TOWARDS THE EXPLOITATION OF THE AMERICAN HAZELNUT (*CORYLUS AMERICANA*)

BY

RONALD SCOTT REVORD

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2019

Urbana, Illinois

Doctoral Committee:

Associate Professor Sarah Lovell, Chair
Associate Professor Thomas Molnar
Professor Youfu Zhao
Assistant Professor Santiago Mideros

ABSTRACT

In the Midwest U.S. landscape dominated by the corn-soybean rotation, agroforestry systems can be particularly valuable for increasing the provisioning and regulatory capacity of the agricultural landscape. However, these systems have not yet been broadly integrated into the landscape of this region since they are mostly relegated to marginal lands. A growing body of literature suggests a path to increase the adoption of agroforestry in the Midwest U.S. lies in the incorporation of low-input food-producing tree species that provide economic incentives for farmers. While existing varieties and breeding selections of tree fruits and nuts provide the opportunity for initial system development and integration, their broad adaptability to the Midwest U.S. requires genetic improvement with respective to target environments. This dissertation begins by summarizing literature on the genetic improvement of underutilized temperate U.S. tree crops and their wild relatives, with emphasis on their strategic integration into the Midwest U.S. agricultural landscape.

Subsequently, hazelnut is the focus of the three experimental chapters, with the theme of characterizing american hazelnut (*Corylus american*) germplasm for eastern filbert blight (EFB) resistance. EFB, caused by the fungus *Anisogramma anomala*, is a primary limitation to european hazelnut (*Corylus avellana*) cultivation in eastern North America. *C. americana* is the endemic host of *A. anomala* and, despite its tiny, thick-shelled nuts, is a potentially valuable source of EFB resistance and climatic adaptation. Interspecific hybrids (*C. americana* × *C. avellana*) have been explored for nearly a century as a means to combine EFB resistance with wider adaptability and larger nuts. While significant progress was made in the past, the genetic diversity of the starting material was limited, and additional improvements are needed for expansion of hazelnut production outside of Oregon, where 99% of the U.S. crop is currently produced. Towards this end, this Ph.D. research sought to expand the availability of characterized *C. americana* germplasm through: i) evaluating american and interspecific hybrid hazelnut (*C. americana* × *C. avellana*) germplasm for EFB resistance, ii) evaluating wild american hazelnut germplasm for genetic diversity and structure, and iii) mapping EFB resistance quantitative trait loci (QTL) in *C. americana* OSU 403.040.

In the first study, to improve our understanding of *C. americana* as a donor of EFB resistance, 29 diverse EFB-resistant *C. americana* accessions were crossed with EFB-susceptible *C. avellana* selections (31 total progenies) to produce 2031 F_1 plants. Additionally, new *C. americana* germplasm was procured from across the native range of the species – 1335 plants from 122 seed lots representing 72 counties and 22 states. The interspecific hybrid progenies and a subset of the american collection (616 trees from 62 seed lots) were field planted and evaluated for EFB response following inoculations and natural disease spread over seven growing seasons. EFB was rated on a scale of 0 (no EFB) to 5 (all stems containing cankers). Results showed that progeny means of the interspecific hybrids ranged from 0.96 to 4.72. Fourteen of the 31 progenies were comprised of at least one-third EFB-free or highly tolerant offspring (i.e., ratings 0 to 2), transmitting a significant level of resistance/tolerance. Several corresponding *C. americana* accessions that imparted a greater degree of resistance to their hybrid offsprings were also identified. In addition, results showed that 587 of the 616 (95.3%) *C. americana* plants evaluated remained completely free of EFB, confirming reports that the species rarely expresses signs or symptoms of the disease and should be further studied and used in breeding.

In the second study, the genetic diversity and structure of new *C. americana* (272 individuals) collected from 33 seedlots across the species' native range are reported. Two-thousand fifty-three SNPs were discovered using a genome-by-sequencing approach and support a heterozygous collection (H_E = 0.276, H_O = 0.280) with moderate differentiation (F_{ST} = 0.108) and low inbreeding (F_{IS} = -0.136). Bayesian model-based and neighbor-joining (NJ) clustering corroborate an uppermost clustering level of K = 3. The NJ dendrogram depicts many small subgroups equally distant from common ancestry. Discriminant analysis of principal components reveals between-sub-group variation (K = 15) within the NJ dendrogram and allows the identification of 19 consensus subgroups. Fifty-one accessions were selected for inclusion within a core set based upon 95% representation of the observed allelic variation. Breeders can now exploit the breadth of genetic diversity held within this collection during development of interspecific hybrids.

In the final study, a genetic linkage map was developed using a genome-by-sequencing approach and used to identify QTL associated with EFB resistance from the *C. americana* selection OSU 403.040 from Nebraska U.S. A bi-parental mapping population comprised of 121 seedling trees was evaluated for EFB under high disease pressure in New Jersey, where *A. anomala* is endemic and highly genetically diverse. With EFB response represented by the percent of diseased wood, a total of three QTLs were discovered on linkage groups (LG) 3, 6, and 11 that respectively represent 62.6%, 23.3%, and 11.1% of the phenotypic variation. EFB resistance from OSU 403.040 appears new based upon it being only the second mapped source from *C. americana* and due to it mapping to three loci – all other sources of EFB resistance in *Corylus* spp. are monogenic and map to a single locus. Additionally, OSU 403.040 likely exhibits resistance to a broad range of *A. anomala* given the genetically diverse *A. anomala* environment under which it was selected. Such durability is requisite for the development of a feasible commercial variety for the eastern U.S. and highlights a priority for its inclusion in gene pyramiding schemes with resistant *C. avellana*.

ACKNOWLEDGEMENTS

I would like to express deep appreciation to my advisor Dr. Sarah Lovell, who beyond supporting my Ph.D. studies was a dedicated and thoughtful mentor. I would also like to thank my doctoral committee, Dr. Thomas Molnar, Dr. Frank Zhao, and Dr. Santiago Mideros, for their guidance and feedback with experimental design as well as very helpful discussions related to the fundamental biology of the *Corylus*-*Anisogramma* pathosystem. Additionally, a special and sincere thank you to Dr. Molnar and members of the Rutgers University hazelnut breeding program who have been exceptional collaborators during my studies and enabled this body of work through generous access to their germplasm. I am also grateful to Dr. Patrick Brown, Shilpa Manjunatha, and Dr. Crystal Sorgini as well as Rutgers collaborators Dr. Josh Honig and Jennifer Vaiciunas for their generous guidance and collaboration.

This dissertation is dedicated to my loving wife Sara and my family, who have all been incredibly supportive and engaged during the course of my graduate studies. I am most grateful for your continued love and support.

TABLE OF CONTENTS

CHAPTER 1

GERMPLASM DEVELOPMENT OF UNDERUTILIZED TEMPERATE U.S. TREE CROPS

1.1 ABSTRACT

In the Midwest U.S. dominated corn-soybean landscape, agroforestry systems can be particularly valuable for increasing the provisioning and regulatory capacity of the agricultural landscape. However, these systems have not yet been broadly integrated into the landscape of this region since they are mostly relegated to marginal lands. A growing body of literature suggests a path to increase the adoption of agroforestry in the Midwest U.S. lies in the incorporation of low-input food-producing tree species that provide economic incentives for farmers. Studies of the system-level integration of such approaches have proceeded by using the currently available cultivars and breeding selections of various tree nut and fruit species. While existing varieties and breeding selections provide the opportunity for initial system development and integration, their broad adaptability to the Midwest U.S. and its marginal land-types is unexplored. Thus, a second tier of research includes the genetic improvement and adaptation of tree crop selections to their respective target environments throughout the Midwest U.S. Fortunately, select tree crops of interest are amendable to systematic breeding and have wild relatives that are endemic across the region. In this chapter, we discussed the value of these wild relatives for broadening the adaption of cultivated tree crop selections by using the hazelnut as an example species. We presented a framework using geospatial tools to define and prioritize target environments for breeding and, in turn, exploiting wild relative germplasm.

1.2 INTRODUCTION

The benefits of agroforestry's regulatory services are well characterized for the Midwest U.S., where annual cropping systems such as the corn-soybean rotation dominate the landscape (JOSE, 2009; TORRALBA et al., 2016; TSONKOVA et al., 2012). Under the appropriate conditions, select agroforestry systems have even demonstrated superior production capacity and profitability compared to their respective

annual grain counterparts (BRANDES et al., 2016, 2017; WOLZ AND DELUCIA, 2018b), which brings pragmatism to the strategic diversification within Midwest U.S. Nevertheless, adoption of agroforestry continues to be relatively limited throughout this region. A growing body of research suggests a path to increase the adoption of U.S. agroforestry systems lies in the integration of low-input fruit and nut producing tree species (henceforth, referred to as tree crops) (JOSE, 2009; LOVELL et al., 2017; MATTIA, 2016, 2017; MOLNAR et al., 2013; MORI et al., 2017; RHODES et al., 2016; WOLZ AND DELUCIA, 2018a; WOLZ et al., 2017). Such systems, described by Lovell et al. (2017), provide a unique opportunity to integrate new food production capacity into the Corn Belt while simultaneously providing regulatory services (LOVELL et al., 2017; WOLZ et al., 2017). A primary constraint in considering these systems more broadly is the limited availability of tree crop germplasm and the extent to which cultivated selections can be improved and adapted to target environments across the Midwest U.S.

Numerous tree crop species are available for integration into Midwest U.S. agroforestry systems at some scale and range (MOLNAR et al., 2013). In recent years, cultivated selections of tree crop species have been tested or released for particular regions of the Midwest U.S. (Table 1.1), which has subsequently led to new regional crop markets (reviewed in part by MORI et al., 2017). These germplasm improvement efforts provide valuable genetic resources that are foundational for tree crop development, but a systematic framework is needed to effectively exploit these genetic resources within environments across the Midwest U.S. Tree crops (such as those in Table 1.1), which are often underutilized species and have varying degrees of assembled genetic resources. Additionally, the development of these tree crops will, in many cases, be restricted to the marginal lands of row-crop regions, which have not been studied for suitability with respect to productivity of specific tree crops. Fortunately, many tree crops are good targets for schematic breeding, and they have wild relatives that are endemic to the range of the Midwest U.S. To be relevant for broad adoption within the region, a framework for tree crop improvement must include systematic steps to target selection environments within the marginal lands of row-crops and, in doing so, identify the associated adaptive traits required of tree crop wild relatives. Forming and implementing such a framework persists as a critical gap in tree crop development.

In this review, we synthesize literature to support a case for the broad integration of select tree crops into the Midwest agricultural landscape. We first contextualize the roles of tree crops within the landscape and then present focal considerations for species development. Thereafter, we make a case for the value of tree crop wild relatives and conceptually define target environments for broad integration into the landscape especially lands that are marginal to row-crops. Lastly, we present a framework using geospatial tools to systematically target environments of the Midwest U.S. for tree crop adaptation using corresponding wild relatives.

1.3 FINDING A PLACE FOR TREE CROPS IN THE MIDWEST U.S.

The suitability of tree crops to the Midwest U.S. relates to the species' capacity to yield both provisioning and regulatory services when cultivated within the marginal lands of the maize-soybean rotation. Moreover, an element of this capacity is the species' propensity for genetic improvement within these respective marginal lands. Henceforth, we refer to the marginal lands of the maize-soybean rotation in the Midwest U.S. just as marginal lands, despite the fact that these areas can still be productive and might not be considered marginal in comparison to lands in other regions.

In the Midwest U.S., maize and soybean yields have increased linearly over the past several decades (FISCHER et al., 2014; LOBELL AND AZZARI, 2017). Recent satellite mapping suggests that maize production gains are disproportionally occurring on well-suited parcels and within-field locations (LOBELL AND AZZARI, 2017), as opposed to low-yielding lands or marginal lands. This trend is in part due to the responsiveness of well suited land-types to the maize system's prominent scientific improvements (i.e., high-density planting, variable rate technologies, and selection of maize hybrids within high-yielding environments) (LOBELL AND AZZARI, 2017). In fact, maize yield improvements have been accompanied by a higher sensitivity to drought stress, which may be induced more rapidly in low-yielding, marginal environments (LOBELL et al., 2014). Nevertheless, the relative lack of maize yield improvement within low-yielding and marginal lands alone provides justification to consider tree crops that are better suited to these environments (BRANDES et al., 2016; LOBELL AND AZZARI, 2017).

Considering tree crops for cultivation on marginal lands requires a procedural approach to pair discrete locations with corresponding well-suited tree crops, so that their adoption and subsequent improvement can proceed accordingly. Low crop productivity and profitability are often used to define marginal land-types (BARBIER, 1989; KANG et al., 2013; WIEGMANN et al., 2008). Collectively, such marginal lands represent millions of underutilized hectares throughout the Midwest U.S. (NIU AND DUIKER, 2006), which possess varying characteristics that cause marginality. Select tree crops have been described as compliments to the complex row-crop landscape (LOVELL et al., 2017; MOLNAR et al., 2013; WOLZ et al., 2017) in that they are low-input (OLSEN, 2001, 2013; STANEK, 2019) and suited to attributes typical of certain marginal land types (e.g., highly erodible, sloped lands) (MOLNAR et al., 2013; SMITH, 2013). However, in practice, the extent to which particular tree crop selections are well suited to marginal lands remains to be determined. Filling this knowledge gap will guide development initiatives by clarifying the extent to which a given tree crop can be integrated into marginal lands, as well as the geographic distribution of those lands and the secondary considerations those locations might carry.

Geospatial mapping and analysis could offer insight regarding the extent to which marginal lands exhibit characteristics that match the suitability of respective tree crops, with a potential to estimate profitability. To date, bioenergy grasses have been the primary perennial crops modeled as alternatives for the Midwest U.S., and they demonstrate the viability of agricultural diversification with perennial species for the region (BRANDES et al., 2016, 2017). Spatial mapping and economic analysis have demonstrated improvement of farm-level profitability through targeted subfield cultivation of perennial grasses, switchgrass (*Panicum virgatum*), and miscanthus (*Miscanthus giganteus*), in place of low-yielding maize and soybean (BRANDES et al., 2016, 2017). A recent study by Wolz and DeLucia (2018b) utilized similar spatial mapping of timber-based black walnut (*Juglans nigra*) plantations and alley cropping systems to demonstrate increased profitability through targeted system placement. Across four states in the Midwest U.S., they showed that black walnut alley cropping systems could increase profitability for landowners on 23.4% of cultivated land, assuming a 5% discount rate in future yields (WOLZ AND DELUCIA, 2018b). In effect, the studies by Brandes et al. (2016, 2017) and Wolz and DeLucia (2018b) provide evidence for

considering the targeted integration of other low-input perennials, which extends to a growing number of tree crop species.

Select tree crops demonstrate potential as new or emerging commercial crops for the Midwest U.S. Research to advance the commercialization of these tree crops has progressed over the past decade, with emphasis on adapting cultivated selections to particular regions across the Eastern and Midwest U.S. (Table 1.1). Variety improvement has stimulated new regional tree crop industries (MORI et al., 2017) and, in some cases, assembled resources requisite to access domestic markets (MOLNAR et al., 2018a), e.g., diverse and characterized breeding material. Systematic exploitation of existing resources offers an opportunity to scale these crops and capitalize on regional and domestic market growth (MORI et al., 2017; TECHNAVIO, 2017). Foremost, this exploitation entails expanding the production range well-suited for cultivated tree crop selections to additional regions throughout the Midwest U.S., which is discussed at depth in subsequent sections.

The benefits of incorporating tree cultivation and conservation into agriculture (i.e., agroforestry) are diverse and well-documented (LEAKEY, 2014; MOSQUERA-LOSADA et al., 2011; RIGUEIRO-RODRÍGUEZ et al., 2009), even for the Midwest U.S. (JOSE, 2009, 2012; LOVELL AND JOHNSTON, 2009; SCHOENEBERGER et al., 2009, 2012, 2017; UDAWATTA AND JOSE, 2012). Strategic integration of trees into temperate croplands adds substantial capacity to mitigate climate change by reducing agricultural greenhouse gas emissions (AMADI et al., 2016; KIM et al., 2016; SCHOENEBERGER et al., 2012) and sequestering carbon (MOSQUERA-LOSADA et al., 2011; UDAWATTA AND JOSE, 2012). Udawatta and Jose (2012) provide an in-depth synthesis of the carbon sequestration capacity of temperate North American agroforestry systems and, conservatively, estimate the sequestration potential of agroforestry systems at 12.4 Mg C ha^{-1} $year^{-1}$ to a landscape-level capacity of 65.7 Tg C $year^{-1}$, excluding silvopastoral systems. Diversifying the tree species that are common to North American agroforestry systems—beyond timber species like walnut (*Juglans* sp.) and poplar (*Populus* sp.)—to include tree crops will help realize agroforestry's sequestration potential (WOLZ et al., 2017).

Temperate agroforestry systems also provide services to filter run-off (UDAWATTA et al., 2002), reduce erosion (GARRETT et al., 2009; MATTIA, 2017), and create habitat that fosters biodiversity (JOSE, 2009; TORRALBA et al., 2016; TSONKOVA et al., 2012). Deep tree roots and a longer growing season enable agroforestry systems to reduce nitrate leaching compared to crop systems and perennial pasture (ALLEN et al., 2004; BAMBO et al., 2009; UDAWATTA et al., 2002), with increasing reductions of leaching at greater soil depths (ALLEN et al., 2004). Consistent with the concept of integrating agroforestry on marginal lands, various studies have demonstrated that subfield row-crop areas that are the lowest yielding coincide with those of the highest environmental risks (MUTH, 2014), including soil erosion (MATTIA, 2017), water, and soil quality (BRANDES et al., 2016; LERCH et al., 2005). Disproportional reductions in soil erosion and nitrate leaching have been demonstrated by targeting these areas with perennial cropping systems (MATTIA, 2017).

1.4 CONSIDERATIONS FOR TREE CROPS DEVELOPMENT

Numerous tree crops are endemic or somewhat well-adapted to the U.S. (FORD-LLOYD et al., 2011; KHOURY et al., 2013). Prioritizing species with sufficient potential to warrant development is essential (MAYES et al., 2011). Measures to prioritize the development of underutilized species are seldom discussed with respect to the temperate U.S. with the exception of several related studies that have systematically categorized the relative importance of wild crop relatives for collection (CASTAÑEDA-ÁLVAREZ et al., 2016; KHOURY et al., 2013, 2017) and generally detailed the use of wild crop relatives in systematic breeding (DEMPEWOLF et al., 2017; PROHENS et al., 2017).

Opportunities exist to bring new crop species into a production region when a respective trait outperforms the region's major crop under select conditions. Referred to as a trait-based approach, the strategy is useful to gauge the potential of a particular underutilized species, in which the production or quality of a specific commercial trait is compared to the major crop when cultivated on a particular marginal land-type or under relevant abiotic stresses (e.g., heat and drought stress) (MAYES et al., 2011). When considering such opportunities, the economic value of the tree crop or its trait of interest must be apparent

and exploitable within existing markets (MAYES et al., 2011). Trait-based comparisons may be most applicable to tree crop species with high value markets (e.g., oil, protein, vitamin, or phytonutrient). One such example is elderberry (*Sambucus* sp.) and a growing market for healthy plant-based food dyes made from phytonutrients (CERNUSCA AND GOLD, 2013). High phytonutrient content is a trait intrinsic to the elderberry genus (LEE AND FINN, 2007; ÖZGEN et al., 2010). The concentration of total phytonutrients in elderberry are naturally higher than the relevant accessions of maize and various vegetable species.

Furthermore, elderberry is adapted to a wide range soils in the Midwest U.S. (CHARLEBOIS et al., 2010), including those that are sloped and moderately fertile. In these environments, elderberry's field-level phytonutrient production can be compared to that of maize and vegetable selections. If elderberry provides a production advantage within these environments, there is precedent for crop integration.

While intuitive, basic market trends provide valuable support for prioritization of tree crops for improvement. Market data may reveal shifts in global production or indicate whether markets can stably accommodate an increased supply. Drivers of market growth or gaps are also valuable considerations since they shed light on constraints for market access or even competitive opportunities for market newcomers. Box 1.1 uses the hazelnut to emphasize the value such information offers. Collectively, such material is valuable to improve stakeholder engagement and garner industry development.

The genetic resources of a given species and its wild relatives are fundamental parameters that determine the outlook of tree crop development. Tree crops with superior wild selections or suitable breeding selections that can serve as first generation varieties hold an invaluable advantage to help facilitate grower adoption and prime markets. The availability of tree crop wild relatives adapted to the targeted cultivation regions is critical. Sufficient genetic diversity and ability to discover adaptive traits is an essential parameter to species prioritization.

Box 1.1. An economic case to broaden hazelnut cultivation in the U.S.

Commercial cultivation of hazelnut has been an ongoing pursuit in the eastern and midwestern U.S., as the global hazelnut market is forecasted to grow substantially from $4.15 to $5.75 billion between 2017 and 2021 (TECHNAVIO, 2017). Although the U.S. hazelnut market is small compared to the european market, the americas are the fastest growing market (7.57% annual growth, $0.262 billion) (TECHNAVIO, 2017). Ferrero (Alba, Italy) – manufacturer of Nutella® and Ferrero Rocher® and purchaser of 25% of the annual global hazelnut supply –has helped drive new growth for hazelnut-based products in the U.S. confectionary markets, which has spurred new purchasing of hazelnut by american chocolate companies (TECHNAVIO, 2017).

Growth in the global hazelnut market has incentivized minor hazelnut producing countries to support development initiatives aimed at expanding cultivation (BALDWIN, 2015; ELLENA et al., 2012; GRAU AND BASTIAS, 2004; MELHENBACHER, 2004; MOLNAR AND CAPIK, 2012; XIE et al., 2012). This incentive has been furthered by price peaks caused by constraints to the Turkish hazelnut supply, which accounts for 60% to 70% of the global supply. Turkish cultivation relies on wild selections or landraces, and most orchards are over a half-century old and gradually declining in productivity (GÖNENC et al., 2006). Moreover, new high-yielding varieties are not available for a new orchard establishment because of the absence of variety improvement programs in the region (MEHLENBACHER, 2008). Further, the Turkish hazelnut supply has greatly fluctuated in recent years due to abiotic stresses (USTAOĞLU, 2012). In 2013 and 2014, severe winter frost and subsequent drought damaged the Turkish crop to cause 30% and 50% reductions to production, respectively (FAOSTAT, 2014). The shortages led to price peaks, motivating Ferrero to lessen its dependency on Turkish supply. Ferrero has heavily invested in hazelnut industry development in non-traditional nations (e.g., Canada and Australia).

These market trends highlight an opportunity for new growing regions to gain a share of the hazelnut market. The U.S. maintains mature and dedicated variety improvement programs, while development initiatives in other countries primarily focus on variety trials of traditional cultivars. These U.S. breeding programs are a competitive advantage for supporting the growth of U.S. hazelnut production. The endemic american hazelnut species also offers a competitive advantage in that it provides adaptability traits requisite for the adaptation of the commercial european hazelnut to the eastern and midwest U.S.

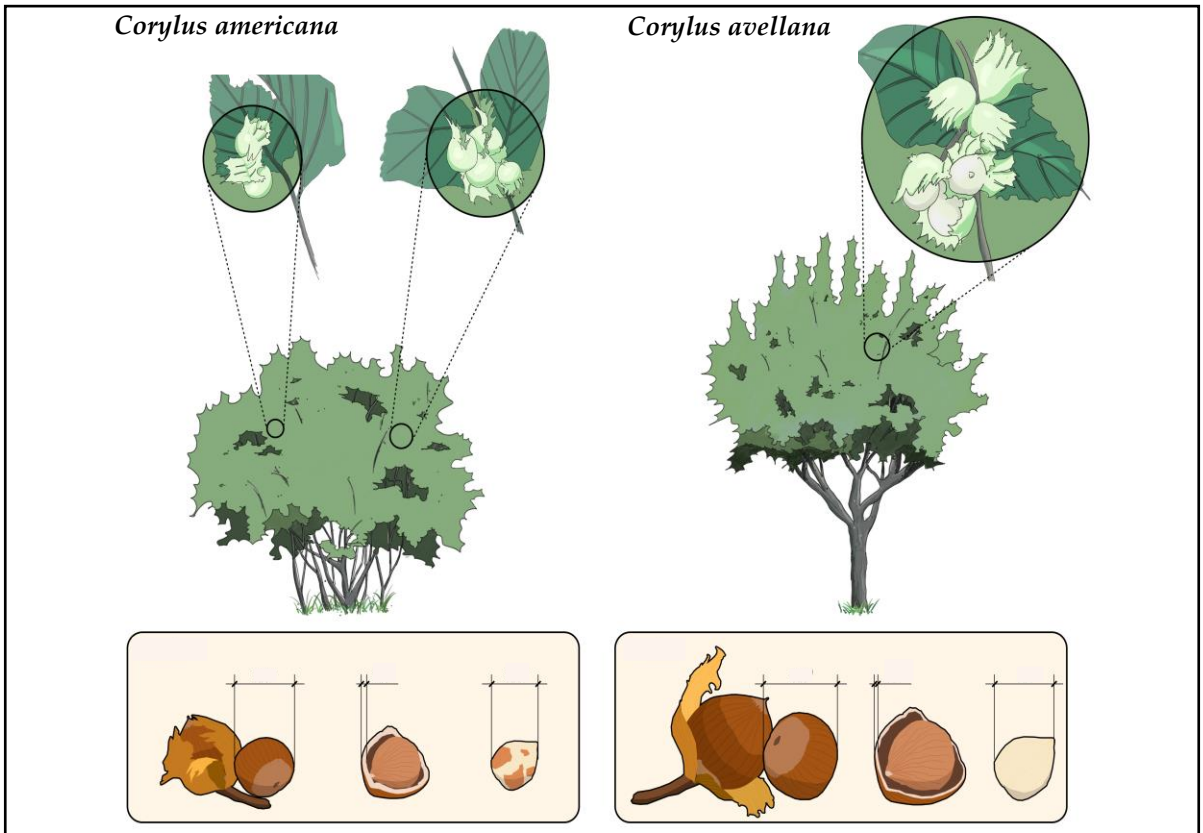
1.5 EXPANDING THE CULTIVATED RANGE OF TREE CROPS USING WILD RELATIVES

Wild crop relatives are increasingly discussed as a source of novel traits relevant to the agriculture's challenges (DEMPEWOLF et al., 2014, 2017; KOLE et al., 2015; PROHENS et al., 2017; WARSCHEFSKY et al., 2014). Such germplasm offers the ability to introduce traits from the far side of the domestication bottleneck and expand breeding pools with alleles that contribute to improved yields, adaptation to abiotic stresses, and an expanded cultivation range. While cultivated tree species have undergone less severe domestication bottlenecks than annual crops (MILLER AND GROSS, 2011), the systematic introduction of wild germplasm

still provides novel allelic diversity to breeding pools. The wild relatives of the tree crops discussed in this case carry a variety of traits of commercial value, including fruit quality, disease resistance, and abiotic stress tolerance (Table 1.2). For example, the american hazelnut, *Corylus americana*, is a wild relative of growing value since breeders seek to expand the cultivated range of the european hazelnut to non-traditional climates within the U.S. (MOLNAR et al., 2018b). The species is endemic to a vast and continuous area of the eastern U.S. and represents a diverse source of adaptability traits to the primary stressors that limit commercial cultivation—eastern filbert blight resistance and climatic stressors (discussed in Box 1.2). The introgression of such traits in perennial systems is increasingly more feasible and efficient due to genomic-assisted techniques that accelerate allele discovery and progeny selection (KOLE et al., 2015; MCCLURE et al., 2014; MIGICOVSKY AND MYLES, 2017)

A majority of tree crop species are amenable to clonal propagation (i.e., asexual and vegetative), and, as a consequence, many tree crop progenitors and even cultivated varieties are clonal propagates of superior wild selections (MILLER AND GROSS, 2011). The ability for "instant" domestication through clonal propagation means that current tree crop varieties are typically only a few generations removed from their wild relatives, and, collectively, these selections often maintain similar levels of neutral genetic variation compared to their wild populations (MILLER AND GROSS, 2011). Additionally, woody perennial varieties often remain genetically indistinct from their uncultivated wild counterparts (MIGICOVSKY AND MYLES, 2017). Comparative multivariate analysis of genome-wide DNA markers shows that cultivated individuals of apple (CORNILLE et al., 2012), grape (MIGICOVSKY et al., 2016; MYLES et al., 2011), and european hazelnut (ÖZTÜRK et al., 2017) appear undifferentiated from either their progenitors or wild relatives, even following selection. Such genetic similarities suggest ongoing gene flow between cultivated selections and wild germplasm following domestication (MYLES et al., 2011), which occurs very regularly (ELLSTRAND et al., 1999). Although not requisite, genetic similarities between these two categories of germplasm provide an optimistic outlook for the role of wild relatives in the introgression of adaptive traits from wild germplasm into breeding selections.

Box 1.2. Expanding the european hazelnut's cultivated range using the american hazelnut.



Commercial hazelnut cultivation is an ongoing pursuit in the eastern U.S. (MOLNAR et al., 2018a), and development initiatives center on the adaptation of european hazelnut varieties through the introgression of climatic adaptations and eastern filbert blight (EFB) resistance (MOLNAR, 2011; MOLNAR et al., 2018b). The american hazelnut species is native to a vast extent of eastern North America (DRUMKE, 1964), suggesting that the species holds a diverse source of EFB resistance and climatic adaptations (CAPIK AND MOLNAR, 2012; FULLER, 1908). In nature, the fungus of EFB co-exists with the american hazelnut, occasionally producing minor cankers (CAPIK AND MOLNAR, 2012; WESCHCKE, 1953), while the infection of the european hazelnut is lethal (JOHNSON AND PINKERTON, 2002). Additionally, recent study of the fungus depicts exceptionally high genetic diversity (MUEHLBAUER et al., 2018) and supports the need for integrating durable multi-genic resistance like that found within *C. americana* into breeding programs.

While *C. americana* is a diverse source of climatic adaptations and EFB resistance, its nut quality and plant architecture are not suited for commercial production and existing markets. Compared to cultivated *C. avellana*, the kernels of *C. americana* are much smaller, enclosed in thicker shells and involucres, which complicates their retrieval during harvest and processing, and retain a higher percent of their pellicle following blanching (depicted above). The species' shrub architecture and aggressive basal sprouts also deviate from standard single-stemmed hazelnut orchards. Promisingly, reciprocal crosses between the species are viable (ERDOGAN AND MEHLENBACHER, 2000), and breeders have observed quick recovery of the european hazelnut's commercial traits in one to two backcross generations (BHATTARAI et al., 2017).

While clonal propagation has been advantageous for prompt domestication and cultivation, it has negated the need to comprehensively introduce wild germplasm into breeding programs. Thus, to-date, wild relatives have been used only sparingly and often just for select traits (BRUMLOP et al., 2013; MAXTED et al., 2012; WARSCHEFSKY et al., 2014). Despite their historical underutilization, wild relatives of tree crops hold extensive phenotypic and genetic diversity critical to broadening the cultivated range of existing tree crop selections. This diversity arises as a consequence of tree reproductive biology (e.g., outcrossing, wind-dispersed pollen, long lifespan) that causes extensive gene flow (MILLER AND GROSS, 2011). In turn, tree species often hold high within-population diversity and a weak population structure (DUMINIL et al., 2007; DUMINIL et al., 2009; HAMRICK et al., 1992; LOVELESS AND HAMRICK, 1984; MILLER AND GROSS, 2011).

Even with minimal among-population structure and high gene flow, divergent selection between geographically distinct natural tree populations still gives rise to local adaptations (MILLER AND GROSS, 2011; PETIT, 2006). Local adaptation can arise as a response to a variety of geographically specific selection pressures, e.g., climate, photoperiod, soil characteristics, and pathogens (HEDRICK, 2006; KAWECKI AND EBERT, 2004; LINHART AND GRANT, 1996). While different traits respond variably to selection, many local adaptations of natural tree populations evolve through gradual allele frequency changes at small effects loci of polygenic traits (LE CORRE AND KREMER, 2012; PRITCHARD AND DI RIENZO, 2010).

The highly polygenetic architecture of adaptive traits provides insight into how local populations could diverge despite high gene flow. Divergent selection pressure occurs on the many small effects loci individually rather than the polygenic trait itself (AITKEN et al., 2008). Therefore, adaptive traits arise from a collection of allele frequency changes at the small effects loci. These allelic changes diverge weakly at individual loci and are not reflected in population structure (LE CORRE AND KREMER, 2012; SAVOLAINEN et al., 2013). Additionally, adaptive traits within local tree populations are often genetically diverse. Due to high levels of outcrossing, linkage disequilibrium (the pattern of non-random allele assortment) is quite low in natural tree populations (BROWN et al., 2004; HEUERTZ et al., 2006; NEALE AND SAVOLAINEN, 2004). Therefore, the recombination rate among small effects loci that comprise adaptive traits is high. Consequently, a diversity of recombinant offspring are produced annually in healthy tree populations with

high fecundity, giving rise to genotypes with different allelic combinations for a given adaptive trait (AITKEN et al., 2008).

The local adaptations, as well as other traits of interest (e.g., Table 1.2) from wild germplasm, can be leveraged through their systematic integration into breeding pools (WARSCHEFSKY et al., 2014). The desirable traits of wild relatives must be integrated into the genetic background of cultivated selections without the loss of traits of commercial importance. Backcrossing is a common approach for trait introgression, where selected progeny that inherited the wild trait of interest are crossed recurrently to advanced breeding selections. Numerous adaptive traits of both monogenic and polygenic architecture have been introgressed into elite varieties. Linkage drag, however, can produce undesirable progeny in backcross generations, if undesirable phenotypes from wild germplasm are linked to the introgressed loci (VARSHNEY et al., 2014). Increasing progeny size and additional backcross generations can facilitate the development of rare recombinants that break linkage drag.

Additionally, rapid linkage disequilibrium decay (i.e., the decrease of non-random segregation) is common in high diversity tree crops (LIJAVETZKY et al., 2007; MIGICOVSKY et al., 2016; NEALE AND SAVOLAINEN, 2004). This decay is the consequence of high levels of ancestral recombination occurring in natural outcrossing tree populations, which improves genetic mapping resolution and enables genome-wide association studies (GWAS) (PARCHMAN et al., 2012). GWAS can proceed using existing plant populations of unrelated individuals (i.e., natural populations or germplasm collections) and, thus, circumvents the time needed to develop and grow linkage-mapping populations. Consequently, GWAS creates the potential to quickly identify markers tightly linked with the causal loci and accelerate the identification of rare recombinants in backcrossing generations (SORK et al., 2013).

When considering the introgression of adaptive traits, recombinant progeny are most effectively identified when selection occurs within the target environment (e.g., marginal lands) (CECCARELLI, 2015; SIMMONDS, 1991). However, much of the historic selection has prioritized the most ideal, productive environments (CECCARELLI, 2015). Decentralizing selection to the targeted cultivation environment(s) ultimately makes the selection environment as similar to the cultivated environment as possible (ATLIN,

2001). As a result, decentralization improves the response to selection for specific adaptations within the respective environments of the target region (compared to selecting for wide adaption across the region) (ANNICCHIARICO et al., 2005). While decentralizing selection to improve the selection response is intuitive, the selection environment for public plant breeding programs has historically been restricted to centralized research stations, and on-farm testing is restricted to the trialing of only a few varieties (CECCARELLI, 2015). The decentralization of selection is particularly important in the breeding of tree crops for the Midwest U.S., where breeding goals center on expanding the cultivated range of respective species.

1.6 DEFINING TARGET ENVIRONMENTS FOR DECENTRALIZED SELECTION

While the decentralization of selection is conceptually straightforward, its implementation is complex. Target environment selection is based upon an assortment of characteristics that determine whether a given location is a suitable cultivation environment for the tree crop. These characteristics are highly heterogeneous among and within environments across the landscape, which leads to the recombination of environmental stressors and the creation of unique environments (LOBELL AND AZZARI, 2017). Despite high heterogeneity, breeders must choose the most opportune series of target environments that are representative of the targeted cultivation region. Subsequently, actual selection sites that represent each target environment must be identified for the placement of progeny tests. To-date, a systematic approach is lacking to accomplish the identification of these target environments.

In this case, we propose a brief logic model to identify target environments and corresponding locations for selection. (A) The approach begins by defining the spatial distribution of the soil characteristics required for the commercial production of cultivated selections within the cultivation region via geospatial suitability mapping. (B) Prioritized target environments can be selected from within the suitable areas using criteria to maximize the impact. (C) The specific adaptability traits required to expand tree crop cultivation to the defined targeted environments can be discretely specified and used to guide a breeding program. High-resolution geospatial techniques, such as this, clarify the spatial distribution of each target environment and create the potential to add additional variables (e.g., photoperiod and climate)

for recurrent assessment of variation within target environments to test for sub-environments (BRANDES et al., 2017; KIDD et al., 2015; WOLZ AND DELUCIA, 2018b).

Geospatial mapping of soil suitability characteristics is immensely informative for guiding decentralized selection. Foremost, the suitable cultivation areas are discretely identified from within the greater region, which significantly narrows the potential target environments (Figure 1.1). Additionally, the spatial distribution and predominant clustering of suitable locations immediately enable some prioritization (or de-prioritization) of the adaptability traits to source from wild relatives (e.g., photoperiod or climatic) as well as the general geographies from which wild germplasm should be sourced, if needed.

The geospatial mapping of soil characteristics requires a well-defined suitability index for the tree crop species. Suitability indices estimate productivity (i.e., growth or yield) of cultivated selections from traditional growing regions as a function of soil and climate characteristics. If sufficient data exists, a suitability index can be a quantitative, continuous index (WALLACE AND YOUNG, 2008).However, discrete suitability classes are currently more common for tree crops ((KIDD et al., 2015), Table A.1.). The suitability index can then be mapped using existing soil and/or climate geospatial databases. For the Midwest U.S., pre-existing geospatial data is available through the Soil Survey Geographic (SSURGO) Database (NRCS) for suitability mapping (SSURGO, 2017).

The specific characteristics that negatively affect suitability (and warrant adaptation) vary significantly among and within regions (LOBELL AND AZZARI, 2017). This variability has the potential to create many discrete target environments within respective local geographies. Consequently, a systematic framework to statistically differentiate these distinct clusters of target environments is needed. A multivariate approach is intriguing in that individual map units can retain their identity and the dataset complexity can be reduced to principal components. Subsequently, cluster analysis could classify individual map units into similar target environments based on the principal component variation representative of different limitations in soil characteristics.

Following the classification of target environments, there are several criteria that can contribute to the prioritization of target environments for breeding.

(1) Spatial extent: As clusters of prospective target environments emerge, their respective sizes and the amenability of their required adaptive traits to systematic improvement can guide the priority in which they are targeted. The target environments should be focused on specific land types and regions that are abundant.

(2) Productivity/profitability of row-crops: A high-resolution index (30 m × 30 m) of row-crop productivity in the Midwest U.S. is available via the National Commodity Crop Productivity Index (NCCPI) (DOBOS et al., 2012). This index can be used in a comparative profitability analysis to ascertain if any of the target environments overlap with low production row-crop environments. Productivity alone, however, is insufficient in determining whether an alternative crop can outcompete row-crops under specific conditions. Instead, high-resolution profitability surfaces are now used to evaluate row-crop suitability (BRANDES et al., 2016, 2017). In this light, it is important to note that the soil and climate suitability of tree crops and row-crops are not necessarily correlated (WOLZ AND DELUCIA, 2018b).

(3) Provision of regulatory ecosystem services: The prioritization of target environments could proceed based upon specific land types and regions that provide disproportionally large regulatory ecosystem services. Subsequent analysis of suitable map units could prioritize locations based upon slope, erosion, water quality, and more (BRANDES et al., 2017; MATTIA, 2017).

To proceed with breeding efforts, adequate wild germplasm with adaptability traits corresponding to the identified target environment must be acquired and characterized. Once discrete target environments are identified, they can be surveyed for the presence and availability of local wild relatives. If distinct clusters were produced from the cluster analysis, the geo-reference points nearest to the cluster centroids can serve as starting locations for germplasm collection. Cluster centroids can also guide the placement of progeny within target environments including to locations that best represent the collective map units of the cluster.

1.7 DISCUSSION AND CONCLUSIONS

Considerable opportunity exists to diversify the agriculture landscape in the Midwest U.S. and, thereby, increase both agricultural productivity and ecological functioning. Crop diversification is particularly pragmatic when viewed through the lens of economically marginal locations. Such marginal lands often coincide with high priority areas for ecological rehabilitation and disproportional potential returns with land-use change (BRANDES et al., 2016; RICHARDS et al., 2014). While the term "marginal" can imply a general lack of suitability for crop cultivation, in this context, the term is specific to the profitability of the maize-soybean rotation in the Midwest U.S. Consequently, areas considered marginal under this definition could be productive for other crop species especially those that are low-input. Additionally, these areas are compositionally diverse (LOBELL AND AZZARI, 2017) and, in turn, offer a variety of opportunities for the targeted development of alternate crops that impart ecological benefits, of which select tree crops are ideal candidates.

Tree crops introduce a variety of innate ecological functions (e.g., carbon sequestration and run-off filtration) while providing economic incentive to justify their long-term maintenance (LOVELL et al., 2017; RHODES et al., 2016). Recent breeding efforts (Table 1.1) have stimulated regionally-specific tree crop adoption (MORI et al., 2017), and, in the process, this work has built a foundational base of germplasm to enable future breeding endeavors. Moreover, wild relatives of these tree crops occupy broad endemic ranges that are inclusive of the Midwest U.S. and offer the genetic resources needed to expand the cultivated range of existing selections throughout the Midwest U.S. However, while existing germplasm collections are accessible, collection gaps are prevalent. Table 1.3 highlights the severity of these collection gaps by summarizing the number of existing accessions held at either the USDA-Agriculture Research Station or universities and the relative degrees to which these collections represent the endemic range of the species. While closing collection gaps is a lofty pursuit, we proposed a GIS-based workflow to give the collection structure as well as guide the subsequent use of regionally adapted germplasm in breeding programs via decentralized selection.

The broad integration of the tree crops into the marginal lands of the Midwest U.S. agricultural landscape would contribute substantially towards food security and environmental goals especially to climate change mitigation and adaption. The integration of trees into the temperate agricultural landscape drives considerable carbon sequestration in both woody biomass and soil (MOSQUERA-LOSADA et al., 2011; UDAWATTA AND JOSE, 2012) as well as the reduction of non-CO_2 greenhouse gases (AMADI et al., 2016; KIM et al., 2016; WOLZ et al., 2018). Tree biomass within the agricultural landscape is an important resource for carbon storage, comprising 75% of carbon stored (34.2 petagrams C) within the global agricultural landscape (ZOMER et al., 2016), even though tree cover only occupies more than 10% of space on roughly 40% of agricultural land. In only the temperate Midwest U.S., Udawatta and Jose (2011) estimated the sequestration capacity of conservative, low-density tree integration on just 10% of row-crop lands (or 15.4 million ha) at 52.4 Tg C $year^{-1}$. Despite the promise, the integration of trees into the agricultural landscape is far from optimized and remains an often overlooked medium for which to increase the landscape's carbon sequestration capacity (ZOMER et al., 2016).

Beyond mitigation functions, tree integration provides a portfolio of climate change adaptation functions to the agricultural landscape. Trees adjacent to row-crops can abate the effect of weather and climate on the crops by blocking wind stress (BÖHM et al., 2014), moderating air and soil temperatures through shade and evaporative cooling (LIN, 2007), and reducing evaporation of soil moisture (SIRIRI et al., 2013). Furthermore, tree-based systems have apparent advantages under increasing interannual variability in rainfall and heat – two focal challenges for future maize yield stability in the Midwest U.S. (LOBELL et al., 2013, 2014). First, deep-rooted systems have access to larger areas of water and nutrients, which is beneficial under drought conditions. Similar benefits are transferred to adjacent row crops through increased soil porosity from the tree roots that, in turn, improves water infiltration and storage leading into drought periods (ANDERSON et al., 2009). Furthermore, trees have higher evapotranspiration rates than row-crops, increasing the capacity for evaporative cooling under extreme heat but also the aeration of soils following periods of excessive rainfall and flooding (VERCHOT et al., 2007).

Equally important, diversification via tree crops also introduces a level of economic resiliency against interannual climatic variability. Tree fruits and nuts are often of higher value than commodity grains, and, even when integrated at comparatively small scales at the farm-level, tree crop revenues can buffer against the increasing risks of grain crop losses due to climate change. Additionally, while the consumer preference for horticultural crops can be somewhat plastic, the decentralized approach offers an opportunity to amend the breeding pipeline accordingly. Decentralization is used often as a strategy to integrate selective end-user and grower feedback related to marketability into the breeding pipeline. If systematically integrated, decentralization can be an effective tactic to help breeding objectives evolve with consumer and grower preferences.

The vision to broadly develop tree crop germplasm complimentary to the Midwest U.S. agricultural landscape is tangible and timely. New multi-disciplinary and multi-institutional collaborations are needed to coordinate existing resources and leverage the knowledge-base of regional breeders and tree crop stakeholders (e.g., farmers, buyers, processors). Foremost, the collaborators should prioritize suitability mapping (focusing on soil characteristics) of respective tree crops across the geographic range of the Midwest U.S. Prioritizing suitability mapping will create the foundational structure and direction for new collaborations. For example, the mapping will identify priority regions better suited for respective species, and, thus, where to collect germplasm and to seek out farmer participants to house progenies. With this direction, collaborations can justify comprehensive narratives for seeking funding. The vision's multifunctional outputs and broad scope has the potential to engage a diversity of funding sources, including foundational and private stakeholders and eventually aspire to committed state-level support. While there is ample opportunity to pursue competitive state and federal grants to support components of the program(s), emphasis should be placed on seeking private industry or foundational funding to support the largely applied breeding component that is core for this endeavor.

1.8 TABLES AND FIGURE

Table 1.1. Progress to date in breeding underutilized tree crops with relevance to the Midwest U.S.

Tree Crop	Cultivated Species	Breeding Objective	Breeding Stage	Adapted Regions	References
Elderberry	*Sambucus nigra* L. *subspecies canadensis* R. Bolli	Identify adapted varieties, selection for site specific conditions, fruit quality and yield, reduced inter-annual variability, and late bud break.	Multi-location trials of traditional varieties and new germplasm.	MO.	(FINN et al., 2008; THOMAS et al., 2013)
Aronia	*Aronia melanocarpa* (Michx.) Elliot	Identify adapted varieties, narrower and shorter growth habit, total phenolics and anthocyanins, total yield, and low chilling.	Trials of traditional varieties and new germplasm.	CT; NE.	(BRAND, 2010, 2013)
Chinese chestnut	*Castanea mollissima* Blume	Identify adapted varieties, kernel size and quality, yield quantity and consistency.	Variety trials; initiated pedigree breeding.	CT; MO.	(ANAGNOSTAKIS, 1999; HUNT et al., 2004; MORI et al., 2017)
Eastern black walnut	*Juglans nigra* L.	Identify adapted varieties, kernel quality and yield (e.g., nutmeat/shell ratio), alternate bearing, define host resistance to pests and disease, early flowering, and spur-type growth habit.	Multi-location variety trials; pedigree breeding.	MO; KS.	(COGGESHALL, 2010; REID et al., 2004; WARMUND AND COGGESHALL, 2009)
European hazelnut	*Corylus avellana*	Eastern filbert blight resistance, cold hardiness, commercial kernel quality and yield.	Screening wild germplasm; modified backcrossing.	NJ; NY.	(MOLNAR et al., 2011, 2018a; MOLNAR AND CAPIK, 2012)
Northern Pecan	*Carya illinoinensis* (Wangenh) K. Koch	High yield, perocity, kernel quality, kernel percentage, ease of shelling, disease resistance, and reduced masting.	Multi-location trials; controlled crossing.	GA; MO; TX.	(GRAUKE et al., 2016; THOMPSON AND CONNER, 2012)

Table 1.2. Wild relatives and wild utilized species with known traits of value.

Species	Known Traits of Value in Wild Relatives [1]
Sambucus nigra L. *subspecies canadensis* R. Bolli	Commercial yields, late flowering, short ripening period, (FINN et al., 2008), high phenolics and anthocyanins, acylated forms of cyanidin-glycosides, and regional and local adaptation (THOMAS et al., 2013).
Aronia arbutifolia (L.) Pers.	Lower chill hours, fruit ripening date, fruit size, and ripe fruit color (BRAND, 2013; TAHERI et al., 2013).
Aronia melanocarpa (Michx.) Elliot	High anthocyanin and unique profiles (BRAND et al., 2017; TAHERI et al., 2013), plant habit (e.g., prostrate), fruit ripening date, fruit size (BRAND, 2013), and diverse microclimate adaptation (BRAND, 2010).
Aronia prunifolia (Marshall) Rehder	High phenolics (BRAND et al., 2017; TAHERI et al., 2013).
Castanea dentata (Marshall) Borkh.	Regional and local adaptation along with rare genetic diversity (ALEXANDER, 2005).
Castanea mollissima Blume [2]	Resistance to chestnut blight, winter hardiness, kernel size and quality, yield quantity, late bud break, flowering date, early nut maturity (GUO-TIAN et al., 2009; HUNT et al., 2004), naturalize populations (MILLER et al., 2014), and some regional and local adaptation (HUNT et al., 2004).
Juglans nigra L.	Local adaptation, late flowering, anthracnose resistance (MCGRANAHAN AND LESLIE, 2009), and rootstock (NPGS, 2018).
Corylus americana Walter	Eastern filbert blight resistance (THOMPSON et al., 1996), local adaptation (MOLNAR, 2011), cold hardiness (SATHUVALLI AND MEHLENBACHER, 2012), heat and drought tolerance (MOLNAR, 2011), and genetic diversity (DEMCHIK et al., 2017; SATHUVALLI AND MEHLENBACHER, 2012).
Carya illinoinensis (Wangenh) K. Koch	Disease resistance, drought and heat tolerance, cold hardiness, and tree size reduction (GRAUKE et al., 2016).

[1] Traits were synthesized from the referenced literature and adapted in part from Khoury et al. (2013).

[2] The Chinese chestnut is included because populations of the endemic chestnut (*Castanea dentata*) have been severely reduced due to *Cryphonectria parasitica*, and, as a consequence, the chinese chestnut has been introduced and naturalized in the eastern U.S. (MILLER et al., 2014). Performance trials of chinese chestnut accessions also demonstrate suitable adaptability to the Eastern U.S. (ANAGNOSTAKIS, 1999; HUNT et al., 2004).

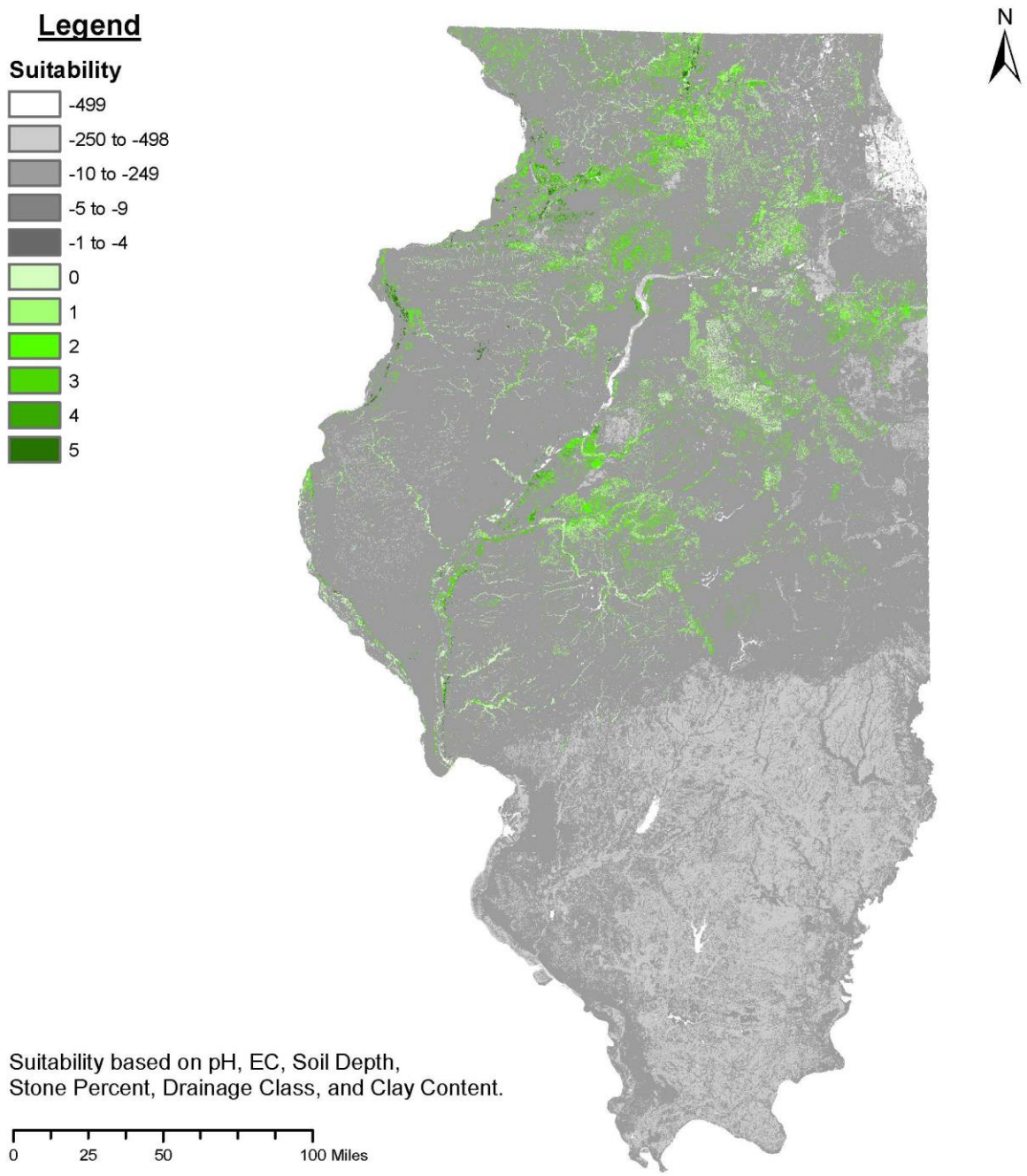


Figure 1.1. A suitability analysis of the soil characteristics relevant to the successful cultivation of the european hazelnut *Corylus avellana* in the state of Illinois in the Midwest U.S. Soil characteristics and corresponding discrete suitability classes, along with details of the analysis, are listed in the supplemental Table A.1. Suitability rankings were not weighted. Green shaded map units indicate that all six soil characteristics are classified as suited or well-suited. Grey shading indicates marginal or unsuited map units, with rankings below zero.

Table 1.3. Collected tree crop wild relatives and wild utilized species endemic to the Eastern U.S.

Species	Accessions		States [1] Represented in Endemic Range	States Represented in Germplasm Collections	
				Populations Collected	
	USDA-ARS [2]	University [3]		>10	1 to 10
Sambucus nigra L. *subspecies canadensis* R. Bolli	38 [4]	55 [5]	52	1	8
Aronia arbutifolia (L.) Pers.	20	19 [6]	29	0	13
Aronia melanocarpa (Michx.) Elliot	50	57 [6]	36	1	16
Aronia prunifolia (Marshall) Rehder	28	41 [6]	31	2	6
Castanea dentata (Marshall) Borkh.	1	Not published [7]	30	-	-
Castanea mollissima Blume [8]	239 [9]	65 [10]	N/A	N/A	N/A
Juglans nigra L.	27	64	44	0	8
Corylus americana Walter	43	~100 [11]	39	2	27
Carya illinoinensis (Wangenh) K. Koch	3615		21	Data not available	

[1] U.S. states and Canadian territories.

[2] Data was obtained from the USDA-ARS Germplasm Resources Information Network (GRIN) taxonomy for plants database (NPGS, 2018).

[3] Figures are based on those collections in which published records were available in the literature.

[4] (BUSHAKRA et al., 2013)

[5] (BYERS et al., 2005)

[6] (ANAGNOSTAKIS, 1999).

[7] The American Chestnut Foundation has conducted a longstanding backcrossing program to develop blight resistant *Castanea dentata*. Germplasm collections also exist at the University of Connecticut and the University of Tennessee (ALEXANDER, 2005).

[8] The chinese chestnut is included because populations of the endemic chestnut (*Castanea dentata*) have been severely reduced due to *Cryphonectria parasitica*, and, as a consequence, the chinese chestnut has been introduced and naturalized in the Eastern U.S. (MILLER et al., 2014). Performance trials of chinese chestnut accessions also demonstrate suitable adaptability to the Eastern U.S. (ANAGNOSTAKIS, 1999; HUNT et al., 2004).

[9] (GUO-TIAN et al., 2009).

[10] (HUNT et al., 2004; MORI et al., 2017).

[11] (SATHUVALLI et al., 2012)

1.9 REFERENCES

Aitken, S.N., S. Yeaman, J.A. Holliday, T. Wang, and S. Curtis-McLane. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. Evol. Appl. 1:95-111.

Alexander, M.T. 2005. Conservation of *Castanea dentata* genetic resources of the southeastern United States. University of Tennessee at Chattanooga, Environmental Science.

Allen, S.C., S. Jose, P.K.R. Nair, B.J. Brecke, P. Nkedi-Kizza, and C.L. Ramsey. 2004. Safety-net role of tree roots: evidence from a pecan (*Carya illinoensis* K. Koch)–cotton (*Gossypium hirsutum* L.) alley cropping system in the southern United States. Forest Ecol. and Mgt. 192:395-407.

Amadi, C.C., K.C. Van Rees, and R.E. Farrell. 2016. Soil–atmosphere exchange of carbon dioxide, methane and nitrous oxide in shelterbelts compared with adjacent cropped fields. Agr., Ecosystems Environ. 223:123-134.

Anagnostakis, S.L. 1999. Cultivars of chestnuts. Annu. Rpt. Northern Nut Growers Assn. 90:16-31.

Anderson, S.H., R.P. Udawatta, T. Seobi, and H.E. Garrett. 2009. Soil water content and infiltration in agroforestry buffer strips. Agroforestry systems 75:5-16.

Annicchiarico, P., F. Bellah, and T. Chiari. 2005. Defining subregions and estimating benefits for a specific-adaptation strategy by breeding programs. Crop Sci. 45:1741-1749.

Atlin, G.N. 2001. Breeding for suboptimal environments.

Baldwin, B.J. 2015. The growth and productivity of hazelnut cultivars (*Corylus avellana* L.) in Australia.

Bambo, S.K., J. Nowak, A.R. Blount, A.J. Long, and A. Osiecka. 2009. Soil nitrate leaching in silvopastures compared with open pasture and pine plantation. J. Environ. Qual. 38:1870-1877.

Barbier, E.B. 1989. Sustaining agriculture on marginal land: a policy framework. Environ.: Sci. and Policy Sustainable Dev. 31:12-40.

Bhattarai, G., S.A. Mehlenbacher, and D.C. Smith. 2017. Eastern filbert blight disease resistance from *Corylus americana* 'Rush' and selection 'Yoder #5' maps to linkage group 7. Tree Genet. Genomes 13.

Böhm, C., M. Kanzler, and D. Freese. 2014. Wind speed reductions as influenced by woody hedgerows grown for biomass in short rotation alley cropping systems in Germany. Agroforestry systems 88:579-591.

Brand, M. 2010. Aronia: Native shrubs with untapped potential. Arnoldia 67:14-25.

Brand, M.H. 2013. Breeding Better Aronia Plants©. Proceedings of the International Plant Propagators Society-2013 1055:269-272.

Brand, M.H., B.A. Connolly, L.H. Levine, J.T. Richards, S.M. Shine, and L.E. Spencer. 2017. Anthocyanins, total phenolics, ORAC and moisture content of wild and cultivated dark-fruited *Aronia* species. Scientia Hort. 224:332-342.

Brandes, E., G.S. McNunn, L.A. Schulte, I.J. Bonner, D. Muth, B.A. Babcock, B. Sharma, and E.A. Heaton. 2016. Subfield profitability analysis reveals an economic case for cropland diversification. Environ. Res. Lett. 11:014009.

Brandes, E., G.S. McNunn, L.A. Schulte, D.J. Muth, A. VanLoocke, and E.A. Heaton. 2017. Targeted subfield switchgrass integration could improve the farm economy, water quality, and bioenergy feedstock production. GCB Bioenergy.

Brown, G.R., G.P. Gill, R.J. Kuntz, C.H. Langley, and D.B. Neale. 2004. Nucleotide diversity and linkage disequilibrium in loblolly pine. Proc. National Acad. Sci. 101:15255-15260.

Brumlop, S., W. Reichenbecher, B. Tappeser, and M.R. Finckh. 2013. What is the SMARTest way to breed plants and increase agrobiodiversity? Euphytica 194:53-66.

Capik, J.M. and T.J. Molnar. 2012. Assessment of host (*Corylus* sp.) resistance to eastern filbert blight in New Jersey. J. Amer. Soc. Hort. Sci. 137:157-172.

Castañeda-Álvarez, N.P., C.K. Khoury, H.A. Achicanoy, V. Bernau, H. Dempewolf, R.J. Eastwood, L. Guarino, R.H. Harker, A. Jarvis, N. Maxted, J.V. Müller, J. Ramirez-Villegas, C.C. Sosa, P.C. Struik, H. Vincent, and J. Toll. 2016. Global conservation priorities for crop wild relatives. Nature Plants 2. DOI 10.1038/NPLANTS.2016.22

Ceccarelli, S. 2015. Efficiency of plant breeding. Crop Sci. 55:87-97.

Cernusca, M. and M. Gold. 2013. Breaking down market barriers for elderberry growers and producers. Acta Hort. 1061:269-278.

Charlebois, D., P.L. Byers, C.E. Finn, and A.L. Thomas. 2010. 4 Elderberry: Botany, Horticulture, Potential. Hort. Reviews 37:213.

Coggeshall, M. 2010. Use of microsatellite markers to develop new eastern black walnut cultivars in Missouri, USA. Acta Hort. 918:221-226.

Cornille, A., P. Gladieux, M.J. Smulders, I. Roldan-Ruiz, F. Laurens, B. Le Cam, A. Nersesyan, J. Clavel, M. Olonova, and L. Feugey. 2012. New insight into the history of domesticated apple: secondary contribution of the

european wild apple to the genome of cultivated varieties. PLoS Genet. 8. DOI 10.1371/journal.pgen.1002703.

Demchik, M., A. Kern, L. Braun, J. Fischbach, and K. Turnquist. 2017. Genetic diversity of american hazelnut in the Upper Midwest, USA. Agroforestry Systems 92:1507-1516.

Dempewolf, H., G. Baute, J. Anderson, B. Kilian, C. Smith, and L. Guarino. 2017. Past and future use of wild relatives in crop breeding. Crop Sci. 57:1070-1082.

Dempewolf, H., R.J. Eastwood, L. Guarino, C.K. Khoury, J.V. Müller, and J. Toll. 2014. Adapting Agriculture to Climate Change: A Global Initiative to Collect, Conserve, and Use Crop Wild Relatives. Agroecology and Sustainable Food Systems 38:369-377.

Dobos, R., H. Sinclair, and K. Hipple. 2012. National commodity crop productivity index (NCCPI) user guide v2. 0. USDA NRCS National Soil Survey Center, Lincoln, NE.

Drumke, J. 1964. A systematic survey of *Corylus* in North America. Univ. of Tenn., Knoxville, TN, PhD diss.

Duminil, J., S. Fineschi, A. Hampe, P. Jordano, D. Salvini, G.G. Vendramin, and R.J. Petit. 2007. Can population genetic structure be predicted from life-history traits? The American Naturalist 169:662-672.

Duminil, J., O.J. Hardy, and R.J. Petit. 2009. Plant traits correlated with generation time directly affect inbreeding depression and mating system and indirectly genetic structure. BMC Evolutionary Biol. 9:177.

Ellena, M., P. Sandoval, A. Gonzalez, J. Jequier, M. Contreras, and P. Grau Beretta. 2012. Chilean hazelnut situation and perspectives. Acta Hort. 1052:329-342.

Ellstrand, N.C., H.C. Prentice, and J.F. Hancock. 1999. Gene flow and introgression from domesticated plants into their wild relatives. Annu. Rev. of Ecol. Systematics 30:539-563.

Erdogan, V. and S.A. Mehlenbacher. 2000. Interspecific hybridization in hazelnut (*Corylus*). J. Amer. Soc. Hort. Sci. 125:489-497.

FAOSTAT. 2014. Food and Agriculture Organization of the United Nations Cropping Database.

Finn, C.E., A.L. Thomas, P.L. Byers, and S. Serçe. 2008. Evaluation of american (*Sambucus canadensis*) and european (*S. nigra*) elderberry genotypes grown in diverse environments and implications for cultivar development. HortScience 43:1385-1391.

Fischer, R., D. Byerlee, and G. Edmeade. 2014. Crop yields and global food security. ACIAR: Canberra, ACT.

Ford-Lloyd, B.V., M. Schmidt, S.J. Armstrong, O. Barazani, J. Engels, R. Hadas, K. Hammer, S.P. Kell, D. Kang, and K. Khoshbakht. 2011. Crop wild relatives—undervalued, underutilized and under threat? BioScience 61:559-565.

Fuller, A. 1908. The nut culturist. Orange Judd, New York, NY.

Garrett, H., R. McGraw, and W. Walter. 2009. Alley cropping practices. North American Agroforestry: Integrated Sci. and Practice 2nd edition:133-162.

Gönenc, S., H. Tanrıvermis, and M. Bülbül. 2006. Economic assessment of hazelnut production and the importance of supply management approaches in Turkey. J. Agri. and Rural Dev. Tropics and Subtropics 107:19-32.

Grau, P. and R. Bastia. 2004. Productivity and yield efficiency of hazelnut (*Corylus avellana* L.) cultivars in Chile.

Grauke, L., B.W. Wood, and M.K. Harris. 2016. Crop vulnerability: *Carya*. HortScience 51:653-663.

Guo-tian, L., A. Cheng-xiang, Z. Li-si, W. Hai-rong, and L. Qing-zhong. 2009. Chestnut genebank in China national clonal plant germplasm repository. Acta Hort.:199.

Hamrick, J.L., M.J.W. Godt, and S.L. Sherman-Broyle. 1992. Factors influencing levels of genetic diversity in woody plant species, p. 95-124, Population genetics of forest trees. Springer.

Hedrick, P.W. 2006. Genetic polymorphism in heterogeneous environments: the age of genomics. Annu. Rev. Ecol. Evol. Syst. 37:67-93.

Heuertz, M., E. De Paoli, T. Källman, H. Larsson, I. Jurman, M. Morgante, M. Lascoux, and N. Gyllenstrand. 2006. Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce (*Picea abies* (L.) Karst). Genet. 174:2095-2105.

Hunt, K.L., M.A. Gold, and M.R. Warmund. 2004. Chinese chestnut cultivar performance in Missouri. Acta Hort. 693:145-148.

Johnson, K. and J. Pinkertons. 2002. Eastern filbert blight. Compendium of Nut Crop Dis. in Temperate Zones. Amer. Phytopathol. Soc. Press, St. Paul, Minn:44-46.

Jose, S. 2009. Agroforestry for ecosystem services and environmental benefits: an overview. Agroforestry Systems 76:1-10.

Jose, S. 2012. Agroforestry for conserving and enhancing biodiversity. Agroforestry Systems 85:1-8.

Kang, S., W. Post, D. Wang, J. Nichols, V. Bandaru, and T. West. 2013. Hierarchical marginal land assessment for land use planning. Land Use Policy 30:106-113.

Kawecki, T.J. and D. Ebert. 2004. Conceptual issues in local adaptation. Ecol. Lett. 7:1225-1241.

Khoury, C.K., S. Greene, J. Wiersema, N. Maxted, A. Jarvis, and P.C. Struik. 2013. An inventory of crop wild relatives of the United States. Crop Sci. 53:1496-1508.

Khoury, C.K., S.L. Greene, K.A. Williams, C. Sosa, and C. Richards. 2017. Conservation Priorities for Tree Crop Wild Relatives in the United States.

Kidd, D., M. Webb, B. Malone, B. Minasny, and A. McBratney. 2015. Digital soil assessment of agricultural suitability, versatility and capital in Tasmania, Australia. Geoderma Regional 6:7-21.

Kim, D.-G., M.U. Kirschbaum, and T.L. Beedy. 2016. Carbon sequestration and net emissions of CH4 and N2O under agroforestry: Synthesizing available data and suggestions for future studies. Agri., Ecosystems Environ. 226:65-78.

Kole, C., M. Muthamilarasan, R. Henry, D. Edwards, R. Sharma, M. Abberton, J. Batley, A. Bentley, M. Blakeney, J. Bryant, H. Cai, M. Cakir, L.J. Cseke, J. Cockram, A.C. de Oliveira, C. De Pace, H. Dempewolf, S. Ellison, P. Gepts, A. Greenland, A. Hall, K. Hori, S. Hughes, M.W. Humphreys, M. Iorizzo, A.M. Ismail, A. Marshall, S. Mayes, H.T. Nguyen, F.C. Ogbonnaya, R. Ortiz, A.H. Paterson, P.W. Simon, J. Tohme, R. Tuberosa, B. Valliyodan, R.K. Varshney, S.D. Wullschleger, M. Yano, and M. Prasad. 2015. Application of genomics-assisted breeding for generation of climate resilient crops: Progress and prospects. Frontiers Plant Sci. 6.

Le Corre, V. and A. Kremer. 2012. The genetic differentiation at quantitative trait loci under local adaptation. Mol. Ecol. 21:1548-1566.

Leakey, R.R.B. 2014. The role of trees in agroecology and sustainable agriculture in the tropics, p. 113-133, Annu. Rev. Phytopathol.

Lee, J. and C.E. Finn. 2007. Anthocyanins and other polyphenolics in american elderberry (*Sambucus canadensis*) and european elderberry (*S. nigra*) cultivars. J. Sci. Food and Agri. 87:2665-2675.

Lerch, R., N. Kitchen, R. Kremer, W. Donald, E. Alberts, E. Sadler, K. Sudduth, D. Myers, and F. Ghidey. 2005. Development of a conservation-oriented precision agriculture system: Water and soil quality assessment. J. Soil Water Conserv. 60:411-421.

Lijavetzky, D., J.A. Cabezas, A. Ibáñez, V. Rodríguez, and J.M. Martínez-Zapater. 2007. High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. BMC Genomics 8:424.

Lin, B.B. 2007. Agroforestry management as an adaptive strategy against potential microclimate extremes in coffee agriculture. Agri. and Forest Meteorol. 144:85-94.

Linhart, Y.B. and M.C. Grant. 1996. Evolutionary significance of local genetic differentiation in plants. Annual review of ecology and systematics 27:237-277.

Lobell, D.B. and G. Azzari. 2017. Satellite detection of rising maize yield heterogeneity in the US Midwest. Environ. Res. Lett. 12:014014.

Lobell, D.B., G.L. Hammer, G. McLean, C. Messina, M.J. Roberts, and W. Schlenkers. 2013. The critical role of extreme heat for maize production in the United States. Nature Climate Change 3:497-501.

Lobell, D.B., M.J. Roberts, W. Schlenker, N. Braun, B.B. Little, R.M. Rejesus, and G.L. Hammer. 2014. Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest. Science 344:516-519.

Loveless, M.D. and J.L. Hamrick. 1984. Ecological determinants of genetic structure in plant populations. Annual review of ecology and systematics 15:65-95.

Lovell, S.T., C. Dupraz, M. Gold, S. Jose, R. Revord, E. Stanek, and K.J. Wolz. 2017. Temperate agroforestry research: considering multifunctional woody polycultures and the design of long-term field trials. Agroforestry Systems:1-19.

Lovell, S.T. and D.M. Johnston. 2009. Creating multifunctional landscapes: How can the field of ecology inform the design of the landscape? Frontiers Ecol. Envir. 7:212-220.

Mattia, C.M. 2017. Identifying targets for multifunctional perennial cropping systems implementation through a landowner adoption preferences survey and a spatial suitability analysis.

Mattia, C.M., S.T. Lovell, and A. Davis. 2016. Identifying barriers and motivators for adoption of multifunctional perennial cropping systems by landowners in the Upper Sangamon River Watershed, Illinois. Agroforestry Systems:1-15.

Maxted, N., S. Kell, B. Ford-Lloyd, E. Dulloo, and Á. Toledo. 2012. Toward the Systematic Conservation of Global Crop Wild Relative Diversity. Crop Sci. 52:774-785.

Mayes, S., F.J. Massawe, P.G. Alderson, J.A. Roberts, S.N. Azam-Ali, and M. Hermann. 2011. The potential for underutilized crops to improve security of food production. J. Exp. Bot. 63:1075-1079.

McClure, K.A., J. Sawler, K.M. Gardner, D. Money, and S. Myles. 2014. Genomics: A potential panacea for the perennial problem. Amer. J. Bot. 101:1780-1790.

McGranahan, G. and C. Leslie. 2009. Breeding walnuts (*Juglans regia*), p. 249-273, Breeding plantation tree crops: Temperate species. Springer.

Mehlenbacher, S.A. 2008. Genetic resources for hazelnut: state of the art and future perspectives.

Melhenbacher, S. 2004. The hazelnut situation in Oregon.

Migicovsky, Z. and S. Myles. 2017. Exploiting wild relatives for genomics-assisted breeding of perennial crops. Frontiers Plant Sci. 8.

Migicovsky, Z., J. Sawler, D. Money, R. Eibach, A.J. Miller, J.J. Luby, A.R. Jamieson, D. Velasco, S. von Kintzel, J. Warner, W. Wührer, P.J. Brown, and S. Myless. 2016. Genomic ancestry estimation quantifies use of wild species in grape breeding. BMC Genomics 17:478.

Miller, A.C., K.E. Woeste, S.L. Anagnostakis, and D.F. Jacobss. 2014. Exploration of a rare population of chinese chestnut in North America: stand dynamics, health and genetic relationships. AoB Plants 6:plu065.

Miller, A.J. and B.L. Gross. 2011. From forest to field: Perennial fruit crop domestication. Amer. J. Bot. 98:1389-1414.

Molnar, T., J. Lombardoni, M. Muehlbauer, J. Honig, S. Mehlenbacher, and J. Capik. 2018a. Progress breeding for resistance to eastern filbert blight in the eastern United States. Acta Hort. 1226: 79-86.

Molnar, T.J. 2011. *Corylus*, p. 15-48, Wild Crop Relatives: Genomic and Breeding Resources. Springer, Berlin, Heidelberg.

Molnar, T.J. and J. Capik. 2012. Advances in hazelnut research in North America. Acta Hort. 940:57-65.

Molnar, T.J., J.A. Honig, A. Mayberry, R.S. Revord, S.T. Lovell, S.A. Mehlenbacher, and J.M. Capik. 2018b. *Corylus americana*: a valuable genetic resource for developing hazelnuts adapted to the eastern United States. Acta Hort 1226:115-121.

Molnar, T.J., P.C. Kahn, T.M. Ford, C.J. Funk, and C.R. Funk. 2013. Tree crops, a permanent agriculture: Concepts from the past for a sustainable future. Resources 2:457-488.

Mori, G.O., M. Gold, and S. Jose. 2017. Specialty Crops in Temperate Agroforestry Systems: Sustainable Management, Marketing and Promotion for the Midwest Region of the USA, p. 331-366, Integrating Landscapes: Agroforestry for Biodiversity Conservation and Food Sovereignty. Springer.

Mosquera-Losada, M., D. Freese, and A. Rigueiro-Rodríguez. 2011. Carbon sequestration in european agroforestry systems, p. 43-59, Carbon sequestration potential of agroforestry systems. Springer.

Muehlbauer, M., J. Tobia, J. Honig, B. Hillman, N. Zhang, K.M. Gold, and T. Molnar. 2018. Population differentiation within *Anisogramma anomala* in North America. Phytopathol. 109:1074-1082.

Muth, D. 2014. Profitability versus environmental performance: Are they competing? J. Soil Water Conserv. 69:203A-206A.

Myles, S., A.R. Boyko, C.L. Owens, P.J. Brown, F. Grassi, M.K. Aradhya, B. Prins, A. Reynolds, J.-M. Chia, and D. Ware. 2011. Genetic structure and domestication history of the grape. Proc. National Acad. Sci. 108:3530-3535.

Neale, D.B. and O. Savolainen. 2004. Association genetics of complex traits in conifers. Trends in Plant Sci. 9:325-330.

Niu, X. and S.W. Duikers 2006. Carbon sequestration potential by afforestation of marginal agricultural land in the Midwestern US. Forest Ecol. and Mgt. 223:415-427.

NPGS, U. 2018. Genetic Resources Information Service. GRIN taxonomy, https://npgsweb.ars-grin.gov/gringlobal/taxon/taxonomysearch.aspx.

Olsen, J. 2001. Nutrient management guide—Hazelnut. Oregon State University Extension Service, EM 8786.

Olsen, J. 2013. Orchard nutrition. Growing hazelnuts in the Pacific Northwest. Oregon State University Extension Service Corvallis, OR:1-5.

Özgen, M., J.C. Scheerens, R.N. Reese, and R.A. Miller. 2010. Total phenolic, anthocyanin contents and antioxidant capacity of selected elderberry (*Sambucus canadensis* L.) accessions. Pharmacognosy magazine 6:198.

Öztürk, S.C., H.İ. Balık, S.K. Balık, G. Kızılcı, Ö. Duyar, S. Doğanlar, and A. Frary. 2017. Molecular genetic diversity of the Turkish national hazelnut collection and selection of a core set. Tree Genet. Genomes 13:113.

Parchman, T.L., Z. Gompert, J. Mudge, F.D. Schilkey, C.W. Benkman, and C.A. Buerkle. 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. Mol. Ecol. 21:2991-3005.

Petit, R.J.a.H., A. 2006. Some Evolutionary Consequences of Being a Tree. Annu. Rev. Ecol. Evol. and Systematics 37:187-214.

Pritchard, J.K. and A. Di Rienzo. 2010. Adaptation – not by sweeps alone. Nature Rev. Genet. 11:665.

Prohens, J., P. Gramazio, M. Plazas, H. Dempewolf, B. Kilian, M.J. Díez, A. Fita, F.J. Herraiz, A. Rodríguez-Burruezo, S. Soler, S. Knapp, and S. Vilanova. 2017. Introgressiomics: a new approach for using crop wild relatives in breeding for adaptation to climate change. Euphytica 213.

Reid, W., M.V. Coggeshall, and K.L. Hunt. 2004. Cultivar evaluation and development for black walnut orchards.

Rhodes, T.K., F.X. Aguilar, S. Jose, and M. Gold. 2016. Factors influencing the adoption of riparian forest buffers in the Tuttle Creek Reservoir watershed of Kansas, USA. Agroforestry Systems:1-19.

Richards, B.K., C.R. Stoof, I.J. Cary, and P.B. Woodbury. 2014. Reporting on marginal lands for bioenergy feedstock production: a modest proposal. BioEnergy Res. 7:1060-1062.

Rigueiro-Rodríguez, A., E. Fernández-Núñez, P. González-Hernández, J. McAdam, and M.R. Mosquera-Losada. 2009. Agroforestry systems in Europe: productive, ecological and social perspectives, p. 43-65, Agroforestry in Europe. Springer.

Sathuvalli, V.R. and S.A. Mehlenbacher. 2012. Characterization of american hazelnut (*Corylus americana*) accessions and *Corylus americana* × *Corylus avellana* hybrids using microsatellite markers. Genet. Resources and Crop Evolution 59:1055-1075.

Savolainen, O., M. Lascoux, and J. Merilä. 2013. Ecological genomics of local adaptation. Nature Rev. Genet. 14:807.

Schoeneberger, M., G. Bentrup, H. De Gooijer, R. Soolanayakanahally, T. Sauer, J. Brandle, X. Zhou, and D. Current. 2012. Branching out: Agroforestry as a climate change mitigation and adaptation tool for agriculture. J. Soil Water Conserv. 67:128A-136A.

Schoeneberger, M.M.s. 2009. Agroforestry: working trees for sequestering carbon on agricultural lands. Agroforestry systems 75:27-37.

Schoeneberger, M.M., G. Bentrup, and T. Patel-Weynands. 2017. Agroforestry: enhancing resiliency in US agricultural landscapes under changing conditions. Gen. Tech. Report WO-96. Washington, DC: US Department of Agriculture, Forest Service 96.

Simmonds, N. 1991. Selection for local adaptation in a plant breeding programme. Theoretical and Appl. Genet. 82:363-367.

Siriri, D., J. Wilson, R. Coe, M. Tenywa, M. Bekunda, C. Ong, and C. Black. 2013. Trees improve water storage and reduce soil evaporation in agroforestry systems on bench terraces in SW Uganda. Agroforestry systems 87:45-58.

Smith, J.R. 2013. Tree crops: A permanent agriculture. Island Press.

Sork, V., S. Aitken, R. Dyer, A. Eckert, P. Legendre, and D. Neale. 2013. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. Tree Genet. Genomes 9:901-911.

Staff, S. 2017. Soil Survey Geographic (SSURGO) Database for Illinois. Natural Resources Conservation Service, United States Department of Agriculture.

Stanek, E.C., Lovell, S.T. & Reisner, A. 2019. Designing multifunctional woody polycultures according to landowner preferences in Central Illinois. Agroforestry Systems.

Taheri, R., B.A. Connolly, M.H. Brand, and B.W. Bollings. 2013. Underutilized chokeberry (*Aronia melanocarpa*, *Aronia arbutifolia*, *Aronia prunifolia*) accessions are rich sources of anthocyanins, flavonoids, hydroxycinnamic acids, and proanthocyanidins. J. Agri. Food Chem. 61:8581-8588.

Technavio. 2017. Global Hazelnut Market 2017-2021. EMIS Intelligence Database.

Thomas, A., P. Byers, J. Avery, JD, M. Kaps, and S. Gu. 2013. Horticultural performance of eight american elderberry genotypes at three Missouri locations. 1061:237-244.

Thompson, M.M., H.B. Lagerstedt, and S.A. Mehlenbacher. 1996. Hazelnuts, p. 125-184, Fruit breeding.

Thompson, T.E. and P.J. Conner. 2012. Pecan, p. 771-801, Fruit Breeding. Springer.

Torralba, M., N. Fagerholm, P.J. Burgess, G. Moreno, and T. Plieninger. 2016. Do european agroforestry systems enhance biodiversity and ecosystem services? A meta-analysis. Agr., Ecosystems Environ. 230:150-161.

Tsonkova, P., C. Böhm, A. Quinkenstein, and D. Freese. 2012. Ecological benefits provided by alley cropping systems for production of woody biomass in the temperate region: a review. Agroforestry Systems 85:133-152.

Udawatta, R.P. and S. Jose. 2011. Carbon sequestration potential of agroforestry practices in temperate North America, p. 17-42, Carbon Sequestration Potential of Agroforestry Systems. Springer.

Udawatta, R.P. and S. Jose. 2012. Agroforestry strategies to sequester carbon in temperate North America. Agroforestry Systems 86:225-242.

Udawatta, R.P., J.J. Krstansky, G.S. Henderson, and H.E. Garrett. 2002. Agroforestry practices, runoff, and nutrient loss. J. Environ. Qual. 31:1214-1225.

Ustaoğlu, B. 2012. The effect of climatic conditions on hazelnut (*Corylus avellana*) yield in Giresun (Turkey).

Varshney, R.K., R. Terauchi, and S.R. McCouch. 2014. Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. PLoS Biol. 12:e1001883.

Verchot, L.V., M. Van Noordwijk, S. Kandji, T. Tomich, C. Ong, A. Albrecht, J. Mackensen, C. Bantilan, K. Anupama, and C. Palm. 2007. Climate change: linking adaptation and mitigation through agroforestry. Mitigation and Adaptation Strategies Global Change 12:901-918.

Wallace, D.C. and F.J. Young. 2008. Black walnut suitability index: a natural resources conservation service national soil information system based interpretive model. In: Jacobs, Douglass F.; Michler, Charles H., eds. 2008. Proceedings, 16th Central Hardwood Forest Conference; 2008 April 8-9; West Lafayette, IN. Gen. Tech. Rep. NRS-P-24. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station: 589-595.

Warmund, M. and M. Coggeshall. 2009. Flowering and fruit characteristics of black walnut cultivars at the University of Missouri repository. Acta Hort. 861:177-182.

Warschefsky, E., R. Varma Penmetsa, D.R. Cook, and E.J.B. Von Wettberg. 2014. Back to the wilds: Tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. Amer. J. Bot. 101:1791-1800.

Weschcke, C. 1953. Growing nuts in the north. Webb, St. Paul, MN.

Wiegmann, K., K.J. Hennenberg, and U.R. Fritsche. 2008. Degraded land and sustainable bioenergy feedstock production.

Wolz, K.J., B.E. Branham, and E.H. DeLucia. 2018. Reduced nitrogen losses after conversion of row crop agriculture to alley cropping with mixed fruit and nut trees. Agriculture, Ecosystems Environ. 258:172-181.

Wolz, K.J. and E.H. DeLucia. 2018a. Alley cropping: Global patterns of species composition and function. Agriculture, Ecosystems Environ. 252:61-68.

Wolz, K.J. and E.H. DeLucia. 2018b. Black walnut alley cropping is economically competitive with row crops in the Midwest USA. Ecol. Applications 29 DOI 10.1002/eap.1829.

Wolz, K.J., S.T. Lovell, B.E. Branham, W.C. Eddy, K. Keeley, R.S. Revord, M.M. Wander, W.H. Yang, and E.H. DeLucia. 2017. Frontiers in alley cropping: Transformative solutions for temperate agriculture. Global Change Biol. 24:883-894.

Xie, M., J. Zheng, and D. Wang. 2012. Achievements and perspective in hazelnut breeding in China. Acta Hort. 1052:41-44.

Zomer, R.J., H. Neufeldt, J. Xu, A. Ahrends, D. Bossio, A. Trabucco, M. van Noordwijk, and M. Wang. 2016. Global Tree Cover and Biomass Carbon on Agricultural Land: The contribution of agroforestry to global and national carbon budgets. Scientific Rpt. 6:29987.

CHAPTER 2

EASTERN FILBERT BLIGHT RESISTANCE IN AMERICAN AND INTERSPECIFIC HYBRID HAZELNUT (*CORYLUS AMERICANA* × *C. AVELLANA*)

2.1 ABSTRACT

Eastern filbert blight (EFB), caused by the fungus *Anisogramma anomala*, is a primary limitation to european hazelnut (*Corylus avellana*) cultivation in eastern North America. *Corylus americana*, the american hazelnut, is the endemic host of *A. anomala* and, despite its tiny, thick-shelled nuts, is a potentially valuable source of EFB resistance and climatic adaptation. Interspecific hybrids (*C. americana* × *C. avellana*) have been explored for nearly a century as a means to combine EFB resistance with wider adaptability and larger nuts. While significant progress was made in the past, the genetic diversity of the starting material was limited and additional improvements are needed for expansion of hazelnut production outside of Oregon, where 99% of the U.S. crop is currently produced. To improve our understanding of *C. americana* as a donor of EFB resistance, 29 diverse EFB-resistant *C. americana* accessions were crossed with EFB-susceptible *C. avellana* selections (31 total progenies) to produce 2031 F_1 plants. Additionally, new *C. americana* germplasm was procured from across the native range of the species. The new collection of 1335 plants from 122 seed lots represents 72 counties and 22 states. The interspecific hybrid progenies and a subset of the american collection (616 trees from 62 seed lots) were field planted and evaluated for EFB response following field inoculations and natural disease spread over seven growing seasons. EFB was rated on a scale of 0 (no EFB) to 5 (all stems containing cankers). Results showed that progeny means of the interspecific hybrids ranged from 0.96 to 4.72. Fourteen of the 31 progenies were comprised of at least one-third EFB-free or highly tolerant offspring (i.e., ratings 0 to 2), transmitting a significant level of resistance/tolerance. Several corresponding *C. americana* accessions that imparted a greater degree of resistance to their hybrid offspring were also identified. In addition, results showed that 587 of the 616 (95.3%) *C. americana* plants evaluated remained completely free of EFB, confirming reports that the species rarely expresses signs or symptoms of the disease and should be further studied and used in breeding.

2.2 INTRODUCTION

The *Corylus* L. genus includes thirteen polymorphic deciduous shrub and tree species and all bear edible nuts (BASSIL et al., 2013; ERDOGAN AND MEHLENBACHER, 2000b). The genus demonstrates wide morphological diversity and environmental adaptability, with species adapted to forest habitats throughout the Northern hemisphere (MEHLENBACHER, 1991; MOLNAR, 2011). Current cultivation depends almost entirely on the european hazelnut, *C. avellana*, and is restricted to Mediterranean-like climates, typically near large bodies of water, that allow consistent yields. The scale of global hazelnut production closely mirrors that of pistachios (*Pistacia vera*) at approximately one million tonnes of annual in-shell production (FAOSTAT, 2017), and its market is experiencing steady growth (TECHNAVIO, 2017). Production in the U.S. occurs primarily in the Willamette Valley of Oregon (USDA, 2018), which produces around 5% of the global supply, though there is considerable interest in expanding cultivation to other regions in the U.S. and Canada (FISCHBACH AND BRAUN, 2017; MOLNAR AND CAPIK, 2012a).

Climatic suitability and the broad presence of EFB limit the cultivation of european hazelnut in the eastern and midwestern U.S. (THOMPSON et al., 1996). The causal organism is *Anisogramma anomala* (Peck) E. Müller, an obligate biotrophic ascomycete in the order Diaporthales that is endemic to eastern North America (JOHNSON AND PINKERTON, 2002). In nature, *A. anomala* co-exists with *C. americana*, occasionally producing minor, inconsequential stem cankers (CAPIK AND MOLNAR, 2012; FULLER, 1908; WESCHCKE, 1953). Conversely, infection of *C. avellana* generally results in severe, perennial stem cankers that lead to branch dieback, yield decline, and in many cases plant death (JOHNSON AND PINKERTON, 2002). *Corylus americana* is native to much of North America east of the Rocky Mountains, extending from Maine to Northern Florida in the east, over to its western boundary that extends from eastern Oklahoma to North Dakota and southern Manitoba (DRUMKE, 1964; GLEASON AND CRONQUIST, 1963). This extensive geography suggests that *C. americana* may represent a diverse source of resistance and tolerance to EFB, and moreover, a source of adaptation to a spectrum of soil types and climatic zones that nearly span the latitudinal range of the U.S.

Thus, if systematically exploited for breeding interspecific hybrids, *C. americana* may enable the expansion of commercial hybrid hazelnut production to a wide portion of the eastern and midwestern U.S. (MOLNAR, 2011; MOLNAR et al., 2005). Unfortunately, the plant architecture and nuts of *C. americana* are not well-suited for commercial production and existing markets. The shrubby architecture and aggressive production of basal sprouts by *C. americana* also present challenges for maintaining the desired single-stemmed form found in modern european hazelnut orchards. Compared to those of *C. avellana* cultivars, the nuts of *C. americana* are very small and enclosed in thick, clasping involucres, which complicates harvest. Further, the shells tend to be very thick. Nevertheless, *C. americana* and *C. avellana* can be crossed in either direction and the resulting hybrids are viable and fertile (ERDOGAN AND MEHLENBACHER, 2000a). Past breeders have noted the potential for continued improvement, with the recovery of most *C. avellana* traits in selected individuals of the first or second backcross generation following interspecific hybridization (MOLNAR, 2011; MOLNAR et al., 2018).

Attempts have been made to develop interspecific hybrids since the early 20th century (reviewed in MOLNAR, 2011; MOLNAR AND CAPIK, 2012b). In 1919, J.F Jones of Lancaster, PA, made the first reported *C. americana* × *C. avellana* F_1 hybrids using the wild selection 'Rush' as the *C. americana* parent crossed with a number of *C. avellana* cultivars. Compared to other *C. americana*, 'Rush' appeared to be better adapted and produced higher yields of larger nuts, and continued to be used as parent by C.A. Reed of the Bureau of Plant Industry, US Department of Agriculture in Beltsville, MD, and G.H. Slate of the New York State Agricultural Experiment Station in Geneva, NY. From 1928 to 1930, Reed and Slate each produced around 2000 F_1 hybrids from crosses of 'Rush' as the female parent with pollen of several *C. avellana* cultivars, and then oversaw their long-term evaluation.

S.H. Graham of Ithaca, NY, also built on Jones' work, evaluating seedlings derived from open-pollinated 'Rush' F_1 hybrids. Additionally, Graham made controlled crosses with a wild *C. americana* selection from Iowa called 'Winkler'. Unfortunately, much of Graham's materials were eventually lost to EFB.

Throughout the 1930s and 1940s, Carl Weschcke of River Falls, WI, also used 'Winkler' to develop many interspecific hybrids. Over time, fatal EFB infections spread throughout most of Weschcke's progenies and no cultivars were released, although there were some survivors (WESCHCKE, 1953, 1963). Later, seeds from select, EFB-resistant individuals that remained in Weschcke's infected plantings were collected by Philip Rutter of Canton, MN, to begin a mass selection program. Rutter subsequently supplemented his population with other wild selections and improved hybrid germplasm, including plants derived from 'Rush'. From this work, populations were developed that expressed EFB resistance, cold hardiness, and improved nut traits (RUTTER, 1987). Seedlings from this population have been disseminated widely throughout the Upper Midwest, where ~130 growers have over 30000 shrubs in production (DEMCHIK et al., 2011). The Arbor Day Foundation (ADF) (Nebraska City, NE) established 5000 seedlings purchased from Rutter in 1996, and from this planting several consistently high yielding, EFB-resistant selections have been identified (CAPIK AND MOLNAR, 2012; HAMMOND, 2006). Additional clonal selections are under replicated evaluation across Wisconsin and Minnesota by the Upper Midwest Hazelnut Development Initiative (BRAUN et al., 2014).

Each of the aforementioned efforts identified EFB-resistant hybrids with improved commercial traits, and Oregon State University (OSU) and the U.S. Department of Agriculture – Agriculture Research Services' National Clonal Germplasm Repository (NCGR) in Corvallis, OR, have conserved a number of these selections for future use. However, while thousands of seedlings and several clonal selections have been sold and distributed by nurseries, none from this early work has supported commercial-scale planting to date. Additionally, these interspecific hybrids were developed from a small number of parents and represent a relatively narrow genetic base. Sathuvalli and Mehlenbacher (2012) characterized 67 *C. americana* × *C. avellana* hybrids jointly held by the NCGR, OSU, and the ADF using microsatellite markers in comparison to a collection of 87 pure *C. americana*, and found nearly all of the hybrid accessions clustered with either 'Rush' or 'Winkler' and the Weschcke hybrid group. While levels of diversity were similarly high in these interspecific hybrid groups and wild *C. americana*, the interspecific hybrids had fewer alleles at the microsatellite marker loci (SATHUVALLI AND MEHLENBACHER, 2012). Moreover, these

early reports presented little data on the inheritance of EFB resistance, but their records mention instances in which large seedling populations were almost entirely lost to EFB (RUTTER, 1991; SLATE, 1969). The absence of systematic studies and a poor understanding of the inheritance of EFB resistance from *C. americana* is a challenge for reaching breeding objectives, which expand beyond disease resistance and include kernel traits, yield, climatic adaption, etc.

Several recent studies have begun to shed light on the inheritance of EFB resistance from *C. americana*. Molnar and Capik (2012b) reported that three 'Rush'-related progenies segregated in a ratio of 1 resistant : 1 susceptible, indicating a dominant allele at a single locus. Bhattarai et al. (2017a) confirmed this mode of inheritance from 'Rush' and the closely related hybrid 'Yoder #5', and placed the resistance locus on linkage group 7. Molnar and Capik (2012b) reported segregation data from bi-parental crosses between eight advanced-generation *C. americana* × *C. avellana* hybrids of Rutter origin and susceptible *C. avellana*. Little to no resistance or tolerance was recovered in seven of these progenies, which was surprising as five of the hybrid parents had been completely resistant to EFB at the experimental site. In the eighth progeny, however, about 50% of the seedlings were rated tolerant and 15% were free of disease. A similar response pattern was observed in three of the six progenies derived from wild *C. americana* crossed to susceptible *C. avellana*, while seedlings in the other three showed little tolerance (MOLNAR AND CAPIK, 2012b). These studies and the results of past breeding efforts demonstrate that EFB resistance from *C. americana* can be accessed through interspecific hybridization. However, the species transmits resistance to its interspecific offspring variably, with both major and minor genes for resistance and tolerance likely present in the germplasm pool. Furthermore, the disease response of the *C. americana* parent may not indicate its ability to transmit resistance or tolerance to its progeny.

The current collections of the NCGR and OSU contain a total of approximately 80 *C. americana* accessions. These accessions were derived primarily from wild seed procured by S.A. Mehlenbacher in the late 1980s with help from members of the Northern Nut Growers Association (NNGA). From among several hundred seedlings, selections were made based on geographic origin, nut characteristics, and a reduced tendency to biennial bearing. Based on genetic characterization of the collection using

microsatellite markers, SATHUVALLI AND MEHLENBACHER (2012) described these accessions as highly diverse and valuable as initial germplasm for breeding. However, given the extensive native range of *C. americana*, this collection does not represent the full extent of the genetic and trait diversity present within the species, and further collection is warranted.

The lack of understanding regarding the transmission of EFB resistance combined with poor nut quality and other negative production traits of *C. americana* complicates the process of developing hybrid cultivars that are competitive with *C. avellana* with respect to nut quality. Fortunately, intraspecific hybridization became a viable strategy for developing disease-resistant cultivars following the identification of EFB-resistant *C. avellana* 'Gasaway' in the 1970s, which was found to be heterozygous for a dominant allele at a single resistance locus (MEHLENBACHER et al., 1991). While 'Gasaway', an obsolete late-shedding pollenizer, has many horticultural deficiencies, it was subsequently used to develop several EFB-resistant cultivars ('Santiam', 'Yamhill', 'Jefferson', 'Dorris', 'Wepster', and 'McDonald') (MEHLENBACHER, 2018; MEHLENBACHER et al., 2007, 2009, 2011, 2013, 2014, 2016). These cultivars have supported a landmark expansion of the Oregon hazelnut industry, which grew from 11700 ha to greater than 31000 ha from 2009 to 2018 (N. Wiman, personal communication).

Concerns over the long-term durability of a single source of resistance as well as potential pathogen diversity led breeders to concurrently assemble and screen vast germplasm collections for EFB response. Initial screens of collected germplasm identified additional sources of resistance (CHEN et al., 2007; COYNE et al., 1998; LUNDE et al., 2000), and linkage mapping confirmed monogenic resistance from multiple sources (e.g., 'Ratoli' and 'Culplà' from Spain, OSU 759.010 from Georgia, OSU 495.027 from Russia, OSU 408.040 from Minnesota, USA, 'Crvenje' and 'Uebov' from Serbia) (BHATTARAI et al., 2017b; CHEN et al., 2005; COLBURN et al., 2015; SATHUVALLI et al., 2011a, 2011b). Resistance was mapped to three different linkage groups (LGs), although most sources map to a cluster on LG 6 in the region near the 'Gasaway' *R* gene (SATHUVALLI et al., 2014). These selections are now being used by OSU and Rutgers University in further breeding and gene pyramiding schemes in an attempt to improve the durability of resistance (MOLNAR et al., 2018; SATHUVALLI et al., 2014).

Continued screening efforts by both OSU and Rutgers University have identified EFB-resistant *C. avellana* from Estonia, Latvia, Lithuania, Poland, Russia, Crimea, Georgia, Moldova, and Turkey (CAPIK AND MOLNAR, 2012; CAPIK et al., 2013; LEADBETTER et al., 2016; MOLNAR et al., 2018; MOLNAR et al., 2007; MUEHLBAUER et al., 2014). The continued discovery of EFB resistance in *C. americana* is additive to the work previously completed for *C. avellana* and may expand, through interspecific hybridization, potential for enhanced climatic adaptability to the more severe summer and winter conditions of the Midwest U.S.

In this study, a diverse collection of 29 *C. americana* accessions were evaluated as donor parents of EFB resistance in interspecific hybridization with EFB-susceptible *C. avellana*. Additionally, an extensive new *C. americana* germplasm collection was assembled to expand the genetic base of the material available for breeding. The F_1 hybrids and new *C. americana* plant materials were established in the field in New Jersey, exposed to high EFB pressure, and assessed for their response to disease after seven growing seasons.

2.3 MATERIALS AND METHODS

Plant materials and culture

Corylus americana × *C. avellana* F_1 progenies. Twenty-nine *C. americana* accessions were crossed with EFB-susceptible *C. avellana* to examine transmission of resistance to their F_1 offspring (Table 2.1, Table A.2.). Crosses were made from 2009 to 2011 following methods described in Mehlenbacher (1994). Twelve interspecific crosses were made at Rutgers and the remainder at OSU. A total of 2031 plants representing 31 different progenies were evaluated in the field. The *C. americana* parents originated from the seed collection made by S. Mehlenbacher (described earlier) and are held in the collections at OSU and the USDA-ARS-NCGR, with a subset of grafted trees held at Rutgers University. Twenty-two of the 29 *C. americana* parents were evaluated and rated free of disease in NJ (as noted in Table 2.1; CAPIK AND MOLNAR, 2012), with the remaining parents only tested in Oregon where they have shown no signs or symptoms of EFB (S.A. Mehlenbacher, unpublished data). The *C. avellana* parents were selected based

on their known susceptibility to EFB and improved nut and kernel characteristics with the aim of producing F_1 offspring that would segregate for disease response as well as improved production traits. One *C. avellana* × *C. avellana* cross was included as a control (OSU 11041) and was expected to segregate for quantitative resistance/tolerance.

The resulting hybrid seeds were collected in mid-Aug. of each year and kept in cold storage until October when they were placed in moist peat moss and stratified at 4 °C until Mar. of the following year. Seeds were germinated in the greenhouse (24 °C day/18 °C night with 16-h day length) in wooden planting boxes (61×91×15 cm) containing a peat-based medium. The seedlings were transplanted after 6 weeks into 2.8 L containers using the same media and top-dressed with 5 g of slow-release fertilizer (Osmocote Plus 15N-9P-12K with micronutrients, 5 to 6 months; The Scotts Co., Marysville, OH). Plants were moved outdoors in late May for acclimation under shade cloth (40% shade) until field planting in Oct. of the same year. Tree spacing was ~1.0 m within the row by ~3.5 m between rows. Plants from each progeny were planted consecutively in rows at the Rutgers University Horticultural Farm 1 in New Brunswick, NJ, and the Cream Ridge Fruit Research and Extension Station, Cream Ridge, NJ. Weed control, irrigation, and fertilizer were provided as needed but no chemical control of pests or diseases was applied. The seedling bushes were not pruned.

Corylus americana **germplasm**

A total of 1335 *C. americana* seedlings from 122 seed lots representing 22 states and 72 counties was procured. The collection effort was initiated and assembled through the help of partners, colleagues, and the interested public, especially members of the Northern Nut Growers Association and the Arbor Day Foundation (Molnar et al., 2018b). Open-pollinated seed was obtained from many locations across the native range of the species and sent to Rutgers for germination and subsequent field planting. The collection locations closely reflect the native range of the species (Figure 2.1). A subset of the *C. americana* collection that included 616 bushes from 62 seed lots was evaluated in this study. This subset included all bushes of the collection that had been under long-term disease pressure for 6 or more years and thus clearly expressed

mature disease phenotypes. These were trees from seeds obtained prior to 2011, which were planted at the Cream Ridge Fruit Research and Extension Station in 2012 under conditions similar to those described above for the interspecific hybrid seedlings in rows adjacent to a majority of them, thus providing a direct basis for comparison.

Exposure to eastern filbert blight

All interspecific hybrid seedlings and the subset of *C. americana* germplasm were exposed to EFB through field inoculations. Hazelnut stems harboring EFB stromata were tied into the canopy of every fifth tree in Apr. at the time of leaf budbreak for the first three years following planting (MOLNAR et al., 2007). The hazelnut stems used as inoculum were collected from infected trees growing in Rutgers field plots. Disease pressure was also provided by natural spread from adjacent breeding nurseries and experimental plots harboring hundreds of infected plants. Additionally, the susceptible seedlings within this planting added to the amount of inoculum as the study progressed.

Evaluation of disease response

Disease ratings followed the six-point rating index developed by PINKERTON et al., (1992): 0 indicates no visible EFB; 1 indicates a single canker or sunken lesion; 2 indicates multiple cankers on a single branch; 3 indicates a tree with several cankered branches; 4 indicates greater than 50% of the tree's branches have cankers; and 5 indicates that all branches contain cankers, except for the basal sprouts. Disease ratings were collected in Jan. 2017 and again in 2018 following the plants' sixth, seventh, or eighth growing season. Previous studies have demonstrated that under field inoculations, five growing seasons provide sufficient time to observe plant phenotypes while minimizing escapes (CAPIK AND MOLNAR, 2012; LEABETTER et al., 2015, 2016; MOLNAR et al., 2007, 2009).

To contextualize the EFB ratings, rating 0 is considered complete resistance, and ratings 1 and 2 are considered highly tolerant as the canker growth does not typically cause abnormal tree growth or reduced nut production. Rating 3 is considered tolerant, with branch dieback and yield reduction over time.

Ratings 4 and 5 are for susceptible individuals where yield reduction occurs soon following cankering and perennial spread of cankers leads to plant death in five to seven years.

Disease response ratings were assigned to each seedling, and the means tabulated for each progeny (Table 2.3). The statistical significance of differences among progeny means were ascertained through a one-way analysis of variance using the AOV function in R 3.3.2 (R Core TEAM, 2014). A post hoc multiple mean comparison was then conducted using the Tukey-Kramer test and the HSD.test function. The disease class scores were presented in histograms indicating percentages of the total for each progeny to aid visualization and comparisons.

2.4 RESULTS AND DISCUSSION

Corylus americana × *C. avellana* F_1 progenies

Useful resistance and tolerance was transmitted to seedlings in a majority of the 31 progenies. Across the entire study, 24% of the offspring were rated free of EFB, and 45% exhibited a degree of tolerance (i.e., rating 1-3). The remaining offspring were susceptible with a rating of 4 or 5, accounting for a respective 20% and 10% of the F_1 bushes. However, the rate at which resistance and/or tolerance was transmitted varied widely by *C. americana* parent and across a full spectrum from high transmission to almost none (Figure 2.2). Disease response ratings, progeny means (Table 2.3), and the percentages of seedlings for each disease rating per progeny are shown (Figure 2.2). This wide variation in transmission of resistance indicates that the disease phenotype of the *C. americana* parent, most of which were evaluated in New Jersey under high disease pressure and found free of EFB (CAPIK AND MOLNAR, 2012), is not a clear indicator of progeny performance. Progeny tests are necessary to evaluate a parent's breeding utility. Results also suggest that the *C. avellana* parent can play a role in the disease response of the progeny.

Twenty-seven progenies retained some level of resistance and/or tolerance while four progenies were effectively lost to EFB (Table 2.3). Ratings of 0 and 3 were the most prevalent in the progenies, representing 26% and 33% of the seedlings, respectively, and a continuum of ratings from 0 to 5 was

recorded in most progenies. Upon closer inspection, we noted three patterns for the disease rating distributions: continuous, bimodal, and no transmission (Figure 2.3).

The first pattern is a continuous distribution, where a major intermediate peak was observed at a single rating within a more or less continuous distribution across the classes. The distribution reflected the "bell-shaped curve" typically observed for segregation of quantitative traits. Thirteen of the progenies exhibited a major peak at rating 3 or across ratings 3 and 4, suggesting that the respective *C. americana* parents carry quantitative (multi-genic) resistance. The trait is partially recovered in the F_1 progeny in the form of tolerance (rating 3), which manifests itself as a high frequency of plants with intermediate phenotypes between the completely resistant *C. americana* and susceptible *C. avellana* parents. This pattern fits the previous descriptions of EFB on *C. americana* as an occasional occurrence of small cankers and general ability to abate perennial canker spread following infection (CAPIK AND MOLNAR, 2012; FULLER, 1908; WESCHCKE, 1953). The disease class with the highest frequency varied in the progenies with continuous distributions. For eight progenies, the peak occurred exclusively at rating 3, while the remaining five progenies of this distribution type held similarly high frequencies of progeny across rating 3 and 4. As mentioned above, rating 4 is representative of a much higher degree of EFB (not suitable for reliable nut production) compared to rating 3, and thus crossing this "ratings threshold" likely represents a loss of loci required for adequate quantitative resistance. Nevertheless, in clonal crops, individuals are selected, and even when ratings 3 or 4 were most prevalent, the progenies still yielded some seedlings with ratings 0, 1, or 2, which would be the target for further breeding or evaluation.

Interestingly, the choice of susceptible *C. avellana* parent appears to also influence the degree to which resistance is recovered, as previously observed by Molnar and Capik (2012b) and Muehlbauer et al. (2018). Here, the influence can be seen by comparing progenies OSU 09044 and Rutgers 11530 (Figure 2.2), which both used the *C. americana* PA OSU 533.069 as the pistillate parent. The male parents of OSU 09044, which displayed a major peak across ratings 3 and 4, come from a *C. avellana* pollen mix (Table 2.3). The male parent of Rutgers 11530, with a peak at rating 3, is *C. avellana* 'Tonda di Giffoni', a cultivar shown to be somewhat tolerant to EFB in prior studies in Oregon (i.e., equivalent to rating 3) and New Jersey (i.e.,

equivalent to rating 4) (CAPIK AND MOLNAR, 2012; MEHLENBACHER et al., 2000, 2001, 2008; PINKERTON et al., 1993). This comparison indicates that the *C. avellana* parent can contribute to the quantitative resistance and disease response of interspecific hybrids.

A bimodal distribution type was observed in 14 progenies, which exhibited both a high frequency of EFB-free (rating 0) and tolerant individuals with a major peak at rating 3. Compiled histograms of these progenies are displayed (Figure 2.4). Tolerance (ratings 1-3) was obtained at similar frequencies to the previously described progenies exhibiting continuous distributions (Table 2.3). However, the frequency of plants exhibiting resistance in these progenies was much higher and resulted in a defined peak at rating 0, which is unlike what would be expected if disease response were under strict multigenic control. The frequency of resistant offspring was higher than expected. This distribution type is generally to be expected if a dominant allele (or two) was segregating in a background of quantitative resistance. *Corylus americana* and *A. anomala* share an extended co-evolutionary history, in where long-term disease pressure often gives rise to a complex pathosystem with a variety of *R* genes and corresponding pathogen effectors (PETIT-HOUDENOT AND FUDAL, 2017). In such systems, the co-segregation of two or more *R* genes (e.g., two-gene epistatic model) is common, and both alleles must be recovered to obtain a rating of 0. Two progenies with a similar bimodal distribution were previously observed by Molnar and Capik (2012b). However, while the regularity of this pattern of transmission had not yet been observed from a broad collection of *C. americana*, a single dominant gene for resistance has recently been mapped to LG 7 in *C. americana* 'Rush' (BHATTARAI et al., 2017a). The frequent observation of this bimodal pattern suggests a promising outlook for the use of *C. americana* in systematic crossing by showing that amenable resistance is available within a geographically diverse collection of *C. americana* accessions.

These bimodal segregation patterns highlight parental genotypes that may be preferential for future breeding. The corresponding *C. americana* parents return a greater number of EFB-free and tolerant offspring and thus increase the opportunity to identify rare recombinant individuals during backcross generations. These accessions represent seedlings from nine different states, stretching the geographic range from ND to NJ. Three accessions [OSU 532.076 (MI), OSU 557.125 (WI), and OSU 531.043 (ND)]

(Table 2.2) appear the most promising as donor parents as their disease rating means are significantly lower than a majority of the other bimodal progenies (Table 2.3).

In four progenies, plus the *C. avellana* × *C. avellana* control progeny (OSU 11041), resistance and tolerance were recovered in few seedlings. The distribution of disease ratings for these progenies (Figure 2.2) shows peaks at rating 4 or across ratings 4 and 5 (average progeny mean of 4.45). This segregation pattern suggests that quantitative resistance in these *C. americana* accessions is highly polygenic. While this pattern characterizes only a small group of the tested progenies, it highlights the extent to which the recovery of *C. americana's* quantitative resistance can vary and emphasizes the need for such characterization of *C. americana* in testcrosses prior to substantial investment in their use as breeding parents.

Across our study, nine progenies used the pollen parent 'Tonda di Giffoni' (Table 2.2), allowing separation of the *C. americana* female parents' contribution to the disease phenotypes of the half-sib progenies. For example, Rutgers 11545 exhibits a continuous distribution with a progeny mean of 3.03 while Rutgers 11547 exhibits a bimodal pattern with a progeny mean of 0.96. Disease incidence in the remaining seven *C. americana* × 'Tonda di Giffoni' varies between *C. americana* parents, distinguishing the various contributions from the *C. americana* parents. Significantly lower progeny means were observed for some (but not all) of the *C. americana* × 'Tonda di Giffoni' crosses compared to other progenies in their respective distribution classes (Table 2.2). The contribution of 'Tonda' di Giffoni' to resistance and tolerance becomes clearer when comparing the progeny means of pooled disease ratings across all *C. americana* × 'Tonda di Giffoni' progenies to that of the other progenies. For continuous distributions, pooled progenies of 'Tonda di Giffoni' have a mean rating of 2.59 compared to a mean of 3.04 for the pooled ratings of all other progenies. The lower mean of 'Tonda di Giffoni' progenies is primarily due to a 12% increase in tolerant offspring (rating 1-3) and a 20% reduction in susceptible offspring. For the bimodal distributions, pooled means of 'Tonda di Giffoni' progenies were again lower (1.26) than that of other progenies (2.26). Quite interestingly, this difference is due to a 23% increase in EFB-free progeny in the 'Tonda di Giffoni' crosses and a corresponding decrease in the quantity of susceptible progeny; an equal

amount of tolerant progeny (42%) were obtained in both groups. Taken together, the 'Tonda di Giffoni' progenies are additional evidence that the *C. avellana* parent can contribute to the transmission of resistance and tolerance from *C. americana* to hybrid offspring, and furthermore, these bimodal progenies provide evidence for the hypothesis of interplay between qualitative and quantitative pathways contributed by both *C. americana* and *C. avellana*. This also indicates value in selecting *C. avellana* genotypes known to express a high level of tolerance, such as 'Sacajawea' (Mehlenbacher et al., 2008), for use in future interspecific hybrid crosses.

There are several underlying mechanisms of resistance that could allow for such high levels of complete resistance without clear goodness-of-fit to Mendelian segregation patterns. The combination of multiple quantitative trait loci can give an additive, strong quantitative defense response to confer complete resistance (NIKS et al., 2015), which was observed with *Xanthomonas campestris* in tomato (STALL et al., 2009). In some cases, highly effective basal resistance can also yield complete resistance (DANGL AND JONES, 2001; NIKS et al., 2015). Pattern recognition receptor (e.g., receptor-like kinases) signaling of pathogen-associated molecular pattern (PAMP)-trigger immunity with segregating downstream pathways is another possible explanation for this distribution pattern (CORWIN AND KLIEBENSTEIN, 2017; POLAND et al., 2009). It is also possible that *R* genes of the *Anisogramma-Corylus* pathosystem behave outside of the traditional gene-for-gene model that incites effector triggered immunity (ETI) in which complete resistance is the consequence of a rapid hypersensitive response (i.e., localized cell death) following the recognition of a pathogen effector by a single *R* gene (i.e., a pattern recognition receptor) (COLL et al., 2011; JONES AND DANGL, 2006; ZHOU et al., 2017).

While bimodal distributions may, or may not, reflect the presence of major *R* genes in our study, segregation distortion has been observed in many crosses of susceptible with EFB-resistant selections, with resistance conferred by a dominant alleles in heterozygous state (MEHLENBACHER, 2018). The percentage of resistant offspring recovered in these progenies ranged from 20% to 75% (BHATTARAI et al., 2017b; COLBURN et al., 2015; LUNDE et al., 2006; SATHUVALLI et al., 2011b). While the cause of distortion is not clear, it is hypothesized that reciprocal translocations are involved. Reciprocal translocations occur

commonly in prevalent cultivars (SALESSES AND BONNET, 1988; TORELLO MARINONI et al., 2018) and could lead to chromosomal rearrangements in gametic cells that distort segregation in either direction. Other researchers suggest this deviation is caused by "modifying" or transcription factors that must be co-inherited with the major gene, and when absent, minor EFB infections (i.e., rating 1 and 2) can occur despite the presence of an *R* gene (MUEHLBAUER et al., 2018). Similar segregation distortion has been observed with apple scab (CROSBY et al., 1990; GESSLER AND PERTOT, 2012) as well as other pathosystems involving biotrophic ascomycete pathogens like *Zymoseptoria tritici* (CHARTRAIN et al., 2005; SAINTENAC et al., 2018), *Leptosphaeria maculans* (PARLANGE et al., 2009), and *Verticillium dahliae* (CASTROVERDE et al., 2017; HAYES et al., 2011). Such distortions can complicate the interpretation of bimodal inheritance from new *Corylus* germplasm based upon phenotyping and present an added challenge in readily identifying new *R* genes, although DNA markers have been very useful to deconvolute segregation (TORELLO MARINONI et al., 2018). Nevertheless, the patterns provide a basis for interplay between qualitative and quantitative pathways and insight as to the possible models for resistance.

Corylus americana germplasm collection

The collection of 1,335 bushes representing 122 seed lots originates from 72 counties and 22 states (Table 2.2), and its distribution is depicted by the geographic information systems map (Figure 2.1). Previously collected germplasm preserved in the USDA-ARS-NCGR and university holdings, in addition to herbarium records of wild *C. americana*, are indicated to show the extent to which this new collection complements them. A subset of 616 bushes underwent long-term evaluation for EFB response and disease ratings are reported by seed lot (Table A.2.). At the final evaluation, 587 (95%) of these trees remained free of disease symptoms (i.e., rating 0). The remaining 29 trees were represented by ratings 1 through 5, where 20 individuals were tolerant (rating 1-3) and the remaining 9 susceptible (rating 4-5). Disease-free trees were present in all 62 seed lots. The evaluated plots of *C. americana* were directly adjacent to a majority of the hybrid seedlings discussed subsequently (those planted in 2012) and provide a stark contrast

of disease infection under similar conditions. The multiple years of exposure reduce the likelihood that trees free of EFB simply escaped infection.

Given that the collection was derived from open-pollinated seeds from a broad geographic range, the results validate that *C. americana* populations carry a high level of innate resistance to EFB. These results also corroborate prior evaluation of clonal *C. americana* in New Jersey, where 51 accessions from the OSU and NGCR collections were exposed to EFB and 49 remained free of disease at the conclusion of the study (CAPIK AND MOLNAR, 2012). Some of the 49 clones that remained free of EFB were used to develop the interspecific hybrid progenies evaluated in this current study. Further, the collection reported here is a robust expansion of available *C. americana* germplasm (Figure 2.1). *Corylus americana* is a highly heterozygous, obligate outcrossing species that is wind-pollinated and sporophytically self-incompatible (DEMCHIK et al., 2018; SATHUVALLI AND MEHLENBACHER, 2012). As a result, seeds of a wild plant are typically derived from a multitude of pollen parents. The reproductive biology of *C. americana*, coupled with the wide geographic range represented in this collection and the high genetic variation in the pathogen (MUEHLBAUER et al., 2018), indicate that this collection holds plants representing diverse sources of genetic resistance to EFB.

2.5 CONCLUSIONS

Our study confirms *C. americana* as an abundant source of EFB resistance or tolerance and demonstrates the recovery of resistance or tolerance in interspecific F_1 offspring from a wide diversity of parents. While EFB-free individuals were common in the studied germplasm collection as well as the parents used to make the crosses, the transmission of resistance to interspecific progeny was variable and at times absent. Thus, transmission from specific *C. americana* parents was not predictable based on its phenotype, and some form of test crossing is required to identify the most promising genotypes for use in breeding. Due to the long maturity times of hazelnut where test crosses are not always feasible, a practical approach may be to make many crosses using a diversity of parents, select the best offspring, and move forward expecting that a portion of the offspring or entire progenies (in some cases) won't be useful.

Continuous and bimodal distributions were commonly observed in the interspecific hybrid progenies, while no transmission was observed in four interspecific crosses. In the continuous distribution, bell-shaped curves indicative of quantitative resistance were observed, with the prominent disease rating intermediate between the parental phenotypes. These continuous distributions occurred with progeny means ranging from 2.38 to 3.60. The bimodal distributions similarly displayed this bell-shaped continuum, suggesting a background presence of quantitative resistance. However, these 14 progenies also had a higher frequency of offspring rated 0 (ranging between 20% to 61%), which suggests the frequent presence of major genes. Additionally, the *C. avellana* parents were observed to influence transmission, as seen when comparing progenies OSU 09047, Rutgers 11529, and the 'Tonda di Giffoni' related crosses, in general, which demonstrates the importance of the *C. avellana* parents chosen to breed advanced generation hybrids.

Our results support the continued use of selected *C. americana* as sources of EFB resistance in systematic breeding. In particular, the *C. americana* parents of the bimodal distributions represent a promising expansion to the *C. americana* selections that carry EFB resistance. Further, the accessions OSU 532.076, OSU 557.125, and OSU 531.043 appear to be superior contributors of both EFB resistance and tolerance and will be targeted for breeding with data provided to the NCGR database. Although the genetic control and mechanisms of resistance within these accessions are not yet clear, and F_2 and backcross hybrids need to also be evaluated in future studies, these parent materials represent a newly identified pool of potential *R* genes from *C. americana.* This premise is bolstered given their wide geographic origins, and the fact that only one source of EFB resistance from *C. americana* has been mapped to date (BHATTARAI et al., 2017a).

This study represents a significant step towards better informed utilization of *C. americana* in hybrid hazelnut breeding and points to a need for additional research to elucidate the genetic control and mechanisms of resistance. The results of the F_1 progenies demonstrate that a wider collection of wild *C. americana* accessions should be tested as donor parents, which can now be pursued further using the collection reported here. A greater diversity of *C. avellana* parents with known phenotypes should be systematically incorporated into such studies to better ascertain the species' effect on inheritance. Bhattarai

et al. (2017a) noted that *C. americana* × *C. avellana* hybrids of the OSU breeding program typically resemble *C. avellana* morphologically by the second backcross generation. Consequently, the inheritance from interspecific F_1 hybrids with known pedigrees should be tested at least until the BC_2 generation to observe the maintenance of resistance. Now that the recovery of resistance from *C. americana* has been demonstrated using a diversity of parents, resources can be dedicated with more confidence toward developing larger progenies during future inheritance studies so that bi-parental progenies are available for subsequent genetic mapping. Additionally, genotypes from this study could be useful in studies of resistance mechanisms, a task that remains challenging because of the uniquely long latent period of asymptomatic growth of the pathogen in which the cellular pattern of hyphae colonization remains elusive. Nevertheless, this study reports new knowledge on EFB resistance from *C. americana* that will be useful in breeding hybrid hazelnuts adapted to the eastern half of North America.

2.6 TABLES AND FIGURES

Table 2.1. Parentage of *C. americana* × *C. avellana* progeny rated for eastern filbert blight (EFB) disease in New Jersey.

Progeny Identification (no.)[z]	Parentage[y,x]
OSU 09041	OSU 403.028 (*C. americana* NE, PI 87145) × OSU Mix A 2009[w]
OSU 09042	OSU 531.016[v] (*C. americana* MI) × OSU Mix B 2009[w]
OSU 09044	OSU 533.069[v] (*C. americana* PA) × OSU Mix B 2009
OSU 09045	OSU 557.075[v] (*C. americana* PA) × OSU Mix B 2009
OSU 09046	OSU 557.102 (*C. americana* WI) × OSU Mix B 2009
OSU 09047	OSU 557.125[v] (*C. americana* WI) × OSU Mix A 2009
OSU 10052	OSU 405.038[v] (*C. americana* NJ) × OSU Mix 2010[v]
OSU 10058	*C. avellana* 'Clark' × OSU 400.033[v] (*C. americana* IN) (CCOR 684.001, PI 617251)
OSU 10059	*C. avellana* 'Clark' × OSU 405.006 (*C. americana* PA)
OSU 11041[u]	OSU 1197.113 × OSU 1155.009
OSU 11050	OSU 405.043[v] (*C. americana* NJ) × OSU Mix 2011
OSU 11051	OSU 532.082 (*C. americana* MI) × OSU Mix 2011
OSU 11052	OSU 588.044[v] (*C. americana* IL) × OSU Mix 2011
OSU 11053	CCOR 857 (*C. americana* IL) × OSU Mix 2011
OSU 11055	*C. avellana* 'Clark' × OSU 400.042 (*C. americana* WI)
OSU 11058	*C. avellana* 'Clark' × OSU 531.037[v] (*C. americana* WI) (CCOR 676.001, PI 617243)
OSU 11059	*C. avellana* 'Clark' × OSU 531.043[v] (*C. americana* ND) (CCOR 677.001, PI 617244)

Table 2.1. (cont.d)

OSU 11060	*C. avellana* 'Clark' × OSU 532.048 (*C. americana* KY) (CCOR 680.001, PI 617248)
OSU 11061	*C. avellana* 'Clark' × CCOR 59.001 [v] (*C. americana* MS) (PI 433984)
OSU 11062	*C. avellana* 'Clark' × CCOR 847 (*C. americana* IL) (PI 641150)
Rutgers 11525	OSU 403.040[v] (*C. americana* NE) × *C. avellana* 'Tonda di Giffoni'
Rutgers 11529	OSU 557.125[v] (*C. americana* WI) × *C. avellana* 'Tonda di Giffoni'
Rutgers 11530	OSU 533.069[v] (*C. americana* PA) × *C. avellana* 'Tonda di Giffoni'
Rutgers 11533	OSU 366.060[v] (*C. americana* MS) (CCOR 59.002, PI 433984) × *C. avellana* 'Tonda Gentile delle Langhe' (PI 557035, CCOR 31.001)
Rutgers 11534	OSU 557.026[v] (*C. americana* VA) × *C. avellana* 'Tonda Gentile delle Langhe'
Rutgers 11535	OSU 531.027[v] (*C. americana* IN) × *C. avellana* 'Tonda Romana'
Rutgers 11540	OSU 557.122[v] (*C. americana* WI) (CCOR 710.001, PI 617273) × *C. avellana* 'Tonda di Giffoni'
Rutgers 11541	OSU 531.043[v] (*C. americana* ND) (CCOR 677.001, PI 617244) × *C. avellana* 'Tonda di Giffoni'
Rutgers 11543	OSU 557.153[v] (*C. americana* WI) (CCOR 713.001, PI 617276) × *C. avellana* 'Tonda di Giffoni'
Rutgers 11545	CCOR 507.001[v] (*C. americana* MN) (PI 557023) × *C. avellana* 'Tonda di Giffoni'
Rutgers 11547	OSU 532.076[v] (*C. americana* MI) (CCOR 682.001, PI 617249) × *C. avellana* 'Tonda di Giffoni'
Rutgers 11550	OSU 405.047[v] (*C. americana* MN) (CCOR 694.001, PI 617261) × *C. avellana* 'Tonda di Giffoni'

[z]Oregon State University (OSU), Corvallis, OR; Rutgers University, New Brunswick, NJ; Controlled crosses were made in the year indicated by the first two digits of the progeny identification number.

Table 2.1. (cont.d)

[y](*C. americana* State) denotes the state in which seeds of the *C. americana* parent were collected.

[x]The OSU selection ID number (i.e., OSU 403.028) represents the location (row.tree) at the OSU Smith Horticulture Research Farm, Corvallis, OR.

[w]*C. avellana* pollen mixtures were comprised of three OSU EFB-susceptible breeding selections to ensure compatibility with the *C. americana* accessions that at the time carried unknown (S) alleles. The 2009A pollen mixture consisted of one-third each of the following selections (S alleles are listed with the dominant allele underlined): 'Sacajawea' (1 22), OSU 786.091 (2 4), and OSU 806.051 (8 19). The 2009B pollen mixture consisted of one-third each of the following selections: OSU 1039.010 (15 21), OSU 1051.038 (2 14), and OSU 1033.068 (4 8). The 2010 pollen mixture consisted of one-third each of the following selections: OSU 995.042 (2 3), OSU 1156.105 (8 10) and OSU 1158.109 (22 25). The 2011 pollen mixture consisted of one-third each of the following selections: OSU 1156.105 (8 10), OSU 1213.038 (1 2), and OSU 1224.065 (12 22).

[v]*Corylus americana* parents were rated as free of EFB in New Jersey (Capik and Molnar 2012). Other selections have not been tested in New Jersey but remain free of EFB in Oregon.

[u]A pure *C. avellana* × *C. avellana* control segregating for quantitative resistance. OSU 1197.113 has quantitative resistance, while OSU 1155.009 is susceptible.

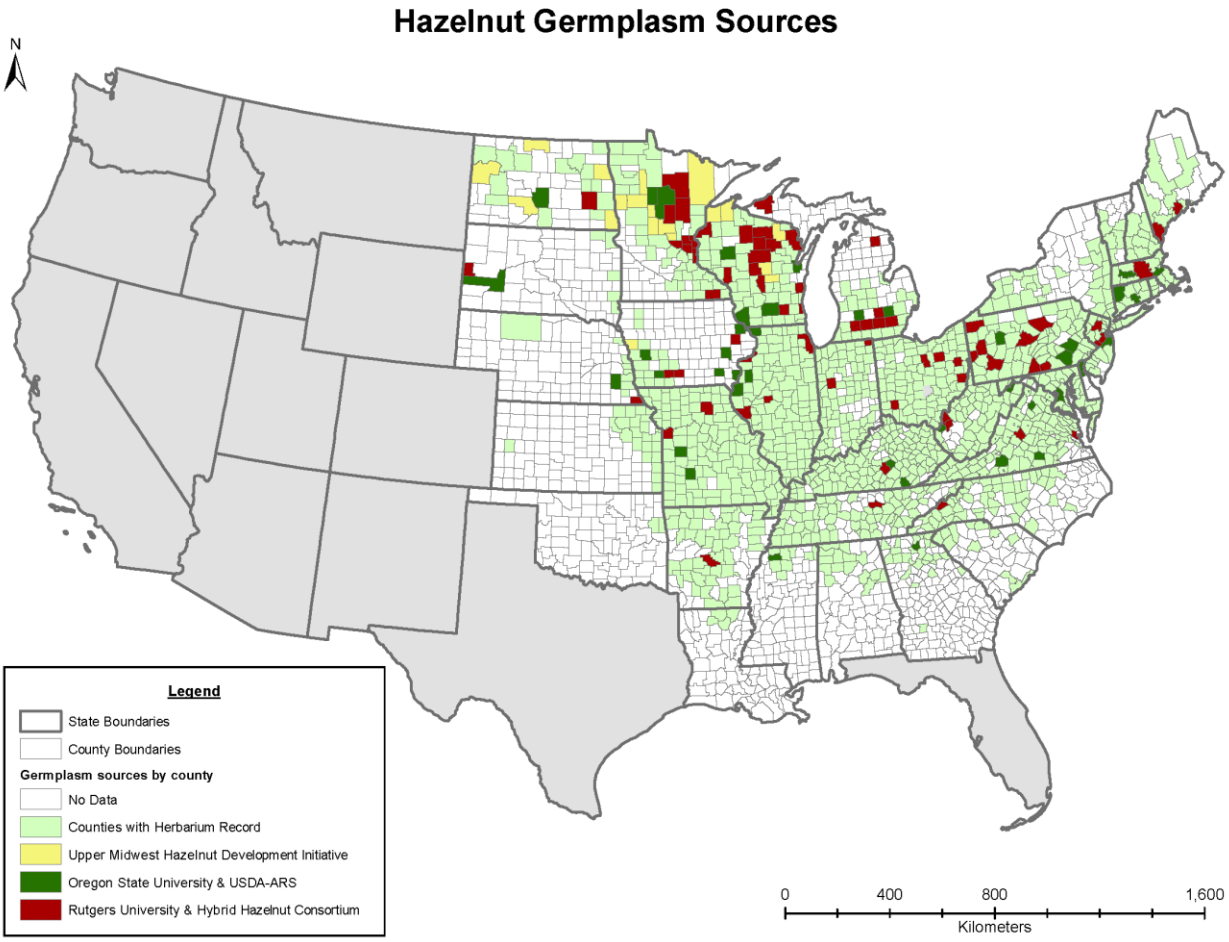


Figure 2.1. Geographic information services map depicting the native distribution of *Corylus americana* and the distribution of the existing university and USDA-ARS collections. Herbarium records compiled from the USDA (plants.usda.gov) are shown in light green. Collections sites of the Upper Midwest Hazelnut Development Initiative (DEMCHIK et al., 2017) are shown in yellow, of Oregon State University and USDA-ARS-NCGR (SATHUVALLI AND MEHLENBACHER, 2012) are shown in green, and the Rutgers University collection is shown in red.

Table 2.2. *Corylus americana* germplasm at the Rutgers University Research and Extension Center in Cream Ridge, NJ, by state and county of origin.

State	County	Seedlings (no.)	Seed lot[z]
AR		1	
	Saline	1	12556
IA		45	
	Clarke	21	11607
	Jones	23	12508
	Union	1	11594
IL		34	
	Cook	11	12533
	Menard	20	11567
	Pike	1	10543
	Rock Island	2	11612
IN		40	
	Steuben	32	14553, 15563
	Tippecanoe	8	12560
KY		30	
	Casey	30	11568
MA		22	
	Nantucket	22	10532, 11605, 11606
ME		2	
	Lincoln	2	10520
MI		116	
	Calhoun	7	11602

Table 2.2. (cont.d)

	Ingham	4	14542, 14543, 16518 - 16522
	Jackson	43	14537, 14539, 14540
	Kalamazoo	7	14538
	Montmorency	16	14544
	Ontonagon	16	11588, 12540, 16518 -16522
	Washtenaw	23	11587
MN		199	
	Aiken	49	11563, 11571
	Anoka	2	12553
	Chisago	34	11579
	Crow Wing	3	11608
	Fillmore	9	11599
	Isanti	64	14551, 15568, 15569
	Mississippi River	29	11565, 15566
	Sherbune	7	11601
	Washington	2	12509
MO		33	
	Jackson	29	11613
	Macon	4	11583
NC		10	
	Madison	10	12561
ND		56	
	Burleigh	40	12547, 15564
	Barnes	16	15567

Table 2.2. (cont.d)

NE		119	
	Lancaster	1	11600
	Richardson	118	10556 - 10579, 12507
NJ		43	
	Mercer	27	11570
	Middlesex	6	11554
	Morris	10	11578A
OH		207	
	Belmont	23	11569
	Butler	48	15571 - 15577
	Carroll	128	15572 - 15576
	Richland	5	11610
	Warren	0	13552 - 13555
	Wayne	3	12552
		134	
PA	Adams	1	13564, 13565
	Beaver	19	14555, 15570
	Butler	11	11564, 11593
	Centre	17	10525 - 10529
	Crawford	21	11575
	Franklin	25	10537
	Lancaster	4	12550
	Lycoming	31	11574
	Westmore	5	12557

Table 2.2. (cont.d)

SD		2	
	Lawrence	2	11611
TN		11	
	Putnam	11	*C. americana* TN
VA		8	
	Gloucester	3	12548
	Nelson	5	13561
WI		166	
	Adams	23	10542
	Burnett	14	10530
	Dane	20	11591
	Jefferson	3	12543
	Langlade	8	12538
	Marathon	1	12546
	Marinette	27	12510, 12539, 12541, 12551
	Oneida	13	11585, 12535, 12536
	Price	24	11577
	Sheboygan	9	12534
	Trempealeau	2	11581
	Wood	22	11580, 11603
WV		14	
	Mason	5	11614
	Putnam	9	11584

[z]Seed lots received from cooperators were assigned numbers by the Rutgers University breeding program.

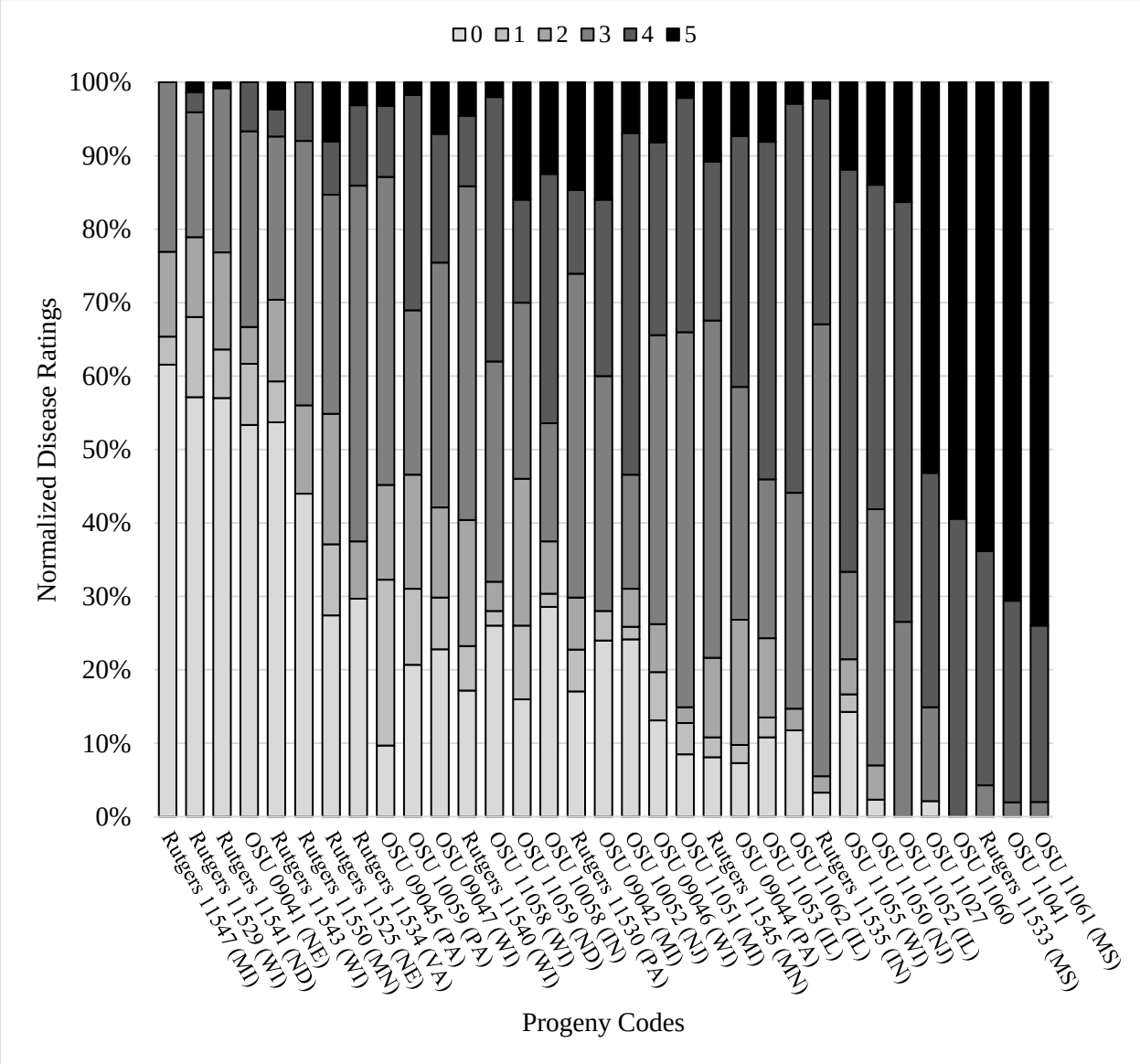


Figure 2.2. A bar chart of disease ratings by progeny, showing the proportion (%) of offspring in each disease category. Disease ratings ranged from 0 (no disease) to 5 (much disease). Plants with ratings of 1-3 were considered tolerant of EFB. Resistance and tolerance of *Corylus americana* is transmitted to *C. americana* × *C. avellana* F_1 hybrids but along a wide spectrum. Most seedlings had a rating of 3 or lower, and EFB-free progeny were common.

Table 2.3. Eastern filbert blight (EFB) disease ratings in *Corylus americana* × *C. avellana* progenies exposed to *Anisogramma anomala* in New Jersey. Individual shrubs were rated on a scale of 0 (no disease) to 5 (disease on all branches).

Progeny[z] (no.) and state	Total no. of plants	Progeny mean[y]	No. seedlings for each disease rating[x,w]					
			0	1	2	3	4	5
Continuous distribution								
OSU 09045 (PA)	31	2.29[de]	3	7	4	13	3	1
Rutgers 11540[t] (WI)	198	2.38[de]	34	12	34	90	19	9
OSU 11059 (ND)	50	2.58[def]	8	5	10	12	7	8
Rutgers 11530[t] (PA)	211	2.71[def]	36	12	15	93	24	31
OSU 09046 (WI)	61	2.84[defg]	8	4	4	24	16	5
OSU 11051 (MI)	47	3.00[efgh]	4	2	1	24	15	1
Rutgers 11545[t] (MN)	37	3.03[efgh]	3	1	4	17	8	4
OSU 09044 (PA)	41	3.05[efgh]	3	1	7	13	14	3
OSU 11053 (IL)	37	3.14[efgh]	4	1	4	8	17	3
OSU 11062 (IL)	34	3.21[fgh]	4	0	1	10	18	1
Rutgers 11535 (IN)	91	3.23[fgh]	3	0	2	56	28	2
OSU 11055 (WI)	42	3.26[fghi]	6	1	2	5	23	5
OSU 11050 (NJ)	43	3.60[ghij]	1	0	2	15	19	6
Bimodal distribution								
Rutgers 11547[t] (MI)	26	0.96[a]	16	1	3	6	0	0
Rutgers 11529[t] (WI)	147	1.01[a]	84	16	16	25	4	2
Rutgers 11541[t] (ND)	121	1.04[a]	69	8	16	27	0	1
OSU 09041 (NE)	60	1.25[ab]	32	5	3	16	4	0

Table 2.3. (cont.d)

Rutgers 11543[t] (WI)	54	1.28[abc]	29	3	6	12	2	2
Rutgers 11550[t] (MN)	25	1.64[abcd]	11	0	3	9	2	0
Rutgers 11525 (NE)	124	2.04[bcd]	34	12	22	37	9	10
Rutgers 11534[t] (VA)	64	2.20[cde]	19	0	5	31	7	2
OSU 10059 (PA)	58	2.34[de]	12	6	9	13	17	1
OSU 09047 (WI)	57	2.37[de]	13	4	7	19	10	4
OSU 11058 (WI)	50	2.54[def]	13	1	2	15	18	1
OSU 10058 (IN)	56	2.63[def]	16	1	4	9	19	7
OSU 09042 (MI)	25	2.76[defg]	6	1	0	8	6	4
OSU 10052 (NJ)	58	2.79[defg]	14	1	3	9	27	4
Low transmission of resistance								
OSU 11052 (IL)	49	3.90[hijk]	0	0	0	13	28	8
OSU 11060 (KY)	37	4.59[jk]	0	0	0	0	15	22
Rutgers 11533 (MS)	47	4.60[jk]	0	0	0	2	15	30
OSU 11061 (MS)	50	4.72[k]	0	0	0	1	12	37
Pooled progeny	2031	2.68	485	105	189	632	406	214
Pooled (%)			0.24	0.05	0.09	0.31	0.20	0.11
C. avellana × *C. avellana*								
OSU 11041[s]	51	4.69[k]	0	0	0	1	14	36

[z]Oregon State University (OSU), Corvallis, OR; Rutgers University, New Brunswick, NJ; Control crosses were made in the year indicated by the first two digits of the progeny identification number.

[y]The same letter following the progeny means indicates a lack of significant difference ($P < 0.05$).

Table 2.3. (cont.d)

[x]Disease was rated in 2018 during the dormant season six to eight years after field establishment and correspond to phenotypes as follows: 0 = no visible EFB, 1 = a single canker, 2 = multiple cankers on a single branch, 3 = multiple branches with cankers, 4 = more than 50% of branches have cankers, and 5 = all branches have cankers or the plant has died from EFB.

[w]Bimodal refers to a distribution that exhibited both a high frequency of EFB-free and tolerant individuals with major peaks occurring at ratings 0 and 3.

[u]Continuous refers to a distribution where a major peak at a single intermediate disease class was most prominent within a more or less continuous distribution among the other classes.

[t]Progenies with 'Tonda di Giffoni' as the pollen parent.

[s]OSU 11041 is a control cross made in Oregon and expected to segregate for quantitative resistance from *C. avellana* based on its parentage. It is a cross of OSU 1197.113 x OSU 1155.009.

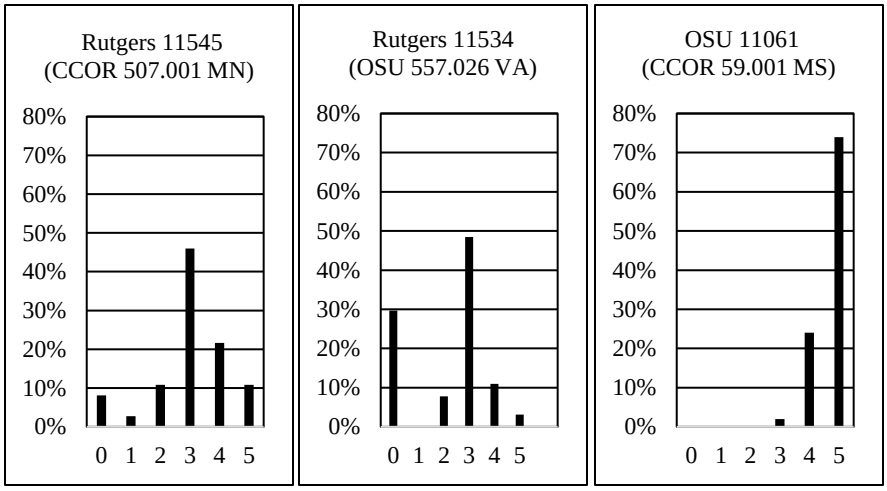


Figure 2.3. Histograms illustrating the three distribution types observed in the *C. americana* × *C. avellana* F_1 progenies, showing the proportion (%) of offspring for each disease rating. The continuous distribution of Rutgers 11545 (left) was observed in 13 progenies where the largest peak occurred at either rating 3 or across ratings 3 and 4. The bimodal distribution of Rutgers 11534 (center) was observed in 14 progenies, with peaks occurring at ratings 0 and 3 or ratings 0 and 3/4. Very little transmission, as in OSU 11061 (right), occurred in four progenies.

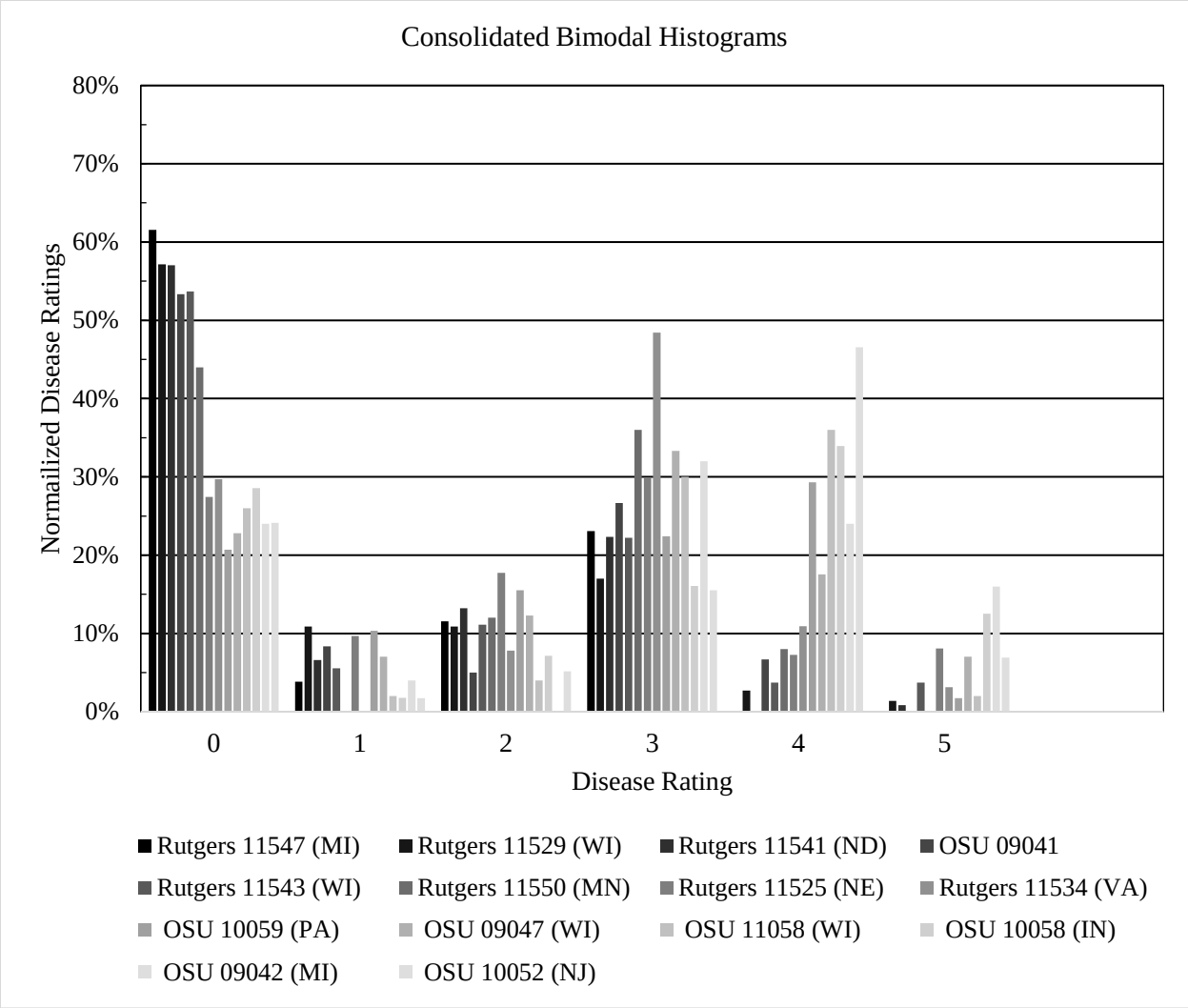


Figure 2.4. Consolidated histograms of the *C. americana* × *C. avellana* F_1 progenies with bimodal distributions. A wide range of percentages of seedlings with complete resistance (rating 0) was observed but without consensus for the segregation of monogenic resistance.

2.7 REFERENCES

Bassil, N., P. Boccacci, R. Botta, J. Postman, and S. Mehlenbachers. 2013. Nuclear and chloroplast microsatellite markers to assess genetic diversity and evolution in hazelnut species, hybrids and cultivars. Genet. Resources and Crop Evol. 60:543-568.

Bhattarai, G., S.A. Mehlenbacher, and D.C. Smiths. 2017a. Eastern filbert blight disease resistance from *Corylus americana* 'Rush' and selection 'Yoder #5' maps to linkage group 7. Tree Genet. and Genomes 13.

Bhattarai, G., S.A. Mehlenbacher, and D.C. Smith. 2017b. Inheritance and linkage mapping of eastern filbert blight disease resistance in 'Uebov' hazelnut. J. Amer. Soc. Hort. Sci. 142:289-297.

Braun, L.C., M.C. Demchik, J.A. Fischbach, K. Turnquist, and A. Kerns. 2018. Yield, quality and genetic diversity of hybrid hazelnut selections in the Upper Midwest of the USA. Agroforestry Systems.

Capik, J.M. and T.J. Molnar. 2012. Assessment of host (*Corylus* sp.) resistance to eastern filbert blight in New Jersey. J. Amer. Soc. Hort. Sci. 137:157-172.

Capik, J.M., M. Muehlbauer, A. Novy, J.A. Honig, and T.J. Molnar. 2013. Eastern filbert blight-resistant hazelnuts from Russia, Ukraine, and Poland. HortScience 48:466-473.

Castroverde, C.D.M., X. Xu, J. Blaya Fernández, R.N. Nazar, and J. Robbs. 2017. Epistatic influence in tomato *Ve1*-mediated resistance. Plant Biol. 19:843-847.

Chartrain, L., P.A. Brading, and J.K.M. Browns. 2005. Presence of the *Stb6* gene for resistance to septoria tritici blotch (*Mycosphaerella graminicola*) in cultivars used in wheat-breeding programmes worldwide. Plant Pathol. 54:134-143.

Chen, H., S.A. Mehlenbacher, and D.C. Smith. 2007. Hazelnut accessions provide new sources of resistance to eastern filbert blight. HortScience 42:466-469.

Chen, H.L., S.A. Mehlenbacher, and D.C. Smiths. 2005. AFLP markers linked to eastern filbert blight resistance from OSU 408.040 hazelnut. J. Amer. Soc. Hort. Sci. 130:412-417.

Colburn, B.C., S.A. Mehlenbacher, V.R. Sathuvalli, and D.C. Smith. 2015. Eastern filbert blight resistance in hazelnut accessions 'Culplà', 'Crvenje', and OSU 495.072. J. Amer. Soc. Hort. Sci. 140:191-200.

Coll, N., P. Epple, and J. Dangl. 2011. Programmed cell death in the plant immune system. Cell Death and Differentiation 18:1247.

Corwin, J.A. and D.J. Kliebenstein. 2017. Quantitative resistance: more than just perception of a pathogen. The Plant Cell:tpc. 00915.02016.

Coyne, C.J., S.A. Mehlenbacher, and D.C. Smith. 1998. Sources of resistance to eastern filbert blight in hazelnut. J. Amer. Soc. Hort. Sci. 123:253-257.

Crosby, J., J. Janick, P. Pecknold, S. Korban, P. O'Connor, S. Ries, J. Goffreda, and A. Voordeckerss. 1990. Breeding apples for scab resistance: 1945–1990. Acta Horticulturae 317:43-70.

Cui, H., K. Tsuda, and J.E. Parker. 2015. Effector-triggered immunity: from pathogen perception to robust defense. Annu. Rev. Plant Biol. 66:487-511.

Dangl, J.L. and J.D.G. Jones. 2001. Plant pathogens and integrated defence responses to infection. Nature 411:826-833.

Demchik, M., A. Kern, L. Braun, J. Fischbach, and K. Turnquist. 2018. Genetic diversity of American hazelnut in the Upper Midwest, USA. Agroforestry Systems 92:1507-1516.

Demchik, M., U.-S. Point, and D. Wyse. 2011. Hazelnut production potential in the Upper Midwest: A report on hybrid hazelnut yields. UW Exten. Res. Bul. 17.

Drumke, J.S. 1964. A systematic survey of *Corylus* in North America. Univ. of Tenn., Knoxville, TN, PhD diss.

Erdogan, V. and S.A. Mehlenbacher. 2000a. Interspecific hybridization in hazelnut (*Corylus*). J. Amer. Soc. Hort. Sci. 125:489-497.

Erdogan, V. and S.A. Mehlenbacher. 2000b. Phylogenetic relationships of *Corylus* species (Betulaceae) based on nuclear ribosomal DNA ITS region and chloroplast matK gene sequences. Systematic Bot. 25:727-737.

FAOSTAT. 2017. Agricultual production, crops, Food and Agriculture Organization of the United Nations, Rome. May 30, 2017. <http://www.fao.org/faostat/en/#data/QC/visualize>.

Fischbach, J. 2017. A production and economic model for hedgerow hazelnut production in the midwestern United States. University of Wisconsin - Extension, Bayfield, WI. Feb. 1, 2017. <www.midwesthazelnuts.org>.

Fuller, A.S. 1908. The nut culturist. Orange Judd, New York, NY.

Gessler, C. and I. Pertot. 2012. *Vf* scab resistance of *Malus*. Trees 26:1-14.

Gleason, H.A. and A. Cronquist. 1963. Manual of vascular plants of northeastern United States and adjacent Canada. Van Nostrand, New York.

Hammond, E. 2006. Identifying superior hybrid hazelnut plants in southeast Nebraska. University of Nebraska-Lincoln, Lincoln, Nebraska, M.S. Thesis.

Hayes, R.J., L.K. McHale, G.E. Vallad, M.J. Truco, R.W. Michelmore, S.J. Klosterman, K. Maruthachalam, and K.V. Subbarao. 2011. The inheritance of resistance to Verticillium wilt caused by race 1 isolates of *Verticillium dahliae* in the lettuce cultivar La Brillante. Theor. Appl. Genet. 123:509-517.

Johnson, K. and J. Pinkerton. 2002. Eastern filbert blight, p. 44-46. In: J.W. Pscheidt (ed.), Compendium of nut crop diseases in temperate zones. St. Paul, MN, APS Press, Amer. Phyopathol. Soc.

Jones, J.D.G. and J.L. Dangl. 2006. The plant immune system. Nature 444:323-329.

Leabetter, C.W., J.M. Capik, S.A. Mehlenbacher, and T.J. Molnar. 2016. Hazelnut accessions from Russia and Crimea transmit resistance to eastern filbert blight. J Amer. Pomol. Soc. 70:92-109.

Leadbetter, C.W., J.M. Capik, M. Pisetta, and T.J. Molnar. 2015. Sources of resistance to eastern filbert blight in hazelnuts from the Republic of Georgia. Scientia Hort. 193:269-275.

Lunde, C.F., S.A. Mehlenbacher, and D.C. Smith. 2000. Survey of hazelnut cultivars for response to eastern filbert blight inoculation. HortScience 35:729-731.

Lunde, C.F., S.A. Mehlenbacher, and D.C. Smith. 2006. Segregation for resistance to eastern filbert blight in progeny of 'Zimmerman' hazelnut. J. Amer. Soc. Hort. Sci. 131:731-737.

Mehlenbacher, S. 2018. Breeding hazelnuts resistant to eastern filbert blight. Acta Hort. 1226:73-78.

Mehlenbacher, S.A. 1991. Hazelnuts (*Corylus*). In: J.N. Moore and J.R. Ballington (eds.). Genetic resources in temperate fruit and nut crops. Acta Hort. 290:789-836.

Mehlenbacher, S.A., A.N. Azarenko, D.C. Smith, and R. McCluskey. 2000. 'Lewis' hazelnut. HortScience 35:314-315.

Mehlenbacher, S.A., A.N. Azarenko, D.C. Smith, and R. McCluskey. 2001. 'Clark' hazelnut. HortScience 36:995-996.

Mehlenbacher, S.A., A.N. Azarenko, D.C. Smith, and R. McCluskey. 2007. ‘Santiam’ hazelnut. HortScience 42:715-717.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2008. ‘Sacajawea’hazelnut. HortScience 43:255 257.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2009. 'Yamhill' hazelnut. HortScience 44:845-847.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2011. 'Jefferson' hazelnut. HortScience 46:662-664.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2013. ‘Dorris’ hazelnut. HortScience 48:796-799.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2014. 'Wepster' hazelnut. HortScience 49:346-349.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2016. 'McDonald' hazelnut. HortScience 51:757-760.

Mehlenbacher, S.A., M.M. Thompson, and H.R. Cameron. 1991. Occurrence and inheritance of resistance to eastern filbert blight in Gasaway hazelnut. HortScience 26:410-411.

Molnar, T., J. Goffreda, and C. Funk. 2005. Developing hazelnuts for the eastern United States. Acta Hort. 686:609-618.

Molnar, T., J. Lombardoni, M. Muehlbauer, J. Honig, S. Mehlenbacher, and J. Capik. 2018. Progress breeding for resistance to eastern filbert blight in the eastern United States. Acta Hort 1226:79-86.

Molnar, T.J. 2011. *Corylus*, p. 15-48. In: C. Kole (ed). Wild crop relatives: Genomic and breeding resources. Springer, Berlin, Heidelberg.

Molnar, T.J. and J. Capik. 2012a. Advances in hazelnut research in North America. Acta Hort. 940:57-65.

Molnar, T.J. and J.M. Capik. 2012b. Eastern filbert blight susceptibility of american × european hazelnut progenies. HortScience 47:1412-1418.

Molnar, T.J., J.M. Capik, and J.C. Goffreda. 2009. Response of hazelnut progenies from known resistant parents to *Anisogramma anomala* in New Jersey, USA. Acta Hort. 845:73-82.

Molnar, T.J., D.E. Zaurov, J.C. Goffreda, and S.A. Mehlenbacher. 2007. Survey of hazelnut germplasm from Russia and Crimea for response to eastern filbert blight. HortScience 42:51-56.

Muehlbauer, M., J.M. Capik, T.J. Molnar, and S.A. Mehlenbacher. 2018. Assessment of the 'Gasaway' source of resistance to eastern filbert blight in New Jersey. Scientia Hort. 235:367-372.

Muehlbauer, M.F., J.A. Honig, J.M. Capik, J.N. Vaiciunas, and T.J. Molnar. 2014. Characterization of eastern filbert blight-resistant hazelnut germplasm using microsatellite markers. J. Amer. Soc. Hort. Sci. 139:399-432.

Niks, R.E., X. Qi, and T.C. Marcel. 2015. Quantitative resistance to biotrophic filamentous plant pathogens: concepts, misconceptions, and mechanisms. Ann. Rev. Phytopathol. 53:445-470.

Parlange, F., G. Daverdin, I. Fudal, M.L. Kuhn, M.H. Balesdent, F. Blaise, B. Grezes-Besset, and T. Rouxel. 2009. *Leptosphaeria maculans* avirulence gene *AvrLm4-7* confers a dual recognition specificity by the *Rlm4* and *Rlm7* resistance genes of oilseed rape, and circumvents *Rlm4*-mediated recognition through a single amino acid change. Mol. Microbiol. 71:851-863.

Petit-Houdenot, Y. and I. Fudal. 2017. Complex interactions between fungal avirulence genes and their corresponding plant resistance genes and consequences for disease resistance management. Frontiers Plant Sci. 8:1072.

Pinkerton, J.N., K.B. Johnson, S.A. Mehlenbacher, and J.W. Pscheidts. 1993. Susceptibility of european hazelnut clones to eastern filbert blight. Plant Dis. 77:261-266.

Pinkerton, J.N., K.B. Johnson, K.M. Theiling, and J.A. Griesbach. 1992. Distribution and characteristics of the eastern filbert blight epidemic in Western Oregon. Plant Dis. 76:1179-1182.

Poland, J.A., P.J. Balint-Kurti, R.J. Wisser, R.C. Pratt, and R.J. Nelson. 2009. Shades of gray: the world of quantitative disease resistance. Trends in Plant Sci. 14:21-29.

Rutter, M. 1991. Variation in resistance to eastern filbert blight in hybrid hazelnuts. Ann. Rpt. Northern Nut Growers Assn. 82:159-162.

Rutter, P. 1987. Badgersett research farm - plantings, projects, and goals. Annu. Rep. Northern Nut Growers Assn. 78:173-186.

Saintenac, C., W.-S. Lee, F. Cambon, J.J. Rudd, R.C. King, W. Marande, S.J. Powers, H. Bergès, A.L. Phillips, C. Uauy, K.E. Hammond-Kosack, T. Langin, and K. Kanyuka. 2018. Wheat receptor-kinase-like protein *Stb6* controls gene-for-gene resistance to fungal pathogen *Zymoseptoria tritici*. Nature Genet. 50:368-374.

Salesses, G. and A. Bonnet. 1988. Étude cytogénétique d'hybrides entre variétés de noisetier (*Corylus avellana*) porteuses d'une translocation à l'état hétérozygote. Cytologia 53:407-413.

Sathuvalli, V.R., H. Chen, S.A. Mehlenbacher, and D.C. Smith. 2011a. DNA markers linked to eastern filbert blight resistance in 'Ratoli' hazelnut (*Corylus avellana* L.). Tree Genet. Genomes 7:337-345.

Sathuvalli, V.R. and S.A. Mehlenbacher. 2012. Characterization of american hazelnut (*Corylus americana*) accessions and *Corylus americana* × *Corylus avellana* hybrids using microsatellite markers. Genet. Resources Crop Evol. 59:1055-1075.

Sathuvalli, V.R., S.A. Mehlenbacher, B.C. Peterschmidt, and D.C. Smith. 2014. Current status of eastern filbert blight resistance sources and mapping. Acta Hort. 1052:27-34.

Sathuvalli, V.R., S.A. Mehlenbacher, and D.C. Smith. 2011b. DNA markers linked to eastern filbert blight resistance from a hazelnut selection from the Republic of Georgia. J. Amer. Soc. Hort. Sci. 136:350-357.

Slate, G.L. 1969. Filberts - incluing varieties grown in the east, p. 287-293. In: R.A. Jaynes (ed.), Handbook of North American nut trees. Northern Nut Growers Assn., Knoxville, TN.

Stall, R.E., J.B. Jones, and G.V. Minsavage. 2009. Durability of resistance in tomato and pepper to *Xanthomonads* causing bacterial spot. Annu. Rev. Phytopathology 47:265-284.

Stotz, H.U., G.K. Mitrousia, P.J.G.M. de Wit, and B.D.L. Fitt. 2014. Effector-triggered defence against apoplastic fungal pathogens. Trends in Plant Sci. 19:491-500.

Team, R.C. 2014. The R project for statistical computing. Available at www. R-project.org

Technavio. 2017. Global hazelnut market 2017-2021. EMIS Intelligence Database.

Torello Marinoni, D., N. Valentini, E. Portis, A. Acquadro, C. Beltramo, S.A. Mehlenbacher, T.C. Mockler, E.R. Rowley, and R. Botta. 2018. High density SNP mapping and QTL analysis for time of leaf budburst in Corylus avellana L. PLOS ONE 13:e0195408.

Thompson, M.M., H.B. Lagerstedt, and S.A. Mehlenbacher. 1996. Hazelnuts. p. 125-184. In: J. Janick and J.N. Moore (eds.). Fruit breeding. Vol. 3. Nuts. John Wiley & Sons, New York.

U.S. Department of Agriculture. 2018. Noncitrus Fruits and Nuts 2017 Summary, National Agricultural Statistics Service. U.S. Dept. Agr., Washington, D.C.

Weschcke, C. 1953. Growing nuts in the north. Webb, St. Paul, MN.

Weschcke, C. 1963. Forty-three years of active work in nutgrowing. Annu. Rep. of the Northern Nut Growers Assn. 54:63-65.

Win, J., A. Chaparro-Garcia, K. Belhaj, D. Saunders, K. Yoshida, S. Dong, S. Schornack, C. Zipfel, S. Robatzek, S. Hogenhout and S. Kamoun. 2012. Effector biology of plant-associated organisms: concepts and perspectives. Cold Spring Harbor Symp. Quant. Biol. 77: 235-247.

Zhou, J.-M., D. Tang, and G. Wang. 2017. Receptor kinases in plant pathogen interactions: more than pattern recognition. The Plant Cell:tpc. 00891.02016.

CHAPTER 3

CHARACTERIZATION OF THE GENETIC DIVERSITY AND STRUCTURE OF AMERICAN HAZELNUT (*CORYLUS AMERICANA*) GERMPLASM

3.1 ABSTRACT

The american hazelnut (*Corylus americana*) is native to a broad range of the eastern United States and Southern Canada, where it is the endemic host of the pathogenic fungus *Anisogramma anomala* – the causal agent of eastern filbert blight (EFB) disease. Initial studies indicate that *C. americana* harbors high genetic diversity as well as durable resistance and tolerance to EFB. While *C. americana* has thick-shelled, tiny nuts not suited for commercial production, it is cross-compatible with the hazelnut species of commercial consequence (*Corylus avellana*) and can serve as a valuable donor of EFB resistance and climate adaptability traits. To-date, however, only a narrow set of *C. americana* parents have been used in interspecific hybrid development and, based upon the vast endemic range of *C. americana*, existing germplasm does not fully represent the genetic diversity of species. In recent years, U.S. hazelnut breeders have expanded the availability of characterized *C. americana* germplasm. Here, we report the genetic diversity and structure of new *C. americana* (272 individuals) collected from 33 seedlots across the species' native range. Two-thousand fifty-three SNPs were discovered using a genome-by-sequencing approach and support a heterozygous collection (H_E = 0.276, H_O = 0.280) with moderate differentiation (F_{ST} = 0.108) and low inbreeding (F_{IS} = -0.136). Bayesian model-based and neighbor-joining (NJ) clustering corroborate an uppermost clustering level of K = 3, with two minimally distant major groups and one smaller, more distant group. The NJ dendrogram depicts many small subgroups equally distant from common ancestry. Discriminant analysis of principal components reveals between-sub-group variation (K = 15) within the NJ dendrogram and allows the identification of 19 consensus subgroups. Fifty-one accessions were selected for inclusion within a core set based upon 95% representation of the observed allelic variation. Breeders can now exploit the breadth of genetic diversity held within this collection to develop interspecific hybrids and expand the cultivated range of hazelnut throughout eastern North America.

3.2 INTRODUCTION

The european hazelnut (*Corylus avellana*) is a high-value tree nut with a growing global market where demand exceeds supply. Turkey continues to produce the lion's share of the global hazelnut supply due to an ideal temperate maritime climate along the Black Sea that has served as a natural habitat for cultivated hazelnut for thousands of years. From 2013 to 2017, Turkey's hazelnut production was roughly two-thirds of the world production at an average of 548,000 t (FAOSTAT, 2017). During this period, Italy was the second largest producer (108,320 t), followed by the U.S. (34,110 t), Azerbaijan (34,088 t), Georgia (31,940 t), China (24,916 t) (FAOSTAT, 2017). Chile, Australia, Iran, France, Spain, and others also produce hazelnut and look to expand their cultivation. Although Turkey is perennially the top producer, its production has oscillated between 420,000 and 800,791 t over the past decade, frequently causing steep international price increases (FAOSTAT, 2017). Reoccurring late spring frosts, aging orchards, and small-scale cultural practices are the primary causes of regular Turkish crop losses (ERDOGAN, 2017). Consequently, new orchards have been established in both traditional and non-traditional cultivation regions as nations seek to increase their market share (MEHLENBACHER, 2018a).

World production of hazelnut is still largely dependent upon local wild selections from traditional Mediterranean-like growing regions (ERDOGAN, 2017). Many geographies where new hazelnut orchards are being planted have climates that are less suited for commercial yielding and lack committed germplasm improvement programs to adapt cultivated selections (MEHLENBACHER, 2018a). The two exceptions are northeastern China and the eastern U.S. where breeders have made hybridizations between *C. avellana* and wild *Corylus* in effort to expand the suitable range of hazelnut cultivation to their regions (MEHLENBACHER, 2018a). Since the late 1990's, Rutgers University (New Brunswick, NJ) has used wild *Corylus* germplasm to breed eastern filbert blight (EFB) resistant selections that meet the quality standards of the global confectionary market (MOLNAR et al., 2018a; MOLNAR AND CAPIK, 2012). *Corylus avellana* from non-traditional cultivation regions is of foremost use as many sources of resistance have been discovered since the initial discovery of the 'Gasaway' gene (MEHLENBACHER, 2018b; MEHLENBACHER et al., 1991; MOLNAR et al., 2018a), and the morphologies of wild *C. avellana* most closely match those of cultivated

selections. *Corylus americana* (the american hazelnut) is also a source of EFB resistance and high tolerance (CAPIK AND MOLNAR, 2012; FULLER, 1908; WESCHCKE, 1953), as the endemic host of the disease's causal organism *Anisogramma anomala* PECK E. MüLLER (MOLNAR et al., 2018b). The development of interspecific hybrids (*C. americana* × *C. avellana*) has been explored in tandem to intraspecific hybrids; however, the current germplasm pool requires improvements to support a viable commercial industry outside of Oregon, where 99% of U.S. hazelnut production occurs (USDA, 2018).

After the discovery of resistance in *C. avellana* 'Gasaway' in the 1970s, intraspecific hybridization between cultivated *C. avellana* and EFB-resistant *C. americana* became a feasible strategy to develop improved EFB-resistant cultivars (MEHLENBACHER, 2018a). The 'Gasaway' cultivar is heterozygous for a single dominant allele (MEHLENBACHER et al., 1991), and despite its horticultural deficiencies, has been the premier donor of resistance used in U.S. cultivar development for the in-shell market – leading to the development of 'Santiam', 'Yamhill', 'Jefferson', 'Dorris', 'Wepster', and 'McDonald' cultivars (MEHLENBACHER et al., 2007, 2009, 2011, 2013, 2014, 2016). The cultivars have supported an industry expansion in Oregon that has caused cultivated lands to rise from 11,700 ha to 31,000 ha between 2009 and 2018 (N. WIMAN, personal communication). With knowledge of EFB-resistance in *C. avellana*, breeders concurrently collected and screen vast *C. avellana* germplasm sourced throughout eastern and northern Europe (i.e., Estonia, Latvia, Lithuania, Poland, Russia, Crimea, Georgia, Moldova, and Turkey) (CAPIK AND MOLNAR, 2012; CAPIK et al., 2013; LEABETTER et al., 2016; MOLNAR et al., 2007; MUEHLBAUER et al., 2014). Over 5,000 resulting trees have undergone long-term field evaluation under high EFB pressure in NJ, and approximately 100 new EFB-resistant accessions have been identified that are genetically and geographically diverse (MOLNAR et al., 2018a; MUEHLBAUER et al., 2014). These selections represent a considerable expansion to the germplasm base of EFB-resistant accessions. In turn, this germplasm augments ongoing efforts to develop cultivated selections with durable EFB-resistance (MOLNAR et al., 2018a) and expand commercial hazelnut cultivation to climates in the northeast U.S. that *C. avellana* can tolerate – like the New Jersey Fruit Belt or areas buffered by large bodies of water. Expanding hazelnut

cultivation beyond the amenable climates of the northeast U.S. requires sourcing new adaptive traits, particularly to areas with colder winter temperatures and higher interannual variability.

Corylus americana is native to a continuous and vast majority of U.S. states east of the Rocky Mountains, extending from Maine to Georgia along the Atlantic coast and from Oklahoma to North Dakota and Manitoba, Canada on the western boundary of the endemic range (DRUMKE, 1964; GLEASON, 1998). The natural habitat of *C. americana* spans a wide range of climates, growing zones, and soil types. Thus, beyond *C. americana*'s primary breeding utility as a source of durable EFB resistance, the species represents a largely underutilized genetic resource of diverse adaptations. Native stands of *C. americana* hold tremendous morphological diversity and superior selections have been made for yield and nut morphologies (HAMMOND, 2006); however, as a species, its morphologies deviates quite severely from commercial standards, with a shrub-like growth habit, small kernels, thick shells, and husk encased shells. Fortunately, reciprocal crosses between *C. americana* and *C. avellana* occur very readily (ERDOGAN AND MEHLENBACHER, 2000) and offer great potential for breeding broadly adapted, EFB-resistant interspecific hybrids that produce large, commercially suitable kernels similar to *C. avellana* (MOLNAR et al., 2005; MOLNAR, 2011).

Breeding efforts to develop *C. americana* × *C. avellana* hybrids have been explored since the early 20th century, starting in 1917 with J.F Jones' (Lancaster, PA) use of his wild selection 'Rush' (Crane, 1937). Other early breeders include C.A. Reed (Washington, DC), G.L. Slate (Geneva, NY), and C. Weschcke (St. Paul, MN). Jones, Reed, and Slate's breeding efforts were each heavily dependent upon 'Rush' as the *C. americana* parent. Weschcke's crosses prominently used *C. americana* 'Winkler' (a wild selection from Iowa), but other local selections from Wisconsin were used as well. More recently, P.A. Rutter (Canton, MN) established a mass selection program using select, EFB-resistant individuals that remained in Weschcke's infected plantings (RUTTER, 1987). This program was later supplemented with other wild selections and improved hybrid germplasm, including progeny of 'Rush' (RUTTER, 1991). This collective body of early breeding work produced a number of EFB-resistant interspecific hybrid selections with improved commercial traits amongst adapted germplasm. Many of these selections are conserved for

at Oregon State University (OSU) and the U.S. Department of Agriculture – Agriculture Research Services' National Clonal Germplasm Repository (NCGR) in Corvallis, OR. Despite the broad distribution and sale of seed-based plants and several clonal selections by nurseries, germplasm derived from this early work has not supported commercial-scale planting to date because the germplasm lacks a variety of traits requisite for commercial cultivation, e.g. small kernels encased in thick shells and husks.

The interspecific hybrid germplasm produced from early efforts was developed from a narrow set of both *C. americana* and *C. avellana* parents and thus represent a narrow genetic base (MOLNAR, 2011; MOLNAR AND CAPIK, 2012). Consequently, university breeders and the NCGR have collaborated on expanding *C. americana* collection and conservation to enable genetic improvement goals for interspecific hybrids. At present, the OSU and NCGR core collection has grown to approximately 80 *C. americana* accessions. These accessions originated as selections from amongst several hundred seed-derived trees, procured with help from members of the Northern Nut Growers Association (NNGA). S.A. Mehlenbacher made selections based upon geographic origin, nut characteristics, and yield in Corvallis, Oregon. Sathuvalli and Mehlenbacher (2012) genetically characterized 87 pure *C. americana* from this collection as well as 67 *C. americana* × *C. avellana* hybrids held by OSU, the NCGR, and the Arbor Day Foundation. The study found that nearly all of the hybrid accessions clustered with either 'Rush' or 'Winkler' and the Weschcke hybrid group, confirming a narrow genetic base. Additionally, the *C. americana* accessions are highly genetically diverse and hold a greater degree of allelic variation compared to the hybrids (SATHUVALLI AND MEHLENBACHER, 2012). This germplasm expanded the genetic base of EFB-resistance selection, and especially those adapted to severe and fluctuating climates. Nevertheless, given the extensive native range of *C. americana*, this core collection likely does not represent the full extent of the genetic and trait diversity present within the species, and continued collection of wild *C. americana* is warranted.

In 2009, an extensive *C. americana* collection and evaluation effort was initiated by the Hybrid Hazelnut Consortium, which is comprised of OSU, Rutgers University, University of Nebraska – Lincoln, and the Arbor Day Foundation. Open-pollinated seed from wild *C. americana* plants has been annually procured through collaboration with a network of participatory collectors and established in a field

repository at Rutgers University. At present, the collection is composed of 618 trees of bearing age from 62 seed lots that span 38 counties and 15 states. Revord et al. (*in press*) recently reported extensive EFB-resistance within this collection. Additionally, the collection is currently under evaluation for yield, nut morphology, and kernel quality, with the overarching goal to identify superior individuals that offer a greater potential for the prompt recovery of commercial traits during pseudo-backcrossing. As a whole, this germplasm bolsters *C. americana* as genetic resources for breeders and continued characterization efforts will inform the selection of breeding parents and the conservation of core collection.

In this study, we used a genotyping by sequencing (GBS) approach to derive 2,653 single nucleotide polymorphisms (SNPs) and investigate the genetic diversity and structure of a subset of this new *C. americana* collection (272 individuals). In addition, we selected a core set of the most diverse material for more detailed morphological characterization and association studies.

3.3 MATERIALS AND METHODS

Plant materials and DNA isolation

Seed-based *C. americana* was assembled with the help of participatory colleagues and the interested public (MOLNAR et al., 2018b). Open-pollinated seed was obtained from many locations across the species' native range and sent to Rutgers for germination and subsequent field planting at the Cream Ridge Fruit Research and Extension Station (Cream Ridge, NJ) in 2012. A subset of 272 trees that reached phenotypic maturity were selected for this study, representing 55 seedlots across 35 counties and 15 states (Figure 3.1, Table A.3.). Leaf tissue was collected from this field-established collection in the spring and immediately frozen with liquid nitrogen. Frozen tissue was shipped on dry ice to the University of Illinois – Urbana-Champaign (Urbana, IL). DNA isolation used a cetyltrimethylammonium bromide (CTAB) extraction protocol optimized for 96-well plates. DNA quantification was completed using the PicoGreen assay (Molecular Probes, Eugene, OR).

GBS library development and sequence analysis

The reduced representation libraries were made using the double restriction enzyme (*Pst*I and *Msp*I) version of the Elshire et al. (2011) GBS protocol (POLAND et al., 2012). The libraries were sequenced at the Roy J. Carver Biotechnology Center at the University of Illinois in Urbana, IL, USA using 100 bp single-end (SE) reads in one lane on an Illumina HiSeq 2500 System (San Diego, CA, USA). Version 2.0 of the *de novo* GBS SNP-Calling Reference Optional Pipeline (GBS-SNP-CROP) (MELO et al., 2016) was used for the analysis of the raw Illumina sequences and genotyping. In the first stage of GBS-SNP-CROP, the raw reads (238.6 million) were parsed, trimmed based on quality, and demultiplexed into individuals FASTQ files per genotype. A *de novo* mock reference was then assembled in the second stage of the pipeline, using two genotypes with the highest read counts. In stage three, the processed reads were then stringently filtered base on quality and mapped to the pseudo-reference (as detailed at https://github.com/halelab/GBS-SNP-CROP). The acceptable proportion of missing data per variant was maintained at 25%, and the minimum average depth of an acceptable variant was raised to 6 reads. With use of the SAMTools (LI et al., 2009), reads were retained for variant calling only if they held high mapping quality (q>30) and no supplementary alignments. Bi-allelic SNPs and insertion/deletions were called in step four. For all downstream diversity analyses, we retained only a single SNP per centroid of the pseudo-reference (i.e., consensus GBS fragments, i.e., simplex SNPs. PGDSpider 2.1.1.3 was used to convert files for downstream analysis (LISCHER AND EXCOFFIER, 2011).

Analysis of genetic structure and differentiation

The genotypic data was analyzed to generate general descriptive parameters of genetic diversity within and among groups of the germplasm collection, including the number of effective alleles (N_E), the Minor Allele Frequency (MAF), the observed (H_O) and unbiased expected heterozygosities (H_E), genetic differentiation (F_{ST}), and the fixation index (F_{IS}). Each of these calculations was made using the R package 'hierfstat' software (GOUDET AND JOMBART, 2015).

Principal components analysis (PCA) was performed on a variance-covariance matrix of allele frequencies using the dudi.pca function in R (package 'adegenet') to gain perspective of the collection's structure. To determine the number of distinct groups and sub-groups underlying the collection we applied Bayesian model-based clustering followed by discriminant analysis of principal components (DAPC). The former analysis was executed using Structure v.2.3.4 (PRITCHARD et al., 2003) to evaluate the hypothetical number of sub-groups (*K*) and to quantify the ancestry partitioned to each genotype from the inferred sub-groups. The admixture ancestry model was used with correlated allele frequencies and a burn-in length of 100,000 iterations followed by 100,000 Markov chain Monte Carlo run iterations each i (1 to 15). Prior information about the geographic origin of accessions was not considered. The optimal K value was chosen using the Δ*K* method (EVANNO et al., 2005) analyzed with Structure harvester (EARL, 2012), which examines the rate of change of consecutive posterior probabilities over the spectrum of tested K values. Individuals were assigned to a group if its proportion of ancestry was ≥ 0.80; individuals with an ancestry coefficient < 0.80 were considered admixed. Subsequently, DAPC based on a sequential K-means method was performed using the find.clusters() function of the R package 'adegenet' (JOMBART AND AHMED, 2011; JOMBART et al., 2010). In DAPC, the minimum Bayesian information criterion (BIC) was used to designate the optimal number of clusters. A scatter plot of the DAPC-derived clusters was made in 'adegenet'.

The validity of the Structure and DAPC defined groups was tested by comparing member assignment to the neighbor-joining (NJ) dendrograms. An unweighted NJ tree was constructed based upon Euclidean distance using the nj() function of the 'ape' R package (POPESCU et al., 2012). Consensus sub-groups were identified using the NJ tree and DAPC in tandem. Analysis of molecular variance was performed (10,000 permutations) with GenAlEx 6.5.03 software (PEAKALL AND SMOUSE, 2006) to assess the hierarchical partitioning of genetic variance among the defined groups and sub-groups. The boot.ppfst() function of the R package 'hierfstat' was used with 1,000 bootstraps to conduct pairwise genetic differentiation (F_{ST}) tests and produce 99% confidence intervals (GOUDET AND JOMBART, 2015).

Core collection assembly

GenoCore software (JEONG et al., 2017) was used to assemble a core collection representative of 95% of the common genetic variation within the collection and thus reduce redundant allelic representation. GenoCore was run with the following parameters: -d 0.01% -cv 95%. The software was run with 10 replications, and a consensus set was selected based on repeated selection by GenoCore. Individuals captured by GenoCore were compared to the NJ tree to evaluate the coreset's representation of the collection's structure.

3.4 RESULTS AND DISCUSSION

Genotyping and basic statistics

Genotypes 11599.08028 MN and 11569.09161 OH were selected to construct the mock reference genome, as they held a respective 5.15 and 4.37 million reads compared to an average of 605,300 reads for the remaining genotypes. GBS-SNP-CROP produced 2,653 high-confidence simplex SNPs (with an average read depth of 58) for downstream analyses. SNPs were filtered to one SNP per centroid of the mock reference, and the overall heterozygosity, homozygosity, and missing data across these loci are reported in Table 3.1. Basic statistics characterizing the genetic diversity of the collection is reported in Table 3.2. The unbiased expected heterozygosity (H_E) for the collection was 0.276, and the observed heterozygosity was higher at 0.280 (H_O). Genetic differentiation is moderate (F_{ST} = 0.108), and inbreeding levels are low with a mean of F_{IS} = -0.136, which was expected given that the species is self-incompatible and wind-pollinated. Minor allele frequencies were greater than 0.1 for 82.6% of loci.

Genetic structure and differentiation

The PCA base on a variance-covariance matrix of allele frequencies depicts spatial separation of accessions, but their clear discrimination into discrete clusters is not apparent (Figure 3.2). Principal component (PC) 1 demonstrates a close relationship amongst about 50% of accessions with the remaining accessions incrementally distant from this large group in a single direction. PC 2 separates the collection

into subsequent halves, although representing a small amount of the dataset's variation. Variation is somewhat evenly distributed across the PCs. PC 1 through 3 hold a respective 13.2%, 2.1%, and 1.8% of the dataset's variation, showing that variation is rather evenly distributed across the PCs and a high level of uniqueness amongst the individual genotypes.

A Bayesian model-based clustering method was applied to the 272 unique genotypes to ascertain the genetic structure that underlies this collection. Evanno's ΔK distinctly suggested $K = 3$ as the optimal uppermost level of stratification of the collection (Figure A.1.), with minor peaks at $K = 5$, 9, and 12 suggesting some level of sub-clustering. Low standard deviation across the iterative runs of $K = 3$ substantiated its selection. Figure 3.3 depicts these $K = 3$ groups, and the inferred ancestry of each accession to these groups. The mean ancestry coefficient of accessions to their inferred group was $Q = 0.87$. Setting an ancestral relationship threshold of $Q \geq 0.80$, 181 genotypes (67%) were assigned to a group: 46 genotypes (17%) were assigned to group 1 (red), 9 genotypes (3%) were assigned to group 2 (green), and 126 genotypes (46%) were assigned to group 3 (blue). Admixed ancestry characterized 91 genotypes (33%). This partitioning of genotypes was asymmetric. Additionally, group 1 and 3 showed continuity through the continuous distribution of individuals with admixed ancestry coefficients between them.

NJ clustering validates that of Structure with a clear uppermost clustering level of $K = 3$ and consistent group assignment (Figure 3.4). Two minimally distant major groups corresponding to Structure group 1 (red) and 3 (blue) comprise a bulk of the collection, and a smaller, more distant group corresponds with group 2 (green) of Structure. Admixed individuals are represented by a lighter shade of the group in which they share the highest ancestry coefficient. Interestingly, 97% of admixed individuals clustered broadly within group 1. Beneath the uppermost clustering level, the architecture of the dendrogram is complex with many small subgroups that are relatively equally distant from common ancestry. Additionally, while the NJ dendrogram depicted some tendency to cluster by seedlot, no geographic pattern of clustering was observed; accessions from seedlots of distant and varying geographic origin frequently clustered together (Table A.3., Figure A.2.). Taken together, these clustering patterns suggest the dendrogram is comprised of many highly heterozygous, unique sub-groups of a single ancestral population.

The biology of *C. americana* explains such patterns of variation, as the species is a long-lived obligate outcrosser that is self-incompatible and wind-pollinated. Consequently, wild populations experience long-distance gene-flow and weak population structure in the absence of geographic barriers or admixture with cultivated germplasm (i.e., local landraces or cultivars) near the origins of domestication. In *Corylus* spp., weak population structure is often epitomized by a very high partitioning of variance within populations rather than among populations. For example, a recent study of 1140 wild *C. americana* shrubs from 25 populations that span South Dakota, Minnesota, Iowa, and Wisconsin attributed 90% of variance to within populations and only 10% to among population (DEMCHIK et al., 2018). Similar observations have been made with *C. avellana* Ireland (BROWN et al., 2016) and Europe (MARTINS et al., 2015) as well as *C. mandshurica* in China (YANG et al., 2018; ZONG et al., 2015). Such even distributions of variation can result in the clustering of genotypes from across wide geographies, as seen in Sathuvalli and Mehlenbacher (2012) and Öztürk et al. (2017). Despite the absence of geographic structure, the small sub-groups within the NJ tree offer an opportunity to define a greater level genetic structure within the collection.

DAPC overlooks within-group variation while maximizing between-group variation to disentangle the collection's sub-structure. Seven PCs were retained for the analysis based on the criteria that the variable explains >1.0% of the dataset's variation. Collectively, PC 1-7 explained 22% of the variation. An optimal $K = 15$ was chosen based upon selecting the lowest BIC value followed by an increase in the subsequent value (Figure A.1.). DAPC spatially forced the collection into 15 sub-groups (Figure 3.5). Six groups are spatially distinct, seven groups are adjacent to a neighboring group(s) with minor admixture, and two groups at the center of the scatter plot represent individuals retained by the original Structure groups 1 (red) and 3 (blue) (Table A.3., Figure 3.5). Noting the admixture displayed in the scatter plot, DAPC sub-group assignment coordinates quite well with the NJ tree (Figure 3.6). Additionally, it mirrors the placement of the Structure-defined groups while further assigning the formerly admixed individuals into sub-groups. DAPC and NJ clustering were then used in tandem to select 19 consensus groups (Table 3.3, Figure 3.6). Consensus group assignment was guided mainly by DAPC placement; however, it allowed for a reassignment of accessions that bordered multiple subgroups and clustered into an admixed branch of the

NJ tree (e.g., DAPC groups 11 and 13), or the joining of distinct NJ sub-groups characterized by the DAPC placement (e.g., consensus group 16).

Genetic differentiation measures deem these groups and sub-groups informative organizations of genetic diversity, in where sub-groups represent discrete and hierarchical partitioning of variance that breeders can exploit accordingly. Pairwise F_{ST} values between the sub-groups range from very greatly differentiated (>0.25) to little differentiation (<0.05), with all values significantly different from zero based on bootstrap derived 95% confidence intervals (Table 3.4). Consensus sub-groups (1-7), representing spatially distinct DAPC groups with minimal admixture, were often greatly (0.15-0.25) or very greatly differentiated from one another while maintaining great to moderate (0.05-0.15) differentiation from larger sub-groups containing greater admixture (8-19) (Table 3.4). Differentiation between these admixed sub-groups was lower and ranged between moderate and little. AMOVA revealed that 47.8% of variation was partitioned to these sub-groups, while 48.0% of variation is explained by to the individual genotypes (Table 3.5). Together, F_{ST} and AMOVA show the relatively high magnitude of genetic variation that is explained due to differences within these sub-groups. In effect, these sub-groups enable breeders to utilize the breadth of genetic diversity within the collection and effectively expand the genetic base of *C. americana* parents used in interspecific hybridization. Additionally, this structure will assist with avoiding inbreeding depression in subsequent generations (i.e., F_2 and F_2BC_X), which is common in species that are clonally propagated and outcrossing.

Constructing a core set

A core set was defined based upon genetic diversity, where individuals were selected to represent 95% of the common alleles within the collection. Fifty-one individuals were selected, which represent 18 of the 19 consensus sub-groups (Figure 3.7), 35 of the 55 seedlots, and 14 of the 15 states (Table A.3.). This core set will be clonally propagated for conservation in field repositories at Rutgers University, Oregon State University, and the University of Missouri. Assessments of EFB response, flowering and vegetative

bud break phenological diversity, cold tolerance, yield, and kernel quality are ongoing and will result in additional selections to assemble a full core collection.

3.5 CONCLUSIONS

This analysis represents a marked step in broadening the genetic base of *C. americana* parentage. Highly heterozygous wild *C. americana* from a broad geographic range were organized into groups and sub-groups based on genetic structure. This organization provides ample genetic differentiation of the collection, where breeders can now discretely sample across the collection's diversity. This organization is essential for the effective use of *C. americana* because geographic origin does not correspond to genetic relatedness, and breeders can now circumvent inbreeding depression by selecting more distantly related individuals during down-stream crosses. As additional efforts to phenotypically characterize this collection proceed, superior individuals will be selected for use as breeding parents to develop interspecific hybrids with durable EFB resistance and wider adaptation to eastern North America.

3.6 FIGURES AND TABLES

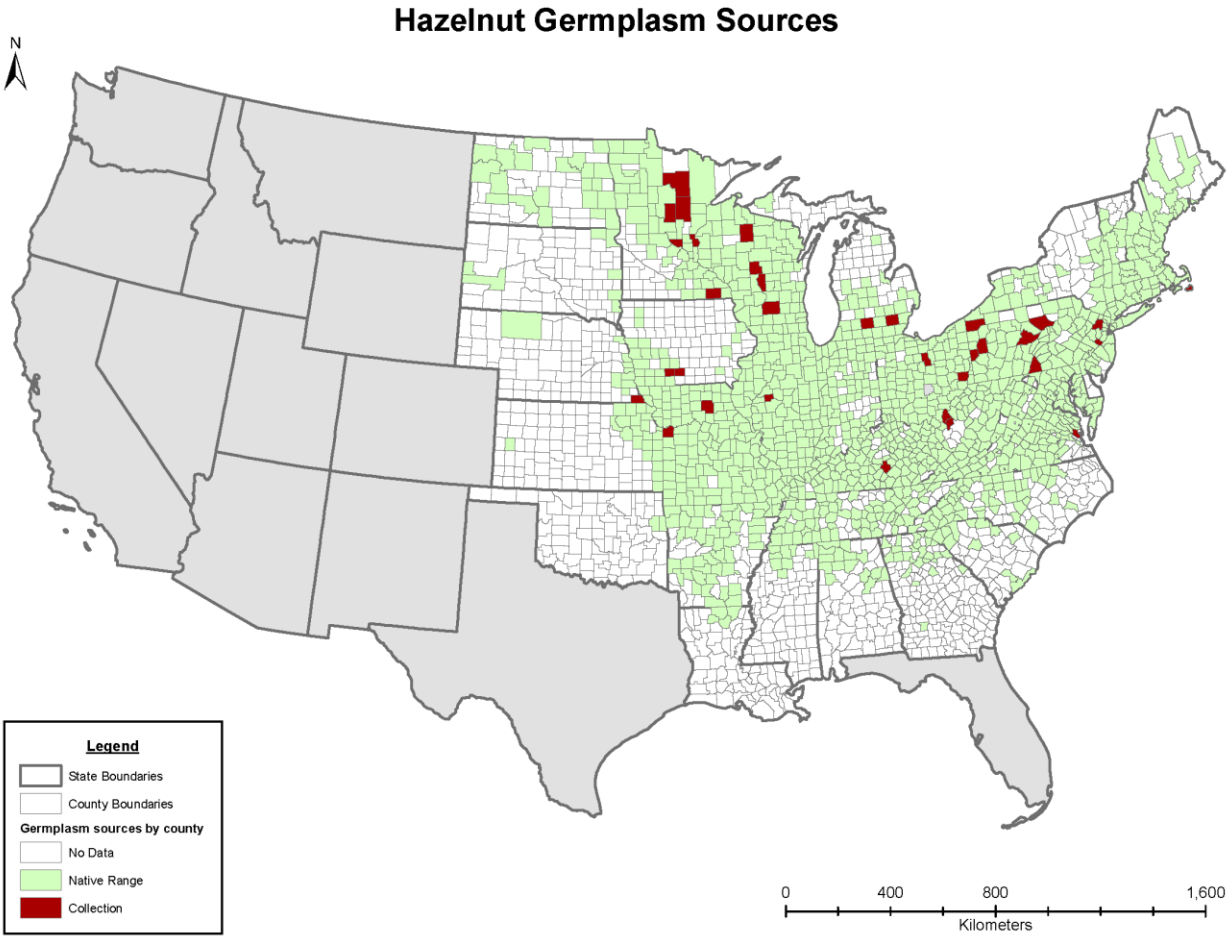


Figure 3.1. A geographic information services map depicting the native distribution of *Corylus americana* based upon herbarium records (plants.usda.gov) (light green), and the distribution of the seedlots represented in this study (red).

Table 3.1. Parameters characterizing the GBS-SNP-CROP-derived set of SNPs used to analyze the *Corylus americana* collection.

N[a]	SE reads[b]	SNPs[c]	D[d]	D > 20[e]	Hetero[f]	Homo[g]	Na[h]
273	238,669,945	2,653	58.6	100	28.6	71.4	19.8

[a]Number of genotypes sequenced

[b]Number of single-end (SE) reads used for SNP calling

[c]SNPs called following genotyping criteria and filtering

[d]Average read depth per called SNP

[e]Percentage of SNPs with an average read depth greater than 20

[f]Percentage of heterozygote genotypes

[g]Percentage of homozygote genotypes

[h]Percentage of missing genotypes across all SNP-accession combinations

Table 3.2. Parameters characterizing the genetic diversity of the *Corylus americana* germplasm collection.

MAF[a]	N_E[b]	H_O[c]	H_E[d]	F_{ST}[e]	F_{IS}[f]
82.6	1.443	0.280	0.276	0.108	-0.136

[a]Percent of loci with a minor allele frequency greater than 10%

[b]Average number of effective alleles per locus

[c]Observed heterozygosity

[d]Unbiased expected heterozygosity

[e]Genetic differentiation

[f]Inbreeding coefficient

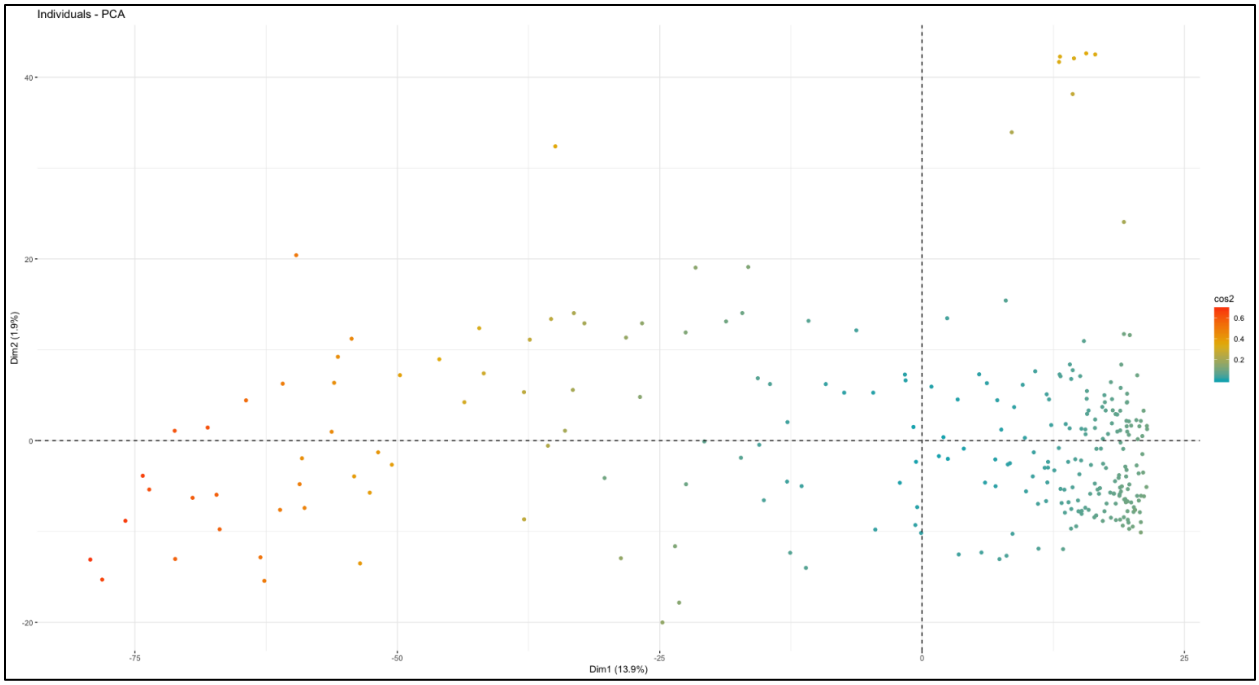


Figure 3.2. A scatter plot of principal component (PC) 1 x PC 2. Principal component analysis using the 2653 SNPs separates some of the accessions, but discrete clusters are not apparent. PC 1 (13.9 %) describes variation that is shared by approximately half of the individuals, while PC 2 (1.9 %) separates the accessions into two halves. Variation is rather evenly distributed across the principal components, suggesting many unique individuals.

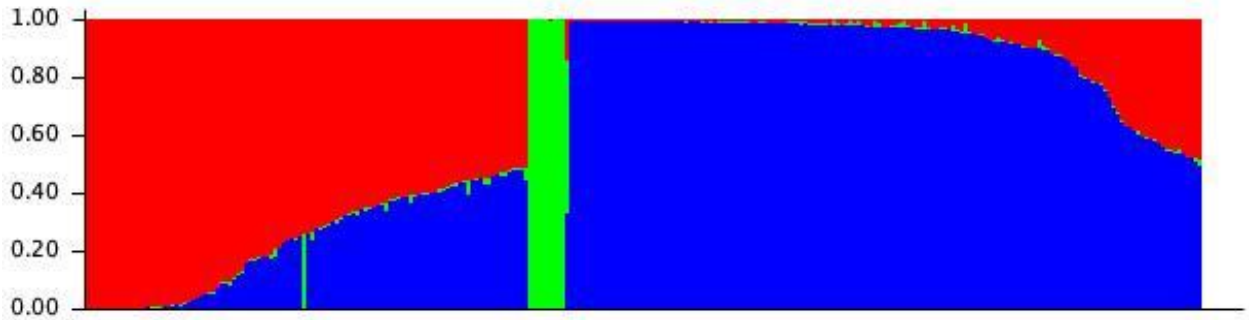


Figure 3.3. The inferred uppermost structure of the germplasm collection based upon the Bayesian model-based program STRUCTURE (PRITCHARD et al., 2000). Each accession is represented by a corresponding vertical bar that is partitioned into $K = 3$ color segments, which represent the ancestral relationship of the genotype to the three clusters. While ΔK supports $K = 3$ (Figure A.1.), the structure-defined group 1 (red) and group 3 (blue) show continuity through a continuous distribution of individuals with admixed ancestry coefficients.

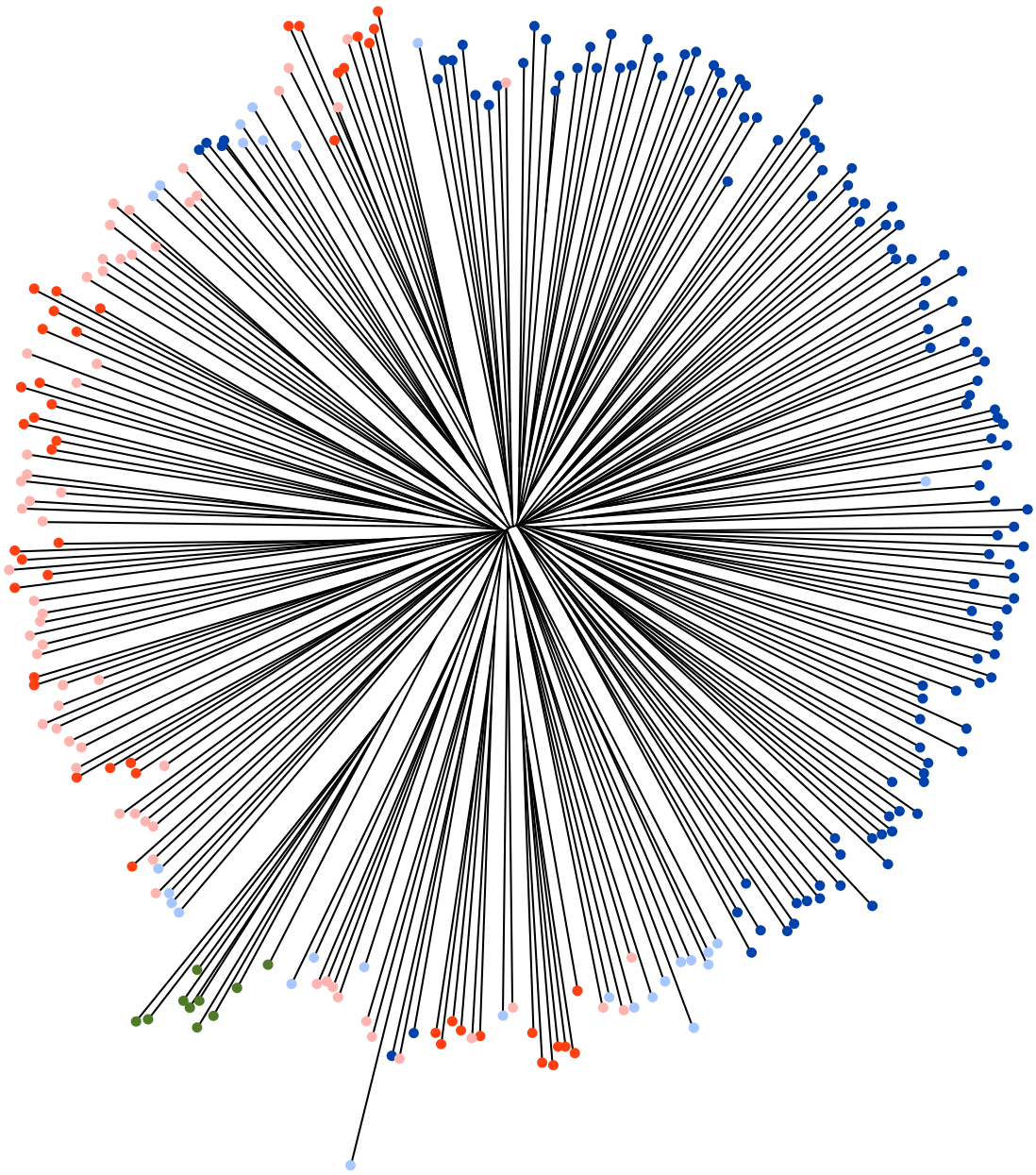


Figure 3.4. A neighbor-joining dendrogram based on an euclidean distance matrix calculated from 2653 SNPs. The 272 individuals clustered into three groups that represent the uppermost genetic structure of the collection and correspond to the Structure-defined groups (colored circles). The tree's architecture is complex, depicting two minimally distant major groups (red and blue) and one smaller, more distant group (green). The major groups are composed of many small sub-groups that are relatively equally distant from a common ancestor. Lightly colored individuals reflect admixed individuals with a coefficient of ancestry less than 0.80 as defined by Structure.

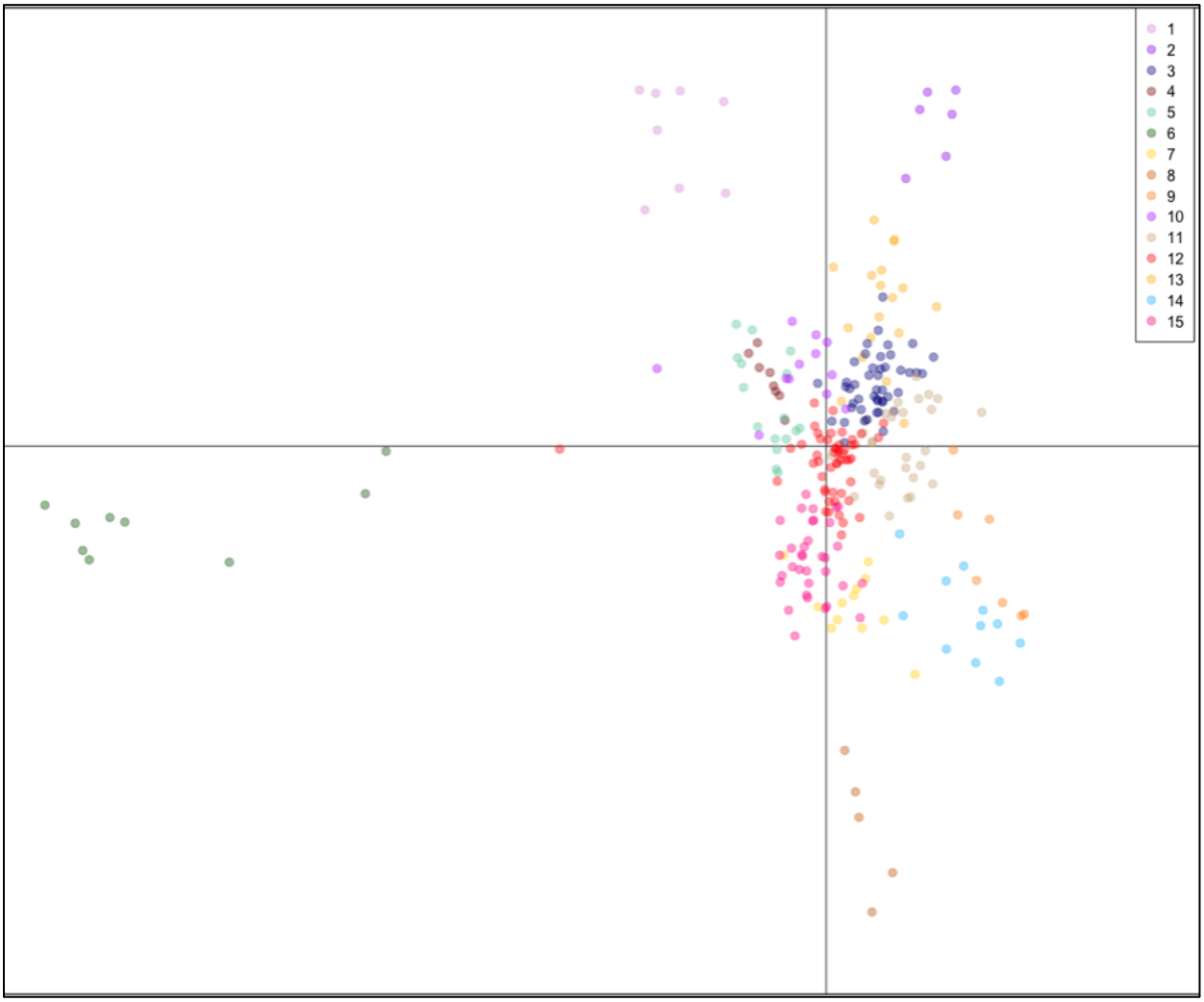


Figure 3.5. A scatter plot of principal component (PC) 1 x PC 2 (horizontal x vertical) from discriminant analysis of principal components, which depicts between-group variation while overlooking within-group variation. Seven PCs where retained for this analysis, which represent 21% of the variation within the data. While many sub-groups are distinct, most sub-groups are adjacent to each other and show some admixture.

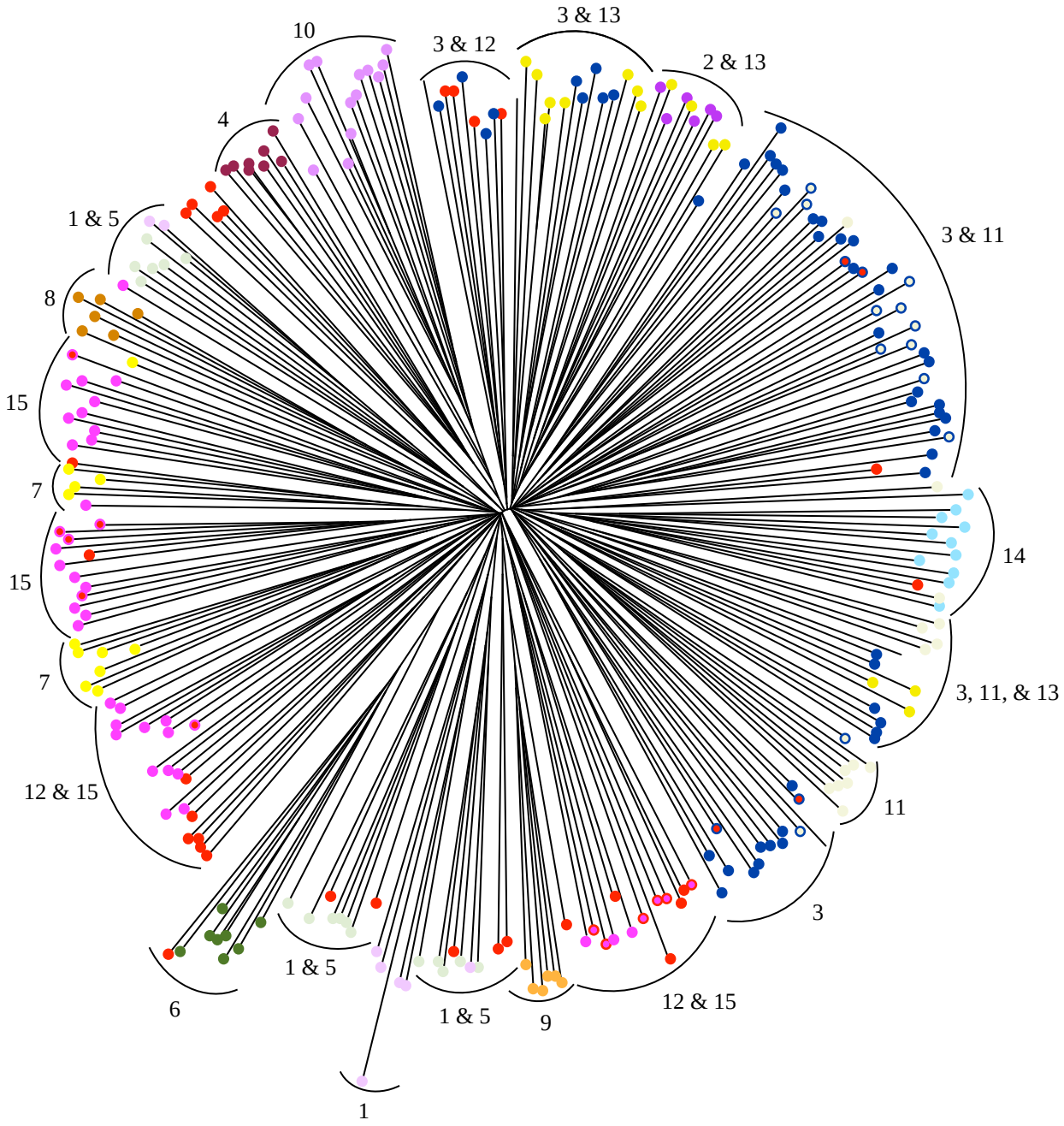


Figure 3.6. A neighbor-joining (NJ) dendrogram based on an euclidean distance matrix calculated from 2653 SNPs with *K* = 15 groups depicted base upon discriminant analysis of principal components (DAPC) (colored circles). Circles with a different color interior represent accessions with a >40 % relationship to a second sub-group. Sub-groups defined by DAPC frequently match those of the NJ dendrogram, with admixed sub-groups also consistently reflected. Consensus groups (using DAPC and the NJ tree) are bracketed and assigned a "consensus group" number.

Table 3.3. Consensus sub-groups based on discriminant analysis of principal components and neighbor-joining tree clustering.

Consensus groups	DAPC
1	4
2	6
3	8
4	9
5	10
6	14
7	1
8	1 & 5
9	2 & 13
10	3 & 13
11	3
12	11
13	3 & 11
14	3 & 12
15	3, 11, & 12
16	7
17	12
18	12 & 15
19	15

Table 3.4. A matrix of pairwise F_{ST} values representing genetic differentiation between the consensus sub-groups.

Consensus groups

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	-	0.314	0.193	0.252	0.162	0.129	0.144	0.094	0.094	0.173	0.086	0.081	0.131	0.102	0.092	0.124	0.144	0.078	0.090
2	0.281	-	0.350	0.397	0.310	0.282	0.289	0.236	0.216	0.346	0.245	0.215	0.287	0.257	0.235	0.265	0.317	0.206	0.222
3	0.167	0.314	-	0.253	0.198	0.178	0.191	0.136	0.114	0.213	0.133	0.132	0.181	0.141	0.139	0.133	0.158	0.094	0.089
4	0.220	0.355	0.222	-	0.254	0.220	0.234	0.187	0.160	0.279	0.189	0.168	0.233	0.175	0.178	0.196	0.244	0.142	0.153
5	0.142	0.277	0.172	0.225	-	0.150	0.165	0.126	0.111	0.185	0.113	0.106	0.160	0.122	0.113	0.130	0.157	0.094	0.105
6	0.111	0.250	0.152	0.191	0.131	-	0.116	0.071	0.084	0.150	0.065	0.055	0.098	0.069	0.056	0.103	0.105	0.068	0.079
7	0.124	0.258	0.164	0.205	0.144	0.099	-	0.052	0.092	0.162	0.071	0.068	0.111	0.081	0.066	0.117	0.124	0.078	0.092
8	0.080	0.209	0.114	0.161	0.108	0.059	0.040	-	0.049	0.110	0.030	0.026	0.066	0.041	0.028	0.070	0.067	0.035	0.048
9	0.080	0.190	0.094	0.136	0.094	0.071	0.078	0.039	-	0.077	0.044	0.044	0.084	0.043	0.043	0.050	0.059	0.026	0.028
10	0.147	0.308	0.182	0.244	0.159	0.124	0.135	0.088	0.057	-	0.110	0.101	0.154	0.116	0.103	0.138	0.139	0.092	0.099
11	0.071	0.217	0.111	0.161	0.095	0.052	0.057	0.022	0.034	0.086	-	0.022	0.059	0.033	0.024	0.062	0.056	0.031	0.044
12	0.069	0.190	0.112	0.144	0.091	0.046	0.056	0.020	0.036	0.080	0.016	-	0.055	0.029	0.021	0.060	0.062	0.031	0.041
13	0.112	0.256	0.155	0.204	0.139	0.083	0.095	0.054	0.071	0.128	0.046	0.045	-	0.084	0.062	0.110	0.111	0.070	0.081
14	0.086	0.227	0.119	0.151	0.105	0.057	0.066	0.030	0.033	0.094	0.023	0.020	0.069	-	0.032	0.068	0.081	0.035	0.043
15	0.077	0.208	0.117	0.152	0.097	0.045	0.053	0.020	0.033	0.079	0.015	0.015	0.050	0.022	-	0.064	0.063	0.031	0.040
16	0.106	0.235	0.112	0.168	0.112	0.085	0.099	0.057	0.041	0.113	0.050	0.050	0.093	0.054	0.052	-	0.081	0.040	0.037
17	0.118	0.276	0.127	0.208	0.129	0.080	0.099	0.046	0.041	0.106	0.034	0.042	0.085	0.060	0.042	0.059	-	0.042	0.048
18	0.066	0.181	0.077	0.119	0.079	0.057	0.064	0.028	0.021	0.073	0.022	0.025	0.057	0.025	0.024	0.031	0.025	-	0.017
19	0.076	0.196	0.072	0.129	0.089	0.066	0.077	0.038	0.022	0.078	0.033	0.033	0.067	0.033	0.031	0.029	0.030	0.012	-

Key	F_{ST}
Little	<0.05
Moderate	0.05 - 0.15
Great	0.15 - 0.25
Very great	>0.25

[a]All pairwise F_{ST} were significantly different than 0.

[b]95% confidence intervals are represented with the upper and lower boundaries represented on the corresponding sides of the matrix.

Table 3.5. AMOVA partitioning of variance in allele frequencies within the collection and consensus sub-groups

Source of variation	df	SS	MS	Variance	Percent Variance
All					
Among groups	18	29534.4	1640.8	26.3	4.1
Among genotypes	251	228847.0	911.7	303.5	47.8
Within genotypes	270	82250.5	304.6	304.6	48.0

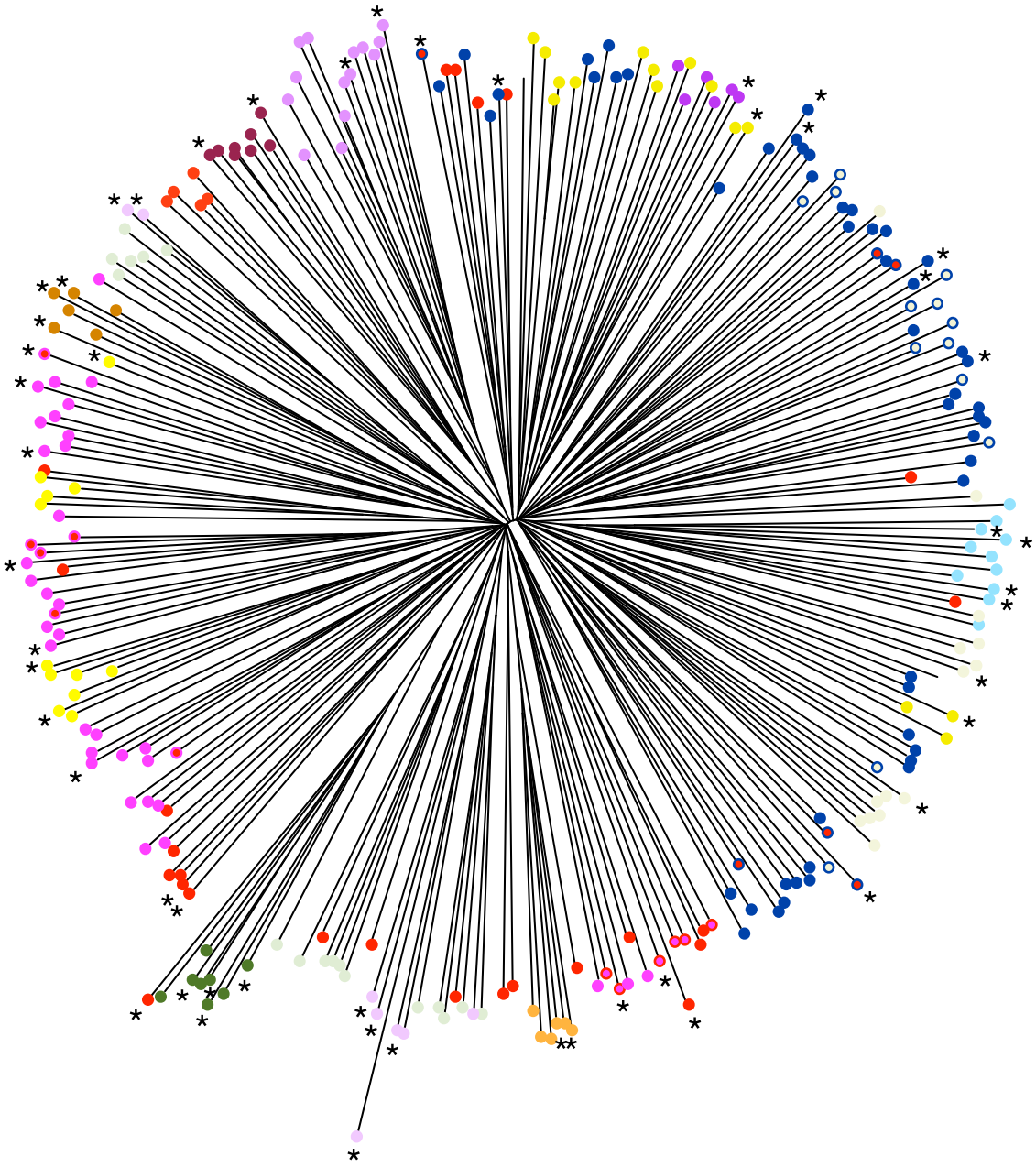


Figure 3.7. Neighbor-joining tree depicting the relationship of individuals selected for the core set to the collection. Individuals selected by Gencore (JEONG et al., 2017) are denoted with an asterisk and collectively represent 95% of the allelic variation observed amongst the 272 studied individuals.

3.7 REFERENCES

Brown, J.A., G.E. Beatty, W.I. Montgomery, and J. Provan. 2016. Broad-scale genetic homogeneity in natural populations of common hazel (*Corylus avellana*) in Ireland. Tree Genet. Genomes 12:122.

Capik, J.M. and T.J. Molnar. 2012. Assessment of host (*Corylus* sp.) resistance to eastern filbert blight in New Jersey. J. Amer. Soc. Hort. Sci. 137:157-172.

Capik, J.M., M. Muehlbauer, A. Novy, J.A. Honig, and T.J. Molnar. 2013. Eastern filbert blight-resistant hazelnuts from Russia, Ukraine, and Poland. HortScience 48:466-473.

Crane, H.L., Reed C.A., and Wood, M.N. . 1937. Yearbook of agriclture, p. 835-844. Nut Breeding.

Demchik, M., A. Kern, L. Braun, J. Fischbach, and K. Turnquist. 2018. Genetic diversity of american hazelnut in the Upper Midwest, USA. Agroforestry Systems 92:1507-1516.

Drumke, J. 1964. A systematic survey of *Corylus* in North America. Univ. of Tenn., Knoxville, TN, PhD diss.

Earl, D.A. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genet. Resources 4:359-361.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS ONE 6:e19379.

Erdogan, V. 2017. Hazelnut production in Turkey: current situation, problems and future prospects. Acta Hort. 1226:13-24.

Erdogan, V. and S.A. Mehlenbacher. 2000. Interspecific hybridization in hazelnut (*Corylus*). J. Amer. Soc. Hort. Sci. 125:489-497.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611-2620.

FAOSTAT. 2017. Agricultual production, crops, Food and Agriculture Organization of the United Nations, Rome.

Fuller, A. 1908. The nut culturist. Orange Judd, New York, NY.

Gleason, H.A. and A. Cronquist. 1963. Manual of vascular plants of Northeastern United States and adjacent Canada. The New York Botanical Gardens, Bronx, NY.

Goudet, J. and T. Jombart. 2015. Hierfstat: estimation and tests of hierarchical F-statistics. R package version 0.04-22.

Hammond, E. 2006. Identifying Superior Hybrid Hazelnut Plants in Southeast Nebraska. University of Nebraska-Lincoln, Lincoln, Nebraska, M.S. Thesis.

Jeong, S., J.-Y. Kim, S.-C. Jeong, S.-T. Kang, J.-K. Moon, and N. Kim. 2017. GenoCore: A simple and fast algorithm for core subset selection from large genotype datasets. PloS ONE 12:e0181420.

Jombart, T. and I. Ahmed. 2011. Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070-3071.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11:94.

Leabetter, C.W., J.M. Capik, S.A. Mehlenbacher, and T.J. Molnar. 2016. Hazelnut Accessions from Russia and Crimea Transmit Resistance to Eastern Filbert Blight. J Amer. Pomol. Soc. 70:92-109.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079.

Lischer, H.E. and L. Excoffier. 2011. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298-299.

Martins, S., F. Simões, D. Mendonça, J. Matos, A.P. Silva, and V. Carnide. 2015. Western european wild and landraces hazelnuts evaluated by SSR markers. Plant Mol. Biol. Rptr. 33:1712-1720.

Mehlenbacher, S. 2018a. Advances in genetic improvement of hazelnut. Acta Hort. 1226:1-12.

Mehlenbacher, S. 2018b. Breeding hazelnuts resistant to eastern filbert blight. Acta Hort. 1226:73-78.

Mehlenbacher, S.A., A.N. Azarenko, D.C. Smith, and R. McCluskey. 2007. ‘Santiam’hazelnut. HortScience 42:715-717.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2009. 'Yamhill' hazelnut. HortScience 44:845-847.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2011. 'Jefferson' hazelnut. HortScience 46:662-664.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2013. ‘Dorris’ hazelnut. HortScience 48:796-799.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2014. ‘Wepster’ hazelnut. HortScience 49:346-349.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2016. ‘McDonald’ hazelnut. HortScience 51:757-760.

Mehlenbacher, S.A., M.M. Thompson, and H.R. Cameron. 1991. Occurrence and inheritance of resistance to eastern filbert blight in Gasaway hazelnut. HortScience 26:410-411.

Melo, A.T., R. Bartaula, and I. Hale. 2016. GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. BMC bioinformatics 17:29.

Molnar, T., J. Goffreda, and C. Funk. 2005. Developing hazelnuts for the eastern Unites States. Acta Hort. 686:609-618.

Molnar, T., J. Lombardoni, M. Muehlbauer, J. Honig, S. Mehlenbacher, and J. Capik. 2018a. Progress breeding for resistance to eastern filbert blight in the eastern United States. Acta Hort 1226:79-86.

Molnar, T.J. 2011. *Corylus*, p. 15-48. In: C. Kole (ed). Wild crop relatives: Genomic and breeding resources. Springer, Berlin, Heidelberg.

Molnar, T.J. and J. Capik. 2012. Advances in hazelnut research in North America. Acta Hort. 940:57-65.

Molnar, T.J. and J.M. Capik. 2012. Eastern filbert blight susceptibility of american × european hazelnut progenies. HortScience 47:1412-1418.

Molnar, T.J., J.A. Honig, A. Mayberry, R.S. Revord, S.T. Lovell, S.A. Mehlenbacher, and J.M. Capik. 2018b. *Corylus americana*: a valuable genetic resource for developing hazelnuts adapted to the eastern United States. Acta Hort. 1226:115-122.

Molnar, T.J., D.E. Zaurov, J.C. Goffreda, and S.A. Mehlenbacher. 2007. Survey of hazelnut germplasm from Russia and Crimea for response to eastern filbert blight. HortScience 42:51-56.

Muehlbauer, M.F., J.A. Honig, J.M. Capik, J.N. Vaiciunas, and T.J. Molnar. 2014. Characterization of eastern filbert blight-resistant hazelnut germplasm using microsatellite markers. J. Amer. Soc. Hort. Sci. 139:399-432.

Öztürk, S.C., H.İ. Balık, S.K. Balık, G. Kızılcı, Ö. Duyar, S. Doğanlar, and A. Frary. 2017. Molecular genetic diversity of the Turkish national hazelnut collection and selection of a core set. Tree Genet. Genomes 13:113.

Peakall, R. and P.E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6:288-295.

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.-L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PloS ONE 7:e32253.

Popescu, A.-A., K.T. Huber, and E. Paradis. 2012. Ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. Bioinformatics 28:1536-1537.

Pritchard, J.K., W. Wen, and D. Falush. 2003. Documentation for STRUCTURE software: Version 2.

Rutter, M. 1991. Variation in resistance to eastern filbert blight in hybrid hazelnuts. Ann. Rpt. Northern Nut Groweres Assn. 82:159-162.

Rutter, P. 1987. Badgersett research farm - plantings, projects, and goals. Annu. Rep. Northern Nut Growers Assoc. 78:173-186.

Sathuvalli, V.R. and S.A. Mehlenbacher. 2012. Characterization of american hazelnut (*Corylus americana*) accessions and *Corylus americana* × *Corylus avellana* hybrids using microsatellite markers. Genet, Resources and Crop Evol. 59:1055-1075.

U.S. Department of Agriculture. 2018. Noncitrus Fruits and Nuts 2017 Summary, National Agricultural Statistics Service. U.S. Dept. Agr., Washington, D.C.

Weschcke, C. 1953. Growing nuts in the north. Webb, St. Paul, MN.

Yang, Z., L. Wang, and T. Zhao. 2018. High genetic variability and complex population structure of the native chinese hazelnut. Brazilian J. Bot. 41:687-697.

Zong, J.-W., T.-T. Zhao, Q.-H. Ma, L.-S. Liang, and G.-X. Wang. 2015. Assessment of genetic diversity and population genetic structure of *Corylus mandshurica* in China using SSR markers. Plos One 10:e0137528.

CHAPTER 4

THE IDENTIFICATION AND MAPPING OF EASTERN FILBERT BLIGHT RESISTANCE QTL IN *CORYLUS AMERICANA* OSU 403.040 USING GENOTYPING-BY-SEQUENCING

4.1 ABSTRACT

The american hazelnut, *Corylus americana*, is a valuable genetic resource for developing durable resistance to the disease eastern filbert blight (EFB). EFB, caused by the ascomycete *Anisogramma anomala* [Peck] E. Muller, is the primary limitation to expanding commercial hazelnut cultivation into regions within the eastern U.S. While the disease is inconsequential to *C. americana* and occasionally produces small cankers, the species has tiny, thick-shelled nuts and is not commercially cultivated. The european hazelnut (*Corylus avellana*), however, is the species of global commerce, and most cultivars are susceptible. Genetic resistance is the primary and most economical means to manage the disease, with the *R*-gene of 'Gasaway' providing the resistance in species that support the hazelnut industry in Oregon. While 'Gasaway' resistance is effective against EFB in the Pacific Northwest (PNW) U.S., recent studies show that 'Gasaway'-carrying genotypes do develop EFB in parts of the eastern U.S. This disease incidence suggests there is limited pathogenic variation of *A. anomala* in the PNW. In this study, we developed a genetic linkage map using a genome-by-sequencing approach and identified quantitative traits loci (QTL) associated with EFB resistance from the selection OSU 403.040 *C. americana* from Nebraska U.S. A bi-parental mapping population comprised of 121 seedling trees was evaluated for EFB under high pressure in New Jersey, where *A. anomala* is endemic and highly genetically diverse. With EFB response represented by the percent of diseased wood, a total of three QTLs were discovered on linkage groups (LG) 3, 6, and 11 that respectively represent 62.6%, 23.3%, and 11.1% of the phenotypic variation. EFB resistance from OSU 403.040 appears to represent a new source of resistance based upon it being only the second mapped source from *C. americana* and due to it mapping to three loci – all other sources of EFB resistance in *Corylus* spp. are monogenic and map to a single locus. Additionally, OSU 403.040 likely exhibits resistance to a broad range of *A. anomala* given the genetically diverse *A. anomala* environment under which it was selected. Such durability is requisite for the development of a feasible commercial

variety for the eastern U.S. and highlights a priority for its inclusion in gene pyramiding schemes with resistant *C. avellana*.

4.2 INTRODUCTION

World hazelnut production is predominant in Mediterranean countries – Turkey (420,000 t), Italy (120,572 t), Azerbaijan (33,941 t), and the Republic of Georgia (29,500 t) – as well as the United States (34,473 t) (FAOSTAT, 2017). The bulk of the global market is comprised of high-quality kernels that are sold into various confectionary industries, with a remaining 10% of hazelnuts sold to the in-shell market (CIARMIELLO et al., 2014; PETRICCIONE et al., 2010). The U.S. prioritizes this in-shell market, and 99% of its commercial production occurs in the Willamette Valley of Oregon (USDA, 2018). In recent years, demand for hazelnut has consistently exceeded supply, and many nations have implemented programs to expand their commercial hazelnut cultivation. U.S. hazelnut cultivation reached 31,809 ha in 2018 – a 2.7-fold increase since 2008 (FAOSTAT, 2019) – supported by new *Corylus avellana* varieties released by Oregon State University (OSU) with eastern filbert blight (EFB) resistance from the 'Gasaway' cultivar. The cultivation of *C. avellana* in eastern North America is prevented by the species' susceptibility to EFB, its lack of climatic adaptation, and greater diversity in the virulence of the EFB pathogen. In turn, the U.S. has sought to diversify sources of EFB resistance, particularly with germplasm from climates with potential to expand U.S. cultivation beyond the Mediterranean-like climate of Oregon.

EFB is a prolific stem canker disease caused by the fungal pathogen *Anisogramma anomala* [Peck] E. Muller (JOHNSON AND PINKERTON, 2002). The disease is often lethal to *C. avellana*, with systemic hyphae spread that leads to perennial cankering and branch girdling (JOHNSON AND PINKERTON, 2002). Breeding host resistance is the economically preferred means of disease control (JULIAN et al., 2008), which is practical – as seen through the effective use of the 'Gasaway' gene. The obsolete pollinizer *C. avellana* 'Gasaway' was revealed as EFB resistant during a disease epidemic in the Willamette Valley (CAMERON, 1976). Subsequently, the cultivar's resistance was discovered to be conferred by a dominant allele for resistance at a single locus (MEHLENBACHER et al., 1991). The resistance gene of 'Gasaway' maps to

linkage group 6 (LG6) (MEHLENBACHER et al., 2006) and has been a prevalent source within the OSU hazelnut breeding program, contributing to the recent release of many advanced generation cultivars (MEHLENBACHER et al., 2009, 2011, 2013, 2014, 2016). However, initial long-term evaluation of 'Gasaway' and its offspring revealed minor infections by *A. anomala* isolates from outside of Oregon – New Jersey, Minnesota, and Michigan (MOLNAR et al., 2010a, 2010b). Consequently, breeders share concerns over the reliability of a single resistance source as well as the long-term durability of monogenic sources of resistance, in Oregon and the broader eastern U.S.

Complicating these concerns, *A. anomala* in the Pacific Northwest (PNW) is narrow in genetic diversity and virulence, and thus, 'Gasaway' was likely screened and selected against limited fungal isolates. EFB is thought to have originated in the PNW from a single introduction (GOTTWALD AND CAMERON, 1980; JOHNSON et al., 1996; PINKERTON et al., 1993). Recent genetic diversity studies of the pathogen support this hypothesis, showing no genetic structure in the PNW isolates and the converse amongst isolates representing the eastern U.S. where populations are diverse and differentiated (CAI et al., 2013; MUEHLBAUER et al., 2014, 2018; TOBIA et al., 2017). While initial greenhouse and field screening of the 'Gasaway' gene against a broad diversity of eastern U.S. isolates revealed that resistance was either maintained or small, inconsequential cankers developed (CAPIK AND MOLNAR, 2012; CAPIK et al., 2013; MOLNAR et al., 2010a, 2010b). Continued evaluation under high disease pressure at Rutgers University in New Jersey recently revealed susceptible phenotypes on 'Gasaway' and its offspring, with cankers commonly over 1 m (T.J. MOLNAR, unpublished data). The pathogen is endemic to this region, and genetic diversity and variation in virulence are documented here (CAI et al., 2013; MOLNAR et al., 2010b; MUEHLBAUER et al., 2014, 2018; unpublished data). This occurrence suggests the possibility of a new virulent isolate and is compelling support that prolonged pressure from a diverse *A. anomala* population on monogenic resistance creates a selection pressure for new virulent isolates. Interestingly, this has occurred even without the heightened selection pressure of large monocultures of the *R*-gene(s).

OSU and Rutgers University have made concerted collection expeditions and evaluations in pursuit of diverse and durable sources of EFB resistance. This collaboration pursues durability in two fashions: i)

pyramiding unique sources of monogenic resistance or major quantitative trait loci (QTLs), and ii) selecting donor parents with ancillary quantitative resistance in the background of major loci. The first step toward this objective involved expanding the genetic base of EFB resistance through the extensive collection of seed-based *C. avellana* across eastern Europe and the Caucasus Republics. A resulting 5,000 seedlings, representing 200 seed lots and ten countries, were subsequently put through long-term field evaluations in New Jersey. Around 150 selections of diverse genetic structure and origin maintain resistance or high tolerance to EFB populations with diverse virulence (MOLNAR et al., 2018). A collection of traditional *C. avellana* cultivars has also been screened (CAPIK AND MOLNAR, 2012; CHEN et al., 2007; COYNE et al., 1998; LUNDE et al., 2000), and monogenic resistance was mapped to one of three LGs (2, 6, or 7) in 6 genotypes other than 'Gasaway'. Half of these sources – the Spanish cultivar 'Culplà', the Serbian cultivar 'Crvenje', and the Russian seedling OSU 495.072 – map to a similar region as the 'Gasaway' gene on LG6 despite diverse geographic origin (COLBURN et al., 2015). One might suspect that these sources comprise an *R*-gene cluster, possibly with similar modes of action. Nevertheless, resistance from 'Ratoli' (Spanish cultivar) and *C. avellana* OSU 759.010 (Republic of Georgia selection) add diversity and respectively map to LG7 and LG2 (SATHUVALLI et al., 2011a, 2011b). Recently, an eighth source of monogenic resistance has also been mapped to LG7 in the *C. americana* 'Rush' selection (BHATTARAI et al., 2017).

C. americana is a highly underutilized yet valuable genetic resource for U.S. hazelnut breeders. The species is an endemic host of *A. anomala*, and thus it shares a long co-evolutionary history with the pathogen. Initial natural observations and evaluations indicate *C. americana* is a prevalent source of EFB resistance and high tolerance (CAPIK AND MOLNAR, 2012; FULLER, 1908; WESCHCKE, 1953). Additionally, it is an ideal source of resistance for U.S. breeders in that it is cross-compatible with *C. avellana* (ERDOGAN AND MEHLENBACHER, 2000) and widely adapted to the target region of the eastern U.S. (DRUMKE, 1964; GLEASON AND CRONQUIST, 1963). Initial backcrosses to *C. avellana* indicate an optimistic outlook of recovering the commercial traits of *C. avellana* in one to two generations (MEHLENBACHER, pers. comm.). Although only a relatively narrow base of *C. americana* parentage has been used in interspecific hybrid development to date (SATHUVALLI AND MEHLENBACHER, 2012), there is a growing desire to use *C.*

americana as a source of resistance to improve durability. Comprehensive collection and characterization efforts have proceeded in recent years to identify superior breeding selections that also have EFB resistance amenable to systematic crossing schemes (MOLNAR et al., 2018, unpublished data). Recently, a subset of the *C. americana* accessions held by the OSU and USDA – Agriculture Research Service repositories were discovered to transmit both resistance and tolerance to their respective interspecific progenies in a bimodal pattern, where resistance is recovered along with a secondary peak representing quantitative resistance (REVORD et al., *in press*). In intraspecific specific crosses between resistant and susceptible *C. avellana*, a bimodal distribution is also common but with the second peak reflecting a susceptible phenotype rather than tolerance. This transmission pattern is intriguing as it offers an ancillary background of quantitative resistance to the major loci, which is valuable for maintaining high levels of tolerance should monogenic or major loci be overcome in the future and also for lessening the selection pressure for new virulent *A. anomala* isolates. EFB resistance from a greater number of *C. americana* must be mapped to elucidate the genetic control of the trait(s) and enable accelerated selection while using *C. americana* as a donor parent.

The objective of this study was to generate a genetic linkage map using a genotyping-by-sequencing (GBS) approach and subsequently conduct QTL mapping experiments in the *C. americana* × *C. avellana* interspecific progeny Rutgers 11525, which segregates for resistance/tolerance in a continuous and quantitative manner. The *C. americana* parent OSU 403.040 is a wild seedling selection from Nebraska that expresses an EFB-free phenotype, while the *C. avellana* parent 'Tonda di Giffoni' is a traditional italian cultivar that expresses a susceptible phenotype in New Jersey.

4.3 MATERIALS AND METHODS

Plant Material

Mapping population parents

A controlled cross between *C. americana* OSU 403.040 Nebraska and *C. avellana* 'Tonda di Giffoni' produced the bi-parental mapping population used in this study. Both parents are heterozygous individuals. OSU 403.040 was the female parent and is an EFB-resistant selection, rated as EFB-free after

long-term trials in New Jersey (CAPIK AND MOLNAR, 2012). OSU 403.040 originated from seed collected from an open-pollinated wild plant in Nebraska. It was subsequently evaluated amongst a large seed-based *C. americana* germplasm collection procured by S.A. Mehlenbacher at OSU and selected for retention in a core collection based on its EFB resistance, nut morphologies, and geographic origin. OSU 403.040 remained free of EFB after greenhouse exposure in Oregon as well as in the field (MEHLENBACHER, unpublished data). Grafted clones of OSU 403.040 were exchanged with Rutgers University and remained free of EFB after more than 10 years of evaluation in New Jersey (CAPIK AND MOLNAR, 2012), where multiple virulent *A. anomala* isolates reside and pathogen genetic diversity is high (CAI et al., 2013; MOLNAR et al., 2010c; MUEHLBAUER et al., 2014, 2018; TOBIA et al., 2017). 'Tonda di Giffoni' was selected as the male parent for the cross based on its susceptibility to EFB, improved nut and kernel characteristics, and high yields. 'Tonda di Giffoni' is rated as a susceptible phenotype in New Jersey (CAPIK AND MOLNAR, 2012), although a minor degree of basal resistance is apparent in comparison to the truly susceptible genotype (i.e., no basal resistance).

Mapping population development

The bi-parental cross OSU 403.040 × 'Tonda di Giffoni' was controlled and conducted as details in the Mehlenbacher (1994) protocols. OSU 403.040 is located at Rutgers University Horticultural Research Farm 3 (East Brunswick, NJ), and OSU provided pollen of 'Tonda di Giffoni'. Pollen was collected in January 2010 and stored at -28.9 ˚C prior to shipping on dry ice by overnight mail before its use in February 2011 at Rutgers. In September 2010, the resulting seeds were harvested, provided a moist chilling period at 4˚ C for 4 months, then germinated and grown in the greenhouse according to Molnar and Capik (2012) protocols. Seedlings were moved outdoors for acclimation under a shade cloth (40% shade) in June 2011 and field planted in November 2011 at Cream Ridge Fruit Research and Extension Station (Cream Ridge, NJ). Tree spacing was ~1.0 m within the row by ~3.5 m between rows. Herbicides, irrigation, and fertilizer were applied as needed for weed control, with no use of insecticides or fungicides. The seedlings were not pruned.

Field inoculation and evaluation of the EFB response

For the first three years following planting, hazelnut stems harboring EFB stromata were collected from infected trees growing in Rutgers field plots and tied into the canopy of every fifth shrub at the time of leaf budbreak in Apr. (MOLNAR et al., 2007). Additionally, many susceptible seedlings from within the experimental plot housing this mapping population, as well as adjacent breeding nurseries and experimental plots harboring hundreds of infected plants, added to the amount of inoculum and natural disease spread as the study progressed.

All shrubs were evaluated in January 2018 and January 2019 for response to EFB, per a modified version of the Pinkerton et al. (1992) 0 to 5 rating index : 0 = no detectable EFB (inclusive of the "sunken lesions" phenotype – small sunken cankers that lack fungal stromata) = 0% of stems diseased; 1 = single canker (with fully formed stromata) ≅ 1% of stems diseased; 2 = multiple cankers on a single branch ≅ 5% of stems diseased; 3 = multiple branches with cankers ≅ 25% of stems diseased; 4 = 50% branches with cankers ≅ 50% of stems diseased; 5 = all branches contain cankers (except basal sprouts) = 100% of stems diseased. Plants scored 0 or 1 were considered resistant to infection by *A. anomala*. The 0 – 5 scale was converted to percent disease, as described above, for the QTL analyses (HONIG et al., *in press*).

In addition, individual cankers were measured (cm) on each progeny. Total canker length was calculated for each progeny by summing the length of individual cankers. Additionally, total stem length was measured by summing the length of every stem of a genotype and used to determine the percent of diseased wood with respect to individual progeny (total canker length divided by total stem length). As in previous studies on this pathosystem, square root transformations were made for each of these measures prior to calculation (COYNE et al., 2000; MEHLENBACHER et al., 2008). Both the disease rating conversions and the percent diseased wood (square root) were used for QTL analyses.

DNA extraction

Genomic DNA was isolated from both parents and 121 seedling samples using a Qiagen DNeasy Plant Mini Kit (Germantown, MD), following the manufacturer's instructions. A Nanodrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA) was used to assess DNA quality and quantity. Extracted DNA was used to prepare GBS libraries for all samples using a double restriction enzyme digestion approach.

GBS library development and sequence analysis

GBS libraries for OSU 403.040, 'Tonda di Giffoni', and 121 mapping population seedlings were made using the double restriction enzyme (*Pst*I and *Msp*I) version of the Elshire et al. (2011) GBS protocol (POLAND et al., 2012). Sequencing pools were constructed with progeny samples divided equally across three pools. OSU 403.040 and 'Tonda di Giffoni' were sequenced at 10× across sequencing runs. Pools were sequenced on three respective lanes of a HiSeq 2500 (Illumina, San Diego, CA) (2 x 150 paired-end high-output sequencing) by Genewiz, Inc. (South Plainfield, NJ). Each Hiseq sequencing run was loaded with 30% PhiX.

Genomic DNA (200 ng) from each sample was double digested with the rare-cutting *PstI* (NEB, Ipswich, MA) and the common-cutting *MspI* restriction enzymes for 2 h at 37°C. Forward *PstI* adapters and reverse *MspI* Y-adapters with unique barcodes (5 – 10 bp) were ligated to the digested DNA in a mastermix [200 U of T4 DNA ligase, 2 uL of 10 × NEBuffer 4, and 4 uL of ATP (10 mM) per sample] (NEB, Ipswich, MA). The ligation reaction was incubated for 2 h at 22°C, with ligase inactivated by a subsequent incubation of 20 min at 65°C. Samples were then "cleaned-up" with 0.5 v/v Agencourt Ampure XP magnetic beads (Beckman Coulter, Brea, CA), and DNA fragments smaller than 300 bp were removed by washing with 70% ethanol. Individual clean library samples were PCR amplified using primers containing sequences for Illumina (San Diego, CA) next-generation sequencing (NGS) flow cell binding and the following thermalcycling conditions: initial denaturation of 95 °C for 30 s; followed by 16 cycles of 95 °C for 30 s, 62 °C for 20 s, 68 °C for 15 s; with a final extension of 68 °C for 5 min. DNA in each

library was quantified using a Qubit 3.0 flourometer (Thermo Fisher Scientific, Waltham, MA), normalized to 5 ng/µL, and pooled for Illumina short-read sequencing. Prior to sequencing, the magnetic bead clean-up and wash steps were repeated for the pooled libraries. An Agilent 2100 Bioanalyzer System (Santa Clara, CA) assessed the quality of final pooled libraries.

GBS SNP maker calling

Version 4.0 of the *de novo* GBS SNP-Calling Reference Optional Pipeline (GBS-SNP-CROP) (MELO et al., 2016; MELO AND HALE, 2018) was used to analyze the raw Illumina sequences and call genotypes. Commands are listed in Text A4.1. In the first stage of GBS-SNP-CROP, the three pools were processed individually and useable raw read pairs were parsed, trimmed based on quality, and demultiplexed into FASTQ files respective to genotype. A *de novo* mock reference was then assembled using the two parent genotypes OSU 403.040 and 'Tonda di Giffoni', which were sequenced at 10x and held a respective 26.0 and 25.7 million raw reads. In stage three, the processed reads were then stringently filtered base on quality and mapped to the pseudo-reference (as detailed at https://github.com/halelab/GBS-SNP-CROP). The acceptable proportion of missing data per variant was maintained at 25% missing. With use of the SAMTools (LI et al., 2009), reads were retained for variant calling only if they held high mapping quality (Q >30) and no supplementary alignments. Bi-allelic SNPs and insertion/deletions were called in step four. For all downstream analyses, we retained only a single SNP per centroid of the pseudo-reference (i.e., simplex SNPs). VCF2Mapmaker (developed by A. Bombarely) was used to convert the VCF file of GBS-SNP-CROP to JoinMap format using the cross-pollination format (- cp) (https://github.com/aubombarely/GenoToolBox/tree/master/SNPTools).

Linkage Map Construction

The software JoinMap 4.1 was utilized to construct the genetic linkage map, employing an extension of the multipoint maximum likelihood mapping algorithm to cross-pollinated (CP) full-sib populations (VAN OOIJEN, 2006). A single map was made by coding parental SNP markers as heterozygous

in one parent and homozygous in the other parent (*lm* × *ll* and *nn* × *np*, respectively), while common markers were coded as (*hk* × *hk*). Individual markers were subjected to Chi-square (χ^2) analysis to test for goodness of fit to expected segregation ratios (1:1). Markers with distorted segregation ($P \leq 0.05$) were excluded from analyses. Marker groupings were tested over a range of logarithm of odds (LOD) values from 2 to 30, with a step of 1. A final LOD value of 13 was used to group loci. Subsequently, loci order was determined using the multipoint maximum likelihood mapping algorithm with a Gibbs sampling procedure (VAN OOIJEN, 2006). Spatial sampling was conducted with five threshold values (0.1, 0.05, 0.03, 0.02, and 0.01), with three map optimization runs for each spatial sampling. Map optimization parameters included: chain length of 1000; cooling control parameter of 0.001; and chain termination after 10,000 chains without improvement. The Kosambi mapping function was used to convert map distances from recombination frequency (RF) to centiMorgans (cM).

QTL analysis

MapQTL 6.0 software (VAN OOIJEN, 2009) was used for QTL analysis of EFB response phenotypes. As described by van Ooijen (2009), markers of the two-way pseudo-testcross approach were recoded as double haploid (DH) population type and used to construct two separate parental maps due to memory constraints with the CP mapping population format. These new parental maps were then used individually for QTL mapping (MCADAM et al., 2013; STUDER et al., 2006; VAN HEERDEN et al., 2014; ZYPRIAN et al., 2016). Interval mapping (IM) was conducted first and followed by iterative rounds of multiple QTL mapping (MQM) to refine QTL position and magnitude. In short, IM was used to identify putative QTLs and in turn, proximal SNPs to be used as initial cofactors in the MQM analysis. With these initial cofactors, a backward elimination procedure was implemented in MQM analysis to select additional cofactors that carried over into subsequent rounds of MQM. The procedure was completed once QTL positions were stabilized. A permutation test (n = 1000) was conducted to determine LOD thresholds for QTL significance, using a genome-wide significance level of 0.05. MapChart 2.30 (VOORRIPS, 2002) was used to visualize QTL positions and their magnitude.

4.4 RESULTS AND DISCUSSION

Segregation for EFB resistance

The EFB response of the OSU 403.040 × 'Tonda di Giffoni' progeny was a continuous distribution where the most frequent disease class was a tolerant phenotype intermediate to both parental phenotypes. Histograms displaying phenotype distributions for both the EFB ratings and the percent diseased wood are displayed in Figure 4.1 and have respective progeny means of 2.93 and 0.21. Using the EFB ratings, 11% of individuals were rated resistant (i.e., ratings 0 and 1), 26% and 31% were rated highly tolerant and tolerant at rating 2 and 3, respectively, and 18% and 14% of offspring were classified as a susceptible phenotype with a respective rating of 4 or 5. The EFB ratings were somewhat normally distributed, with ratings slightly negatively skewed towards the higher disease classes. The percent diseased wood phenotypes were normally distributed across six bins, which comprised intervals of disease phenotypes that ranged 0.08 (e.g., 0.00 – 0.08) (Figure 4.1B.). Both distributions support the segregation of quantitative resistance, with some resistance offspring.

SNP discovery and polymorphic loci development

The GBS libraries were sequenced across three pools, each on a single lane of the Hiseq instrument platform. The pools held a respective 173.8, 176.7, and 168.5 million reads for a total of 519 million reads and 114 Gb of DNA sequences. OSU 403.040 and 'Tonda di Giffoni' were used to make the pseudo-mock-reference genome, where 26.0 and 25.7 million reads were aligned after demultiplexing and removal of barcodes, adapters, and low-quality bases (Q <30). The mock reference contained 453,165 clusters of an average length of 267 base pairs and was collectively 121.0 million base pairs in length. The average number reads aligned to the mock reference per offspring was 2.2 million, ranging from 913,936 to 2.99 million. This read volume produced on average ≅11.9× the reads in the parental samples compared to the offspring, which supports high SNP calling accuracy.

The GBS-SNP-CROP pipeline produced 16,291 SNPs with an average depth of 53.9. Thirty-nine percent of loci were heterozygous, and 50.9% were homozygous, with 9.4% missing data. Filtering variants

to one per cluster reduced the SNP set to 8,130, and after further filtering based on a heterozygosity depth of 5 and 10% missing data per progeny, 1,935 SNPs were retained. Within JoinMap, 556 additional SNPs were removed prior to linkage grouping due to duplicated loci, suspect linkages, failure to group during linkage mapping, and segregation distortion ($P \leq 0.05$). Of the remaining 1,397 SNPs, 1,373 SNPs segregated 1:1 (*lm* × *ll* or *nn* × *np*) and 24 SNPs segregated 1:2:1 (*hk* × *hk*).

Linkage map construction

The complete linkage map is depicted in Figure 4.2, and individual groups that display SNP marker positions are displayed in Figure A.3.–A.13. Eleven integrated linkage groups were returned at a LOD of 13, consistent with the haploid chromosome number of hazelnut ($n = x = 11$) (Figure 4.1; Figure A.3.–A.13.) (MEHLENBACHER et al., 2006). The map's 1,397 markers spanned a total genetic distance of 1262.7 cM with an average spacing of 0.90 cM. LGs averaged 114.8 cM in distance, ranging from 88.5 cM (147 markers) for LG9 to 159.5 cM (152 markers) for LG1, and each contained common markers (*hk* × *hk*). In many previous genetic mapping studies of hazelnut, which are primarily of *C. avellana* × *C. avellana* progenies, LG assignments were made through comparing the placement of SSR anchor markers to previous linkage maps (BELTRAMO et al., 2016; BHATTARAI et al., 2017, 2018; COLBURN et al., 2015; GÜRCAN AND MEHLENBACHER, 2010; GÜRCAN et al., 2010; IVES et al., 2014; MEHLENBACHER et al., 2006; SATHUVALLI et al., 2011a; SATHUVALLI AND MEHLENBACHER, 2011, 2012; TORELLO MARINONI et al., 2018). The identification of polymorphic SSR markers that amplify in both *C. americana* and *C. avellana*, and particularly this mapping population, is an ongoing pursuit. In the only previous mapping study using a *C. americana* × *C. avellana* progeny (BHATTARAI et al., 2017), a limited number of SSR markers were used and thus anchoring SSR markers could not be leveraged from this previous work. In this previous study, Bhattarai et al. (2017) studied a population segregating for monogenic resistance, where SSR markers previously associated with monogenic resistance were first tested. Disease scores were highly correlated with markers proximal to monogenic resistance on LG7, and subsequent mapping was performed with just nine SSR markers known to straddle the putative loci. Ongoing efforts to genotype the 121

hazelnut seedlings of this study's OSU 403.040 × 'Tonda di Giffoni' progeny with SSR markers will allow for syntenic comparisons of the linkage groups reported here to those of previous studies.

QTL analysis

QTL analysis was performed in MapQTL 6.0 in two phases – IM and MQM – with both the conversion of EFB ratings as well as percent diseased wood. For both phenotyping methods, IM identified an initial putative marker/trait association on a similar region of LG3. The association was discovered when using the parental map of the EFB-resistant OSU 403.040 *C. americana* parent – via the DH two-way pseudo-testcross approach (VAN OOIJEN, 2009). Subsequent cofactor selection and MQM mapping refined the QTL position and magnitude, and in the case of percent diseased wood, MQM mapping also revealed secondary QTLs (Figure 4.3). Permutation tests (n = 10,000) were conducted per linkage group and with a genome-wide significance level of 0.05 to determine the LOD thresholds for QTL significance (dashed line). The putative EFB resistance QTL of LG3 stabilized following iterative MQM and became prominently clear at SNP marker 2900, for both the converted EFB ratings (Figure 4.3A) and percent diseased wood (Figure 4.3B). Interestingly, the QTL LOD scores (solid red line) were higher when using percent diseased wood (24.9 to 6.8), and a greater percent of phenotypic variation was explained by the loci (62.6% to 14.5%). When using percent diseased wood, two additional QTLs were discovered on LG6 (Figure 4.3C) at SNPs 1894 and 762 and LG11 (Figure 4.3D) at SNPs 385 and 463. The QTL on LG6 held a LOD score of 12.2 and explained an additional 23.3% of the phenotypic variation. The QTL on LG11 held a LOD score of 6.6 and explained a further 11.1% of the phenotypic variation. These three QTLs account for 97% of the phenotypic variation observed in the progeny's disease response, and their positions are denoted on the integrated link map by green boxes (Figure 4.2).

While the more traditional and expedient EFB rating (and conversion) method can successfully identify major EFB resistant QTL in *C. americana*, percentage diseased wood is clearly more informative and offers a near-complete picture of the QTL's genetic control underlying the phenotypic response in this case. Additionally, SSR anchoring markers have yet to be added to this linkage map, and we thus cannot

make a distinction regarding the novelty of the resistance loci discovered in this study based upon LG position. However, the polygenic control of EFB resistance from OSU 403.040 supports it is a unique source of resistance/tolerance and additive to the sources that have been previously mapped, which are predominantly monogenic. Further, OSU 403.040 is only the second mapped source of resistance from *C. americana*. As breeders pursue durable EFB resistance, it is immensely valuable to discover that quantitative resistance in select *C. americana* is amenable to QTL analysis and that polygenic EFB resistance can be governed by few QTLs (in this case 3) as opposed to many. These results give an optimistic outlook for the efficient and effective use of select *C. americana* as donors of EFB resistant QTLs and open a new avenue for pyramiding loci that contribute to the durability of resistance.

Many of the sources of monogenic resistance to EFB were selected in Oregon, where pathogen virulence is limited, as the population is likely clonal and originated from a single source. Additionally, while the genetic base of monogenic EFB resistance has expanded in recent years, with eight sources now mapped, these sources map to similar regions with respect to three LGs. When trialed in high-pressure environments where the pathogen population is highly diverse and variably virulent, these monogenic sources express different EFB responses, which extend from the maintenance of resistance to the occasional minor canker to extensive cankering. This body of work suggests *A. anomala* isolates elicit varying EFB responses from these *R*-genes. This research also highlights the need to develop germplasm with durable resistance, where *R*-genes are pyramided into a single genotype or *R*-genes are discovered that confer resistant to multiple isolates. A first step toward achieving this goal has been screening germplasm under high disease pressure composed of genetically diverse isolates in an attempt to identify selections resistant to a variety of fungal isolates. Many sources of resistance or high tolerance have been identified in this manner, and recently, the first of such monogenic resistance sources was mapped in the Rutgers' *C. avellana* selection (HONIG et al., *in press*). The mapping of OSU 403.040 represents the second source of mapped EFB resistance selected within such an environment.

Through this study, we substantiate the ability to identify select *C. americana* in the screening manner mentioned above that are amenable to gene pyramiding schemes. *C. americana* offers advantages

in pursuit of durable resistance in that the species has co-evolved with the pathogen across a broad endemic range. Such fungal pathogen systems with long-lived hosts offer an opportunity for a variety of *R*-gene/effector protein interactions to evolve. These populations can be exceptionally complex, often segregating for several *R*-genes that have corresponding pathogen effectors, and the presence or absence of a given *R*-gene or pathogen effector results in some degree of disease. Although this is not necessarily what was observed in OSU 403.040, this biology suggests that select *C. americana* might be advantageous for gene pyramiding pursuits. Further, *C. americana* is characterized as having high basal resistance and select accessions offer an opportunity to introgress ancillary tolerance into the background of monogenic loci or major QTLs, which should be advantageous as an "insurance policy" should major loci be overcome during the 40+ year cultivation life of the tree crop. Finally, the discovery of DNA markers tightly linked to EFB resistance genes or QTLs is essential for the effective pyramiding of loci via marker-assisted selection and should persist as an ongoing priority.

4.5 FIGURES

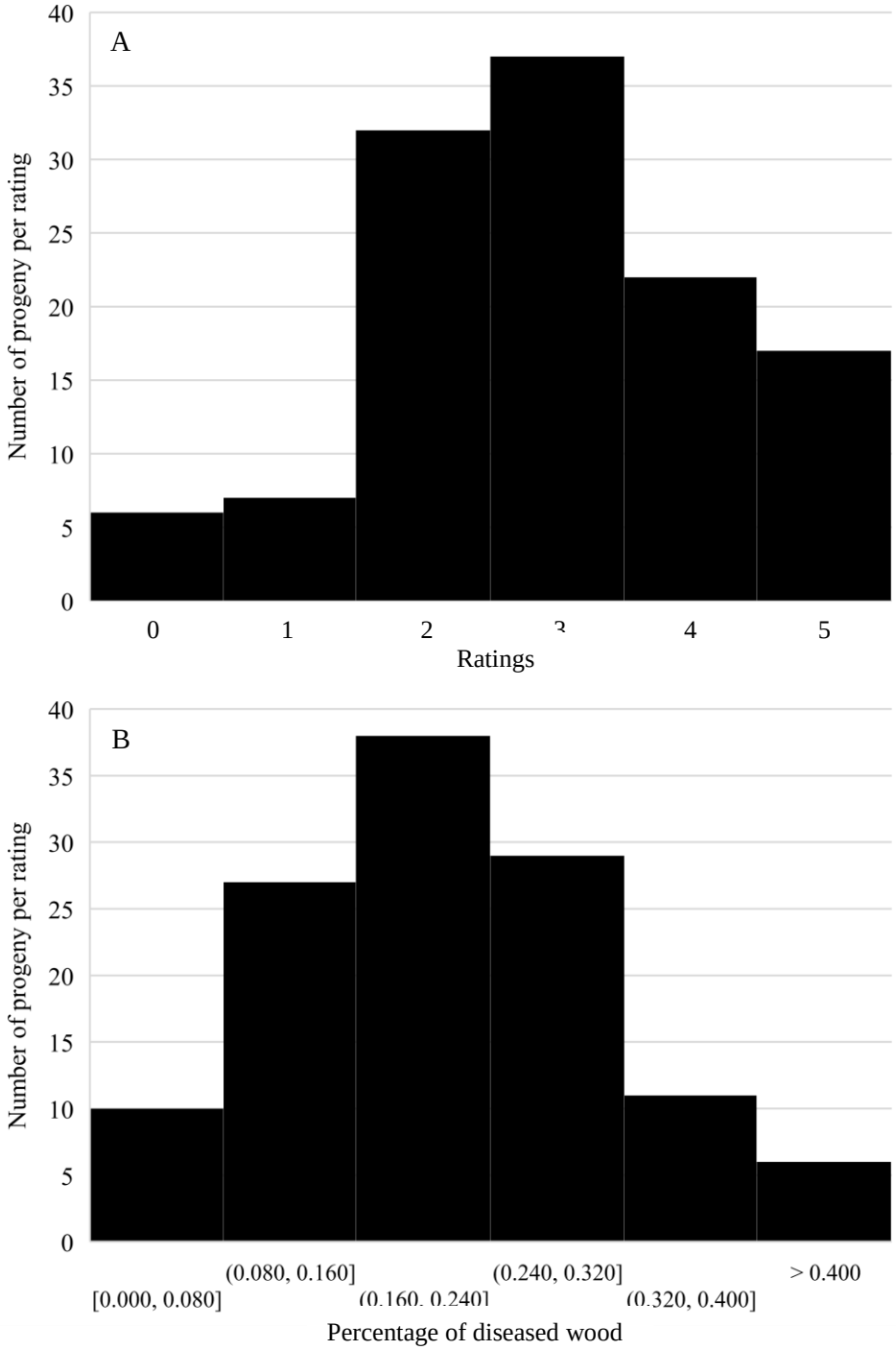


Figure 4.1. Histograms of the OSU 403.040 *C. americana* × 'Tonda di Giffoni' *C. avellana* progeny using the categorical ratings (A) and percentage of diseased wood (B). Both measurements of EFB show continuous and relatively normal distributions indicative of a polygenic trait.

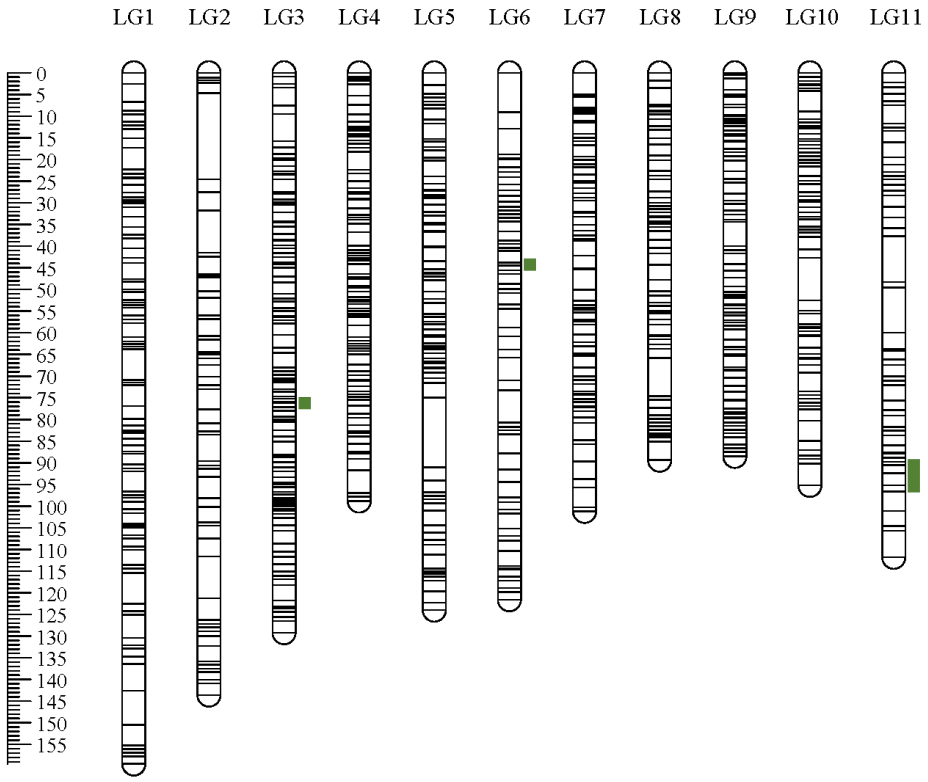


Figure 4.2. A genetic linkage map was constructed using of 1357 SNPs and the 121 hazelnut seedlings of the OSU 403.040 × 'Tonda di Giffoni' cross. Genetic distance (Kosambi) is in centimorgans (cM). EFB disease resistance QTLs detected using the percentage of diseased wood phenotype are represented by green boxes on LG3, LG6, and LG11. The QTL on LG3 was also detected using the conversion of the categorical EFB ratings.

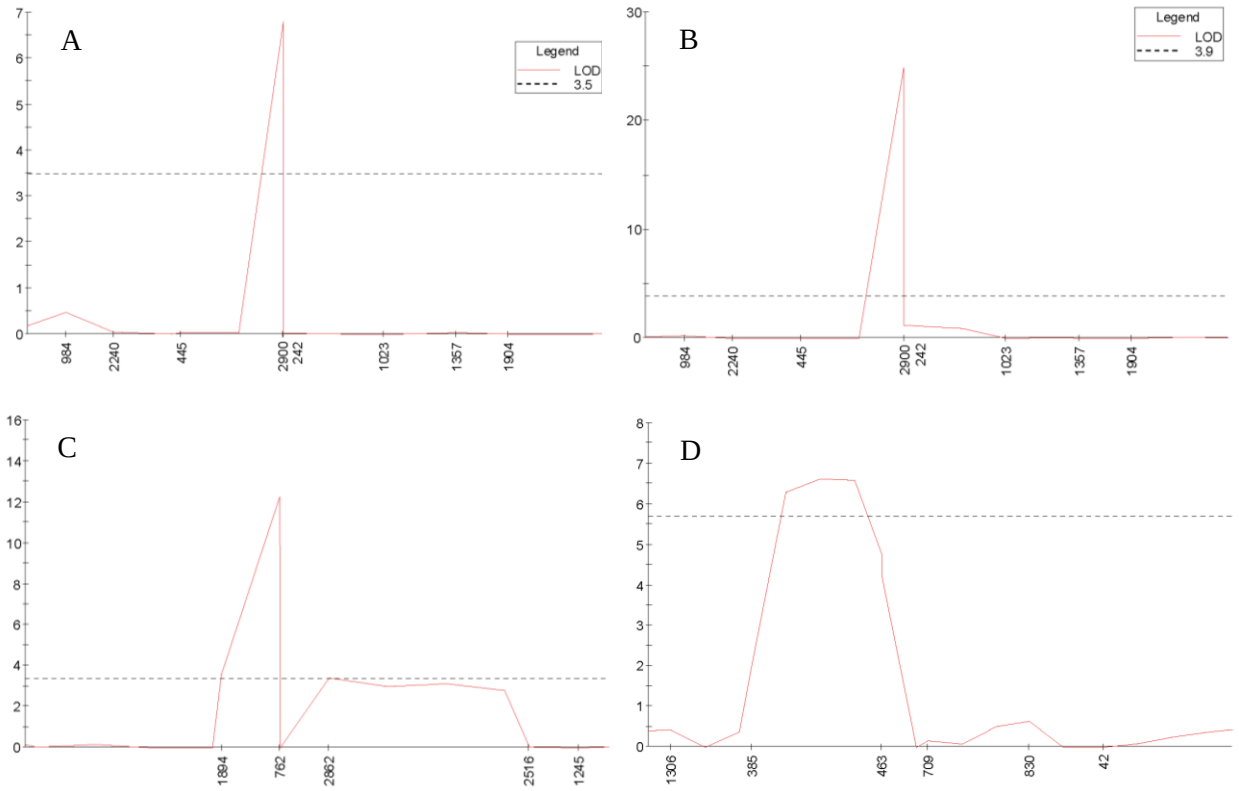


Figure 4.3. The quantitative trait loci (QTL) on their respective genetic linkage maps for the OSU 403.040 × 'Tonda di Giffoni' cross. A permutation test (n = 10,000) per linkage group determined LOD thresholds for QTL significance (dashed line) with a genome wide significance level of 0.05. An EFB disease resistance QTL was discovered on linkage group (LG) 3 at SNP marker 2900 using both phenotyping methods – the conversion of the categorical disease ratings (A) and the percentage of diseased wood (B); the QTL LOD scores (solid red line) were higher when using percentage of diseased wood (24.9 to 6.8), and a greater percent phenotypic variation was explained by the loci (62.6% to 14.5%). Two additional QTLs were discovered when using percentage of diseased wood on LG6 (C) at SNPs 1894 and 762 and LG11 (D) at SNPs 385 and 463. The QTL on LG6 held a QTL LOD score of 12.2 and explained 23.3% of the phenotypic variation. The QTL on LG11 held a QTL LOD score of 6.6 and explained 11.1% of the phenotypic variation.

4.6 REFERENCES

Beltramo, C., N. Valentini, E. Portis, D. Torello Marinoni, P. Boccacci, M.A. Sandoval Prando, and R. Botta. 2016. Genetic mapping and QTL analysis in european hazelnut (*Corylus avellana* L.). Mol. Breeding 36:27. DOI 10.1007/s11032-016-0450-6

Bhattarai, G., S.A. Mehlenbacher, and D.C. Smith. 2017. Eastern filbert blight disease resistance from *Corylus americana* 'Rush' and selection 'Yoder #5' maps to linkage group 7. Tree Genet. Genomes 13:45. DOI 10.1007/s11295-017-1129-9

Bhattarai, G. and S.A. Mehlenbacher. 2018. Discovery, characterization and linkage mapping of simple sequence repeat markers in hazelnut. J. Amer. Soc. Hort. Sci. 143:347-362. DOI 10.21273/JASHS04461-18

Cameron, H.R. 1976. Eastern filbert blight established in the Pacific Northwest. Plant Dis. Rptr. 60:737–740.

Cai, G., C.W. Leadbetter, M.F. Muehlbauer, T.J. Molnar, and B.I. Hillman. 2013. Genome-wide microsatellite identification in the fungus *Anisogramma anomala* using illumina sequencing and genome assembly. PLoS ONE 8. DOI 10.1371/journal.pone.0082408

Capik, J.M. and T.J. Molnar. 2012. Assessment of host (*Corylus* sp.) resistance to eastern filbert blight in New Jersey. J. Amer. Soc. Hort. Sci. 137:157-172.

Capik, J.M., M. Muehlbauer, A. Novy, J.A. Honig, and T.J. Molnar. 2013. Eastern filbert blight-resistant hazelnuts from Russia, Ukraine, and Poland. HortScience 48:466-473.

Chen, H., S.A. Mehlenbacher, and D.C. Smith. 2007. Hazelnut accessions provide new sources of resistance to eastern filbert blight. HortScience 42:466-469.

Ciarmiello, L.F., M.F. Mazzeo, P. Minasi, A. Peluso, A. De Luca, P. Piccirillo, R.A. Siciliano, and V. Carbone. 2014. Analysis of different european hazelnut (*Corylus avellana* L.) cultivars: authentication, phenotypic features, and phenolic profiles. J. Agri. Food Chem. 62:6236-6246.

Colburn, B.C., S.A. Mehlenbacher, V.R. Sathuvalli, and D.C. Smith. 2015. Eastern filbert blight resistance in hazelnut accessions 'Culplá', 'Crvenje', and OSU 495.072. J. Amer. Soc. Hort. Sci. 140:191-200.

Coyne, C.J., S.A. Mehlenbacher, K.B. Johnson, J.N. Pinkerton, and D.C. Smith. 2000. Comparison of two methods to evaluate quantitative resistance to eastern filbert blight in European hazelnut. J. Amer. Soc. Hort. Sci. 125:603-608.

Coyne, C.J., S.A. Mehlenbacher, and D.C. Smith. 1998. Sources of resistance to eastern filbert blight in hazelnut. J. Amer. Soc. Hort. Sci. 123:253-257.

Drumke, J. 1964. A systematic survey of *Corylus* in North America. Univ. of Tenn., Knoxville, TN, PhD diss.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS ONE 6:e19379. DOI 10.1371/journal.pone.0019379

Erdogan, V. and S.A. Mehlenbacher. 2000. Interspecific hybridization in hazelnut (*Corylus*). J. Amer. Soc. Hort. Sci. 125:489-497.

FAOSTAT. 2017. Agricultual production, crops, Food and Agriculture Organization of the United Nations, Rome. June 30, 2017. <http://www.fao.org/faostat/en/#data/QC/visualize>

FAOSTAT. 2019. Agricultual production, crops, Food and Agriculture Organization of the United Nations, Rome. June 30, 2019. <http://www.fao.org/faostat/en/#data/QC/visualize>

Fuller, A.S. 1908. The nut culturist. Orange Judd, New York, NY.

Gleason, H.A. and A. Cronquist. 1963. Manual of vascular plants of northeastern United States and adjacent Canada. Van Nostrand, New York.

Gottwald, T. and H. Cameron. 1980. Disease increase and the dynamics of spread of canker caused by *Anisogramma anomala* in european filbert in the Pacific Northwest. Phytopathol. 70:1087-1092.

Gürcan, K. and S.A. Mehlenbacher. 2010. Development of microsatellite marker loci for european hazelnut (*Corylus avellana* L.) from ISSR fragments. Mol. Breeding 26:551-559.

Gürcan, K., S.A. Mehlenbacher, R. Botta, and P. Boccacci. 2010. Development, characterization, segregation, and mapping of microsatellite markers for European hazelnut (*Corylus avellana* L.) from enriched genomic libraries and usefulness in genetic diversity studies. Tree Genet. Genomes 6:513-531.

Ives, C., V.R. Sathuvalli, B.C. Colburn, and S.A. Mehlenbachers. 2014. Mapping the incompatibility and style color loci in two hazelnut progenies. HortScience 49:250-253.

Johnson, K. and J. Pinkerton. 2002. Eastern filbert blight, p. 44-46. In: J.W. Pscheidt (ed.), Compendium of nut crop diseases in temperate zones. St. Paul, MN, APS Press, Amer. Phyopathol. Soc.

Johnson, K.B., S.A. Mehlenbacher, J.K. Stone, J.W. Pscheidt, and J.N. Pinkertons. 1996. Eastern filbert blight of European hazelnut: It's becoming a manageable disease. Plant Dis. 80:1308-1316.

Julian, J.W., C.F. Seavert, and J.L. Olsen. 2008. Orchard economics: The costs and returns of establishing and producing hazelnuts in the Willamette Valley. Corvallis, Or.: Extension Service, Oregon State University.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079.

Lunde, C.F., S.A. Mehlenbacher, and D.C. Smiths. 2000. Survey of hazelnut cultivars for response to eastern filbert blight inoculation. HortScience 35:729-731.

McAdam, E.L., J.S. Freeman, S.P. Whittock, E.J. Buck, J. Jakse, A. Cerenak, B. Javornik, A. Kilian, C.-H. Wang, and D. Andersen. 2013. Quantitative trait loci in hop (*Humulus lupulus* L.) reveal complex genetic architecture underlying variation in sex, yield and cone chemistry. BMC Genomics 14:360.

Mehlenbacher, S.A. 1994. Genetic improvement of the hazelnut. Acta Hort. 351: 23-38.

Mehlenbacher, S.A., R.N. Brown, E.R. Nouhra, T. Gökirmak, N.V. Bassil, and T.L. Kubisiak. 2006. A genetic linkage map for hazelnut (*Corylus avellana* L.) based on RAPD and SSR markers. Genome 49:122-133.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2008. ‘Sacajawea’hazelnut. HortScience 43:255-257.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2009. 'Yamhill' hazelnut. HortScience 44:845-847.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2011. 'Jefferson' hazelnut. HortScience 46:662-664.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2013. ‘Dorris’ hazelnut. HortScience 48:796-799.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2014. ‘Wepster’hazelnut. HortScience 49:346-349.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2016. ‘McDonald’ hazelnut. HortScience 51:757-760.

Mehlenbacher, S.A., M.M. Thompson, and H.R. Cameron. 1991. Occurrence and inheritance of resistance to eastern filbert blight in Gasaway hazelnut. HortScience 26:410-411.

Melo, A.T., R. Bartaula, and I. Hale. 2016. GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. BMC Bioinformatics 17:29.

Melo, A.T. and I. Hale. 2018. Expanded functionality, increased accuracy, and enhanced speed in the de novo genotyping-by-sequencing pipeline GBS-SNP-CROP. BMC Bioinformatics 35:1783-1785. DOI 10.1093/bioinformatics/bty873

Molnar, T., J. Lombardoni, M. Muehlbauer, J. Honig, S. Mehlenbacher, and J. Capik. 2018. Progress breeding for resistance to eastern filbert blight in the eastern United States. Acta Hort. 1226:79-86.

Molnar, T.J., J. Capik, S. Zhao, and N. Zhang. 2010a. Disease notes: First report of eastern filbert blight on *Corylus avellana* 'gasaway' and 'VR20-11' caused by *anisogramma anomala* in new jersey. Plant Dis. 94:1265.

Molnar, T.J., J. Capik, S. Zhao, and N. Zhang. 2010b. First report of eastern filbert blight on *Corylus avellana* 'Gasaway' and 'VR20-11' caused by *Anisogramma anomala* in New Jersey. Plant Dis. 94:1265-1265.

Molnar, T.J. and J.M. Capik. 2012. Eastern filbert blight susceptibility of american × european hazelnut progenies. HortScience 47:1412-1418.

Molnar, T.J., J.C. Goffreda, and C.R. Funk. 2010c. Survey of *Corylus* resistance to *Anisogramma anomala* from different geographic locations. HortScience 45:832-836.

Molnar, T.J., D.E. Zaurov, J.C. Goffreda, and S.A. Mehlenbacher. 2007. Survey of hazelnut germplasm from Russia and Crimea for response to eastern filbert blight. HortScience 42:51-56.

Muehlbauer, M., T. Molnar, J. Honig, K. Morey, and N. Zhang. 2014. Genetic diversity of *Anisogramma anomala* and its implications for breeding eastern filbert blight resistant hazelnuts. (Abstr.) HortScience 49(9):S134.

Muehlbauer, M., J. Tobia, J. Honig, B. Hillman, N. Zhang, K.M. Gold, and T. Molnar. 2018. Population differentiation within *Anisogramma anomala* in North America. Phytopathol. DOI 10.1094/PHYTO-06-18-0209-R

Petriccione, M., L.F. Ciarmiello, P. Boccacci, A. De Luca, and P. Piccirillo. 2010. Evaluation of 'Tonda di Giffoni'hazelnut (*Corylus avellana* L.) clones. Scientia Hort. 124:153-158.

Pinkerton, J.N., K.B. Johnson, S.A. Mehlenbacher, and J.W. Pscheidt. 1993. Susceptibility of european hazelnut clones to eastern filbert blight. Plant Dis. 77:261-266.

Pinkerton, J.N., K.B. Johnson, K.M. Theiling, and J.A. Griesbach. 1992. Distribution and characeteristics of the eastern filbert blight epidemic in Western Oregon. Plant Dis. 76:1179-1182.

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.-L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PloS one 7:e32253. DOI 10.1371/journal.pone.0032253

Sathuvalli, V.R., H. Chen, S.A. Mehlenbacher, and D.C. Smith. 2011a. DNA markers linked to eastern filbert blight resistance in "Ratoli" hazelnut (*Corylus avellana* L.). Tree Genet. Genomes 7:337-345.

Sathuvalli, V.R. and S.A. Mehlenbacher. 2011. A bacterial artificial chromosome library for 'Jefferson' hazelnut and identification of clones associated with eastern filbert blight resistance and pollen-stigma incompatibility. Genome 54:862-867.

Sathuvalli, V.R. and S.A. Mehlenbacher. 2012. Characterization of American hazelnut (*Corylus americana*) accessions and *Corylus americana* × *Corylus avellana* hybrids using microsatellite markers. Genet. Resources and Crop Evol. 59:1055-1075.

Sathuvalli, V.R., S.A. Mehlenbacher, and D.C. Smith. 2011b. DNA markers linked to eastern filbert blight resistance from a hazelnut selection from the Republic of Georgia. J. Amer. Soc. Hort. Sci. 136:350-357.

Sathuvalli, V.R., S.A. Mehlenbacher, and D.C. Smith. 2012. Identification and mapping of DNA markers linked to eastern filbert blight resistance from OSU 408.040 hazelnut. HortScience 47:570-573.

Studer, B., B. Boller, D. Herrmann, E. Bauer, U.K. Posselt, F. Widmer, and R. Kölliker. 2006. Genetic mapping reveals a single major QTL for bacterial wilt resistance in italian ryegrass (*Lolium multiflorum* Lam.). Theor. Appl. Genet. 113:661-671.

Tobia, J., M. Muehlbauer, J. Honig, J. Pscheidt, and T. Molnar. 2017. Cluster analysis of *Anisogramma anomala* isolates collected from the Pacific Northwest and New Jersey. (Abstr.) Phytopathol. 107:S5.125. DOI 10.1094/PHYTO-107-12-S5.125

Torello Marinoni, D., N. Valentini, E. Portis, A. Acquadro, C. Beltramo, S.A. Mehlenbacher, T.C. Mockler, E.R. Rowley, and R. Botta. 2018. High density SNP mapping and QTL analysis for time of leaf budburst in *Corylus avellana* L. PLoS ONE 13:e0195408. DOI 10.1371/journal.pone.0195408

U.S. Department of Agriculture. 2018. Noncitrus Fruits and Nuts 2017 Summary, National Agricultural Statistics Service. U.S. Dept. Agr., Washington, D.C

van Heerden, C.J., P. Burger, A. Vermeulen, and R. Prins. 2014. Detection of downy and powdery mildew resistance QTL in a 'Regent'×'RedGlobe'population. Euphytica 200:281-295. DOI 10.1007/s10681-014-1167-4

van Ooijen, J. 2006. JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands.

van Ooijen, J.. 2009. MapQTL 6. Software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma B.V., Wageningen, Netherlands.

Voorrips, R. 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. J. Heredity 93:77-78.

Weschcke, C. 1953. Growing nuts in the north. Webb, St. Paul, MN.

Zyprian, E., I. Ochßner, F. Schwander, S. Šimon, L. Hausmann, M. Bonow-Rex, P. Moreno-Sanz, M.S. Grando, S. Wiedemann-Merdinoglu, and D. Merdinoglu. 2016. Quantitative trait loci affecting pathogen resistance and ripening of grapevines. Mol. Genet. Genomics 291:1573-1594.

APPENDIX A: SUPPLEMENTARY TABLES AND FIGURES

Table A.1. Suitability parameters of cultivated hazelnut (adapted from KIDD et al. 2015).

Suitability class	Soil depth (cm)	pH (0-15cm)	EC (ds/m) (0-15cm)	Clay % (0-15 cm)	Soil drainage class	Stone % (>20 cm)	Rainfall, mean August (mm)
Well suited	>50	6.5–6.599	<0.15	30–50	Well to moderate	<10	<80
Suited	40–50	5.6–6.499	<0.15	30–50	Imperfect	10–20	<50
Marginally suited	30–40	6.6–7.199	<0.15	30–50	Imperfect	10–20	<50
Unsuited	<30	<5.599 or >7.2	<0.15	>50 or <10	Poor to very poor	>20	>50

The suitability analysis was performed for Illinois using the National Soil Survey Geographic Database (gSSURGO) that provides soil trait data at 10 × 10 m resolution (SSURGO, 2017). Only soil characteristics were used in calculating suitability, with rainfall later included with soil traits for a multivariate statistical analysis. For the geospatial analysis, each soil trait was classified into discrete suitability classes, where a given map unit is assigned a value of 1, 0, −1, or −100 based on how that soil trait relates to successful hazelnut production. Suitability was calculated as the unweighted sum of the classified soil traits. If one of the soil traits was classified as Unsuited (−100), then the entire map unit was identified as unsuitable.

The entirety of Illinois is included in the spatial analysis, in that all map units with available data for Illinois from gSSURGO were included. The suitability of a given map unit for actual hazelnut production is subject to whether it is located in an agricultural field.

Table A.2. Disease ratings for the subset of the *Corylus americana* germplasm collection in winter 2018 after seven seasons of growth under high disease pressure from *Anisogramma anomala*.

				Disease rating[w,v]						
Seed lot[z]	County[y]	State[x]	Total no. of plants[v]	Ave.	0	1	2	3	4	5
10520	Jefferson	ME	2	0	2	0	0	0	0	0
10530	Grantsburg	WI	15	0	15	0	0	0	0	0
10532	Nantucket	CT	16	0	16	0	0	0	0	0
10537	Willow Hill	PA	25	1.60	14	0	0	5	5	1
10542	Adams	WI	23	0	23	0	0	0	0	0
10543	Pike	IL	1	0	1	0	0	0	0	0
10556	Richardson	NE	5	0	5	0	0	0	0	0
10557	Richardson	NE	5	0	5	0	0	0	0	0
10559	Richardson	NE	4	0	4	0	0	0	0	0
10560	Richardson	NE	5	0	5	0	0	0	0	0
10561	Richardson	NE	4	0	4	0	0	0	0	0
10562	Richardson	NE	3	0	3	0	0	0	0	0
10563	Richardson	NE	5	0	5	0	0	0	0	0
10564	Richardson	NE	3	0	3	0	0	0	0	0
10565	Richardson	NE	4	0	4	0	0	0	0	0
10566	Richardson	NE	4	0	4	0	0	0	0	0
10567	Richardson	NE	2	0	2	0	0	0	0	0
10568	Richardson	NE	3	0	3	0	0	0	0	0
10569	Richardson	NE	4	0	4	0	0	0	0	0
10570	Richardson	NE	3	0	3	0	0	0	0	0
10571	Richardson	NE	5	0	5	0	0	0	0	0
10572	Richardson	NE	5	0	5	0	0	0	0	0
10573	Richardson	NE	4	0	4	0	0	0	0	0
10574	Richardson	NE	5	0	5	0	0	0	0	0
10575	Richardson	NE	1	0	1	0	0	0	0	0
10576	Richardson	NE	3	0	3	0	0	0	0	0
10577	Richardson	NE	2	0	2	0	0	0	0	0
10578	Richardson	NE	4	1.00	3	0	0	0	1	0
10579	Richardson	NE	4	0	4	0	0	0	0	0
10581	Centre	PA	9	0	9	0	0	0	0	0
11554	Cream Ridge	NJ	6	1.83	2	1	1	0	2	0
11563	Aiken	MN	22	0	22	0	0	0	0	0
11564	Butler	PA	3	0	3	0	0	0	0	0
11565	Mississippi River	MN	21	0	21	0	0	0	0	0
11567	Menard	IL	25	0.04	23	0	1	1	0	0
11568	Casey	KY	23	0.04	22	1	0	0	0	0
11569	Belmont	OH	24	0	24	0	0	0	0	0

Table A.2. (cont.d)

11570	Mercer	NJ	24	0.04	23	1	0	0	0	0
11571	Aiken	MN	22	0	22	0	0	0	0	0
11574	Hughesville	PA	20	0	20	0	0	0	0	0
11575	Titusville	PA	15	0.13	14	0	1	0	0	0
11577	Prentice	WI	24	0	24	0	0	0	0	0
11578A	Morris	NJ	10	0.30	9	0	0	1	0	0
11579	Stacy	MN	24	0	24	0	0	0	0	0
11583	Ethel	MO	7	0	7	0	0	0	0	0
11585	Oneida	WI	1	0	1	0	0	0	0	0
11587	Pittsfield	MI	14	0.08	13	1	0	0	0	0
11591	Madison	WI	23	0	23	0	0	0	0	0
11593	West Subury	PA	8	0	8	0	0	0	0	0
11594	Union	IA	1	1	0	1	0	0	0	0
11601	Livonia	MN	10	0	10	0	0	0	0	0
11602	Calhoun	MI	7	0	7	0	0	0	0	0
11603	Arpin	WI	24	0.04	23	1	0	0	0	0
11604	Venturini	VA	7	0	7	0	0	0	0	0
11605	Nantucket	CT	3	0	3	0	0	0	0	0
11606	Nantucket	CT	9	0	9	0	0	0	0	0
11607	Clarke	IA	21	0	21	0	0	0	0	0
11608	Crosslake	MN	3	0	3	0	0	0	0	0
11610	Mansfield	OH	4	1.75	1	0	2	1	0	0
11612	East Moline	IL	4	0	4	0	0	0	0	0
11613	Kansas City	MO	24	0.08	23	0	1	0	0	0
11614	Pt. Pleasant	WV	5	0	5	0	0	0	0	0
Totals			616		587	6	6	8	8	1

[z]Seed lots received by cooperators were assigned numbers by the Rutgers University breeding program.

[y]The county in which the *C. americana* seeds were collected.

[x]The state in which the *C. americana* seeds were collected.

[w]Number of living seedlings at the Rutgers University Research and Extension Center in Cream Ridge, NJ.

[v]Disease ratings correspond to phenotypes: 0 = no visible EFB, 1 = a single canker, 2 = multiple cankers on a single branch, 3 = multiple branches with cankers, 4 = more than 50% of branches have cankers, and 5 = all branches have cankers or the plant has died from EFB.

Table A.3. Accession codes and group and sub-group assignment per the cluster method.

	Accession code			Group assignment			
NJ tree pos.[a]	Seedlot	Field location	State	Structure	DAPC[b]	Consensus[c]	Core Set
1	11603	10086	WI	Admixed	Admixed	14	1
2	11613	10021	MO	3 (Blue)	3 (Blue)	14	-
3	11593	09120	PA	3 (Blue)	12 (Red)	14	-
4	11614	09111	MN	3 (Blue)	12 (Red)	14	-
5	11579	10138	MN	Admixed	3 (Blue)	14	-
6	11606	09135	MA	3 (Blue)	12 (Red)	14	-
7	10569	11125	NE	3 (Blue)	3 (Blue)	14	-
8	10567	11180	NE	3 (Blue)	3 (Blue)	14	2
9	10560	11175	NE	Admixed	12 (Red)	14	-
10	11571	11037	MN	3 (Blue)	Admixed	10	-
11	11584	08045	WV	3 (Blue)	13 (Gold)	10	-
12	11567	10055	IL	3 (Blue)	13 (Gold)	10	-
13	10576	11162	NE	3 (Blue)	13 (Gold)	10	-
14	10571	11145	NE	3 (Blue)	13 (Gold)	10	-
15	11566	09052	MN	3 (Blue)	13 (Gold)	10	-
16	11575	09011	PA	3 (Blue)	3 (Blue)	10	-
17	11571	11034	MN	3 (Blue)	3 (Blue)	10	-
18	11579	10122	MN	3 (Blue)	3 (Blue)	10	-
19	11599	08032	MN	3 (Blue)	3 (Blue)	10	-
20	11577	10001	WI	3 (Blue)	3 (Blue)	10	-
21	11599	08030	MN	3 (Blue)	13 (Gold)	10	-
22	11568	11065	KY	3 (Blue)	13 (Gold)	10	-

Table A.3. (cont.d)

23	11568	11048	KY	3 (Blue)	13 (Gold)	10	-
24	11566	09056	MN	3 (Blue)	2 (Dark purple)	9	-
25	11570	10163	NJ	3 (Blue)	13 (Gold)	9	-
26	11602	11070	MI	3 (Blue)	2 (Dark purple)	9	-
27	11591	10100	WI	3 (Blue)	2 (Dark purple)	9	-
28	11601	09107	MN	3 (Blue)	13 (Gold)	9	-
29	11587	09028	MI	3 (Blue)	2 (Dark purple)	9	-
30	11565	10213	MN	3 (Blue)	2 (Dark purple)	9	-
31	11604	09068	VA	3 (Blue)	2 (Dark purple)	9	3
32	10578	11148	NE	3 (Blue)	13 (Gold)	9	-
33	11574	09191	PA	3 (Blue)	13 (Gold)	9	4
34	11575	09005	PA	3 (Blue)	3 (Blue)	9	-
35	11602	11069	MI	3 (Blue)	3 (Blue)	13	-
36	10560	11174	NE	3 (Blue)	3 (Blue)	13	5
37	10526	12004	-	3 (Blue)	3 (Blue)	13	6
38	10532	12108	MA	3 (Blue)	3 (Blue)	13	-
39	10537	12028	PA	3 (Blue)	3 (Blue)	13	-
40	11565	09212	MN	3 (Blue)	3 (Blue)	13	-
41	11568	11047	KY	3 (Blue)	3 (Blue)	13	-
42	11591	10103	WI	3 (Blue)	3 (Blue)	13	-
43	11565	10194	MN	3 (Blue)	3 (Blue)	13	-
44	10537	12027	PA	3 (Blue)	3 (Blue)	13	-
45	11583	09080	MO	3 (Blue)	3 (Blue)	13	-
46	11591	10098	WI	3 (Blue)	3 (Blue)	13	-

Table A.3. (cont.d)

47	11604	09065	VA	3 (Blue)	3 (Blue)	13	-
48	10526	12008	-	3 (Blue)	3 (Blue)	13	-
49	10526	12009	-	3 (Blue)	3 (Blue)	13	-
50	11601	09099	MN	3 (Blue)	3 (Blue)	13	-
51	10577	11156	NE	3 (Blue)	3 (Blue)	13	-
52	11591	10106	WI	3 (Blue)	3 (Blue)	13	-
53	11605	11080	MA	3 (Blue)	3 (Blue)	13	7
54	11577	11009	WI	3 (Blue)	3 (Blue)	13	8
55	11587	09025	MI	3 (Blue)	3 (Blue)	13	-
56	11577	11014	WI	3 (Blue)	3 (Blue)	13	-
57	11565	10203	Mn	3 (Blue)	3 (Blue)	13	-
58	11579	10130	MN	3 (Blue)	3 (Blue)	13	-
59	10542	12056	WI	3 (Blue)	3 (Blue)	13	-
60	11571	11027	MN	3 (Blue)	3 (Blue)	13	-
61	11574	09201	PA	3 (Blue)	3 (Blue)	13	-
62	11569	09174	OH	3 (Blue)	3 (Blue)	13	-
63	11610	09155	OH	3 (Blue)	3 (Blue)	13	9
64	11568	11049	KY	3 (Blue)	3 (Blue)	13	-
65	11574	09189	PA	3 (Blue)	3 (Blue)	13	-
66	11569	09161	OH	3 (Blue)	3 (Blue)	13	-
67	11565	10207	Mn	3 (Blue)	3 (Blue)	13	-
68	11577	11012	WI	3 (Blue)	3 (Blue)	13	-
69	11614	09110	MN	3 (Blue)	3 (Blue)	13	-
70	11565	10205	MN	3 (Blue)	3 (Blue)	13	-

Table A.3. (cont.d)

71	11606	09140	MA	3 (Blue)	3 (Blue)	13	-
72	11568	11063	KY	3 (Blue)	3 (Blue)	13	-
73	10581	12046	PA	Admixed	Admixed	13	-
74	11575	09001	PA	3 (Blue)	3 (Blue)	13	-
75	11579	10131	MN	3 (Blue)	11 (Tan)	6	-
76	11587	09041	MI	3 (Blue)	14 (Light blue)	6	-
77	11566	09062	MN	3 (Blue)	14 (Light blue)	6	-
78	11599	08029	MN	3 (Blue)	14 (Light blue)	6	10
79	10526	11003	-	3 (Blue)	14 (Light blue)	6	11
80	10564	11117	NE	3 (Blue)	14 (Light blue)	6	-
81	11578A	09146	NJ	3 (Blue)	14 (Light blue)	6	-
82	11570	10145	NJ	3 (Blue)	14 (Light blue)	6	-
83	10564	11120	NE	3 (Blue)	14 (Light blue)	6	-
84	11594	09016	IA	3 (Blue)	14 (Light blue)	6	12
85	11571	11026	MN	3 (Blue)	14 (Light blue)	6	13
86	11603	10066	WI	3 (Blue)	12 (Red)	6	-
87	11563	10177	MN	3 (Blue)	11 (Tan)	6	-
88	11575	08003	PA	3 (Blue)	14 (Light blue)	6	-
89	11584	08037	WV	3 (Blue)	11 (Tan)	15	-
90	11564	09038	PA	3 (Blue)	11 (Tan)	15	-
91	11578A	09141	NJ	3 (Blue)	11 (Tan)	15	-
92	11578A	09142	NJ	3 (Blue)	11 (Tan)	15	14
93	11570	10152	NJ	3 (Blue)	Admixed	15	-
94	11570	10148	NJ	3 (Blue)	3 (Blue)	15	-

Table A.3. (cont.d)

95	11591	10099	WI	3 (Blue)	3 (Blue)	15	-
96	11566	09054	MN	3 (Blue)	13 (Gold)	15	15
97	10563	11128	NE	3 (Blue)	13 (Gold)	15	-
98	11604	09069	VA	3 (Blue)	13 (Gold)	15	-
99	11613	10033	MO	3 (Blue)	3 (Blue)	15	-
100	11578A	09149	NJ	3 (Blue)	3 (Blue)	15	-
101	11579	10124	MN	3 (Blue)	3 (Blue)	15	-
102	11567	10052	IL	3 (Blue)	3 (Blue)	15	-
103	11607	11103	IA	3 (Blue)	3 (Blue)	15	-
104	11603	10071	WI	3 (Blue)	11 (Tan)	12	16
105	11566	09055	MN	3 (Blue)	11 (Tan)	12	-
106	11608	10063	MN	3 (Blue)	11 (Tan)	12	-
107	11603	10069	WI	3 (Blue)	11 (Tan)	12	-
108	11604	09070	VA	3 (Blue)	11 (Tan)	12	-
109	11567	10040	IL	3 (Blue)	11 (Tan)	12	-
110	11587	09037	MI	3 (Blue)	11 (Tan)	12	-
111	10565	11134	NE	3 (Blue)	3 (Blue)	11	-
112	11599	08036	MN	3 (Blue)	3 (Blue)	11	-
113	11563	10175	MN	3 (Blue)	3 (Blue)	11	17
114	11563	10174	MN	3 (Blue)	3 (Blue)	11	-
115	11577	11008	WI	3 (Blue)	3 (Blue)	11	-
116	11599	08035	MN	3 (Blue)	3 (Blue)	11	-
117	10526	11002	-	3 (Blue)	3 (Blue)	11	-
118	11579	10118	MN	3 (Blue)	3 (Blue)	11	-

Table A.3. (cont.d)

119	10557	11158	NE	3 (Blue)	3 (Blue)	11	-
120	11591	10109	WI	3 (Blue)	3 (Blue)	11	-
121	11610	09153	OH	3 (Blue)	3 (Blue)	11	-
122	10570	11137	NE	3 (Blue)	3 (Blue)	11	-
123	11593	09126	PA	3 (Blue)	3 (Blue)	11	-
124	11587	09023	MI	3 (Blue)	3 (Blue)	11	-
125	11599	08034	MN	Admixed	12 (Red)	18	-
126	10579	11170	NE	Admixed	12 (Red)	18	-
127	11575	08005	PA	Admixed	12 (Red)	18	-
128	11607	11094	IA	Admixed	12 (Red)	18	-
129	11603	10090	WI	Admixed	12 (Red)	18	-
130	11607	11101	IA	Admixed	12 (Red)	18	18
131	11604	09067	VA	Admixed	12 (Red)	18	19
132	10581	12040	PA	Admixed	15 (Pink)	18	-
133	10581	12037	PA	Admixed	12 (Red)	18	-
134	10537	12023	PA	Admixed	15 (Pink)	18	-
135	10526	12003	-	Admixed	12 (Red)	18	20
136	10581	12039	PA	Admixed	12 (Red)	18	-
137	10537	12018	PA	Admixed	15 (Pink)	18	-
138	10532	12100	MA	1 (Red)	12 (Red)	18	-
139	10532	12102	MA	1 (Red)	9 (Orange)	4	-
140	10571	11143	NE	1 (Red)	9 (Orange)	4	21
141	10576	11163	NE	1 (Red)	9 (Orange)	4	22
142	10542	12064	WI	1 (Red)	9 (Orange)	4	-

Table A.3. (cont.d)

143	10526	12005	-	1 (Red)	9 (Orange)	4	-
144	11605	11079	MA	1 (Red)	9 (Orange)	4	-
145	11593	09122	PA	Admixed	12 (Red)	8	-
146	11569	09165	OH	Admixed	12 (Red)	8	-
147	11565	10197	MN	1 (Red)	5 (Sea green)	8	-
148	11593	09116	PA	Admixed	1 (Plum)	8	-
149	11602	11071	MI	1 (Red)	5 (Sea green)	8	-
150	10532	12104	MA	1 (Red)	12 (Red)	8	-
151	10561	11113	NE	1 (Red)	5 (Sea green)	8	-
152	11566	09059	MN	1 (Red)	5 (Sea green)	8	-
153	11575	09008	PA	3 (Blue)	5 (Sea green)	8	-
154	10560	11173	NE	Admixed	1 (Plum)	7	-
155	11567	10044	IL	3 (Blue)	1 (Plum)	7	23
156	10563	11130	NE	Admixed	1 (Plum)	7	24
157	11567	10058	IL	Admixed	1 (Plum)	7	25
158	11587	09029	MI	Admixed	1 (Plum)	7	26
159	10565	11133	NE	Admixed	12 (Red)	8	-
160	11569	09160	OH	Admixed	5 (Sea green)	8	-
161	10561	11115	NE	Admixed	5 (Sea green)	8	-
162	11568	11060	KY	Admixed	5 (Sea green)	8	-
163	11613	10014	MO	Admixed	5 (Sea green)	8	-
164	11570	10160	NJ	Admixed	12 (Red)	8	-
165	11603	10068	WI	Admixed	5 (Sea green)	8	-
166	10561	11114	NE	2 (Green)	5 (Sea green)	8	-

Table A.3. (cont.d)

167	10570	11141	NE	2 (Green)	6 (Green)	2	27
168	10573	12168	NE	2 (Green)	6 (Green)	2	-
169	11607	11091	IA	2 (Green)	6 (Green)	2	28
170	10559	11110	NE	2 (Green)	6 (Green)	2	29
171	11607	11089	IA	2 (Green)	6 (Green)	2	-
172	10526	12007	-	2 (Green)	6 (Green)	2	30
173	10579	11171	NE	2 (Green)	6 (Green)	2	-
174	10532	12103	MA	2 (Green)	6 (Green)	2	-
175	10542	12058	WI	Admixed	12 (Red)	2	31
176	11575	09009	PA	Admixed	12 (Red)	18	32
177	11608	10065	MN	Admixed	12 (Red)	18	33
178	11567	10042	IL	Admixed	12 (Red)	18	-
179	11583	09077	MO	Admixed	12 (Red)	18	-
180	11584	08042	WV	Admixed	12 (Red)	18	-
181	11613	10031	MO	Admixed	15 (Pink)	18	-
182	11612	09093	IL	1 (Red)	15 (Pink)	18	-
183	10579	11169	NE	Admixed	12 (Red)	18	-
184	10537	12012	PA	Admixed	15 (Pink)	18	-
185	10532	12107	MA	Admixed	15 (Pink)	18	-
186	10542	12060	WI	Admixed	15 (Pink)	18	-
187	11579	10120	MN	Admixed	15 (Pink)	18	-
188	11563	10181	MN	1 (Red)	15 (Pink)	18	-
189	10526	11004	-	1 (Red)	15 (Pink)	18	-
190	11577	11010	WI	1 (Red)	15 (Pink)	18	-

Table A.3. (cont.d)

191	11570	10144	NJ	1 (Red)	15 (Pink)	18	34
192	11574	09192	PA	Admixed	15 (Pink)	18	-
193	11568	11054	KY	Admixed	15 (Pink)	18	-
194	11603	10082	WI	Admixed	15 (Pink)	18	-
195	11563	10171	MN	Admixed	7 (Yellow)	16	-
196	11571	11019	MN	Admixed	7 (Yellow)	16	35
197	10542	12067	WI	Admixed	7 (Yellow)	16	-
198	11568	11052	KY	Admixed	7 (Yellow)	16	-
199	10571	11144	NE	Admixed	7 (Yellow)	16	-
200	11606	09132	MA	1 (Red)	7 (Yellow)	16	-
201	11566	09078	MN	1 (Red)	7 (Yellow)	16	36
202	10532	12113	MA	Admixed	15 (Pink)	19	37
203	11607	11086	IA	Admixed	15 (Pink)	19	-
204	10542	12057	WI	Admixed	15 (Pink)	19	-
205	11565	10199	MN	Admixed	15 (Pink)	19	-
206	11578A	09148	NJ	Admixed	15 (Pink)	19	-
207	11565	10201	MN	Admixed	15 (Pink)	19	-
208	11583	09076	MO	1 (Red)	15 (Pink)	19	-
209	10537	12021	PA	1 (Red)	15 (Pink)	19	-
210	11607	11095	IA	Admixed	15 (Pink)	19	38
211	11567	10039	IL	1 (Red)	15 (Pink)	19	-
212	11603	10077	WI	1 (Red)	15 (Pink)	19	-
213	11587	09033	MI	1 (Red)	15 (Pink)	19	-
214	10542	12052	WI	Admixed	15 (Pink)	19	-

Table A.3. (cont.d)

215	11579	10140	MN	Admixed	7 (Yellow)	16	-
216	11577	11016	WI	Admixed	7 (Yellow)	16	-
217	11577	11015	WI	Admixed	7 (Yellow)	16	-
218	11591	10095	WI	Admixed	7 (Yellow)	16	-
219	11578A	09150	NJ	Admixed	12 (Red)	19	-
220	10577	11152	NE	Admixed	15 (Pink)	19	39
221	11579	10134	MN	1 (Red)	15 (Pink)	19	-
222	11564	09039	PA	1 (Red)	15 (Pink)	19	-
223	10557	11159	NE	1 (Red)	15 (Pink)	19	-
224	11605	11077	MA	1 (Red)	15 (Pink)	19	-
225	11599	08028	MN	1 (Red)	15 (Pink)	19	-
226	11567	10046	IL	1 (Red)	15 (Pink)	19	40
227	11587	09036	MI	1 (Red)	15 (Pink)	19	-
228	11571	11023	MN	Admixed	15 (Pink)	19	-
229	11613	10016	MO	Admixed	15 (Pink)	19	41
230	11584	08043	WV	Admixed	7 (Yellow)	19	42
231	11599	08033	MN	1 (Red)	8 (Brown)	3	43
232	11563	10172	MN	1 (Red)	8 (Brown)	3	-
233	11614	09112	MN	1 (Red)	8 (Brown)	3	-
234	11591	10096	WI	1 (Red)	8 (Brown)	3	44
235	11571	11039	MN	1 (Red)	8 (Brown)	3	45
236	11600	08026	NE	1 (Red)	8 (Brown)	3	-
237	10542	12051	WI	Admixed	15 (Pink)	8	-
238	10559	11107	NE	Admixed	5 (Sea green)	8	-

Table A3. (cont.d)

239	10537	12017	PA	Admixed	5 (Sea green)	8	-
240	10559	11109	NE	Admixed	5 (Sea green)	8	-
241	11607	11088	IA	Admixed	5 (Sea green)	8	-
242	10526	12006	-	Admixed	5 (Sea green)	8	-
243	11607	11092	IA	Admixed	5 (Sea green)	8	-
244	11603	10079	WI	Admixed	1 (Plum)	8	46
245	11613	10027	MO	Admixed	1 (Plum)	8	47
246	11567	10054	IL	Admixed	12 (Red)	17	-
247	11603	10070	WI	Admixed	12 (Red)	17	-
248	11602	11074	MI	Admixed	12 (Red)	17	-
249	10537	12014	PA	Admixed	12 (Red)	17	-
250	11568	11055	KY	Admixed	12 (Red)	17	-
251	10564	11118	NE	3 (Blue)	4 (Dark red)	1	48
252	11564	09040	PA	3 (Blue)	4 (Dark red)	1	-
253	11575	09004	PA	3 (Blue)	4 (Dark red)	1	-
254	11575	09002	PA	3 (Blue)	4 (Dark red)	1	-
255	11565	10195	MN	Admixed	4 (Dark red)	1	-
256	11575	08010	PA	Admixed	4 (Dark red)	1	-
257	11577	11011	WI	Admixed	4 (Dark red)	1	-
258	11604	09064	VA	Admixed	4 (Dark red)	1	49
259	10570	11138	NE	Admixed	10 (Purple)	5	-
260	10576	11166	NE	Admixed	10 (Purple)	5	-
261	10565	11147	NE	Admixed	10 (Purple)	5	-
262	11570	10147	NJ	1 (Red)	10 (Purple)	5	-

Table A.3. (cont.d)

263	11568	11056	KY	1 (Red)	10 (Purple)	5	-
264	10526	12002	-	1 (Red)	10 (Purple)	5	-
265	10526	12001	-	Admixed	10 (Purple)	5	-
266	11607	11093	IA	1 (Red)	10 (Purple)	5	-
267	11571	11021	MN	1 (Red)	10 (Purple)	5	50
268	11613	10025	MO	Admixed	10 (Purple)	5	-
269	10581	12041	PA	1 (Red)	10 (Purple)	5	-
270	10578	11151	NE	1 (Red)	10 (Purple)	5	-
271	11565	09210	MN	1 (Red)	10 (Purple)	5	-
272	11601	09102	MN	1 (Red)	10 (Purple)	5	51

[a]Number corresponds to the position in the neighbor-joining tree, as depicted in Table S2.

[b]Discriminant analysis of principal components (DAPC)

[c]Consensus groups were derived by comparing group assignment between the DAPC and neighbor-joining methods.

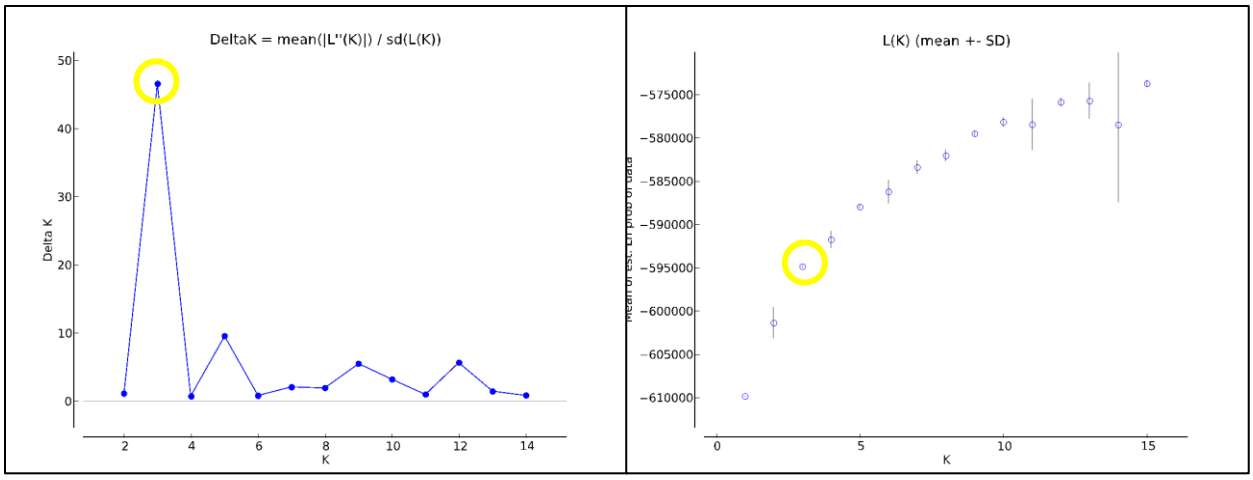


Figure A.1. Plotted Δ*K* values (left) from STRUCTURE HARVESTER (Evanno et al., 2005), where the optimal number of clusters was determined by the highest delta *K* (*K* = 3). Small peaks at *K* = 5, 9, and 12 suggest sub-clusters. Plotted means (and standard deviations) of Ln probability for each tested *K* is depicted (right). Low standard deviation at *K* = 3 supports Δ*K* = 3.

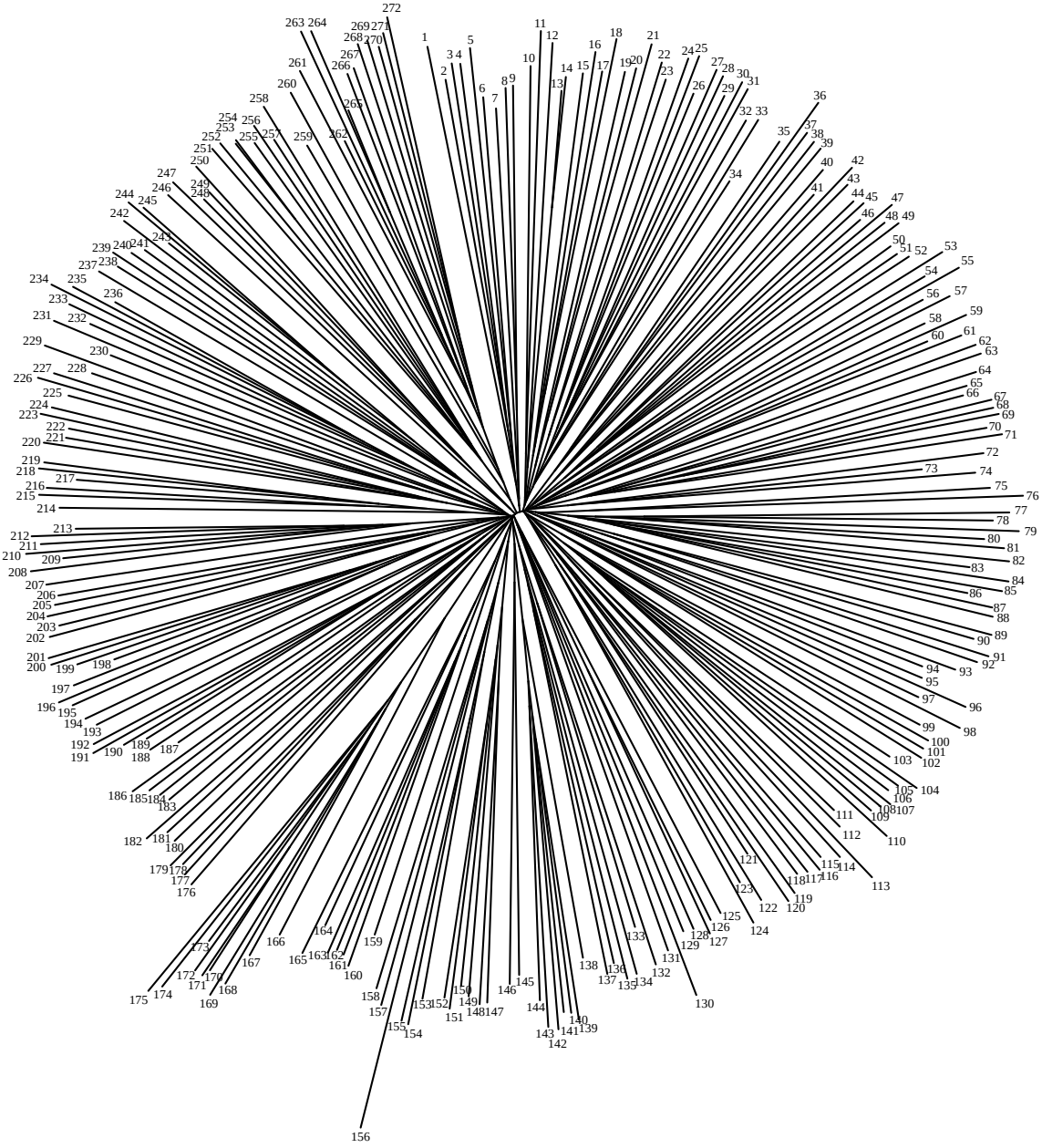


Figure A.2. The neighbor-joining dendrogram with labels that correspond to Table A.3.

Figure A.3. OSU 403.040 *C. americana* × 'Tonda di Giffoni' *C. avellana* LG1

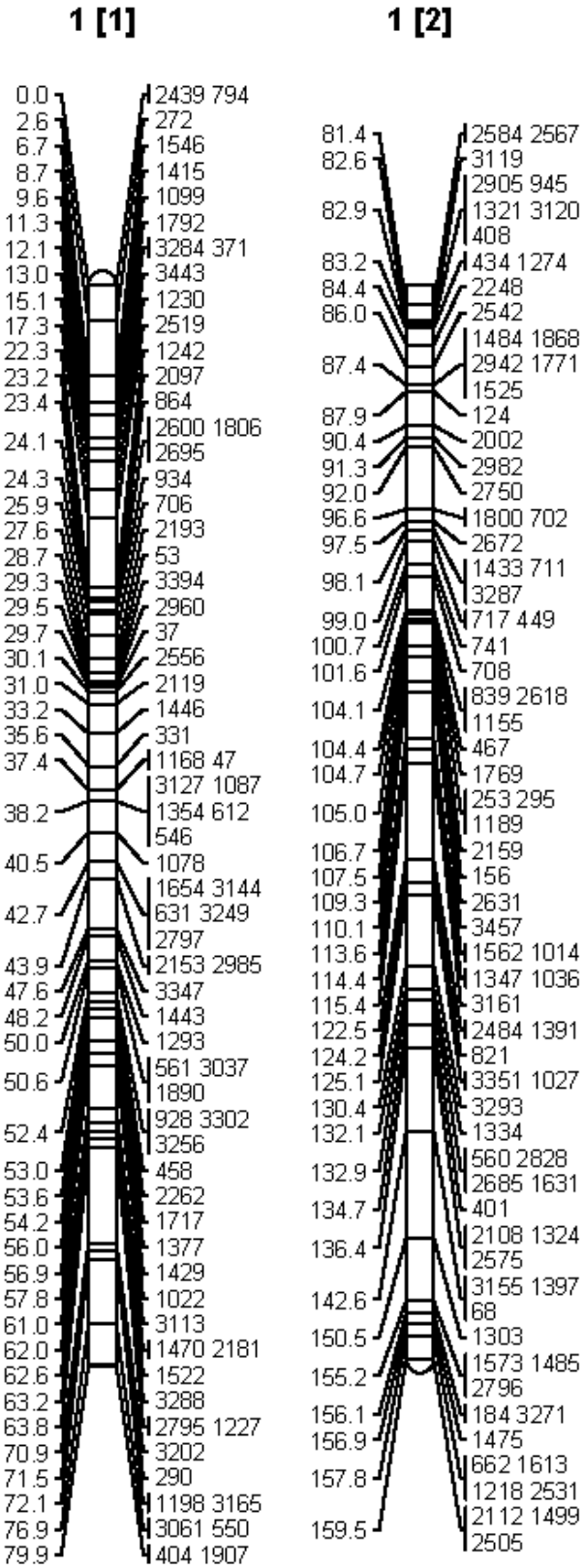


Figure A.4. OSU 403.040 *C. americana* × 'Tonda di Giffoni' *C. avellana* LG2

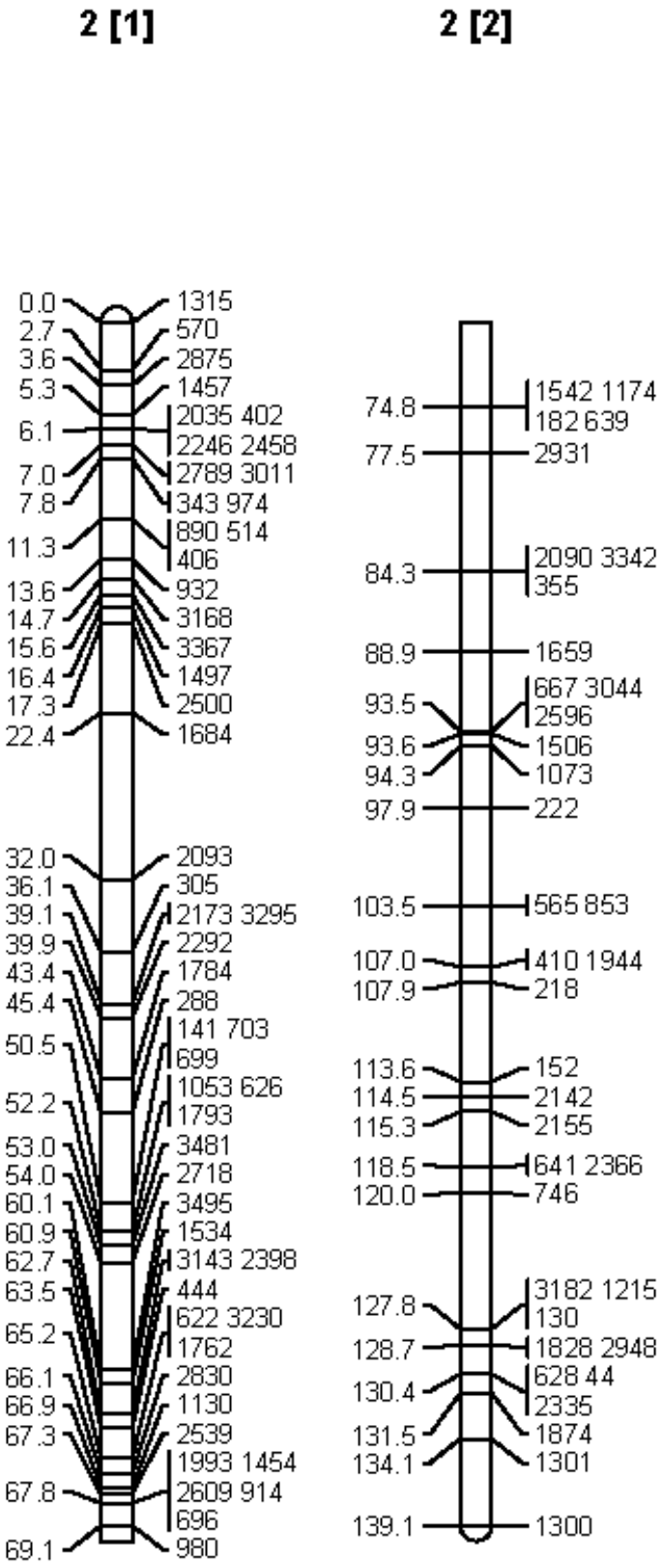


Figure A.5. OSU 403.040 *C. americana* × 'Tonda di Giffoni' *C. avellana* LG3

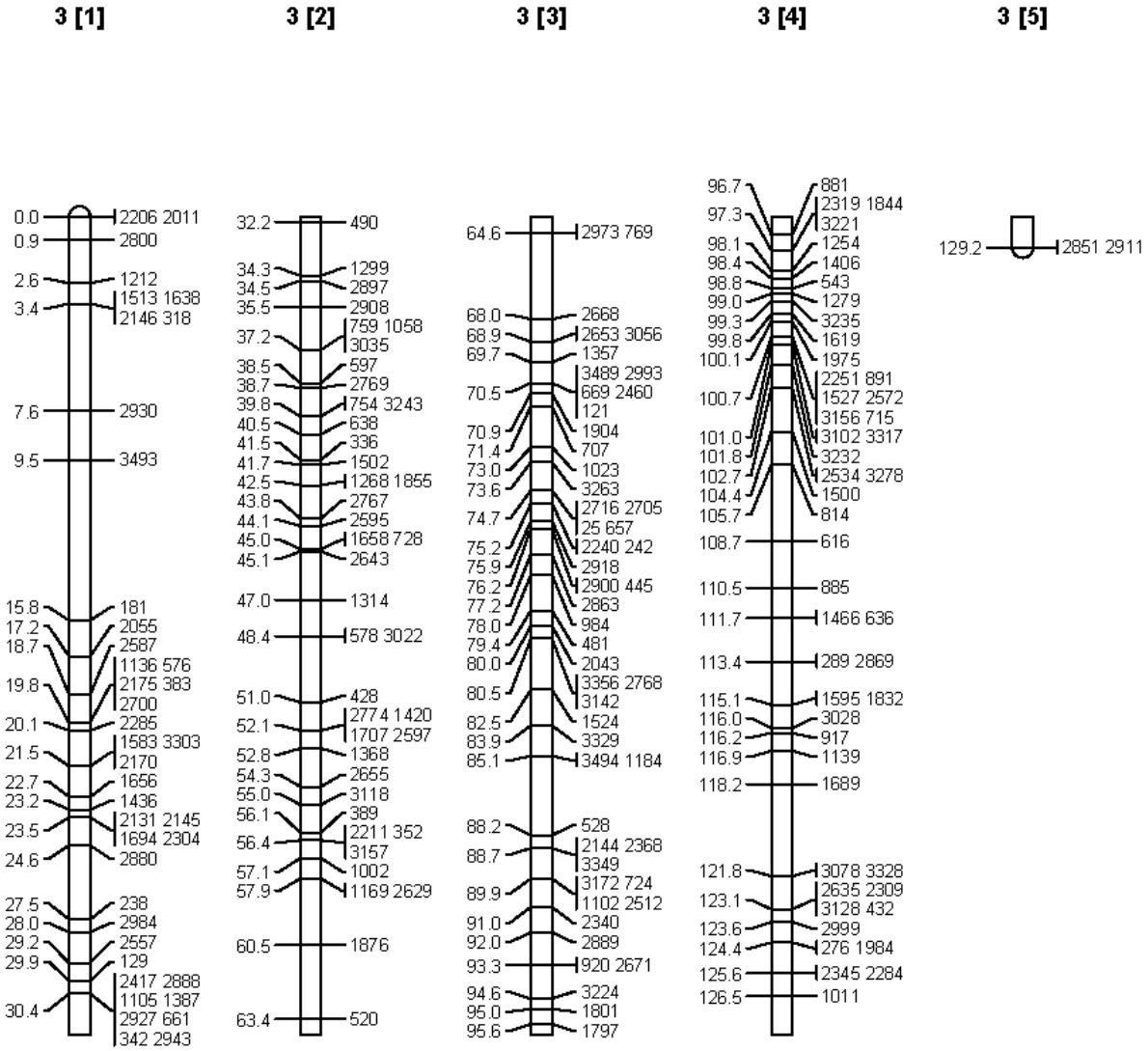


Figure A.6. OSU 403.040 *C. americana* × 'Tonda di Giffoni' *C. avellana* LG4

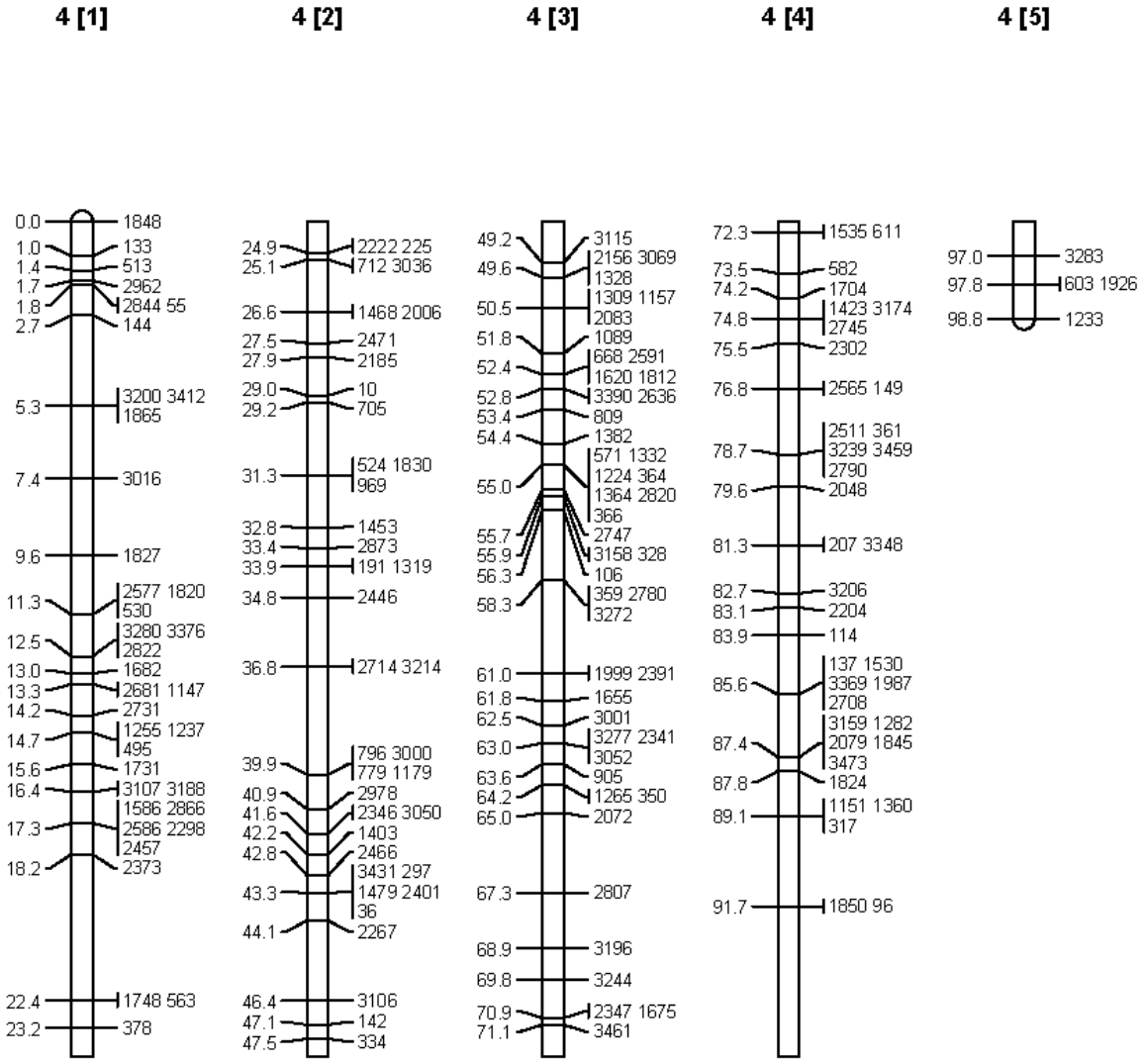


Figure A.7. OSU 403.040 *C. americana* × 'Tonda di Giffoni' *C. avellana* LG5

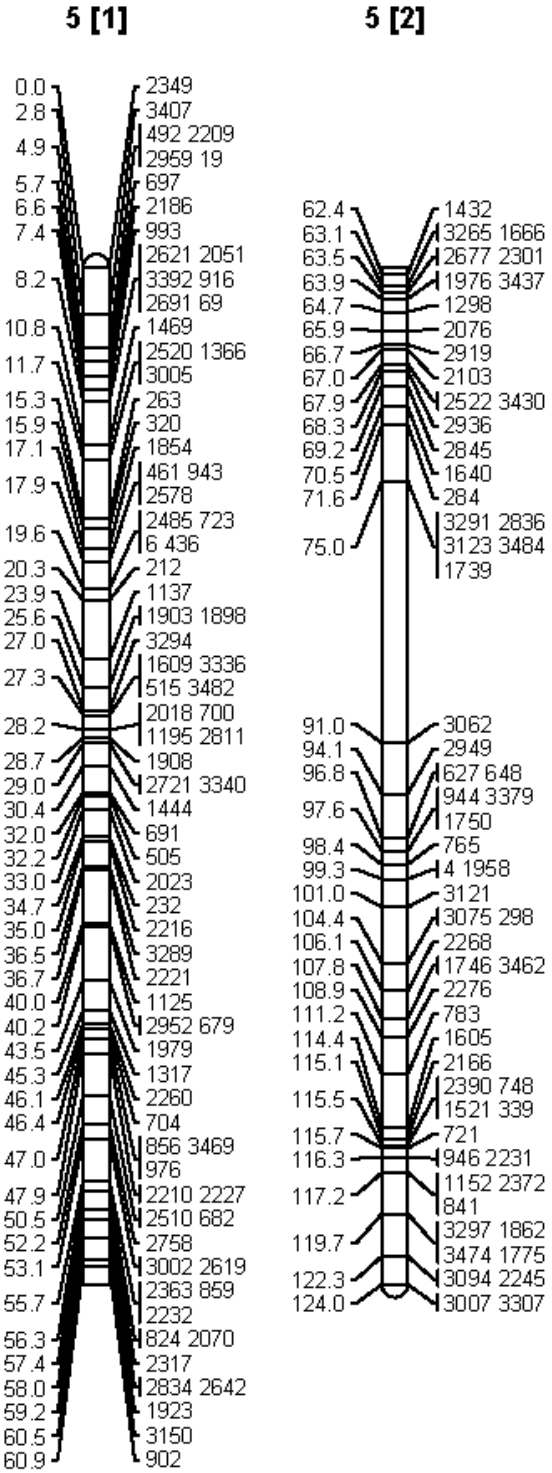


Figure A.8. OSU 403.040 *C. americana* × 'Tonda di Giffoni' *C. avellana* LG6

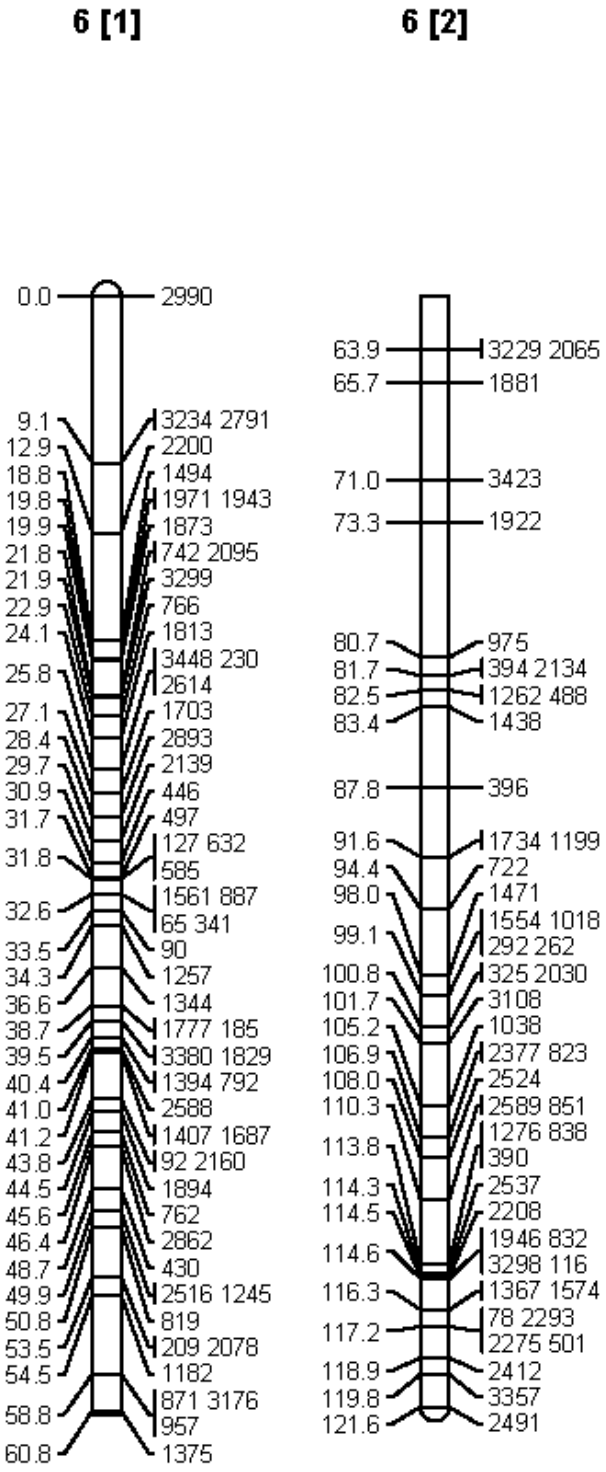


Figure A.9. OSU 403.040 *C. americana* × 'Tonda di Giffoni' *C. avellana* LG7

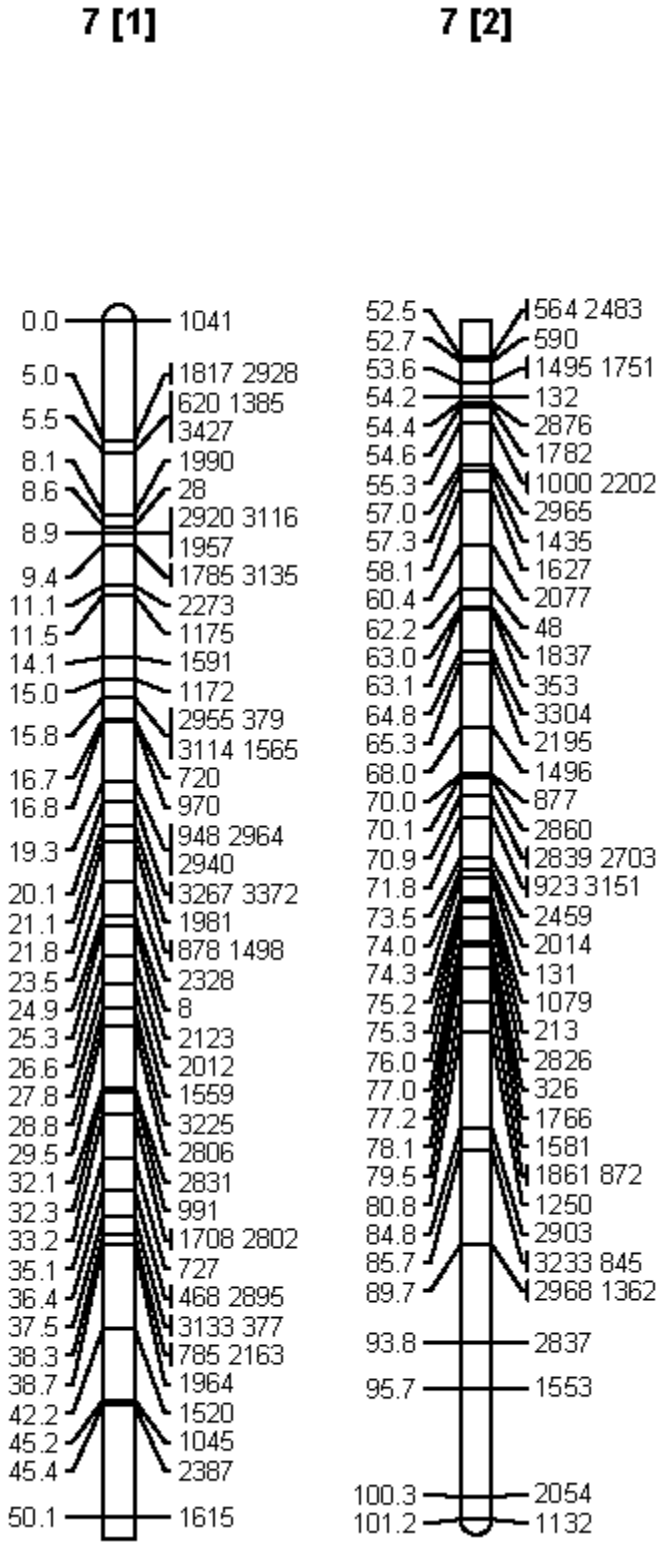


Figure A.10. OSU 403.040 *C. americana* × 'Tonda di Giffoni' *C. avellana* LG8

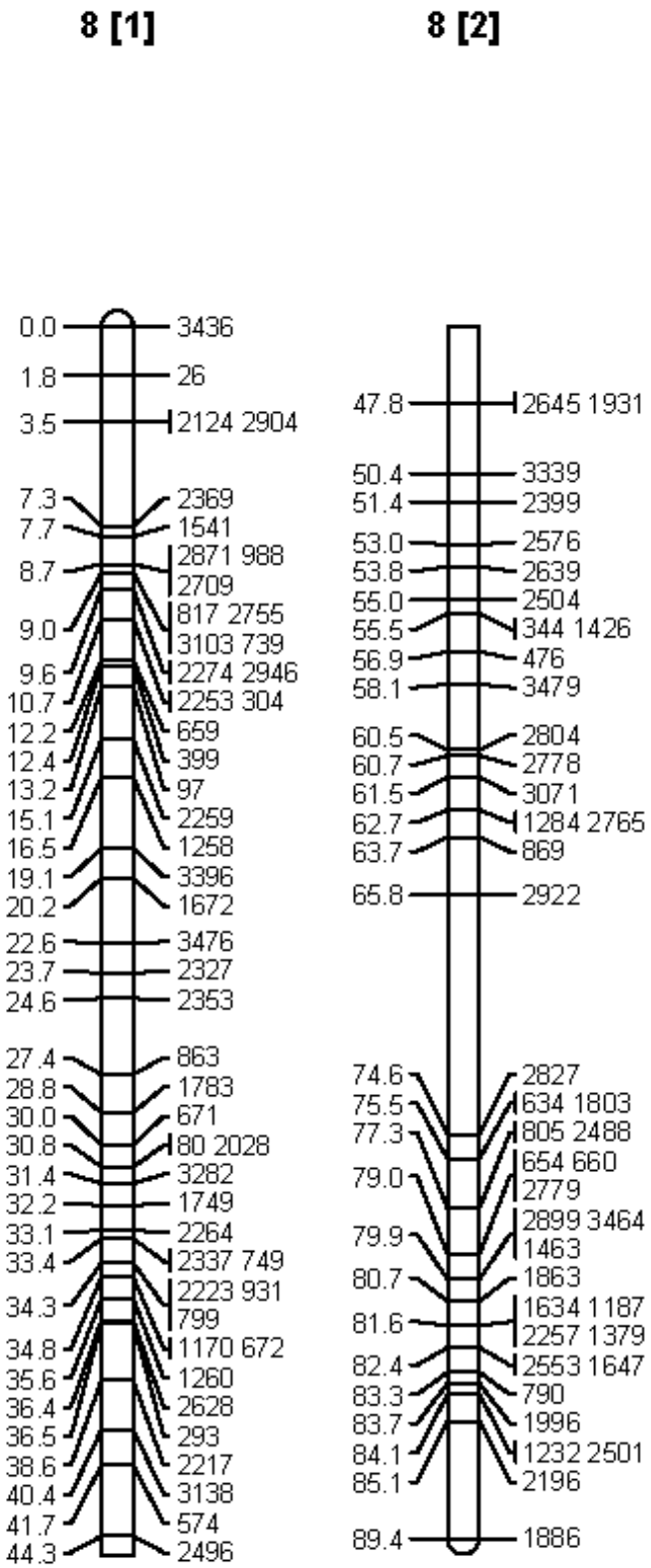


Figure A.11. OSU 403.040 *C. americana* × 'Tonda di Giffoni' *C. avellana* LG9

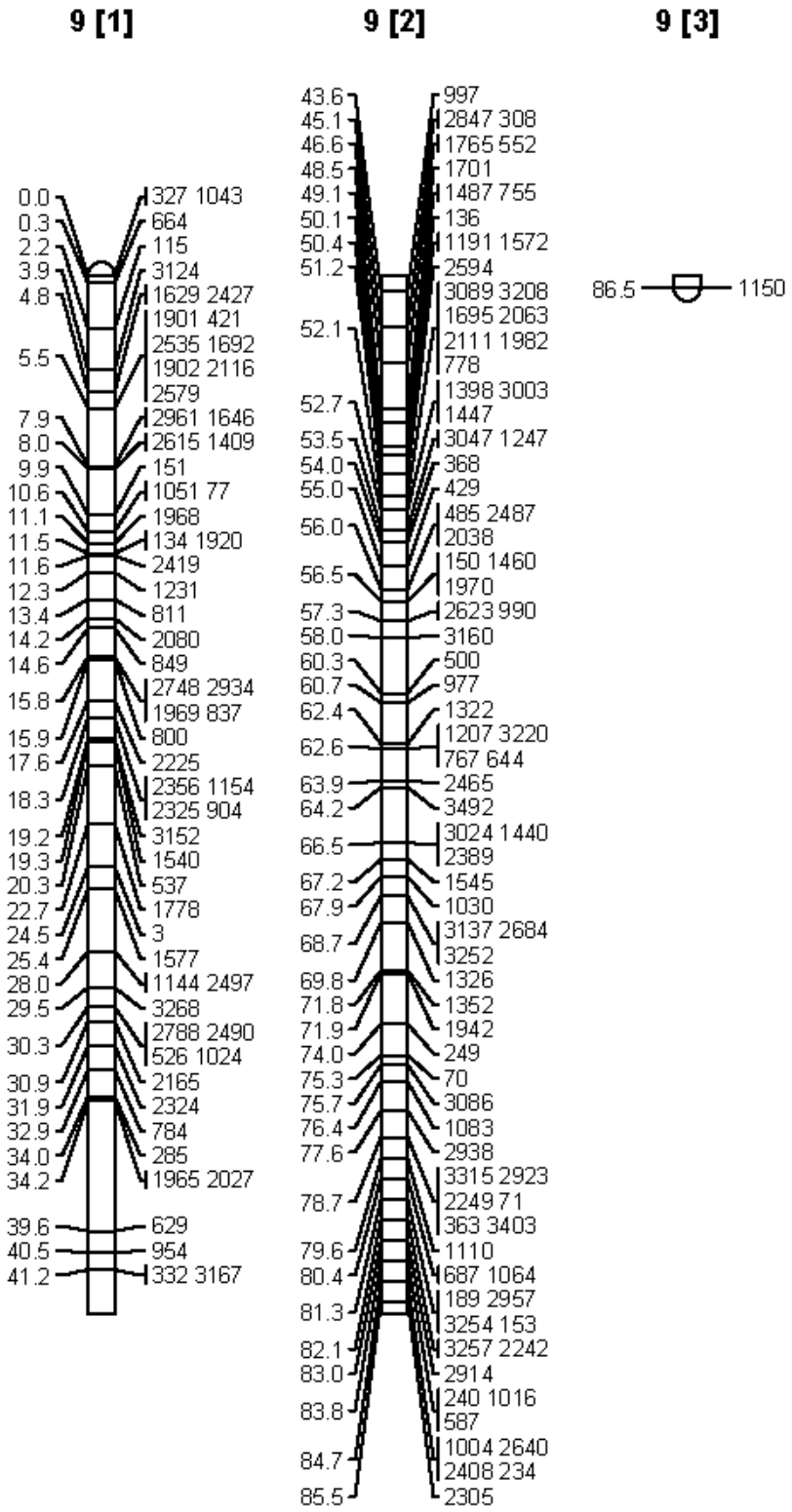


Figure A.12. OSU 403.040 *C. americana* × 'Tonda di Giffoni' *C. avellana* LG10

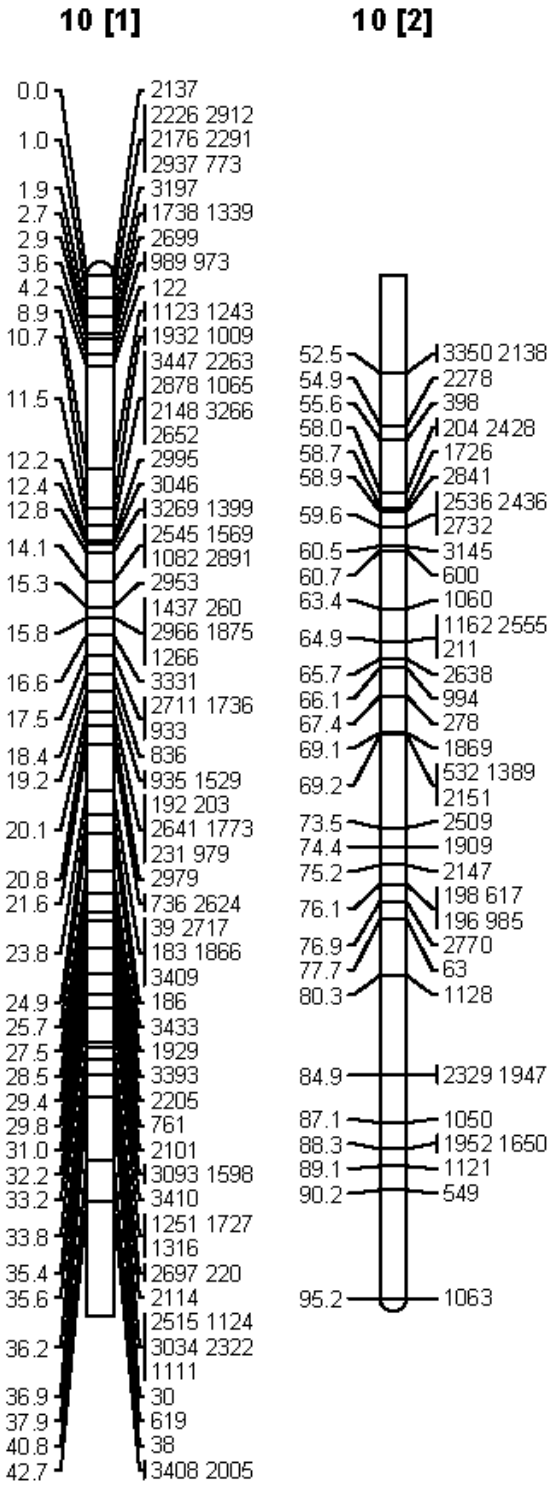
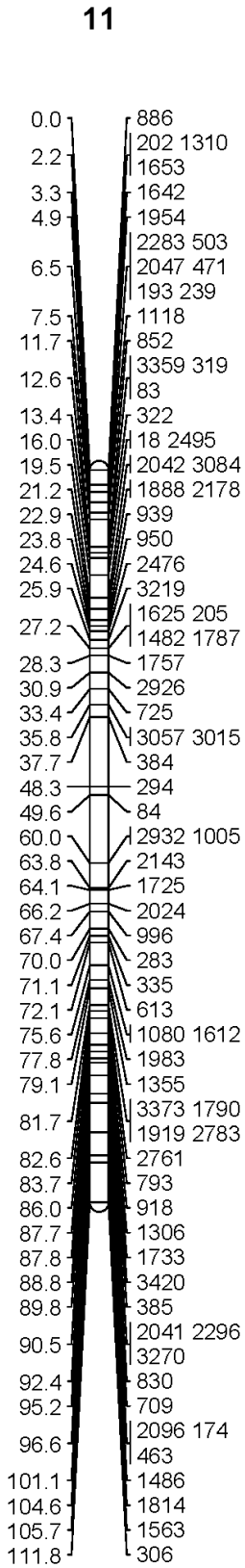


Figure A.13. OSU 403.040 *C. americana* × 'Tonda di Giffoni' *C. avellana* LG11



APPENDIX B: SUPPLEMENTARY TEXT

GBS-SNP-CROP command lines

- GBS-SNP-CROP-1.pl (Parse the raw reads):

$ perl GBS-SNP-CROP-1.pl -d PE -b barcodeID.Pool1.txt -fq 1 -s 1 -e 1 -enz1 TGCA -enz2 CGG -t 24

$ perl GBS-SNP-CROP-1.pl -d PE -b barcodeID.Pool2.txt -fq 2 -s 1 -e 1 -enz1 TGCA -enz2 CGG -t 24

$ perl GBS-SNP-CROP-1.pl -d PE -b barcodeID.Pool3.txt -fq 3 -s 1 -e 1 -enz1 TGCA -enz2 CGG -t 24

- GBS-SNP-CROP-2.pl (Trim based on quality):

$ perl GBS-SNP-CROP-2.pl -d PE -fq 1 -t 10 -ph 33 -ad TruSeq3-PE.fa:2:30:10 -l 30 -sl 4:30 -tr 30 -m 32

$ perl GBS-SNP-CROP-2.pl -d PE -fq 2 -t 10 -ph 33 -ad TruSeq3-PE.fa:2:30:10 -l 30 -sl 4:30 -tr 30 -m 32

$ perl GBS-SNP-CROP-2.pl -d PE -fq 3 -t 10 -ph 33 -ad TruSeq3-PE.fa:2:30:10 -l 30 -sl 4:30 -tr 30 -m 32

- GBS-SNP-CROP-3.pl (Demultiplex):

$ perl GBS-SNP-CROP-3.pl -d PE -b barcodeID.Pool1.txt -fq 1

$ perl GBS-SNP-CROP-3.pl -d PE -b barcodeID.Pool2.txt -fq 2

$ perl GBS-SNP-CROP-3.pl -d PE -b barcodeID.Pool3.txt -fq 3

- GBS-CNP-CROP-4.pl (Mock Reference Build):

$ perl GBS-SNP-CROP-4.pl -pr pear -vs vsearch -d PE -b barcodeIDMerged.txt -t 10 -cl consout -rl 150 -pl 32 -p 0.01 -id 0.93 -min 32 -MR GSC

- GBS-SNP-CROP-5.pl (Align with BWA-mem and process with SAMtools):

$ perl GBS-SNP-CROP-5.pl -bw bwa -st samtools -d PE -b barcodeIDMerged.txt -ref GSC.Genome.fa -Q 30 -q 30 -F 2308 -f 2 -t 10 -Opt 0

- GBS-SNP-CROP-6.pl (Parse mpileup files and discovery variants):

$ perl GBS-SNP-CROP-6.pl -b barcodeIDMerged.txt -out GSC.MasterMatrix.txt -p indel -t 10

- GBS-SNP-CROP-7.pl (Filter variants and call genotypes):

$ perl GBS-SNP-CROP-7.pl -in GSC.MasterMatrix.txt -out GSC.GenoMatrix.txt -p indel -mnHoDepth0 5 -mnHoDepth1 20 -mnHetDepth 3 -altStrength 0.8 -mnAlleleRatio 0.25 -mnCall 0.75 -mnAvgDepth 3 -mxAvgDepth 200

- GBS-SNP-CROP-8.pl (Creating input files for downstream analyses):

$ perl GBS-SNP-CROP-8.pl -in GSC.GenoMatrix.txt -out GSC -b barcodeIDMerged.txt -formats T,R,P,V,H

- GBS-SNP-CROP-9.pl (Provide descriptors for all called variants based on Mock Ref):

$ perl GBS-SNP-CROP-9.pl -in GSC.GenoMatrix.txt -out GSC -ref GSC.MR.Cluster.fa