



Genotyping-by-sequencing shows high genetic diversity in *Corylus avellana* germplasm resistant to eastern filbert blight

J. J. Lombardoni^{1,3} · J. A. Honig¹ · J. N. Vaiciunas¹ · C. Kubik¹ · J. M. Capik¹ · S. A. Mehlenbacher² · T. J. Molnar¹ 

Received: 4 March 2024 / Revised: 20 September 2024 / Accepted: 10 October 2024 / Published online: 29 October 2024
© The Author(s) 2024

Abstract

European hazelnut (*Corylus avellana*) is an anemophilous, dichogamous, self-incompatible tree nut species. It is native to a large portion of Europe, Turkey, and the Caucasus region, across which a wealth of plant genetic resources is present. The objective of this study was to evaluate the genetic diversity of a core set of *C. avellana* representing the world's germplasm using genotyping-by-sequencing derived single nucleotide polymorphism (SNP) markers and to classify novel eastern filbert blight (EFB) resistant or tolerant accessions. Two-hundred-twenty-two accessions underwent next-generation sequencing (NGS) to generate SNP markers. From this, 1,250 SNP markers were used to construct a neighbor-joining (NJ) dendrogram and perform a STRUCTURE and discriminant analysis of principal coordinates (DAPC) analyses. In the dendrogram, five major groups were established, which generally corresponded to geographic origins of the plant materials studied. In STRUCTURE, support was found for groupings at $(K)=3$, $(K)=6$, and $(K)=10$ populations, with the greatest $\Delta(K)$ value occurring at $(K)=10$. Although the three different analyses indicated slightly different solutions, the overall results were generally consistent from the standpoint of identifying similar accession groupings. For many of the accessions, recorded origins tended to correspond with their genetic grouping, although there was also evidence of intermixing and likely movement of plant materials. Interestingly, in all three analyses, a vast majority of the new accessions from the Republic of Georgia formed their own distinct group, highlighting this geographic region as a unique pool of *C. avellana* genetic resources. Overall, EFB resistant/tolerant accessions were placed across a wide range of genetic backgrounds. Thus, our results indicate EFB resistance/tolerance is present across a wide spectrum of *C. avellana* genetic resources, with the Georgian accessions representing a new and relatively unique germplasm pool that can be incorporated into hazelnut breeding programs.

Keywords Hazelnut · Genetic diversity · Genotyping-by-sequencing · Eastern filbert blight

Introduction

European hazelnut (*Corylus avellana*) is the most economically prominent and well-studied species of the *Corylus* genus, which current taxonomic revisions suggest holds

13 species assigned to four subsections (Botta et al. 2019; Molnar 2011). It is the primary species grown in commercial hazelnut production worldwide for both the in-shell and kernel markets. Major producing countries include Turkey, Italy, the United States, Azerbaijan, and the Republic of Georgia (FAO, 2021). Currently, 99% of production within the United States takes place in the Willamette Valley of Oregon, which comprised about 3.2% of total world production of 1,077,117 metric tonnes in 2021 (FAO, 2021). While current commercial production is limited to areas bordered by large bodies of water, such as the Black Sea or the Mediterranean Sea, the native range of *C. avellana* spans a considerable region from coastal Norway south through much of Europe to Morocco and the Caucasus, bordered in the West by the Atlantic Ocean and the East by the Ural Mountains of Russia to parts of Iran (Gökirmak et al. 2009; Mehlenbacher 1991).

Communicated by D. Chagne.

✉ T. J. Molnar
thomas.molnar@rutgers.edu

¹ Department of Plant Biology, Rutgers – The State University of New Jersey, Foran Hall, 59 Dudley Road, New Brunswick, NJ 08901, USA

² Department of Horticulture, Oregon State University, 4017 Agricultural and Life Sciences Building, Corvallis, OR 97331, USA

³ Chicago Botanic Garden, 1000 Lake Cook Road, Glencoe, IL 60022, USA

The disease eastern filbert blight (EFB), caused by the fungus *Anisogramma anomala*, has severely limited the development of a European hazelnut industry in eastern North America (Thompson et al. 1996). This pathogen is harbored by American hazelnut (*C. americana*), a shrubby species with small, thick-shelled nuts, whose native range spans much of the eastern half of North America. While *C. americana* is largely resistant or tolerant of the disease, most European hazelnut trees are highly susceptible, where the pathogen causes destructive stem cankers, branch die-back, and tree death within a few years of exposure (Fuller 1908; Capik and Molnar 2012; Johnson and Pinkerton 2002; Revord et al. 2020). Historically, EFB was restricted to North America east of the Rocky Mountains. Around 1960, however, *A. anomala* was inadvertently introduced into southwestern Washington, where it decimated orchards prior to the development of control measures. It has since spread south throughout the entire Willamette Valley of Oregon, where it represents a significant production challenge (Gottwald and Cameron 1980; Johnson et al. 1996; Pinkerton et al. 1992, 1993).

Genetic host resistance is widely considered to be the most effective, long-term approach for managing this pathogen (Johnson et al. 1996; Thompson et al. 1996). The first source of genetic resistance identified was ‘Gasaway’, an obsolete, late-shedding pollinizer found free of disease in a heavily infected ‘DuChilly’ orchard in Battle Ground, Washington (Cameron 1976; Thompson et al. 1996). ‘Gasaway’ resistance was found to be controlled by a dominant allele at a single locus, which facilitated successful breeding efforts (Mehlenbacher et al. 1991). Today, after several decades of breeding utilizing the ‘Gasaway’ resistance source, many EFB resistant cultivars and pollinizers have been released from Oregon State University (OSU) that have led to the revival of a hazelnut industry in Oregon (Mehlenbacher et al. 2023). Production in Oregon has increased from 11,700 ha to more than 35,000 ha from 2009 to 2023 (N.G. Wiman, personal communication).

Studies have shown *A. anomala* to be highly genetically diverse across its native range but limited in the Pacific Northwest, where there is believed to have been a single point of introduction (Cai et al. 2013; Muehlbauer et al. 2019; Tobia et al. 2017). For durable resistance, breeders have sought to identify additional sources of resistance. Germplasm screening efforts have been ongoing at OSU and, fortunately, several additional sources of resistance have since been identified (Mehlenbacher et al. 2023). Genetic linkage mapping has been part of the effort to understand genetic control of resistance and find sources on different linkage groups (LGs) to aid possible resistance gene (R-gene) pyramiding. ‘Gasaway’ was the first source of resistance mapped to LG6 (Mehlenbacher et al. 2006). Other sources of resistance were later mapped, including

the selection OSU 759.010 from the Republic of Georgia (Sathuvalli et al. 2011a) and accession H3R07P25 from southern Russia (Honig et al. 2019) which both mapped to LG2, as well as ‘Ratoli’ from Spain (Sathuvalli et al. 2011b) and *C. americana* ‘Rush’ (Bhattarai et al. 2017) which both mapped to LG7. Additional accessions whose resistance mapped to LG6 include OSU 495.072, ‘Culpla’ (Colburn et al. 2015) and several others summarized in Mehlenbacher and Molnar (2021) and Mehlenbacher et al. (2023). Further, Lombardoni et al. (2022) found QTLs for tolerance from ‘Tonda di Giffoni’ that mapped to LG10.

Beyond this work, sources of resistance identified in Oregon were tested at Rutgers University, where the pathogen is native. These efforts found that some accessions deemed resistant in Oregon, including ‘Gasaway’, can succumb to EFB when exposed to the pathogen in New Jersey and several other locations (Capik and Molnar 2012; Molnar et al. 2010a, 2010b).

Alongside evaluation of accessions deemed resistant in Oregon, seed-based *C. avellana* germplasm collection efforts were made, and trees were grown at Rutgers under high disease pressure to screen for new sources of resistance. Over a period of five years, open-pollinated seeds were collected from many locations across Russia, Crimea, Poland, Latvia, Lithuania, Estonia, Moldova, the Republic of Georgia, Turkey, and Italy. From these efforts, which evaluated approximately 5,000 trees in the field in New Jersey, over 100 EFB resistant and tolerant accessions were identified that originate from over 60 different locations (Capik and Molnar 2012; Capik et al. 2013; Leadbetter et al. 2015; Molnar et al. 2005, 2007, 2018). Among these resistant/tolerant accessions is a subset of 64 that originated from different regions of the Republic of Georgia. These accessions have not been genotyped previously and hold great promise for breeding efforts, especially if they represent sources of resistance genetically distinct from existing germplasm.

European hazelnut is monoecious, dichogamous, anemophilous, and self-incompatible; these characteristics lead to an obligately outcrossing mating system (Mehlenbacher 1991, 1994). The hazelnut genome is relatively small at ~378 Mb, but the outcrossing nature of hazelnuts leads to a highly heterozygous genome and the preservation of genetic diversity (Rowley et al. 2012, 2018). Hazelnut genetic diversity has been studied previously using both chloroplast and nuclear molecular markers and the grouping of accessions was often related to their geographic origin (Bassil et al. 2013; Boccacci et al. 2006; Gökirmak et al. 2009; Muehlbauer et al. 2014). Most hazelnut diversity studies have made use of microsatellite or simple sequence repeat (SSR) markers, as they are abundant throughout the genome, codominant, and highly polymorphic. However, as next-generation sequencing (NGS) becomes more affordable and single nucleotide polymorphism (SNP) markers become

less cumbersome to develop, SNP markers are becoming used more widely (Elshire et al. 2011).

Defined and recurring hazelnut genetic groups in studies using nuclear SSR markers include a Spanish-Italian group, a Turkish/Black Sea group, an English group, and a Central European group, among other groups that occur less consistently across studies (Boccacci et al. 2006; Gökirmak et al. 2009; Gürcan et al. 2010; Muehlbauer et al. 2014). Two studies of Turkish germplasm found a high level of genetic diversity within the country and some clustering by geographic region (Gürcan et al. 2010; Öztürk et al. 2017). A later study of Turkish germplasm using SNP markers also supported geographic clustering of accessions and found that cultivated hazelnut remained highly heterozygous and did not undergo a significant genetic bottleneck during domestication (Helmstetter et al. 2019). Outcrossing in hazelnut has likely played a large role in preserving diversity in the species during its domestication history. This may prove especially beneficial to breeders, who can capitalize on this diversity to combine different traits into more desirable cultivars that remain heterozygous. Understanding this genetic diversity can also aid in parental selection and help avoid the possible narrowing of genetic diversity when a reduced number of parents from a limited pool of disease resistant selections is used in a population improvement program.

The objective of this study was to use SNP markers from double digest restriction-side associated DNA sequencing (ddRADSeq) to evaluate the genetic diversity and relationships in a large collection of European hazelnut accessions expressing resistance or tolerance to EFB. This collection included new germplasm from the Republic of Georgia and well-known cultivars and other accessions believed to represent the known diversity within *C. avellana*.

Materials and methods

Germplasm

Two-hundred-twenty-two accessions were genotyped in this study (Table 1). Roughly half of these represent EFB resistant and tolerant germplasm identified from open-pollinated seed collection efforts made across the native range of *C. avellana*. These selected plants are now maintained at Rutgers University with a subset also established at OSU (Capik et al. 2013; Molnar et al. 2007, 2018). Included in this study are 64 new accessions from the Republic of Georgia (Leadbetter et al. 2015), as well as multiple accessions from Turkey and Latvia, that have not been genotyped previously. The remaining accessions were included to represent the known genetic diversity of European hazelnut based on previous studies. Roughly half of the samples in this study had been previously fingerprinted with SSR markers and thus serve

as points of comparison with SNP markers (Gökirmak et al. 2009; Muehlbauer et al. 2014). One hornbeam (*Carpinus sp.*) accession was included to serve as an outgroup within the Betulaceae (Birch) family.

Disease response phenotyping

The EFB disease response of many cultivars and selections has been documented in previous studies (Muehlbauer et al. 2014). For uncharacterized accessions, disease response was rated on the 0–5 scale established by Pinkerton et al., (1992). Briefly, 0 = no EFB detected, 1 = one canker, 2 = multiple cankers on one branch, 3 = cankers on multiple branches, 4 = cankers on over 50% of branches, 5 = cankers on all branches. Over a period of several years, it is common to see ratings of three or more progress to a higher rating and ultimately tree death. As such, trees rated zero were categorized as resistant, while those rated one or two were categorized as tolerant and those rated three to five were categorized as susceptible.

Sample collection and DNA extraction

Leaf samples were collected on dry ice from Rutgers University Horticultural Research Farms 1 and 3 (New Brunswick and East Brunswick, NJ, respectively), the Rutgers Specialty Crop Research and Extension Center (Cream Ridge, NJ) in May 2017, and the USDA National Clonal Germplasm Repository (Corvallis, OR) and OSU in June 2017. Leaf tissue was flash frozen in liquid nitrogen and approximately 0.20 g of tissue per sample was ground in liquid nitrogen using a mortar and pestle. DNA extractions were conducted using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions, and DNA quality and quantity were measured using a spectrophotometer (NanoDrop; Thermo-Fischer Scientific, Waltham, MA). Samples were then diluted to 10 ng/μL to be used for ddRADSeq library preparation.

Library preparation and sequencing

Restriction site associated DNA sequencing (RADseq) libraries were prepared using a protocol modified from Poland et al. (2012). Library preparation began with a double restriction digest of 200 ng of DNA per sample using the rare-cutting enzyme *PstI* (NEB, Ipswich, MA) and common-cutting enzyme *MspI* at 37 °C for 2 h. *PstI* adapters with unique barcodes of 5–10 base pairs and common *MspI* Y-adapters were ligated to the digested DNA in a mastermix containing 200 U of T4 DNA ligase, 2 μL of 10×NEBuffer 4, and 4 μL of ATP (10 mM) (NEB, Ipswich, MA) per sample by incubating at 22 °C for 2 h, followed by 20 min at 65 °C to inactivate the ligase. Samples then

Table 1 European hazelnut accession names, Rutgers identifier code or United States Department of Agriculture National Plant Germplasm System Plant Inventory (PI) number, countries of origin, eastern filbert blight (EFB) responses, dendrogram group placements, STRUCTURE group placements, and DAPC group placement of the 222 accessions included in this study. The accessions are ordered by their placement in dendrogram

Accession name	Identifier code/ PI code	Country of origin	EFB response	Dendrogram group	STRUCTURE $K=10$ grouping	DAPC grouping
Georgia Kakheti 1	HF1BR21P93	Georgia	resistant	Georgian	Georgian 1	Group 2—Georgian
Tartu Home Garden 1	H4BR22P55	Estonia	resistant	Georgian	Georgian 1	Group 2—Georgian
Georgia Samegrelo 1	HF1BR28P99	Georgia	resistant	Georgian	Georgian 1	Group 2—Georgian
Georgia Adjara 4	HF1BR26P21	Georgia	resistant	Georgian	Georgian 1	Group 2—Georgian
Georgia Guria 7	HF1BR28P237	Georgia	resistant	Georgian	Georgian 1	Group 2—Georgian
Georgia Adjara 5	HF1BR26P37	Georgia	resistant	Georgian	Georgian 1	Group 2—Georgian
Georgia Guria 6	HF1BR28P236	Georgia	resistant	Georgian	Georgian 1	Group 2—Georgian
Anakliuri	Anakliuri	Georgia	susceptible	Georgian	Georgian 1	Group 2—Georgian
Sochi unknown 2	CRR05P32	Russia	resistant	Georgian	Georgian 1	Group 2—Georgian
Georgia Kakheti 9	HF1BR26P233	Georgia	resistant	Georgian	Georgian 1	Group 2—Georgian
Khachapura	Khachapura	Georgia	susceptible	Georgian	Georgian 1	Group 2—Georgian
Oregon State University (OSU) 1090.013	OSU 1090.013	Georgia	susceptible	Georgian	Georgian 1	Group 2—Georgian
OSU 1120.034	OSU 1120.034	Georgia	tolerant*	Georgian	Georgian 1	Group 2—Georgian
Georgia Imereti 1	HF1BR21P151	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
Georgia Imereti 6	HF1BR23P235	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
Georgia Imereti 2	HF1BR23P219	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
Georgia Imereti 5	HF1BR23P232	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
Georgia Imereti 3	HF1BR23P220	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
Georgia Imereti 4	HF1BR23P223	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
Georgia Imereti 13	HF1BR32P31	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
Georgia Imereti 12	HF1BR32P27	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
Georgia Imereti 15	HF1BR32P209	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
Georgia Imereti 16	HF1BR32P217	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
Ata Baba	PI 557422	Azerbaijan	susceptible	Georgian	Georgian 2	Group 2—Georgian
OSU 1120.020	OSU 1120.020	Georgia	susceptible	Georgian	Georgian 2	Group 2—Georgian
OSU 1120.025	OSU 1120.025	Georgia	susceptible*	Georgian	Georgian 2	Group 2—Georgian
OSU 1217.005	OSU 1217.005	Georgia	tolerant*	Georgian	Georgian 2	Group 2—Georgian
Georgia Samegrelo 4	OSU 1225.046	Georgia	susceptible	Georgian	Georgian 2	Group 2—Georgian
Georgia Samegrelo 5	OSU 1225.067	Georgia	tolerant*	Georgian	Georgian 2	Group 2—Georgian
Georgia Adjara 1	HF1BR21P100	Georgia	tolerant	Georgian	Georgian 2	Group 2—Georgian
Dedoplistiti	Dedoplistiti	Georgia	susceptible	Georgian	Georgian 2	Group 2—Georgian
Georgia Samegrelo 2	HF1BR30P34	Georgia	tolerant	Georgian	Georgian 2	Group 2—Georgian
Badem OP 3	H3R03P12	Russia	resistant	Georgian	Georgian 2	Group 2—Georgian
Georgia Samegrelo 3	HF1BR30P37	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
Sochi unknown 3	CRR03P11	Russia	resistant	Georgian	Georgian 2	Group 2—Georgian
Georgia Guria 3	HF1BR23P70	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
Georgia Imereti 7	HF1BR28P11	Georgia	tolerant	Georgian	Georgian 2	Group 2—Georgian
Georgia Adjara 8	HF1BR30P197	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
OSU 495.072	PI 557421	Russia	resistant	Georgian	Georgian 2	Group 2—Georgian
B X 3 OP 1	CRR01P116	Russia	resistant	Georgian	Georgian 2	Group 2—Georgian

Table 1 (continued)

Accession name	Identifier code/ PI code	Country of origin	EFB response	Dendrogram group	STRUCTURE $K=10$ grouping	DAPC grouping
Georgia Adjara 17	HF1BR32P128	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
Georgia Adjara 2	HF1BR21P163	Georgia	resistant	Georgian	Georgian 3	Group 2—Georgian
Georgia Samegrelo 6	OSU 1225.026	Georgia	tolerant*	Georgian	Georgian 3	Group 2—Georgian
Georgia Adjara 10	HF1BR32P101	Georgia	resistant	Georgian	Georgian 3	Group 2—Georgian
Georgia Adjara 12	HF1BR32P105	Georgia	resistant	Georgian	Georgian 3	Group 2—Georgian
OSU 1090.080	OSU 1090.080	Georgia	tolerant*	Georgian	Georgian 3	Group 2—Georgian
Georgia Adjara 11	HF1BR32P103	Georgia	resistant	Georgian	Georgian 3	Group 2—Georgian
Georgia Kakheti 12	HF1BR32P159	Georgia	resistant	Georgian	Georgian 3	Group 2—Georgian
Georgia Kakheti 16	HF1BR32P169	Georgia	resistant	Georgian	Georgian 3	Group 2—Georgian
Georgia Guria 2	HF1BR23P14	Georgia	resistant	Georgian	Georgian 3	Group 2—Georgian
Georgia Kakheti 14	HF1BR32P166	Georgia	resistant	Georgian	Georgian 3	Group 2—Georgian
Georgia Adjara 13	HF1BR32P112	Georgia	tolerant	Georgian	Georgian 3	Group 2—Georgian
Georgia Adjara 14	HF1BR32P116	Georgia	resistant	Georgian	Georgian 3	Group 2—Georgian
Georgia Kakheti 17	HF1BR32P171	Georgia	resistant	Georgian	Georgian 3	Group 2—Georgian
Georgian 759.010	OSU 759.010	Georgia	tolerant	Georgian	Georgian 3	Group 2—Georgian
Gulshishvela	Gulshishvela	Georgia	susceptible	Georgian	Georgian 3	Group 2—Georgian
Georgia Adjara 16	HF1BR32P120	Georgia	resistant	Georgian	Georgian 3	Group 2—Georgian
Georgia Kakheti 13	HF1BR32P163	Georgia	resistant	Georgian	Georgian 3	Group 2—Georgian
Georgia Adjara 9	HF1BR32P99	Georgia	resistant	Georgian	Georgian 3	Group 2—Georgian
Georgia Kakheti 15	HF1BR32P168	Georgia	resistant	Georgian	Georgian 3	Group 2—Georgian
Georgia Adjara 3	HF1BR23P162	Georgia	tolerant	Georgian	Georgian 3	Group 2—Georgian
Georgia Adjara 7	HF1BR30P189	Georgia	resistant	Georgian	Georgian 4	Group 2—Georgian
Georgia Kakheti 10	HF1BR30P205	Georgia	resistant	Georgian	Georgian 4	Group 2—Georgian
Georgia Adjara 15	HF1BR32P119	Georgia	resistant	Georgian	Georgian 4	Group 2—Georgian
Georgia Guria 8	HF1BR32P135	Georgia	resistant	Georgian	Georgian 4	Group 2—Georgian
OSU 1090.112	OSU 1090.112	Georgia	tolerant*	Georgian	Georgian 4	Group 2—Georgian
Georgia Guria 1	HF1BR21P192	Georgia	resistant	Georgian	Georgian 4	Group 2—Georgian
Georgia Guria 4	HF1BR28P63	Georgia	resistant	Georgian	Georgian 4	Group 2—Georgian
Georgia Guria 5	HF1BR28P68	Georgia	tolerant	Georgian	Georgian 4	Group 2—Georgian
Georgia Abkhazia 1	HF1BR26P140	Georgia	resistant	Georgian	Georgian 4	Group 2—Georgian
OSU 1089.024	OSU 1089.024	Georgia	tolerant*	Georgian	Georgian 4	Group 1 – Spanish-Italian / English
OSU 1133.053	OSU 1133.053	Georgia	susceptible*	Georgian	Georgian 4	Group 2—Georgian
Georgia Kakheti 4	HF1BR26P138	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
OSU 1133.009	OSU 1133.009	Georgia	tolerant*	Georgian	Georgian 2	Group 2—Georgian
Georgia Kakheti 5	HF1BR26P151	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
Georgia Kakheti 6	HF1BR26P157	Georgia	tolerant	Georgian	Georgian 2	Group 2—Georgian
Georgia Kakheti 7	HF1BR26P162	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
Georgia Kakheti 8	HF1BR26P163	Georgia	resistant	Georgian	Georgian 2	Group 2—Georgian
OSU 1089.004	OSU 1089.004	Georgia	susceptible	Georgian	Georgian 2	Group 2—Georgian
Russia unknown 1	HF1BR21P182	Russia	resistant	Central European	Central European 1	Group 3 – Central European
DZ 118 Dacha 4 Moscow	H4BR21P137	Russia	resistant	Central European	Central European 1	Group 3 – Central European
Latvia Dobelle 3	CRTR05P140	Latvia	resistant	Central European	Central European 1	Group 3 – Central European
Moscow N26	Moscow N26	Russia	resistant	Central European	Central European 1	Group 3 – Central European

Table 1 (continued)

Accession name	Identifier code/ PI code	Country of origin	EFB response	Dendrogram group	STRUCTURE $K=10$ grouping	DAPC grouping
Estonia unknown OP 1	H4BR22P75	Estonia	tolerant	Central European	Central European 1	Group 3 – Central European
Estonia unknown OP 2	H4BR22P81	Estonia	tolerant	Central European	Central European 1	Group 3 – Central European
Agusalu Wild 1	H4BR22P114	Estonia	tolerant	Central European	Central European 1	Group 3 – Central European
OSU 408.040	PI 617266	Minnesota, USA	resistant	Central European	Central European 1	Group 3 – Central European
Latvia Dobelle 2	CRTR05P87	Latvia	tolerant	Central European	Central European 1	Group 3 – Central European
Fusco Rubra	PI 557057	Germany	susceptible	Central European	Central European 1	Group 3 – Central European
Gasaway	PI 557042	Washington, USA	resistant*	Central European	Central European 1	Group 3 – Central European
Delta	PI 637883	Oregon, USA	resistant*	Central European	Central European 1	Group 3 – Central European
Cutleaf	PI 557306	England	susceptible	Central European	Central European 1	Group 3 – Central European
Albania 55	PI 617217	Albania	susceptible	Central European	Central European 1	Group 3 – Central European
Aurea	PI 557050	France	susceptible	Central European	Central European 1	Group 3 – Central European
Sweden COR 627	PI 617209	Sweden	susceptible	Central European	Central European 1	Group 3 – Central European
Moscow N01	Moscow N01	Russia	susceptible	Central European	Central European 1	Group 3 – Central European
Simferopol Market 1A 2	CRXR18P08	Crimea	tolerant	Central European	Central European 1	Group 3 – Central European
Finland COR 187	PI 557080	Finland	tolerant	Central European	Central European 1	Group 3 – Central European
Moscow N23	Moscow N23	Russia	resistant*	Central European	Central European 1	Group 3 – Central European
Moscow N37	Moscow N37	Russia	resistant	Central European	Central European 1	Group 3 – Central European
Simferopol Market 5 2	CRXR19P04	Crimea	tolerant	Central European	Central European 1	Group 3 – Central European
Simferopol Market 4 1	CRXR19P21	Crimea	resistant	Central European	Central European 1	Group 3 – Central European
OSU 1133.068	OSU 1133.068	Georgia	resistant*	Central European	Central European 1	Group 3 – Central European
Skierniewice mix 7	H4AR18P73	Poland	resistant	Central European	Central European 1	Group 3 – Central European
Riga Market 3 OP 2	CRTR06P100	Latvia	resistant	Central European	Central European 1	Group 3 – Central European
Louisens Zellernuss	PI 557086	Germany	susceptible	Central European	Central European 1	Group 3 – Central European
Riga Market 3 OP 1	CRTR06P95	Latvia	resistant	Central European	Central European 1	Group 3 – Central European
Warsaw Market 4 1	H4AR20P88	Poland	resistant	Central European	Central European 2	Group 3 – Central European
Syrena	PI 617237	Poland	susceptible	Central European	Central European 2	Group 3 – Central European
Warsaw mix 2	H4AR21P05	Poland	resistant	Central European	Central European 2	Group 3 – Central European

Table 1 (continued)

Accession name	Identifier code/ PI code	Country of origin	EFB response	Dendrogram group	STRUCTURE $K=10$ grouping	DAPC grouping
Warsaw mix 1	H4AR21P03	Poland	resistant	Central European	Central European 2	Group 3 – Central European
Warsaw mix 3	H4AR21P43	Poland	tolerant	Central European	Central European 2	Group 3 – Central European
Cozia	PI 617172	Romania	unknown	Central European	Central European 2	Group 1 – Spanish-Italian / English
Aveline d'Angleterre	PI 557194	France	susceptible	Central European	Central European 2	Group 3 – Central European
Maria	PI 617236	Poland	susceptible	Central European	Central European 2	Group 3 – Central European
Lenka 3	PI 217233	Poland	unknown	Central European	Central European 2	Group 3 – Central European
Lech	PI 617232	Poland	unknown	Central European	Central European 2	Group 3 – Central European
Halls Giant	PI 557027	Germany/France	susceptible	Central European	Central European 2	Group 3 – Central European
Gustavs Zellernuss	PI 557085	Germany	susceptible	Central European	Central European 2	Group 3 – Central European
Truchsess Zeller-nuss	PI 557087	Germany	susceptible	Central European	Central European 2	Group 3 – Central European
Frango 4	PI 617228	Poland	susceptible	Central European	Central European 2	Group 3 – Central European
Latvia Dobelle 1	CRTR04P166	Latvia	tolerant	Central European	Central European 1	Group 3 – Central European
Blumberger	PI 557084	Germany	susceptible	Central European	Central European 1	Group 3 – Central European
Moscow unknown	Moscow unk	Russia	unknown	Central European	Central European 1	Group 3 – Central European
Moscow N02	Moscow N02	Russia	resistant	Central European	Central European 1	Group 3 – Central European
Kudashovski OP 7	CRXR13P91	Russia	tolerant	Central European	Central European 1	Group 3 – Central European
Kudashovski OP 1	CRXR13P78	Russia	resistant	Central European	Central European 1	Group 3 – Central European
Moskovskii Rubin OP 11	CRR04P48	Russia	tolerant	Central European	Central European 1	Group 3 – Central European
Moskovskii Rubin OP 7	CRR04P28	Russia	resistant	Central European	Central European 1	Group 3 – Central European
OSU 26.072	PI 323961	Russia	tolerant*	Central European	Central European 1	Group 3 – Central European
Tonda Rossa	PI 557118	Italy	susceptible	Central European	Central European 1	Group 3 – Central European
Gubener Barcel-loner	PI 557148	Germany	susceptible	Central European	Central European 1	Group 3 – Central European
Gunslebert	PI 557191	Germany	susceptible	Central European	Central European 1	Group 3 – Central European
Tonnolella	PI 557218	Italy	unknown	Central European	Central European 1	Group 3 – Central European
Luisa OP 1	CRTR05P48	Lithuania	tolerant	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Siciliana	PI 557175	Italy	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Sodlinger	PI 557212	Yugoslavia	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English

Table 1 (continued)

Accession name	Identifier code/ PI code	Country of origin	EFB response	Dendrogram group	STRUCTURE $K=10$ grouping	DAPC grouping
Willamette	PI 557234	Oregon, USA	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Bianca	PI 557182	Italy	unknown	Spanish-Italian	Spanish-Italian	Group 3 – Central European
Culpla	PI 557107	Spain	resistant	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Morell	PI 270338	Spain	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Tapparona di SCC	PI 617239	Italy	unknown	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Tonda G.d. Langhe	PI 557035	Italy	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Pinyolenc 1	PI 654964	Spain	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Ribet	PI 557055	Spain	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Negret	PI 270340	Spain	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Alcover	PI 557418	Italy	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Comen	PI 557178	Spain	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Ratoli	PI 557167	Spain	resistant	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Zugdui OP3	CRR06P02	Russia	tolerant	Spanish-Italian	Black Sea	Group 2—Georgian
Tonda Romana	PI 557025	Italy	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Casina	PI 557033	Spain	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Sant Jaume	PI 557103	Spain	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Garrofi	PI 557165	Spain	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Barcelona	PI 557037	Spain	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Macrocarpa	PI 557137	England	susceptible*	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Segorbe	PI 557046	Spain	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Artellet	PI 557108	Spain	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Tomasina	PI 557200	Spain	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Martorella	PI 557203	Spain	susceptible	Spanish-Italian	Spanish-Italian	Group 1 – Spanish-Italian / English
Contorta	PI 557049	England	susceptible	English	English	Group 1 – Spanish-Italian / English
Skorospelka	PI 617175	Georgia	unknown	English	English	Group 1 – Spanish-Italian / English
Nottingham	PI 557140	England	susceptible	English	English	Group 1 – Spanish-Italian / English
Ugbrooke	PI 557100	New Zealand	susceptible	English	English	Group 1 – Spanish-Italian / English

Table 1 (continued)

Accession name	Identifier code/ PI code	Country of origin	EFB response	Dendrogram group	STRUCTURE $K=10$ grouping	DAPC grouping
Whiteheart	PI 557101	New Zealand	unknown	English	English	Group 1 – Spanish-Italian / English
Frizzled Filbert	PI 557097	England	susceptible	English	English	Group 1 – Spanish-Italian / English
Henneman #3	PI 557427	Oregon, USA	susceptible	English	English	Group 1 – Spanish-Italian / English
The Shah	PI 557149	England	susceptible	English	English	Group 1 – Spanish-Italian / English
DuChilly	PI 557099	England	susceptible	English	English	Group 1 – Spanish-Italian / English
Ennis	PI 557045	Washington, USA	susceptible	English	English	Group 1 – Spanish-Italian / English
Daviana	PI 557040	England	susceptible	English	English	Group 1 – Spanish-Italian / English
Cosford	PI 557039	England	susceptible	English	English	Group 3 – Central European
Garibaldi	PI 557162	England	susceptible	English	English	Group 1 – Spanish-Italian / English
Rote Zellernuss	PI 271280	Netherlands	susceptible	English	English	Group 1 – Spanish-Italian / English
Webb's Prize Cobb	PI 557160	England	susceptible	English	English	Group 1 – Spanish-Italian / English
Bandnuss	PI 557157	England	susceptible	English	English	Group 1 – Spanish-Italian / English
Barr's Zellernuss	PI 557158	England	susceptible	English	English	Group 3 – Central European
Georgia Kakheti 3	HF1BR26P77	Georgia	resistant	Black Sea	Georgian 4	Group 2—Georgian
Simferopol Market 1B 1	H3R14P26	Crimea	resistant	Black Sea	Georgian 4	Group 1 – Spanish-Italian / English
Imperiale de Trebi-zonde	PI 271105	Turkey	tolerant*	Black Sea	Georgian 4	Group 1 – Spanish-Italian / English
Burchardt's Zellernuss	PI 557089	Germany	susceptible	Black Sea	Georgian 4	Group 1 – Spanish-Italian / English
Holmskij Market 1 1	H3R13P40	Russia	resistant	Black Sea	Georgian 4	Group 1 – Spanish-Italian / English
Sochi Institute mix 2	H3R10P94	Russia	resistant	Black Sea	Georgian 4	Group 1 – Spanish-Italian / English
Georgia Adjara 6	HF1BR28P199	Georgia	resistant	Black Sea	Turkish	Group 1 – Spanish-Italian / English
Giresun 286 1	CRTR04P14	Turkey	resistant	Black Sea	Turkish	Group 1 – Spanish-Italian / English
Palaz	PI 304632	Turkey	susceptible	Black Sea	Turkish	Group 1 – Spanish-Italian / English
San Benedetto	PI 557401	Italy	tolerant*	Black Sea	Turkish	Group 1 – Spanish-Italian / English
Kudryavchik	PI 217177	Georgia	unknown	Black Sea	Turkish	Group 1 – Spanish-Italian / English
Tombul Ghiaghli	PI 304634	Turkey	susceptible	Black Sea	Turkish	Group 1 – Spanish-Italian / English
Tombul Ocak 1	PI 304634	Turkey	susceptible	Black Sea	Turkish	Group 1 – Spanish-Italian / English
Ordu	PI 557417	Turkey	susceptible	Black Sea	Turkish	Group 1 – Spanish-Italian / English

Table 1 (continued)

Accession name	Identifier code/ PI code	Country of origin	EFB response	Dendrogram group	STRUCTURE $K=10$ grouping	DAPC grouping
Pioneer	PI 617178	Russia	susceptible	Black Sea	Turkish	Group 1 – Spanish-Italian / English
Giresun 112 2	CRTR02P138	Turkey	resistant	Black Sea	Turkish	Group 1 – Spanish-Italian / English
Giresun 191 1	CRTR02P131	Turkey	tolerant	Black Sea	Turkish	Group 1 – Spanish-Italian / English
Giresun 112 1	CRTR02P134	Turkey	tolerant	Black Sea	Turkish	Group 1 – Spanish-Italian / English
Giresun 146 1	CRTR02P186	Turkey	tolerant	Black Sea	Turkish	Group 1 – Spanish-Italian / English
Georgia Imereti 14	HF1BR32P148	Georgia	resistant	Black Sea	Admixed	Group 1 – Spanish-Italian / English
Istrski Duguljasti	PI 557400	Croatia	susceptible	Black Sea	Admixed	Group 1 – Spanish-Italian / English
Badem	PI 304630	Turkey	susceptible	Black Sea	Admixed	Group 1 – Spanish-Italian / English
Giresun 194 1	CRTR04P124	Turkey	resistant	Black Sea	Admixed	Group 1 – Spanish-Italian / English
Giresun 362 1	CRTR02P68	Turkey	resistant	Black Sea	Admixed	Group 1 – Spanish-Italian / English
Bulgaria XI 8	PI 557219	Bulgaria	tolerant	Black Sea	Admixed	Group 1 – Spanish-Italian / English
Giresun 362 2	CRTR02P71	Turkey	tolerant	Black Sea	Admixed	Group 1 – Spanish-Italian / English
Giresun 380 1	CRTR03P158	Turkey	tolerant	Black Sea	Admixed	Group 1 – Spanish-Italian / English
Giresun 233 1	CRTR02P03	Turkey	resistant	Black Sea	Admixed	Group 1 – Spanish-Italian / English
Georgia Kakheti 11	HF1BR32P74	Georgia	tolerant	Black Sea	Black Sea	Group 2—Georgian
Sochi Market 5 4	CRXR14P117	Russia	tolerant	Black Sea	Black Sea	Group 2—Georgian
Giresun 173 1	CRTR04P44	Turkey	susceptible	Black Sea	Black Sea	Group 1 – Spanish-Italian / English
Cherkesskii II	PI 617176	Russia	susceptible	Black Sea	Black Sea	Group 2—Georgian
Holmskij Market 3 3	CRXR15P59	Russia	resistant	Black Sea	Black Sea	Group 2—Georgian
Holmskij Market 4 2	H3R07P25	Russia	resistant	Black Sea	Black Sea	Group 2—Georgian
Amestec de Sovier OP 1	H4AR27P28	Moldova	tolerant	Black Sea	Black Sea	Group 2—Georgian
Adighei OP 2	H4AR26P102	Moldova	tolerant	Black Sea	Black Sea	Group 1 – Spanish-Italian / English
Georgia Imereti 8	HF1BR30P71	Georgia	resistant	Black Sea	Georgian 2	Group 2—Georgian
Georgia Imereti 11	HF1BR30P179	Georgia	resistant	Black Sea	Georgian 2	Group 2—Georgian
Georgia Imereti 9	HF1BR30P84	Georgia	resistant	Black Sea	Georgian 2	Group 2—Georgian
Georgia Imereti 10	HF1BR30P103	Georgia	tolerant	Black Sea	Georgian 2	Group 2—Georgian
Rimski OP 2	CRR02P41	Russia	resistant	Black Sea	Georgian 2	Group 2—Georgian
B X 3 OP 2	CRR02P96	Russia	tolerant	Black Sea	Georgian 2	Group 2—Georgian
Maykop VIR 1	CRXR17P48	Russia	resistant	Black Sea	Georgian 2	Group 2—Georgian
Georgia Kakheti 2	HF1BR23P52	Georgia	tolerant	Black Sea	Georgian 2	Group 2—Georgian
Carpinus betulus	C. betulus	New Jersey, USA	n/a	Outgroup	Outgroup	Outgroup

underwent a size-selection clean-up by mixing 0.5 v/v solid phase reversible immobilization magnetic beads (Agen-court Ampure XP; Beckman Coulter, Brea, CA) with the samples. The 96-well plate was then placed on a magnetic stand where the supernatant was removed and followed by a 70% ethanol wash to remove fragmented DNA < 300 bp. The magnetic bead and ethanol wash was repeated twice for a total of three washes. Individual library samples next went through PCR amplification using primers containing sequences that would allow for flow cell binding during Illumina (San Diego, CA) sequencing. Thermal cycling conditions for amplification were as follows: initial denaturation at 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 library samples were then quantified using a fluorometer (Qubit 3.0; Thermo Fisher Scientific). Samples were diluted to 5 ng/μL and pooled in 48-plex. Pooled libraries underwent another cleanup using the previously described procedure but with only two ethanol washes. Final pooled libraries were quantified using the fluorometer and were checked for quality using high-resolution automated electrophoresis (2100 Bioanalyzer System; Agilent Technologies, Santa Clara, CA). Pools were developed so that all samples were divided evenly across five HiSeq 2500 lanes (Illumina) (2 × 150 paired-end) and sequenced by Genewiz, Inc. (South Plainfield, NJ) with a 30% PhiX spike-in.

Marker calling

Raw FASTQ files of Illumina reads were filtered using the software pipeline GBS-SNP-CROP v.4.1 (Melo et al. 2016). The 5' end of reads were first scanned for the barcode sequence and expected cut site and trimmed while appending the barcode to the read for further processing. The 3' ends of reads were similarly scanned for the common adapter and expected cut site and truncated. Trimmomatic (Bolger et al. 2014) was then used to remove adapter sequences and trim reads according to default quality filters. Retained high quality reads were then searched for appended barcodes and demultiplexed. A FASTQ file was compiled for each of the two reads (R1 and R2) for each sample. The GBS-SNP-CROP software then proceeded with the de novo assembly protocol following default parameters to generate the mock reference genome.

After this, the BWA algorithm (Li and Durbin 2009) was called on to align each sample's reads to the mock reference genome based on sequence similarity. SAMtools (Li et al. 2009) was then used to sort the output of BWA, select reads that aligned properly, and convert the sequence alignment/map (SAM) files to binary alignment/map (BAM) files. Putative SNPs and their read counts were stored, and each genotype was parsed so that only reads polymorphic to the reference were kept. Once samples were processed individually,

they were mined to produce one non-redundant master matrix of all potential SNPs. The master matrix was filtered so that retained reads had a minimum depth of the putative secondary allele. Only bi-allelic markers were retained by comparing the secondary depth to the total depth at a given locus. Additional filtering steps included retaining only markers with ≤ 0.10 missing data across all 222 accessions and a minor allele frequency (MAF) > 0.05 to ensure that initial SNP calls were not due to PCR or sequencing errors. Polymorphism information content (PIC) values were calculated for the entire set of filtered biallelic SNP markers using the following formula derived from Botstein et al. (1980):

$$PIC = 1(2MAF^2 + (1 - MAF)^2) - (2MAF^2(1 - MAF)^2)$$

Where 1 – squared allele frequencies is the calculated heterozygosity per SNP marker minus the identical heterozygous trios ($2MAF^2(1 - MAF)^2$). The 1,250 most polymorphic SNP markers, as determined by the PIC formula above, were chosen for all subsequent downstream data analyses. This more limited set of SNP markers was utilized to reduce computational time and memory constraints associated with large SNP dataset inputs for the NJ and STRUCTURE analyses and to ensure the use of identical datasets for all three analyses (NJ, STRUCTURE, and DAPC).

Data analysis

NTSYSpc 2.21 software was used to conduct genetic similarity estimates and a clustering analysis using the neighbor-joining (NJ) method (Saitou and Nie 1987; Gascuel 1997). Briefly, pairwise comparisons of SNP allele calls between individual hazelnut samples were transformed into a genetic distance matrix using the SIMGEND module. Clustering analysis was performed with the genetic distance matrix generated above using the neighbor-joining NJOIN module. The CPH (cophenetic values converted to a matrix) and MXCOMP (Mantel test statistic for matrix comparisons) modules were utilized to assess the goodness of fit between the NJ dendrogram and the original genetic distance matrix (cophenetic correlation test) using 999 permutations. Bootstrap values for the dendrogram were generated from 1,000 resampled genetic distance matrices using the RESAMPLE module and were calculated using the CONSENS module with the majority rule option and a minimum support value of 0.500.

STRUCTURE 2.3.4 (Pritchard et al. 2000) was used to conduct model-based clustering and determine the optimal population structure for grouping the accessions. The same 1,250 SNP loci were utilized, and all alleles were assumed to be both independent and in linkage equilibrium. STRUCTURE was first run across a range of (*K*) values from 2 to 25 (in increments of one, with 10 replicates within each

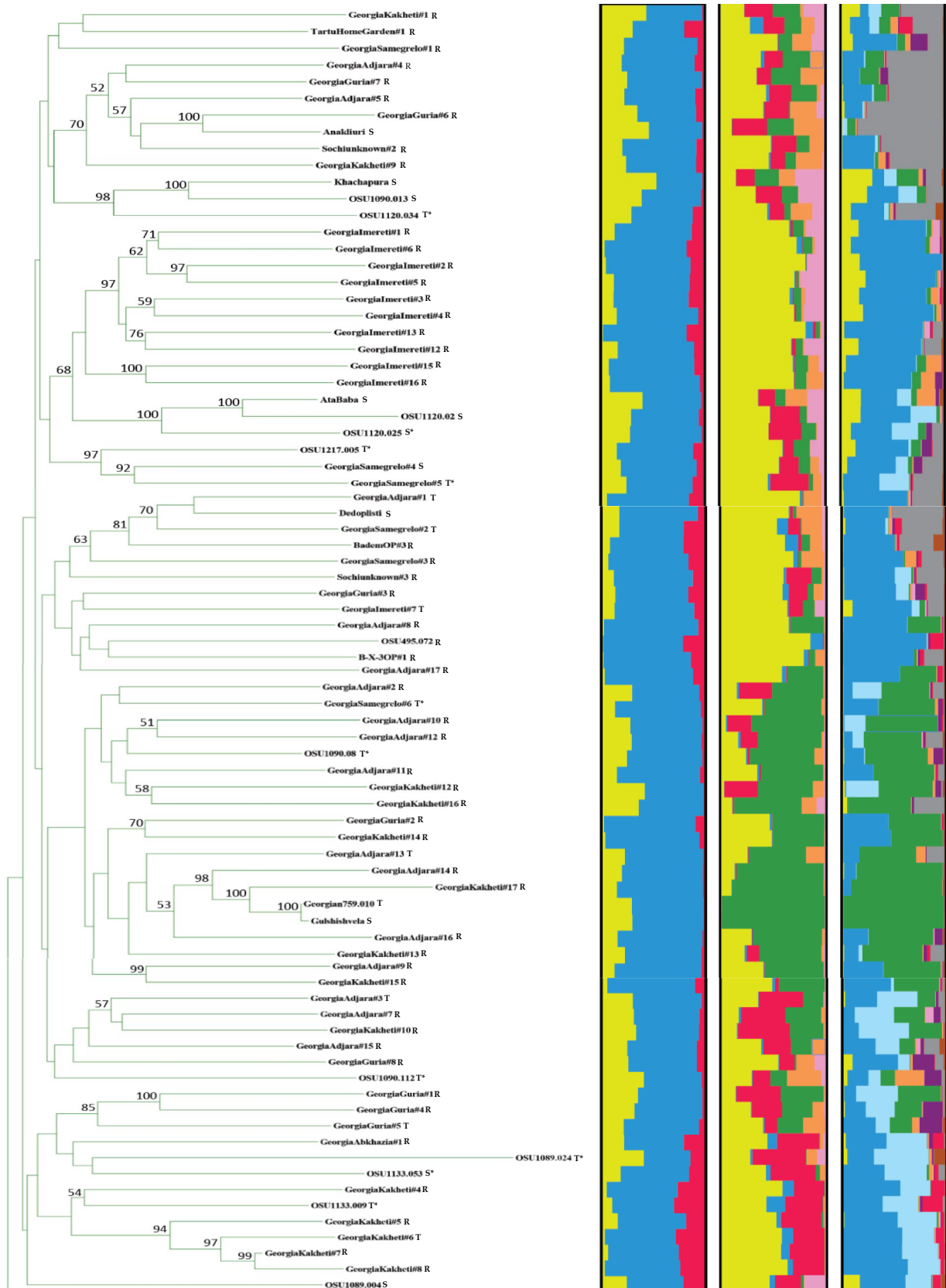


Fig. 1 Georgia hazelnut group node of dendrogram ($n=79$), with $K=3$, $K=6$, and $K=10$ STRUCTURE results aligned. Bootstrap values >50 are presented in the dendrogram. R denotes resistant, T denotes tolerant, S denotes susceptible. Disease responses with an * were recorded in Corvallis, OR

run) to find the most parsimonious value of (K). Settings included 20,000 run iterations for burn-in length and 50,000 Markov chain Monte Carlo run iterations following the burn-in for each (K) value. Default settings were used for all other parameters. Once the optimal (K) value was determined using the $\Delta(K)$ method (Evanno et al. 2005) as implemented in STRUCTURE SELECTOR (Li and Liu 2017), figures were generated for each selected (K) value solution (see results; Fig. 1–5 STRUCTURE panels) in which each row represents one accession and each color represents a genetic population. STRUCTURE software can capture genetic admixture within individuals using the admixture ancestry model, so individual entries may be represented by multiple colors corresponding to more than one genetic population background in their respective rows.

Population structure was further investigated with the same 1,250 SNP loci in the NJ and STRUCTURE analyses utilizing discriminant analysis of principal components (DAPC) in the *adegenet* package (Jombart 2008) for R software (R core Team 2016), where multilocus genotype data are dimensionally reduced first using principal component analysis (PCA) followed by linear discriminant analysis (DA). Briefly, the DAPC analysis was initiated using PCA analysis to determine PCA eigenvalues. The next step was determining the number of clusters in the data, without prior group information, utilizing the *find.clusters()* function which runs sequential k -means clustering over increasing values of (k), followed by optimal cluster determination using the Bayesian information criterion (BIC). The optimal number of PCs to be retained was cross-validated using the *xvalDapc()* function (Jombart and Collins 2015). The resultant optimal clusters were visualized using a DAPC scatterplot, while group membership was visualized using composition plot assignment in *adegenet*.

Results

Sequencing and SNP marker data

Sequencing data, including the total number of bases and markers generated, are summarized in Table 2. The 39,251 SNP and Indel markers generated from the ddRADseq sequencing libraries were filtered and sorted by PIC, where a final marker dataset of the 1,250 biallelic SNP markers with the highest PIC value were utilized in the current analyses. PIC values for the sorted 1,250 SNP markers ranged from a

high of 0.375 (maximum PIC value for biallelic SNP markers) to 0.284, with an average PIC value per marker across all study accessions of 0.307. This high average PIC value per marker indicated that the SNP marker dataset utilized in the current study was capable of capturing sufficient genetic diversity amongst all accessions in the current analyses.

Neighbor-joining dendrogram

The results of the NJ clustering analysis are presented in Figs. 1–5. The 221 hazelnut accessions formed five broad, distinct groups in the NJ dendrogram which largely corresponded to their geographic regions of origin, while most of the standard cultivars aligned with previous studies. Moving from the least to most basal group of the dendrogram, the first and largest group in the dendrogram ($n=79$) held most of the accessions from the Republic of Georgia and thus was labeled the Georgian group (Fig. 1). Accessions spanned several growing regions across the country, including Abkhazia, Adjara, Guria, Kakheti, Imereti, and Samegrelo. Except for the Kakheti region, all these regions occupy the western half of the Republic of Georgia, with Abkhazia, Adjara, Guria, and Samegrelo bordering the Black Sea. The specific region of origin was unknown for some Georgian samples. This group contained several standard cultivars grown across multiple regions of Georgia, including ‘Anakliuri’, ‘Dedoplistiti’, and ‘Gulshishvela’. In total, seven accessions within this group did not originate directly from Georgia. These included five obtained from seed collected in southern Russia, one from Azerbaijan, and one from Estonia. Except for the single Estonian accession (Tartu Home Garden #1), these accessions come from regions immediately bordering the Republic of Georgia, so their clustering with the Georgian germplasm is unsurprising since the accessions may have originated in Georgia or may be open-pollinated progeny of a Georgian accession.

The adjacent group ($n=55$) held accessions that mostly originated across central Europe and was labeled the Central European group (Fig. 2). Nearly all accessions within this group had known origins spanning Germany, Poland, and Moscow (Russia). A few accessions with origins in Latvia, Estonia, and Crimea were placed in this group, while just one or two accessions with origins in Albania (Albania 55), Romania (‘Cozia’), Sweden (Sweden COR 627), Finland (Finland COR 187), England (‘Cutleaf’), France (‘Aurea’ and ‘Aveline d’ Angleterre’), Italy (‘Tonda Rossa’ and ‘Tonnolella’), and Georgia (OSU 1133.068) were also placed in the Central European group. Accessions from the United States placed in this group included ‘Gasaway’ (from Washington, but of unknown origin), its offspring ‘Delta’ (from ‘Montebello’ x ‘Gasaway’, though not first generation), and OSU 408.040 (from Minnesota, but of unknown origin).

Table 2 Summary of raw and filtered Illumina sequencing data of the 222 accessions included in this study

Sequencing data characteristics	Value	All Sequencing Sets
Phred quality (Q) score	Minimum	36.3
	Maximum	37.2
	Mean	36.8
Q score > 30 (%)	Minimum	84.5
	Maximum	87.8
	Mean	86.5
Sequence data (Gb)	Total	~ 630
	Average per sample	~ 2.60
Reads (no.)	Total	2,086,416,858
	Average per sample	8,693,404
	Usable paired reads	1,911,430,731
	Unique Clusters/Centroids	10,924
	Total Markers (including SNPs and INDELS)	39,251

The next adjacent group ($n = 26$) contained primarily Spanish and Italian cultivars and was therefore labeled the Spanish-Italian group (Fig. 3). Some of the cultivars from Spain or Italy included Siciliana, Culpla, Tonda Gentile delle Langhe, Negret, Ratoli, Tonda Romana, and Barcellona. Single accessions from other countries within this group include Luisa OP#1 from seed obtained in Lithuania, 'Sodliger' from the former country of Yugoslavia, Zugdui OP#3 from seed originating in southern Russia, and 'Macrocarpa' from England. The OSU cultivar Willamette was also included; its female parent is Montebello (syns. Siciliana, Nocchione).

The next adjacent group ($n = 17$) contained mostly English cultivars such as Nottingham, DuChilly, Daviana, Cosford, and Garibaldi, and was therefore labeled the English group (Fig. 4). Accessions from outside of England within this group included 'Skorospelka' from the Republic of Georgia, 'Rote Zellernuss' from the Netherlands, 'Ugbrooke' and 'Whiteheart' from New Zealand, and 'Henneman #3' (from Oregon, but of unknown origin).

The final and most basal group ($n = 44$) contained accessions with origins spanning several countries that either bordered the Black Sea or were in close proximity and was therefore labeled the Black Sea group (Fig. 5). Many accessions in this group originated in Turkey, southern Russia, and Georgia, while other countries each contributed just one or two accessions: Bulgaria ('Bulgaria XI-8'), Croatia ('Istrski Duguljasti'), Germany ('Burchardt's Zellernuss'), Italy ('San Benedetto'), Moldova ('Adighei OP #2' and 'Amestec de Sovier OP #1'), and Crimea ('Pioneer' and Simferopol Market 1B #1) each contributed just one or two accessions.

Structure populations

For the STRUCTURE clustering analysis, accessions were arranged in the same order as listed in the dendrogram. In determining the most parsimonious number of populations, the greatest support, as determined by the largest $\Delta(K)$ values, was found at $(K) = 3$, $(K) = 6$, and $(K) = 10$, with $(K) = 10$ having the most support (Supp Fig. 1). For all (K) value solutions, and in particular for $(K) = 6$ and $(K) = 10$, many accessions displayed significant admixture, wherein portions of their genomes originated from different genetic populations, making them difficult to place in any one population, especially when accessions had less than 50% of their genetic backgrounds in any one population.

At $(K) = 3$, accessions were organized into three large groups that tended to follow broad geographic origins. For example, the first group held accessions of mostly Georgian origin while also holding several accessions from the nearby Black Sea region regions of southern Russia and Turkey (Fig. 1). The second group was a mix of accessions with primarily central European and Russian origins, with many accessions from Germany, Poland, and Russia, as well as a few from other nearby countries (Fig. 2). The final group was a mix of known Spanish-Italian and Turkish accessions, with smaller contributions from Russia as well as other southern European countries (Fig. 3). Interestingly, accessions known to originate from England displayed fairly even admixture between the central European and Spanish-Italian/Turkish groups (Fig. 4).

At $(K) = 6$, some of the genetic populations from $(K) = 3$ were broken down into subgroups; however, the major delineations resolved at $(K) = 3$ were still present in the overall organization of accessions. The Georgian

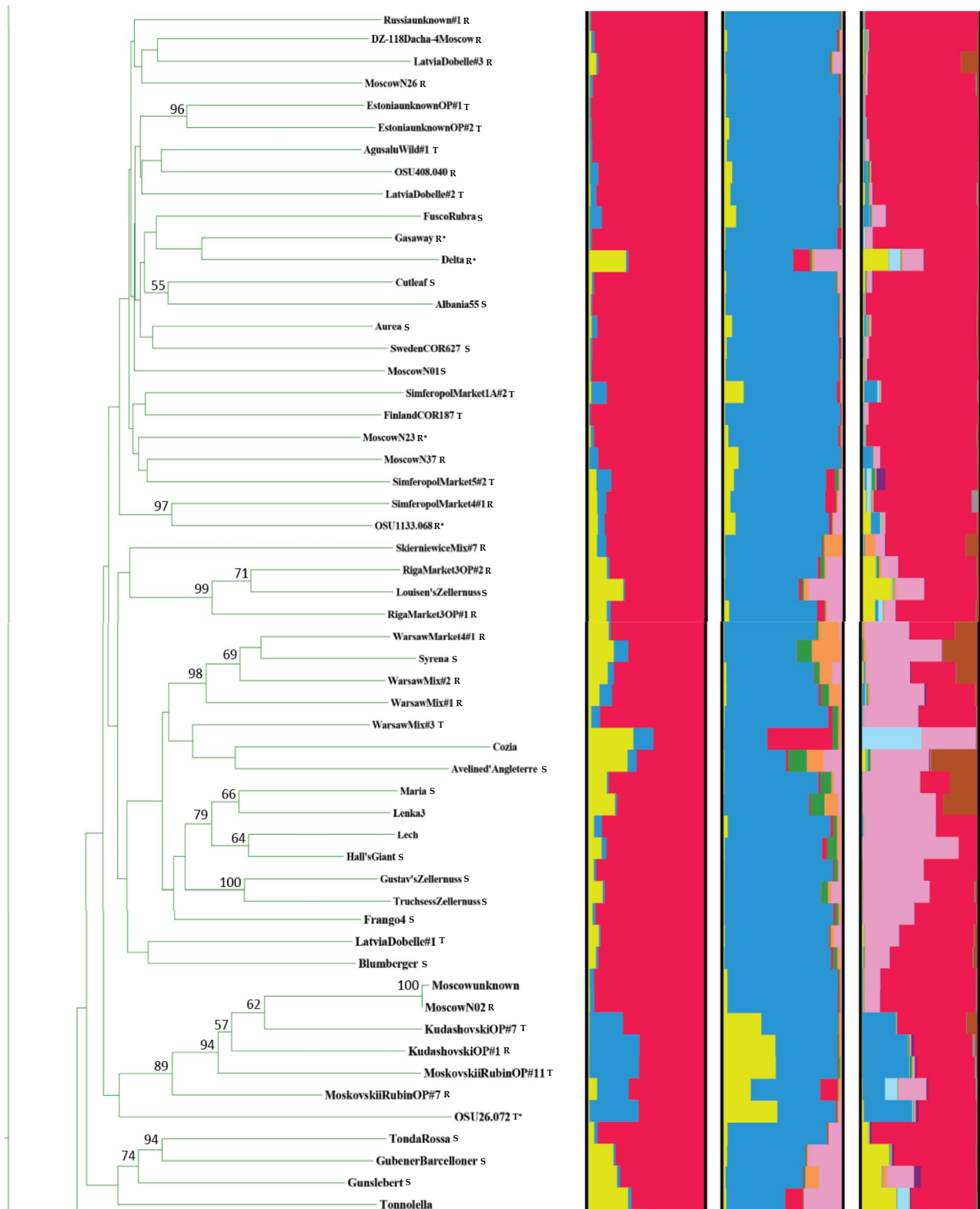


Fig. 2 Central European hazelnut group node of dendrogram ($n=55$), with $K=3$, $K=6$, and $K=10$ STRUCTURE results aligned. Bootstrap values > 50 are presented in the dendrogram. R denotes resistant, T

denotes tolerant, S denotes susceptible. Disease responses with an * were recorded in Corvallis, OR

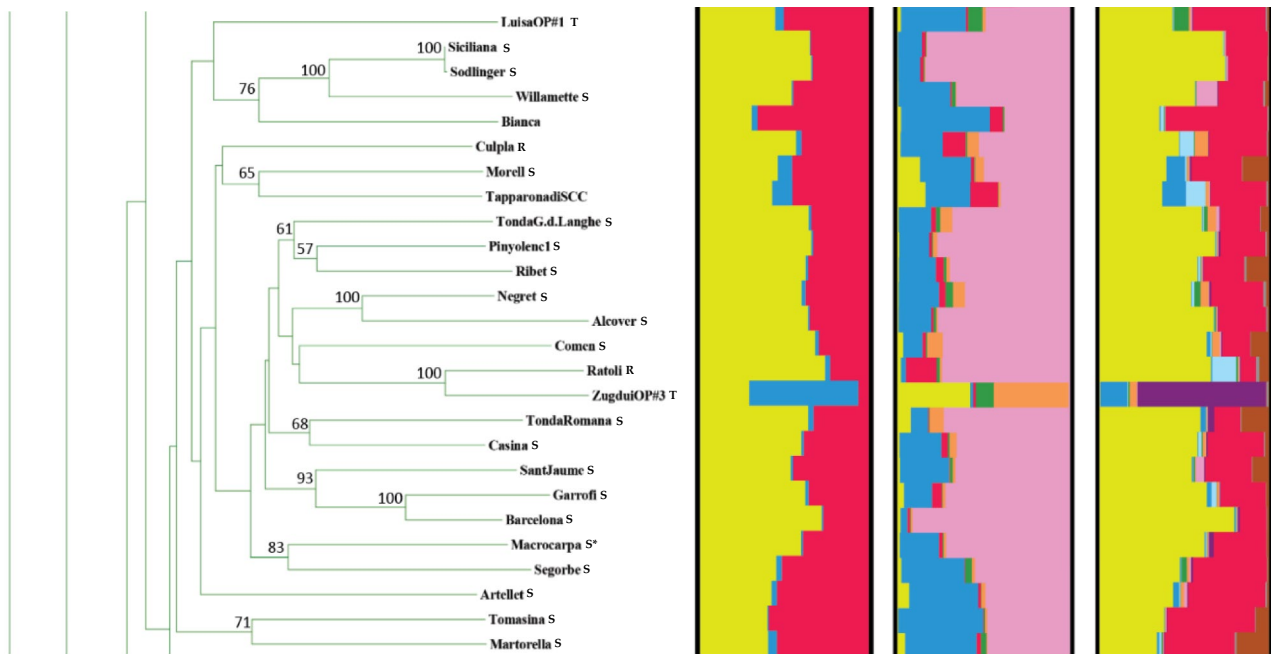


Fig. 3 Spanish-Italian hazelnut group node of dendrogram ($n=26$), with $K=3$, $K=6$, and $K=10$ STRUCTURE results aligned. Bootstrap values > 50 are presented in the dendrogram. R denotes resistant, T

denotes tolerant, S denotes susceptible. Disease responses with an * were recorded in Corvallis, OR

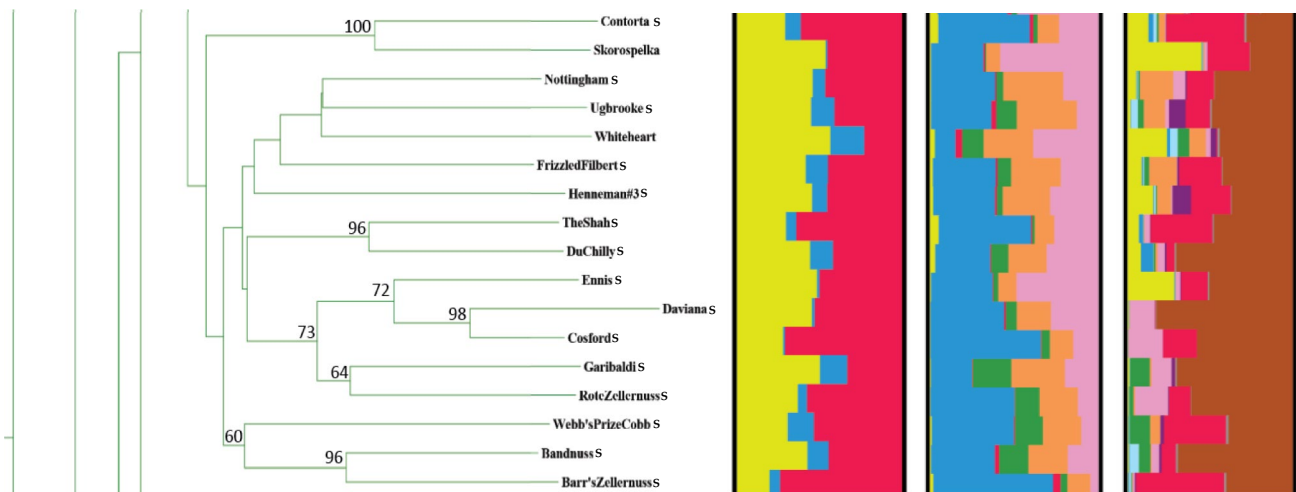


Fig. 4 English hazelnut group node of dendrogram ($n=17$), with $K=3$, $K=6$, and $K=10$ STRUCTURE results aligned. Bootstrap values > 50 are presented in the dendrogram. R denotes resistant, T

denotes tolerant, S denotes susceptible. Disease responses with an * were recorded in Corvallis, OR

group of (K) = 3 was divided into two subgroups (Fig. 1), the larger of which held Georgian accessions and a few accessions from southern Russia, while the smaller held only accessions from Georgia. Both Georgian groups contained accessions collected from multiple regions within the Republic of Georgia, many of which were selected from open-pollinated seed. As such, these groups did not appear to be associated with specific regions of the

country and suggest movement and/or mixing of plant materials between regions.

The Central European/Russian group found at (K) = 3 formed once again at (K) = 6, with little change found in admixture of the samples within the group (Fig. 2). The remaining group at (K) = 3 that held the Spanish-Italian and Turkish accessions was broken into three subgroups at (K) = 6 (Fig. 3–5). Those of Spanish-Italian origin were

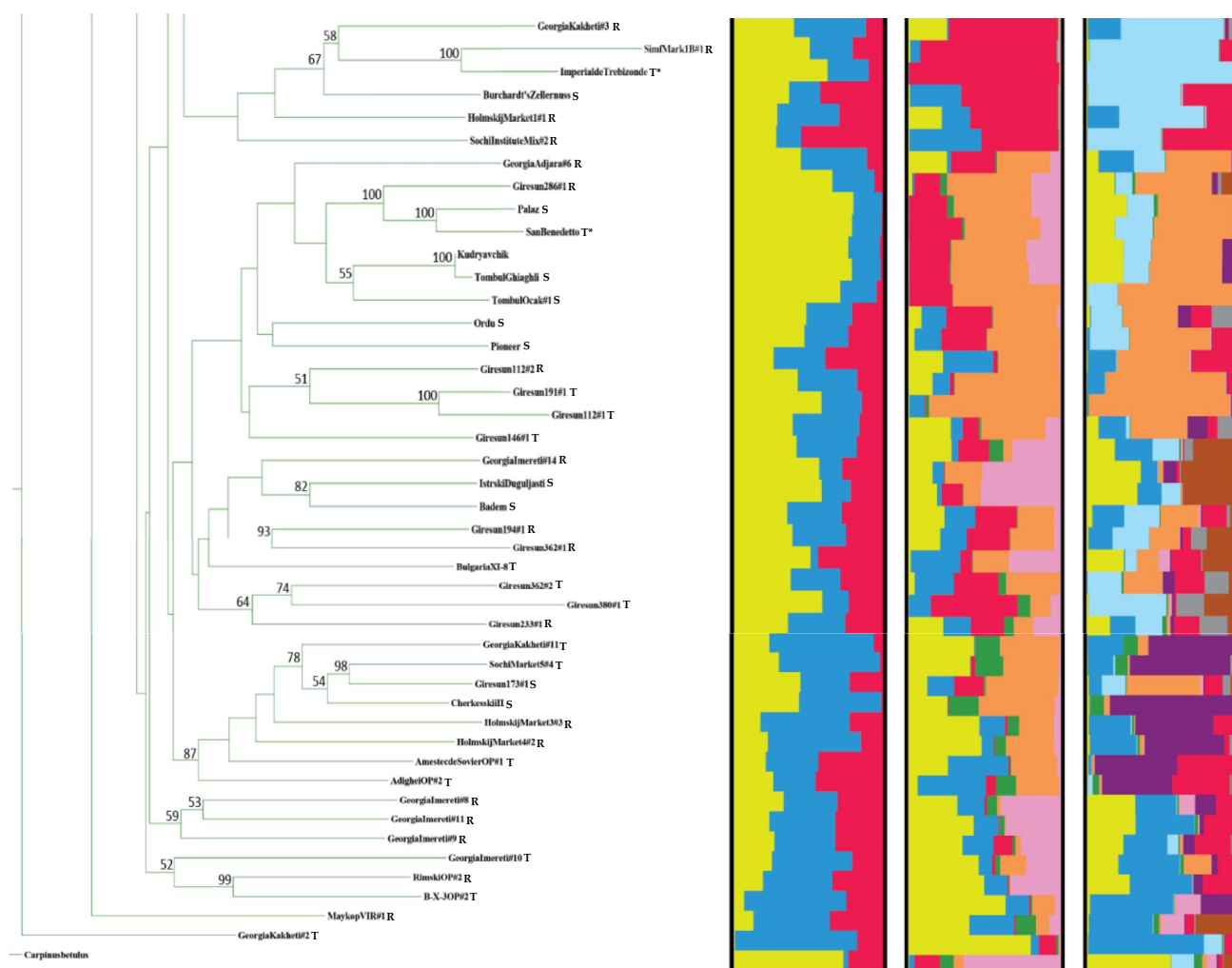


Fig. 5 Black Sea hazelnut group node of dendrogram ($n=44$), with $K=3$, $K=6$, and $K=10$ STRUCTURE results aligned. Bootstrap values > 50 are presented in the dendrogram. R denotes resistant, T

denotes tolerant, S denotes susceptible. Eastern filbert blight disease responses with an * were recorded in Corvallis, OR

separated from the accessions with known Turkish heritage, with the newly formed Spanish-Italian group containing single accessions from Russia, Lithuania, England, Georgia, Croatia, Turkey, and the former Yugoslavia (Fig. 3). Further, two separate groups were resolved that held accessions with known Black Sea origins (Fig. 5). Both had several accessions from Russia and Turkey, with very minor contributions from other countries. Specific geographic regions around the Black Sea could not be established for the two groups based on accession assignments. Many accessions in the Black Sea groups displayed significant admixture between the two Black Sea groups and with other groups, making their placements less definitive. Accessions from England once again displayed significant admixture, this time between the Central European group, the Spanish-Italian group, and one of the Black Sea groups, with the greatest contribution from the Central European group (Fig. 4).

At (K) = 10, further subgrouping was observed while still maintaining the broader organization present at both (K) = 3 and (K) = 6 group assignments. For example, four groups were now established within the (K) = 3 Georgian group (Fig. 1). Again, each Georgian subgroup contained accessions from multiple regions across the country, suggesting the movement or mixing of plant materials within the Republic of Georgia. An example of such movement includes many orchards in Kakheti, which were planted using rooted suckers from western Georgia. The (K) = 10 analysis also continued to support the premise that the Georgian material was genetically unique within the broader population of plants included in this study. Further, each of the four groups held a few accessions collected from Russia, Turkey, Crimea, Azerbaijan, Estonia, or Germany, suggesting that these accessions may have originated in Georgia.

Also at (K) = 10, two Central European groups were formed (Fig. 2). The larger group included accessions with origins that spanned a greater geographic area, had a significant number from Russia, and included accessions from Germany, Poland, Latvia, Estonia, and Crimea, and a few other countries. The smaller Central European subgroup was dominated by accessions of German and Polish origin, with just one accession each from Romania and France.

The same group of primarily Spanish-Italian cultivars that was resolved in (K) = 6 formed again at (K) = 10 (Fig. 3). The English accessions that previously showed a high degree of admixture between two or more groups formed their own group at (K) = 10 (Fig. 4). This group contained a Russian accession and a Dutch accession, as well as one from Oregon and two from New Zealand. A small group of Turkish germplasm was resolved that contained just one accession each from Georgia, Russia, and Crimea each (Fig. 5). Additionally, a small Black Sea group formed containing accessions from Russia, the Republic of Georgia, and Moldova (Fig. 5). Finally, it should be noted that the (K) = 10 STRUCTURE solution (the K value with the highest $\Delta(K)$ support; Supp Fig. 1) best matched the finer structural grouping details shown in the NJ dendrogram (Figs. 1–5).

DAPC analysis

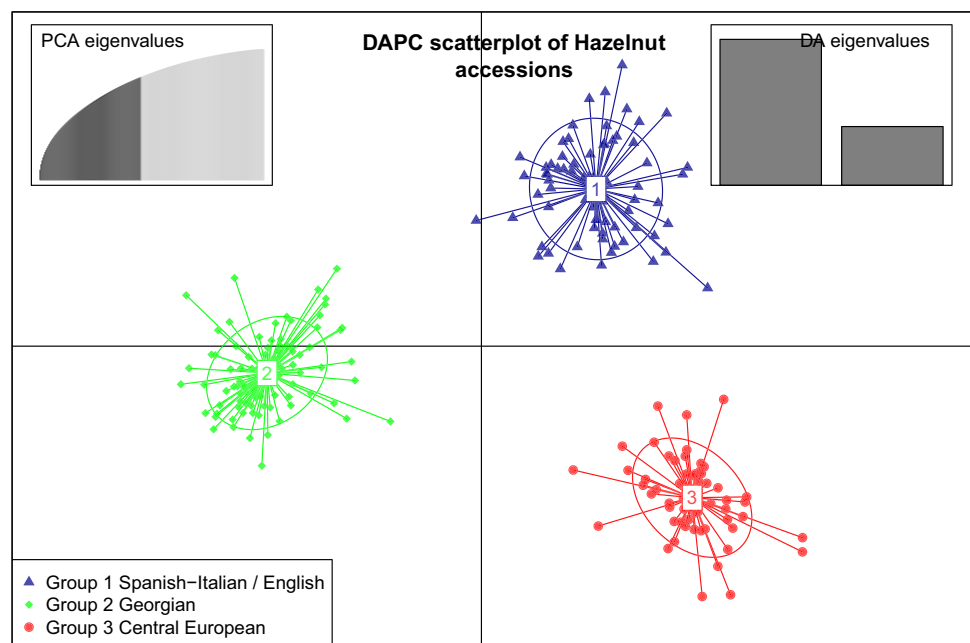
Sequential k -means clustering (*find.clusters()*), BIC results (Supp Fig. 2), *xvalDapc()* cross-validation, and the DAPC scatterplot and compoplots (Fig. 6 and Supp Fig. 3, respectively) all supported 3 primary groups of hazelnut accessions. These results best matched the (K) = 3 solution of the STRUCTURE results as well as the broadest delineations

shown in the NJ dendrogram (Figs. 1–5). DAPC group 1 (n = 70) was comprised primarily of Spanish-Italian and English accessions (DAPC Group 1 Spanish-Italian/English – Fig. 6; Table 1). This DAPC group corresponded to the combination of most accessions listed as the Spanish-Italian and the English groups in the (K) = 10 STRUCTURE solution (Figs. 1–5; Table 1). DAPC group 2 (n = 94) was comprised of Georgian accessions (DAPC Group 2 Georgian – Fig. 6; Table 1). This DAPC group corresponded to the combination of most accessions listed as Georgian groups 1–4 as well as a limited number of Black Sea accessions (n = 6) in the (K) = 10 STRUCTURE solution (Figs. 1–5; Table 1). DAPC group 3 (n = 57) was comprised of primarily Central European accessions (DAPC Group 3 Central European – Fig. 6; Table 1), and corresponded to the majority of the Central European groups 1 and 2 in the (K) = 10 STRUCTURE solution (Figs. 1–5; Table 1).

Discussion

As shown in the results, assignment of most accessions based on 1,250 SNP markers corresponded to known geographic origins of the plant materials studied, substantiating most findings in previous SSR-based hazelnut diversity studies (Boccacci et al. 2006; Gökirmak et al. 2009; Gürçan et al. 2010; Muehlbauer et al. 2014). Due to the use of many more markers than most previous work, our results strongly reinforce the premise that *C. avellana* germplasm contains high genetic diversity. Additionally, evidence of admixture between genetic groups was also revealed in some accessions.

Fig. 6 Discriminate analysis of principle components (DAPC) of 221 hazelnut accessions and one outgroup species based on 1,250 single nucleotide polymorphism (SNP) markers. DAPC results indicated that the 222 study entries were best described as belonging to 3 populations separated into a Spanish-Italian / English group, a Central European group, and a Georgian group



Overall, there was general consensus was found between the results of the NJ dendrogram, STRUCTURE, and DAPC analyses, as divisions between genetic populations of the STRUCTURE and DAPC analyses occurred along the boundaries between different clusters of the NJ dendrogram (Fig. 1–5; Table 1). Consensus could also be seen between the results of the STRUCTURE analysis across the different (K) values and the 3 DAPC groupings. When increasing the number of genetic populations from three to six and six to ten in the STRUCTURE results, divisions between groups found at lower (K) values remained in similar positions, while the STRUCTURE (K)=3 solution closely matched the 3 groups identified in the DAPC analysis. Additionally, in some cases, groups remained relatively unchanged across two different STRUCTURE (K) values. When groups were subdivided at the highest and most parsimonious STRUCTURE (K)=10 value, the original group boundaries remained but additional subdivisions formed within the previous group.

As the STRUCTURE (K) value increased, more narrow groups of related accessions were often resolved. Regardless of STRUCTURE (K) value, the standard accessions with known origins clustered together as expected, with little change in overall groupings compared to previous SSR studies (Boccacci et al. 2006; Gökirmak et al. 2009; Gürcan et al. 2010; Muehlbauer et al. 2014). For example, at STRUCTURE (K)=3, as well as the 3 groupings identified in the DAPC analysis, very broad groups were established that reflected the known origins of many of the accessions. These groups primarily held accessions with origins from central Europe, southern Europe, and the Republic of Georgia. At (K)=10, where a larger number of genetic groups was resolved, these groupings continued to reflect findings from previous studies, wherein clustering among accessions from more narrow geographic origins was defined for many groups. For example, at (K)=10, accessions with known origins from England, Spain and Italy, and Turkey were placed into their own genetic populations, just as they had been in previous SSR-based studies (Boccacci et al. 2006; Gökirmak et al. 2009; Muehlbauer et al. 2014).

In other cases, the increased number of groups did not lead to increased resolution of geographic origin. This indicates that while many hazelnut accessions' genetic profiles remain unique to their geographic origins, many others represent the movement and intermixing of plant materials among regions over time. The single Georgian group that formed at STRUCTURE (K)=3 and in the DAPC analysis was divided into four genetic populations at STRUCTURE (K)=10, but accessions in each of the four groups spanned many of the same regions. As such, these genetic populations could not be separated by the region in which the seed originated. A similar situation occurred with the central European germplasm. These accessions were grouped

together at STRUCTURE (K)=3 and in the DAPC analysis but were not fully resolved at STRUCTURE (K)=10. Instead, two Central European subgroups formed at (K)=10, with one group completely overlapping the other in terms of geographic location. While the Georgian germplasm had not been studied previously, many of the accessions that formed the two Central European subgroups were included in Muehlbauer et al. (2014) (SSR-based) where they formed just one Central European subgroup.

Significant genetic admixture was detected at all three STRUCTURE (K) values but, as expected, was most prominent at (K)=10. Additionally, at STRUCTURE (K)=10, many accessions had significant contributions from more than one genetic population, and some accessions did not clearly belong to any one population. Known geographic origin, along with the dendrogram, STRUCTURE, and DAPC results, was used to tentatively place many of the highly admixed samples into genetic populations. The widespread admixture signals in the STRUCTURE results were an unsurprising result, as hazelnut is an outcrossing crop with sporophytic self-incompatibility (Thompson 1979). As such, following human vectoring, previously separated hazelnut accessions from different genetic populations could freely cross with each other. A small group of accessions clustered together in the dendrogram displayed an exceptionally high degree of admixture in the STRUCTURE (K)=10 analysis. These included Giresun 194 #1, Giresun 233 #1, Giresun 362 #1, Giresun 362 #2, Giresun 380 #1, and 'Badem' from Turkey, 'Istrski Duguljasti' from Croatia, 'Bulgaria XI-8' from Bulgaria, and Imereti #14 from Georgia. The accessions from Giresun originated as open-pollinated seed collected from the Hazelnut Research Institute in Giresun, so their highly admixed backgrounds are unsurprising. With the exception of 'Istrski Duguljasti' and 'Badem', all of these admixed accessions are EFB tolerant or resistant. Most of these accessions are from the Black Sea region, with several accessions from Turkey. Turkey is the top hazelnut producer in the world, and the high degree of diversity retained in the country has been previously documented (Gürcan et al. 2010; Helmstetter et al. 2019). Beyond Turkey, the Black Sea region serves as a confluence of several different hazelnut producing areas and their respective genetic populations, with Spain and Italy to the west, central Europe and Russia to the north, and Georgia to the east. Because of its central location relative to these other regions, the Black Sea region may have served as a melting pot of hazelnut germplasm with a high degree of hybridization between other genetic groups. It is worth noting that plants with origins in the Black Sea region did also form their own genetically distinct population at STRUCTURE (K)=10, though this group was very small.

Human vectoring could be responsible for some of the accessions that had unexpected assignments, even when

admixture was not a factor. For example, Zugdui OP #3 was collected in Sochi, Russia, but was placed in the middle of the Spanish-Italian group by the dendrogram (although the (K) = 10 analysis supports it as a Black Sea accession, while the DAPC analysis supports it as a Georgian accession). This accession was the only instance in which the dendrogram and STRUCTURE analyses had a clearly different result. Other accessions with placements different from what was expected based on geographic origin included Tartu Home Garden #1 from Estonia (Georgian group for all three analyses), OSU 1133.068 from Georgia (Central European group for all three analyses), 'Tonda Rossa' and 'Tonnolella' from Italy (Central European group for all three analyses), 'Luisa OP #1' from Lithuania (Spanish-Italian group in the NJ dendrogram and STRUCTURE analyses and Spanish-Italian/English group in the DAPC analysis), 'Macrocarpa' from England (Spanish-Italian group), 'Skorospelka' from Georgia (English group in the NJ dendrogram and STRUCTURE analyses and Spanish-Italian/English group in the DAPC analysis), and 'Burchardt's Zellernuss' from Germany (Black Sea group in the NJ analysis, Georgian group in the STRUCTURE analysis, and Spanish-Italian/English in the DAPC analysis). Accessions with unknown origins grown in areas beyond the native range of *C. avellana* (such as those from the United States and New Zealand) were transported to these regions and could therefore have origins in any native region within Europe or the Caucasus. Of these eight accessions (OSU 408.040, 'Gasaway', 'Delta', 'Willamette', 'Henneman #3', 'Ennis', 'Ugbrooke', and 'Whiteheart') five were included in Muehlbauer et al. (2014). Four of these were placed in groups not formed in this study. While 'Ennis' was placed in the English subgroup in the NJ dendrogram and STRUCTURE analyses and the Spanish-Italian/English group in the DAPC analysis in this study, it was placed in the Spanish-Italian subgroup in the SSR-based Muehlbauer et al. (2014) study.

The Georgian germplasm made up most the accessions that had not been genotyped previously. When included alongside the other germplasm representing the world's standard cultivars, these accessions remained in separate genetic groups from most other accessions for all three analyses in the current study. This is seen in the dendrogram (Fig. 1) as well as the STRUCTURE (K) = 3 analysis and DAPC analysis (Table 1), where the vast majority of Georgian accessions clustered to form their own group. This was a significant result, as allelic frequencies needed to differ prominently from other accessions for the Georgia germplasm to form its own group at the lowest STRUCTURE (K) value. As the (K) value increased to six and ten, the number of Georgian genetic populations increased to two and four additional groups, respectively.

Similar to the results of Muehlbauer et al. (2014), accessions expressing resistance/tolerance to EFB were found

in a majority of the genetic clusters resolved in this study. While Muehlbauer et al. (2014) found resistance/tolerance in all 11 of their consensus groups, we found resistance/tolerance in four of the five dendrogram groups, nine of the ten STRUCTURE groups, and all three DAPC groups (Table 1). The only group lacking an accession with resistance/tolerance to EFB was the NJ dendrogram and STRUCTURE (K) = 10 English defined group of 17 accessions (same set of accessions for both the dendrogram and STRUCTURE (K) = 10 analysis). Two accessions within the English group, 'Skorospelka' from the Republic of Georgia and 'Whiteheart' from New Zealand, had unknown EFB responses, so it is possible this group may possess EFB resistance/tolerance after further field testing. It is important to note that finding resistance or tolerance in different genetic backgrounds does not necessarily mean that different genes are responsible for each group's resistance or tolerance. Regardless, resistant/tolerant accessions found in novel genetic groups should be explored as potentially new sources of R-genes. Resistance screening and mapping efforts summarized in Mehlenbacher and Molnar (2021) and Mehlenbacher et al. (2023) demonstrate the increasing list of R-genes placed on new linkage groups as a wider array of plant materials is examined. The R-genes of many of the accessions included in our study have yet to be mapped, but their placement in different genetic groups could be used as a starting point for further study.

In total, 1,250 bi-allelic SNP markers were used to characterize the relatedness of accessions across a wide sampling of *C. avellana* genetic origins, whereas most previous hazelnut studies used less than 25 SSR markers, although each SSR marker was comparatively more polymorphic (greater allelic diversity for any single SSR versus a single bi-allelic SNP). Due to the nearly thousandfold increase in number of markers used, the genome of each accession was captured in far greater detail than SSR-based studies. This allowed for higher resolution grouping of accessions and better capture of genetic admixture patterns in the STRUCTURE analysis.

However, even with the greater number of markers used here, our results largely reflected and corroborated the findings of previous studies. Many of the same groups that were identified as having specific geographic origins in previous studies were also established here for the NJ dendrogram and the STRUCTURE (K) = 10 analyses. These include the Spanish-Italian, English, Central European, and Turkish/Black Sea groups. Of course, the number of groups identified and the geographic regions represented depend on the specific accessions included, the number of accessions from each origin, and the markers used in each study, so some differences are to be expected. Boccacci et al. (2006) surveyed 78 hazelnut cultivars using 16 SSR markers and established a large Spanish-Italian group, a Turkish group, an English group, and a small mixed group. Gökirmak et al. (2009)

Table 3 European hazelnut accessions with new assignments compared to those of previous studies

Accession	Geographic Origin	Previous study	Previous study assignment	New assignment
Gunslebert	Germany	Boccacci et al. 2006	Spanish-Italian	Central European
Contorta	England	Boccacci et al. 2006	Turkish	English
Sodlinger	Yugoslavia	Boccacci et al. 2006	English	Spanish-Italian
Gasaway	Washington	Boccacci et al. 2006	Mixed	Central European
Halls' Giant	France/Germany	Boccacci et al. 2006	Mixed	Central European
Fusco Rubra	Germany	Gökirmak et al. 2009	No group	Central European
Gasaway	Washington	Gökirmak et al. 2009	No group	Central European
Albania 55	Albania	Gökirmak et al. 2009	No group	Central European
Cutleaf	England	Gökirmak et al. 2009	No group	Central European
Finland COR 187	Finland	Gökirmak et al. 2009	No group	Central European
Sweden COR 627	Sweden	Gökirmak et al. 2009	No group	Central European
OSU 408.040	Minnesota	Gökirmak et al. 2009	No group	Central European
Aurea	France	Gökirmak et al. 2009	No group	Central European
Aveline d'Angleterre	France	Gökirmak et al. 2009	No group	Central European
Syrena	Poland	Gökirmak et al. 2009	No group	Central European
OSU 26.072	Russia	Gökirmak et al. 2009	No group	Central European
OSU 495.072	Russia	Gökirmak et al. 2009	No group	Georgia
Rote Zellernuss	Netherlands	Gökirmak et al. 2009	No group	English
Henneman #3	Oregon	Gökirmak et al. 2009	No group	English
Barr's Zellernuss	England	Gökirmak et al. 2009	Central European	English
Cozia	Romania	Gökirmak et al. 2009	Black Sea	Central European
Whiteheart	New Zealand	Gökirmak et al. 2009	Black Sea	England
Ugbrooke	New Zealand	Gökirmak et al. 2009	Black Sea	England
Webb's Prize Cobb	England	Gökirmak et al. 2009	Black Sea	England
Skorospelka	Russia	Gökirmak et al. 2009	Black Sea	England
Artellet	Spain	Gökirmak et al. 2009	English	Spanish-Italian
Badem	Turkey	Gökirmak et al. 2009	Spanish-Italian	Black Sea
Delta	Oregon	Gökirmak et al. 2009	Spanish-Italian	Central European
Tonnolella	Italy	Gökirmak et al. 2009	Spanish-Italian	Central European
Tonda Rossa	Italy	Gökirmak et al. 2009	Spanish-Italian	Central European
Gubener Barcelloner	Germany	Gökirmak et al. 2009	Spanish-Italian	Central European
Tonnolella	Italy	Gürçan et al. 2010	Caucasian	Central European
Negret	Spain	Gürçan et al. 2010	Caucasian	Spanish-Italian
Ordu	Turkey	Gürçan et al. 2010	No group	Black Sea
Khachapura	Georgia	Gürçan et al. 2010	Turkish	Georgian
Ribet	Spain	Gürçan et al. 2010	No group	Spanish-Italian
Tapparona di SCC	Italy	Gürçan et al. 2010	No group	Spanish-Italian
Tonda Rossa	Italy	Gürçan et al. 2010	Spanish-Italian	Central European
Skorospelka	Russia	Muehlbauer et al. 2014	Black Sea	English
Zugdui OP #3	Russia	Muehlbauer et al. 2014	Black Sea	Spanish-Italian
Sochi unknown #2	Russia	Muehlbauer et al. 2014	Black Sea	Georgian
OSU 495.072	Russia	Muehlbauer et al. 2014	Black Sea	Georgian
Georgian 759.010	Georgia	Muehlbauer et al. 2014	Black Sea	Georgian
B-X-3 OP #1	Russia	Muehlbauer et al. 2014	Black Sea	Georgian
Ata Baba	Azerbaijan	Muehlbauer et al. 2014	Black Sea	Georgian
Badem OP #3	Russia	Muehlbauer et al. 2014	Black Sea	Georgian
B-X-3 OP#2	Russia	Muehlbauer et al. 2014	No group	Black Sea
Rimski OP #2	Russia	Muehlbauer et al. 2014	Spanish-Italian	Black Sea
Maykop VIR #1	Russia	Muehlbauer et al. 2014	Wild <i>C. avellana</i>	Black Sea
Contorta	England	Muehlbauer et al. 2014	Mixed	English

Table 3 (continued)

Accession	Geographic Origin	Previous study	Previous study assignment	New assignment
Duchilly	England	Muehlbauer et al. 2014	Mixed	English
Bulgaria XI-8	Bulgaria	Muehlbauer et al. 2014	Mixed	Black Sea
Barr's Zellernuss	England	Muehlbauer et al. 2014	Mixed	English
Ugbrooke	New England	Muehlbauer et al. 2014	Mixed	English
Gasaway	Washington	Muehlbauer et al. 2014	Gasaway	Central European
Cutleaf	England	Muehlbauer et al. 2014	Wild <i>C. avellana</i>	Central European
Aurea	France	Muehlbauer et al. 2014	Wild <i>C. avellana</i>	Central European
OSU 408.040	Minnesota	Muehlbauer et al. 2014	Wild <i>C. avellana</i>	Central European
Finland COR 187	Finland	Muehlbauer et al. 2014	Wild <i>C. avellana</i>	Central European
Henneman #3	Oregon	Muehlbauer et al. 2014	Mixed	English
Istrski Duguljasti	Croatia	Muehlbauer et al. 2014	Mixed	Black Sea
Badem	Turkey	Muehlbauer et al. 2014	Mixed	Black Sea
Daviana	England	Muehlbauer et al. 2014	Spanish-Italian	English
Ennis	Washington	Muehlbauer et al. 2014	Spanish-Italian	English
Bianca	Italy	Muehlbauer et al. 2014	Wild <i>C. avellana</i>	Spanish-Italian
Moskovskii Rubin OP #7	Russia	Muehlbauer et al. 2014	Moscow	Central European
Moskovskii Rubin OP #11	Russia	Muehlbauer et al. 2014	Moscow	Central European
Kudashovski OP #1	Russia	Muehlbauer et al. 2014	Moscow	Central European
Kudashovski OP #7	Russia	Muehlbauer et al. 2014	Moscow	Central European
Moscow #2	Russia	Muehlbauer et al. 2014	Moscow	Central European
Rote Zellernuss	Netherlands	Muehlbauer et al. 2014	Central European	English
Fusco Rubra	Germany	Muehlbauer et al. 2014	Wild	Central European

conducted a more expansive study that genotyped 270 cultivars with 21 SSRs (although 72 accessions were found to be duplicates). Their study established a large Spanish-Italian group with nine subgroups, two Black Sea groups, two English groups, and a Central European group. Gürçan et al. (2010) conducted a diversity study focused specifically on germplasm from the Black Sea region, although some Spanish-Italian cultivars were included for reference. Their study found a Caucasian group with an Azerbaijan subgroup, a large Turkish group with five subgroups, and a Spanish-Italian group.

Muehlbauer et al. (2014) genotyped 323 accessions with 17 SSR markers. The study included representatives of a wide sampling of European hazelnut diversity and accessions of other *Corylus* species, as well as several *C. americana* × *C. avellana* and *C. colurna* × *C. avellana* hybrids. It also included a very large set of European accessions originating in Russia, Crimea, and Poland. Groups formed among the European germplasm in the UPGMA dendrogram of Muehlbauer et al. (2014) included a Central European group, a Spanish-Italian group, a Black Sea group, a Moscow group, a Gasaway-related group, a mixed group, and a wild *C. avellana* group. Many of these same groups formed in their STRUCTURE analysis as well, although the Black Sea group split into two groups and the mixed and Gasaway-related groups dissolved in that analysis.

While most accessions in this study and previous studies were assigned to similar genetic groups with known geographic origins, several reassignments were also made. In general, clustering of accessions across hazelnut studies has not been very strong due to the much larger degree of genetic variation within clusters compared to variation between clusters (Muehlbauer et al. 2014). In total, 54 accessions in this study had been assigned to different groups compared to at least one of four previous studies. Some accessions had different assignments across multiple previous studies when considering the current NJ dendrogram and STRUCTURE ($K=10$) results. This resulted in a total of 72 instances in which an assignment was different between this study and one of four previous studies (Table 3). In some previous studies, accessions fell outside of the main geographic clusters in their respective dendrograms. Enough accessions from a given geographical region are needed in a study to form a clear group within a dendrogram, which may explain why previously ungrouped accessions were placed into groups in this study. This also may explain why Muehlbauer et al. (2014) resolved a Moscow group while our study resolved a Georgian group for the first time. Although the current study also included several accessions with diverse Russian origins, they did not cluster together and instead dispersed across the Central European, Black Sea, and

Georgian groups. Further, 11 accessions that were not placed in any group in Gökirmak et al. (2009) were placed in the Central European group in our study. This was an interesting result, as Gökirmak et al. (2009) did have a Central European group, although only half the size of the same group in our study. Other small sets of accessions with similar origins were once again placed together but were assigned to different groups in our study. Sets of accessions that changed groups between Gökirmak et al. (2009) and our study included four accessions from the Black Sea group moving to the English group and four accessions from the Spanish-Italian group moving to the Central European group. Sets of accessions that changed groups between Muehlbauer et al. (2014) and our study included six accessions moving from the Black Sea group to the Georgian group, four accessions moving from the mixed group to the English group, four accessions moving from the wild *C. avellana* group to the Central European group, and five accessions moving from the Moscow group to the Central European group. Because our study used a larger number of markers than previous studies to profile each accession, these updated assignments should more accurately reflect each accession's genetic relationships, which may better reflect their historical origins.

The current DAPC analysis (Fig. 6; Supp Figs. 2–4; Table 1) is a cogent, more simplified representation of the overall broad population structure/grouping definitions from many of the previous studies discussed above. This is not unexpected, as the dimensional reduction inherent in this analysis prioritizes among group/population variation over within group/population variation (Campoy et al. 2016; Dvorak et al. 2022; Miller et al. 2020; Thia 2022). This is likely the reason that admixture patterns were high in the STRUCTURE results (Figs. 1–5), while minimal or non-existent in the DAPC compoplot results (Supp Fig. 3). It is also important to note that it is not necessarily unexpected that different genetic population structure analyses often identify different optimal k -value solutions, as a “true k ” can be difficult to discern; as such various analyses may be better described as tools that can be utilized as a means to summarize data with some tools providing more efficient summaries of a dataset than other tools (Jombart and Collins 2015). From a higher-level perspective, our study shows that wide genetic diversity exists among many different sources of EFB resistance/tolerance. This indicates that these germplasm sources can be used in breeding to help increase or maintain overall genetic diversity rather than narrowing genetic diversity when breeding with a limited number of resistant accessions. Irrespective of identifying a “true k ” across different analysis methods (NJ vs. STRUCTURE vs. DAPC), a major finding of the current work is that the Georgian population(s) include several EFB resistant/tolerant

accessions and are clearly genetically divergent from the remainder of the germplasm included in this study.

The Georgian germplasm should be studied further and in multiple locations for both EFB resistance and additional traits, as this germplasm clustered together in the NJ dendrogram, STRUCTURE results, and DAPC solutions, indicating its genetic uniqueness compared to the other groups representing a wide sampling of diversity of *C. avellana*. Although the use of NGS and SNP markers did not drastically alter the population structure/groupings that were established with SSR markers, SNP markers are likely to be the marker of choice moving forward, as they are less labor-intensive per marker obtained and their increased abundance helps interrogate genomes in greater detail.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11295-024-01679-1>.

Author contribution Study conception and design were done by Thomas Molnar. Material preparation, data collection and analysis were performed by Justin Lombardoni, Jennifer Vaiciunas, Christine Kubik, and Josh Honig. The first draft of the manuscript was written by Justin Lombardoni and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This research was supported by US Department of Agriculture (USDA)–National Institute of Food and Agriculture (NIFA) Specialty Crops Research Initiative Competitive grant 2016–51181-25412. Research at Rutgers University is also supported by the USDA-NIFA Hatch project NJ12171 through the New Jersey Agricultural Experiment Station. Research at Oregon State University is also supported by state, Hatch Act, and Oregon Hazelnut Commission funds with additional support provided by a specific cooperative agreement with the USDA.

Data archiving statement Reads from Illumina sequencing of all accessions this study are held in the National Center for Biotechnology Information (Bethesda, MD) Sequence Read Archive (SRA) database [PRJNA758015 (Leinonen et al., 2010)].

Declarations

Competing Interests The authors have no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bassil N, Boccacci P, Botta R, Postman J, Mehlenbacher S (2013) Nuclear and chloroplast microsatellite markers to assess genetic diversity and evolution in hazelnut species, hybrids and cultivars. *Genet Resour and Crop Evol* 60:543–568. <https://doi.org/10.1007/s10722-012-9857-z>
- Bhattarai G, Mehlenbacher SA, Smith DC (2017) Eastern filbert blight disease resistance from *Corylus americana* ‘Rush’ and selection ‘Yoder# 5’ maps to linkage group 7. *Tree Genet and Genomes* 13:45. <https://doi.org/10.1007/s11295-017-1129-9>
- Boccacci P, Akkac A, Botta R (2006) DNA typing and genetic relations among European hazelnut (*Corylus avellana* L.) cultivars using microsatellite markers. *Genome* 49:598–611. <https://doi.org/10.1139/g06-017>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinform* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Amer J of Hum Genet* 32:314–331
- Botta R, Molnar TJ, Erdoğan V, Valentini N, Torello Marinoni D, Mehlenbacher SA (2019) Hazelnut (*Corylus* spp.) Breeding. In: Al-Khayri JM, Jain SM, Johnson DV (eds) *Advances in Plant Breeding Strategies: Nut and Beverage Crops*, vol 4. Springer, Cham, pp 157–219. https://doi.org/10.1007/978-3-030-23112-5_6
- Cai G, Leadbetter CW, Muehlbauer MF, Molnar TJ, Hillman BI (2013) Genome-wide microsatellite identification in the fungus *Anisogramma anomala* using Illumina sequencing and genome assembly. *PLoS ONE* 8(11):e82408. <https://doi.org/10.1371/journal.pone.0082408>
- Cameron HR (1976) Eastern filbert blight established in the Pacific northwest. *Plant Dis Rep* 60:737–740. <https://doi.org/10.5555/19761331602>
- Campoy JA, Lerigoleur-Balsemin E, Christmann H, Beauvieux R, Girollet N, Quero-Garcia J, Dirlwanger E, Barreneche T (2016) Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of *Prunus avium* L. landraces and bred cultivars. *BMC Plant Biol* 16:49–63. <https://doi.org/10.1186/s12870-016-0712-9>
- Capik JM, Molnar TJ (2012) Assessment of host (*Corylus* sp) resistance to eastern filbert blight in New Jersey. *J Amer Soc Hortic Sci* 137:157–172. <https://doi.org/10.21273/JASHS.137.3.157>
- Capik JM, Muehlbauer M, Novy A, Honig JA, Molnar TJ (2013) Eastern filbert blight-resistant hazelnuts from Russia, Ukraine, and Poland. *HortSci* 48:466–473. <https://doi.org/10.21273/HORTS.CI.48.4.466>
- Colburn BC, Mehlenbacher SA, Sathuvalli VR, Smith DC (2015) Eastern filbert blight resistance in hazelnut accessions ‘Culplà’, ‘Crvenje’, and OSU 495.072. *J Amer Soc Hortic Sci* 140:191–200. <https://doi.org/10.21273/JASHS.140.2.191>
- Dvorak J, Korecky J, Faltinova Z, Zadravova D (2022) Genetic diversity of sessile oak populations in the Czech Republic. *J. Forest Sci* 68:8–18. <https://doi.org/10.17221/99/2021-JFS>
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Food and Agricultural Organization of the United Nations (2021) FAOSTAT, 15 Oct 2022. <https://www.fao.org/faostat/en/#data/QCL>
- Fuller AS (1908) *The nut culturist*. Orange Judd, New York
- Gascuel O (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol and Evol* 14:685–695. <https://doi.org/10.1093/oxfordjournals.molbev.a025808>
- Gökirmak T, Mehlenbacher SA, Bassil NV (2009) Characterization of European hazelnut (*Corylus avellana*) cultivars using SSR markers. *Genet Resour and Crop Evol* 56:147–172. <https://doi.org/10.1007/s10722-008-9352-8>
- Gottwald TR, Cameron HR (1980) Infection site, infection period, and latent period of canker caused by *Anisogramma anomala* in European filbert. *Phytopathol* 70:1083–1087. <https://doi.org/10.1094/Phyto-70-1083>
- Gürçan K, Mehlenbacher SA, Erdoğan V (2010) Genetic diversity in hazelnut (*Corylus avellana* L.) cultivars from Black Sea countries assessed using SSR markers. *Plant Breed* 129:422–434. <https://doi.org/10.1111/j.1439-0523.2009.01753.x>
- Helmstetter AJ, Oztolan-Erol N, Lucas SJ, Buggs RJA (2019) Genetic diversity and domestication of hazelnut (*Corylus avellana* L.) in Turkey. *Plants People Planet* 2:326–339. <https://doi.org/10.1002/ppp3.10078>
- Honig JA, Muehlbauer MF, Capik JM, Kubik C, Vaiciunas JN, Mehlenbacher SA, Molnar TJ (2019) Identification and mapping of eastern filbert blight resistance quantitative trait loci in European hazelnut using double digestion restriction site associated DNA sequencing. *J Amer Soc Hortic Sci* 144:295–304. <https://doi.org/10.21273/JASHS04694-19>
- Johnson KB, Mehlenbacher SA, Stone JK, Pscheidt JW, Pinkerton JN (1996) Eastern filbert blight of European hazelnut—it’s becoming a manageable disease. *Plant Dis* 80:1308–1316
- Johnson KB, Pinkerton JN (2002) Eastern filbert blight. In: Teviotdale BL, Michailides TJ, Pscheidt JW (eds) *Compendium of nut crop diseases in temperate zone*. APS Press, St. Paul, pp 44–46
- Jombart T (2008) *adeigenet*: A R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jombart T, Collins C (2015) A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0.0. Available from: <https://adegenet.r-forge.rproject.org/files/tutorial-dapc.pdf>
- Leadbetter CW, Capik JM, Pisetta M, Molnar TJ (2015) Sources of resistance to eastern filbert blight in hazelnuts from the Republic of Georgia. *Scientia Hortic* 193:269–275. <https://doi.org/10.1016/j.scienta.2015.07.017>
- Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration (2010) The sequence read archive. *Nucleic Acids Res* 39:D19–D21. <https://doi.org/10.1093/nar/gkq1019>
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinform* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinform* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li Y-L, Liu J-X (2017) STRUCTURESELECTOR: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Mol Ecol Resour* 18:176–177. <https://doi.org/10.1111/1755-0998.12719>
- Lombardoni JJ, Honig JA, Vaiciunas JN, Revord RS, Molnar TJ (2022) Segregation of eastern filbert blight disease response and single nucleotide polymorphism markers in three European-American interspecific hybrid hazelnut populations. *J Amer Soc Hortic Sci* 147:196–207. <https://doi.org/10.21273/JASHS05112-22>
- Mehlenbacher SA (1991) Hazelnuts (*Corylus*). In: Moore JN, Ballington JR (eds) *Genetic resources of temperate fruit and nut crops*. Wageningen, Int Soc Hortic Sci, pp 789–836. <https://doi.org/10.17660/ActaHortic.1991.290.18>

- Mehlenbacher SA (1994) Genetic improvement of the hazelnut. *Acta Hort* 351:23–38. <https://doi.org/10.17660/ActaHortic.1994.351.1>
- Mehlenbacher SA, Brown RN, Nouhra ER, Gökirmak T, Bassil NV, Kubisiak TL (2006) A genetic linkage map for hazelnut (*Corylus avellana* L.) based on RAPD and SSR markers. *Genome* 49:122–133. <https://doi.org/10.1139/g05-091>
- Mehlenbacher SA, Heilnis BJ, Mooneyham RT, Snelling JW (2023) Breeding hazelnuts resistant to eastern filbert blight. *Acta Hort* 1362:557–562. <https://doi.org/10.17660/ActaHortic.2023.1362.75>
- Mehlenbacher SA, Molnar TJ (2021) Hazelnut (*Corylus*) Breeding. In: Goldman I (ed) *Plant Breeding Reviews*. John Wiley and Sons, Hoboken, pp 9–141. <https://doi.org/10.1002/9781119828235.ch2>
- Mehlenbacher SA, Thompson MM, Cameron HR (1991) Occurrence and inheritance of resistance to eastern filbert blight in “Gasaway” hazelnut. *HortSci* 26:410–411. <https://doi.org/10.21273/HORTSCI.26.4.410>
- Melo ATO, Bartaula R, Hale I (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 Bioinform* 17:29. <https://doi.org/10.1186/s12859-016-0879-y>
- Miller JM, Cullingham CI, Peery RM (2020) The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. *Heredity* 125:269–280. <https://doi.org/10.1038/s41437-020-0348-2>
- Molnar TJ (2011) *Corylus*. In: Kole C (ed) *Wild crop relatives: Genomic and breeding resources*. Springer, Berlin/Heidelberg, pp 15–48. https://doi.org/10.1007/978-3-642-21250-5_2
- Molnar TJ, Capik JM, Zhao S, Zhang N (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. <https://doi.org/10.1094/PDIS-06-10-0445>
- Molnar TJ, Goffreda JC, Funk CR (2005) Developing hazelnuts for the eastern United States. *Acta Hort* 686:609–618. <https://doi.org/10.17660/ActaHortic.2005.686.81>
- Molnar TJ, Goffreda JC, Funk CR (2010a) Survey of *Corylus* resistance to *Anisogramma anomala* from different geographic locations. *HortSci* 45:832–836. <https://doi.org/10.21273/HORTSCI.45.5.832>
- Molnar TJ, Lombardoni JJ, Muehlbauer MF, Honig JA, Mehlenbacher SA, Capik JM (2018) Progress breeding for resistance to eastern filbert blight in the eastern United States. *Acta Hort* 1226:79–86. <https://doi.org/10.17660/ActaHortic.2018.1226.11>
- Molnar TJ, Zurov DE, Goffreda JC, Mehlenbacher SA (2007) Survey of hazelnut germplasm from Russia and Crimea for response to eastern filbert blight. *HortSci* 42:51–56. <https://doi.org/10.21273/HORTSCI.42.1.51>
- Muehlbauer MF, Honig JA, Capik JM, Vaiciunas JN, Molnar TJ (2014) Characterization of eastern filbert blight-resistant hazelnut germplasm using microsatellite markers. *J Amer Soc Hortic Sci* 139:399–432. <https://doi.org/10.21273/JASHS.139.4.399>
- Muehlbauer MF, Tobia J, Honig JA, Zhang N, Hillman BI, Gold KM, Molnar TJ (2019) Population differentiation within *Anisogramma anomala* in North America. *Phytopathol* 109:1074–1082. <https://doi.org/10.1094/PHYTO-06-18-0209-R>
- Öztürk SC, Balık Hİ, Balık SK, Kızılcı G, Duyar Ö, Doğanlar S, Fray A (2017) Molecular genetic diversity of the Turkish national hazelnut collection and selection of a core set. *Tree Genet and Genomes* 13:113. <https://doi.org/10.1007/s11295-017-1195-z>
- Pinkerton JN, Johnson KB, Mehlenbacher SA, Pscheidt JW (1993) Susceptibility of European hazelnut clones to eastern filbert blight. *Plant Dis* 77:261–266. <https://doi.org/10.1094/PD-77-0261>
- Pinkerton JN, Johnson KB, Theiling KM, Griesbach JA (1992) Distribution and characteristics of the eastern filbert blight epidemic in western Oregon. *Plant Dis* 76:1179–1182. <https://doi.org/10.1094/PD-76-1179>
- Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7(2):e32253. <https://doi.org/10.1371/journal.pone.0032253>
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genet* 155:945–959. <https://doi.org/10.1093/genetics/155.2.945>
- R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Available from: <https://www.R-project.org/>
- Revord RS, Lovell ST, Capik JM, Mehlenbacher SA, Molnar TJ (2020) Eastern filbert blight resistance in American and interspecific hybrid hazelnuts. *J Amer Soc Hortic Sci* 145:162–173. <https://doi.org/10.21273/JASHS04732-19>
- Rowley ER, Fox SE, Bryant DW, Sullivan CM, Priest HD, Givan SA, Mehlenbacher SA, Mockler TC (2012) Assembly and characterization of the European hazelnut ‘Jefferson’ transcriptome. *Crop Sci* 52:2679–2686. <https://doi.org/10.2135/cropsci2012.02.0065>
- Rowley ER, VanBuren R, Bryant DW, Priest HD, Mehlenbacher SA, Mockler TC (2018) A draft genome and high-density genetic map of European hazelnut (*Corylus avellana* L.). *BioRxiv* 469015. <https://doi.org/10.1101/469015>
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol and Evol* 4:406–425
- Sathuvalli VR, Mehlenbacher SA, Smith DC (2011a) DNA markers linked to eastern filbert blight resistance from a hazelnut selection from the Republic of Georgia. *J Amer Soc Hortic Sci* 136:350–357. <https://doi.org/10.21273/JASHS.136.5.350>
- Sathuvalli VR, Chen H, Mehlenbacher SA, Smith DC (2011b) DNA markers linked to eastern filbert blight resistance in “Ratoli” hazelnut (*Corylus avellana* L.). *Tree Genet and Genomes* 7:337–345. <https://doi.org/10.1007/s11295-010-0335-5>
- Thia JA (2022) Guidelines for standardizing the application of discriminant analysis of principal components to genotype data. *Mol Ecol Resour* 23:523–538. <https://doi.org/10.1111/1755-0998.13706>
- Thompson MM (1979) Genetics of incompatibility in *Corylus avellana* L. *Theor and Appl Genet* 54:113–116. <https://doi.org/10.1007/BF01159464>
- Thompson MM, Lagerstedt HB, Mehlenbacher SA (1996) Hazelnuts. In: Janick J, Moore JN (eds) *Fruit breeding*, vol 3. Wiley, New York, pp 125–184
- Tobia J, Muehlbauer M, Honig J, Pscheidt J, Molnar TJ (2017) Cluster analysis of *Anisogramma anomala* isolates collected from the Pacific northwest and New Jersey. *Phytopathol* 107(S5):125

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.