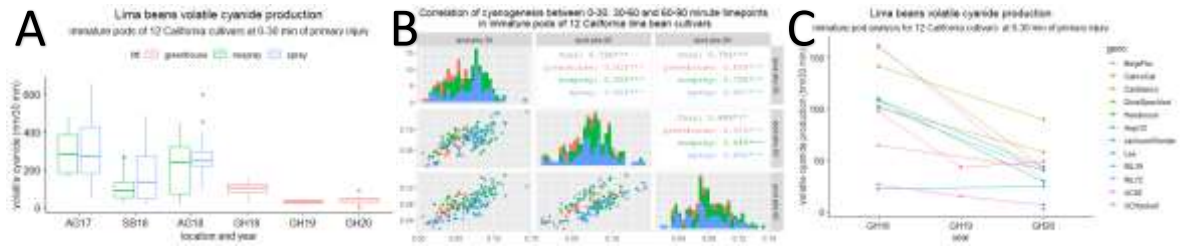

Figure 3.8 Volatile cyanide production for immature pod tissue following primary injury at the 0-30 min time interval A. Across different locations and treatments B. Correlation of volatile cyanide production (absorbance) between different time intervals across different treatments. '***' denotes that p < 0.001 C. Across different genotypes and greenhouse locations.

There were significant correlations among the three different time intervals for the primary injury of both the floral bud and immature pod tissue at the $p < 0.001$ significance level; these correlations were strongest across the different greenhouse environments (Figure

Table 3.4 Descriptive statistics for volatile cyanide production in the immature pods of 12 California cultivars across different treatments and locations for the primary injury and 0-30 time interval.

	Trait	**Immature Pod HCN (primary 0-30 min)**				
	units	nM volatile HCN				
	Loc	all	GH	2017	2018-1	2018-2
F values	Geno	10.83***	7.92***	1.15ns	4.58***	
	Trt	49.63***	-	0.45ns	4.99*	
	Loc	19.95***	15.09***	-	-	-
	Geno*Trt	2.69**	-	-	-	-
	Geno*Loc	1.55.	-	-	-	-
	Geno*Trt*Loc	2.91**	-	-	-	-
Geno means	Beija Flor	224.1bc	48.3ab	210a	224.8bc	318bc
	Calico Cat	239.1abc	96.6ab	569a	87.3abc	220ab
	Cariblanco	134.6ab	74.4ab	335a	130.9ab	177a
	Dixie Speckled	305.5c	110.3b	285a	333.4c	454c
	Henderson	153.7abc	63.4ab	211a	114.5ab	278abc
	Hopi12	85.0a	71.5ab	215a	62.6ab	118a
	Jackson Wonder	140.1ab	18.2ab	224a	155.1abc	213ab
	Lee	157abc	26.6ab	234a	-	-
	RIL 39	162.3abc	65.9ab	485a	76.8ab	233abc
	RIL 72	164.3abc	91.0ab	463a	79.3ab	171ab
	UC 92	119.1ab	16.0a	379a	83.3ab	143ab
	UC Haskell	89.2a	63.4ab	155a	40.8ab	171ab
Trt means	No spray	188.7b	-	327a	81.5a	207a
	Spray	235.7b	-	301a	158.5b	247a
	Greenhouse	69.1a	-	-	-	-
Loc means	2017-FD	298.7c	-	-	-	-
	2018-FD1	140.1b	-	-	-	-
	2018-FD2	228.9c	-	-	-	-
	2018-GH	94.6ab	96.4b	-	-	-
	2019-GH	82.0abc	51.7ab	-	-	-
	2020-GH	38.2a	38.3a	-	-	-
CV (%)		12.61	11.38	14.61	22.87	17.44
R^2		0.88	0.91	0.59	0.60	0.54
H^2		0.12	0.32	0.32	0.40	0.4

'***' denotes that p < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.10, 'ns' > 0.10
'abc' letters represent significantly different adjusted lsmeans groups at the p < 0.05 significance level

3.7B and Figure 3.8B). There was a general conservation of genotype rank for volatile cyanide production of the 12 cultivars is displayed across the three different greenhouse environments in Figure 3.7C and Figure 3.8C for floral bud and immature pod tissues, respectively.

Status diversity study (wild, landrace, escape and cultivar)

An analysis of variance for volatile cyanide production in the floral buds and immature pods of lima bean accessions according to status (wild, landrace, escape and cultivar) during the primary injury and initial 0 – 30 min time interval revealed that there were significant differences according to the accession status for both tissue types at the $p < 0.001$ significance level. When comparing the least squares means for volatile cyanide production in the floral bud tissue, the cultivar accessions had volatile cyanide production that were significantly higher than those of landrace and wild accessions at the $p < 0.05$ significance level. For the volatile cyanide production in the immature bud tissue, the volatile cyanide production of landrace accessions was significantly greater than that of the wild accessions, at the $p < 0.05$ significance level.

Table 3.5 Descriptive statistics for volatile cyanide production in the floral buds and immature pods of different lima bean statuses across greenhouse locations for the primary injury and 0-30 time interval.

	Trait	Floral Bud HCN	Immature Pod HCN
	units	nM volatile HCN	nM volatile HCN
F values	Status	9.30***	10.45***
	Loc	2.31ns	19.43***
	Status*Loc	2.23.	4.31**
Status means	Cultivar	169.4b	62.7ab
	Escape	1.7ab	-
	Landrace	84.1a	89.3b
	Wild	88.0a	46.8a
Loc means	2018-GH	72.2a	115.5b
	2019-GH	72.3a	56.6a
	2020-GH	102.8b	26.9a
CV (%)		24.88	34.73
R^2		31.06	49.9
H^2		58.72	28.88

'***' denotes that p < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.10, 'ns' > 0.10
'ab' letters represent significantly different adjusted lsmeans groups at the p < 0.05 significance level

There was a significant positive correlation between the production of volatile cyanide in the floral bud and immature pod tissues at the $p < 0.001$ significance level; the strongest correlation was observed among the wild accessions and can be seen in Figure 3.9E.

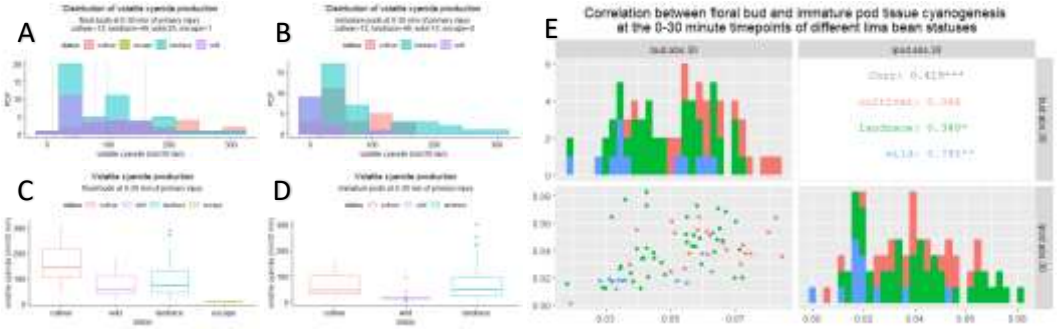

Figure 3.9 Distribution of volatile cyanide production for different statuses following primary injury at the 0-30 min time interval for A. floral bud tissue and B. immature pod tissue. Volatile cyanide production for different statuses following primary injury at the 0-30 min time interval for C. floral bud tissue and D. immature pod tissue E. Correlation between floral bud and immature pod volatile cyanide production (absorbance) across statuses. '***' denotes that p < 0.001, '**' < 0.01, '*' < 0.05

Split-Plot RIL experiment

Volatile cyanide production from floral bud and immature pod tissue in the primary injury during the 0-30 minute time interval were used to identify three significant loci controlling these traits through QTL mapping. These three QTL were incorporated in the mixed linear model to normalize the residual errors of the bimodal distributions of data, to perform an analysis of variance on these two traits in this population.

In the floral bud tissue, significant differences were detected in the UC 92 – UC Haskell recombinant inbred line population between different alleles at three relevant QTLs on chromosomes Pl05, Pl08 and Pl10, between the applications of different insecticidal treatments, between different locations, and among genotype-environment interactions for

volatile cyanide production. Results from the analysis of variance on the floral bud tissue showed that the QTL on chromosome Pl05 was the most significant factor for volatile cyanide production at the *p < 0.001* significance level, followed by the QTL on chromosome Pl08 and the interaction of the QTLs on chromosomes Pl05 and Pl10 and the QTLs on chromosomes Pl08 and Pl10 at the *p < 0.001* significance level. Insecticide treatment and location were also significant at the *p < 0.001* significance level and the interaction between the two were significant at the *p < 0.05* significance level. Effects of the different QTL allelic combinations, as well as the location and treatments are shown in Figure 3.10, and the descriptive statistics of the analysis of variance on the floral bud tissue following the primary injury during the 0-30 minute time interval are included in Table 3.6.

Table 3.6 Descriptive statistics for volatile cyanide production in the floral buds of the UC 92 – UC Haskell RIL population for the primary injury and 0-30 time interval at different locations and under different treatments.

Factor	F values			Factor	Level	Floral Bud HCN Means nM volatile HCN		
Loc	all	2018-FD1	2018-FD2	Loc		all	2018-FD1	2018-FD2
chr5.QTL	937.28***	343.32***	778.34***	RIL summary	UC 92	148.71	156.94	366.25
chr8.QTL	73.66***	34.69***	29.99***		UC Haskell	315.62	265.00	391.29
chr10.QTL	0.85ns	0.66ns	4.66*		RIL average	204.73	168.68	248.62
Trt	18.77***	4.68*	16.15***		RIL min	1.72	0.86	2.30
Loc	19.86***	-	-		RIL max	840.54	855.59	1223.04
Trt*Loc	4.54*	-	-					
chr5.QTL*chr8.QTL	0.41ns	2.47ns	4.69*	QTL means	chr5.QTL.A	48.6a	39.3a	49.9a
chr5.QTL*chr10.QTL	56.02***	38.21***	13.84***		chr5.QTL.B	389.4b	278.7b	514.2b
chr8.QTL*chr10.QTL	19.61***	11.79***	2.18ns					
chr5.QTL*Trt	0.05ns	1.16ns	1.51ns		chr8.QTL.A	283b	203b	369b
chr5.QTL*Loc	41.06***	-	-		chr8.QTL.B	155a	115a	195a
chr8.QTL*Trt	5.70*	0.59ns	6.12*					
chr8.QTL*Loc	0.001ns	-	-		chr10.QTL.A	196a	129a	255a
chr10.QTL*Trt	1.07ns	0.89ns	0.61ns		chr10.QTL.B	242a	189a	309b
chr10.QTL*Loc	0.28ns	-	-					
chr5.QTL*chr8.QTL*chr10.QTL	11.58***	6.19*	1.97ns	Trt means	Spray	180a	149a	222a
					No Spray	258b	169b	342b
CV (%)	20.58	21.71	18.91					
R^2	0.90	0.89	0.92	Loc means	2018-FD1	157a	-	-
H^2	0.89	0.91	0.95		2018-FD2	281b	-	-

'***' denotes that p < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.10, 'ns' > 0.10
'ab' letters represent significantly different adjusted lsmeans groups at the p < 0.05 significance level
A = UC 92 haplotype, B = UC Haskell haplotype.

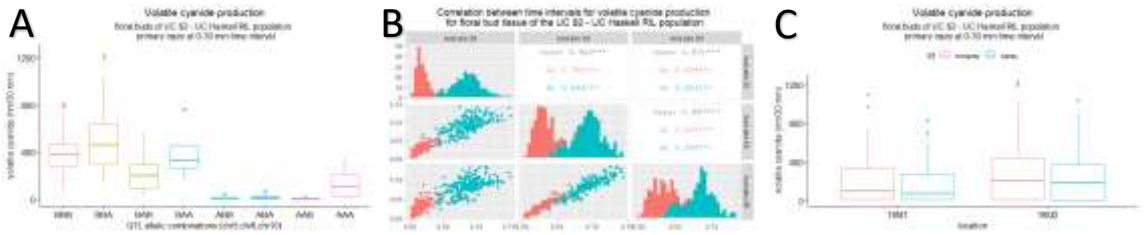

Figure 3.10 Volatile cyanide production in the floral bud tissue of the UC 92 – UC Haskell RIL population A. across the different QTL allelic combinations B. (Absorbance) correlations between different time intervals, and correlations by chromosome 5 QTL allele. '***' denotes that p < 0.001 C. across the different locations and treatments. A = UC 92 haplotype, B = UC Haskell haplotype.

In the immature pod tissue, significant differences for volatile cyanide production were detected among the UC 92 – UC Haskell recombinant inbred lines between different alleles at relevant QTLs, between the applications of different insecticidal treatments, among different locations, and among genotype-environment interactions. Results from the analysis of variance on the volatile cyanide production in immature pod tissue showed that the QTL on chromosome Pl05 was the most significant factor for the volatile cyanide production at the $p < 0.001$ significance level, and the interaction of the QTLs on chromosomes Pl05 and Pl10 and the QTLs on chromosomes Pl05 and Pl01 were significant at the $p < 0.01$ and $p < 0.05$ significance levels, respectively. Treatment was significant at the at the $p < 0.01$ significance levels, and the

Figure 3.11 Volatile cyanide production in the immature pod tissue of the UC 92 – UC Haskell RIL population A. across the different QTL allelic combinations B. (Absorbance) correlations between different time intervals, and correlations by chromosome 5 QTL allele '***' denotes that p < 0.001 C. across the different locations and treatments. A = UC 92 haplotype, B = UC Haskell haplotype.

interaction between treatment and location and the QTL on chromosome Pl05 and treatment and location were also significant at the $p < 0.10$ and $p < 0.01$ significance levels, respectively. The effects of the different QTL allelic combinations, as well as the location and treatments are shown in Figure 3.11, and the descriptive statistics of the analysis of variance on the immature pod tissue following the primary injury during the 0-30 minute time interval are included in Table 3.7.

Table 3.7 Descriptive statistics for volatile cyanide production in the immature pods of the UC 92 – UC Haskell RIL population for the primary injury and 0-30 time interval at different locations and under different treatments.

Factor	F values			Factor	Level	Immature Pod HCN Means		
				units		nM volatile HCN		
Loc	all	2018-FD1	2018-FD2	Loc		all	2018-FD1	2018-FD2
					UC 92	241.65	89.80	311.73
chr5.QTL	240.88***	120.75***	139.49***		UC Haskell	135.18	85.90	184.46
chr10.QTL	0.56ns	0.26ns	0.07ns	RIL summary	RIL average	147.77	134.72	172.70
chr11.QTL	5.30*	1.18ns	3.98.		RIL min	1.09	1.28	0.75
chr1.QTL	2.95.	0.47ns	4.25*		RIL max	691.03	923.30	1034.76
Trt	7.07**	0.40ns	11.34**					
Loc	1.34ns	-	-		chr5.QTL.A	9.78a	3.68a	7.64a
Trt*Loc	3.04.	-	-		chr5.QTL.B	271.2b	245.38b	276.74b
chr5.QTL*chr10.QTL	8.08**	3.16.	5.52*					
chr5.QTL*chr11.QTL	0.71ns	0.94ns	0.05ns		chr10.QTL.A	124a	102a	121a
chr5.QTL*chr1.QTL	5.33*	4.24*	2.68ns		chr10.QTL.B	157a	147a	148a
chr5.QTL*Trt	4.21*	0.55ns	15.61***	QTL means				
chr5.QTL*Loc	0.01ns	-	-		chr11.QTL.A	165b	142a	163a
chr5.QTL*Trt*Loc	8.97**	-	-		chr11.QTL.B	116a	107a	106a
chr5.QTL*chr10.QTL*chr11.QTL*chr1.QTL	2.91**	2.41*	1.72ns					
					chr1.QTL.A	177b	151.2a	179.4b
					chr1.QTL.B	104a	97.9a	89.7a
				Trt means	Spray	159b	122a	173.4b
CV (%)	37.15	36.72	33.74		No Spray	122a	128a	95.7a
R^2	0.75	0.73	0.80					
H^2	0.91	0.98	0.85	Loc means	2018-FD1	123a	-	-
					2018-FD2	158a	-	-

'***' denotes that p < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.10, 'ns' > 0.10.
'ab' letters represent significantly different adjusted lsmeans groups at the p < 0.05 significance level.

Comparing the least squares means in the floral bud and immature pod tissue in Tables 3.6 and 3.7, the allelic contributions of UC Haskell at the chromosome Pl05 QTL led to significant increases in the volatile cyanide production compared to the allelic contribution of UC 92, at the $p < 0.05$ significance level. The allelic contribution of UC Haskell at the

chromosome Pl10 QTL led to a significant increase in volatile cyanide production in floral bud tissue, and an insignificant increase in immature pod tissue at the $p < 0.05$ significance level. For the chromosome Pl08 QTL, the allelic contribution of UC 92 led to a significant increase in the volatile cyanide production in floral bud tissue only and was insignificant in immature pod tissue, at the $p < 0.05$ significance level. The effects of these QTL allelic combinations can be observed in Figure 3.10A and Figure 3.11A for the floral bud and immature pod tissues, respectively. The broad sense heritability of volatile cyanide production was high for both reproductive tissue types, ranging between 89 - 95% for the floral bud tissue and 79 95% for the immature pod tissue. The broad sense heritability, H^2, estimates for floral bud and immature pod tissue are included in Tables 3.6 and 3.7, respectively.

Comparing the least squares means in the floral bud and immature pod tissues, there was a significant increase in the volatile cyanide production for the unsprayed treatments compared to the sprayed treatments at the $p < 0.05$ significance level, which was consistent across all locations. There was a 43% and 51% increase in volatile cyanide production from sprayed to unsprayed treatments, for the floral bud and immature pod tissue, respectively. There was a significant increase in the volatile cyanide production at the 2018-FD2 location compared to the 2018-FD1 location at the $p < 0.05$ significance level, which resulted in an increase in volatile cyanide production of 79% and 63% for the floral bud and immature pod tissue, respectively.

There was a positive correlation of 77% between the volatile cyanide production of the floral bud and immature pod tissue in the recombinant inbred line population across treatments, which was significant at the *p < 0.001* significance level. There was also a positive correlation of 25% between yield and volatile cyanide production in the immature pod tissue, for RILs of the lower volatile cyanide producing UC 92 haplotype at the peak QTL on chromosome Pl05, which is significant at the *p < 0.01* significance level. There was a negative correlation of 18% between yield and volatile cyanide production in the floral bud tissue, for RILs of the higher volatile cyanide producing UC Haskell haplotype at the peak QTL on chromosome Pl05, which was significant at the *p < 0.05* significance level. Correlations between yield and volatile cyanide production in the floral bud and immature pod tissue are shown in Figure 3.12.

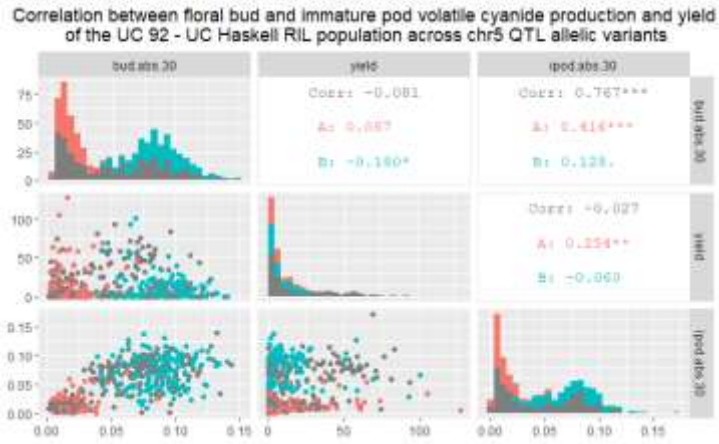

Figure 3.12 Correlations between yield (grams plant^{-1}) and volatile cyanide production (absorbance) in floral bud and immature pod tissue for the UC 92 – UC Haskell RIL population. A = UC 92 haplotype, B = UC Haskell haplotype at the chromosome 5 QTL peak.

Another correlation analysis was performed between yield, yield reduction, and the volatile cyanide production in the floral bud tissue and the immature pod tissue for genotypic averages in the two different insecticide treatments. There were no significant correlations between yield reduction with volatile cyanide production in either tissues or either treatments. The two significant correlations identified between yield and volatile cyanide production was between the immature pod tissue in the no-spray treatment and yield in the spray treatment

and yield in the no-spray treatment, which were significant at the $p < 0.05$ and $p < 0.10$ significance levels, respectively, and only for the RILs that contained the UC 92 haplotype at the peak QTL for chromosome Pl05. The correlations of volatile cyanide production in floral bud and immature pod tissue with yield and yield reduction are shown in Figure 3.13.

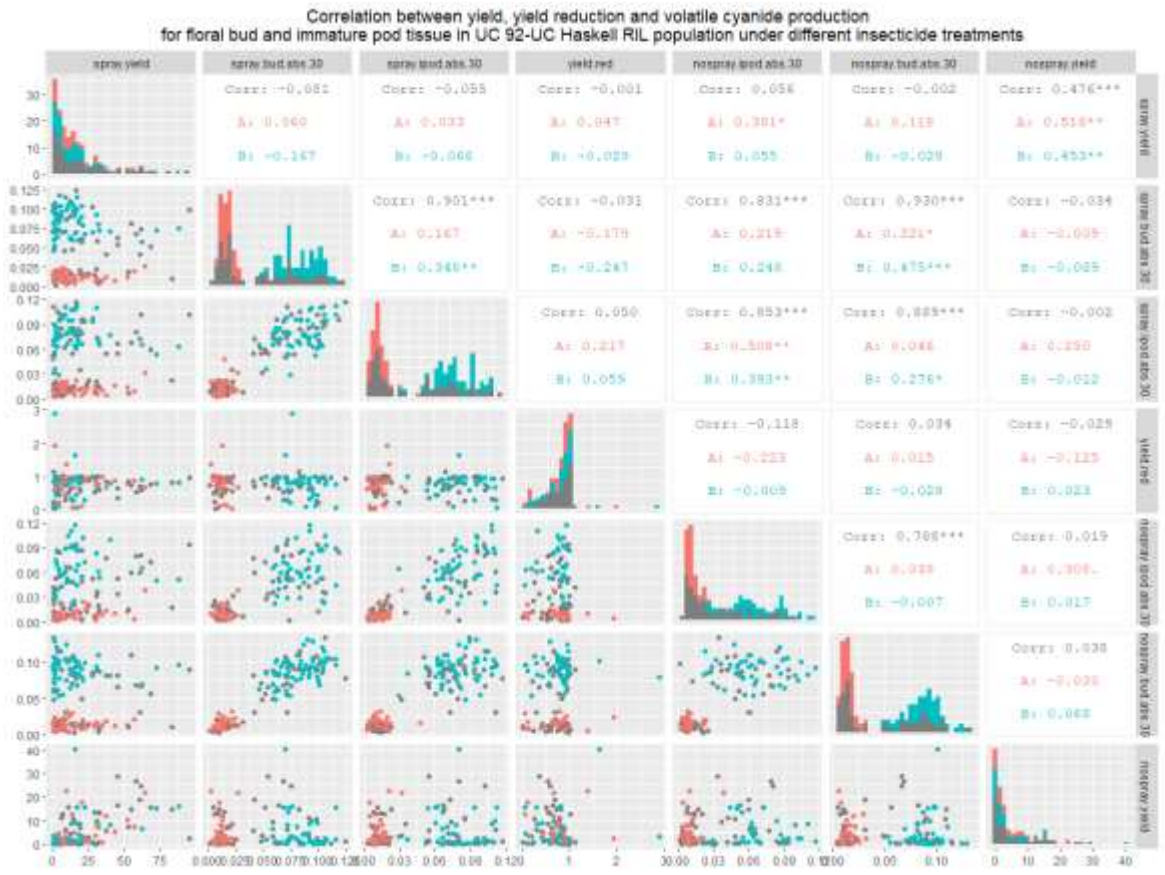

Figure 3.13 Correlations between genotypic averages of yield (grams plant^{-1}), percent yield reduction and volatile cyanide production (absorbance) in floral bud and immature pod tissue for the UC 92 – UC Haskell RIL population under a 'spray' and 'no spray' insecticide treatment.
A = UC 92 haplotype, B = UC Haskell haplotype at the chromosome 5 QTL peak.

QTL analysis of the UC 92 – UC Haskell RIL population

QTL analyses were run for volatile cyanide production in the floral bud and immature pod tissue for the primary injury and the 0-30 min time interval. Multiple significant QTL peaks and significant QTL interactions were identified for both the floral bud and immature pod tissue volatile cyanide production (Figure 3.14 and Table 3.8).

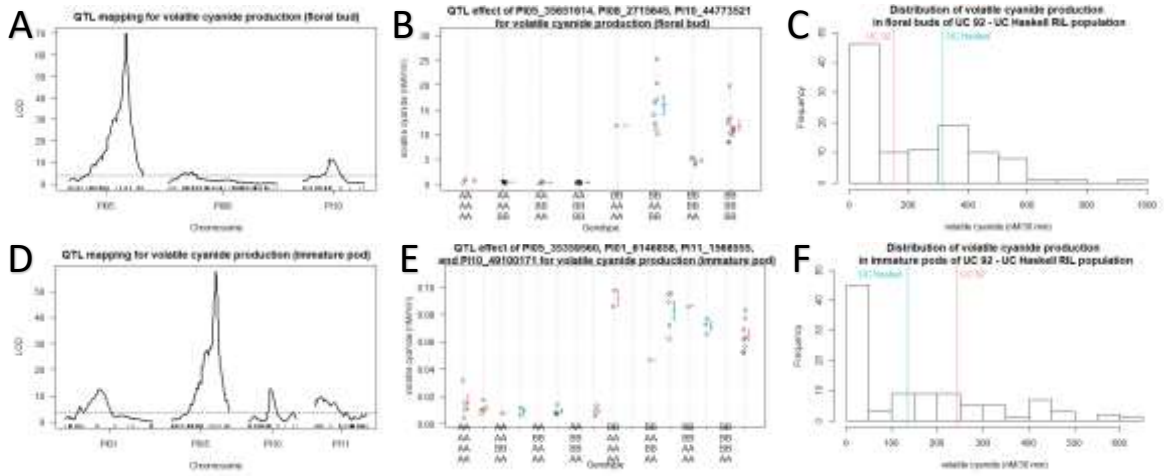

Figure 3.14 QTL peaks for volatile cyanide production in A. floral bud and D. immature pod tissue. QTL effects of the closest markers to the QTL peaks of volatile cyanide production in B. floral bud and E. immature pod tissue. Distribution of volatile cyanide production in the UC 92 – UC Haskell RIL population for C. floral bud and F. immature pod tissue.

QTL mapping identified the most significant peaks for volatile cyanide production in both the floral bud tissue and immature pod tissue in the same region of chromosome Pl05; the peak areas were around 200 kbp and only 1 cM apart for this recombinant inbred line population. The percent of the phenotypic variation explained for volatile cyanide production by the major QTL on chromosome Pl05 is 93% and 87% in the floral bud and immature pod tissue, respectively. The positive allele is contributed by UC Haskell in both tissue types and

leads to an increase in 134 and 106 nM of volatile HCN 30 min^{-1} in the floral bud and immature pod tissue, respectively.

Table 3.8 Significant QTLs, flanking markers around QTL peak, LOD scores, percent phenotypic variation explained, QTL effects, and the positive effect allele for volatile cyanide production (nM HCN 30 min^{-1}) in the floral bud and immature pod tissue for the primary injury 0-30 min time interval.

Trait	Tissue	QTL Peak	Flanking Markers of QTL Peak	LOD	% PVE	QTL Effect (nM HCN 30 min^{-1})	QTL Effect (+) Allele
Volatile Cyanide Production (nM HCN 30 min^{-1})	Floral Bud	Pl05.57.4	Pl05_35520814 - Pl05_35804750	68.33	93.26	133.81	UC Haskell
		Pl10.27	Pl10_46286504 - Pl10_46521900	9.98	2.79	5.21	UC Haskell
		Pl08.24.6	Pl08_2715645 - Pl08_2791797	5.98	1.53	0.92	UC 92
		Pl05.57.4 * Pl10.27	-	4.74	1.18	-	-
		Pl05.57.4 * Pl08.24.6	-	1.53	0.35	-	-
	Immature Pod	Pl05.56.8	Pl05_35359560 - Pl05_35364729	53.84	88.69	106.12	UC Haskell
		Pl10.29	Pl10_49100171 - Pl10_50371482	8.30	3.72	2.87	UC Haskell
		Pl01.42	Pl01_6146858 - Pl01_10207081	8.04	3.58	6.62	UC 92
		Pl11.17	Pl11_1100133 - Pl11_1568555	5.91	2.50	0.63	UC Haskell
		Pl05.56.8 * Pl10.29	-	4.25	1.72	-	-
		Pl05.56.8 * Pl01.42	-	2.37	0.92	-	-
		Pl05.56.8 * Pl11.17	-	1.85	0.71	-	-

Only QTL interactions significant at the $p < 0.05$ significance level are included.

The second most significant QTL peaks for volatile cyanide production in both the floral bud tissue and immature pod tissue also collocated to the same region of chromosome Pl10; they were around 3 Mbp and 2 cM apart in this recombinant inbred line population. This minor QTL explained 3% and 4% of the phenotypic variation for volatile cyanide production in the floral bud and immature pod tissue, respectively. Similar to the major QTL on chromosome Pl05, the positive allele for this minor QTL was contributed by the UC Haskell parent as well and led to an increase of 5% and 3% nM of volatile HCN 30 min^{-1} in the floral bud and immature pod tissues, respectively.

The two QTLs for volatile cyanide production on chromosomes Pl05 and Pl10, combined with the other minor QTLs – on chromosome Pl08 for floral bud tissue and on chromosomes

Pl01 and Pl11 for immature pod tissue - explained over 98% and 99% of the phenotypic variation, respectively. All of the interactions between the major QTL on chromosome Pl05 with the minor QTLs were significant at the $p < 0.05$ significance level in the multiple QTL model established for both tissues volatile cyanide production (Table 3.8).

Results of the principal component analysis (Figure 3.15) suggest that the haplotype groups of the major QTL on chromosome Pl05 are responsible for over 95% of the principal variation between floral bud and immature pod volatile cyanide production. The UC Haskell haplotype group has greater variation than the UC 92 haplotype group across both the PC1 and PC2 axes of the analysis.

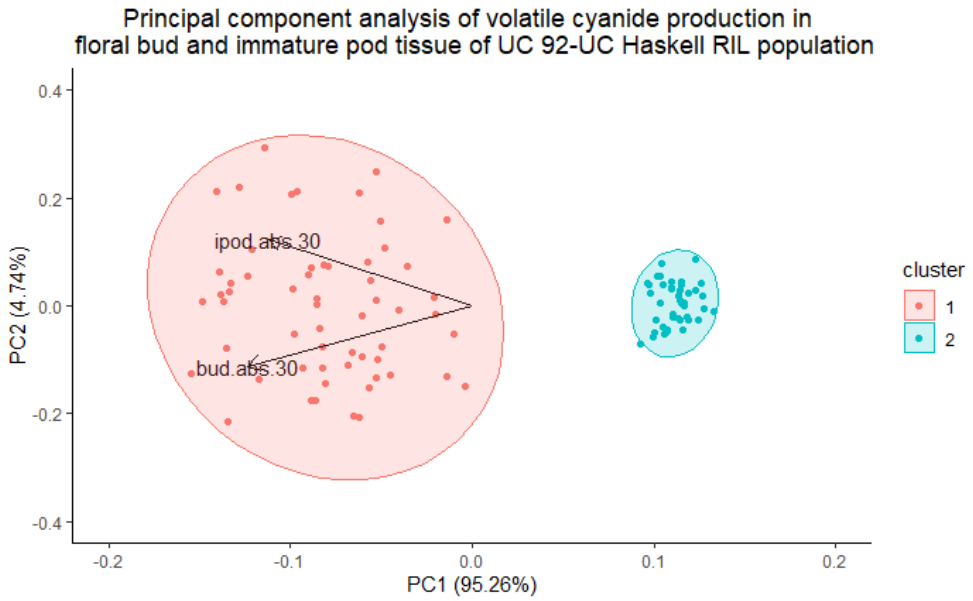

Figure 3.15 Principal component analyses between volatile cyanide production in the floral bud tissue and the immature pod tissue of the UC 92 – UC Haskell RIL population with clustering of allelic variance at the QTL peak on chromosome 5. 1 = UC Haskell haplotype, 2 = UC 92 haplotype.

Discussion

This research evaluates the differences in volatile cyanide production in different reproductive tissues of lima beans, following repeated injury, across different time intervals post-injury, in different environments, for different diversity statuses (wild, landraces, and cultivars), and different genotypes of lima beans. This research also evaluates the effects of lygus pressure on volatile cyanide production and the impact on lima bean yield by analyzing a recombinant inbred line population.

Differences in volatile cyanide production in different reproductive tissues

The results of this study demonstrate the significant differences in the volatile cyanide production of different reproductive tissues of lima bean cultivars. The production of volatile cyanide was higher in the immature pod and floral bud tissue and dropped dramatically and significantly in the mature pod and fresh seed tissue. Previous research reports the mobilization and translocation of cyanogenic compounds from senescing to developing tissue, which may explain why there is a decline in volatile cyanide production in the mature pod and fresh seed tissue in the cultivars included in this study (Ballhorn et al., 2008; Siritunga et al., 2004). Variability in cyanogenesis expression in plants vary with the developmental stages of the plant or the tissue studied and is supported in our findings (Ballhorn et al., 2006; Frehner et al., 1990; Schappert & Shore, 2000). While there were significant differences in the volatile cyanide produced among the different tissue types, there was a strong positive correlation in the volatile cyanide production between floral bud and immature pod tissue.

Within each of the different tissue types analyzed for volatile cyanide production across the California-adapted cultivars, significant differences were identified across genotypes and environments supporting the genotypic variability and phenotypic plasticity of cyanogenesis in lima beans (Ballhorn et al., 2006, 2008). Data for these analyses were collected across different years, pesticide applications and in different field and greenhouse locations, with the greatest stability and highest heritability of volatile cyanide production observed across the greenhouse locations.

The volatile cyanide production declined from the initial tissue injury following repeated secondary and tertiary injury in the floral bud and immature pod tissues. Previous studies have clarified the cyanogenic potential and cyanogenic capacity of plant tissue, which are dictated by the cyanogenic glucosides stored within the plant tissue and the activity of the β-glucosidase enzyme in releasing volatile cyanide as a product, respectively (Conn, 1969; Frehner et al., 1990; Poulton, 1990). Our results support the limitations of volatile cyanide production, likely by the storage and availability of cyanogenic glucosides as the limiting resource, following repeated tissue damage in lima beans young reproductive tissue.

Differences in diversity status

There was no significant difference in the expression of volatile cyanide production in the floral bud tissue between landraces and wild accessions; however, the California-adapted cultivars had greater variation in their volatile cyanide production, which exceeded the landrace and wild accessions in the floral bud and immature pod tissue. While previous research has not studied the reproductive tissues of lima bean, studies done on the cyanogenic potential of lima

bean leaves supports the greater range of cyanogenesis for cultivars compared to wild accessions of both Andean and Mesoamerican origin (Ballhorn et al., 2008). While cyanogenic compounds in the edible seed tissue have been selected against throughout the domestication process of lima beans, cyanogenic production in other tissues may have been selected for as a mechanism of herbivore resistance (Ballhorn et al., 2008).

QTL mapping and candidate genes

There was significant variability in the volatile cyanide production in the floral bud and immature pod tissue of the UC 92 – UC Haskell population, resulting in bimodal distributions for both traits. Results of this research have identified a major locus for both traits, which underlie the bimodality of the distribution for both traits and are collocated in the same region of chromosome Pl05. Assuming a single locus with pleiotropic effect on volatile cyanide production in the two tissues, this locus accounts for 93% and 89% of the phenotypic variation explained for floral bud and immature pod volatile cyanide production, respectively, and may contribute strongly to the high heritability of these traits, ranging from 79-95%, across different environments.

This major locus on chromosome Pl05 in the UC 92 – UC Haskell recombinant inbred line population maps to a region on the lima bean reference genome encoding an annotated β-glucosidase cyanogenic gene, Pl05G0000264700.v1, and nine additional β-glucosidase genes annotated in this region. β-glucosidase enzymes react with cyanogenic glucoside molecules following tissue damage and are important in the production of volatile hydrogen cyanide (Figure 3.2). These ten annotated β-glucosidase genes cluster in a 162 kb region on the long

arm of chromosome Pl05, with eight of the nine β-glucosidase genes ranging from 2059-3271 bp in size and the ninth gene, Pl05G0000264400.v1, spanning nearly 64 kb (Figure 3.16). The peak QTL markers for volatile cyanide production in floral bud and immature pod tissue are located 100kb upstream and downstream from the cluster of these ten β-glucosidase genes, respectively. A descriptive summary of the candidate model genes for this major locus on chromosome Pl05 is included in Table 3.9.

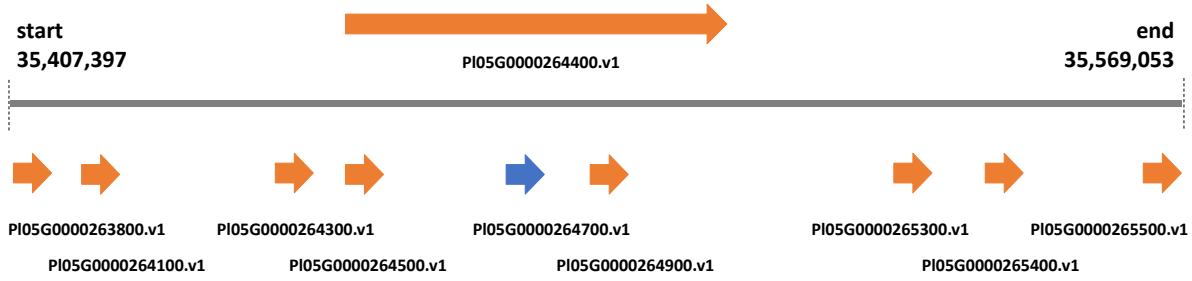

Figure 3.16 Cluster of nine annotated β-glucosidase genes on the long arm of chromosome 5 on the *Phaseolus lunatus* reference genome. The orange references the BGL24 gene family of β-glucosidases and the blue references the BGLT gene family of cyanogenic β-glucosidases.

The second most significant locus for volatile cyanide production in both floral bud and immature pod tissue both collocated on the long arm of chromosome Pl010 and explained 3% and 4% of the phenotypic variation, respectively. This minor locus identified in the UC 92 – UC Haskell recombinant inbred line population mapped to a region on the lima bean reference genome encoding three annotated β-glucosidase gene models within 6 Mbp of the peak QTL (Table 3.9).

Table 3.9 Candidate genes, function and location significant QTL for volatile cyanide production in the floral bud and immature pod tissue of the UC 92 – UC Haskell recombinant inbred line population.

Trait	Tissue	Chromosome	Gene	Gene Family and Function	Start (bp)	End (bp)	Length (bp)	Exons
Volatile Cyanide Production	Floral Bud and Immature Pod	5	PI05G0000263800.v1	BGL24: beta-glucosidase	35407397	35410281	2884	13
		5	PI05G0000264100.v1	BGL24: beta-glucosidase	35417790	35420213	2423	11
		5	PI05G0000264300.v1	BGL24: beta-glucosidase	35437573	35440489	2916	13
		5	PI05G0000264400.v1	BGL24: beta-glucosidase	35453032	35516811	63779	26
		5	PI05G0000264500.v1	BGL24: beta-glucosidase	35447863	35449922	2059	8
		5	PI05G0000264700.v1	BGLT: cyanogenic beta-glucosidase	35484472	35486972	2500	6
		5	PI05G0000264900.v1	BGL24: beta-glucosidase	35496238	35499146	2908	13
		5	PI05G0000265300.v1	BGL24: beta-glucosidase	35542544	35545712	3168	24
		5	PI05G0000265400.v1	BGL24: beta-glucosidase	35552244	35555166	2922	13
		5	PI05G0000265500.v1	BGL24: beta-glucosidase	35565782	35569053	3271	13
	Floral Bud and Immature Pod	10	PI10G0000236000.v1	BGH3B: beta-glucosidase	40043289	40046173	2884	9
		10	PI10G0000255300.v1	BGL10: beta-glucosidase	42829224	42834084	4860	5
		10	PI10G0000348100.v1	E1314: beta-glucosidase	53150492	53152300	1808	3
	Floral Bud	8	PI08G0000013100.v1	E13B: beta-glucosidase	1134273	1137037	2764	4
		8	PI08G0000031800.v1	3AT1: anthocyanidin 3-O-glucoside	2510853	2512256	1403	1
		8	PI08G0000034100.v1	MAT1: phenolic glucoside malonyltransferase	2662322	2665632	3310	3
		8	PI08G0000034200.v1	PMAT2: phenolic glucoside malonyltransferase	2673822	2675169	1347	1
		8	PI08G0000034300.v1	PMAT2: phenolic glucoside malonyltransferase	2679155	2680622	1467	1
		8	PI08G0000034400.v1	PMAT2: phenolic glucoside malonyltransferase	2682617	2701703	19086	2
	Immature Pod	1	PI01G0000067600.v1	E134: beta-glucosidase	6339660	6342716	3056	4
		1	PI01G0000068400.v1	E134: beta-glucosidase	6379454	6383801	4347	9
		1	PI01G0000069100.v1	E135: beta-glucosidase	6437444	6440884	3440	2
		1	PI01G0000099200.v1	BGL24: beta-glucosidase	10080308	10126327	46019	13
		1	PI01G0000099500.v1	BGL24: beta-glucosidase	10123081	10126101	3020	14
	Immature Pod	11	PI11G0000027000.v1	E13K: beta-glucosidase	1887441	1889094	1653	2
		11	PI11G0000052700.v1	E137: beta-glucosidase	3845011	3849669	4658	5
		11	PI11G0000063200.v1	BGL12: beta-glucosidase	4661408	4664492	3084	13
		11	PI11G0000063400.v1	BGL12: beta-glucosidase	4678376	4683518	5142	13
		11	PI11G0000063500.v1	BGL24: beta-glucosidase	4695681	4698874	3193	13
		11	PI11G0000063800.v1	BGL24: beta-glucosidase	4705537	4710048	4511	12
		11	PI11G0000063900.v1	BGL24: beta-glucosidase	4712724	4718655	5931	13
		11	PI11G0000064000.v1	BGL24: beta-glucosidase	4728187	4732146	3959	13
		11	PI11G0000064100.v1	BGL12: beta-glucosidase	4738590	4743883	5293	13
		11	PI11G0000064200.v1	BGL12: beta-glucosidase	4746761	4750704	3943	13

In addition to the shared QTL on chromosomes Pl05 and Pl10 for volatile cyanide production in floral bud and immature pod tissue, there were other minor loci identified for the different tissue types. A minor locus for volatile cyanide production in floral bud tissue was located on the short arm of chromosome Pl08 and mapped within 1.5 Mbp of an annotated gene model for a β-glucosidase gene and five annotated gene models involved in cyanogenic glucoside synthesis. Two additional minor loci for volatile cyanide production in immature pod tissue mapped to regions with multiple β-glucosidase annotated gene models on the short arms of chromosome Pl01 and Pl11, within the QTL peak and within 3 Mbp of the QTL peak, respectively.

For the volatile cyanide production in floral bud tissue, the UC Haskell allele had a positive additive effect for the QTLs located on chromosomes Pl05 and Pl10, whereas the UC 92 allele has a positive additive effect for the QTL on chromosome Pl08. Additionally, there were multiple significant epistatic interactions among the QTLs, particularly between the QTLs on chromosomes Pl05 and Pl10, between the QTLs on chromosomes Pl08 and Pl10 and among all three of the QTLs on chromosomes Pl05, Pl08, and Pl10. For the volatile cyanide production in immature pod tissue, the UC Haskell allele has a positive additive effect for the QTLs located on chromosomes Pl05, ,Pl10 and Pl11, whereas the UC 92 allele has a positive additive effect for the QTL on chromosome Pl01. Additionally, there were multiple significant epistatic interactions among the QTLs, particularly between the QTLs on chromosomes Pl05 and Pl10, between the QTLs on chromosomes Pl05 and Pl01, and among all four of the QTLs on chromosomes Pl05, Pl10, Pl11, and Pl01. For both the floral bud and immature pod tissues, the epistasis among QTLs for volatile cyanide production within the recombinant inbred line

population appeared to be responsible for the transgressive nature of the segregation (Figure 3.14C and 3.14F).

Cyanogenesis is contingent upon the presence of both the cyanogenic glucoside substrate, predominantly linamarin and to a lesser extent lotaustralin in lima beans, and a β-glucosidase enzyme (Butlers & Conn, 1964; Lai et al., 2020). While this population was able to identify significant QTLs collocating with gene models for multiple β-glucosidase enzymes necessary for cyanogenesis, there was only a single minor locus for the volatile cyanide production in the immature pod tissue alone that maps near annotated genes for cyanogenic glucoside synthesis. However, this locus also contains an annotated gene model for beta-glucosidase, suggesting this recombinant inbred line population may not segregate for any cyanogenic glucosides necessary for volatile cyanide production in its reproductive tissues.

Phenotypic plasticity

The variability of volatile cyanide production across genotypes and the correlation of volatile cyanide production with yield in the recombinant inbred line population, suggests that this trait may be a defense mechanism against herbivory for lygus in lima beans, as has been shown with leaf feeding herbivores in lima beans (Ballhorn et al., 2008). The increase in volatile cyanide production in the floral bud tissue in field environments that did not control lygus pressure with insecticide treatments, supports the inducibility of cyanogenesis in the presence of herbivore pressure reported in other research (Ballhorn et al., 2006, 2009; Frehner et al., 1990). Additionally, volatile cyanide production was greater in tissues collected in field environments compared to tissues collected in greenhouse environments, suggesting that general abiotic/biotic stressors of an outdoor environment may induce volatile cyanide

production in the immature reproductive tissue in lima beans. Volatile cyanide production was variable across locations and years in field environments, while the greenhouse environment was most consistent for the expression of volatile hydrogen cyanide across years. While there are genotypic differences in the expression of volatile cyanide in the immature reproductive tissues of lima bean, this trait also maintains phenotypic plasticity based on the stressors present in the environment.

High-throughput cyanide quantification

This research presents novel results on volatile cyanide production in lima bean reproductive tissues, but also introduces a method to enable high-throughput quantification of volatile cyanide production using automated data imagery analysis. This new protocol builds upon the classic Feigl-Anger method of volatile cyanide quantification but allows for high-throughput scaling of samples by adapting the sampling to a 96-well plate and reducing labor and time requirements by replacing any maceration of the tissue samples with a simple freeze-thaw method to trigger volatile cyanide production. Previous studies have been limited in the ability to quantify the results of the Feigl-Anger method, but by utilizing an automated data-imagery analysis the limitations of a qualitative and typically binary (cyanogenic/acyanogenic) analysis can be translated into quantitative data. The availability of quantitative data for volatile cyanide production allows for a more robust analysis that draws more accurate conclusions for any experiment.

Conclusion

This work presents the results of multiple studies, including a multi-year and -location study of California-adapted cultivars, different diversity statuses, and a split-plot analysis of the effects of insecticidal applications and lygus pressure on a segregating population of lima beans on the production of volatile cyanide in different reproductive tissues. The results confirm that the expression of volatile hydrogen cyanide upon tissue damage is influenced by the environment, both locations and insecticidal applications/lygus pressure, and demonstrates the phenotypic plasticity of this trait. The results show significant genotypic variability in the volatile cyanide production of different reproductive tissues and that volatile cyanide production is highly heritable in the floral bud and immature pod tissue. For floral bud and immature pod volatile cyanide expression, over 98% and 99% of the phenotypic variation is explained by three QTL and four QTL, respectively, two of which collocate for the major QTL on chromosome Pl05 and the minor QTL on chromosome Pl10. The high heritability of volatile cyanide production, the transgressive expression in the recombinant inbred population, and the correlation in volatile cyanide production in the reproductive tissues with yield supports volatile cyanide production as a promising mechanism for lygus resistance in lima beans. The QTLs identified offer hope in utilizing marker-assisted selection for these traits with further breeding potential to be achieved to confer a level of resistance to lygus damage in lima beans.

References

Ballhorn, D. J., Heil, M., & Lieberei, R. (2006). Phenotypic plasticity of cyanogenesis in lima bean Phaseolus lunatus - Activity and activation of β-glucosidase. *Journal of Chemical Ecology*, *32*(2), 261–275. https://doi.org/10.1007/s10886-005-9001-z

Ballhorn, D. J., Kautz, S., Heil, M., & Hegeman, A. D. (2009). Cyanogenesis of wild lima bean (Phaseolus lunatus L.) is an efficient direct defence in nature. *PLoS ONE*, *4*(5), 1–7. https://doi.org/10.1371/journal.pone.0005450

Ballhorn, D. J., Kautz, S., Lion, U., & Heil, M. (2008). Trade-offs between direct and indirect defences of lima bean (Phaseolus lunatus). *Journal of Ecology*, *96*(5), 971–980. https://doi.org/10.1111/j.1365-2745.2008.01404.x

Ballhorn, D. J., Lieberei, R., & Ganzhorn, J. U. (2005). Plant cyanogenesis of Phaseolus lunatus and its relevance for herbivore-plant interaction: The importance of quantitative data. *Journal of Chemical Ecology*, *31*(7), 1445–1473. https://doi.org/10.1007/s10886-005-5791-2

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Broman, K. W. (2009). *Statistics for Biology and Health A Guide to QTL Mapping with R / qtl*. https://doi.org/10.1007/978-0-387-92125-9

Butlers, G. W., & Conn, E. E. (1964). *Biosynthesis of the Cyanogenic Glucosides Linamarin and Lotaustralin I. LABELING STUDIES IN VIVO WITH LINUM USITATISSIMUM *. The Journal of Biological Chemistry* (Vol. 239). Retrieved from http://www.jbc.org/

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005, January). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*. https://doi.org/10.1007/s10681-005-1681-5

Conn, E. E. (1969). Cyanogenic glycosides. *Journal of Agricultural and Food Chemistry*, *17*(3), 519–526. https://doi.org/10.1021/jf60163a014

Dohle, S. (2017). *Development of Resources for Lima Bean (Phaseolus lunatus) Breeding and Genetics Research*. University of California, Davis.

Fox, J., & Weisbery, S. (2019). *An R Companion to Applied Regression - John Fox, Sanford Weisberg - Google Books*. *Sage Publications*. Retrieved from https://books.google.com/books?hl=en&lr=&id=uPNrDwAAQBAJ&oi=fnd&pg=PP1&dq=fox+weisberg+car&ots=MwF64I2A61&sig=PZXVVsfUb19NftmcF9vV1uuahhU#v=onepage&q=fox weisberg car&f=false

Frehner, M., & Conn, E. E. (1987). The Linamarin β-Glucosidase in Costa Rican Wild Lima Beans (Phaseolus lunatus L.) Is Apoplastic . *Plant Physiology*, *84*(4), 1296–1300. https://doi.org/10.1104/pp.84.4.1296

Frehner, M., Scalet, M., & Conn, E. E. (1990). Pattern of the Cyanide-Potential. *Plant Physiology*, *94*, 28–34.

Halkier, B. A., & Du, L. (1997). The biosynthesis of glucosinolates. *Trends in Plant Science*, *2*(11), 425–431. https://doi.org/10.1016/S1360-1385(97)01128-X

Holland, J. B., Nyquist, W. E., & Cervantes-Martinez, C. T. (2003). Estimating and interpreting heritability for plant breeding: An update. *Plant Breeding Reviews*, *22*. Retrieved from https://www.google.com/search?q=holland+et+al+2003+heritability&rlz=1C1CHBF_enUS767US767&oq=holland&aqs=chrome.0.69i59j46l4j69i60j69i61j69i60.1248j0j7&sourceid=chrome&ie=UTF-8

Jones, D. A. (1998). Why are so many food plants cyanogenic? *Phytochemistry*, *47*(2), 155–162. https://doi.org/10.1016/S0031-9422(97)00425-1

Lai, D., Maimann, A. B., Macea, E., Ocampo, C. H., Cardona, G., Pičmanová, M., … Rook, F. (2020). Biosynthesis of cyanogenic glucosides in Phaseolus lunatus and the evolution of oxime-based defenses. *Plant Direct*, *4*(8). https://doi.org/10.1002/pld3.244

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*. https://doi.org/10.18637/jss.v069.i01

Long, R., Temple, S., Meyer, R., Schwankl, L., Godfrey, L., Canevari, M., & Roberts, P. (2014). Lima Bean Production in California. *Lima Bean Production in California*, (May). https://doi.org/10.3733/ucanr.8505

Lüdecke, D. (2020). Package 'sjstats'. Collection of convenient functions for common statistical computations. *Statistical Functions for Regression Models*.

Miller, J. M., & Conn, E. E. (1980). Metabolism of Hydrogen Cyanide by Higher Plants. *Plant Physiology*, *65*(6), 1199–1202. https://doi.org/10.1104/pp.65.6.1199

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Poulton, J. E. (1990). *Cyanogenesis in Plants1*. *Plant Physiol* (Vol. 94). Retrieved from www.plantphysiol.org

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundatrion for Statistical Computing. Retrieved from http://www.r-project.org/

Schappert, P. J., & Shore, J. S. (2000). Cyanogenesis in Turnera ulmifolia L. (Turneraceae): II. Developmental expression, heritability and cost of cyanogenesis. *Evolutionary Ecology Research*, *2*, 337–352. Retrieved from http://www.evolutionary-ecology.com/abstracts/v02/1183.html

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012, July 28). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*. Nature Publishing Group. https://doi.org/10.1038/nmeth.2089

Scott, D. R. (1977). An annotated list of host plants for Lygus hesperus Knight. *Bull. Entomol. Soc. Am*, *23*, 19–22.

Shlichta, J. G., Glauser, G., & Benrey, B. (2014). Variation in Cyanogenic Glycosides Across Populations of Wild Lima Beans (Phaseolus lunatus) Has No Apparent Effect on Bruchid Beetle Performance. *Journal of Chemical Ecology*, *40*(5), 468–475. https://doi.org/10.1007/s10886-014-0434-0

Siritunga, D., Arias-Garzon, D., White, W., & Sayre, R. T. (2004). Over-expression of hydroxynitrile lyase in transgenic cassava roots accelerates cyanogenesis and food detoxification. *Plant Biotechnology Journal*, *2*(1), 37–43. https://doi.org/10.1046/j.1467-7652.2003.00047.x

Swezey, S. L., Nieto, D. J., & Bryer, J. A. (2007). Control of Western Tarnished Plant Bug Lygus hesperus Knight (Hemiptera: Miridae) in California Organic Strawberries Using Alfalfa Trap Crops and Tractor-Mounted Vacuums. *Environmental Entomology*, *36*(6), 1457–1465. https://doi.org/10.1093/EE/36.6.1457

Takos, A., Lai, D., Mikkelsen, L., Hachem, M. A., Shelton, D., Motawia, M. S., … Rook, F. (2010). Genetic screening identifies cyanogenesis-deficient mutants of Lotus japonicus and reveals enzymatic specificity in hydroxynitrile glucoside metabolism. *Plant Cell*, *22*(5), 1605–1619. https://doi.org/10.1105/tpc.109.073502

USDA ERS. (2004). *Economic research service*. *Choice Reviews Online* (Vol. 42). https://doi.org/10.5860/choice.42-2206

USDA NASS. (2019). *USDA/NASS QuickStats*. Retrieved from https://quickstats.nass.usda.gov/results/9979DE70-2420-3A8E-A718-4D3E4C830346

Vetter, J. (2000). Plant cyanogenic glycosides. *Toxicon*, *38*(1), 11–36. https://doi.org/10.1016/S0041-0101(99)00128-2

Walling, L. L. (2000). The myriad plant responses to herbivores. *Journal of Plant Growth Regulation*, *19*(2), 195–216. https://doi.org/10.1007/s003440000026

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., & Zemla, J. (2017). R package “corrplot”: Visualization of a Correlation Matrix. *Statistician*, *56*, 316–324. Retrieved from https://peerj.com/articles/9945/Supplemental_Data_S10.pdf

Young, O. P. (1986). Host Plants of the Tarnished Plant Bug, Lygus lineolaris (Heteroptera: Miridae). *Annals of the Entomological Society of America*, *79*(4), 747–762. https://doi.org/10.1093/aesa/79.4.747

Chapter 4: Polygalacturonase-inhibiting proteins as a putative mechanism of resistance to lygus herbivory in lima bean

Abstract

The Western tarnished plant bug, *Lygus hesperus* Knight, is a common pest for many important crops grown in California, including lima beans, *Phaseolus lunatus*. Lygus insects parasitize plant tissue by depositing digestive enzymes for extra-oral digestion. These enzymes target structural components of the plant cell wall. In response, multiple plant species have developed constitutive and/or induced defense responses to these digestive enzymes and produce inhibiting proteins to mitigate this digestion of the plant cell wall. This chapter will explore the role of polygalacturonase-inhibiting proteins in the plant defense response of lima beans to digestive enzymatic activity caused by lygus and another fungal source of polygalacturonase for comparison. The results will identify the heritability and plasticity of polygalacturonase-inhibiting proteins in a population and panel of lima beans grown under different environmental conditions and the role of this mechanism in developing lima beans resistant to lygus damage.

Introduction

Lima bean, *Phaseolus lunatus*, is the most important grain legume in California, a source of protein, fiber, and nutrients for people throughout the world, and an important nitrogen-fixing crop for many growers. Lima beans account for nearly half of the total dry bean production in California, with Stanislaus, San Joaquin, and Sutter counties leading lima bean production over the past decade. Dry lima bean production in the US is valued around $30 million dollars annually over the past decade, accounting for 60-80% of total global production,

with California as the sole producer of dry lima bean production nationally (Long et al., 2014; USDA ERS, 2004; USDA NASS, 2019). The most significant factor impacting lima bean production in California in recent year is the pest, *Lygus hesperus* Knight. Lygus is a proliferate pest in California due to its high mobility, its wide range of both native and cultivated host plants, and its ability to inflict significant damage to yields with low population levels (Long et al., 2014).

Lygus are piercing and sucking polyphagous bugs that use stylets to penetrate plant tissue, deposit salivary enzymes for extra-oral digestion, and siphon degraded plant tissue for consumption, a process known as lacerate-and-flush (Miles & Taylor 1994). Damage caused by lygus can be particularly destructive depending on the stage of the growing season for the crop and the duration and scale of the lygus infestation. During the growing season, lygus insects target the tender apical bud, floral bud, and developing pod tissue in lima bean with this lacerate-and-flush method of feeding, often resulting in necrosis and abortion. When lygus feed on the developing floral bud and pod tissue of lima bean, and the plant is unable to recover and regenerate the damaged tissues, there is a crop loss and a total reduction of grain yield. When lygus feed on the maturing pods and beans at the end of the growing season, this results in blemished and irregular beans and a reduction in marketable grain yield.

Lygus belong to the Heteroptera order and Miridae family. The latter is unique among the most common insect species evaluated in its ubiquitous expression of polygalacturonases (Frati, De Lorenzo, Salerno, Galletti, & Conti, 2006). The deposition of pectinases to parasitize plant tissue occurs with other insect predators (*Lygus lineolaris, Sitophilus oryzae*), as well as fungal (*Aspergillus niger, Fusarium moniliforme, Sclerotinia sclerotiorum*) and bacterial pathogens (*Bacillus subtilis, Erwinia carotovora)* (Pickersgill et al., 1998; Van Santen et al., 1999;

Vasconcellos et al., 2016). Polygalacturonase enzymes have been extensively studied in their ability to effectively digest the plant tissue of many economically important crops, including alfalfa, cotton, and strawberry (Long et al., 2014; Shackel et al., 2005).

Lygus saliva initiates the process of tissue degradation in the plant cell walls using pectinase enzymes, which includes pectolyase, pectozyme and polygalacturonase enzymes (Strong 1970). The salivary enzyme polygalacturonase hydrolyzes the α-1,4-polygalacturonic acid molecules, which are an integral structural component of the middle lamella of the plant cell wall comprised of repetitive chains of galacturonic acid molecules (Celorio-Mancera et al., 2008). Polygalacturonic acid molecules are hydrolyzed by polygalacturonase enzymes in two ways: (1) endo-polygalacturonases target o-glycosidic bonds randomly within the polygalacturonic acid chain producing oligogalacturonide residues, and (2) exo-polygalacturonases target the non-reducing end of the polygalacturonic acid chain producing monosaccharides of galacturonic acid (Celorio-Mancera et al., 2008). The enzymatic digestion of polygalacturonic acid molecules results in these monomer and oligomer molecules, which are more easily digestible by Lygus and other parasitic organisms.

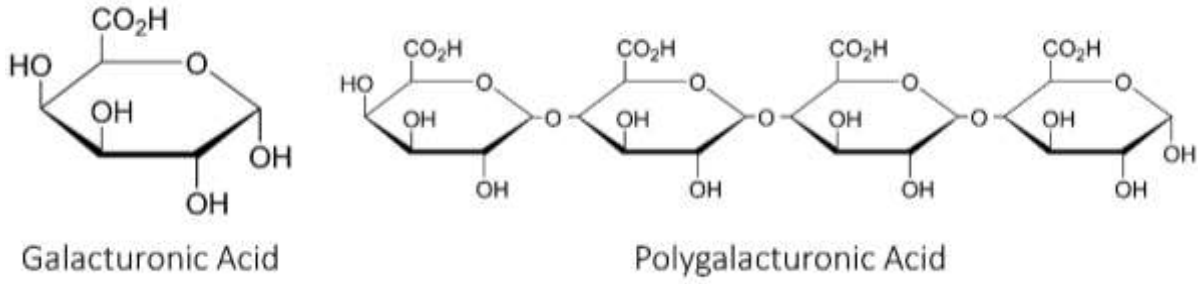

Figure 3.1 The monosaccharide, galacturonic acid, and a four-unit example of the polysaccharide, polygalacturonic acid.

Plants have developed mechanisms of resistance to this parasitic damage, which include the production of polygalacturonase-inhibiting proteins (PGIPs) in response to this enzymatic stress. Polygalacturonase-inhibiting protein production is induced by the monomer and oligomer products that result from polygalacturonase digestion of the polygalacturonic acid in the cell wall, and was first characterized in *Phaseolus vulgaris* plant extracts (Albersheim & Anderson, 1971; Aziz et al., 2004). Polygalacturonase-inhibiting proteins offer a direct response to polygalacturonase digestion when they target, bind to, and mitigate damage caused by polygalacturonase enzymes in the cell wall (Bergmann et al., 1994).

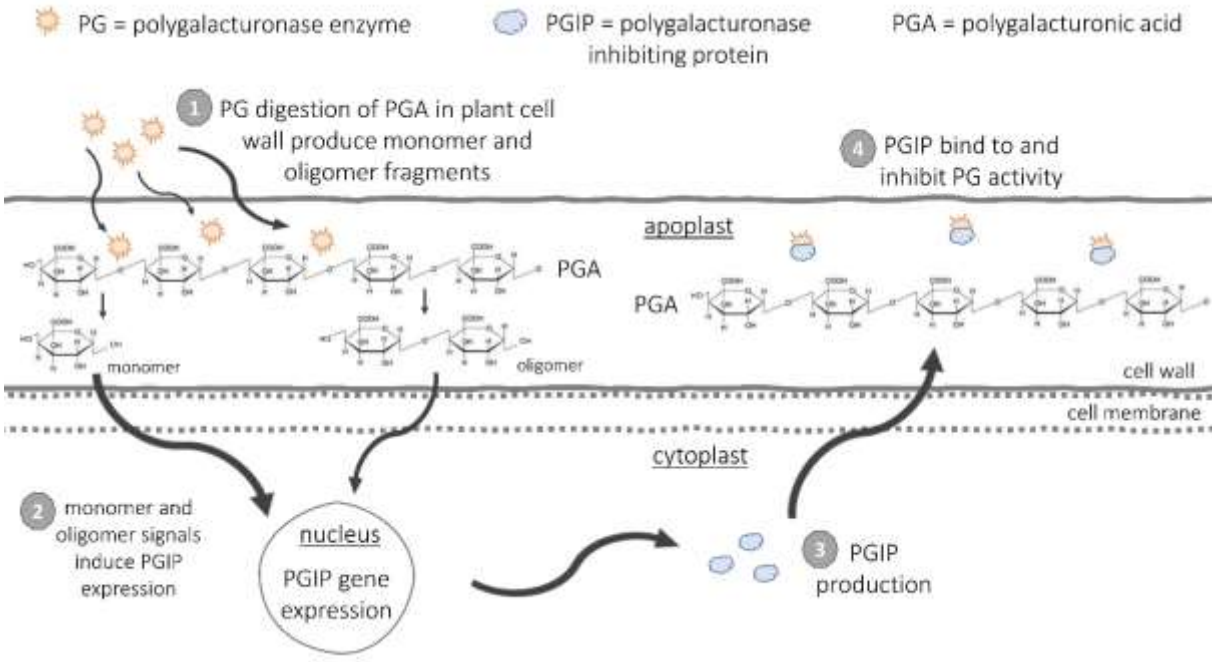

Figure 3.2 Digestion of polygalacturonic acid in the cell wall by polygalacturonase, which result in monomer and oligomer products that trigger the production of polygalacturonase inhibiting proteins to mitigate further polygalacturonase enzymatic activity in the cell wall.

Polygalacturonase-inhibiting proteins offer an indirect response to these enzymes, as they induce the expression of other stress response compounds such as jasmonic acid, salicylic

acid, and volatile organic compounds, the former two of which are thought to regulate polygalacturonase-inhibiting protein expression (Ferrari et al., 2003; Frati et al., 2006; R. Li et al., 2003).

Genes characterizing polygalacturonase-inhibiting proteins have been identified in a wide range of agricultural crop species, including alfalfa, apple, common bean, garbanzo bean, onion, pear, raspberry, rice, soybean, tomato, and wheat (De Lorenzo et al., 2001; Kalunke et al., 2015). In common bean, four genes (*PvPGIP1, PvPGIP2, PvPGIP3, PvPGIP4)* are clustered across a 50 kb region on the Pv02 linkage group and have a high level of nucleotide conservation across these four genes (Vasconcellos et al., 2016). The high conservation among the four polygalacturonase-inhibiting proteins genes is partially attributed to a duplication of the *PvPGIP1*/*PvPGIP2* and *PvPGIP3*/*PvPGIP4* genes , however there is variability in the expression of these genes across genotypes and types of tissue damage (D'Ovidio, Mattei, Roberti, & Bellincampi, 2004; Vasconcellos et al., 2016). The *PvPGIP1*/*PvPGIP2* genes demonstrate a strong phylogenetic relationship to the *GmPGIP3* gene, suggesting that these genes were strongly conserved and that the duplication in common bean occurred prior to the separation of the *Glycine* and *Phaseolus* genera (D'Ovidio, Mattei, et al., 2004; Kalunke et al., 2015). Duplications of PGIP genes are observed in multiple plant species, including rice, common bean, soybean, chickpea, barrel clover and thale cress, and offer an evolutionary advantage to these species through varying polygalacturonase specificities (Kalunke et al., 2015). The duplication and variation between genes within a species is partially attributed to the variation in the stress stimuli and induced elicitors that provide a broader range of

specificity and biochemical protection against polygalacturonase injury (Favaron, Sella, & D'Ovidio, 2004).

Previous work by Dashner, identified a series of sequence variants in common bean and lima bean in the amino acid sequence for *PvPGIP2* and *PlPGIP2*, with a high conservation of amino acids in the leucine-rich regions between these two species (Dashner, 2016). The leucine-rich regions for *PvPGIP2* encode amino acid residues of important structural components comprising the hydrophobic core of the enzyme (Matteo et al., 2003). Dashner's research work also identified variability in polygalacturonase inhibition among common bean and lima bean samples, and particularly between the UC 92 and UC Haskell lima bean cultivars. One nucleotide substitution (A580T) resulted in a nonsynonymous amino acid substitution (T194K) between the UC 92 and UC Haskell cultivars, which results in exposed surface accessibility of the synthesized protein in the UC 92 cultivar but not in UC Haskell (Dashner, 2016). This sequence variation may be responsible for the observed variation in polygalacturonase inhibition between them.

Previous research established that the expression of the PGIP genes are induced following polygalacturonase damage two days post infection, and that transformed common bean plants overexpressing the PGIP genes have a significant reduction in damage at the wound site compared to untransformed plants three days post infection (D'Ovidio et al., 2004; Manfredini et al., 2005). While the expression and production of polygalacturonase-inhibiting proteins is induced upon polygalacturonase wounding locally for up to three days, it is unknown how effective they are against repeated and future polygalacturonase damage.

Questions remain about the systemic effects of polygalacturonase-inhibiting proteins in plants exposed to polygalacturonase damage and herbivory throughout the lifetime of the plant and the impact that this defense mechanism may have on yield productivity. This body of research explores the variation in polygalacturonase inhibition in a panel of California-adapted cultivars and a biparental recombinant inbred line population in different environments subject to different biotic stress and identifies through QTL mapping loci contributing to polygalacturonase-inhibiting protein production and polygalacturonase inhibition.

Materials and Methods

Enzymatic assays

An adapted colorimetric assay was used to measure the enzymatic activity of lygus head pronotum extract and purified *Aspergillus niger* and the inhibition of this enzymatic activity by lima bean plant crude extract (Dashner, 2016; Ortiz et al., 2011). The colorimetric assay quantifies the amount of polygalacturonic acid digested by the polygalacturonase enzymes based on the binding properties of ruthenium red to undigested polygalacturonic acid. Final results for calculating the inhibition of polygalacturonase on the digestion of polygalacturonic acid were produced following multiple steps of extract preparation and assays. Key steps include (1) the preparation of head pronotum extract as the polygalacturonase enzyme, (2) the preparation of crude plant extract as the inhibiting proteins, (3) the quantification of protein levels per crude plant extract sample via the Bradford Assay, (4) the quantification of polygalacturonic acid digestion via the Colorimetric Assay, and (5) the final calculation of the percentage of polygalacturonase inhibition per sample.

Preparation of polygalacturonase enzymes

Head pronotum extract was sourced from live lygus insects that were collected from a conventionally managed field on the campus of the University of California, Davis in 2017 and obtained from a colony maintained by K. Gibson in the Gepts Lab (Plant and Environmental Sciences Building, University of California, Davis) in 2018. Adult lygus insects were frozen at -20°C for a minimum of 12 hours. The lygus head pronotum was separated from the rest of their body using a razor on a surface that was cooled by liquid nitrogen, using the UCANR lygus identification guide as a reference for the appropriate location of separation between the pronotum and scutellum (Mueller et al., 2003). The head pronotums were stored in separate Eppendorf tubes in liquid nitrogen and finely ground with a plastic pellet pestle (Z359947, Sigma Aldrich). Head pronotum extract was prepared in a ratio of 10 ml sodium acetate extraction buffer (0.1M sodium acetate, 1M NaCl, 5mM β-mercaptoethanol, pH 5.5) to 0.3 g of ground head pronotum, mixed vigorously on a vortexer (VWR Advanced Microplate Vortex Mixer), homogenated with a bead beater and two stainless steel beads per Eppendorf tube, and then centrifuged at 16,000 g for 15 minutes at 4°C (Dashner, 2016; Shackel et al., 2005). The supernatant was collected without disturbing the pellet and pooled into a Vivaspin-2 Polyethersulfone 30kDa MWCO Centrifugal Concentrator (#490013-726, VWR). The sample was centrifuged at 3,000 g for 8 minutes at 4°C, followed by 2 minute cycles at 3,000 g at 4°C until the volume recovered was half of the initial sample to concentrate the sample 10-fold. The concentrator caps were inverted and centrifuged again at 3,000 g for 2 minutes at 4°C to recover the concentrated head pronotum extract. The sample was aliquoted into separate

containers, stored at 4°C and used for the colorimetric assay within 72 hours, and then stored at -20°C for later use.

Purified pectinase of *Aspergillus niger* was selected as a secondary enzymatic source of polygalacturonase for evaluation across the sample of lima bean floral buds. Enzymatic activity of purified *A. niger* pectinase was evaluated at different dilutions [undiluted, 400-fold, 800-fold and 3200-fold dilutions] to determine the optimal enzymatic activity for this colorimetric assay. The 400-fold dilution of *A. niger* was selected as the secondary enzymatic source for colorimetric assays on the complete set of lima bean crude plant extract samples.

Plant material and experimental design

Tissue samples for colorimetric assay were sampled from plants grown in the field at the UC Davis Plant Sciences Field Facility under different insecticide treatments during the summer of 2017 and 2018. Genotypes include a collection of 12 California-adapted lima bean cultivars and a population of 180 recombinant inbred lines originating from lygus-resistant (UC Haskell) and lygus-susceptible (UC 92) parents.

More specifically, a collection of 12 California-adapted lima bean cultivars were planted in June 2017 in duplicated continuous rows. The seeds were planted as double rows on a 60-inch [1.52 m] planting bed with a 4-inch [10.16 cm] between-plant spacing. The conventionally managed field was split in the application of pesticide treatments to mitigate lygus damage and no pesticide treatments targeting lygus insects. The pesticide treated rows received applications of Warrior II (3.75 oz/acre) when lygus pressure was present. The samples of floral buds, immature pods, and mature pod tissue were collected from the first 20 feet [60.10 m] of

each row. The cultivars included in the study were Calico Cat (G26451), Dixie Speckled, Henderson Bush, Hopi 12 (G25623), Jackson Wonder, RIL 39, RIL 72, UC 92, UC Beija Flor, UC Cariblanco N, UC Haskell and UC Lee.

A recombinant inbred line (RIL) lima bean population was created from a cross of a lygus-resistant, small-seeded vine-type California cultivar, UC Haskell, and a lygus-susceptible, large-seeded bush-type California cultivar, UC 92. An experiment of 180 RILs were planted in a randomized augmented balanced incomplete block split-plot design at two locations in April and May of 2018. Each plot was a 4-foot [1.22 m] double-row planting on 60-inch [1.52 m] beds with a 4-inch [10.16 cm] between-plant spacing of 24 seeds. Due to the differences in management of different plant habits, bush RILs and vine RILs were separated into distinct experiments. The vine experiment consisted of 110 RILs with a between-plot spacing of 6-feet [1.83 m], four checks (Calico Cat, Hopi 12, RIL 39, RIL 72), which were replicated twice, and two controls [UC Haskell (lygus-resistant) and Cariblanco (lygus-susceptible)], which were replicated in each of three balanced incomplete blocks. To prevent cross-contamination of vine RILs across beds, a double-row bed of black-seeded bush-type common beans, *Phaseolus vulgaris*, were planted between the beds of vine RILs. The bush experiment had 61 RILs with a between-plot spacing of 3-feet [0.91 m], three checks (Dixie Speckled, Henderson, and Jackson Wonder) that were replicated twice, and two controls [UC 92 (lygus-susceptible) and UC Beija Flor (lygus-resistant)], which were replicated four times in each of the three balanced incomplete blocks.

The split-plot experimental design was created for the application of pesticide treatments to mitigate lygus damage on one plot and no pesticide treatments targeting lygus insects on the other plot. The pesticide treatment plots received applications of Warrior II (3.75

oz/acre) and Mustang (4 oz/acre) when more than three lygus bugs were observed on more than half of the control plots. The no-pesticide treatment main plots did not receive either of these pesticide treatments that would target lygus insects throughout the season.

Preparation of polygalacturonic-acid inhibiting proteins from plant crude extracts

For each plot, three subsamples were collected for the floral bud and immature pod tissues into a 96-well plate and immediately placed on ice and stored in a -80°C freezer for later use; the year, location, genotype, treatment and plot identification were recorded for each sample. Plant tissue was ground with plastic pellet pestles (Sigma Aldrich) with the sample tube cooled in liquid nitrogen. Plant crude extract was prepared in a ratio of 1 ml sodium acetate extraction buffer (0.1M sodium acetate, 1M NaCl, 5mM β-mercaptoethanol, pH 5.5) to 0.3 g of ground plant tissue. Ground plant tissue was homogenized in extraction buffer with plastic pellet pestles (Z359947, Sigma Aldrich), mixed on a vortexer (VWR Advanced Microplate Vortex Mixer) for 60 seconds at 2500 rpm, and placed on a shaker on ice for 1 hour. Plant tissue was mixed again on a vortexer (VWR Advanced Microplate Vortex Mixer) for 30 seconds at 2500 rpm and then centrifuged at 6,000 g for 7 minutes at 4°C. Supernatant was carefully aliquoted without disturbing the pellet and either used immediately or stored at -20°C for further assays.

Bradford assay

Protein concentration of the crude plant extracts for floral bud and immature pod tissue samples from lima bean was calculated using the Bradford Assay (Wu et al., 2012). To each well in a 200 µl clear-bottom 96-well microtiter plate (#41-12-21-02, Fisher Scientific), a 10 µl aliquot of crude plant extract for each sample was added to 190 µl of the Coomassie brilliant blue

reagent (0.01% (w/v) Coomassie Brilliant Blue G-250 (#97063-852, VWR), 4.75% (w/v) ethanol, 8.5% (w/v) phosphoric acid). Samples were then mixed on a vortexer (VWR Advanced Microplate Vortex Mixer) for 30 seconds at 500 rpm. After 10 minutes, their absorbance at 595 nm was measured using a Gen5 Microplate Reader and Image Software.

Standards of Pierce bovine serum albumin (#Pl23210, Thermo Scientific) [1mg/ml, 0.8 mg/ml, 0.6 mg/ml, 0.4 mg/ml, 0.2 mg/ml], were diluted with sodium extraction buffer (0.1M sodium acetate, 1M NaCl 5mM, pH 5.5) and mixed in the same ratio as samples with 10 µl of standard solution to 190 µl of the Coomassie brilliant blue reagent (0.01% (w/v) Coomassie Brilliant Blue G-250 (#97063-852, VWR), 4.75% (w/v) ethanol, 8.5% (w/v) phosphoric acid). Standard concentrations were run in triplicate and the absorbance was measured at 595 nm using the Gen5 Microplate Reader and Image Software. A calibration curve of protein concentration was summarized in a linear regression.

Colorimetric assay

The colorimetric assay was conducted using two sources of enzymatic digestion, the head pronotum extract derived from our targeted pest, *Lygus hesperus*, and a purified source of pectinase from the fungal pathogen, *Aspergillus niger* (#P4716, Sigma-Aldrich). The crude plant extract of lima bean tissue was used as a source of potentially inhibitory proteins acting on the enzymatic digestion of the head pronotum extract and *A. niger* pectinase on the substrate, polygalacturonic acid. The substrate solution for this assay was 0.5% polygalacturonic acid, created by suspending polygalacturonic acid sodium salt (#P1656, Spectrum Chemical) in sodium acetate extraction buffer (0.1M sodium acetate, 1M NaCl 5mM) and mixing under 60°C

heat for 72 hours. Following this suspension, the pH of the solution was adjusted to 5.5 using 6 ml of glacial acetic acid.

The inhibitory effects of plant crude extract from the two parental genotypes, UC 92 and UC Haskell, were evaluated at different concentration levels [0 µg/µl, 0.5 µg/µl, 1 µg/µl, 1.5 µg/µl, 2 µg/µl]. The different combinations of plant crude extract protein concentrations were assayed in triplicate samples with lygus head pronotum extract and *A. niger* 400-fold dilution as sources of polygalacturonase. The different plant crude extract protein concentrations were evaluated to identify the optimum crude plant extract protein concentration to inhibit the enzymatic activity of different sources of polygalacturonase.

The colorimetric assays were run with a total volume of 20 µl per reaction across 200 µl 96-well PCR microplates (#14222322, Fisher Scientific). Each well was loaded with 8 µl polygalacturonic acid, followed by 4 µl of crude plant extract from lima bean samples, followed by 8 µl of either *L. hesperus* or *A. niger* polygalacturonase enzyme loaded onto the sides of the wells. The plates were then centrifuged for 10 seconds to allow the enzyme to uniformly contact the polygalacturonic acid-crude plant extract solutions, then mixed on a vortexer (VWR Advanced Microplate Vortex Mixer) for 30 seconds at 1500 rpm and centrifuged again for 10 seconds to assure the liquid was at the bottom of each well. The PCR microplates were placed on a PCR machine for a cycle of 20 minutes at 40°C, followed by 1 minute at 4°C, and then programmed to remain at 4°C until removed from the PCR machine. To each well, 40 µl of 1.125 mM Ruthenium Red (#R2751, Sigma-Aldrich) was added, covered and mixed on a vortexer (VWR Advanced Microplate Vortex Mixer) for 30 seconds at 1500 rpm, then 100 µl of 8mM NaOH was added, covered and mixed on a vortexer (VWR Advanced Microplate Vortex

Mixer) for 30 seconds at 1500 rpm. To each well in a 200 µl clear-bottom 96-well microtiter plate (#41-12-21-02, Fisher Scientific), a 25 µl aliquot of each sample from the PCR microplate was added to 175 µl of milliQ water, then mixed on a vortexer (VWR Advanced Microplate Vortex Mixer) for 30 seconds at 500 rpm and let sit for 10 minutes before measuring the absorbance at 535 ηm using the Gen5 Microplate Reader and Image Software.

Standards for polygalacturonase activity were created from different final concentrations of polygalacturonic acid [0 µg, 10 µg, 20 µg, 30 µg, and 40 µg] per 20 µl reaction. The standard concentrations were diluted with sodium acetate extraction buffer (0.1M sodium acetate, 1M NaCl 5mM, pH 5.5), and contained crude plant extract (50% UC Haskell: 50% UC92) and either denatured lygus head pronotum extract or 400-fold diluted *A. niger* enzyme in the same ratio (8 µl of PGA: 4 µl of CPE: 8 µl of enzyme) for the total 20 µl reaction volume of the colorimetric assay. The *L. hesperus* and *A. niger* enzymes were previously denatured by placing the Eppendorf tubes in an incubator set to 80°C for a minimum of 2 hours. Standard concentrations were run in triplicate and underwent post-PCR steps before measuring the absorbance at 535 ηm using the Gen5 Microplate Reader and Image Software, and a calibration curve of polygalacturonic acid was summarized in a linear regression.

Final calculation of enzymatic inhibition

Enzymatic inhibition was calculated as the percentage of the initial 40 µg of polygalacturonic acid retained following enzymatic incubation, and the equation for the calculation is as follows:

$$\% \text{ enzymatic inhibition} = \left(\frac{(A_{sample} - \beta_0)}{\beta_1} \right) * 40\ \mu g^{-1}$$

where A_{sample} is the 535 ηm absorbance of each reaction sample containing the polygalacturonic acid - enzyme - sample crude plant extract in a ratio of 8 µl: 8 µl: 4 µl in a total volume of 20 µl, and β_0 is the y-intercept of the calibration curve and β_1 is the slope of the calibration curve of 535 ηm absorbance to polygalacturonic acid concentrations containing the inactive enzyme and crude plant extract in a ratio of 8 µl: 8 µl: 4 µl in a total volume of 20 µl.

Statistical analysis

An Analysis of Variance was performed to determine significant differences among variables for putative traits for lygus resistance. Linear models were created for percent enzymatic inhibition for *L. hesperus* and *A. niger* polygalacturonase sources incorporating genotype, treatment, location and protein concentration as fixed effect factors, and year, block and plate as random effect factors. Genotype interactions with location and treatment were also included as fixed effect factors in the linear models. A linear model was created for protein concentration of the crude plant extract incorporating genotype, treatment and location as fixed effect factors, and block and plate as a random effect factor. Genotype interactions with location and treatment were also included as fixed effect factors in the linear model.

For the linear models, genotype refers to the UC 92 – UC Haskell recombinant inbred lines, parental controls, and California-adapted cultivars, treatment refers to the application of insecticide to control for lygus pressure (multiple applications of insecticide or no application of insecticide), location refers to the location of the field (Loc 1 (SB), Loc 2 (AG2), Loc 3 (AG1)), protein concentration as an independent variable and factor refers to the rounded value of the crude plant extract protein concentration converted to a factor (1, 2, 3, 4, 5), plate refers to the

96-well sampling plate that the tissues were collected and assays were performed on (1-12), and block refers to the field blocking relative to the irrigation source (1-5).

The additive mixed effect models were created using the 'lme4' package and the ANOVA was performed with the 'stats' package in R statistical software (Bates et al., 2015; R Core Team, 2013). Prior to results being reported, the assumptions of independence, normality and homoscedasticity were tested by running the Shapiro-Wilk test of normality and the Levene's test for homogeneity of variance using the 'stats' and 'car' packages in R statistical software, respectively (Fox & Weisbery, 2019; R Core Team, 2013). Phenotypic data that deviated from the assumptions were set to missing. The adjusted R^2 incorporating the fixed and random effects of the linear models and the coefficient of variation of the linear models were calculated using the 'piecewiseSEM' and 'sjstats' packages in R statistical software, respectively (Lüdecke, 2020; Nakagawa & Schielzeth, 2013). The least significant differences between fixed effect factors were calculated using the 'lsmeans' package, the broad sense heritability of each trait was calculated following the heritability estimator equation across multiple environments and treatments, and the correlations between traits were calculated using the 'corrplot' package in R statistical software (J. B. Holland et al., 2003; Lenth, 2016; Wei et al., 2017).

QTL mapping

QTL mapping was conducted on the UC 92 – UC Haskell RIL population using the 'R/qtl' package in R statistical software (Broman, 2009). A genome-wide scan for single QTLs was performed using the 'cim' function with the extended Haley-Knott regression method, the Kosambi mapping function, and 1000 permutations to identify the position and LOD score of

QTLs above the 95% significance threshold. The 'makeqtl' and 'fitqtl' functions were used to identify the optimal multiple QTL model and to calculate peak LOD scores, percent of phenotypic variation explained by the QTL and the QTL effects. Major QTLs are defined as those that account for > 10% of the phenotypic variation explained, and minor QTLs are defined as those that account for < 10% of the phenotypic variation explained (Collard et al., 2005).

Results

Evaluation of California cultivars

Significant differences for the three traits [Crude Extract Plant Protein Concentration, Enzymatic Inhibition (*L. hesperus*), Enzymatic Inhibition (*A. niger*)] were detected among the 12 California-adapted cultivars of lima bean, between the applications of insecticide treatments, and among the different locations; significant genotype-location interactions were observed for all three traits for crude plant extract protein concentration and the enzymatic inhibition of *L. hesperus* and *A. niger* polygalacturonase from floral bud tissue. The application of insecticide treatment was the most significant factor for all three traits across all locations. Genotype was significant for all three traits across all locations, excluding location 1 for the enzymatic inhibition of *A. niger* polygalacturonase. Significant genotype-by-treatment interactions were observed for crude plant extract protein concentration at all locations at the $p < 0.001$ significance level, and for the enzymatic inhibition of *A. niger* polygalacturonase at the $p < 0.05$ significance level for the collective location analysis but was not significant at either of the individual locations. Genotype-by-treatment interactions were not significant for the enzymatic inhibition of *L. hesperus* polygalacturonase, except for the 2017 location at the $p < 0.10$

significance level. Genotype-by-treatment-by-location interactions were observed for crude plant extract protein concentration and the enzymatic inhibition of *A. niger* polygalacturonase at the $p < 0.001$ and $p < 0.05$ significance levels, respectively, but not for the enzymatic inhibition of *L. hesperus* polygalacturonase. Effects of the different treatments across locations and genotypes for the three traits are shown in Figure 3.3, and descriptive statistics for the three traits collectively and at each location are included in Table 3.1.

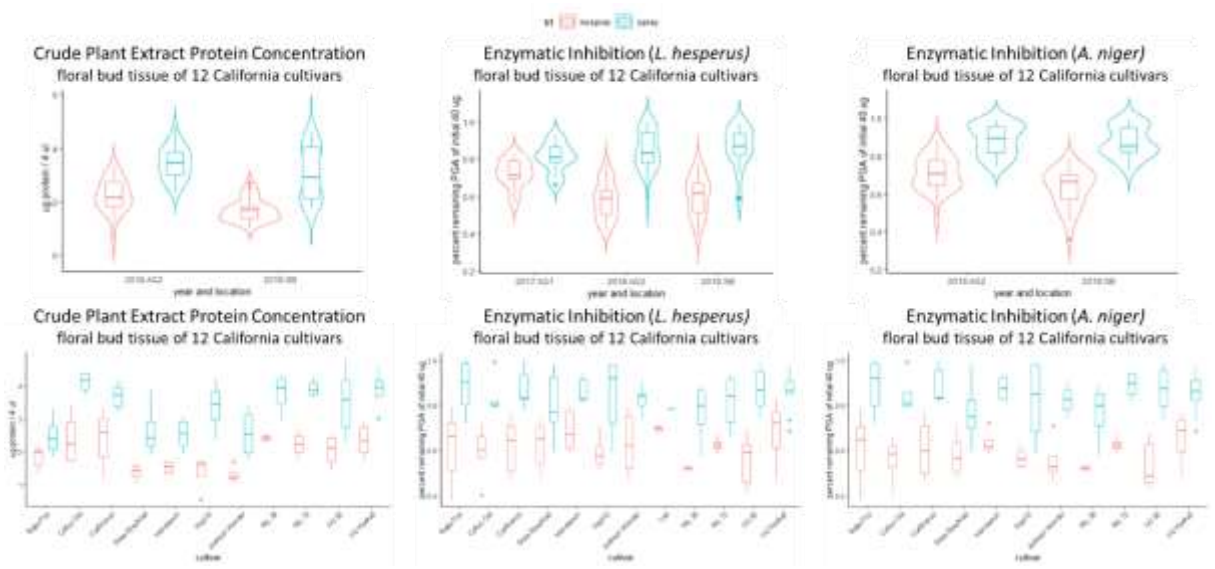

Figure 3.3 Boxplots of the five-number summary for crude plant extract protein concentration and the inhibition of polygalacturonase activity *Lygus* and *A.niger* sources and across locations (top row) and genotypes (bottom row) in the floral bud tissue of 12 California-adapted cultivars. Red boxplots represent the 'no-spray' insecticidal treatment, and blue boxplots represent the 'spray' insecticidal treatment.

Comparing the least squares means of all three traits, treatments were significantly different from each other at all locations at the $p < 0.05$ significance level. For the unsprayed treatments, there was a reduction in the crude plant extract protein concentration and enzymatic inhibition for *L. hesperus* and *A. niger* polygalacturonase sources of 39%, 21% and 26%, respectively. Comparing the least squares means of all three traits, locations were

significantly different from each other at the $p < 0.05$ significance level. There was a 18% increase in crude plant extract protein concentration and an 8% and 4% reduction in enzymatic inhibition of *L. hesperus* and *A. niger* polygalacturonase, respectively, from location 1 to location 2. For the enzymatic inhibition of *L. hesperus* polygalacturonase, there was a reduction of 16% and 22% from location 3 to location 1 and location 2, respectively.

Table 3.1 Descriptive statistics for crude plant extract protein concentration, enzymatic inhibition of *L. hesperus* polygalacturonase and enzymatic inhibition of *A. niger* polygalacturonase in the floral bud tissue of 12 California cultivars.

	Trait	Protein Concentration			Enzymatic Inhibition (*L. hesperus*)				Enzymatic Inhibition (*A. niger*)		
	units	ug/4ul			% 40ug PGA				% 40ug PGA		
	Loc	all	Loc 1	Loc 2	all	Loc 1	Loc 2	Loc 3	all	Loc 1	Loc 2
F values	Geno	22.28***	25.21***	9.01***	1.48.	2.40*	2.68*	5.88**	4.66***	1.28ns	8.44***
	Trt	380.50***	313.67***	168.38***	46.27***	7.93**	33.60***	68.04***	59.46***	15.31***	45.85***
	Loc	33.29***	-	-	4.28*	-	-	-	6.96*	-	-
	Prot	-	-	-	2.38.	1.79ns	2.65*	-	0.66ns	0.69ns	0.62ns
	Geno*Trt	3.70***	7.93***	4.92***	0.73ns	1.72ns	1.16ns	2.85.	2.33*	1.80ns	1.55ns
	Geno*Loc	4.38***	-	-	1.92*	-	-	-	4.92***	-	-
	Geno*Trt*Loc	7.83***	-	-	0.96ns	-	-	-	2.03*	-	-
Geno means	Beija Flor	2.09 a	1.73 a	2.41 a	0.82 b	0.76 ab	0.78 b	0.82 b	0.80 bc	0.72 a	0.85 cd
	Calico Cat	3.24 d	2.88 bcd	3.61 c	0.73 ab	0.70 ab	0.68 ab	0.76 ab	0.71 ab	0.80 a	0.74 abc
	Cariblanco	3.03 cd	2.80 cd	3.26 bc	0.78 ab	0.79 ab	0.74 ab	0.74 ab	0.69 a	0.82 a	0.68 a
	Dixie Speckled	2.02 ab	1.64 a	2.40 ab	0.77 ab	0.68 a	0.76 ab	0.82 b	0.72 ab	0.66 a	0.76 abc
	Henderson	2.03 ab	1.73 ab	2.34 ab	0.84 b	0.80 b	0.77 ab	0.83 b	0.77 abc	0.68 a	0.84 bcd
	Hopi12	2.37 abc	2.54 abcd	2.42 ab	0.76 ab	0.60 a	0.80 b	0.77 ab	0.70 ab	0.71 a	0.79 abcd
	Jackson Wonder	1.94 a	1.57 a	2.31 ab	0.80 ab	0.76 ab	0.70 ab	0.83 b	0.75 abc	0.67 a	0.80 abcd
	Lee	-	-	-	0.76 ab	-	-	0.76 ab	-	-	-
	RIL 39	3.19 cd	3.33 d	2.70 abc	0.67 a	0.66 a	0.53 a	0.72 ab	0.66 a	0.74 a	0.71 ab
	RIL 72	3.23 cd	3.09 cd	3.29 abc	0.74 ab	0.72 ab	0.74 ab	0.61 a	0.73 ab	0.81 a	0.76 abc
	UC 92	2.71 bcd	2.15 abc	3.08 bc	0.76 ab	0.78 ab	0.65 ab	0.73 ab	0.85 c	0.78 a	0.90 d
	UC Haskell	3.09 cd	3.07d	3.15 bc	0.80 ab	0.80 b	0.72 ab	0.87 b	0.67 a	0.72 a	0.77 a
Env means	No spray	1.99 a	1.68 a	2.23 a	0.69 a	0.63 a	0.58 a	0.70 a	0.62 a	0.61 a	0.69 a
	Spray	3.27 b	3.14 b	3.40 b	0.87 b	0.83 b	0.85 b	0.85 b	0.84 b	0.87 b	0.87 b
	Loc 1 (SB)	2.37 a	-	-	0.75 b	-	-	-	0.75 a	-	-
	Loc 2 (AG2)	2.89 b	-	-	0.69 c	-	-	-	0.72 b	-	-
	Loc 3 (AG1)	-	-	-	0.89 a	-	-	-	-	-	-
CV (%)		10.94	9.92	10.37	9.45	8.82	9.22	2.88	6.36	7.69	4.84
R^2		0.87	0.92	0.83	0.70	0.77	0.75	0.87	0.81	0.76	0.85
H^2		0.02	0.07	0.05	0.02	0.16	0.07	0.08	0.05	0.06	0.15

'***' denotes that $p < 0.001$, '**' < 0.01, '*' < 0.05, '.' < 0.10, 'ns' > 0.10.
'abcd' letters represent significantly different adjusted lsmeans groups at the $p < 0.05$ significance level.

There were significant differences among the genotypes at the *p < 0.05* significance level; however, there was a general conservation of enzymatic inhibition performance across

the two sources of polygalacturonase, which is seen in Figure 3.4. Comparing the least squares means of the 12 California adapted cultivars for enzymatic inhibition across the two different sources of polygalacturonase enzymes provides insight into the stability of cultivar enzymatic inhibition across biotic sources. The most notable changes in rank for enzymatic inhibition was observed in UC 92 and UC Haskell, the two parents of the biparental population study.

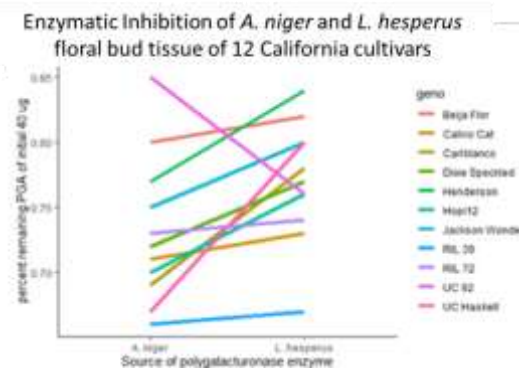

Figure 3.4 Enzymatic inhibition of polygalacturonase activity from *A. niger* and *L. hesperus* sources in the floral bud tissue of 12 California cultivars.

The heritability of all three traits among the California cultivars are relatively low varying between 2-7% for crude plant extract protein concentration, 2-16% for the enzymatic inhibition of *L. hesperus* polygalacturonase, and 5-15% for the enzymatic inhibition of *A. niger* polygalacturonase across locations. The broad sense heritability, H^2, estimates are included in Table 3.1.

Evaluation of L. hesperus and A. niger enzymatic activity on parental lines

A preliminary study identified significant differences between the two recombinant inbred line parental genotypes, UC 92 and UC Haskell, for inhibiting the enzymatic digestion of polygalacturonic acid of both *L. hesperus* and *A. niger* sources at the $p < 0.001$ level. While there was a significant difference between UC 92 and UC Haskell in their inhibition, they showed a difference in magnitude in inhibiting enzymatic activity depending on the enzymatic source. Descriptive statistics for this parental analysis on percent inhibition of enzymatic

activity by *L. hesperus* and *A. niger* are included in Table 3.2, and the performance of the two genotypes across the three different protein concentrations are included in Figure 3.5.

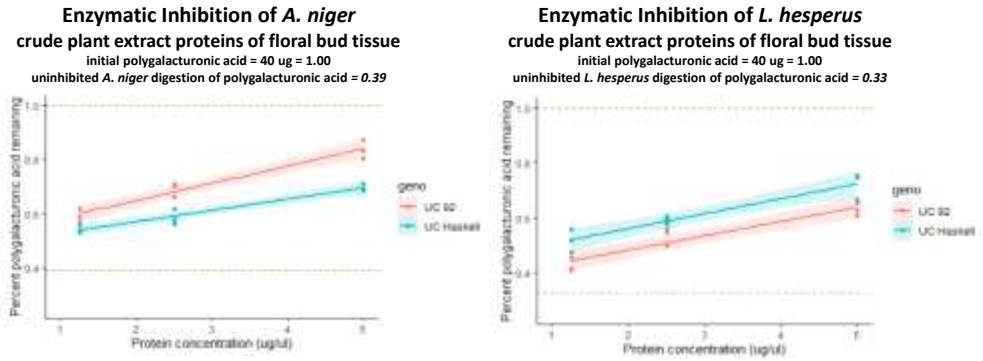

Figure 3.5 Enzymatic inhibition of *L. hesperus* and *A. niger* polygalacturonase activity by crude plant extract protein concentrations of floral bud tissue for UC 92 and UC Haskell.

Significant differences were also identified for crude plant extract protein concentration for both enzymatic sources at the $p < 0.001$ level. A significant interaction was observed between genotype and protein concentration for *A. niger* enzymatic inhibition only at the $p < 0.05$ level. There was a strong positive correlation between crude plant extract protein concentration and enzymatic inhibition of both polygalacturonase sources for both parental genotypes, with adjusted R^2 values above 90% and the coefficient of variation for the linear models of each trait below 5%.

Table 3.2 Descriptive statistics of enzymatic inhibition of *L. hesperus* (*Lh)* polygalacturonase and *A. niger* (*An*)polygalacturonase across different protein concentrations of crude plant extract from floral bud tissue of UC 92 and UC Haskell.

	Trait units	Enz. Inh. (*Lh)* % 40 ug PGA	Enz. Inh. (*An)* % 40 ug PGA
F values	Geno	27.22***	72.01***
	Prot	59.23***	99.96***
	Prot*Geno	0.48ns	4.86*
Geno means	UC 92	0.53 a	0.71 a
	UC Haskell	0.61 b	0.61 b
Prot means	1.25 ug 4ul^{-1}	0.47 a	0.57 a
	2.5 ug 4ul^{-1}	0.57 b	0.64 b
	5.0 ug 4ul^{-1}	0.68 c	0.77 c
CV (%)		4.69	3.06
R^2		0.93	0.96

'***' denotes that p < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.10, 'ns' > 0.10

Statistical analysis of UC 92 - UC Haskell RIL population

Percent of enzymatic inhibition of two different enzymatic sources and crude plant extract protein concentration were evaluated among 180 RIL genotypes replicated in split-plot trials at two different locations. Results from the ANOVA for crude plant extract protein concentration identified genotype and treatment as highly significant factors at the $p < 0.001$ level and location as a significant factor at the $p < 0.01$ level; there was a highly significant genotype by location interaction at the $p < 0.001$ level. Results from the ANOVA for percent enzymatic inhibition of *L. hesperus* polygalacturonase identified genotype and treatment as highly significant factors at the $p < 0.001$ level, protein concentration as a significant factor at the $p < 0.10$ level; significant genotype by location interactions at the $p < 0.01$ *level* and genotype by treatment interactions at the $p < 0.10$ level were identified as well. Results from

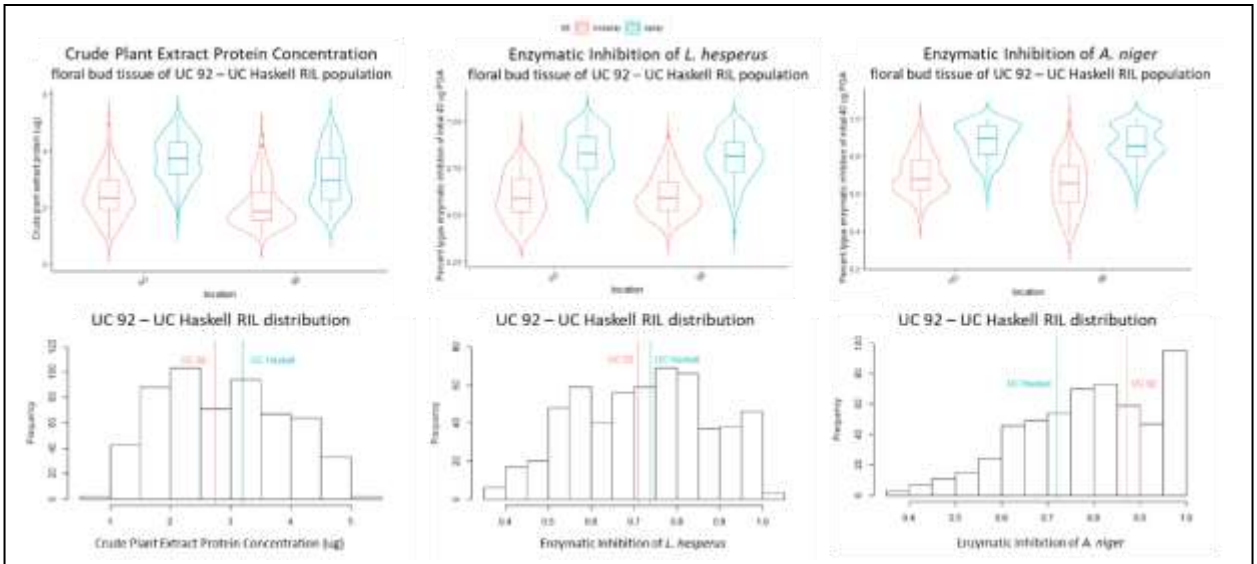

Figure 3.5 Boxplots of the effect of treatment by location on the crude plant extract protein concentration and the inhibition of polygalacturonase from *L. hesperus* and *A. niger* sources (top row). Phenotypic distribution among the UC 92 - UC Haskell RIL population for the crude plant extract protein concentration and the inhibition of polygalacturonase from *L. hesperus* and *A. niger* sources (bottom row). Red boxplots represent the 'no-spray' insecticidal treatment, and blue boxplots represent the 'spray' insecticidal treatment.

the ANOVA for percent enzymatic inhibition of *A. niger* polygalacturonase identified genotype and treatment as highly significant factors at the $p < 0.001$ level, protein concentration as a significant factor at the $p < 0.05$ level, and a significant genotype by location interaction at the $p < 0.05$ level. The effects of the different treatments and distribution among the RIL population for the three traits are shown in Figure 3.5 and descriptive statistics for the three traits are shown in Table 3.3.

Table 3.3 Descriptive statistics for crude plant extract protein concentration, enzymatic inhibition of *L. hesperus* polygalacturonase, and enzymatic inhibition of *A. niger* polygalacturonase.

	Trait	**Protein Concentration**			**Enzymatic Inhibition (*L. hesperus*)**			**Enzymatic Inhibition (*A. niger*)**		
	units	ug/4ul			% 40 ug PGA			% 40 ug PGA		
	Loc	all	Loc 1	Loc 2	all	Loc 1	Loc 2	all	Loc 1	Loc 2
Parent means	UC 92	2.75	1.96	3.33	0.71	0.77	0.64	0.81	0.81	0.87
	UC Haskell	3.13	3.02	3.25	0.76	0.80	0.71	0.70	0.75	0.70
RIL	Mean	2.9	2.66	3.12	0.71	0.70	0.71	0.76	0.76	0.77
	Min	1.43	1.20	0.89	0.48	0.43	0.30	0.57	0.52	0.47
	Max	4.23	4.42	5.61	0.90	0.92	0.95	0.98	0.99	0.99
F values	Geno	3.04***	2.31***	2.79***	2.41***	1.80***	1.93***	1.93***	1.76***	3.59***
	Trt	104.81***	24.75*	278.20*	87.13***	85.59***	106.21***	28.91***	7.36.	91.34***
	Loc	19.49**	-	-	4.58ns	-	-	0.43ns	-	-
	Prot	-	-	-	2.06.	3.66**	2.58*	2.50*	1.38ns	3.70**
	Geno*Trt	1.17ns	-	-	1.23.	-	-	0.93ns	-	-
	Geno*Loc	1.80***	-	-	1.49**	-	-	1.41*	-	-
Env means	No spray	2.26 a	2.05 a	2.46 a	0.62 a	0.62 a	0.61 a	0.68 a	0.67 a	0.70 a
	Spray	3.51 b	3.21 b	3.76 b	0.81 b	0.79 b	0.82 b	0.84 b	0.84 a	0.84 b
	Loc 1 (SB)	2.61 a	-	-	0.71 a	-	-	0.75 a	-	-
	Loc 2 (AG2)	3.15 b	-	-	0.72 a	-	-	0.77 a	-	-
CV (%)		8.52	13.60	10.93	6.82	9.82	9.58	5.78	8.68	5.03
R^2		0.82	0.80	0.77	0.72	0.66	0.68	0.75	0.72	0.80
H^2		0.02	0.08	0.01	0.03	0.02	0.02	0.05	0.15	0.04

'***' denotes that p < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.10, 'ns' > 0.10.
'ab' letters represent significantly different adjusted lsmeans groups at the p < 0.05 significance level.

The heritability of all three traits among the UC 92 – UC Haskell RIL population are relatively low varying between 1-8% for crude plant extract protein concentration, 2-3% for the enzymatic inhibition of *L. hesperus* polygalacturonase, and 4-15% for the enzymatic inhibition

of *A. niger* polygalacturonase across locations. The broad sense heritability, H^2, estimates are included in Table 3.3.

QTL analysis of the UC 92 – UC Haskell RIL population

QTL analyses of crude plant extract protein concentration and the percent of enzymatic inhibition of *L. hesperus* and *A. niger* polygalacturonases, respectively, were run for each of the two insecticide treatments and the average of the two treatments. Significant QTL peaks (Figure 3.6, summarized in Table 3.4) for crude plant extract protein concentration were identified for both treatments and their averages and for the percent of enzymatic inhibition of *A. niger* polygalacturonase for the insecticide treatment only. There were no significant QTL identified for the percent enzymatic inhibition of *L. hesperus* polygalacturonase for either treatment or the treatment average.

Table 3.4 Significant QTLs, flanking markers around QTL peak, LOD scores, percent phenotypic variation explained, QTL effects, and the positive effect allele for crude plant extract protein concentration and the percent enzymatic inhibition of *A. niger* polygalacturonase for different insecticide treatments and average across treatments.

Trait	Treatment	QTL Peak	Flanking Markers of QTL Peak	LOD	% PVE	QTL Effect	QTL Effect (+) Allele
Protein Concentration	Average	PI01.59	PI01_22100649 - PI01_42432553	8.03	19.97	0.68 ug $4ul^{-1}$	UC Haskell
	Spray	PI01.57	PI01_22100649 - PI01_42432553	5.74	14.98	0.66 ug $4ul^{-1}$	UC Haskell
	No Spray	PI01.62.8	PI01_42432553 - PI01_42503850	5.93	18.32	0.67 ug $4ul^{-1}$	UC Haskell
Enzymatic Inhibition (*A.niger*)	Spray	PI01.53	PI01_22100649 - PI01_42432553	4.03	10.57	4.53% 40ug PGA	UC 92

QTL mapping identified major QTLs for crude plant extract protein concentration for all treatments (average, spray and no spray) and for the enzymatic inhibition of *A. niger*

polygalacturonase for the spray treatment alone, which collocate in the same region of chromosome Pl01. The flanking markers, Pl01_22100649 and Pl01_42432553, for 3 of these QTL peaks (protein concentration for the treatment average and spray treatment, and enzymatic inhibition of *A. niger* polygalacturonase under the spray treatment) are the two markers that also flank a ~ 20 Mbp and ~ 20 cM gap in the pericentromeric region of chromosome Pl01 on this recombinant inbred line populations genetic map. The fourth QTL peak falls on the locus containing the latter marker, Pl01_42432553, as the significant region for crude plant extract protein concentration under the no-spray treatment.

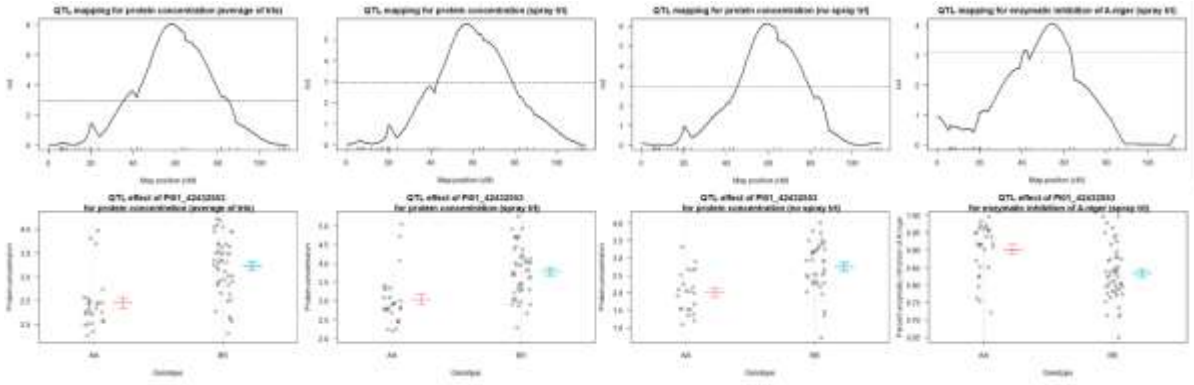

Figure 3.6 QTL peaks for protein concentration and enzymatic inhibition of *A. niger* on chromosome Pl01 and the QTL effect of the closest marker, Pl01_42432553, to the QTL peaks.

The percent of the phenotypic variation explained by the closest flanking marker ranges between 15 – 20 % for the crude plant extract protein concentration depending on the treatment and is highest for the average of the two treatments. The positive allele was contributed by UC Haskell for an increase of 0.66 - 0.68 µg 4 μl^{-1} for crude plant extract protein concentration for all treatments (average, spray, and no-spray). The percent of the phenotypic variation explained by the closest flanking marker for the enzymatic inhibition of *A. niger* polygalacturonase is 11% under the spray treatment. The positive allele was contributed by UC

92 for an increase of 4.5% of 40 µg polygalacturonic acid for the enzymatic inhibition of *A. niger* polygalacturonase under the spray treatment. The reversal of the positive allele contributors for the QTL effects of these crude plant extract protein concentration and the enzymatic inhibition of *A. niger* polygalacturonase is consistent with the lsmeans of the parental averages reported in Table 3.3.

Correlations and principal component analyses

There were significant positive correlations between crude plant extract protein concentration and the enzymatic inhibitions of *L. hesperus* and *A. niger* polygalacturonases in the recombinant inbred line population across treatments. The correlation between crude plant extract protein concentration and the enzymatic inhibitions of *L. hesperus* and *A. niger* polygalacturonases were 39% and 44%, respectively, and were significant at the $p < 0.001$ significance level. There was a significant positive correlation between the enzymatic inhibitions of the two sources of polygalacturonase of 43%, at the $p < 0.001$ significance level. The correlations between crude plant extract protein concentration, enzymatic inhibition of *L.*

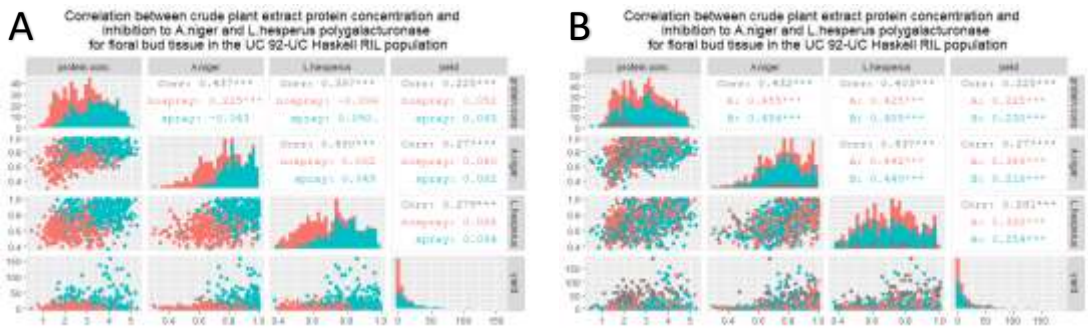

Figure 3.7 Correlations between crude plant extract protein concentration, enzymatic inhibition of *L. hesperus* and *A. niger* polygalacturonase in floral bud tissue of the UC 92 – UC Haskell RIL population A. sub-correlations for the different insecticide treatments B. sub-correlations for the different haplotypes at the peak QTL on chromosome 1. A = UC 92 haplotype, B = UC Haskell haplotype.

hesperus and enzymatic inhibition of *A. niger* with yield were also evaluated: all three biochemical traits showed a significant positive correlation with yield of 23%, 28% and 28%, at the $p < 0.001$ significance level. For all three of these traits, the strength of the correlation within each of the two treatments (insecticide and no insecticide application) was lower than the correlation of the combined treatment data. Within each of the two parental haplotypes at the peak QTL marker, Pl01_42432553, the strength of the correlations were equal or greater than the overall correlations and were all significant at the $p < 0.001$ significance level. Correlations and phenotypic distributions of these traits are shown in Figure 3.7.

Within and across the two treatments (insecticide and no insecticide application), the correlations between crude plant extract protein concentration, enzymatic inhibition of *L. hesperus* and enzymatic inhibition of *A. niger* with yield and yield reduction (between insecticide and no insecticide applications) were also evaluated and are shown in Figure 3.8. There were significant positive correlations between protein concentration in both treatments with the enzymatic inhibitions of both polygalacturonase sources in both treatments, except for the enzymatic inhibition to the *A. niger* polygalacturonase in the no spray treatment only. The only significant correlation identified with yield reduction was the crude plant extract protein concentration and enzymatic inhibition of *A. niger* polygalacturonase in the spray treatments for the UC 92 haplotypes at the peak QTL marker, Pl01_42432553, which were significant at the $p < 0.05$ and 0.10 significance levels.

Figure 3.8 Correlations between crude plant extract protein concentration, enzymatic inhibition of *L. hesperus*, enzymatic inhibition of *A. niger,* and yield for both treatments, for the spray treatment only and for the no spray treatment only. A = UC 92 haplotype, B = UC Haskell haplotype at the peak QTL on chromosome 1.

Principal component analyses were run for the three combinations of all three traits and revealed clustering of data points based on the allelic variant at the peak QTL marker, Pl01_42432553 (Figure 3.9). The clustering is most distinct along the first principal component for the comparisons of crude plant extract protein concentration with the enzymatic inhibition of *L. hesperus* and *A. niger* polygalacturonases, which each account for 98.5% of the variation observed. The clustering is less defined between the inhibition of the two different polygalacturonase sources, with 53% of the variation observed along the first principal component.

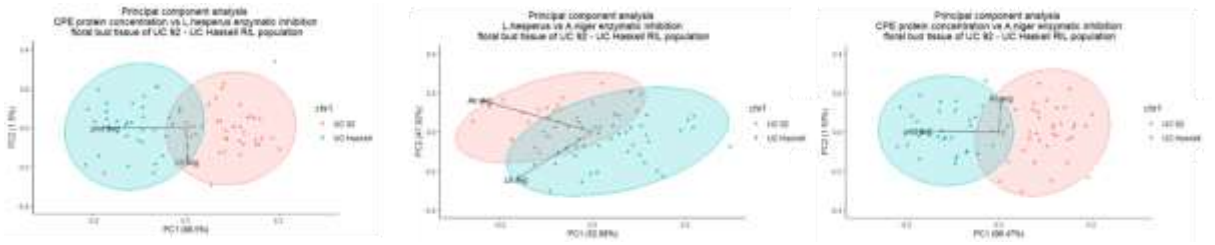

Figure 3.9 Principal component analyses between crude plant extract (CPE) protein concentration, enzymatic inhibition of *L. hesperus* polygalacturonase and enzymatic inhibition of *A. niger* polygalacturonase in floral bud tissue of the UC 92 – UC Haskell RIL population with clustering of haplotypes at the QTL peak on chromosome 1.

Discussion

This research evaluates the effects of lygus pressure and polygalacturonase enzymatic degradation on lima bean reproductive tissue samples and the impact on lima bean yield by analyzing a panel of California adapted cultivars and a recombinant inbred line population. The target biotic pest, *L. hesperus*, was used as a source of polygalacturonase as well as a common black mold, *A. niger*, to identify differences in the specificity of enzymatic inhibition across lima bean genotypes and in different environments.

Treatment effect

One of the findings of this research that is consistent across genotypes, locations and years is that the enzymatic inhibition of *L. hesperus* and *A. niger* polygalacturonases and crude plant extract protein concentration is significantly greater among fields that applied insecticidal treatments to control lygus pressure than fields that did not apply insecticidal treatments. In this research, a significant amount of the variation for each of these three traits is attributed to the insecticidal treatment and suggests there is a strong environmental influence on the crude plant extract protein concentration and the enzymatic inhibition of *L. hesperus* and *A. niger* polygalacturonases.

The inducibility of polygalacturonase-inhibiting proteins or other inhibitory metabolites as a defense mechanism may be a limited defense resource in lima beans and/or pest infestations may reduce the ability of affected plants to produce these proteins or metabolites to inhibit polygalacturonase activity. However, it is difficult to differentiate the effects of increased lygus pressure from the application of insecticides in the split-plot experiment. The insecticide applied in these split-plot trials contains an ingredient that protects against photodegradation of the active ingredient but may also confer some protection against the photodegradation of proteins or metabolites involved in the inhibition of polygalacturonase activity. Minimal literature is available on the effects of this photodegradation-protecting compound or its effects on proteins and metabolites within the plants.

Previous research has studied the effects of herbivore pressure on the expression of polygalacturonase-inhibiting proteins in common bean and found an increase in these proteins

up to 48 hours following herbivore damage (D'Ovidio et al., 2004). A similar upregulation response of polygalacturonase-inhibiting proteins has been observed in different plant species following polygalacturonase damage from different pathogenic species (De Lorenzo et al., 2001; Kalunke et al., 2015; Powell et al., 2000; Sicilia et al., 2005). Further research may be needed in a greenhouse or field setting to study the effects of herbivory or lygus pressure alone on lima bean reproductive tissue, independent of the application of insecticides for herbivore control. These studies confirm that there are significantly different levels of protein concentration and enzymatic inhibition observed in the floral bud tissues.

Correlations among traits

There were positive correlations identified between the crude plant extract protein concentration and the enzymatic inhibition of both polygalacturonase sources across genotypes, locations, and treatments. However, when the data were split by the insecticidal treatment applied, the correlation between crude plant extract protein concentration and the enzymatic inhibition of both polygalacturonase sources within each treatment was insignificant. The latter finding is consistent with previous research that found an insignificant relationship between crude plant extract protein concentration and the enzymatic inhibition of polygalacturonase across different lima bean genotype samples (Dashner, 2016). Since treatment was the most significant factor for all three traits, the initial findings of a positive correlation between crude plant extract protein concentration and the enzymatic inhibition of both polygalacturonase sources may be influenced by the two treatment effects which create unique clusters of the data.

Another consistent finding between this work and Dashner's is the positive correlation within a biological replicate of crude plant extract protein concentration and the enzymatic inhibition of both sources of polygalacturonase for UC 92 and UC Haskell (Dashner, 2016). This positive correlation exists within a genotype's biological replicate; however, across genotypes, different biological replicates, and different environments, the correlation between crude plant extract protein concentration and enzymatic inhibition of polygalacturonase weakens. This observation highlights the variation and environmental influence on these traits.

There were positive correlations between all three biochemical traits and yield across genotypes, locations, and treatments, with the strongest correlation between yield being with the enzymatic inhibition of *L. hesperus*. Inhibition to *L. hesperus* polygalacturonase is the most relevant biochemical trait to the field experiment, where lygus pressure is seen as the greatest threat to lima bean yields. The significance of the correlations between the three biochemical traits and yield may be strongly influenced by the effect of the insecticide treatment on the biochemical traits and yield since there are no significant correlations between any of the biochemical traits and yield within either treatment separately.

Heritability

Although the insecticide treatment was the most significant factor in the California cultivar and recombinant inbred line population studies, there were also significant differences among genotypes in both studies, which supports that there is a genetic component of these traits that confers inhibition of polygalacturonase damage in the floral bud tissue of lima bean. The calculated heritability's of enzymatic inhibition of the two polygalacturonase sources and the crude plant extract protein concentration ranged between 1 – 15 % in the recombinant

inbred population and 2 – 16 % in the California cultivar study and suggests that this is a heritable trait that can be selected for in breeding programs. All three biochemical traits have low broad-sense heritability and demonstrate greater phenotypic plasticity under the different environmental conditions of the split-plot trial. All three biochemical traits declined in environments that did not control for lygus pressure, suggesting that this mechanism of resistance is not inducible or has limitations for its inducibility following biotic stress in this recombinant inbred line population.

QTL mapping and candidate genes

QTL mapping in the UC 92 – UC Haskell recombinant inbred line population identified significant peaks for both crude plant extract protein concentration and for the enzymatic inhibition of *A. niger* polygalacturonase colocating at the same locus on chromosome Pl01. This locus at 42.4 Mbp is the first following a 20 cM and 20 Mbp gap across the pericentromeric region of chromosome Pl01, the lowest coverage region of the entire genome for this mapping population. The identification of these significant co-locating QTLs for two of the three traits studied offers promise in utilizing marker-assisted-selection for these traits that confer a level of resistance to lygus and fungal tissue damage in lima beans.

While the low marker coverage on chromosome Pl01 limits the precision of the QTL mapping across this region, there are several functionally relevant genes within the significance range of our QTL peak. The most notable candidate gene in the significance range for our QTL peak is a *Pgip* gene, Pl01G0000281800.v1 (PAC:44364665), annotated on the lima bean reference genome at 34.9 Mbp. This region on the long arm of chromosome Pl01 in common

bean (*P. vulgaris*) also mapped a *Pgip* locus in BAC clones of the BAT93 genotype (of the Mesoamerican domesticated gene pool), which is not annotated in the G19833 (Andean domesticated gene pool) common bean reference genome (Kalunke et al., 2014; Schmutz et al., 2014). The Pl01G0000281800.1.v1 transcript shares 98% similarity with the PvUl111.01G147900.1 transcript annotated on the common bean whole- genome sequence of genotype Pinto UI111 (Mesoamerican domesticated gene pool, ecogeographic race Durango) ("Phytozome v13 info: P.vulgaris_UI111 v1.1," 2020) for the same family of polygalacturonase inhibitor 1 proteins.

Around the *Pgip* gene on chromosome Pl01 are a series of genes related to the stress response and protein synthesis that may be involved in the signaling pathway and mediation of polygalacturonase inhibition. Prior research has identified different pathways of polygalacturonase inhibition protein promotion, including the role of salicylic acid, jasmonic acid, and ethylene in the chemical defense response (Akagi et al., 2010; Edwards & Singh, 2006; Ferrari et al., 2007; R. Li et al., 2003).

There are other relevant genes on this region of chromosome Pl01 involved in the defense response and potential upregulation of polygalacturonase-inhibiting proteins. There are 7 transcription factor genes including Pl01G0000281900.v1 at 34.9 Mbp related to the wounding stress response involving jasmonic acid and salicylic acid signaling pathways, Pl01G0000305500.v1, Pl01G00003055600.v1 and Pl01G00003055900.v1 at 37.4, 37.4 and 37.5 Mbp related to the defense response involving jasmonic acid and ethylene signaling pathways, Pl01G0000307400.v1 at 37.7 Mbp related to the stress response to jasmonic acid, Pl01G0000310000.v1 at 37.9 Mbp related to the defense and wounding response to bacterium,

jasmonic acid, salicylic acid and ethylene, Pl01G0000322000.v1 at 39.1 Mbp related to the stress response to jasmonic acid, salicylic acid and ethylene. Two additional non-transcription factor genes are involved in the defense response, including Pl01G0000325800.v1 at 39.5 Mbp related to the defense response to bacterium, fungus and jasmonic-acid- and ethylene-dependent systemic resistance, and Pl01G0000313300.v1 at 38.2 Mbp related to transmembrane transport and the response to jasmonic acid, salicylic acid and ethylene.

Other related candidate genes in this region of chromosome Pl01 include genes involved in protein synthesis including Pl01G0000284800.v1 at 35.3 Mbp related to protein-folding chaperones, Pl01G0000296900.v1 at 36.5 Mbp, Pl01G0000303600.v1 at 37.2 Mbp, Pl01G0000339200.v1 at 40.8 Mbp, Pl01G0000342100.v1 and Pl01G0000342200.v1 at 41.1 Mbp related to Kelch-protein synthesis, and Pl01G0000311900.v1 at 38.1 Mbp related to protein glycosylation and salicylic acid signaling pathways.

Enzyme source and protein specificity

The evolutionary adaptations of polygalacturonases and polygalacturonase-inhibiting proteins has resulted in a variety of polygalacturonase-inhibiting protein genes with varying specificities in plants. The differences of the inhibitory effects of polygalacturonase-inhibiting proteins from different tissues within the same plant has been attributed to slight differences in glycosylation in the tissues (Powell et al., 2000). The production response of a given polygalacturonase-inhibiting protein, i.e. PvPGIP1, within a plant does not always provide successful inhibition of the triggering polygalacturonase, and suggests that the protein concentration of crude plant extract may not always be indicative of the inhibitory effects of

the tissue sample (Manfredini et al., 2005). This may also be demonstrated in this study in how genotype tissue samples with a known crude plant extract protein concentration can vary in their enzymatic inhibition depending on the polygalacturonase source, i.e., the reversal of superior inhibition between UC 92 and UC Haskell for *L. hesperus* and *A. niger* polygalacturonase sources.

The crude plant extract protein concentration is purified and quantified using the Bradford Assay, which provides a measurement of general protein concentration and not a measurement of specific polygalacturonase-inhibiting protein concentration. Although the Bradford Assay is widely used in polygalacturonase studies to quantify polygalacturonase-inhibiting proteins, it does not directly or specifically measure polygalacturonase-inhibiting proteins in a sample (Wu et al., 2012). It is acknowledged that due to the high specificity of polygalacturonase-inhibiting proteins interactions to a given source of polygalacturonase, even the validation of the presence of polygalacturonase-inhibiting proteins does not always confer effective inhibition of specific polygalacturonase enzymes (Kalunke et al., 2015; Leckie et al., 1999). The high interaction specificity between the enzyme and inhibiting protein may also contribute to the weak correlations between crude plant extract protein concentration and polygalacturonase enzymatic inhibition.

The significant differences among genotypes in these studies are consistent with the findings of Dashner, which identified significant differences between UC 92 and UC Haskell, with UC 92 having superior enzymatic inhibition for *A. niger* polygalacturonase activity (Dashner, 2016). Dashner also identified non-synonymous nucleotide substitutions on the PlPGIP2 gene between the UC 92 and UC Haskell lines, substitutions which for the PvPGIP2

gene led to structural changes of the polygalacturonase-inhibiting protein that could impact specificity and inhibition against polygalacturonases (Dashner, 2016; Farina et al., 2009; Leckie et al., 1999). Further research needs to be conducted on the non-synonymous nucleotide substitutions in the PlPGIP gene on chromosome Pl01 to understand the relevance of the identified QTL peak.

Colorimetric assay

Previous research conducted on polygalacturonase-inhibiting protein activity has been limited in the number of biological replicates, genotypes and environments evaluated, and also limited in the conclusions drawn related to this mechanism due to the nature of the evaluating assay. The radial diffusion assay is a commonly used method to measure the effects of polygalacturonase activity and polygalacturonase-inhibiting protein activity, however it can limit the number of samples that can be run simultaneously and is time-consuming in the incubation of the enzymatic assay and the final radial measurement for analysis (R. J. Taylor, 1988). The protocol adopted in this research draws upon the adapted microplate assay that Dashner developed, while making minor adjustments based on experimental differences and recommendations from Dashner (Dashner, 2016; Ortiz et al., 2014; Torres et al., 2011). The consistency of this study's findings with Dashner's validates the use of the colorimetric assay for measuring the inhibition of polygalacturonase enzymatic activity, especially for high-throughput phenotyping of large sample sizes (Dashner, 2016).

Conclusion

This work presents the results of a split-plot analysis of the effects of insecticidal applications and lygus pressure on the protein concentration and enzymatic inhibition of floral bud tissue in a collection of California-adapted cultivars and a segregating population of lima beans. The results show a strong environmental influence on both crude plant extract protein concentration and the enzymatic inhibition of two distinct sources of polygalacturonase: the insect, *L. hesperus,* and the fungus, *A. niger.* While this study is unable to fully distinguish the effects of the insecticidal application from the lygus pressure, it does highlight the strong environmental influence on these biochemical traits. This study also reveals significant genotypic variability among lima bean cultivars and recombinant inbred lines, identifies a significant region on chromosome Pl01 underlying two of the three traits studied for polygalacturonase inhibition, establishes a positive correlation between all three biochemical traits with yield, and demonstrates the heritability of polygalacturonase inhibition as a mechanism for lygus resistance in lima beans. Significant co-locating QTLs were identified in the recombinant inbred line population for two of the three traits studied, which both have positive correlations with yield. The QTLs identified offer hope in utilizing marker-assisted-selection for these traits that confer a level of resistance to lygus and fungal tissue damage in lima beans.

References

Akagi, A., Engelberth, J., & Stotz, H. U. (2010). Interaction between polygalacturonase-inhibiting protein and jasmonic acid during defense activation in tomato against Botrytis cinerea. *European Journal of Plant Pathology*, *128*(4), 423–428. https://doi.org/10.1007/s10658-010-9684-z

Albersheim, P., & Anderson, A. J. (1971). Proteins from plant cell walls inhibit polygalacturonases secreted by plant pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, *68*(8), 1815–1819. https://doi.org/10.1073/pnas.68.8.1815

Aziz, A., Heyraud, A., & Lambert, B. (2004). Oligogalacturonide signal transduction, induction of defense-related responses and protection of grapevine against Botrytis cinerea. *Planta*, *218*(5), 767–774. https://doi.org/10.1007/s00425-003-1153-x

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bergmann, C. W., Ito, Y., Singer, D., Albersheim, P., Darvill, A. G., Benhamou, N., … De Lorenzo, G. (1994). Polygalacturonase-inhibiting protein accumulates in Phaseolus vulgaris L. in response to wounding, elicitors and fungal infection. *The Plant Journal*, *5*(5), 625–634. https://doi.org/10.1111/j.1365-313X.1994.00625.x

Broman, K. W. (2009). *Statistics for Biology and Health A Guide to QTL Mapping with R / qtl*. https://doi.org/10.1007/978-0-387-92125-9

Celorio-Mancera, M. de la P., Allen, M. L., Powell, A. L., Ahmadi, H., Salemi, M. R., Phinney, B. S., … Labavitch, J. M. (2008). Polygalacturonase causes lygus-like damage on plants: Cloning and identification of western tarnished plant bug (Lygus hesperus) polygalacturonases secreted during feeding. *Arthropod-Plant Interactions*, *2*(4), 215–225. https://doi.org/10.1007/s11829-008-9050-7

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005, January). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*. https://doi.org/10.1007/s10681-005-1681-5

D'Ovidio, R., Mattei, B., Roberti, S., & Bellincampi, D. (2004, February 12). Polygalacturonases, polygalacturonase-inhibiting proteins and pectic oligomers in plant-pathogen interactions. *Biochimica et Biophysica Acta - Proteins and Proteomics*. Elsevier. https://doi.org/10.1016/j.bbapap.2003.08.012

D'Ovidio, R., Raiola, A., Capodicasa, C., Devoto, A., Pontiggia, D., Roberti, S., … De Lorenzo, G. (2004). Characterization of the complex locus of bean encoding polygalacturonase-inhibiting proteins reveals subfunctionalization for defense against fungi and insects. *Plant Physiology*, *135*(4), 2424–2435. https://doi.org/10.1104/pp.104.044644

Dashner, Z. (2016). *Examination of Lygus Bug Resistance in Lima Bean: Polygalacturonase Inhibiting Proteins as Candidate Traits*. University of California, Davis.

De Lorenzo, G., D'Ovidio, R., & Cervone, F. (2001, November 28). The role of polygalacturonase-inhibiting proteins (PGIPS) in defense against pathogenic fungi. *Annual Review of Phytopathology*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA . https://doi.org/10.1146/annurev.phyto.39.1.313

Edwards, O., & Singh, K. B. (2006). Resistance to insect pests: What do legumes have to offer? *Euphytica*, *147*(1–2), 273–285. https://doi.org/10.1007/s10681-006-3608-1

Farina, A., Rocchi, V., Janni, M., Benedettelli, S., De Lorenzo, G., & D'Ovidio, R. (2009). The bean polygalacturonase-inhibiting protein 2 (PvPGIP2) is highly conserved in common bean (Phaseolus vulgaris L.) germplasm and related species. *Theoretical and Applied Genetics*, *118*(7), 1371–1379. https://doi.org/10.1007/s00122-009-0987-4

Favaron, F., Sella, L., & D'Ovidio, R. (2004). *Relationships Among Endo-Polygalacturonase, Oxalate, pH, and Plant Polygalacturonase-Inhibiting Protein (PGIP) in the Interaction Between Sclerotinia sclerotiorum and Soybean*. *Molecular Plant-Microbe Interactions MPMI* (Vol. 17).

Ferrari, S., Galletti, R., Denoux, C., De Lorenzo, G., Ausubel, F. M., & Dewdney, J. (2007). Resistance to Botrytis cinerea induced in arabidopsis by elicitors is independent of salicylic acid, ethylene, or jasmonate signaling but requires PHYTOALEXIN DEFICIENT3. *Plant Physiology*, *144*(1), 367–379. https://doi.org/10.1104/pp.107.095596

Ferrari, S., Vairo, D., Ausubel, F. M., Cervone, F., & De Lorenzo, G. (2003). Tandemly Duplicated Arabidopsis Genes That Encode Polygalacturonase-Inhibiting Proteins Are Regulated Coordinately by Different Signal Transduction Pathways in Response to Fungal Infection. *The Plant Cell*, *15*, 93–106. https://doi.org/10.1105/tpc.005165

Fox, J., & Weisbery, S. (2019). *An R Companion to Applied Regression - John Fox, Sanford Weisberg - Google Books*. *Sage Publications*. Retrieved from https://books.google.com/books?hl=en&lr=&id=uPNrDwAAQBAJ&oi=fnd&pg=PP1&dq=fox+weisberg+car&ots=MwF64I2A61&sig=PZXVVsfUb19NftmcF9vV1uuahhU#v=onepage&q=fox weisberg car&f=false

Frati, F., De Lorenzo, G., Salerno, G., Galletti, R., & Conti, E. (2006). Activity of endo-polygalacturonases in mirid bugs (Heteroptera:Miridae) and their inhibition by plant cell wall proteins (PGIPs). *Eur. J. Entomol.* https://doi.org/10.14411/eje.2006.067

Holland, J. B., Nyquist, W. E., & Cervantes-Martinez, C. T. (2003). Estimating and interpreting heritability for plant breeding: An update. *Plant Breeding Reviews*, *22*. Retrieved from https://www.google.com/search?q=holland+et+al+2003+heritability&rlz=1C1CHBF_enUS767US767&oq=holland&aqs=chrome.0.69i59j46l4j69i60j69i61j69i60.1248j0j7&sourceid=chrome&ie=UTF-8

Kalunke, R. M., Cenci, A., Volpi, C., O'Sullivan, D. M., Sella, L., Favaron, F., … D'Ovidio, R. (2014).

The pgip family in soybean and three other legume species: Evidence for a birth-and-death model of evolution. *BMC Plant Biology*, *14*(1), 189. https://doi.org/10.1186/s12870-014-0189-3

Kalunke, R. M., Tundo, S., Benedetti, M., Cervone, F., De Lorenzo, G., & D'Ovidio, R. (2015). An update on polygalacturonase-inhibiting protein (PGIP), aleucine-rich repeat protein that protects crop plants against pathogens. *Frontiers in Plant Science*, *6*(MAR), 1–17. https://doi.org/10.3389/fpls.2015.00146

Leckie, F., Mattei, B., Capodicasa, C., Hemmings, A., Nuss, L., Aracri, B., … Cervone, F. (1999). The specificity of polygalacturonase-inhibiting protein (PGIP): A single amino acid substitution in the solvent-exposed β-strand/β-turn region of the leucine-rich repeats (LRRs) confers a new recognition capability. *EMBO Journal*, *18*(9), 2352–2363. https://doi.org/10.1093/emboj/18.9.2352

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*. https://doi.org/10.18637/jss.v069.i01

Li, R., Rimmer, R., Yu, M., Sharpe, A. G., Séguin-Swartz, G., Lydiate, D., & Hegedus, D. D. (2003). Two Brassica napus polygalacturonase inhibitory protein genes are expressed at different levels in response to biotic and abiotic stresses. *Planta*, *217*(2), 299–308. https://doi.org/10.1007/s00425-003-0988-5

Long, R., Temple, S., Meyer, R., Schwankl, L., Godfrey, L., Canevari, M., & Roberts, P. (2014). Lima Bean Production in California. *Lima Bean Production in California*, (May). https://doi.org/10.3733/ucanr.8505

Lüdecke, D. (2020). Package 'sjstats'. Collection of convenient functions for common statistical computations. *Statistical Functions for Regression Models*.

Manfredini, C., Sicilia, F., Ferrari, S., Pontiggia, D., Salvi, G., Caprari, C., … Lorenzo, G. De. (2005). Polygalacturonase-inhibiting protein 2 of Phaseolus vulgaris inhibits BcPG1, a polygalacturonase of Botrytis cinerea important for pathogenicity, and protects transgenic plants from infection. *Physiological and Molecular Plant Pathology*, *67*(2), 108–115. https://doi.org/10.1016/j.pmpp.2005.10.002

Matteo, A. Di, Federici, L., Mattei, B., Salvi, G., Johnson, K. A., Savino, C., … Cervone, F. (2003). *The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense*. *National Acad Sciences*. Retrieved from www.rcsb.org

Mueller, S. C., Summers, C. G., & Goodell, P. B. (2003). A field key to the most common Lygus species found in agricultural fields of the central San Joaquin Valley of California. *University of California Division of Agriculture and Natural Resources*, 1–12. Retrieved from http://cottoninfo.ucdavis.edu/files/133229.pdf

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Ortiz, G. E., Guitart, M. E., Albertó, E., Fernández Lahore, H. M., & Blasco, M. (2014). Microplate assay for endo-polygalacturonase activity determination based on ruthenium red method. *Analytical Biochemistry*, *454*(1), 33–35. https://doi.org/10.1016/j.ab.2014.02.028

Phytozome v13 info: P.vulgaris_UI111 v1.1. (2020). Retrieved December 29, 2020, from https://phytozome-next.jgi.doe.gov/info/PvulgarisUI111_v1_1

Pickersgill, R., Smith, D., Worboys, K., & Jenkins, J. (1998). Crystal structure of polygalacturonase from Erwinia caratovora ssp. carotovora. *Journal of Biological Chemistry*, *273*(38), 24660–24664. https://doi.org/10.1074/jbc.273.38.24660

Powell, A. L. T., Van Kan, J., Ten Have, A., Visser, J., Greve, L. C., Bennett, A. B., & Labavitch, J. M. (2000). Transgenic expression of pear PGIP in tomato limits fungal colonization. *Molecular Plant-Microbe Interactions*, *13*(9), 942–950. https://doi.org/10.1094/MPMI.2000.13.9.942

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundatrion for Statistical Computing. Retrieved from http://www.r-project.org/

Schmutz, J., Jackson, S. A., Wang, M., Wing, R. A., Gao, D., Zhang, M., … Cannon, S. B. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics*, *46*(7), 707–713. https://doi.org/10.1038/ng.3008

Shackel, K. A., Celorio-Mancera, M. de la P., Ahmadi, H., Greve, L. C., Teuber, L. R., Backus, E. A., & Labavitch, J. M. (2005). Micro-injection of lygus salivary gland proteins to simulate feeding damage in alfalfa and cotton flowers. *Archives of Insect Biochemistry and Physiology*, *58*(2), 69–83. https://doi.org/10.1002/arch.20033

Sicilia, F., Fernandez-Recio, J., Caprari, C., De Lorenzo, G., Tsernoglou, D., Cervone, F., & Federici, L. (2005). The Polygalacturonase-Inhibiting Protein PGIP2 of Phaseolus vulgaris Has Evolved a Mixed Mode of Inhibition of Endopolygalacturonase PG1 of Botrytis cinerea 1. *Am Soc Plant Biol*. https://doi.org/10.1104/pp.105.067546

Taylor, R. J. (1988). An Improved Diffusion Assay for Quantifying the Polygalacturonase Content of Erwinia Culture Filtrates. *Phytopathology*, *78*(8), 1101. https://doi.org/10.1094/phyto-78-1101

Torres, S., Sayago, J. E., Ordoñez, R. M., & Isla, M. I. (2011). A colorimetric method to quantify endo-polygalacturonase activity. *Enzyme and Microbial Technology*, *48*(2), 123–128. https://doi.org/10.1016/j.enzmictec.2010.04.004

USDA ERS. (2004). *Economic research service*. *Choice Reviews Online* (Vol. 42). https://doi.org/10.5860/choice.42-2206

USDA NASS. (2019). *USDA/NASS QuickStats*. Retrieved from https://quickstats.nass.usda.gov/results/9979DE70-2420-3A8E-A718-4D3E4C830346

Van Santen, Y., Benen, J. A. E., Schröter, K. H., Kalk, K. H., Armand, S., Visser, J., & Dijkstra, B. W. (1999). 1.68-Å crystal structure of endopolygalacturonase II from Aspergillus niger and

identification of active site residues by site-directed mutagenesis. *Journal of Biological Chemistry*, *274*(43), 30474–30480. https://doi.org/10.1074/jbc.274.43.30474

Vasconcellos, R. C. C., Lima, T. F. C., Fernandes-Brum, C. N., Chalfun-Junior, A., & Santos, J. B. (2016). Expression and validation of PvPGIP genes for resistance to white mold (Sclerotinia sclerotiorum) in common beans (Phaseolus vulgaris L.) of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License. *Genetics and Molecular Research*, *15*(3). https://doi.org/10.4238/gmr.15038269

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., & Zemla, J. (2017). R package "corrplot": Visualization of a Correlation Matrix. *Statistician*, *56*, 316–324. Retrieved from https://peerj.com/articles/9945/Supplemental_Data_S10.pdf

Wu, H., Chen, L., Pan, G., Tu, C., Zhou, X., & Mo, L. (2012). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, *235*(5), 248–254. https://doi.org/10.1016/j.foodcont.2008.03.004

Final Conclusion and Discussion

Lima bean, *Phaseolus lunatus,* is the second most agronomically important *Phaseolus* species, and improvements in varietal development have been hindered by the lack of genetic and genomic resources available for this crop. Without the availability of a high-quality linkage map or reference genome, lima bean research has relied upon the deferment to common bean, *Phaseolus vulgaris,* resources and minimized the differences between the species. The first chapter focuses on the sequencing of a recombinant inbred line population derived from two California-adapted cultivars of differing gene pool origins (Mesoamerican and Andean), and the development of a high-quality molecular linkage map. This linkage map was aligned to both the novel lima bean reference genome and the common bean reference genome, which highlights the chromosomal rearrangements and divergence of these two related *Phaseolus* species. The second chapter focuses on mapping key agronomic and domestication traits in this population and identifying regions of collocation and homology between lima bean and common bean.

Biotic stress due to lygus, *Lygus hesperus* Knight, pressure is a main factor impacting lima bean productivity throughout California, as this insect is found throughout the western half of North America. Biotic stress from lygus as well as other related herbivores may drive adaptation to this stress in wild lima beans and may have been influenced by the domestication process as well. The third and fourth chapters focus on evaluating a recombinant inbred line population in a split-plot trial evaluating the effects of lygus and two potential mechanisms of resistance to lygus herbivory, cyanogenesis and polygalacturonase-inhibiting proteins, and their impact on yield productivity in lima beans. The third chapter additional includes a study

evaluating the differences in cyanogenesis across different statuses of wild and domesticated lima beans, to understand if this trait may have been actively selected for or against during the domestication process.

Development of a linkage map for lima beans and mapping of key traits

The development of this linkage map for lima beans has already enabled the improvement of the first high-quality annotated reference genome assembly for lima bean, which provides a wealth of knowledge relevant for advancements in molecular breeding and comparative genomic analysis (Garcia et al., 2021; "Phytozome v13 info: P.lunatus V1," 2020) In the absence of this linkage map, gaps in the assembly across the pericentromeric regions of chromosomes were inferred as being synonymous with the common bean reference genome. However, one of the most unique rearrangements in the lima bean genome is the complete inversion of the short arm of chromosome 10, which would not have been recognized without the contribution of the linkage map.

The lima bean linkage map highlights the chromosomal rearrangements and divergence between lima bean and its relative, common bean, and highlights the uniqueness of lima bean in having expansive pericentromeric regions for its relatively small genome size. The calculation of recombination rates across the genome allows the delineation of the euchromatic and pericentromeric regions for each chromosome and can better inform breeders of the likelihood of recombination for traits according to their location across the lima bean genome.

There were a few limitations of the linkage map, which include the scarcity of markers across the pericentromeric regions of chromosome 1 and chromosome 9. The scarcity of

markers across the pericentromeric region of chromosome 1 is even more problematic due to the number of important traits that have significant QTLs map to this region. The scarcity of markers provides a lack of clarity in the true location of these QTLs in the region, and the underlying cause of this marker scarcity is still unknown. This absence of markers across this region was consistent when the linkage map was assembled with reference to both the lima bean and common bean reference genomes. The lima bean reference genome is based on a Mesoamerican line and the common bean reference genome is based on an Andean line, which suggest that the distinction of either gene pool is not wholly responsible for the absence of markers. For both parents of the recombinant inbred line, there was consistently lower read depths in this region, however SNPs were still identified throughout, and no significant deletion regions were identified between the reference genomes and either parents. Additionally, since multiple traits have QTLs that map to this region of chromosome 1, it suggest that there are polymorphisms and functional genes across this region, but it is not understood why there is poor marker coverage in this region. The underlying cause of this marker scarcity in this chromosomal region is of particular interest for future lima bean research, since traits related to determinacy, maturity, and biotic stress tolerance map to this region.

Mapping of key agronomic and domestication traits in lima beans reveals that chromosome 1 contains a region of strong importance for domestication and adaptation traits in lima bean, involving the colocation of QTLs related to determinacy, photoperiod sensitivity, flowering time and inflorescence position in the UC 92 – UC Haskell recombinant inbred line population. The location of the QTL for determinacy shares an ortholog with the PvTFL1y locus on common bean, with the phenotype for determinacy being contributed by the Andean

genotype in both species (Campa et al., 2018; Repinski et al., 2012). A strong correlation between the determinate habit, earlier flowering time, and photoperiod insensitivity in lima bean is also corroborated by the strong association between these two traits in common beans (Koinange et al., 1996; Weller et al., 2019).

Transgressive segregation was observed for multiple traits throughout this biparental population, including flowering time, inflorescence position, plant height and pod set density. Several traits had limitations of their transgressive segregation towards one extreme in the progeny, which includes lower seed weights and yields, and presence of seed coat color (instead of white color). The appearance of seed coat color in the progeny, notable because of its absence in the parental phenotypes, involves the complementarity and epistatic interaction of QTLs. Unfortunately, for seed weight and yields, there directionality of this segregation away from the more desirable traits may limit the breeding potential towards larger seeds and greater yields from this population. Strong correlations between plant height and pod set density with yield and the identification of significant QTL for each of these traits, may increase the ability to efficiently select for these phenotypes with associated markers to improve productivity in lima bean breeding.

Evaluation of putative mechanisms of resistance to lygus herbivory

California is a major region for dry lima bean production in the United States, and the largest threat to lima bean yields throughout the state is lygus damage. Two mechanisms of resistance were evaluated to understand their role in biotic stress resistance related to yield productivity, cyanogenesis and polygalacturonase-inhibiting proteins. Cyanogenesis has been

studied as a mechanism of biotic stress resistance in lima beans as a form of resistance to bruchid beetles, and has additionally been studied in white clover, cassava and sorghum, while polygalacturonase-inhibiting proteins has been studied as a mechanism of biotic stress resistance to lygus specifically in cotton and alfalfa (Ballhorn et al., 2008; Gleadow & Møller, 2014; Shackel et al., 2005).

Ontogenetic variability has been reported in plants that express cyanide, and are supported in these results across lima bean reproductive tissues (Frehner et al., 1990; Gleadow & Møller, 2014). Volatile cyanide production was higher in the younger reproductive floral bud and immature pod tissues and decreased to negligible levels in the mature pod and fresh seed tissue. While this is encouraging from a food safety perspective to have negligible cyanide in the edible tissues, it demonstrates the remobilization of cyanogenic glucosides from senescing to developing tissues in lima beans (Frehner et al., 1990; Halkier & Du, 1997).

Genotypic variability was observed across cultivated, wild and landrace status accessions of lima beans for volatile cyanide production in the young reproductive tissues. Results of the variability in volatile cyanide production in the immature reproductive tissues in both domesticated and wild accessions of lima beans suggest that there was not a strong selection against this trait in lima beans during the domestication process. High broad-sense heritability was consistently reported in the volatile cyanide production of the immature reproductive tissues throughout the multiple studies and environments. Genotypic variability was observed among cultivars evaluated and in the recombinant inbred line population, which had a bimodal distribution of volatile cyanide production in both immature reproductive tissues, signaling an underlying major gene. Volatile cyanide production is a highly heritable

trait, and QTL mapping within a biparental recombinant inbred line population identified the single major locus on chromosome 5 for both floral bud and immature pod tissue volatile cyanide production. This major locus with additional minor QTLs explain over 98% of the phenotypic variation for both immature reproductive tissues of this highly heritable trait.

The phenotypic plasticity of this trait under different environmental stressors was also observed with the highest levels of volatile cyanide reported in floral bud and immature pod samples collected from fields without insecticidal treatment to control for lygus pressure. The lowest levels of volatile cyanide across environments were observed in reproductive tissue samples collected from controlled greenhouse environments, which also had the greatest level of stability across different years compared to measurements collected from field environments across different locations and years. The increase in volatile cyanide production in uncontrolled field environments exposed to lygus pressure demonstrates the inducibility of this trait following herbivory pressure in the young lima bean reproductive tissues.

Unfortunately, while volatile cyanide production may offer a level of resistance to herbivory there is an expensive allocation of plant resources to deploy these protective compounds and there was no significant correlation between volatile cyanide production and yield in this study, except within the different haplotype groups of the most significant QTL in the recombinant inbred line population (Gleadow & Møller, 2014). There was a negative correlation between yield and volatile cyanide production in the floral bud tissue for the higher producing volatile cyanide haplotype group, and a positive correlation between yield and volatile cyanide production in the immature pod tissue for the lower producing volatile cyanide haplotype group.

Genotypic variation was observed in the inhibition to different sources of polygalacturonase enzymes across the recombinant inbred population, and each parent had superior dominance to each of the two sources of polygalacturonase. The UC 92 parent of Andean origin had superior inhibition to the *A. niger* source of polygalacturonase, while the UC Haskell parent of Mesoamerican origin had superior inhibition to the *L. hesperus* source of polygalacturonase in their floral bud tissue. The inhibition to either polygalacturonase sources has low broad-sense heritability and greater phenotypic plasticity under the different environmental conditions of the split-plot trial for lygus pressure. The crude plant extract protein concentration and the inhibition to both polygalacturonase sources was greater in the environments that controlled for lygus pressure, suggesting that this mechanism of resistance is not inducible or has limitations for its inducibility following biotic stress in this recombinant inbred line population.

Polygalacturonase-inhibiting proteins may offer a level of resistance to herbivory in lima beans since there were positive correlations between the inhibition of *L. hesperus* polygalacturonase and yield in this study. Additionally, there was a significant positive correlation between the inhibition of *A. niger* polygalacturonase and yield in this study, which may infer resistance to this and other related molds in lima beans and may be important to the succulent lima bean market in other regions. Polygalacturonase inhibition to both of these sources has low heritability, and is more strongly influenced by the environment, which may limit its effectiveness as traits for biotic stress resistance in lima beans.

While polygalacturonase-inhibiting proteins have a significant positive correlation with yield but low heritability, and cyanogenesis has high heritability but a nonsignificant correlation

with yield, it highlights the complication of understanding traits that influence biotic stress resistance in lima beans. Undoubtedly, biotic stress resistance that mitigates yield loss due to lygus pressure in lima beans is a quantitative trait involving multiple physiological and biochemical traits.

The beginning of the long arm of chromosome 1 presents itself as an important region for agronomic, domestication and stress resistance traits in this population of lima beans. Two QTLs related to polygalacturonase inhibition and another QTL related to volatile cyanide production mapped to the same region of chromosome 1 that collocates with the QTLs for determinacy, flowering time, and inflorescence position. Further marker clarity in this region may allow for fine-mapping of important QTLs in this region to identify causative genes for these traits and is emphasized as an important chapter of further lima bean genetic research.

Collectively, this dissertation provides a high-quality genetic map as a key genetic and genomic resource for lima bean research, identifies QTLs for key agronomic and domestication traits that provide insight into the synteny and rearrangements between lima bean and its closest relative, common bean. This genetic map combined with analysis of two putative mechanisms for biotic stress resistance allows QTL mapping of biotic stress resistance traits, which can aid in the development and advancement of lima bean cultivars with improved biotic stress resistance in California.

References

Ballhorn, D. J., Kautz, S., Lion, U., & Heil, M. (2008). Trade-offs between direct and indirect defences of lima bean (Phaseolus lunatus). *Journal of Ecology*, *96*(5), 971–980. https://doi.org/10.1111/j.1365-2745.2008.01404.x

Campa, A., Murube, E., & Ferreira, J. J. (2018). Genetic Diversity, Population Structure, and Linkage Disequilibrium in a Spanish Common Bean Diversity Panel Revealed through Genotyping-by-Sequencing. *Genes*. https://doi.org/10.3390/genes9110518

Frehner, M., Scalet, M., & Conn, E. E. (1990). Pattern of the Cyanide-Potential. *Plant Physiology*, *94*, 28–34.

Garcia, T., Duitama, J., Zullo, S. S., Gil, J., Ariani, A., Dohle, S., … Chacón-Sánchez, M. I. (2021). Comprehensive genomic resources related to domestication and crop improvement traits in Lima bean. *Nature Communications*, *12*(1), 1–17. https://doi.org/10.1038/s41467-021-20921-1

Gleadow, R. M., & Møller, B. L. (2014). Cyanogenic glycosides: Synthesis, physiology, and phenotypic plasticity. *Annual Review of Plant Biology*. Annual Reviews Inc. https://doi.org/10.1146/annurev-arplant-050213-040027

Halkier, B. A., & Du, L. (1997). The biosynthesis of glucosinolates. *Trends in Plant Science*, *2*(11), 425–431. https://doi.org/10.1016/S1360-1385(97)01128-X

Koinange, E. M. K., Singh, S. P., & Gepts, P. (1996). Genetic Control of the Domestication Syndrome in Common Bean. *Crop Science*, *36*(4), 1037–1045. https://doi.org/10.2135/cropsci1996.0011183X003600040037x

Phytozome v13 info: P.lunatus V1. (2020). Retrieved December 30, 2020, from https://phytozome-next.jgi.doe.gov/info/Plunatus_V1

Repinski, S. L., Kwak, M., & Gepts, P. (2012). The common bean growth habit gene PvTFL1y is a functional homolog of Arabidopsis TFL1. *Theoretical and Applied Genetics*, *124*(8), 1539–1547. https://doi.org/10.1007/s00122-012-1808-8

Shackel, K. A., Celorio-Mancera, M. de la P., Ahmadi, H., Greve, L. C., Teuber, L. R., Backus, E. A., & Labavitch, J. M. (2005). Micro-injection of lygus salivary gland proteins to simulate feeding damage in alfalfa and cotton flowers. *Archives of Insect Biochemistry and Physiology*, *58*(2), 69–83. https://doi.org/10.1002/arch.20033

Weller, J. L., Vander Schoor, J. K., Perez-Wright, E. C., Hecht, V., González, A. M., Capel, C., … Santalla, M. (2019). Parallel origins of photoperiod adaptation following dual domestications of common bean. *Journal of Experimental Botany*, *70*(4), 1209–1219. https://doi.org/10.1093/jxb/ery455

Appendix

Appendix 1.1 Chromosome, genetic distance, locus marker counts and physical span for the genetic map aligned to the *Phaseolus lunatus* reference genome

Chromosome	Genetic Distance	Marker Count	Physical Position Start	Physical Position End
Pl01		119	591375	47588051
Pl01	0	2	591375	605537
Pl01	5.378	2	1030589	1048072
Pl01	5.921	4	1099547	1153794
Pl01	6.465	4	1245709	1339043
Pl01	7.008	5	1383508	1394764
Pl01	8.107	3	1459476	1466744
Pl01	8.651	5	1499017	1516185
Pl01	12.105	17	1618693	1857276
Pl01	14.353	1	2025476	2025476
Pl01	16.602	2	2636843	2724483
Pl01	17.146	6	2785811	2838430
Pl01	19.394	7	2885337	3250712
Pl01	19.938	1	3288666	3288666
Pl01	20.481	9	3311070	3503057
Pl01	23.935	3	3705560	3754592
Pl01	38.954	9	5631159	5860224
Pl01	41.798	4	6132586	6146858
Pl01	43.466	5	7720271	22100649
Pl01	62.751	5	42432553	42503850
Pl01	64.999	2	43003865	43097717
Pl01	82.061	1	44758370	44758370
Pl01	85.514	3	45502966	45804304
Pl01	88.968	1	46355610	46355610
Pl01	109.444	4	46956958	46982599
Pl01	111.111	4	47325827	47418065
Pl01	112.21	6	47443970	47476302
Pl01	113.309	4	47498014	47588051
Pl02		1554	720985	51607776
Pl02	0	10	720985	809538
Pl02	0.543	3	821503	837108
Pl02	1.087	3	841694	847135

Pl02	2.186	4	889766	907052
Pl02	3.285	9	939326	1038507
Pl02	4.952	3	1216196	1274680
Pl02	5.496	2	1286425	1309360
Pl02	6.595	7	1378911	1512408
Pl02	22.614	4	2143231	2171172
Pl02	23.158	25	2209219	2564422
Pl02	24.825	8	2591130	2730110
Pl02	27.074	14	2897287	3129039
Pl02	28.741	3	3139218	3256139
Pl02	29.84	7	3297185	3543599
Pl02	30.939	7	3593474	4107092
Pl02	32.607	3	4247039	4369737
Pl02	33.15	4	4430312	4530870
Pl02	33.694	10	4668391	4977650
Pl02	34.793	9	4983343	5115538
Pl02	35.892	16	5244351	6054781
Pl02	37.559	13	6165541	6473855
Pl02	38.103	11	6489179	7040395
Pl02	38.646	8	7125997	7383980
Pl02	39.745	11	7416933	7616023
Pl02	42.589	7	8332041	8419970
Pl02	43.133	11	8469537	8605983
Pl02	44.8	20	8705641	9162361
Pl02	45.343	18	9211618	9470859
Pl02	46.443	2	9543582	9586535
Pl02	48.691	3	9617233	9927448
Pl02	49.235	17	9984302	11430294
Pl02	49.778	36	11459032	11939165
Pl02	50.322	4	11973196	12011203
Pl02	50.865	724	12071342	31851474
Pl02	51.409	183	31861730	33832492
Pl02	51.952	60	33888184	36417138
Pl02	53.051	41	36470174	36940695
Pl02	56.505	1	37978932	37978932
Pl02	57.049	61	37991205	40189448
Pl02	57.592	3	40197555	40283230
Pl02	64.34	4	41185977	41420150
Pl02	76.921	7	43865633	43923800
Pl02	78.02	43	43936558	44653089
Pl02	79.119	45	44672594	45353924

Pl02	80.786	26	46174719	46649988
Pl02	81.33	2	46669148	46729787
Pl02	91.873	6	48344152	48507079
Pl02	92.972	3	48534335	48567957
Pl02	94.639	1	48672065	48672065
Pl02	96.306	5	48762203	48910388
Pl02	99.15	2	48945438	48954432
Pl02	99.694	5	49149228	49291119
Pl02	101.943	2	49542722	49545574
Pl02	102.486	1	50004537	50004537
Pl02	126.917	7	50862952	51012070
Pl02	128.585	1	51069601	51069601
Pl02	131.429	1	51291411	51291411
Pl02	134.273	2	51413724	51445311
Pl02	134.816	1	51456581	51456581
Pl02	136.483	5	51500548	51607776
Pl03		1102	2903	45042368
Pl03	0	7	2903	112355
Pl03	0.543	9	159771	330381
Pl03	1.087	3	388989	400959
Pl03	2.754	6	441793	473265
Pl03	3.298	14	480103	712046
Pl03	4.397	6	800782	858357
Pl03	7.241	9	921635	1081052
Pl03	12.619	5	1513421	1606821
Pl03	13.718	1	1670306	1670306
Pl03	14.261	4	1697454	1814336
Pl03	14.805	8	1877649	1977039
Pl03	16.472	11	2163566	2371096
Pl03	18.721	24	2447356	2859665
Pl03	19.264	36	2871039	3327795
Pl03	21.513	3	3363616	3409668
Pl03	22.612	1	3419879	3419879
Pl03	23.155	4	3458353	3527888
Pl03	24.823	24	3572205	3865041
Pl03	25.366	4	3902723	3909302
Pl03	30.086	27	5235977	5692787
Pl03	31.185	8	5884797	5965546
Pl03	33.434	316	6030026	20756437
Pl03	41.244	68	23136416	23885529
Pl03	43.493	17	23903414	24401012

Pl03	45.16	11	24578389	25517091
Pl03	46.827	4	25757774	25956532
Pl03	49.671	21	26028885	26577294
Pl03	53.125	7	26637359	26863651
Pl03	53.669	8	26903896	27310005
Pl03	55.336	38	27581379	28288116
Pl03	57.585	7	28307125	28426351
Pl03	59.252	30	28433082	29526699
Pl03	59.795	1	29556452	29556452
Pl03	65.173	26	31659714	32757776
Pl03	68.627	16	33145735	33410437
Pl03	70.876	8	33441335	33612624
Pl03	72.543	12	33748992	33907370
Pl03	73.642	22	33985037	34818257
Pl03	74.185	7	34866328	35091929
Pl03	78.264	12	35097194	35708929
Pl03	79.363	16	35803279	36110559
Pl03	80.462	36	36183908	37004816
Pl03	82.13	10	37081485	37484436
Pl03	83.229	2	37902612	37902650
Pl03	83.772	28	38007712	39250747
Pl03	86.616	13	39322988	39498663
Pl03	88.865	26	39536478	40820378
Pl03	90.532	15	40924519	41233880
Pl03	91.076	12	41268052	41587398
Pl03	91.619	10	41616165	41773707
Pl03	92.718	1	41789115	41789115
Pl03	93.262	7	41803332	41889653
Pl03	94.361	13	41928025	42304227
Pl03	97.205	13	42379794	42618222
Pl03	98.304	4	42637725	42709914
Pl03	99.971	2	42874565	42945552
Pl03	100.515	4	43092205	43165267
Pl03	104.594	5	43315462	43488243
Pl03	105.693	2	43673784	43673819
Pl03	109.771	12	44057075	44186659
Pl03	110.315	4	44240096	44240468
Pl03	111.414	2	44251983	44344435
Pl03	111.957	7	44374617	44485837
Pl03	113.625	1	44507582	44507582
Pl03	114.168	5	44633526	44767009

Pl03	114.712	4	44827920	44894623
Pl03	115.255	2	44987843	45019174
Pl03	116.923	1	45042368	45042368
Pl04		887	66315	49299606
Pl04	0	4	66315	84816
Pl04	0.543	13	234042	368408
Pl04	1.643	3	388898	429560
Pl04	2.186	1	434771	434771
Pl04	2.73	11	439812	554109
Pl04	3.829	2	561109	568866
Pl04	4.928	7	584379	913635
Pl04	10.981	14	1252563	1483232
Pl04	12.08	4	1527136	1608088
Pl04	13.179	19	1623314	2092952
Pl04	14.847	2	2131139	2166897
Pl04	16.514	1	2205382	2205382
Pl04	18.763	21	2223098	2594817
Pl04	19.306	2	2601146	2613982
Pl04	22.15	5	2640722	2748475
Pl04	23.249	4	2757687	2874077
Pl04	24.348	5	2884774	2905603
Pl04	24.892	2	2912455	2912461
Pl04	27.14	35	2995411	4371478
Pl04	28.808	9	4489079	4545061
Pl04	29.351	4	4550552	4682070
Pl04	31.019	8	4731770	4809147
Pl04	31.562	13	4832050	5176716
Pl04	35.016	1	5293869	5293869
Pl04	35.559	18	6157149	6340327
Pl04	36.658	10	6388195	6482068
Pl04	37.757	5	6531171	6702707
Pl04	38.857	16	6748726	7054456
Pl04	39.956	6	7065203	7244146
Pl04	40.499	5	7307377	7519657
Pl04	42.166	11	7926675	8425396
Pl04	44.415	214	9379033	36421127
Pl04	44.959	341	36502763	44546389
Pl04	45.502	45	44546459	45141010
Pl04	46.046	2	45158231	45159869
Pl04	46.589	2	45487451	45550270
Pl04	48.256	2	45635918	45659183

Pl04	55.004	1	46476704	46476704
Pl04	67.243	1	47169721	47169721
Pl04	68.91	5	47666611	47713928
Pl04	77.108	2	48331280	48331313
Pl04	81.187	1	48779836	48779836
Pl04	82.286	2	48866482	48866494
Pl04	89.748	2	49240260	49260511
Pl04	91.416	6	49268271	49299606
Pl05		1263	24078	38089793
Pl05	0	22	24078	297553
Pl05	1.099	8	323084	417359
Pl05	1.643	4	440196	457627
Pl05	3.891	2	518656	545310
Pl05	5.559	6	588060	652252
Pl05	6.658	8	685782	941124
Pl05	8.906	12	982420	1187421
Pl05	10.005	7	1189003	1203708
Pl05	12.254	15	1267769	1324678
Pl05	13.921	11	1358119	1777533
Pl05	17.375	1	1819980	1819980
Pl05	20.829	11	2237200	2365063
Pl05	21.372	1	2381337	2381337
Pl05	21.916	4	2404694	2475240
Pl05	23.015	15	2523848	2779196
Pl05	23.559	17	2827337	2967793
Pl05	25.226	24	3029288	3665917
Pl05	26.893	3	3697464	3750102
Pl05	27.437	51	3831991	4589100
Pl05	30.281	33	5296153	5728509
Pl05	30.824	37	5757900	6887969
Pl05	31.923	47	6910649	7634971
Pl05	33.022	56	7701926	8284785
Pl05	34.121	616	8325385	28784132
Pl05	35.22	82	28821738	30802152
Pl05	36.319	36	30916189	31929805
Pl05	38.568	8	31957095	32353063
Pl05	39.112	18	32539894	32875776
Pl05	40.211	12	32881305	33163226
Pl05	47.673	6	34503358	34724625
Pl05	51.127	19	34792683	35085360
Pl05	52.226	3	35150223	35194149

Pl05	56.305	2	35322680	35330207
Pl05	56.848	2	35359560	35364729
Pl05	57.392	9	35520814	35804750
Pl05	61.47	6	36084177	36166560
Pl05	62.57	3	36342660	36568732
Pl05	68.623	7	37021135	37049474
Pl05	69.722	3	37102694	37110861
Pl05	70.266	5	37125475	37231046
Pl05	71.933	4	37369052	37392535
Pl05	73.032	24	37417254	37953886
Pl05	74.699	3	38073638	38089793
Pl06		661	2763	36649143
Pl06	0	48	2763	1372318
Pl06	1.099	81	1554796	19760913
Pl06	3.943	26	20152202	20849862
Pl06	5.042	12	21265843	21534125
Pl06	6.709	34	21617983	22830791
Pl06	10.163	5	23552648	23687211
Pl06	10.707	32	23723716	24528489
Pl06	13.551	3	24721896	24766250
Pl06	14.094	11	24767915	25386527
Pl06	15.193	12	25403707	25694708
Pl06	16.861	20	25795680	26399299
Pl06	19.109	9	26454444	26861866
Pl06	22.563	13	26909383	27044986
Pl06	26.017	14	27077019	27397926
Pl06	27.116	39	27418082	28556499
Pl06	27.659	3	28627042	28675544
Pl06	28.203	3	28881524	28907123
Pl06	29.302	6	28942859	29020574
Pl06	30.969	15	29042838	30084846
Pl06	32.636	18	30271024	30739963
Pl06	33.18	16	30754170	31013220
Pl06	33.723	3	31167823	31243607
Pl06	34.823	13	31509191	31797353
Pl06	36.49	17	32014512	32273428
Pl06	37.033	1	32319653	32319653
Pl06	37.577	13	32327432	32527825
Pl06	38.12	10	32588448	32690669
Pl06	38.664	2	32693746	32740108
Pl06	39.763	20	32772144	33128740

Pl06	42.012	18	33189916	33490361
Pl06	42.555	2	33581374	33592074
Pl06	43.654	3	33649697	33702499
Pl06	44.198	6	33743884	33908089
Pl06	44.741	1	33933465	33933465
Pl06	45.285	10	33950146	34098280
Pl06	46.384	2	34144985	34166079
Pl06	46.927	5	34214044	34225645
Pl06	47.471	1	34256243	34256243
Pl06	49.138	10	34287127	34459460
Pl06	49.682	10	34545433	34701449
Pl06	50.781	18	34761967	34993972
Pl06	51.88	10	35053939	35254084
Pl06	54.724	27	35276778	35776368
Pl06	56.972	17	35789730	35972104
Pl06	59.221	3	36127559	36165163
Pl06	60.32	19	36188538	36649143
Pl07		1604	1778	47820557
Pl07	0	15	1778	204632
Pl07	2.249	7	234701	324112
Pl07	6.327	10	330576	462497
Pl07	7.995	9	680653	877028
Pl07	8.538	3	877132	921811
Pl07	10.787	2	948837	949122
Pl07	11.886	3	973473	1053191
Pl07	12.985	2	1083751	1134796
Pl07	16.439	5	1177523	1181924
Pl07	16.982	14	1197372	1351024
Pl07	20.436	11	1446085	1639346
Pl07	23.28	3	1711255	1723179
Pl07	24.947	6	1968175	2026594
Pl07	26.046	9	2146700	2317582
Pl07	27.714	6	2408177	2574373
Pl07	28.813	9	2580445	2704389
Pl07	29.356	3	2714737	2738903
Pl07	34.734	3	3234371	3387655
Pl07	36.401	16	3457023	4297034
Pl07	37.501	28	4427687	5124721
Pl07	38.6	17	5271744	5602174
Pl07	39.143	4	5723274	5724131
Pl07	40.81	17	5824195	6095746

Pl07	41.909	1	6663731	6663731
Pl07	45.363	32	7070981	7777005
Pl07	45.907	24	7852209	8349132
Pl07	47.006	87	8422355	9743058
Pl07	48.105	162	9757815	13347039
Pl07	49.204	18	13361809	13937488
Pl07	49.747	688	13992329	35623121
Pl07	53.201	38	36386831	37150814
Pl07	53.745	49	37214474	38424421
Pl07	55.993	12	39093498	39465018
Pl07	56.537	33	39475271	40096112
Pl07	57.08	12	40131914	40291338
Pl07	58.748	11	40365830	40686826
Pl07	59.291	3	40706584	40706627
Pl07	59.835	22	40765481	41150813
Pl07	60.934	20	41155435	41929260
Pl07	62.601	15	41954605	42234210
Pl07	64.268	30	42305725	43161814
Pl07	67.112	12	43418049	43618944
Pl07	68.199	22	43714467	44499527
Pl07	69.867	5	44627539	44686035
Pl07	73.945	22	44773369	45045300
Pl07	74.489	5	45062857	45106997
Pl07	75.588	2	45252647	45309847
Pl07	79.042	14	45618920	45864855
Pl07	80.141	10	45932871	46077584
Pl07	81.808	4	46133465	46199200
Pl07	85.262	7	46257596	46375875
Pl07	86.361	13	46426599	46548742
Pl07	88.028	2	46603511	46623081
Pl07	88.572	7	46633760	46714565
Pl07	96.769	13	47285256	47620457
Pl07	98.437	7	47648011	47820557
Pl08		749	35793	57107478
Pl08	0	5	35793	279738
Pl08	2.844	5	368734	488969
Pl08	3.943	2	610854	826431
Pl08	6.192	6	996275	1101669
Pl08	7.859	2	1152559	1160130
Pl08	14.607	8	1344892	1408433
Pl08	15.151	3	1426242	1432632

Pl08	16.818	1	1473202	1473202
Pl08	17.917	4	1514017	1553035
Pl08	19.016	4	1566815	1679261
Pl08	19.559	15	1814013	2048830
Pl08	20.103	2	2096423	2096500
Pl08	21.77	9	2175527	2408937
Pl08	24.019	10	2421738	2608588
Pl08	24.562	4	2715645	2791797
Pl08	25.662	4	2825736	2862297
Pl08	27.91	21	2948470	3303250
Pl08	29.009	1	3365774	3365774
Pl08	29.553	3	3409420	3490921
Pl08	31.802	6	3552422	3725150
Pl08	32.345	5	3801092	3993946
Pl08	33.444	2	4173407	4181548
Pl08	36.288	13	4411312	4731062
Pl08	37.387	1	4901859	4901859
Pl08	38.486	3	4970888	4991111
Pl08	39.03	2	4999491	5030734
Pl08	39.573	17	5130388	5329540
Pl08	40.672	13	5380347	5723700
Pl08	42.34	2	5757473	5759623
Pl08	44.588	20	5829658	6301578
Pl08	45.132	20	6311916	6525883
Pl08	46.799	30	6538964	7650199
Pl08	48.466	21	7682461	9014121
Pl08	50.715	12	9116155	9638847
Pl08	52.382	24	10168001	11372587
Pl08	52.926	3	11457047	11776029
Pl08	53.469	7	11859473	11968115
Pl08	57.548	42	14141262	45171872
Pl08	61.002	43	45222854	46949362
Pl08	64.456	26	47122287	47646650
Pl08	64.999	1	47739433	47739433
Pl08	65.543	17	47813410	48830995
Pl08	67.791	3	48870508	48870530
Pl08	68.335	16	49130100	49377244
Pl08	68.878	87	49414866	51185377
Pl08	70.546	30	51208759	52054042
Pl08	72.213	6	52148513	52301304
Pl08	72.756	6	52360317	52619814

Pl08	73.3	6	52701051	52872926
Pl08	74.399	3	52910706	52910725
Pl08	74.943	2	52923730	52923737
Pl08	75.486	3	53014731	53040970
Pl08	77.735	9	53050887	53260498
Pl08	78.834	23	53286482	53670023
Pl08	79.933	10	53788504	53891931
Pl08	81.032	9	54069991	54220708
Pl08	83.281	6	54270585	54348918
Pl08	85.529	14	54443288	54694386
Pl08	86.628	3	54773683	54782827
Pl08	87.172	2	54835773	54835787
Pl08	87.715	2	54881028	54902743
Pl08	88.259	2	54962308	54986196
Pl08	89.358	24	55006786	55559225
Pl08	91.607	1	55615959	55615959
Pl08	92.706	1	55675953	55675953
Pl08	94.373	2	55709700	55731298
Pl08	96.622	7	55803909	55856483
Pl08	97.165	3	55908530	55956877
Pl08	98.264	12	56022530	56200501
Pl08	99.932	2	56277191	56336107
Pl08	101.599	8	56365425	56435090
Pl08	104.443	2	56602233	56679957
Pl08	109.821	6	56983827	57107478
Pl09		310	2050	40853806
Pl09	0	6	2050	136199
Pl09	4.079	3	328428	374967
Pl09	8.158	3	580172	580179
Pl09	9.257	9	747553	807744
Pl09	10.356	1	837160	837160
Pl09	12.023	7	851900	867841
Pl09	13.122	5	891454	925234
Pl09	14.221	3	953027	1024698
Pl09	14.765	1	1062397	1062397
Pl09	16.432	4	1112013	1134976
Pl09	17.531	22	1157961	1462558
Pl09	18.63	5	1539916	1629505
Pl09	20.879	10	1734163	1912075
Pl09	25.599	2	2065036	2065037
Pl09	26.142	2	2067139	2085168

Pl09	27.241	4	2092091	2176564
Pl09	27.785	3	2231720	2237625
Pl09	28.328	1	2256780	2256780
Pl09	29.996	12	2278047	2426641
Pl09	32.244	1	2452353	2452353
Pl09	32.788	1	2469179	2469179
Pl09	33.331	1	2541950	2541950
Pl09	33.875	8	2549520	2916546
Pl09	35.542	5	2998543	3092891
Pl09	36.641	11	3119655	3584966
Pl09	42.695	18	4276133	5682901
Pl09	43.794	1	5746040	5746040
Pl09	45.461	3	5906816	5974844
Pl09	46.56	6	5985731	6075105
Pl09	47.104	5	6153743	6176886
Pl09	48.203	23	6185896	6557140
Pl09	50.451	23	6615632	7143991
Pl09	51.55	8	7181290	7368190
Pl09	53.799	4	8108289	8146469
Pl09	54.898	44	8179570	10430222
Pl09	58.352	4	11803021	12928597
Pl09	65.814	4	27833533	29512225
Pl09	93.28	2	38045862	38045871
Pl09	93.823	13	38059036	38156515
Pl09	94.922	7	38169067	38265070
Pl09	127.765	8	40204081	40329170
Pl09	133.143	6	40682706	40833707
Pl09	133.687	1	40853806	40853806
Pl10		453	53329	53972946
Pl10	0	11	53329	85254
Pl10	0.543	17	161013	368146
Pl10	2.792	6	405391	587631
Pl10	3.891	23	643955	915203
Pl10	7.345	2	985002	1009385
Pl10	8.444	6	1050331	1128069
Pl10	9.543	1	1152477	1152477
Pl10	13.622	6	1196306	1288559
Pl10	14.165	1	1304858	1304858
Pl10	15.265	22	1336263	1556319
Pl10	16.364	9	1576418	1767346
Pl10	17.463	2	1884771	1895961

Pl10	18.562	9	1967625	2120400
Pl10	20.229	23	2136134	4231211
Pl10	23.073	66	4459113	5346164
Pl10	23.617	11	5358800	5740249
Pl10	24.16	97	5766389	8753908
Pl10	26.409	46	9492837	39674709
Pl10	26.952	42	41115057	46286504
Pl10	27.496	5	46521900	47316056
Pl10	28.595	5	47629455	49100171
Pl10	37.55	10	50371482	50715240
Pl10	39.799	2	50930557	51344605
Pl10	46.547	6	52022988	52183419
Pl10	57.923	4	53576010	53584706
Pl10	62.002	21	53681163	53972946
Pl11		1795	30471	48067523
Pl11	0	20	30471	225257
Pl11	1.099	15	231674	513915
Pl11	1.643	6	516814	556730
Pl11	5.096	2	592315	603711
Pl11	6.195	14	643162	806126
Pl11	7.863	3	999663	1002160
Pl11	10.111	6	1043011	1100133
Pl11	19.067	1	1568555	1568555
Pl11	21.911	2	1727637	1740350
Pl11	30.108	1	2554743	2554743
Pl11	31.776	3	2967279	2975445
Pl11	32.319	2	2995203	3004815
Pl11	36.398	2	3522530	3670878
Pl11	38.647	4	3827374	3916614
Pl11	40.314	2	4174666	4221388
Pl11	44.393	19	4943357	5226374
Pl11	44.936	10	5253443	5390017
Pl11	45.48	82	5414119	6511771
Pl11	46.023	157	6583824	8755414
Pl11	46.567	955	8784053	37898565
Pl11	47.11	134	37907747	39754991
Pl11	47.654	155	39780356	42757508
Pl11	48.197	29	42917167	43245287
Pl11	48.741	31	43265340	43639744
Pl11	50.408	16	43654091	44523326
Pl11	53.862	14	44851602	45121142

Pl11	54.961	14	45178966	45308919
Pl11	57.805	7	45327313	45714049
Pl11	58.348	1	46030489	46030489
Pl11	58.892	36	46047473	46835757
Pl11	59.991	3	46875298	46914006
Pl11	62.835	16	46936504	47364560
Pl11	63.378	1	47393154	47393154
Pl11	64.477	5	47435085	47444708
Pl11	66.726	27	47448882	48067523

Appendix 3.1 Accessions included in the status diversity panel

cultivar	escape	landrace	wild
Calico Cat (G26451)	G26752	G25144	G25227
Dixie Speckled		G25172B	G25586
Henderson Bush		G25191	G25737
Hopi 12 (G25623)		G25236	G25762
Jackson Wonder		G25254	G25785
RIL 39		G25268A	G25816
RIL 72		G25269A	G26309
UC 92		G25270	G26358
UC Beija Flor		G25303	G26459
UC Cariblanco N		G25309	G26517
UC Haskell		G25411	G26615
UC Lee		G25420	G26629
		G25540	G26632
		G25653A	G26635
		G25653B	G26653
		G25653C	G26686
		G25674/L136	G26741
		G25675A	G26742
		G25677	G27298
		G25678	JMC1099
		G25685/L148	JMC1101
		G25705	JMC1148
		G25805	JMC224
		G25837	JMC609
		G25987	JMC715
		G26107	
		G26159A	
		G26161	
		G26173	
		G26184	
		G26215	
		G26222	
		G26354	
		G26438	
		G27353	
		G27399	
		G27433	
		G27455	
		G27456	
		G27457	
		G27458	
		G27459	
		G27461	
		G27462	
		G27540	
		G27574	
		PI347775	