Genomic approaches to accelerate American chestnut restoration

Authors: Jared W. Westbrook[1]*, Joanna Malukiewicz[2,3], Qian Zhang[2], Avinash Sreedasyam[4,5], Jerry W. Jenkins[4], Vasiliy Lakoba[1], Sara Fitzsimmons[1], Jamie Van Clief[1], Kendra Collins[1], Stephen Hoy[1], Cassie Stark[1], Lake Graboski[1], Eric Jenkins[1], Thomas M. Saielli[1], Benjamin T. Jarrett[1], Lucinda J. Wigfield[1], Lauren M. Kerwien[1], Ciera Wilbur[1], Alexander M. Sandercock[2], J. Hill Craddock[1,6], Susanna Keriö[7], Tetyana Zhebentyayeva[8], Shenghua Fan[8], Austin M. Thomas[8], Albert G. Abbott[8], C. Dana Nelson[8,9], Xiaoxia Xia[10], James R. McKenna[11], Caleb Kell[12], Melissa Williams[4], LoriBeth Boston[4], Christopher Plott[4], Florian Carle[1,13], Jack Swatt[1], Jack Ostroff[1], Steven N. Jeffers[10], Kathleen McKeever[9], Erica Smith[9], Thomas J. Ellis[14], Joseph B. James[1], Paul Sisco[1], Andrew Newhouse[15], Erik Carlson[15], William A. Powell[15], Frederick V. Hebard[1], John Scrivani[1,16], Caragh Heverly[1], Martin Cipollini[1,17], Brian Clark[1], Eric Evans[1], Bruce Levine[1,18], John E. Carlson[19], David Goodstein[5], Jack Orebaugh[20], Zamin K. Yang[20], Madhavi Z. Martin[20], Joanna Tannous[20], Tomás A. Rush[20], Nancy L. Engle[20], Timothy J. Tschaplinski[20], Jane Grimwood[4], Jeremy Schmutz[4,5], Jason A. Holliday[2], & John T. Lovell[4,5,21]*

Affiliations:

[1] The American Chestnut Foundation, Asheville, NC 28804, USA
[2] Virginia Tech, Department of Forest Resources and Environmental Conservation, Blacksburg, VA 24061, USA
[3] Research Unit for Evolutionary Immunogenomics, Department of Biology, University of Hamburg, 20146 Hamburg, Germany
[4] Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA
[5] US Department of Energy Joint Genome Institute, Berkeley, CA 94720, USA
[6] University of Tennessee at Chattanooga, Chattanooga, TN 37403 USA
[7] Connecticut Agricultural Experiment Station, Department of Environmental Science and Forestry, New Haven, CT 06511 USA
[8] USDA Forest Service, Southern Research Station, Forest Health Research & Education Center, Lexington, KY 40546
[9] US Forest Service Southern Research Station, Asheville, NC 28806 USA
[10] Clemson University, Clemson, SC 29634 USA
[11] A Breed Above Timber, West Lafayette, IN 47907
[12] Purdue University Hardwood Tree Improvement and Regeneration Center, West Lafayette, IN 47907
[13] Yale University, New Haven, CT 06511 USA
[14] Gregor Mendel Institute of Molecular Plant Sciences, Doktor-Bohr-Gasse 3 1030 Vienna, Austria
[15] State University of New York College of Environmental Science and Forestry, Department of Environmental Biology, Syracuse, NY 13210 USA
[16] Virginia Department of Forestry, Charlottesville, VA 22903 USA
[17] Berry College, Mt Berry, GA 30149 USA
[18] University of Maryland, College Park, MD 20742 USA
[19] Pennsylvania State University, The Schatz Center for Tree Molecular Genetics, University Park, PA 16802
[20] Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831
[21] The Center for Bioenergy Innovation (CBI)
*To whom correspondence should be addressed: jared.westbrook@tacf.org, jlovell@hudsonalpha.org

Abstract: Over a century after two introduced pathogens killed billions of American chestnuts, introgression of resistance alleles from Chinese chestnuts has contributed to the recovery of self-sustaining populations. However, progress has been slow because of the complex genetic architecture of resistance. To better understand blight resistance, we compared reference genomes, gene expression responses, and stem metabolite profiles of the resistant Chinese and susceptible American chestnut species. To accelerate resistance breeding, we conducted large-scale phenotyping and genotyping in hybrids of these species. Simulation and inoculation experiments suggest that significant resistance gains are possible through selectively breeding trees with an average of 70% to 85% American chestnut ancestry. The resources developed here are foundational for breeding to create diverse restoration populations with sufficient disease resistance and competitive growth.

Main text: The demise of the American chestnut (*Castanea dentata*) is a quintessential example of population collapse following the introduction of non-native pathogens (*1*). The necrotrophic fungus that causes chestnut blight (*Cryphonectria parasitica*) was introduced to North America on Asian chestnuts in the late 1800s and spread

quickly by air-borne spores (*2*, *3*). By 1950, chestnut blight killed billions of *C. dentata* trees from Maine to Mississippi (Fig. 1A), eliminating this once-dominant tree from the forest canopy (*4*). The impact of chestnut blight was compounded by the earlier introduction of an invasive oomycete (*Phytophthora cinnamomi*) that causes *Phytophthora* root rot (hereon 'root rot') (*5*). While millions of *C. dentata* persist as root collar sprouts (*6*), they rarely reproduce before succumbing to the blight fungus (Fig. 1B), hence the species is considered functionally extinct (*7*).

There has been over a century of effort to generate *C. dentata* populations with improved resistance to chestnut blight and root rot (*8*). Reforestation with disease-resistant *C. dentata* has many potential benefits, including the return of a regular mast for wildlife and human consumption (*9*), restoration of strip-mined land (*10*), and carbon sequestration (*11*). The blight fungus is ubiquitous throughout *C. dentata*'s former range, while root rot has a patchy distribution (*12*). Therefore, improving blight resistance and, secondarily, root rot resistance are essential restoration priorities. Restoration populations must also have the canopy competitiveness of pre-blight populations (*13*) and sufficient genetic diversity to adapt to a wide geographic range and changing climates (*14*).

Integration of genomic data to evaluate efforts to develop blight-resistant *C. dentata* populations: Three strategies have been employed to improve blight resistance within *C. dentata* (Fig. 1): (1) hybridization with resistant Chinese chestnuts (*C. mollissima*) and subsequent backcrossing to diverse *C. dentata* trees (*8*), (2) intercrossing rare, large surviving American chestnuts (hereon "LSACs"), which are blight-infected trees with stem diameters exceeding 25 cm at breast height (*15*), and (3) constitutive expression of an *oxalate oxidase* (OxO) transgene from wheat, which degrades oxalic acid, a virulence factor secreted by *Cr. parastica* (*16*). We assessed these strategies by inoculating all trees with virulent strains of blight fungus and evaluating the resistance of (1) 3,296 hybrids, 96 susceptible *C. dentata*, and 69 resistant *C. mollissima* controls, (2) 30 first- and 122 second-generation progeny of 14 wild LSACs, and (3) 259 transgenic OxO+ and 217 wild type (WT) full sibling controls. Long-term blight resistance of hybrids and LSACs (average age 14 years) was quantified using a single index trait that integrated seven canker size and morphology resistance traits. The 'blight resistance index' was scaled from 0 (mean of susceptible *C. dentata*) to 100 (mean of resistant *C. mollissima*) after accounting for the effects of tree age, soil, and climate on blight severity. Trees were genotyped at 93,332 single nucleotide polymorphisms (SNPs) to estimate species ancestry and additive genetic variation ('breeding values') in blight resistance.

Within hybrid populations, there was significant heritable variation in blight resistance ($h^2 = 0.41 \pm 0.03$; LRT $P < 0.0001$; Table S1), with breeding values ranging from <0 (susceptible) to >100 (fully resistant). Blight resistance was positively genetically correlated with height growth ($r_g = 0.57 \pm 0.05$, LRT $P < 0.0001$), indicating that resistance is needed for forest competitiveness. However, at a site where blight severity has been limited by hypoviruses that reduce *Cr. parastica* virulence (*17*), *C. dentata* progeny grew tallest, *C. mollissima* shortest, and backcross hybrids with ~75% *C. dentata* ancestry showed intermediate growth (Fig. S1). Therefore, to generate competitive hybrid chestnut populations, selective breeding must enhance resistance and height growth, while maintaining a majority of *C. dentata* ancestry.

Blight resistance was negatively genetically correlated with *C. dentata* ancestry ($r_g = –0.65$, $P < 0.0001$, Fig. 1E) (*18*), indicating that introgression of *C. mollissima* ancestry generally improves blight resistance. To estimate the minimum *C. mollissima* ancestry to improve blight resistance substantially (>50), we inoculated seedling progeny from 30 intercrosses between *C. dentata* backcross hybrids from seven *C. mollissima* resistance sources with varying hybrid ancestry. Four *C. dentata* backcross families had above-average blight resistance ratings (56-85) and majority (64% to 68%) *C. dentata* ancestry (Figure S2; Table S2). Furthermore, among 3,132 hybrids with >70% *C. dentata* ancestry, there was significant heritable variation in blight resistance ($h^2 = 0.38 \pm 0.03$) and tree height ($h^2 = 0.62 \pm 0.06$), implying that selection among majority-*C. dentata* ancestry trees can improve blight resistance and canopy competitiveness.

To evaluate whether LSACs have heritable blight resistance, we compared the blight resistance of 30 first-generation LSAC progeny with that of 36 susceptible *C. dentata* controls planted in an orchard in Virginia, USA. On average, the LSAC progeny had modestly higher resistance breeding values (mean = 8.5, n = 30) than those from non-LSAC parents (mean = -7.5, $n = 36$; $t_{34.7} = 3.24$, $P = 0.0026$); however, four LSAC progeny from this site had breeding values exceeding 50. To confirm heritable resistance, we inoculated 122 second-generation seedling progeny from one controlled ($n = 37$) and six open-pollinated LSAC crosses ($n = 6$–26 per family) from an isolated orchard in Maryland. The trees in this orchard had low to intermediate blight resistance breeding values (0–54, mean 27, $n = 24$). In seedling inoculation trials, four of seven second-generation families also had modestly enhanced average

resistance (32–40), and one family had high resistance (98) (Fig. S3; Table S2). Among 143 trees with 100% *C. dentata* ancestry evaluated for long-term resistance, only seven had resistance breeding values exceeding 50. Our results imply that selective breeding could enhance blight resistance within *C. dentata*, but additional wild resistance sources are needed to maintain a diverse breeding population.

To evaluate the efficacy of the oxalate oxidase transgenic approach, we evaluated blight resistance and growth of three-year-old progeny from the 'Darling 54' *C. dentata* founder, which is currently under federal review (*19*). One year after inoculation, OxO+ progeny (n = 259) had 38% shorter cankers than wild-type ('WT' n = 217) full siblings (Fig. 1E; $F_{1,472} = 108$, $P < 0.0001$). However, canker severity ratings among OxO+ progeny varied widely (Fig. 1B), and cankers on 13% of trees with the highest year one resistance ratings continued to expand in subsequent years (Fig. S4-S5). Furthermore, OxO+ trees grew 22% slower than WT siblings ($F_{1,472} = 113$, $P < 0.0001$; Fig. S4), which may be a pleiotropic effect of constitutive OxO activity (*20*, *21*). It is also noteworthy that in 'Darling 54', the OxO construct is inserted into SAL1 (Caden.04G06260), a single-copy regulator of growth, development, and stress responses in Arabidopsis (*22*, *23*). Consistent with deleterious effects of constitutive OxO activity or SAL1 disruption, homozygous OxO+ Darling 54 progeny were recovered less frequently (8/99, 8.1% observed homozygosity (*24*)) than the expected rate of 25% (binomial P = 0.00002). Despite these limitations, 4% of the hemizygous OxO+ trees had both high resistance ratings and height growth comparable (± 1 standard deviation) to WT controls. Larger field trials are underway to determine if the resistance and growth of 'Darling 54' progeny are sufficient for forest restoration.

Despite the challenges of hybrid, LSAC, and OxO breeding, there remains significant potential to develop disease-resistant, competitive, and diverse populations. Additional generations of selection for resistance and growth in hybrid and LSAC populations may yield substantial gains in these traits. Exploration of additional biotechnology strategies, including multi-gene editing in diverse *C. dentata* backgrounds, has the potential to augment LSAC and hybrid breeding efforts.

Genomic foundations for American chestnut restoration: A lack of genomic resources has historically limited molecular breeding and biotechnology to improve disease resistance in *C. dentata*. Therefore, we developed chromosome-scale *Castanea* genomes for three important founder trees. Our *C. dentata* genome is from 'Ellis-1', the wild-type tree transformed with the OxO construct to create 'Darling 54' (*19*). We also generated *C. mollissima* genomes from two founder trees ('Mahogany' and 'Nanking') that contributed resistance to The American Chestnut Foundation's (TACF) hybrid breeding program (*8*). To construct the genomes, we assembled 112-125× PacBio sequencing, scaffolded the contigs into chromosomes with 76-79× HiC, and polished the assemblies with 50-100× Illumina reads. Differences in sequencing technology necessitated a single haplotype build for 'Ellis-1', whereas we generated haplotype-resolved assemblies for 'Mahogany' and 'Nanking'. The five assemblies are accurate (QV range = Q48-Q58), complete (BUSCO = 97.7-98.3%), and contiguous (contig N_{50} = 12-53Mb, Fig. S6), indicating that these are the highest quality *Castanea* genomes available (*25–30*).

We annotated 25.2k-31.2k protein-coding genes per assembly based on homology, ab initio evidence, and RNA-seq from multiple tissues, including stems infected with chestnut blight. While the vast majority of genes were in families ('orthogroups') that spanned all five genomes (120,244 genes, 90.7%), there was substantial copy number variation ('CNV') between species. We hypothesized that *C. mollissima* blight resistance is, in part, explained by CNV expansion within defense-related gene families, enabling a stronger expression response to pathogen challenge (*31*, *32*). To compare resistant and susceptible gene expression responses, we performed RNA-sequencing on blight-inoculated *C. mollissima* and *C. dentata* stems relative to wound-only controls 3 and 10 days after treatment ($n = 6$ per treatment). Of 1,401 orthogroups that were expanded across all four *C. mollissima* genomes relative to *C. dentata*, 128 had significantly greater expression upregulation in inoculated *C. mollissima* (*t*-test FDR-adjusted $P < 0.05$). These orthogroups contained genes related to multiple stages of the blight resistance response, including lipid signaling (*e.g.,* phospholipase A2), secondary metabolism (e.g., phenazine biosynthesis), and cell wall strengthening (*e.g.,* hydroxyproline-rich glycoproteins), which supports the hypothesis that copy number expansion has contributed to the evolution of blight resistance in *Castanea* (Table S3).

Contrasting gene expression and stem metabolomes in resistant and susceptible hosts: Evolutionary changes in single-copy gene expression responses to the blight fungus may also contribute to interspecific resistance variation (*33*). Across our RNA-seq experiment 6,033 (33.6%) of the 17,959 single-copy orthologs expressed in stems were differentially expressed between control and fungal inoculation treatments. A majority of blight-responsive genes (n = 4,073, 67%) were only differentially expressed in one species, indicating strongly divergent interspecific expression responses (Fig. S7). The majority of species-specific responses (3,025 of 4,073 genes) were found only

in the susceptible *C. dentata* host and were enriched in gene ontology (GO) terms indicative of stress responses, including those triggered by low oxygen during pathogen growth (*34)* (Fig. S8). Of the 548 genes uniquely upregulated in the resistant *C. mollissima* host, overrepresented pathways included those involved in lipid signaling, secondary metabolism, and cell wall strengthening (Fisher's Exact Test, FDR-adjusted $P < 0.05$; Figs. S9–S10), underscoring the importance of these processes in induced resistance.

We hypothesized that the enhanced blight resistance of *C. mollissima* may also be explained by constitutively elevated metabolites that inhibit the initial growth of the blight fungus (*35*). We quantified 99 metabolites constitutively present in the stems of *C. mollissima* and *C. dentata*, and found that *C. mollissima* had elevated concentrations of 41 metabolites (*t*-test FDR-adjusted $P < 0.05$; Table S4). The metabolites with the most significant fold increases in *C. mollissima* included three aromatic compounds (227×homogentisic acid-2-*O*-glucoside, 21×homogentisic acid, 9×β-phenyllactic acid) and three triterpene sterols (*i.e.,* 22×lupeol, 22×betulin, 19×erythrodiol). We tested if these metabolites inhibited the growth of the blight fungus *in vitro* and found that lupeol completely suppressed growth, while erythrodiol partially inhibited growth, and β-phenyllactic acid stimulated growth (Fig. S11; Table S5). Therefore, gene-editing or hybrid breeding strategies to enhance triterpene sterol biosynthesis, particularly lupeol production, could be a promising strategy to improve blight resistance in *C. dentata*.

The genetic architecture of blight resistance: Characterizing the number and effect sizes of loci underlying resistance is crucial when designing optimal breeding strategies. Backcross breeding is most effective for traits governed by a few large-effect loci, while multi-generational intercrossing and selection *(i.e.*, recurrent selection) incrementally improve quantitative traits governed by many small-effect loci (*36*). To determine whether large-effect loci underlie blight resistance*,* we conducted quantitative trait locus (QTL) mapping by associating resistance variation with *C. mollissima* and *C. dentata* ancestry blocks (Fig. S12)*.* Combined, we detected four loci in the two largest backcross populations ($n_{\text{Clapper}} = 1{,}324$, $n_{\text{Graves}} = 1{,}275$; Fig. 2). At three QTL, inheriting a *C. mollissima* allele was estimated to modestly increase blight resistance by 9 to 11 units, while *C. mollisima* ancestry was estimated to decrease blight resistance by 11 units at the fourth locus (Fig. 2). These results imply resistance or susceptibility loci segregate within both *C. mollissima* and *C*. *dentata* populations.

To increase power to detect blight resistance-associated loci, we conducted a genome-wide association study (GWAS) across 3,365 trees with >40% *C. dentata* ancestry. These trees descended from 28 *C. mollissima* resistance sources and 406 wild *C. dentata* parents (Fig. 1A) and represent the population that could be intercrossed to generate progeny exceeding our breeding target of >70% *C. dentata* ancestry. After controlling for population structure (Fig. S13), 17 significantly-associated SNPs (FDR-adjusted $P < 0.05$; Figs. 2 & S14) explained 10% of the total phenotypic and ~25% of the heritable variation. With this complex genetic architecture, recurrent selection is needed to improve blight resistance incrementally.

Root rot resistance in *C. dentata* hybrids: Despite being restricted to poorly drained soils and warmer climates of the southeastern U.S. (*37*, *38*), the root rot pathogen (*P. cinnamomi*) is predicted to spread north with warming winter temperatures (*39*). Unlike blight, no resistance to root rot has been identified in *C. dentata* populations; hence, hybridization with *C. mollissima* is the primary source of resistance. Here, we genotyped 484 hybrid mothers whose root-rot resistance was estimated from progeny inoculations (*40*). Progeny survival was significantly heritable ($h^2 = 0.62 \pm 0.08$; LRT $P < 0.0001$), and 36 mothers with > 70% *C. dentata* ancestry had root rot survival breeding values > 50 on a scale of 0 (*C. dentata* mean) to 100 (*C. mollissima* mean) (Fig. S15). These results show that hybrid populations have inherited substantial resistance to root rot despite selection only for blight resistance.

Root rot resistance is at least partially controlled by a large-effect locus on chromosome five. Among backcross descendants of the 'Graves' founder ($n = 395$), inheriting a *C. mollissima* allele was estimated to improve survival by 16% (Fig. 2). This QTL colocalized with loci detected in a previous study (*41*), demonstrating its replicable contribution (Fig. S16). Additionally, GWAS detected three QTL that explained 21% of the phenotypic and 34% of the heritable variation in survival across 475 backcross families descending from nine *C. mollissima* sources (Figs. 2 & S14). With multi-locus resistance, recurrent selection should be an effective strategy to enhance resistance to both blight and root rot.

Genomic selection for enhanced resistance and majority *C. dentata* ancestry: To accelerate genetic gains, we implemented genomic selection for disease resistance, height growth, and majority *C. dentata* ancestry through GBLUP, which estimates breeding values from genome-wide markers linked to multiple QTLs (*42*, *43*). To assess genomic prediction accuracy, we predicted breeding values for five random subsets (20%) of phenotyped trees with

>70% *C. dentata* ancestry. Relative to pedigree-based prediction, genomic prediction with 25k SNPs was 1.20×, 1.16×, and 1.07× more accurate for blight resistance ($r_g = 0.61 \pm 0.05$), root rot survival ($r_g = 0.62 \pm 0.08$), and height growth ($r_g = 0.57 \pm 0.07$), respectively (Table S6). Downsampling to 6k SNPs yielded <2% accuracy reduction, suggesting medium-density genotyping can lower genotyping costs. While genomic selection was reasonably accurate, breeding values estimated from phenotypes were more accurate (*e.g.*, 0.83–0.97 for blight resistance). To maximize genetic gain, we will use genomic prediction to initially select seedlings for planting in open-pollinated seed orchards. We will subsequently select parents for the next breeding cycle based on resistance and growth phenotyping in replicated field trials (Fig. 3A).

To select parents for the next breeding cycle, we blended pedigree and genomic relationships ('single-step GBLUP') to estimate breeding values for blight resistance in an expanded population of 4,219 inoculated trees, including 703 trees not yet genotyped. We also estimated root rot resistance breeding values for 1,307 trees, including 178 progeny-tested trees not yet genotyped, and 645 genotyped trees yet to be progeny tested (Fig. S15). We then predicted genetic gains from breeding parents with above-average blight resistance (>70) and height growth (>70th percentile) by simulating progeny genotypes from these crosses. Across three simulation replicates composed of 30 crosses each, genomic selection of the top 20% of progeny is expected to increase blight resistance from the current population average of 22 to 65-68, while increasing tree height 1.15-1.17× (Fig. S17). Similarly, genomic selection among progeny of parents with above-average blight resistance (>60), root rot resistance (>40), and height growth (>70th percentile) is predicted to improve population mean blight resistance to 60-65, root rot resistance to 46-49, and height growth 1.20-1.21× (Fig. 3B).

To build future potential for local climate adaptation, we are performing controlled crosses among hybrids derived from three genetically and geographically distinct wild *C. dentata* subpopulations (Fig. 1A). We aim to include 20-50 wild-type *C. dentata* parents within each population to capture >95% of the climate-adaptive allelic variation (*14*). To incorporate additional resistance alleles, we are crossing the most resistant advanced backcross (*e.g.,* BC_3F_2) selections with F_1 hybrids derived from >20 *C. mollissima* sources to generate hybrids with ~70% *C. dentata* ancestry. Large quantities of open-pollinated seeds will likely be available for restoration trials within 7 to 15 years (*44*).

Conclusions: Recurrent selection in hybrid populations remains a key approach to enhance disease resistance and forest competitiveness in genetically diverse *C. dentata* populations. Complementary selection for blight resistance within *C. dentata* would benefit from incorporating additional wild genotypes. Constitutive OxO transgene expression improves blight resistance, although further research is needed to confirm if resistance is sufficient and if selection or additional transformations can reduce growth penalties. Metabolites in Chinese chestnut bark that inhibit *Cr. parasitica* provide further genetic engineering opportunities. As invasive pathogens threaten forests worldwide, it is vital to conserve and improve genetic resistance in surviving populations.

References and Notes:

1. S. L. Anagnostakis, Chestnut blight: The classical problem of an introduced pathogen. *Mycologia* **79**, 23–37 (1987).

2. C. L. Shear, N. E. Stevens, The chestnut-blight parasite (*Endothia parasitica)* from China. *Science* **38**, 295–297 (1913).

3. F. V. Hebard, Developmental histopathology of cankers incited by hypovirulent and virulent isolates of *Endothia parasitica* on susceptible and resistant chestnut trees. *Phytopathology* **74**, 140 (1984).

4. D. E. Davis, *The American Chestnut: An Environmental History* (University of Georgia Press, 2021).

5. B. S. Crandall, G. F. Gravatt, M. Ryan, Root disease of *Castanea* species and some coniferous and broadleaf nursery stocks, caused by *Phytophthora cinnamomi*. *Phytopathology* **35**, 162–180 (1945).

6. H. J. Dalgleish, C. D. Nelson, J. A. Scrivani, D. F. Jacobs, Consequences of shifts in abundance and distribution of American chestnut for restoration of a foundation forest tree. *For. Trees Livelihoods* **7**, 4 (2015).

7. K. M. Potter, B. S. Crane, W. W. Hargrove, A United States national prioritization framework for tree species vulnerability to climate change. *New Forests* **48**, 275–300 (2017).

8. K. C. Steiner, J. W. Westbrook, F. V. Hebard, L. L. Georgi, W. A. Powell, S. F. Fitzsimmons, Rescue of American chestnut with extraspecific genes following its destruction by a naturalized pathogen. *New Forests* **48**, 317–336 (2017).

9. H. J. Dalgleish, R. K. Swihart, American chestnut past and future: Implications of restoration for resource pulses and consumer populations of eastern U.S. forests. *Restor. Ecol.*, 20,490–497 (2012)

10. J. G. Skousen, K. Dallaire, S. Scagline-Mellor, A. Monteleone, L. Wilson-Kokes, J. Joyce, C. Thomas, T. Keene, C. DeLong, T. Cook, D. F. Jacobs, Plantation performance of chestnut hybrids and progenitors on reclaimed Appalachian surface mines. *New Forests* **49**, 599–611 (2018).

11. E. J. Gustafson, B. R. Sturtevant, A. M. G. de Bruijn, N. Lichti, D. F. Jacobs, D. M. Kashian, B. R. Miranda, P. A. Townsend, Forecasting effects of tree species reintroduction strategies on carbon stocks in a future without historical analog. *Glob. Chang. Biol.* **24**, 5500–5517 (2018).

12. Y. Balci, S. Balci, J. Eggers, W. L. MacDonald, J. Juzwik, R. P. Long, K. W. Gottschalk, *Phytophthora* spp. associated with forest soils in eastern and north-central U.S. oak ecosystems. *Plant Dis.* **91**, 705–710 (2007).

13. S. L. Clark, S. E. Schlarbaum, A. M. Saxton, S. N. Jeffers, R. E. Baird, Eight-year field performance of backcross American chestnut (*Castanea dentata*) seedlings planted in the southern Appalachians, USA. *For. Ecol. Manage.* **532**, 120820 (2023).

14. A. M. Sandercock, J. W. Westbrook, Q. Zhang, J. A. Holliday, A genome-guided strategy for climate resilience in American chestnut restoration populations. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2403505121 (2024).

15. G. J. Griffin, F. V. Hebard, R. W. Wendt, J. R. Elkins, Survival of American chestnut trees: Evaluation of blight resistance and virulence in *Endothia parasitica*. *Phytopathology* **73**, 1084–1092 (1983).

16. W. A. Powell, A. E. Newhouse, V. Coffey, Developing blight-tolerant American chestnut trees. *Cold Spring Harb. Perspect. Biol.* **11** (2019).

17. J. W. Westbrook, Q. Zhang, M. K. Mandal, E. V. Jenkins, L. E. Barth, J. W. Jenkins, J. Grimwood, J. Schmutz, J. A. Holliday, Optimizing genomic selection for blight resistance in American chestnut backcross populations: A trade-off with American chestnut ancestry implies resistance is polygenic. *Evol. Appl.* **13**, 31–47 (2020).

18. G. J. Griffin, J. R. Elkins, D. McCurdy, S. L. Griffin, Integrated use of resistance, hypovirulence, and forest management to control blight on American chestnut. *Restoration of American chestnut to forest lands*, 97–107 (2006).

19. A. Newhouse, "Petition for determination of nonregulated status for blight-tolerant Darling 54 American chestnut" (Animal and Plant Health Inspection Service, 2020); https://www.regulations.gov/document/APHIS-2020-0030-17585.

20. X. Hu, D. L. Bidney, N. Yalpani, J. P. Duvick, O. Crasta, O. Folkerts, G. Lu, Overexpression of a gene encoding hydrogen peroxide-generating oxalate oxidase evokes defense responses in sunflower. *Plant Physiol.* **133**, 170–181 (2003).

21. N. Smirnoff, D. Arnaud, Hydrogen peroxide metabolism and functions in plants. *New Phytologist* **221**, 1197–1214 (2019).

22. S. Y. Phua, D. Yan, K. X. Chan, G. M. Estavillo, E. Nambara, B. J. Pogson, The SAL1-PAP pathway: A case study for integrating chloroplast retrograde, light and hormonal signaling in modulating plant growth and development? *Front Plant Sci* **9**, 1171 (2018).

23. G. M. Estavillo, P. A. Crisp, W. Pornsiriwong, M. Wirtz, D. Collinge, C. Carrie, E. Giraud, J. Whelan, P. David, H. Javot, C. Brearley, R. Hell, E. Marin, B. J. Pogson, Evidence for a SAL1-PAP chloroplast retrograde pathway that functions in drought and high light signaling in Arabidopsis. *Plant Cell* **23**, 3992–4012 (2011).

24. T. Klak, H. Pilkey, V. G. May, D. Matthews, A. D. Oakes, E. H. Tan, A. E. Newhouse, Speed breeding transgenic American chestnut trees toward restoration. *bioRxiv*, 2025.05.19.654928 [Preprint] (2025). doi: https://doi.org/10.1101/2025.05.19.654928

25. L. Bianco, P. Fontana, A. Marchesini, S. Torre, M. Moser, S. Piazza, S. Alessandri, V. Pavese, P. Pollegioni, C. Vernesi, M. Malnoy, D. Torello Marinoni, S. Murolo, L. Dondini, C. Mattioni, R. Botta, F. Sebastiani, D. Micheletti, L. Palmieri, The de novo, chromosome-level genome assembly of the sweet chestnut (*Castanea sativa* Mill.) Cv. Marrone Di Chiusa Pesio. *BMC Genom Data* **25**, 64 (2024).

26. K. Shirasawa, S. Nishio, S. Terakami, R. Botta, D. T. Marinoni, S. Isobe, Chromosome-level genome assembly of Japanese chestnut (*Castanea crenata* Sieb. et Zucc.) reveals conserved chromosomal segments in woody rosids. *DNA Res.* **28** (2021).

27. Y. Xing, Y. Liu, Q. Zhang, X. Nie, Y. Sun, Z. Zhang, H. Li, K. Fang, G. Wang, H. Huang, T. Bisseling, Q. Cao, L. Qin, Hybrid de novo genome assembly of Chinese chestnut (*Castanea mollissima*). *Gigascience* **8** (2019).

28. J. Wang, P. Hong, Q. Qiao, D. Zhu, L. Zhang, K. Lin, S. Sun, S. Jiang, B. Shen, S. Zhang, Q. Liu, Chromosome-level genome assembly provides new insights into Japanese chestnut (*Castanea crenata*) genomes. *Front. Plant Sci.* **13**, 1049253 (2022).

29. Y. Sun, Z. Lu, X. Zhu, H. Ma, Genomic basis of homoploid hybrid speciation within chestnut trees. *Nat. Commun.* **11**, 3375 (2020).

30. M. Staton, C. Addo-Quaye, N. Cannon, J. Yu, T. Zhebentyayeva, M. Huff, N. Islam-Faridi, S. Fan, L. L. Georgi, C. D. Nelson, E. Bellis, S. Fitzsimmons, N. Henry, D. Drautz-Moses, R. E. Noorai, S. Ficklin, C. Saski, M. Mandal, T. K. Wagner, N. Zembower, C. Bodénès, J. Holliday, J. Westbrook, J. Lasky, F. V. Hebard, S. C. Schuster, A. G. Abbott, J. E. Carlson, A reference genome assembly and adaptive trait analysis of *Castanea mollissima* "Vanuxem," a source of resistance to chestnut blight in restoration breeding. *Tree Genet. Genomes* **16**, 57 (2020).

31. P. Spealman, C. de Santana, T. De, D. Gresham, Multilevel gene expression changes in lineages containing adaptive copy number variants. *Mol Biol Evol* **42**, msaf005 (2025).

32. V. Suryawanshi, I. N. Talke, M. Weber, R. Eils, B. Brors, S. Clemens, U. Krämer, Between-species differences in gene copy number are enriched among functions critical for adaptive evolution in *Arabidopsis halleri*. *BMC Genomics* **17**, 1034 (2016).

33. R. Assis, Lineage-specific expression divergence in grasses is associated with male reproduction, host-pathogen defense, and domestication. *Genome Biology and Evolution* **11**, 207 (2018).

34. H. Chung, Y.-H. Lee, Hypoxia: A double-edged sword during fungal pathogenesis? *Front. Microbiol.* **11**, 551990 (2020).

35. N. M. Westrick, D. L. Smith, M. Kabbage, Disarming the host: Detoxification of plant defense compounds during fungal necrotrophy. *Front. Plant Sci.* **12**, 651716 (2021).

36. J. E. Rutkoski, "A practical guide to genetic gain" in *Advances in Agronomy* (Elsevier, 2019), pp. 217–249.

37. C. C. Rhoades, S. L. Brosi, A. J. Dattilo, P. Vincelli, Effect of soil compaction and moisture on incidence of Phytophthora root rot on American chestnut (*Castanea dentata*) seedlings. *For. Ecol. Manage.* **184**, 47–54 (2003).

38. E. J. Gustafson, B. R. Miranda, T. J. Dreaden, C. C. Pinchot, D. F. Jacobs, Beyond blight: Phytophthora root rot under climate change limits populations of reintroduced American chestnut. *Ecosphere* **13** (2022).

39. T. I. Burgess, J. K. Scott, K. L. Mcdougall, M. J. C. Stukely, C. Crane, W. A. Dunstan, F. Brigg, V. Andjic, D. White, T. Rudman, F. Arentz, N. Ota, G. E. S. J. Hardy, Current and projected global distribution of

Phytophthora cinnamomi, one of the world's worst plant pathogens. *Glob. Chang. Biol.* **23**, 1661–1674 (2017).

40. J. W. Westbrook, J. B. James, P. H. Sisco, J. Frampton, S. Lucas, S. N. Jeffers, Resistance to *Phytophthora cinnamomi* in American chestnut (*Castanea dentata*) backcross populations that descended from two Chinese chestnut (*Castanea mollissima*) sources of resistance. *Plant Dis.* **103**, 1631–1641 (2019).

41. T. N. Zhebentyayeva, P. H. Sisco, L. L. Georgi, S. N. Jeffers, M. T. Perkins, J. B. James, F. V. Hebard, C. Saski, C. D. Nelson, A. G. Abbott, Dissecting resistance to *Phytophthora cinnamomi* in interspecific hybrid chestnut crosses using sequence-based genotyping and QTL mapping. *Phytopathology* **109**, 1594–1604 (2019).

42. T. H. Meuwissen, B. J. Hayes, M. E. Goddard, Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001).

43. D. Habier, R. L. Fernando, D. J. Garrick, Genomic BLUP decoded: A look into the black box of genomic prediction. *Genetics* **194**, 597–607 (2013).

44. F. L. Paillet, P. A. Rutter, Replacement of native oak and hickory tree species by the introduced American chestnut (*Castanea dentata*) in southwestern Wisconsin. *Can. J. Bot.* **67**, 3457–3469 (1989).

45. S. L. Anagnostakis, Chestnut breeding in the United States for disease and insect resistance. *Plant Dis.* **96**, 1392–1403 (2012).

46. F. V. Hebard, "The Backcross Breeding Program of The American Chestnut Foundation" in *Proceedings of Conference and Workshop*, J. E. Carlson, K. C. Steiner, Eds. (Asheville, NC, 2004).

47. D. E. Beck, "Yellow poplar site index curves" (Res. Note SE-180. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southeastern Forest Experiment Station); https://research.fs.usda.gov/treesearch/5148

48. J. Catchen, P. A. Hohenlohe, S. Bassham, A. Amores, W. A. Cresko, Stacks: An analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).

49. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

50. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021).

51. G. A. Van der Auwera, B. D. O'Connor, *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (O'Reilly Media, Inc., 2020).

52. R. Poplin, V. Ruano-Rubio, M. A. DePristo, T. J. Fennell, M. O. Carneiro, G. A. Van der Auwera, D. E. Kling, L. D. Gauthier, A. Levy-Moonshine, D. Roazen, K. Shakir, J. Thibault, S. Chandran, C. Whelan, M. Lek, S. Gabriel, M. J. Daly, B. Neale, D. G. MacArthur, E. Banks, Scaling accurate genetic variant discovery to tens of thousands of samples, *bioRxiv,* 201178 (2018). doi: https://doi.org/10.1101/201178

53. B. L. Browning, X. Tian, Y. Zhou, S. R. Browning, Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* **108**, 1880–1890 (2021).

54. B. L. Browning, Y. Zhou, S. R. Browning, A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018).

55. R. Corbett-Detig, R. Nielsen, A hidden markov model approach for simultaneously estimating local ancestry and admixture time using next generation sequence data in samples of arbitrary ploidy. *PLoS Genet.* **13**, e1006529 (2017).

56. A. M. Sandercock, J. W. Westbrook, Q. Zhang, H. A. Johnson, T. M. Saielli, J. A. Scrivani, S. F. Fitzsimmons,

K. Collins, M. T. Perkins, J. H. Craddock, J. Schmutz, J. Grimwood, J. A. Holliday, Frozen in time: Rangewide genomic diversity, structure, and demographic history of relict American chestnut populations. *Mol. Ecol.* **31**, 4640–4655 (2022).

57. J. Goudet, HIERFSTAT, a package for R to compute and test hierarchical *F*-statistics. *Mol. Ecol. Notes* **5**, 184–186 (2005).

58. K. W. Broman, H. Wu, S. Sen, G. A. Churchill, R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).

59. D. Grattapaglia, R. Sederoff, Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: Mapping strategy and RAPD markers. *Genetics* **137**, 1121–1137 (1994).

60. D. H. Alexander, K. Lange, Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 246 (2011).

61. J. T. Lovell, A. Sreedasyam, M. E. Schranz, M. Wilson, J. W. Carlson, A. Harkess, D. Emms, D. M. Goodstein, J. Schmutz, GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *Elife* **11** (2022).

62. T. J. Ellis, D. L. Field, N. H. Barton, Efficient inference of paternity and sibship inference given known maternity via hierarchical clustering. *Mol. Ecol. Resour.*, doi: 10.1111/1755-0998.12782 (2018).

63. P. H. Noah, N. L. Cagle, J. W. Westbrook, S. F. Fitzsimmons, Identifying resilient restoration targets: Mapping and forecasting habitat suitability for *Castanea dentata* in Eastern USA under different climate-change scenarios. *Climate Change Ecology* **2**, 100037 (2021).

64. R.J. Hijmans, J. van Etten, M. Mattiuzzi, M. Sumner, J.A. Greenberg, O.P. Lamigueiro, A. Bevan, E.B. Racine, E. B., A. Shortridge, Raster: Geographic data analysis and modeling, version 2.1-49 (2013); https://CRAN.R-project.org/package=raster

65. L. H. Moro Rosso, A. F. de Borja Reis, A. A. Correndo, I. A. Ciampitti, XPolaris: An R-package to retrieve United States soil data at 30-meter resolution. *BMC Res. Notes* **14**, 327 (2021).

66. A. Dahl, V. Iotchkova, A. Baud, Å. Johansson, U. Gyllensten, N. Soranzo, R. Mott, A. Kranis, J. Marchini, A multiple-phenotype imputation method for genetic studies. *Nat. Genet.* **48**, 466–472 (2016).

67. A. J. Wilson, D. Réale, M. N. Clements, M. M. Morrissey, C. A. Walling, L. E. B. Kruuk, D. H. Nussey, An ecologist's guide to the animal model. *Journal of Animal Ecology* **79**, 13–26 (2010).

68. D. G. Butler, B. R. Cullis, A. R. Gilmour, B. J. Gogel, R. Thompson, ASReml-R reference manual version 4. *VSN International Ltd.* (2017).

69. J. Ødegård, T. H. E. Meuwissen, Estimation of heritability from limited family data using genome-wide identity-by-descent sharing. *Genet. Sel. Evol.* **44**, 16 (2012).

70. P. M. VanRaden, Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423 (2008).

71. S. Gezan, D. Murray, A. A. de Oliveira, G. Galli, ASRgenomics: An R package with complementary genomic functions, Version 1.0.0, VSN International (2021); https://CRAN.R-project.org/package=ASRgenomics

72. P. C. Austin, J. Merlo, Intermediate and advanced topics in multilevel logistic regression analysis. *Stat. Med.* **36**, 3257–3277 (2017).

73. M. L. Cipollini, J. P. Moss, W. Walker, N. Bailey, C. Foster, H. Reece, C. Jennings, Evaluation of an alternative small stem assay for blight resistance in American, Chinese, and hybrid chestnuts (*Castanea* spp.). *Plant Dis.* **105**, 576–584 (2021).

74. C. E. Conn, N. Howie, M. Lynch, S. Lee, E. Young, J. Westbrook, J. Holliday, Q. Zhang, M. L. Cipollini, Validation of an alternative small stem assay for blight resistance in chestnut seedlings and recommendations for broader use. *Plant Dis.* **107**, 1576–1583 (2023).

75. K. Baier, C. Maynard, W. Powell, Chestnuts and light: Early flowering in chestnut species induced under high-intensity, high-dose light in growth chambers. *Journal of the American Chestnut Foundation* **26**, 8–10 (2012).

76. B. Dumas, G. Freyssinet, K. E. Pallett, Tissue-specific expression of germin-like oxalate oxidase during development and fungal infection of barley seedlings. *Plant Physiol* **107**, 1091–1096 (1995).

77. Z. Li, S. Parris, C. A. Saski, A simple plant high-molecular-weight DNA extraction method suitable for single-molecule technologies. *Plant Methods* **16**, 38 (2020).

78. C.-L. Xiao, Y. Chen, S.-Q. Xie, K.-N. Chen, Y. Wang, Y. Han, F. Luo, Z. Xie, MECAT: Fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nat. Methods* **14**, 1072–1074 (2017).

79. C.-S. Chin, D. H. Alexander, P. Marks, A. A. Klammer, J. Drake, C. Heiner, A. Clum, A. Copeland, J. Huddleston, E. E. Eichler, S. W. Turner, J. Korlach, Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).

80. N. C. Durand, M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, E. S. Lander, E. L. Aiden, Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).

81. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).

82. R. Vaser, I. Sović, N. Nagarajan, M. Šikić, Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).

83. S. Shu, D. Goodstein, D. Rokhsar, *PERTRAN: Genome-guided RNA-seq read assembler* (2013); https://www.osti.gov/biblio/1241180

84. T. D. Wu, S. Nacu, Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010).

85. J. T. Lovell, J. Jenkins, D. B. Lowry, S. Mamidi, A. Sreedasyam, X. Weng, K. Barry, J. Bonnette, B. Campitelli, C. Daum, S. P. Gordon, B. A. Gould, A. Khasanova, A. Lipzen, A. MacQueen, J. D. Palacio-Mejía, C. Plott, E. V. Shakirov, S. Shu, Y. Yoshinaga, M. Zane, D. Kudrna, J. D. Talag, D. Rokhsar, J. Grimwood, J. Schmutz, T. E. Juenger, The genomic landscape of molecular responses to natural drought stress in *Panicum hallii*. *Nat. Commun.* **9**, 5213 (2018).

86. B. J. Haas, A. L. Delcher, S. M. Mount, J. R. Wortman, R. K. Smith, L. I. Hannick, R. Maiti, C. M. Ronning, D. B. Rusch, C. D. Town, S. L. Salzberg, O. White, Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **19**, 5654-5666 (2003).

87. A. F. A. Smit, R. Hubley, P. Green, P. RepeatMasker Open-3.0. (2015); http://www.repeatmasker.org

88. W. Bao, K. K. Kojima, O. Kohany, Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).

89. A. A. Salamov, V. V. Solovyev, Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res.*, **4**, 516-522 (2000).

90. G. S. C. Slater, E. Birney, Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, doi: 10.1186/1471-2105-6-31 (2005).

91. M. Stanke, O. Keller, I. Gunduz, A. Hayes, S. Waack, B. Morgenstern, AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–9 (2006).

92. R Core Team, "R: A language and environment for statistical computing" (R Foundation for Statistical Computing, Vienna, Austria, 2024); https://www.R-project.org/.

93. D. M. Emms, S. Kelly, OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

94. Y. Wang, H. Tang, J. D. Debarry, X. Tan, J. Li, X. Wang, T.-H. Lee, H. Jin, B. Marler, H. Guo, J. C. Kissinger, A. H. Paterson, MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).

95. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

96. S. Chang, J. Puryear, J. Cairney, A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* **11**, 113–116 (1993).

97. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

98. S. Chen, Y. Zhou, Y. Chen, J. Gu, FASTP: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

99. M. Love, S. Anders, W. Huber, Differential analysis of count data – the DESeq2 package. *Genome Biol.* (2013).

100. D. Tenenbaum, KEGGREST: Client-side REST access to the Kyoto Encyclopedia of Genes and Genomes (KEGG), R package version 1.48.1 (2025); https://bioconductor.org/packages/KEGGREST

101. A. Alexa A, J. Rahnenführer, topGO: Enrichment analysis for gene ontology. Bioconductor (2024).

102. D. J. Roiger, S. N. Jeffers, Evaluation of *Trichoderma spp.* for biological control of Phytophthora crown and root rot of apple seedlings. *Phytopathology* **81**, 910–917 (1991).

103. B. Holmes, Evaluation of *Phytophthora parasitica* var. *nicotianae* for biocontrol of *Phytophthora parasitica* on *Catharanthus roseus*. *Plant Dis* **78**, 193–199 (1994).

104. J.T. Lovell, qtlTools: Data processing and plotting in association with R/qtl (2020); https://github.com/jtlovell/qtlTools

105. K. W. Broman, D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, Ś. Sen, B. S. Yandell, G. A. Churchill, R/qtl2: Software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics* **211**, 495–502 (2019).

106. J. Yang, N. A. Zaitlen, M. E. Goddard, P. M. Visscher, A. L. Price, Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46**, 100–106 (2014).

107. Broman, K. W., & Sen, Ś., *A Guide to QTL Mapping with R/qtl* (Springer, 2009).

108. S. Fan, L. Georgi, F. Hebard, T. Zhebentyayeva, J. Yu, P. Sisco, S. Fitzsimmons, M. Staton, A. Abbott, C. Nelson, Mapping QTLs for blight resistance and morphological traits in inter-species hybrid families of chestnut (*Castanea spp.*). *Front. Plant Sci.* **15** (2024).

109. M. Huang, X. Liu, Y. Zhou, R. M. Summers, Z. Zhang, BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* **8**, giy154 (2018).

110. J. Wang, Z. Zhang, GAPIT Version 3: Boosting power and accuracy for genomic association and prediction. *Genomics, Proteomics & Bioinformatics* **19**, 629–640 (2021).

111. A. L. Price, N. A. Zaitlen, D. Reich, N. Patterson, New approaches to population stratification in genome-wide association studies. *Nature reviews. Genetics* **11**, 459 (2010).

112. G. Covarrubias-Pazaran, Genome-assisted prediction of quantitative traits using the R Package Sommer. *PLOS ONE* **11**, e0156744 (2016).

113. C. Fraslin, J.M. Yáñez, D. Robledo, R.D. Houston, The impact of genetic relationship between training and validation populations on genomic prediction accuracy in Atlantic salmon. *Aquaculture Reports* **23**, 101033 (2022).

114. Y. C. Wientjes, R. F. Veerkamp, M. P. Calus, The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. *Genetics* **193** (2013).

115. H. D. Daetwyler, M. P. L. Calus, R. Pong-Wong, G. de Los Campos, J. M. Hickey, Genomic prediction in animals and plants: Simulation of data, validation, reporting, and benchmarking. *Genetics* **193**, 347–365 (2013).

116. X. Zheng, M. X. Zheng, SNPRelate: A software package for parallel computation of relatedness and principal component analysis of SNP data. R package version 1.30.0 (2013).

117. O. F. Christensen, M. S. Lund, Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* **42**, 2 (2010).

118. I. Aguilar, I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, T. J. Lawlor, Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* **93**, 743–752 (2010).

119. A. Legarra, O. F. Christensen, I. Aguilar, I. Misztal, Single Step, a general approach for genomic selection. *Livest. Sci.* **166**, 54–65 (2014).

120. O. F. Christensen, P. Madsen, B. Nielsen, T. Ostersen, G. Su, Single-step methods for genomic evaluation in pigs. *Animal* **6**, 1565–1571 (2012).

121. J. W. R. Martini, M. F. Schrauf, C. A. Garcia-Baccino, E. C. G. Pimentel, S. Munilla, A. Rogberg-Muñoz, R. J. C. Cantet, C. Reimer, N. Gao, V. Wimmer, H. Simianer, The effect of the H^{-1} scaling factors τ and ω on the structure of H in the single-step procedure. *Genet. Sel. Evol.* **50**, 16 (2018).

122. R. C. Gaynor, G. Gorjanc, J. M. Hickey, AlphaSimR: An R package for breeding program simulations. *G3* **11** (2021).

123. P. Medina, B. Thornlow, R. Nielsen, R. Corbett-Detig, Estimating the timing of multiple admixture pulses during local ancestry inference. *Genetics* **210**, 1089–1107 (2018).

Acknowledgements: Laura Barth and Dan Mckinnon collected tissue for the *C. mollissima* reference genome sequencing. Shanmugan Rajasekar and David Kundra performed high-molecular-weight DNA extractions for the reference genomes. TACF volunteers and interns Israel Golden, Mira Polishook, Vinny Varsalona, Lizzy Dvorak, Daniel Ott, Mike Aucott, and Jay Brenneman assisted with phenotyping. Analyses were performed on Virginia Tech Advanced Research Computing clusters. Christine Wiese performed copy editing.

Funding:

Members and Donors of The American Chestnut Foundation

Oak Hill Fund (JW, JH)

Foundation for the Carolinas (JW, JL, JS, JG)

Colcom Foundation (JW, JS, JG)

Templeton World Charities (AN, EC, WP, JW)

U.S. Department of Agriculture National Institutes of Food and Agriculture projects 1018599, 1027966, and 1025004 (JH, JW, FH)

U.S. Forest Service (JW, SF, VL)

Tucker Foundation (JW)

Ballyshannon Foundation (SF, JW)

Orentreich Foundation for the Advancement of Science (SF, JW)

Office of Science of the U.S. Department of Energy Contract No. DE-AC02-05CH1123 (JS, JG, JL)

U.S. Department of Energy, Office of Science, Biological and Environmental Research, as part of the Plant-Microbe Interfaces (PMI) Scientific Focus Area at Oak Ridge National Laboratory, Contract No. DE-AC05-00OR22725 (TJT, NLE, MZM, TAR, JO)

U.S. Department of Energy, Office of Science, Biological and Environmental Research, as part of the Secure Ecosystem Engineering and Design (SEED) Scientific Focus Area at Oak Ridge National Laboratory, Contract No. DE-AC05-00OR22725 (JT).

The Center for Bioenergy Innovation (CBI), U.S. Department of Energy, Office of Science, Biological and Environmental Research Program under Award Number ERKP886 (JTL)

Author contributions:

Conceptualization: JW, JL, JH

Methodology: JW, JM, VL, SF, QZ, JL, TE, BL, TS, TE, AS, JHC, SK, CP, AS, XX, TJT

Investigation: JW**,** JM, QZ, VL, SFF, JVC, KC, SH, CS, LG, EJ, TS, BJ, LW, LK, CW, PZ, SK, TZ, XX, FC, JS, JO, SJ, KM, ES, MW, LB, JJ, PS, AN, EC, FH, JS, CH, MC, BC, EE, BL, AT, MZM, ZKY, JO, JT, TAR,

Data Curation: JW, AS, JH, DG, TJT

Visualization: JL, JW

Funding acquisition: JW, JH, SF, VL, KC

Project administration: JW, JL, SF, VL, JH

Supervision: JW, JH, JL, SF, VL, AA, CDN, KC, JS, JG, JC, WP, TJT

Writing – original draft: JW, JL, JH

Writing – review & editing: JW, JL, JH, JM, FH, AA, BL, AT, SJ, JS, SK, TJT, JO, JT, TAR

Competing interests: The Research Foundation for SUNY has established non-exclusive licensing agreements to distribute transgenic chestnut trees, pending regulatory approval, with American Castanea Public Benefit Corporation and potentially other entities.

Data and materials availability:

Data are deposited in Dryad https://doi.org/10.5061/dryad.4xgxd25mj. Raw genomic and RNA sequence data used to develop reference genomes can be accessed via NCBI bioproject PRJNA1147634. Chestnut reference genomes are accessible on Phytozome. Raw fastq from files from genotyping-by-sequencing can be accessed through NCBI under bioproject PRJNA507748. All other data are available in the manuscript or the supplementary material.

Supplementary Materials:

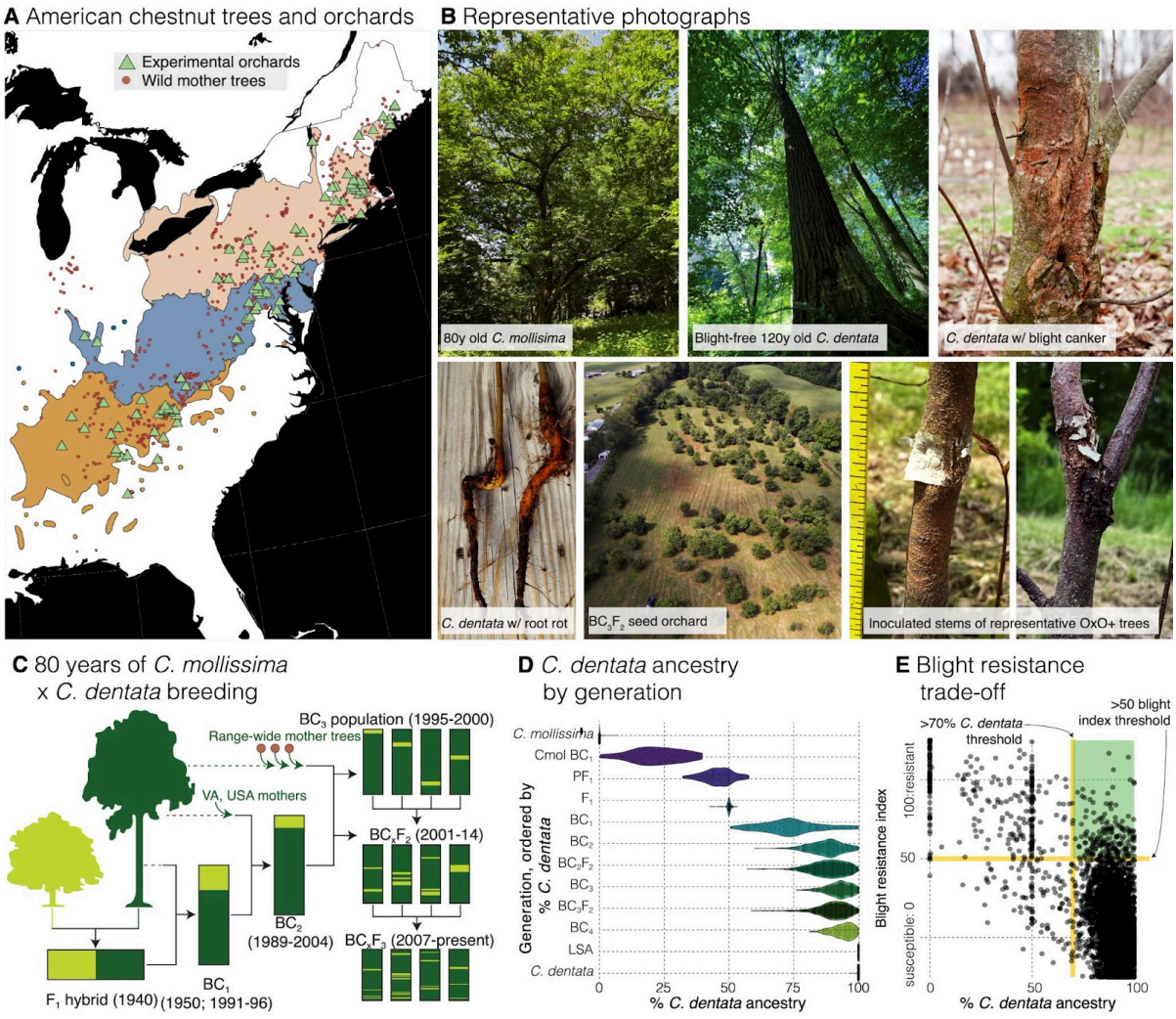

Fig. 1. Biogeography of *C. dentata* and breeding efforts to improve tolerance to its non-native fungal pathogens. **(A)** The native distribution of *C. dentata* has a center of diversity residing along the Appalachian Mountains. Conservation efforts involve breeding with rare reproductive *C. dentata* individuals (red dots) and growing progeny in orchards (green triangles) that span *C. dentata*'s three main climate regions (orange, blue, and pink polygons). **(B)** Photographs (from left to right) of the 'Mahogany' *C. mollissima* tree at the Connecticut Agricultural Experiment Station; a *C. dentata* in the Arboretum of Tervuren in Belgium (Photo: A. Sproul-Lattimer); a blight canker on a susceptible *C. dentata* stem; seedling roots affected by Phytophthora root rot; a backcross hybrid chestnut orchard in Meadowview, VA where trees with low blight resistance have been culled; susceptible (left) and resistant (right) blight canker responses on 'Darling 54' OxO+ trees. **(C)** A schematic of the TACF hybrid/backcross breeding program. Proportional segregation of *C. dentata* (dark green) versus *C. mollissima* (light green) ancestry is shown for F_1, backcross (BC), and intercross (BC_xF_x) generations. Wild mother trees in panel A are parents of the BC_2/BC_3 populations. (**D**) Variation in *C. dentata* ancestry among 12 backcross BC_x and intercross (BC_xF_x) generations, including the wild species progenitors (Cmol & Cden). PF1: 'pseudo-F1s' are progeny of advanced *C. dentata* backcrosses (e.g., BC_3) and *C. mollissima*. LSAC: 'large surviving American chestnuts' are *C. dentata* trees infected with blight, with surviving stems > 25 cm. Cmol BC_1: [*C. dentata* x *C. mollissima*] x *C. mollissima* progeny. **(E)** Correlations between *C. dentata* ancestry and additive genetic variation in an index of seven stem traits related to blight resistance among 3,499 trees that were artificially inoculated with *Cr. parasitica*. Blight resistance index breeding values were scaled from 0 = mean of *C. dentata*; 100 = mean of *C. mollissima*. The light green boxes highlight trees with intermediate to high levels of blight resistance (>50) and a majority of *C. dentata* ancestry (>70%).

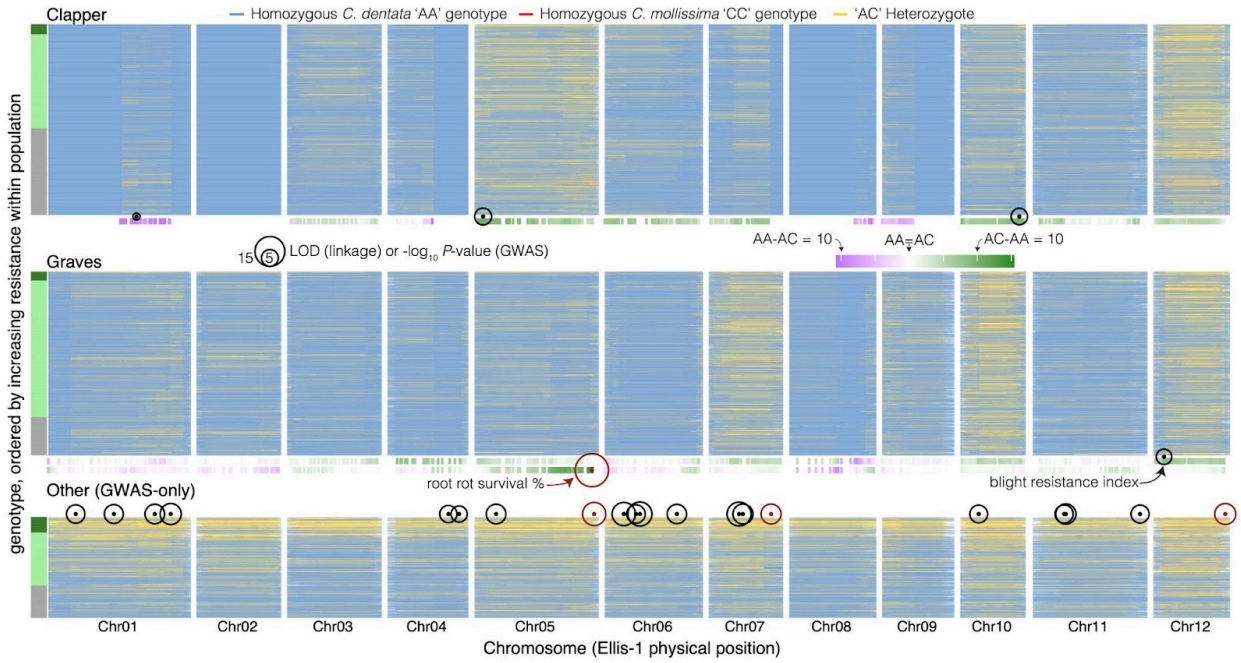

Fig. 2. Genetic architecture of blight and root rot resistance in American chestnut hybrid backcross populations. Recombination of *C. dentata* and *C. mollissima* species ancestry (red/yellow/blue map) is shown for *C. dentata* backcross descendants of The American Chestnut Foundation's two primary sources of resistance ('Clapper' $n = 1324$ and 'Graves' $n = 1275$) and 27 other *C. mollissima* sources of resistance ($n = 434$). In the purple-green horizontal bars below the genotype maps, green denotes regions where inheriting at least one allele from *C. mollissima* ('AC') enhances resistance, whereas purple denotes regions where inheriting both alleles from *C. dentata* ('AA') improves resistance. In addition to ancestry-based QTL mapping, genome-wide association studies (GWAS) for blight ($n = 3{,}365$) and root rot ($n = 475$) resistance were conducted for the combined population of 'Clapper', 'Graves', and other sources. Circles denote marker-trait associations (black - blight resistance, red - root rot resistance) with sizes proportional to the logarithm of the odds (LOD) scores for QTL or $-\log_{10}$ *P*-values for GWAS.

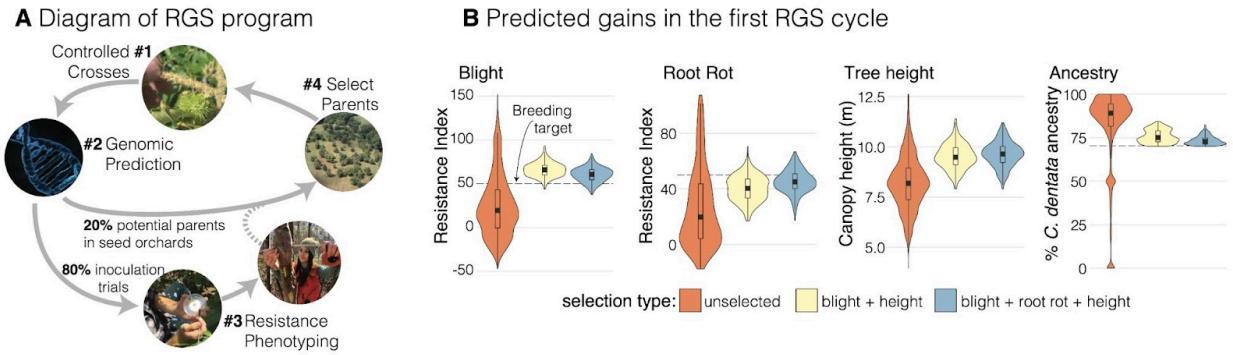

Fig. 3. Next steps for breeding to improve disease resistance and height growth in *C. dentata* hybrid backcross populations. **(A)** Diagram of a single recurrent genomic selection (RGS) cycle. The percentages of individuals passing to each step in the cycle are indicated. The dashed line shows that resistance phenotyping of relatives will be used to determine which parents to use in controlled crosses. **(B)** Genetic gains and *C. dentata* ancestry are predicted for two breeding tracks: 1. breeding parents that maximize blight resistance (>70) and height growth (70th percentile), or 2. breeding a smaller subset of parents that combine above-average blight resistance (>60), root rot resistance (>40), and height growth (>70th percentile). Gains were predicted from breeding the highest-performing parents with >70% *C. dentata* ancestry, followed by genomic selection of 20% of progeny with the highest predicted performance, based on accuracies estimated via cross-validation. Violin plots for the two breeding tracks depict variation in true breeding values for the top 20% selected trees. True breeding values were simulated as variables that were correlated with genomic estimated breeding values in proportion to genomic selection accuracies reported in Table S6. 'Unselected' depicts variation in breeding values estimated from phenotypes for the current population of surviving trees.

Science
AAAS

Supplementary Materials for

Genomic approaches to accelerate American chestnut restoration

Authors: Jared W. Westbrook[1]*, Joanna Malukiewicz[2,3], Qian Zhang[2], Avinash Sreedasyam[4,5], Jerry W. Jenkins[4], Vasiliy Lakoba[1], Sara Fitzsimmons[1], Jamie Van Clief[1], Kendra Collins[1], Stephen Hoy[1], Cassie Stark[1], Lake Graboski[1], Eric Jenkins[1], Thomas M. Saielli[1], Benjamin T. Jarrett[1], Lucinda J. Wigfield[1], Lauren M. Kerwien[1], Ciera Wilbur[1], Alexander M. Sandercock[2], J. Hill Craddock[1,6], Susanna Keriö[7], Tetyana Zhebentyayeva[8], Shenghua Fan[8], Austin M. Thomas[8], Albert G. Abbott[8], C. Dana Nelson[8,9], Xiaoxia Xia[10], James R. McKenna[11], Caleb Kell[12], Melissa Williams[4], LoriBeth Boston[4], Christopher Plott[4], Florian Carle[1,13], Jack Swatt[1], Jack Ostroff[1], Steven N. Jeffers[10], Kathleen McKeever[9], Erica Smith[9], Thomas J. Ellis[14], Joseph B. James[1], Paul Sisco[1], Andrew Newhouse[15], Erik Carlson[15], William A. Powell[15], Frederick V. Hebard[1], John Scrivani[1,16], Caragh Heverly[1], Martin Cipollini[1,17], Brian Clark[1], Eric Evans[1], Bruce Levine[1,18], John E. Carlson[19], David Goodstein[5], Jack Orebaugh[20], Zamin K. Yang[20], Madhavi Z. Martin[20], Joanna Tannous[20], Tomás A. Rush[20], Nancy L. Engle[20], Timothy J. Tschaplinski[20], Jane Grimwood[4], Jeremy Schmutz[4,5], Jason A. Holliday[2], & John T. Lovell[4,5,21]*

Corresponding authors: jared.westbrook@tacf.org and jlovell@hudsonalpha.org

The PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Data S1 to S29 accessible at Dryad: https://doi.org/10.5061/dryad.4xgxd25mj

MATERIALS AND METHODS

Blight phenotyping in hybrid and large surviving American chestnut populations

We phenotyped 5,501 *C. dentata* hybrids and wild-type parental species for 10 long-term blight resistance and growth traits between 2018 and 2023 (Supplementary Data 1). The trees were planted at 91 orchard locations and maintained by volunteers and staff of The American Chestnut Foundation (TACF). To assess blight resistance, trees were inoculated with the weakly virulent SG2,3 strain and/or the strongly virulent EP155 strain of *Cryphonectria parasitica* when they were between 2 and 15 years old (average inoculation age = 8 years). This population represents genetic diversity from a rangewide sample of 573 wild-type *C. dentata* parents. A total of 41 known *C. mollissima* parents contributed resistance to the hybrids; however, the vast majority of the hybrids (4,454 trees) are descendants of two hybrid founders: 'Clapper' a (*C. dentata* × *C. mollissima*) × *C. dentata* BC_1 hybrid that descended from the 'M16' *C. mollissima* parent (*45*), and 'Graves' a (*C. dentata* × *C. mollissima*) F_1 that putatively

descended from the 'Mahogany' *C. mollissima* (*46*). The large surviving American chestnut (LSAC) population descended from a total of 14 blight-infected *C. dentata* trees with surviving main stems > 25 cm in diameter.

Approximately 80% of the trees were inoculated using the cork-borer, agar-disk method (*15*). Non-inoculated trees were phenotyped based on natural chestnut blight infections. Trees were phenotyped when they were an average age of 15 years for presence/absence traits indicative of blight resistance or susceptibility, including a living main stem ('mainstemalive'), cankers >15 cm in length ('largecankers'), expanding cankers ('blightcontained'), exposed wood ('exposedwood'), *Cr. parasitica* fruiting bodies ('sporulation'), and stump sprouts ('stumpsprouts'). Before quantitative genetic analyses, the presence of susceptibility traits was coded as 0 and their absence as 1, so that trees lacking susceptibility phenotypes had higher blight resistance estimates. Blight cankers were also rated based on whether they were sunken (0, a susceptibility response), swollen (1, a partially resistant reaction), or flat (2, indicative of higher resistance) (named 'sunkenswollen'). The percent of the tree canopy that was dead as a result of blight infection was converted into canopy survival proportion ('propcanopysurvival'). The diameter of the largest living stem in centimeters ('dbhlargest_cm') was measured at breast height (1.37 meters) among 3,197 trees. Tree heights in meters ('height_m') were measured for 1,916 trees with a clinometer (Suunto PM 5/360 PC Clinometer) or a laser hypsometer (Nikon Forestry Pro II Laser Rangefinder).

Genetic analyses (*i.e.*, heritability, genome-wide association studies, and genomic prediction) were performed on genotyped hybrids, *C. dentata,* and *C. mollissima* controls. To determine whether analyzing inoculated and naturally infected trees together in these analyses introduced bias, heritability (Table S1), genomic selection accuracy (Table S6), and predicted gains from selection (Fig. S17; Fig. 3) were estimated separately for inoculated and naturally infected trees (n = 3,885) versus only trees that were inoculated (n = 3,499). Results were similar and in the main text we report results using only inoculated trees.

Comparisons of height growth among *C. dentata* backcross hybrids, *C. dentata*, and *C. mollissima*

To compare the height growth of *C. mollissima*, *C. dentata* first backcross (BC_1) hybrids, and LSACs, tree heights were measured in a closed canopy stand of 777 trees varying in age from 22 to 33 years at Lesesne State Forest, VA, that was planted and maintained by the Virginia Department of Forestry. Many of the trees had surviving main stems, possibly because chestnut blight severity was controlled through inoculation and subsequent spread, *Cr. parasitica* infected with hypoviruses that reduce fungal virulence (*18*). To account for age-based differences in tree heights, observed tree heights were divided by age-based expectations for yellow poplar (*Liriodendron tulipifera*) at site index = 90 (*47*). Age-adjusted tree heights were compared for the 30% fastest-growing individuals in each genotype class.

Genotyping

Genotyping-by-sequencing (GBS) was performed on 5,003 *C. dentata* backcross trees and wild-type species controls at Virginia Tech (Supplementary Data 2). Library preparation and bioinformatic methods were described in detail in (*17*). Briefly, young leaves were collected, homogenized in liquid nitrogen, and genomic DNA was extracted using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Samples were digested with ApeKI and barcoded with unique P1 and standard P2 adapters, which were ligated with T4 ligase. Libraries were subsequently PCR-amplified for 18 cycles. Libraries were assessed on 1% agarose gels, quantified with a Qubit fluorometer (Thermo Fisher Scientific, Waltham, USA), and randomly assigned to 96-well pools. Pools were size-selected on a BluePippen instrument (Sage Science, Beverly, USA) and sequenced on an Illumina NovaSeq 6000 S-Prime flow cell in 2x150bp format at the Duke University School of Medicine. FASTQ files were demultiplexed with the process_radtags function in Stacks (*48*) and aligned to the *C. dentata* 'Ellis-1' reference genome with bwa mem (*49*). Aligned sequence was converted to BAM format, sorted, and indexed with Samtools (*50*). Variants were called with GATK v4.2.6.1 (*51*) HaplotypeCaller (*52*) by creating gVCFs per individual and chromosome. The gVCFs were merged in batches of 50 samples with GenomicsDBImport and GenotypeGVCFs functions in GATK. The 50 sample VCF files were merged with bcftools to create a dataset comprising all samples and chromosomes. The VCF for the whole population was filtered to remove sites that did not meet the following quality thresholds: MQ<40.00, QD<2.0, FS>40.000, MQRankSum< -12.500, ReadPosRankSum<-8.000, and SOR>3.0. Variants detected in repetitive regions of the 'Ellis-1' *C. dentata* genome were masked. Remaining variants were further filtered with bcftools to retain 93,332 biallelic SNPs with phred quality scores > 30 and < 20% missing data across individuals. After filtering, the median per-sample sequencing depth was 36.7×. The filtered VCF was imputed and phased in Beagle v 5.4 (*53*, *54*).

Species ancestry inference

Castanea species ancestry in hybrids (Supplementary Data 3) was inferred in Ancestry_HMM (*55*) using whole genome sequence (WGS) data from 21 *C. dentata*, 23 *C. mollissima*, 15 *C. crenata*, 16 *C. sativa*, 27 *C. pumila*, 20 *C. henryi,* 17 *C. seguinii* obtained from (*56*) and (*29*). Sequence was aligned to the *C. dentata* 'Ellis' reference genome, and variant calling was performed using the same methods as for GBS to generate a final VCF containing 143,933,272 variants. Biallelic SNPs genotyped with GBS (n = 93,332) were subset from the WGS data, and these datasets were merged with bcftools. Separate Ancestry_HMM runs were performed for the following putative ancestry combinations: *C. dentata* v. *C. mollissima* (4,504 trees), *C. dentata* v. *C. mollissima* v. *C. crenata* (479 trees), *C. dentata* v. *C. sativa* v. *C. crenata* (6 trees), and *C. dentata* v. *C. pumila* v. *C. henryi* (14 trees). Before running the analyses, per-locus Fst between reference populations was estimated with the R package 'hierfstat' (*57*). Ancestry informative markers with Fst > 0.7 between reference populations and > 1 Kb physical distance between markers were subset, and reference and alternative alleles were counted for each reference population. Genetic distances between markers were estimated by building two parental genetic maps in R/QTL (*58*) from GBS data from 125 full sibling progeny of two wild-type *C. dentata* (GMBig × Horn) using the pseudo-test cross method (*59*) (Supplementary Data 4). The 'Horn' map, composed of 1,003 markers and covering 99.5% of the genome, was used to interpolate genetic map positions for ancestry-informative markers using the program predictGMAP (https://github.com/szpiech/predictGMAP). For the *C. mollissima* v. *C. dentata* ancestry estimates, Ancestry_HMM was run by specifying average global ancestry proportions of 0.88 for *C. dentata* and 0.12 for *C. mollissima*, four pulses of *C. dentata* ancestry (corresponding to F_1 through BC_3 generations) and one pulse of 5% *C. mollissima* ancestry (corresponding to intercrossing in the BC_3F_2 generation). For the three-way ancestry contrasts, population mean ancestry proportions were first estimated with ADMIXTURE (*60*), and three pulses of ancestry were specified corresponding to these proportions. Ancestry calls (*i.e.*, AA = homozygous *C. dentata*, AC = heterozygous, and CC = homozygous *C. mollissima*) were made if a genotype probability was greater than 0.5. Start and end coordinates for species ancestry blocks were obtained using the 'add_rle' function in GENESPACE (*61*). Potentially spurious short ancestry runs (< 20 markers) were converted to missing because the likelihood of two crossover events in such proximity is low, and these short blocks are likely genotyping errors. Global *C. dentata* ancestry percentages for individual trees were estimated as % *C. dentata* = $(2\Sigma L_{AA} + \Sigma L_{AC})/2L$ *100, where ΣL_{AA} and ΣL_{AC} are the summed lengths of AA of the AC genotype blocks and L is the total length of the genome in Mb. Supplementary Data 5 contains the coordinates for the ancestry tracts of individual trees that were used for QTL mapping.

Pedigree reconstruction

Pedigree records for phenotyped or genotyped trees, their parents, and previous generations back to founders were obtained from TACF's online database, dentataBase (https://acf.herokuapp.com/; Supplementary Data 6). Complete maternal and paternal pedigree records were available for 1,965 of 5,501 the blight phenotyped trees that were progeny of controlled pollinations. Paternity was assigned to an additional 2,007 trees using the genotypes of 288 candidate fathers with the program FAPs v. 2.6.4 (*62*). A total of 182 SNPs with less than 10% missing genotype data, that were spaced at least 1.1 Mb apart, and had expected levels of heterozygosity, were selected for paternity analysis. The FAPs 'paternity_array' and 'sibship_cluster' functions were run assuming a mutation rate of 0.0015 and a maximum of four paternity clashes. Paternity was assigned to the highest probability father for 1,110 progeny, and an additional 897 progeny that were missing a single highest probability father were assigned to their second highest probability father.

Estimation of a blight resistance index

For each tree, we estimated a 'blight resistance index' by summing adjusted phenotypes for seven stem severity traits (*i.e.*, 'mainstemalive', 'largecankers', 'blightcontained', 'exposedwood', 'sporulation', 'sunkenswollen', and 'stumpsprouts'). We selected this combination of traits because it was most strongly correlated with average blight canker ratings of progeny (see 'progeny validation of blight resistance'). Phenotypes were adjusted for fixed effects of inoculation status (for analyses of inoculated + naturally infected trees), tree age class, climate, and soil variables important for *C. dentata* habitat suitability (*63*). Average monthly maximum temperature ('tmax') and average monthly precipitation ('prec') for each orchard were retrieved using the 'raster' R package (*64*). Soil pH and soil percent sand content were retrieved from XPolaris in R (*65*). To model continuous tree ages, soil, and climate variables as discrete fixed factors, values were assigned to quartiles based on the distribution across individuals and orchard locations. The residual variation from this model was scaled from 0 to 1. These residuals were then weighted by multiplying by each trait's heritability (see 'heritability estimation'). Missing trait values were imputed in 'phenix' (*66*) from correlations among traits and pedigree relationships among trees. Weighted residuals were

summed from the variables. Blight resistance index phenotypes and breeding values (i) were separately scaled to 0 = mean for susceptible *C. dentata* and 100 = mean for resistant *C. mollissima* using the formula:

$$i = (i' - \mu_{am}/(\mu_{ch} - \mu_{am})$$

Where, μ_{am} and μ_{ch} are the average unscaled blight resistance index values (i') for 96 artificially-inoculated susceptible (non-LSAC) *C. dentata* and 69 resistant *C. mollissima* trees, respectively.

Heritability estimation for blight resistance and growth traits

We used an 'animal model' to partition phenotypic variance in disease resistance and growth traits into additive genetic and environmental components using genomic relationships among individuals (*67*). Variance components were estimated in ASReml-R v. 4.2 (*68*) with the model $y = Xb + Zu + e$, where the vector of individual tree phenotypes (y) is a function of fixed effects (b), random additive genetic effects (*i.e.*, breeding values) u ~ N(0, $\sigma^2_a\mathbf{G}$), and residual error e ~ N(0, $\sigma^2_e\mathbf{I}$). The incidence matrices X and Z relate phenotypes to fixed and random effects, respectively. We used a genomic relationship matrix (**G**) rather than pedigree relationships to estimate additive genetic variance (σ^2_a). The log-likelihood for the model in which additive genetic variance was estimated using genomic relationships was significantly greater ($P < 0.0001$, LRT) than that for the model using pedigree relationships. Furthermore, genomic relationships estimate Mendelian segregation within families and provide less-biased estimates of variance components in cases like ours, where genetic and environmental effects are partially confounded (*69*). The **G** matrix was estimated with the VanRaden method (*70*) in 'ASRgenomics'(*71*) after pruning the SNP matrix to 48,266 markers with minor allele frequencies > 0.05 and genotypic correlations between markers < 0.9.

Heritability, or the proportion of phenotypic variance that is attributable to the additive effects of genome-wide allelic variation, was estimated as $h^2 = \sigma^2_a / (\sigma^2_{a+}\sigma^2_e)$, where σ^2_a is the additive genetic variance and σ^2_e the error variance. To weight traits in the blight resistance index, we first estimated heritability for individual stem severity traits in a model that included the fixed effects of inoculation status, tree age class, tmax, prec, soil pH, and % sand. For binary traits, σ^2_e was assumed to equal the variance of the standard logistic distribution ($\pi^2/3 = 3.29$) (*72*). We compared heritability estimates for inoculated trees spanning the full range of *C. dentata* and *C. mollissima* ancestry (n = 3,499) versus the subset of trees with 70% to 100% *C. dentata* ancestry (n = 3,143). We also compared these heritability estimates to those obtained the larger population of naturally infected + inoculated trees (n = 3,885).

Progeny validation of blight resistance in backcross and large-surviving American chestnut populations

To determine whether parental blight resistance was predictive of progeny blight resistance, we conducted two years of *Cr. parasitica* inoculation experiments on first-year seedling progeny. In 2024, we inoculated 1,015 seedling progeny from 30 families, including three open pollinated (OP) *C. mollissima* families (resistant control), one *C. dentata* × *C. mollissima* F_1 family (intermediate resistance control), one OP *C. dentata* family (susceptible control), ten OP *C. dentata* backcross families, eight controlled-pollinated (CP) backcross-F_2 families, six quasi-BC_1 (F_1 × BC) families, and one LSAC × LSAC family (Supplementary Data 7). Seedlings were grown in 1.5-gallon pots for five months before inoculation in an outdoor nursery at the University of Tennessee Chattanooga (~70% of the seedlings within families) and Berry College in Georgia (30% of the seedlings). Seedlings were inoculated by cutting off the apical leader of the seedlings and placing an agar disk containing inoculum from the highly virulent EP155 strain of *Cr. parasitica*. Ninety days post inoculation, orange zone canker lengths were measured to the nearest mm, and the presence or absence of fungal fruiting bodies was noted (*73*, *74*). To control for outliers in the canker length distribution, canker severity ratings, on a 1 to 5 scale, were derived by assigning trees to quartiles from the canker length distribution and adding one to the rating if fungal fruiting bodies were present. Family mean canker ratings were scaled from 0 (*C. dentata* mean) to 100 (*C. mollissima* mean), and correlations were estimated with average blight resistance index and *C. dentata* ancestry of the parents (Fig. S2).

In 2025, an additional seedling inoculation experiment was conducted to confirm that LSAC families have heritable blight resistance. Open-pollinated seeds were collected from six of 24 first-generation LSAC progeny planted together in the Maryland 'Scrivener' orchard that was isolated from other chestnuts. The LSAC progeny were initially obtained by the landowner from the American Chestnut Cooperators Foundation and were confirmed with genotype data to have 100% *C. dentata* ancestry. The wild LSAC parents were unknown to us, but genomic relationships among these trees ranged from 0.07 to 0.75 (mean 0.30), indicating that the progeny were a mixture of half- and full-siblings. The trees were naturally infected with *Cr. parasitica*, and the population as a whole had blight

resistance index breeding values ranging from 0 to 54 (average 27). A total of 85 second-generation LSACs (n = 6 to 26 per family) were grown for five months at the Baltimore County Agricultural Center along with resistant *C. mollissima* (n = 40), susceptible *C. dentata* (n = 20), and three *C. dentata* backcross families where the parents had 15% to 25% *C. mollissima* ancestry (Supplementary Data 8). Seedlings greater than 4 mm in groundline diameter were inoculated by using parafilm to affix an agar disk of the highly virulent EP155 *Cr. parasitica* strain into a 1 mm wide × five long incision in the stem made four cm above the soil surface. Sixty days post-inoculation, cankers were rated on a scale of 1 (inoculation failure), 2 (minimal canker expansion within 5 mm of inoculation site), 3 (large canker with no sporulation), 4 (large, sunken, sporulating canker), and 5 (tree dead with large cankers). After removing trees with inoculation failures, family mean blight resistance ratings were scaled from 0 (the mean of the susceptible *C. dentata* family) to 100 (the mean of the resistant *C. mollissima family*). Tukey tests were conducted to compare family mean resistance.

Progeny testing of 'Darling 54' OxO transgenic lines

The 'Darling 54' line was developed at the State University of New York (SUNY-ESF) via Agrobacterium-mediated insertion of a wheat oxalate oxidase gene, constitutively expressed with a 35S promoter, in the 'Ellis-1' wild type *C. dentata* background from New York. The OxO gene has subsequently been introgressed into wild-type *C. dentata* populations via five generations of outcrossing (*24*). It was previously assumed that another founder line ('Darling 58') was the donor parent for OxO introgression; however, genome sequencing of first- and second-generation progeny (T1 & T2) revealed that outcrossing was performed using 'Darling 54' as a founder (*24*). In 'Darling 54', OxO is inserted into a chestnut gene (Caden.04G062600) that is homologous (81% identical in amino acid sequence) to the *SAL1* gene in *Arabidopsis thaliana* (AT5G63980) (*19*). To determine if Caden.04G062600 is a single copy gene in chestnut, we blasted the protein sequence against the *C. dentata* 'Ellis-1' genome. The most similar gene shared only 42% amino acid sequence similarity. Furthermore, Caden.04G062600 was assigned to an orthogroup with four other single-copy orthologs from our four *C. mollissima* reference genomes. These results support the interpretation that SAL1 orthologs are single-copy genes across all *Castanea* reference genomes.

To evaluate the blight resistance and growth of 'Darling 54' progeny, pollen from first and second generation (T1 & T2) 'Darling 54' progeny was produced at SUNY-ESF and University of New England by growing seedlings indoors with supplemental light and fertilizer treatments (*24*, *75*). Controlled pollinations between four 'Darling 54' T1 & T2 pollen lines and 18 grafted wild-type *C. dentata* from Indiana were performed in 2019 and 2020 at the Purdue University Martell Forest under a permit from the U.S. Department of Agriculture Animal and Plant Health Inspection Service. To detect the inheritance of OxO in the progeny, a small slice of tissue from the chestnut seeds (*i.e.*, embryonic cotyledons) was tested for the presence of oxalate oxidase activity using a histochemical assay. The assay detects H_20_2 after the addition of oxalic acid to the media (*19*, *76*). Progeny were planted in family row plots with 1.5 feet between trees and 8 feet between rows. Progeny were planted in alternating order of OxO+ and OxO- in a "case-control" experimental design. When the trees were three years old, they were inoculated with *Cr. parasitica* isolated from the Potawatami Park BC_3F_2 hybrid chestnut orchard, which had initially been inoculated with the SG2,3 strain. One year post-inoculation, canker length and widths were measured in millimeters. Presence or absence of fungal fruiting bodies (an indicator of susceptibility) was noted. Canker severity was rated on a scale of 1 (minimal expansion beyond initial lesion) to 5 (large sunken and sporulating). To compare growth rates between OxO+ and OxO- progeny, tree heights were measured in cm, and ground line diameter was measured in millimeters (Supplementary Data 9). Original photographs of cankers on Darling 54 trees from Fig. S5 can be accessed on Dryad (https://doi.org/10.5061/dryad.4xgxd25mj).

High molecular weight DNA extraction for reference genome assemblies

High molecular weight (HMW) DNA was extracted from young leaves at Clemson University ('Ellis-1' *C. dentata)* and Arizona Genomics Institute ('Mahogany' and 'Nanking' *C. mollissima*). For the 'Ellis-1' *C. dentata* genome, tissue culture-generated plantlets were grown at SUNY-ESF in the dark to reduce the accumulation of carbohydrates. For the 'Mahogany' and 'Nanking' *C. mollissima* genomes, newly emerging leaves from mature trees at the Connecticut Agricultural Experiment Station and Meadowview Research Farms, respectively, were wrapped in a shade cloth for 48 hours. Leaves were flash frozen in liquid nitrogen upon collection and stored at -80 °C. HMW DNA extraction was performed using the protocol described in (*77*) with modifications. Briefly, nuclei were isolated using Nuclear Isolation Buffer (NIB) composed of 0.5 M sucrose, 2% of polyvinylpyrrolidone (M.W.40,000), 0.5% Triton X-100, 10 mM KCl, 50 mM EDTA, four mM spermidine trihydrochloride, one mM spermine tetrahydrochloride in 10 mM Tris-HCl at pH 8.0. Before use, the NIB was supplemented with 0.1% 2-Mercaptoethanol. The tissue weight to NIB buffer volume ratio was adjusted to 1:10. Pelleted nuclei were washed

with NIB supplemented with 0.5% Triton X-100 and resuspended in CTAB lysis buffer (2% CTAB, 2M NaCl,100 mM EDTA, 100 mM Tris-HCl, pH 8.0) in the presence of Proteinase K (0.45 mg /ml) either overnight on ice or for 2 hours at 65°C. DNA was phase-separated with an equal volume of chloroform and precipitated with 0.7 volume of cold isopropanol. The DNA was washed with 70% ethanol, dried at room temperature, and resuspended in 0.01 M Tris-HCl, pH 8.0. Then, the DNA was treated with three µL of the Ambion® RNase Cocktail™ (Thermo Fisher Scientific) for 30 min at 37°C, followed by chloroform:isoamyl alcohol (24:1 vol/vol) extraction and precipitation with two volumes of ethanol. The DNA was resuspended in 100 µL of 0.01 M Tris-HCl, pH 8.0. The quality and integrity of the DNA were evaluated using a Qubit 2 Fluorometer (Thermo Fisher Scientific Inc., USA) and a NanoDrop ND-8000 spectrophotometer (Thermo Fisher Scientific Inc., USA), followed by Pulse Field Gel Electrophoresis in agarose.

Genome assembly: *C. dentata* 'Ellis'

We produced 118.58× coverage of PacBio continuous long reads (CLR), which were assembled with MECAT (*78*) and polished with Quiver (*79*). A total of 24 misjoins were corrected in the polished assembly using a 980-marker genetic map from the 'Cranberry' *C. dentata* tree (*41*). Contigs were assembled into larger scaffolds with HiC data using the JUICER pipeline (*80*). Using the genetic map, a total of 160 joins were applied to the scaffolds. The final assembly contained 99.37% of the assembled sequence in 12 chromosomes. Chromosome assemblies were ordered and oriented with respect to the *Quercus robur* genome (http://www.oakgenome.fr/), and the telomeric sequence was found to be correctly oriented. Forty-eight adjacent alternative haplotypes were collapsed using the longest common substring. The raw CLR data were used to correct 76,296 heterozygous SNPs and INDELs (2.8% of the 2,718,565 heterozygous phasing errors). A total of 533 homozygous SNPS and 137,372 homozygous INDELs were corrected across 662,414,933 callable bases using 100X Illumina reads (2x150, 500bp insert).

Genome assembly: *C. mollissima* 'Mahogany' and 'Nanking

We produced 112.82-125.78× of single-haplotype PacBio HiFi coverage, where the average read length was 16,111-23,789 bp across the two genomes. Two haplotypes for each genome were assembled using HiFiAsm+HIC (*81*). The resulting sequence was polished using RACON (*82*). We searched for misjoins in the assemblies using HiC data, but none were identified. Contigs for both haplotypes were oriented, ordered, and joined with JUICER (*80*). Six joins were applied to form the final assembly, in which the 12 chromosomes contained 99.27-99.66% of the assembled sequence. One adjacent alternative haplotype was collapsed in the assemblies using the longest common substring between the two haplotypes. A total of 966/1,655 heterozygous SNPs and INDELs were corrected using the HiFi data. Additionally, 298/554 homozygous SNPs and 12,027/13,402 homozygous INDELs were corrected using 50-51× Illumina reads (2 × 150, 400bp insert). No vectors or contaminants were found in the assembled sequence. Chromosomes were numbered and oriented using the V1 'Ellis-1' *C. dentata* release.

Genome annotations

To annotate gene models, RNA sequencing was performed on female flowers, catkins, leaves, bark, and stems that were wounded and/or infected with *Cr. parasitica* using methods described below in 'RNA extraction and sequencing.' Gene models were annotated using our standard pipeline developed by the DOE Joint Genome Institute and Phytozome. Transcripts were assembled using PERTRAN (*83*) from 4.28 billion 2 × 150-bp stranded paired-end Illumina RNA-seq data. PERTRAN conducts genome-guided transcriptome short-read assembly via GSNAP (v2013-09-30) (*84*) and builds splice alignment graphs after alignment validation, realignment, and correction (*85*). Subsequently, PASA (v2.0.2) (*86*) was used to assemble 181,468, 148,147, 146,457, 153,886, and 153,743 transcripts for the 'Ellis', 'Mahogany' hap1, 'Mahogany' hap2, 'Nanking' hap1, and 'Nanking' hap2 genomes, respectively. The assembled transcripts were aligned to the repeat-soft-masked genomes with up to 2,000 bp of extension at both ends, unless extending into another locus on the same strand. Repeats were identified using RepeatMasker v.4.1.3 (*87*), RepeatModeler v. 1.0.11 (open), and RepBase (*88*). Homology support was provided by alignments to seven publicly available genomes (*Arabidopsis thaliana*, soybean, sorghum, grape, peach, rice, and *Setaria viridis*) and Swiss-Prot proteomes. Gene models were predicted using the homology-based predictors FGENESH+ (v.3.1.0) and FGENESH_EST (*89*), which use expressed sequence tags to compute splice-site and intron information rather than protein/translated open reading frames (ORFs). In addition, gene models were predicted using EXONERATE (v.2.4.0) (*90*) and AUGUSTUS (v.3.1.0) (*91*), both trained with high-confidence PASA assembly ORFs and intron hints from short-read alignments. The selected gene models were analyzed with Pfam, and those with more than 30% Pfam TE domains were removed. We also removed (1) incomplete, (2)

low-homology-supported without full transcriptome support, and (3) short single exon (<300 bp coding DNA sequences) without protein domain or transcript-supported gene models.

Comparative genomics

We used GENESPACE v1.2.3 (*61*) and DEEPSPACE v0.0.1 (https://github.com/jtlovell/DEEPSPACE) in R v4.4.1 (*92*) to conduct all comparative genomics analyses. GENESPACE employs OrthoFinder v2.5.4 (*93*) to build orthogroups, then parses orthogroup-constrained protein-protein BLAST hits into syntenic blocks with MCScanX (*94*). This produced genome-wide synteny-constrained sets of orthologs across the five reference haplotypes (Supplementary Data 10 & 11). DEEPSPACE finds syntenic block coordinates by converting query genomes into windows and aligning these to the reference genomes with minimap2 v2.28 (*95*) (Supplementary Data 12). Despite over 20 million years of divergence between *C. mollissima* and *C. dentata* (35), we expected these genomes to be highly conserved because of the high degree of synteny among the Fagaceae (36). As expected, there was strong sequence collinearity between species (mean = 662.0Mb, 94.7% of the Ellis genome) (Fig. S6). Genome structure within *C. mollissima* was even more conserved (mean = 3.50Mb of inverted sequence), and the two Nanking haplotypes only segregate for a single 110kb inversion. This similar genomic organization provides a strong foundation for comparative genomics and the discovery of candidate resistance or susceptibility alleles.

Experimental design for *Cr. parasitica* RNAseq timecourses

Species-specific responses to *Cr. parasitica* inoculation were compared at 3 and 10 days post-inoculation (Supplementary Data 13). RNA from six biological replicates of *C. mollissima*, *C. dentata*, and their F_1 hybrids was sequenced at each time point. Two genotypes per species or hybrid were sequenced, and half of the biological replicates were derived from each genotype. The *C. dentata* genotypes included 'Ellis-1' and full sib progeny from trees from Tennessee (SM156 × TNCAN1). 'Mahogany' and 'Nanking' were the *C. mollissima* genotypes. The two F_1 genotypes were KY115 and WWC70, progeny of two *C. dentata* trees from Virginia crossed with 'Mahogany' and 'Nanking' *C. mollissima* trees, respectively. To generate biological replicates, all genotypes were clonally propagated by whip-and-tongue grafting, except for the SM156 × TNCAN1 *C. dentata* progeny, for which full siblings served as biological replicates. Grafts and seedlings were grown in an outdoor nursery in two-gallon pots at the University of Tennessee Chattanooga for 6 months before inoculation. For each species and for F_1 hybrids, transcript abundance was compared between two treatments: (1) stem wounding and inoculation with the EP155 strain of *Cr. parasitica* and (2) stem wounding without inoculation (control treatment). All trees were wounded by making a 1 cm long × 2 mm wide excision with a scalpel. Stem segments were sampled by cutting 0.5 cm above and below the wound with Felco pruners and immediately flash-frozen in liquid nitrogen.

RNA extraction and sequencing

Total RNA was extracted at the Connecticut Agricultural Experiment Station after homogenizing tissue with a mortar and pestle in liquid nitrogen. Each RNA sample was extracted from 100 to 200 mg of frozen tissue in RNase-free 5 ml tubes using a CTAB-based RNA extraction buffer utilizing the protocol of (*96*). Each sample was extracted with 3 ml of CTAB buffer with 120 µl of 20% SDS, extracted twice with chloroform:IAA, and precipitated with isopropanol. The pellets were washed with 70% ethanol and resuspended in nuclease-free water. Samples that had at least 400 ng/µl nucleic acids based on Nanodrop 2000 (Thermo Scientific) readings were purified with an RNA extraction kit (PureLink RNA Mini Kit, Invitrogen, cat. No. 12183018A) combined with on-column DNase treatment (PureLink, Invitrogen, cat. 12185-010). RNA quality and quantity were assessed using Agilent TapeStation RNA ScreenTapes. Samples with < 200 ng/µl of RNA were vacuum-concentrated, and the concentration and quality of these samples were re-evaluated with Invitrogen Qubit 4 fluorometer, Qubit BR RNA assays, and Qubit RNA IQ assays. Samples that had RNA Integrity Number (RIN) or RNA IQ score > 7.0, no DNA contamination, and > 200 ng/µl of RNA were sequenced. Messenger RNA was isolated for sequencing at HudsonAlpha Institute for Biotechnology using a Sciclone NGS robotic liquid handling system (PerkinElmer) and Illumina's TruSeq Stranded mRNA HT Poly(A) RNA sample preparation kit. One µg of total RNA per sample was used as starting material, and eight cycles of PCR were performed for library amplification. The prepared libraries were quantified using KAPA Biosystems' next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. Sequencing of the flow cell was performed on the Illumina NovaSeq sequencer using NovaSeq XP v.1 reagent kits and an S4 flowcell, following a 2×150 bp indexed run recipe.

Resistant and susceptible gene expression responses to *Cr. parasitica*

To contrast *Castanea* species gene expression responses to *Cr. parasitica*, reads from *C. dentata* were aligned to the 'Ellis-1' reference, and reads from *C. mollissima* samples were aligned to 'Mahogany' hap1 genome assemblies using STAR 2.7.10a (*97*) after filtering sequencing adaptor and low quality sequences in FASTP v0.23.2 (*98*). Expression responses to *Cr. parasitica* inoculation were quantified within species at each time point by contrasting read counts from the inoculation treatment versus the wounding-only control in DESeq2 (*99*) (Supplementary Data 14 & 15). Within species, the response to *Cr. parasitica* inoculation was deemed significant for genes with $|\log_2$ fold change$| > 1$ and Benjamini-Hochberg FDR-adjusted $P < 0.05$ (Supplementary Data 16).

To identify orthogroups where *C. mollissima*-specific copy number expansion resulted in increased expression responses to *Cr. parasitica*, we first tallied gene copy number within orthogroups from a 'pan genome' file produced by GENESPACE (*61*) that compared copy number for all orthogroups from the 'Mahogany' hap1 genome relative to the other reference genomes (*e.g.*, 'Mahogany' hap2, 'Nanking' hap1, 'Nanking' hap2, and 'Ellis'). For each biological replicate, we summed transcripts per million (TPM) across genes within orthogroups with increased copy number in all *C. mollissima* reference genomes relative to *C. dentata*. We used *t*-tests to identify orthogroups where the inoculated *C. mollissima* samples had significantly greater TPM values at Benjamini-Hochberg FDR-adjusted $P < 0.05$ than the inoculated *C. dentata* samples within each timepoint (Supplementary Data 17).

To compare expression responses between species, single-copy orthologs between the *C. dentata* 'Ellis' and *C. mollissima* 'Mahogany' hap1 reference genomes were inferred in GENESPACE. Single-copy genes were classified as up- or down-regulated in *C. mollissima* only, *C. dentata* only, both species, or not significantly differentially expressed in either species. To identify KEGG pathways that were significantly enriched among genes differentially expressed in individual species, chestnut genes were classified into pathways based on orthology to *Arabidopsis thaliana* using the R package 'KEGGREST' (*100*) (Supplementary Data 18). Similarly, the 'TopGO' R package (*101*) was used to test for enrichment of biological process gene ontology terms. Fisher's Exact tests were performed to determine if KEGG pathways and GO terms were significantly over-represented at Benjamini-Hochberg FDR-adjusted $P < 0.05$.

Metabolomic profiling

Stem metabolite profiles were compared among 8 *C. mollissima* trees (from the 'Cropper' and 'Qing' sources), 11 *C. dentata* trees (from the 'WB275-27', 'Zoar', and 'Lasdon' sources), and 12 (*C. dentata* × *C. mollissima*) × *C. dentata* BC_1 hybrids (from the 'KL' and 'GR68' sources) planted at SUNY-ESF in Syracuse, NY. Trees were two to seven years old at the time of sampling in July 2013. Stem tissue ranging in diameter from 3 to 7 mm from the previous year's growth was collected with sterilized hand pruners and flash frozen. Frozen stems were lyophilized and ground to a fine powder using a Wiley Mini Mill to 40 mesh. The powdered chestnut stems (~50 mg) were weighed into vials and extracted overnight with 2 mL of 80% ethanol to which 75 µL sorbitol (1 mg*mL^{-1} aq.) was added as an internal standard. The extract was removed, and overnight extraction was repeated with an additional 1 mL of 80% ethanol. The second day extract was combined with the first, mixed well, and a 500 µL aliquot was dried under a stream of nitrogen for analysis by gas chromatography-mass spectrometry (GC-MS). The dried extract was dissolved in 500 µL acetonitrile, followed by the addition of 500 µL of N-Methyl-N-(trimethylsilyl)trifluoroacetamide with 1% trimethylchlorosilane (MSTFA + 1% TMCS), and heated for one hour at 70 °C to produce trimethylsilyl (TMS) derivatives. After 2 days, a 1 µL aliquot was injected into an Agilent 7890A GC coupled to a 5975C inert-XL MS. The GC was fitted with a Restek Rtx-5MS column with an Integra Guard (30 m x 0.25 mm I.D. x 0.25 µm df; 5 m Integra Guard). Helium flow was 1 mL per minute. The initial oven temperature was set to 50 °C for 2 minutes, followed by a 20 °C/min ramp to 325 °C, then held for 11 minutes. The MS was operated in electron impact (70 eV) ionization mode with a scan range of 50-650 Da. The source temperature was set to 230 °C, and the quadrupole to 150 °C. Metabolite peaks were extracted using key mass-to-charge selected ions, then scaled back to the total ion chromatogram using previously calculated scaling factors and normalized to the internal standard recovered, extract volume analyzed, mass extracted, derivatization volume, and injection volume, yielding concentrations in µg/g dry weight (DW) (sorbitol equivalents) for most of the metabolites. Absolute concentrations (µg/g DW) were determined for six metabolites that differed the most between species that were subsequently tested in fungal inhibition assays. Response factors for each of these six metabolites relative to sorbitol were generated by using authentic standards. Metabolite identification was performed using the Wiley Registry/NIST Mass Spectral Library 2023, in addition to an extensive user-created library of

TMS-derivatized metabolites. Metabolite concentrations of *C. mollissima* trees and the F_1 hybrid trees were contrasted against *C. dentata* trees using *t*-tests (Supplementary Data 19).

Metabolite fungal inhibition assays

Six compounds present at higher concentrations in *C. mollissima* relative to *C. dentata* stem tissues were evaluated for antifungal activity. Commercially available standards of betulin (Sigma-Aldrich, product number B9757), erythrodiol (Sigma-Aldrich, product number 09258), lupeol (Sigma-Aldrich, product number L5632), homogentisic acid (Sigma-Aldrich, product number H0751), homogentisic acid-2-*O*-glucoside (i.e., 'phaseolodin', MedChemExpress, product number HY-N7400), and β-phenyllactic acid (Sigma-Aldrich, product number P-7251) were incorporated individually into potato dextrose agar (PDA-BD DIFCO™). We initially tested all metabolites for bioactivity at 0.1 mg/ml to determine if there was a response at this preliminary, low concentration. Homogentisic acid was additionally tested at 1 mg/mL to assess dose-dependent effects. Solvent-matched controls (DMSO, DMF, or ethanol) were prepared in parallel using the same concentrations employed to dissolve the metabolite standards. The ATCC 9414 *Cr. parasitica* strain was cultured on treated and control media, and radial growth was measured over 10 days. Fungal growth was analyzed and measured using ImageJ version 1.54g (Java 1.8.0_345, 64-bit), with area calculated relative to a 1 cm² scale. Statistical analyses of growth area measurements were performed in GraphPad Prism (v. 10.5.0 (774) using an unpaired *t*-test to compare each treatment with its corresponding control (Supplementary Data 20).

Phenotyping and heritability estimation for Phytophthora root rot resistance

To assess root rot resistance, 26,288 progeny from 662 open-pollinated mother trees were inoculated with *P. cinnamomi* in yearly trials from 2005 to 2023 (*40*). The phenotyped backcross families descended from the 'Graves' F_1 (458 families), *C. mollissima* 'Mahogany' (36 families), *C. mollissima* 'Nanking' (43 families), and other sources of resistance (22 families). Descendants of the 'Clapper' BC_1 tree had little to no root rot resistance (*40*); therefore, progeny from this source of resistance were not included among the 662 families in subsequent analyses. Twenty-two susceptible *C. dentata* families and 14 resistant *C. mollissima* families were included in yearly trials as controls (Supplementary Data 21).

Seedlings were grown in soilless media (Fafard 3B mix) in D40 pots (Steuwe & Sons) in a completely randomized design in the 2014 through 2016 annual trials at North Carolina State University (Raleigh, NC) and the 2016 to 2023 trials in the U.S. Forest Service Resistance Screening Center (Asheville, NC). For trials conducted from 2005 through 2016 at Chestnut Return Farms (Seneca, SC), seedlings were planted in Rubbermaid 150-gallon tubs in a randomized complete block design. At 3 to 4 months of age, seedlings were inoculated with *P. cinnamomi* isolates grown on sterilized vermiculite moistened with V8 broth (*102*) or on sterilized rice grains (*103*). Seedlings were inoculated with different *P. cinnamomi* isolates are collected each year from locations where the surviving seedlings were planted to avoid moving isolates between locations.

The mother trees were genotyped rather than the individual progeny; therefore, quantitative genetic analyses were performed on progeny survival proportions at 16 to 24 weeks post-inoculation (Supplementary Data 22). Survival proportions were adjusted for the fixed effects of site, year, and isolate. Genomic heritability was estimated in ASReml-R 4.2 using an inverse genomic relationship matrix following the methods described for blight resistance. We estimated heritability within a total population of 484 genotyped and progeny-tested mothers and a subset of 434 mothers with 70% to 100% *C. dentata* ancestry. A likelihood-ratio test was used to determine whether the additive genetic variance component of progeny survival proportions was significantly greater than zero.

Quantitative trait locus mapping

We detected QTL underlying the blight resistance index and root survival phenotypes in backcross populations by associating variation in these traits with ancestry calls for *C. mollissima* versus *C. dentata* from Ancestry_HMM (Supplementary Data 23). Ancestry calls were pruned to 641 loci with a minimum recombination frequency between markers of 0.001 using the 'dropSimilarMarkers' function from 'qtlTools' (*104*). A genetic map for the ancestry calls was estimated with the 'est.map' R/QTL function and Kosambi distances, assuming chromosomes were linkage groups and markers were ordered by physical position in the 'Ellis-1' *C. dentata* genome (Supplementary Data 24).

Blight resistance QTL were mapped separately for 1,324 backcross descendants of the 'Clapper' BC_1 tree and 1,275 descendants of the 'Graves' F_1 tree. Root rot survival QTL were mapped for 395 backcross (primarily BC_2F_2) descendants of 'Graves'. To map QTL specific to the 'Clapper' source of resistance, we extracted all of the haplotype blocks supported by ≥ 100 markers that were homozygous for *C. dentata* ancestry in the Clapper BC_1 founder and excluded backcross descendants that had *C. mollissima* ancestry in these regions. Graves is an F_1 hybrid, so there were no regions where all backcross descendants were expected to be homozygous for *C. dentata* ancestry. To map QTL for 'Graves' descendants, we filtered out open-pollinated BC_2F_2 progeny with parents that descended from other sources of resistance based on the paternity analysis. We also removed individuals with >50% *C. mollissima* ancestry that resulted from open pollination with *C. mollissima* fathers.

We used the 'scan1' function from the R package 'qtl2' (*105*) to estimate LOD scores for marker-trait associations after controlling for polygenic kinship effects with the leave one chromosome out method (*106*). There was a low frequency of homozygous *C. mollissima* ancestry calls (3.7%) in backcross populations. Therefore, to maximize power to detect QTLs, CC calls were converted to 'AC' and QTL analysis was performed specifying a 'backcross' population type. Significance thresholds for QTL were estimated with the 'scan1perm' function and 1,000 permutations. Peak and border markers for QTL intervals were obtained with the 'find_peaks' and 'lod_int' functions specifying a peak drop of two LOD units. Genetic variance in resistance indices explained by the QTL intervals was estimated with $R^2 = 1 - 10^{-2/N * LOD}$, where N is the population size and LOD is the score for the peak marker (*107*). The effect of inheriting a *C. mollissima* allele at the QTL, in resistance index units, was estimated with the 'scan1coef' function (Supplementary Data 25). The genomic extent of the blight and root rot resistance QTL intervals mapped in this study were compared those from previous studies in *C. dentata* backcross populations (*41*, *108*) by blasting flanking sequences from peak and border markers against 'Ellis' *C .dentata* and both haplotypes from the 'Mahogany' and 'Nanking' *C. mollissima* genomes in Phytozome (Supplementary Data 26).

Genome-wide association studies

To increase power to detect loci significantly associated with blight resistance, a GWAS was performed with the blight resistance index phenotypes of 3,365 *Cr. parasitica*-inoculated trees with 40% to 100% *C. dentata* ancestry and 65,128 SNPs with minor allele frequencies > 0.05. The GWAS was conducted on a combined population consisting of hybrid descendants of 'Clapper' and 'Graves' (n = 2,864, including progeny of intercrosses between Clapper and Graves descendants with other resistance sources), and 27 additional *C. mollissima* sources (n = 361). To detect loci associated with root rot resistance, GWAS was performed on survival proportions among the progeny of 475 backcross trees, including 395 descendants of Graves and 80 descendants from seven other *C. mollissima* sources.

Genome-wide association studies were performed using the 'BLINK' algorithm (*109*) implemented in the 'GAPIT' v.3 R package (*110*). 'BLINK' iteratively fits the most significantly associated markers in multiple-locus fixed-effect models. The most parsimonious model with the most significant explanatory power was selected based on the Bayesian Information Criterion. To control for population structure, between 2 and 10 principal components (PCs) estimated from SNP data were fit as covariates. To identify the optimal number of PCs to control population structure (blight n PCs = 8, root rot n PCs = 2), the distribution of observed versus expected P-values was visualized using QQ-plots. Furthermore, potential inflation of P-values due to population structure was estimated using the genomic inflation factor (λGC). Models with PCs resulting in λGC 0.99-1.1 were deemed to control for population structure (111) sufficiently. Markers with Benjamini-Hochberg FDR-adjusted $P < 0.05$ for association were considered significant (Supplementary Data 27). To estimate the proportion of phenotypic variance explained by the significant markers, R^2 values were obtained from a linear model with genotypes as fixed effects (*i.e.,* the 'lm' function in R). To map to extent the intervals likely to contain causal variants, linkage disequilibrium was estimated for markers within three megabases of the associated markers with the 'LD.decay' function in 'Sommer' (*112*), and the distance at which average r^2 decayed to < 0.2 was calculated with the 'loess' function in R.

Accuracy of genomic selection

Five-fold cross validation was performed to evaluate accuracy of genomic selection with varying marker densities among inoculated individuals with > 70% *C. dentata* ancestry (n blight = 3,143, n root rot = 434, n height = 956) and among inoculated + naturally infected trees (blight n = 3,288, root rot survival n = 434, and tree height n = 1,028). Genomic prediction is most accurate when the training population is closely related to the prediction population (*113*, *114*); therefore, related trees were randomly partitioned into training and validation populations to assess genomic prediction accuracy in our pedigree-structured populations. Phenotypes were masked for 20% of the genotyped and phenotyped individuals. Genomic estimated breeding values (GEBVs) were predicted from genomic

relationships with remaining phenotyped individuals. This procedure was repeated five times until GEBVs were obtained for the entire training population. Predictive ability was estimated as the Pearson correlation between GEBVs and phenotypes after accounting for fixed factors ($r_{gebv,\,yadj}$). Genomic prediction accuracy—the theoretical correlation between GEBVs and true breeding values ($r_{gebv,\,tbv}$) —was estimated by dividing predictive ability by the square root of trait heritability (*115*). To assess the minimum marker density required for accurate genomic prediction, we pruned the GBS dataset in snpRelate (*116*) to lower densities based on increasingly stringent D' thresholds within 1000 marker windows (*i.e.*, D' > 0.95 = 25,676 SNPs, 0.9 = 10,522 SNPs, 0.8 = 5,835 SNPs, 0.6 = 3,353 SNPs, 0.4 = 1,087 SNPs, 0.2 = 328 SNPs). To assess the degree to which marker data improved prediction accuracy, we also evaluated prediction accuracy using pedigree relationships alone.

Genomic selection of parents for the next generation

To select parents for the next generation of crosses, we used single-step genomic best linear unbiased prediction (ssGBLUP) (*117–119*). With the ssGBLUP method, we estimated breeding values for both genotyped and non-genotyped individuals from a blend of pedigree and genomic relationships. Pedigree (**A**) and genomic relationships (**G**) were blended into a hybrid relationship matrix (**H**) after scaling the mean diagonal and off-diagonal elements of **G** to those of **A** using the method of (*120*). To minimize bias in breeding value estimation, five-fold cross-validation was performed for each trait using scaling factors τ for $\mathbf{G}^{-1}$ and ω for $\mathbf{A}^{-1}$, varying from 0 to 1 (*121*). For each trait, we selected the combination of scaling factors where the mean slope (b) of the regression of genomic estimated breeding values (x) on phenotypes (y) was 1 ± 0.01 (< 1% bias) across five replicates of cross validation (*i.e.*, $\tau = 1$ for all traits; $\omega = 0.8$ for blight resistance, $\omega = 0.7$ for height growth, and $\omega = 0.6$ for root rot resistance). For individuals genotyped but not phenotyped, breeding values were predicted from blended genomic and pedigree relationships using trees that had been phenotyped. Breeding values for the blight resistance index were estimated for a total population of 4,219 inoculated trees, including 3,499 genotyped and phenotyped individuals and 720 phenotyped but not genotyped individuals. Breeding values for tree height were estimated for 3,745 trees, including 1,406 genotyped and phenotyped, 26 phenotyped only, and 2,313 genotyped only. Breeding values for root rot survival proportions were calculated for 1,307 trees using a model that included 505 genotyped mothers with progeny phenotypes, 157 non-genotyped mothers with progeny phenotypes, and 645 individuals with genotype but no progeny phenotypes. To minimize potential for spurious results, genomic prediction for root rot resistance was only attempted for mothers that descended from the primary sources of resistance, as the mothers whose progeny had been inoculated with *P. cinnamoni* (i.e., 'Graves', 'Mahogany', 'Nanking', and 'Musick'). Breeding values for root rot survival were scaled from 0 = *C. dentata* mean to 100 = *C. mollissima* mean with the same method used for scaling the blight resistance index (Supplementary Data 28).

Predicting gains from genomic selection

Gains in the blight resistance index and root rot survival rates were simulated for one generation of controlled crosses, followed by genomic selection of 20% of progeny with the highest predicted performance. Two *C. dentata* backcross hybrid breeding tracks were simulated where biparental combinations with 70% to 85% average *C. dentata* ancestry were selected to (1) maximize blight resistance and height growth (*i.e.,* 30 crosses where midparent blight resistance index was > 70 and height growth > 70th percentile) or (2) maximize blight resistance, root rot, and height growth, simultaneously (*i.e.,* 20 crosses where midparent blight resistance index was > 60, root rot survival index > 40, and height growth > 70th percentile). Crosses were simulated among trees from the central breeding region (*i.e.*, trees from Virginia, Maryland, and Kentucky). To minimize inbreeding, only crosses where co-ancestry coefficients between parents did not exceed 0.0625 were considered.

To estimate variation in potential breeding outcomes, simulations were replicated three times for crosses among inoculated trees and once for inoculated + naturally infected trees. Crosses were randomly sampled from longer lists of 1,404 potential crosses that met thresholds for the blight + height breeding track and 349 potential crosses that met thresholds for the blight + root rot + height breeding track. Genotypes of 100 progeny per cross were simulated with the 'makeCross' function in 'AlphaSimR' (*122*) using phased parental genotypes from ~93,332 biallelic SNPs and their interpolated genetic map positions. Genomewide *C. dentata* v. *C. mollissima* ancestry proportions of the simulated progeny were estimated with Ancestry_HMM (*55*, *123*). Genomic estimated breeding values (GEBVs) for the blight resistance index, root rot survival, and height growth were predicted for the simulated progeny with GBLUP in ASReml-R v. 4.2. Training populations consisted of trees genotyped and phenotyped for each trait (inoculated trees: $n = 3{,}499$ for blight resistance, $n = 484$ for root rot resistance, and $n = 1{,}152$ for height growth; inoculated + naturally infected trees: $n = 3{,}885$ for blight resistance, $n = 484$ for root rot resistance, and $n = 1{,}355$ for height growth). For each breeding track, GEBVs for each trait were weighted equally in a selection index. The top

20% of progeny with > 70% *C. dentata* ancestry and the highest selection index values were selected. To predict realized gains from selection, true breeding values were simulated as variables that were correlated GEBVs proportional to genomic prediction accuracies (*i.e.*, blight resistance $r_{gebv,bv} = 0.608$, root rot resistance $r_{gebv,bv} = 0.617$, and height growth $r_{gebv,bv} = 0.574$; Table S6). Supplementary Data 29 contains genomic estimated breeding values and ancestry estimates across simulation replicates.

SUPPLEMENTARY FIGURES AND TABLES

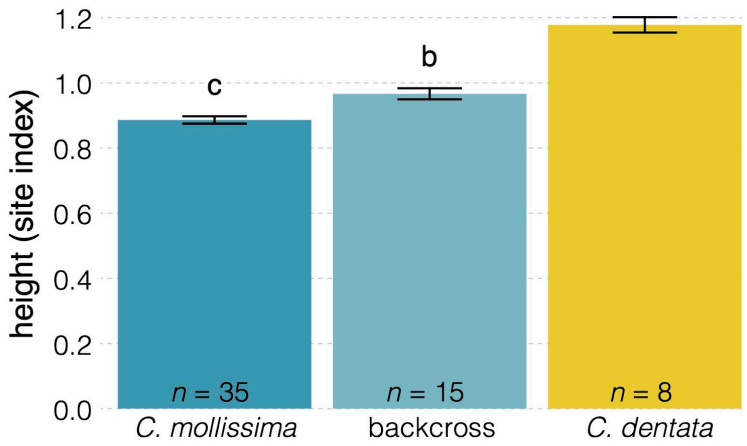

Fig. S1. Comparison of tree heights for *C. mollissima*, *C. dentata* backcross, and large surviving *C. dentata* (LSACs) at a site in Virginia (Lesesne State Forest), where chestnut blight severity has been reduced by the introduction of hypovirulent strains of *Cr. parasitica*. Age-adjusted tree heights were compared for the 30% fastest growing *C. dentata*, *C. dentata* BC_1 (backcross) hybrids, and *C. mollissima* individuals at this site. Trees varied in age from 22 to 33 years old. To account for age-based differences in tree height, observed tree heights were divided by age-based expectations for yellow poplar (*Liriodendron tulipifera*) at site index = 90 (*47*).

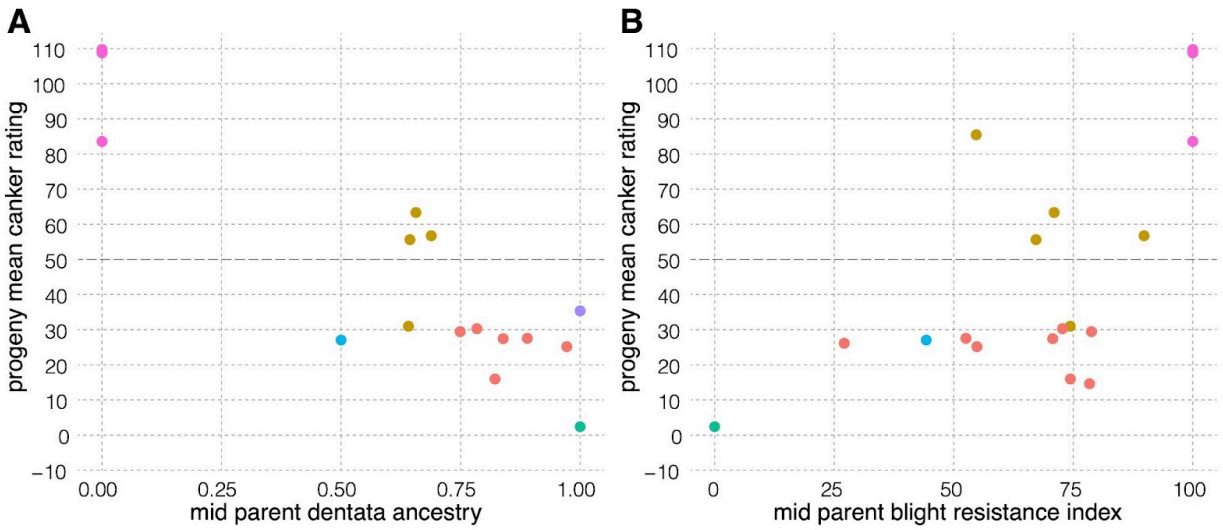

Fig. S2. Correlations between progeny mean blight resistance and parental *C. dentata* ancestry (A) or blight resistance (B). Colors indicate different cross types (pink - *C. mollissima*, light blue - *C. mollissima* × *C. dentata* F_1, green = *C. dentata,* purple - large surviving *C. dentata*, red - *C. dentata* BC_3F_2, brown - *C. dentata* backcross × F_1). Only the subset of families where both parents were genotyped (A, 16 families) or phenotyped (B, 18 families) is shown. A summary of resistance and ancestry data for the entire population of 30 families is presented in Table S2. The dotted line is the desired resistance threshold (>50) for the first generation of selection.

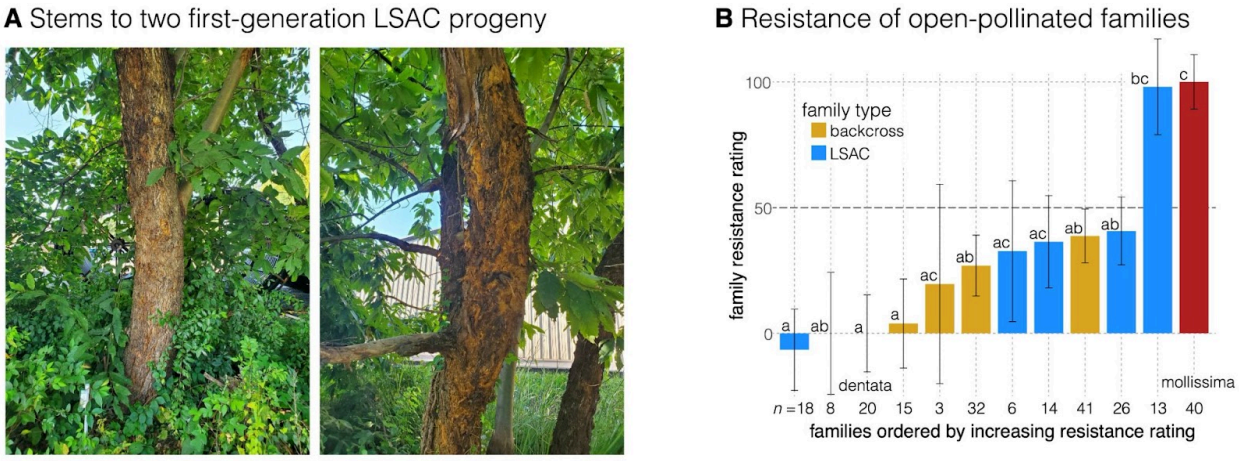

Figure S3: Heritable variation in blight resistance among progeny of large surviving American chestnuts (LSACs). (A) Stems of the first two generations of LSAC progeny planted at the Scrivener Orchard in Maryland, with swollen and callused cankers indicative of partial blight resistance. (B) Family mean blight resistance of six second-generation, open-pollinated (op) LSAC families from the Scrivener orchard. Resistance of LSACs seedlings was compared to susceptible *C. dentata* (*i.e.*, 'mixed op dentata'), resistant *C. mollissima* (*i.e.*, mixed op mollissima), and three *C. dentata* backcross families where the parents had 15% to 25% *C. mollissima* ancestry. Family mean resistance ratings were scaled from 0 (mean of susceptible *C. dentata*) to 100 (mean of resistant *C. mollissma*). The number of individuals phenotyped within each family is indicated below the bars. Pairwise differences between family mean resistance ratings were assessed via Tukey tests. Letters indicate differences significant at $P < 0.05$. Blight resistance of an additional LSAC × LSAC family from controlled pollination is reported in Table S2.

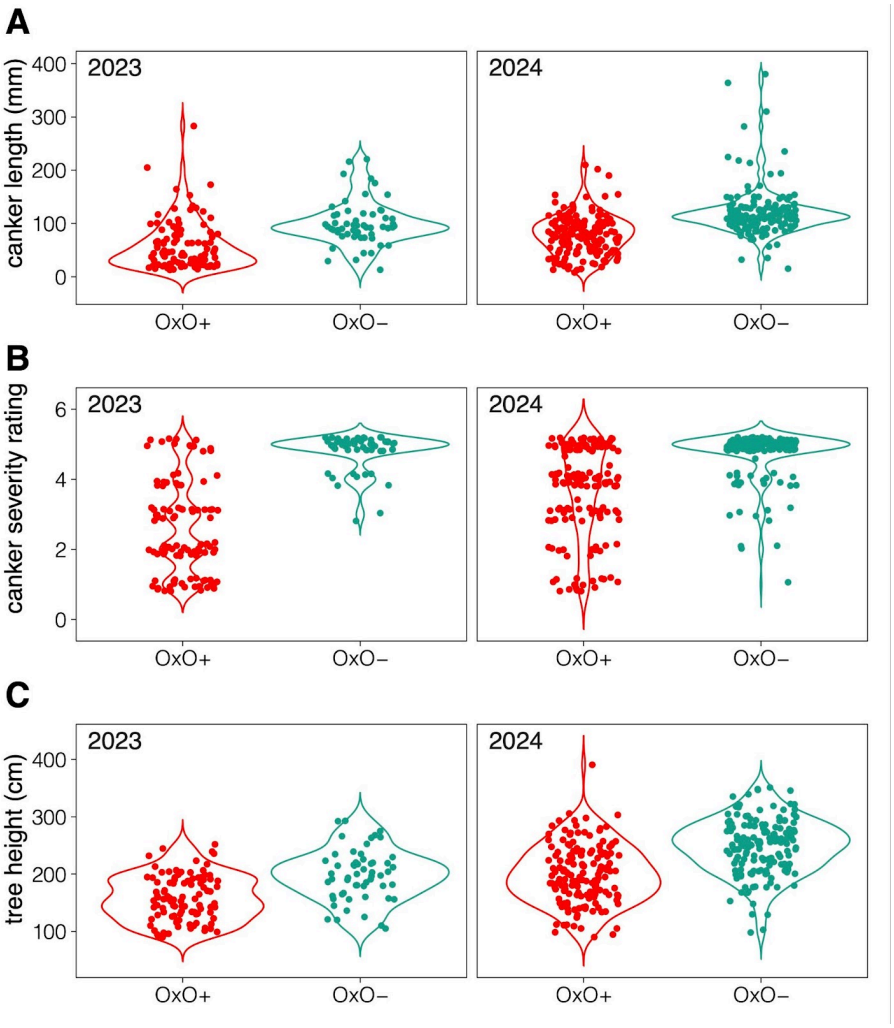

Fig. S4. Canker length (A), canker rating (B), and tree height (C) comparisons among 3-year-old progeny of 'Darling 54' that inherited (OxO+) or did not inherit (OxO-) the oxalate oxidase construct. Two three-year-old 'Darling 54' progeny cohorts were inoculated in 2023 (n = 159) and 2024 (n = 319).

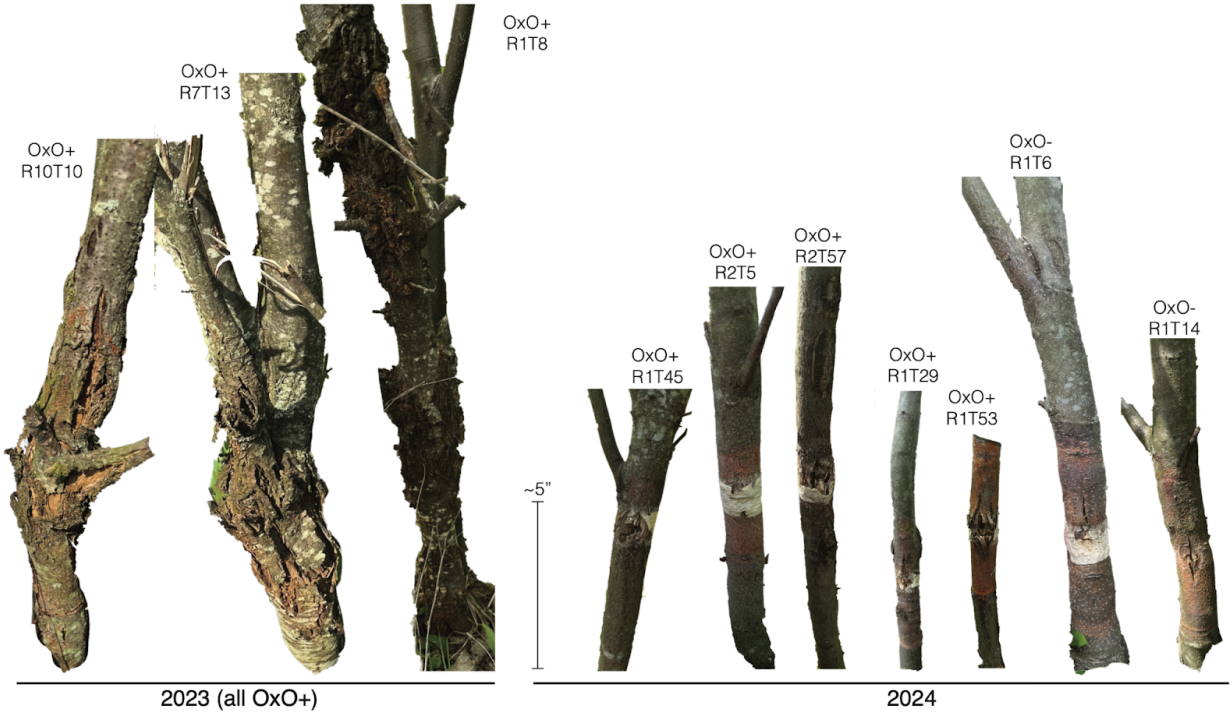

Fig. S5. Chestnut blight cankers on ‘Darling 54’ progeny. Photographs of the stem inoculation site of 10 representative trees. Backgrounds have been removed to better show the cankers via Adobe Express ‘delete’ functionality. The three 2023 trees ‘Darling 54’ OxO+ progeny were selected as having the highest blight resistance from the 2023 inoculations, two years after inoculation. The left five 2024 trees are ‘Darling 54’ OxO+ from 2024 inoculations, highlighting the variation in canker responses. The right two show representative canker response on an OxO- progeny for comparison (Photos: Lake Graboski).

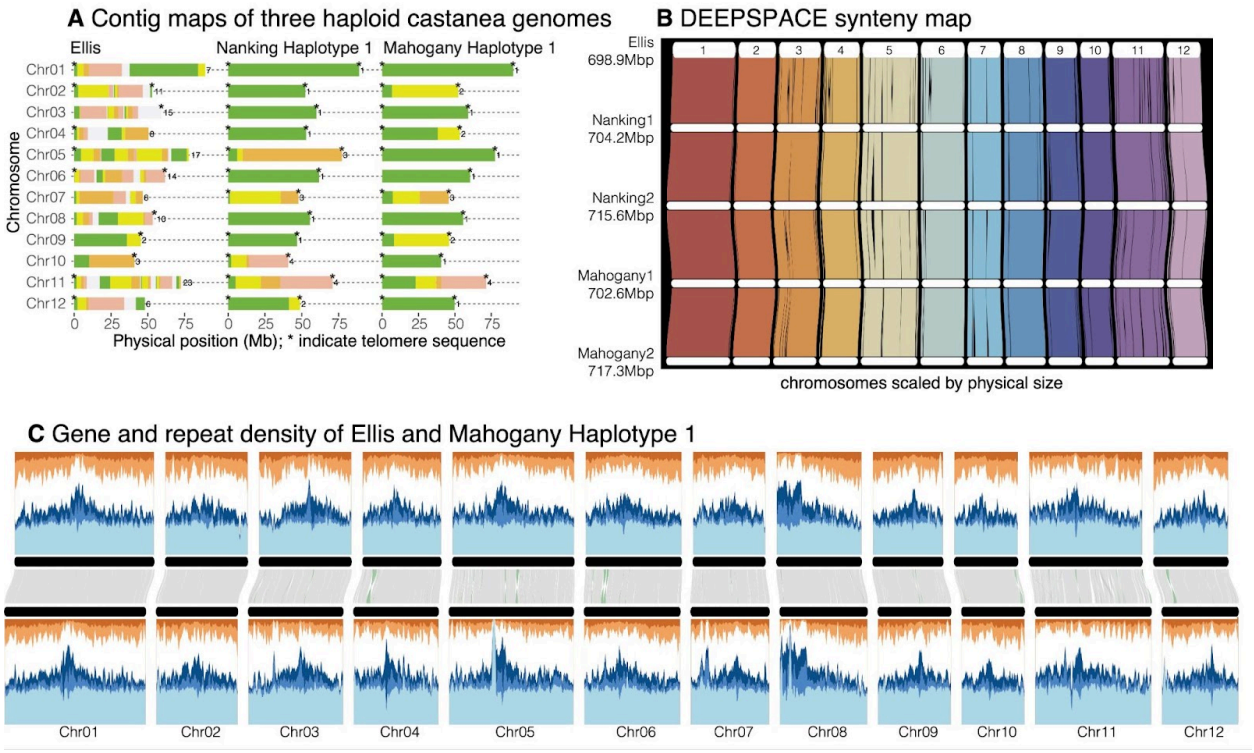

Fig. S6. Genome structure and assembly for three *Castanea* genotypes. (A) Contig maps, as produced by GENESPACE, showing the positions of contig breaks (gaps = where colors change) and telomeres (*) for three of the five genome assemblies. **(B)** DEEPSPACE riparian plot showing syntenic positions across all five assemblies. **(C)** Gene, repeat, and unannotated (white) sliding window densities in 1Mb 900-kb overlapping windows are connected by the syntenic links as were used to generate panel B; inverted syntenic blocks are highlighted in green.

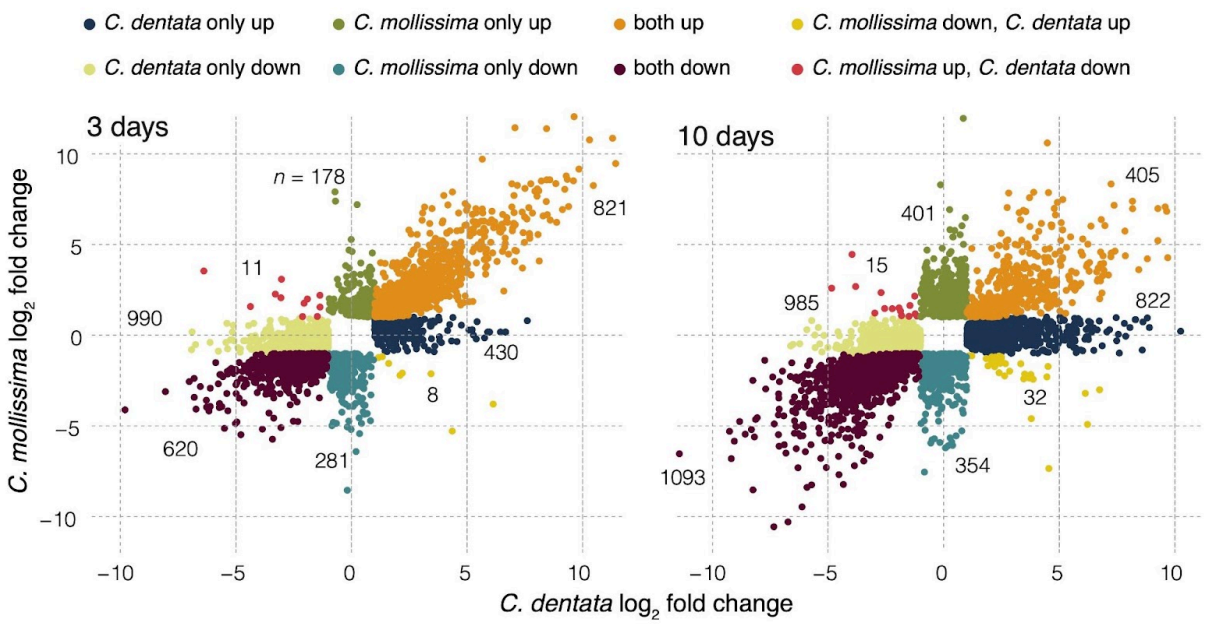

Fig. S7. Comparisons of gene expression responses of *C. mollissima* and *C. dentata* to the chestnut blight fungus at 3 and 10 days post-inoculation. The $\log_2$ fold change in expression for the inoculation treatment relative to the wound-only control is compared across species for putative single-copy orthologs. Numbers indicate the number of differentially expressed genes in each expression category.

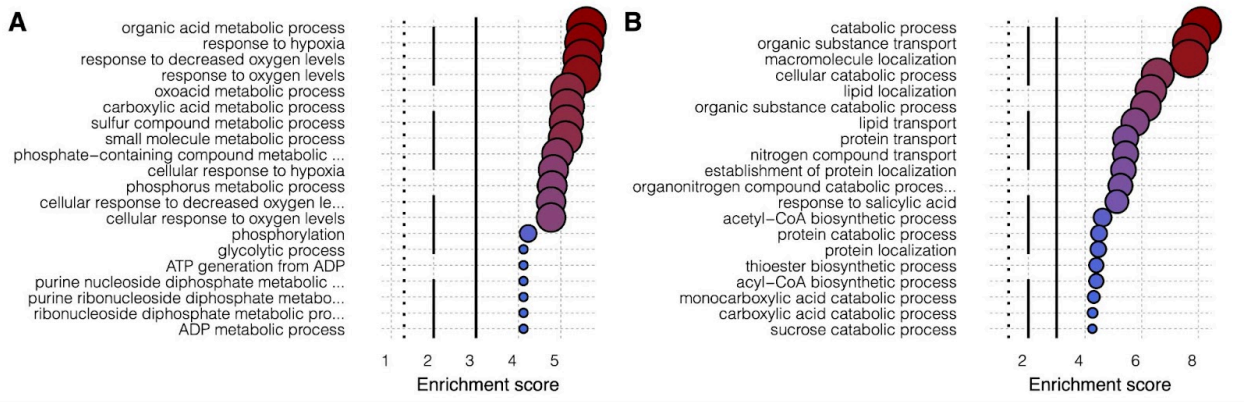

Fig. S8. Overrepresentation of gene ontology (GO) terms among genes specifically upregulated in stems of the susceptible *C. dentata* host at 3 (A) or 10 (B) days post inoculation (dpi) with the chestnut blight fungus. Fisher's exact tests were performed to test for an excess of genes with specific GO terms, and the enrichment score is the $-\log_{10}$ P-value.

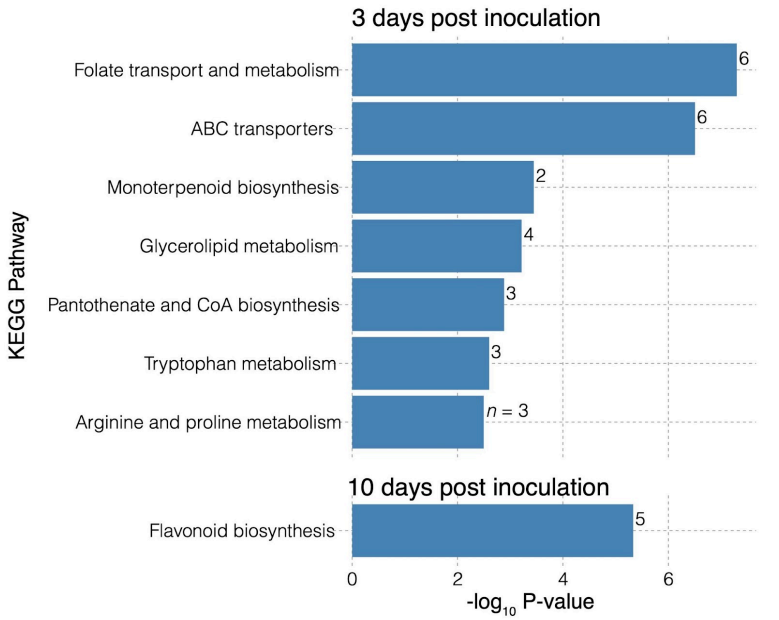

Fig. S9. Overrepresentation of KEGG pathways among genes specifically upregulated in stems of the resistant *C. mollissima* host 3 or 10 days post inoculation (dpi) with the chestnut blight fungus. Fisher's exact tests were used to assess an excess of genes in specific KEGG pathways. Pathways significant at a Benjamini Hochberg FDR-adjusted $P < 0.05$ are shown. A complete list of KEGG pathways represented in the species-specific and common *C. mollissima* v. *C. dentata* responses to chestnut blight fungal inoculation is provided in Supplementary Data 18.

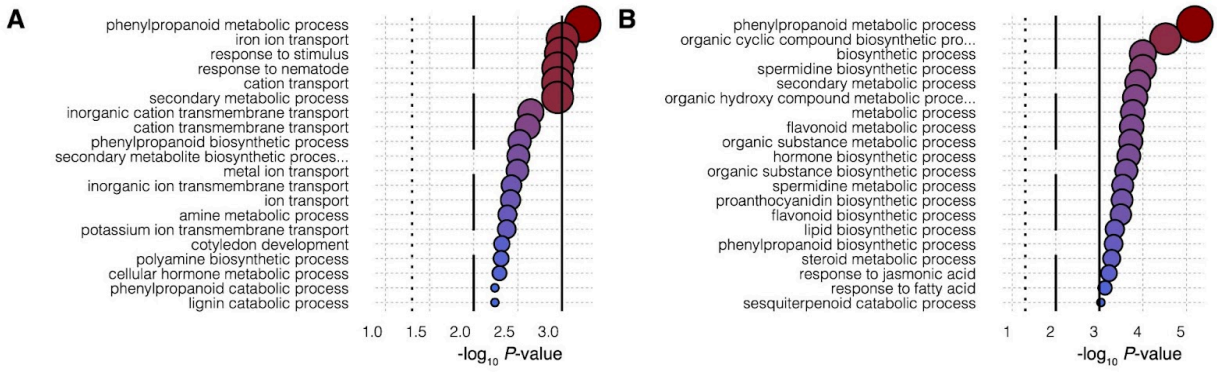

Fig. S10. Overrepresentation of gene ontology (GO) terms among genes specifically upregulated in stems of the resistant C. mollissima host at 3 (A) or 10 (B) days post inoculation (dpi) with the chestnut blight fungus. Fisher's exact tests were performed to test for an excess of genes with specific GO terms, and the enrichment score is the $-\log_{10}$ P-value.

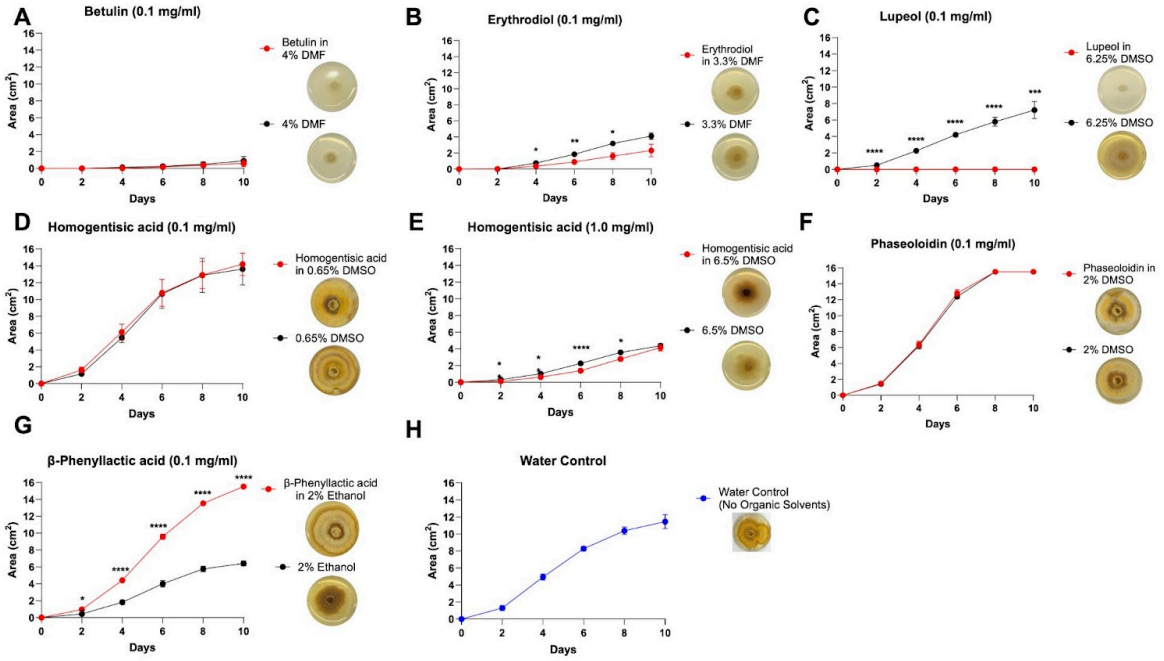

Figure S11. Growth of *Cr. parasitica* with plant metabolite treatments dissolved in organic solvents compared to their respective solvent controls (a-g). The water control shows typical fungal growth on PDA (h). Error bars represent the standard error of the mean. The data represent four biological replicates.

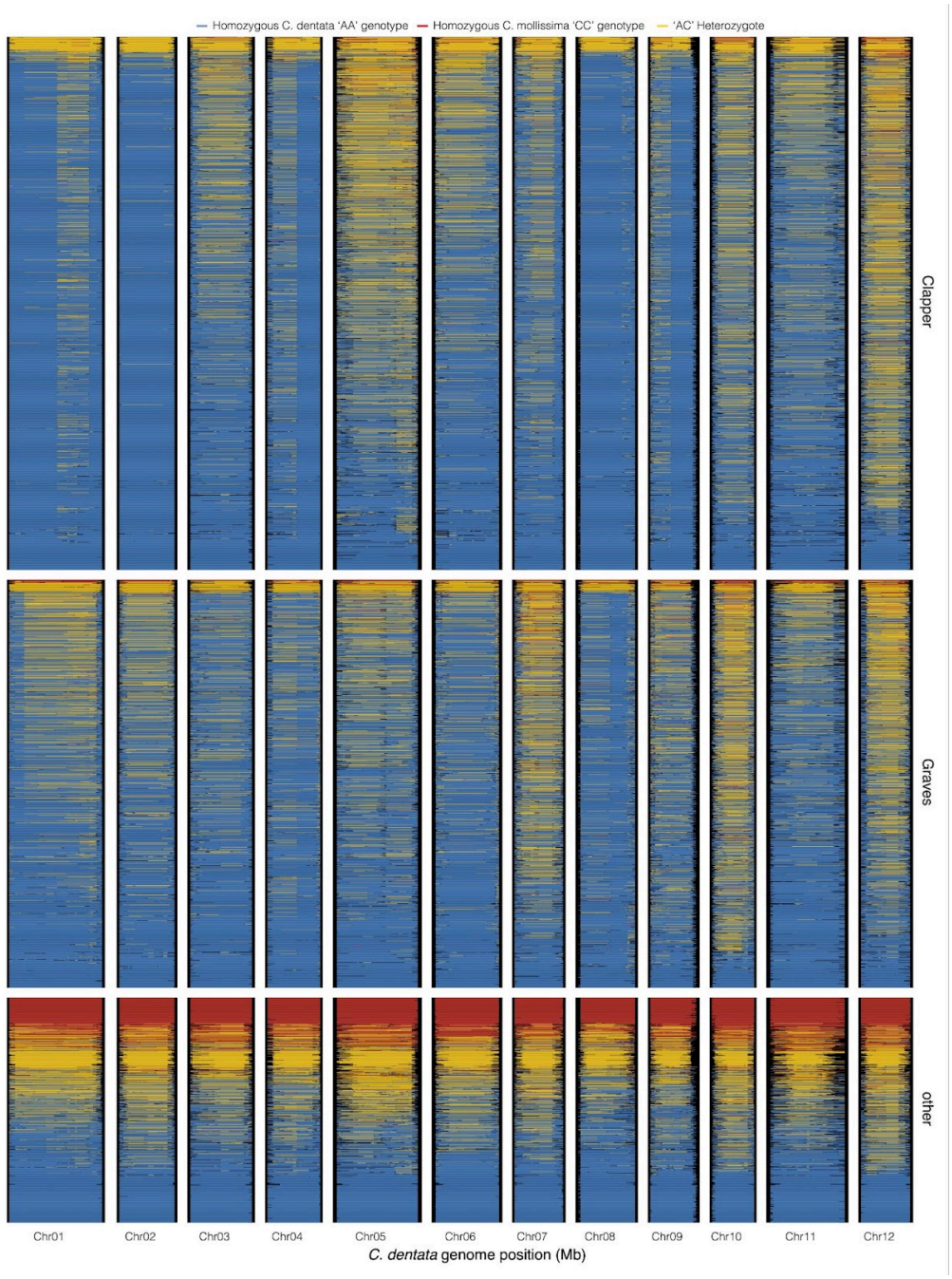

Fig. S12. Local species ancestry among *C. dentata* backcross hybrids. Local ancestry calls were generated for descendants of the Clapper BC_1 and Graves F_1 trees, and for other sources of resistance in Ancestry_HMM. Numbers of trees from each population are depicted on the y-axis. Genomic regions where trees inherited both alleles from *C. dentata* or *C. mollissima* are described as dark blue and red regions, respectively, and regions of heterozygous ancestry are shown as yellow regions. Ancestry-based genotype calls at 641 loci were used for ancestry-based QTL mapping for blight resistance among 1,324 primarily BC_3F_2 descendants of the 'Clapper' BC_1 and 1,275 BC_2F_2 descendants of the 'Graves' F_1 tree.

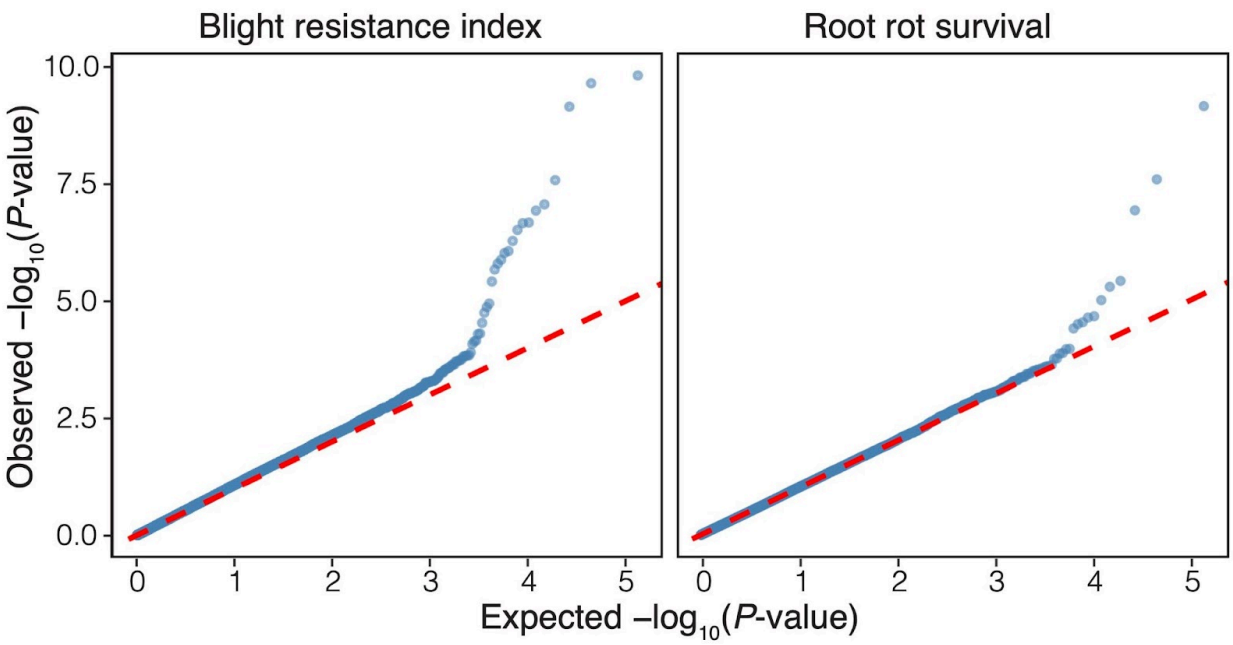

Fig. S13. Quantile-quantile plots of observed versus expected P-values from genome-wide association studies with the blight resistance index and root rot survival. Genomic inflation factors ($\lambda GC\lambda$) for P-values were 1.1 for blight resistance and 0.99 for root resistance, indicating that population structure in *C. dentata* backcross populations did not strongly inflate marker effects.

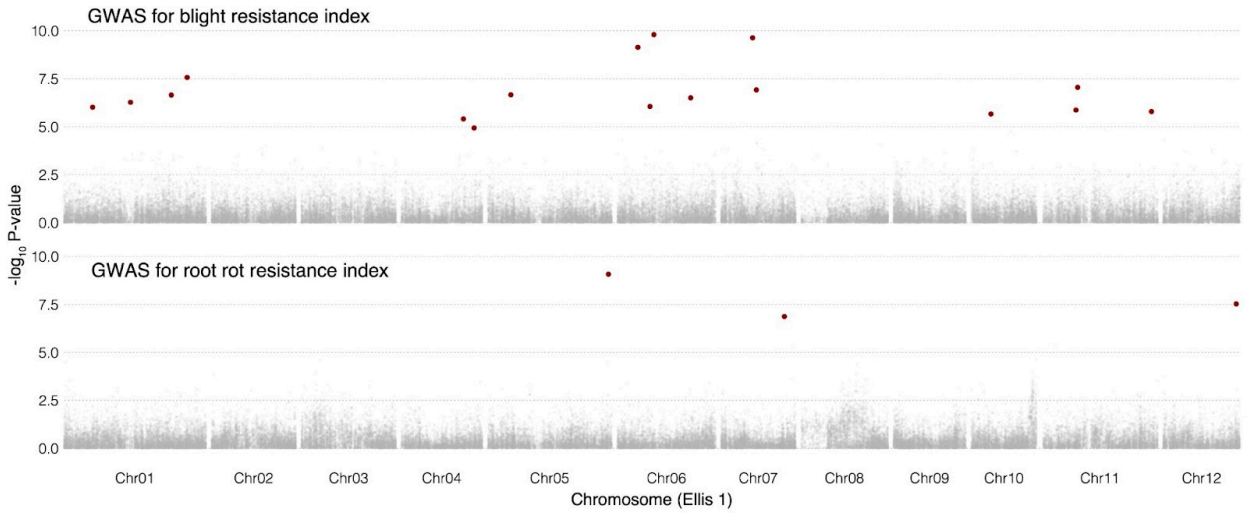

Fig. S14. Genome-wide associations with blight resistance index and root rot survival in *C. dentata* backcross populations. The locations of markers with Benjamini-Hochberg FDR-adjusted P-values < 0.05 (horizontal dashed line) are depicted with red vertical dotted lines. GWAS analyses were performed using blight resistance index phenotypes from 3,365 *C. dentata* backcross hybrids derived from 29 *C. mollissima* sources of resistance. A GWAS for root rot resistance was performed on *P. cinnamomi* inoculation survival rates among open-pollinated progeny of 475 trees descended from nine *C. mollissima* trees.

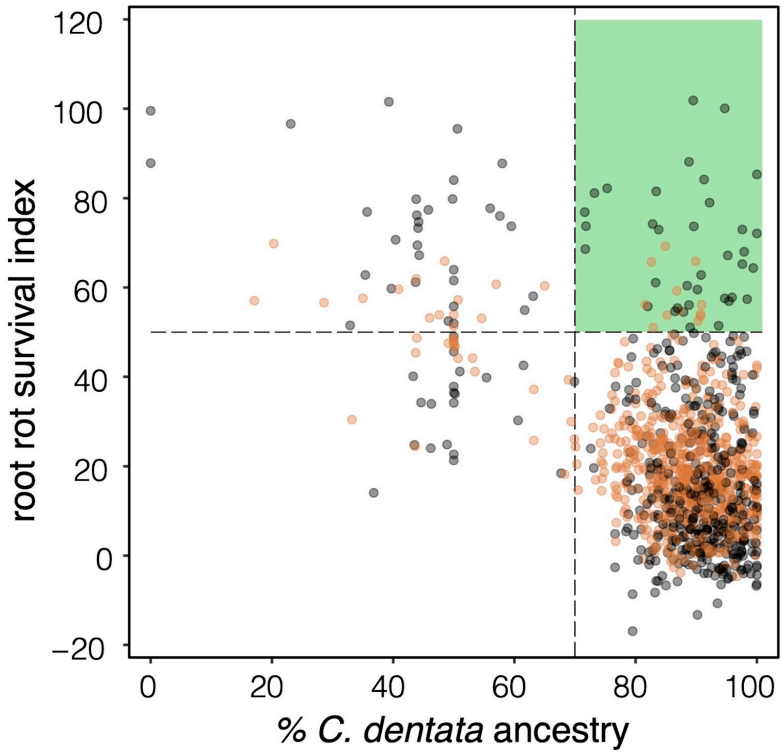

Fig. S15. Root rot survival breeding values versus American chestnut (*C. dentata*) ancestry. Black dots represent breeding values for 484 mothers estimated from survival rates of open-pollinated progeny inoculated with *P. cinnamomi.* Orange dots are breeding values for 645 mothers without progeny-test data, predicted using genomic relationships with phenotyped mothers.

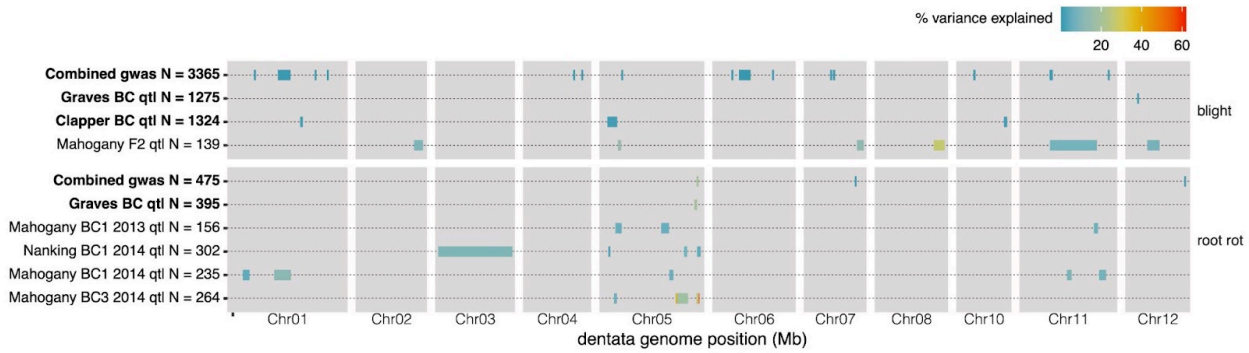

Fig. S16. Comparisons of marker-trait associations from quantitative trait loci (QTL) and genome-wide association study (GWAS) between populations and studies. Populations used in the current study are bolded. Previous blight resistance QTL are from (*108*), and previous root rot QTL are from (*41*, *108*).

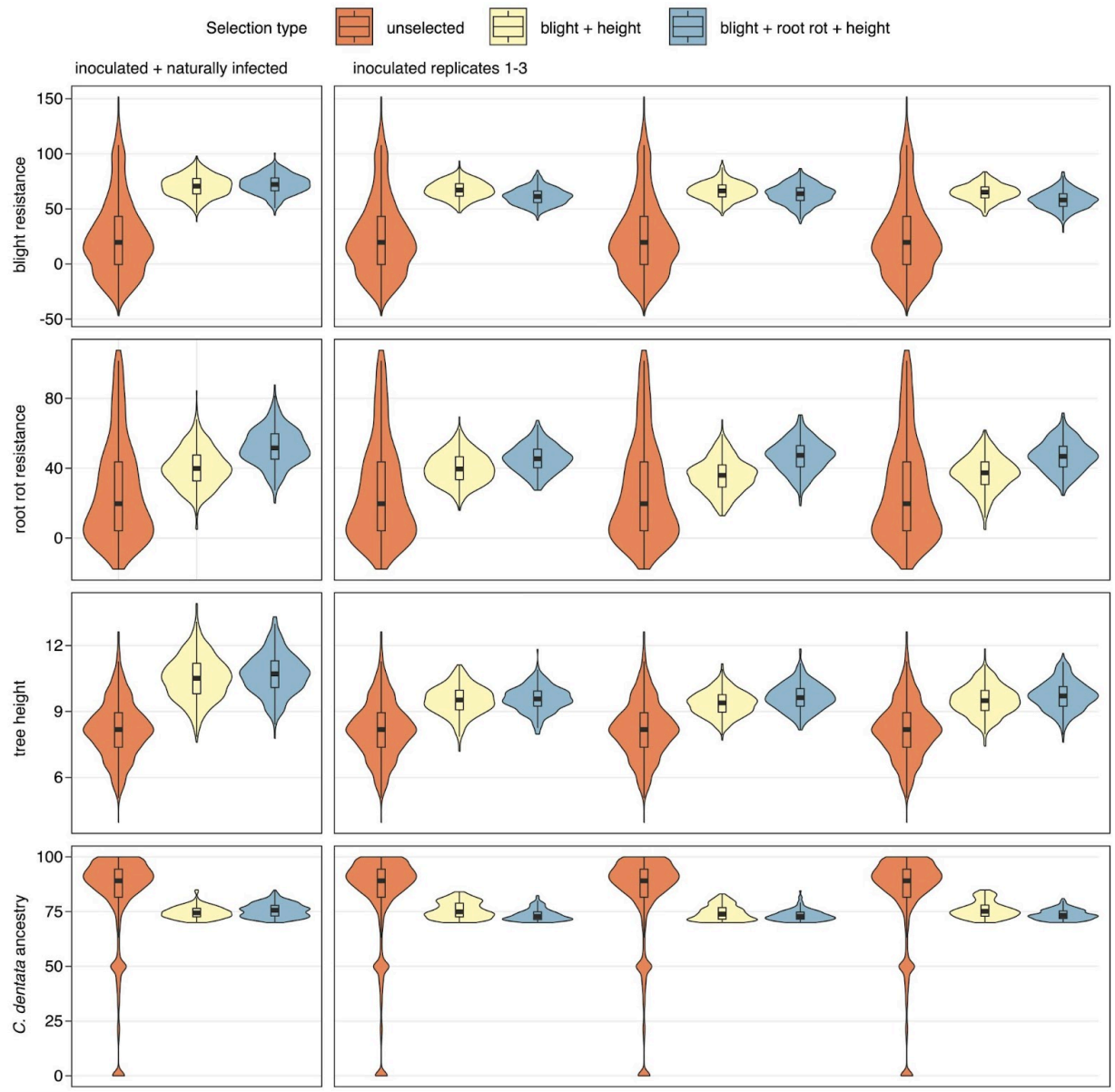

Fig. S17. Predicted genetic gains across four simulation replicates of controlled pollination between selected trees followed by genomic selection of 20% progeny with the highest predicted performance. Two breeding tracks were simulated where (1) 30 random crosses were performed among parents that maximize blight resistance (>70) and height growth (70th percentile) or (2) 20 random crosses were performed among parents that combine above average blight resistance (>60), root rot resistance (>40), and height growth (>70th percentile). Genomic prediction models were trained on a combination of inoculated and naturally infected trees (one replicate, n = 3,885) or only inoculated trees (three replicates, n = 3,499). 'Unselected' depicts genetic variation in the current population of trees.

Table S1. Heritability (h^2) estimates for blight resistance, root rot survival, and tree height growth. Narrow-sense heritability was estimated from genomic relationships among trees. For blight and height growth, heritability estimates were compared for trees artificially inoculated with *Cr. parasitica* versus a larger population of trees inoculated and naturally infected. Heritability estimates are also reported for the subset of trees with ≥ 70% *C. dentata* ancestry. Heritability was estimated as $\sigma_a^2/(\sigma_a^2 + \sigma_e^2)$, where σ_a^2 is the additive genetic variance and σ_e^2 is the error variance estimate. Numbers (n) are sample sizes of genotyped and phenotyped trees used to estimate h^2 for each trait. Heritabilities for individual traits included in the blight resistance index for inoculated trees are reported below the blight resistance index. Likelihood-ratio tests (LRTs) were performed to test the null hypothesis of zero additive genetic variance.

Trait	n	$h^2 \pm SE$	σ_a^2	σ_e^2	$\sigma_a^2 = 0$, LRT P
Blight resistance index (inoculated)	3499	0.41 ± 0.03	665.7	928.7	< 0.0001
(≥ 70% *C. dentata,* inoculated)	3143	0.38 ± 0.03	566.7	906.3	< 0.0001
Blight resistance index (inoc. + natural infection)	3885	0.42 ± 0.03	726.5	986.2	< 0.0001
(≥ 70% *C. dentata,* inoc. + natural infection)	3288	0.41 ± 0.03	641.6	936.7	< 0.0001
Main stem alive	3499	0.20 ± 0.03	0.85	3.29	
Large cankers	3469	0.24 ± 0.06	1.05	3.29	
Exposed wood	3469	0.20 ± 0.03	0.83	3.29	
Sporulation	3495	0.20 ± 0.03	0.87	3.29	
Blight contained	813	0.23 ± 0.06	0.96	3.29	
Stump sprouts	1300	0.23 ± 0.05	0.96	3.29	
Sunken/swollen cankers	3458	0.32 ± 0.03	0.20	0.41	
Tree height (inoculated)	1214	0.62 ± 0.05	3.03	1.83	< 0.0001
(≥ 70% *C. dentata,* inoculated)	956	0.62 ± 0.06	2.92	1.75	< 0.0001
Tree height (inoc. + natural infection)	1355	0.79 ± 0.04	5.40	1.46	< 0.0001
(≥ 70% *C. dentata,* inoc. + natural infection)	1028	0.79 ± 0.04	6.10	1.57	< 0.0001
Root rot progeny survival (all)	484	0.62 ± 0.08	0.013	0.008	< 0.0001
≥ 70% *C. dentata*	434	0.63 ± 0.09	0.013	0.008	< 0.0001

Table S2. Variation in average blight resistance of progeny from controlled and open-pollinated (op) crosses among parents with varying *C. dentata* ancestry and blight resistance estimates. Blight resistance ratings for seedling progeny were scaled from 0 = mean of susceptible *C. dentata* controls to 100 = mean of resistant *C. mollissima* controls. Tukey tests were performed for comparisons of family means, and significant differences at $P < 0.05$ are indicated by different letters. Progeny blight resistance was compared to estimates of parental blight resistance index and *C. dentata* ancestry proportions.

cross	n	type	source	family mean blight resistance	se	tukey	mom blight resistance	mom dentata	dad blight resistance	dad dentata	midparent blight resistance	midparent dentata
GAFL1XopCH x op	42	mollissima	GAFL1	109.7	10.0	a	-	0.00	-	0.00	100	0.00
GAML001-C5 x op	12	mollissima	GAML001-C5	108.8	18.9	ab	-	0.00	-	0.00	100	0.00
AL-TVANW-22-18 x CC-WW01-67-06	20	BC x F1	Clapper, Nanking	85.4	14.5	ac	93.3	-	16.1	0.50	54.7	-
PA-McCe-m x op	25	mollissima	PA-McCe-m	83.6	13.2	ac	-	0.00	-	0.00	100	0.00
D2-10-3 x MD-Tho-6-01	49	BC x F1	Clapper, Musick	63.3	9.3	acd	89	0.48	53	0.84	71	0.66
TN-RC09-3-9 x CCSP-3-50	36	BC x F1	NC10, Clapper	56.7	10.8	ace	91.2	0.87	88.4	0.51	89.8	0.69
TN-RC09-6-46 x CCSP-3-50	41	BC x F1	NC10, Clapper	55.6	10.1	bce	46	0.78	88.4	0.51	67.2	0.64
VA-LCPT-LCU4/2-263 x op	42	BC x op	Clapper	54.2	10.0	bce	84.7	-	-	-	-	-
VA-LCPT-LCU4/2-368 x op	34	BC x op	Clapper	40.5	11.1	bcef	-	-	-	-	-	-
D7-3-14 x op	32	BC x op	Clapper	37.7	11.5	bceg	51.6	0.84	-	-	-	-
VA-LSFB2-R71C2 x PA-OrYo	37	LSAC x LSAC	Ort, RaggedMtn	35.3	10.7	bceg	80.1	1.00	-	-	-	1.00
TN-RC09-6-46 x TN-TTU05-A29	48	BC x F1	Gideon, Clapper	31.0	9.4	bceg	46	0.78	102.7	0.50	74.4	0.64
D6-27-65 x MD-Tho-111	30	BC x BC	Clapper, Musick	30.3	11.9	bceg	73.4	0.87	72.1	0.70	72.8	0.78
D1-2-4 x MD-Tho-111	32	BC x BC	Clapper, Musick	29.5	11.5	bceg	85.5	0.80	72.1	0.70	78.8	0.75
D2-10-3 x op	41	BC x op	Clapper	29.4	10.2	bceg	89	0.48	-	-	-	-
D8-18-4 x D6-27-65	40	BC x BC	Clapper	27.5	10.3	cg	31.7	0.91	73.4	0.87	52.55	0.89
GABE001-297 x TN-RC09-2-22	39	BC x BC	Graves, Clapper	27.4	10.4	cg	73.2	0.81	68.2	0.87	70.7	0.84
TN10 x SS44	11	F1	CD	27.0	19.6	aceg	102	0.00	-13.5	1.00	44.3	0.50
ME-BK13-29 x ME-RW11-65	47	BC x BC	Graves	26.1	9.5	cg	13.4	-	40.8	0.99	27.1	-
ME-BK13-20 x ME-RW11-91	41	BC x BC	Graves	25.2	10.2	cg	48.5	0.94	61.2	1.00	54.9	0.97
D8-18-4 x op	18	BC x op	Clapper	24.6	15.3	bceg	31.7	0.91	-	-	-	-
MD-11-232 x op	27	BC x op	Musick	19.4	12.5	cg	35.5	0.79	-	-	-	-
MD-11-321 x op	25	BC x op	Musick	18.1	13.0	cg	24.7	0.77	-	-	-	-
GABE001-58 x GABE001-297	27	BC x BC	Graves	16.0	12.5	cg	75.6	0.84	73.2	0.81	74.4	0.82
ME-BK13-51 x ME-RW11-44	38	BC x BC	Graves	14.6	10.5	deg	55.7	0.91	101.1	-	78.4	-
MD-11-129 x op	42	BC x op	Musick	4.0	10.0	eg	44.7	0.86	-	-	-	-
VT-HRC-35 x op	39	dentata	NA	2.4	10.5	eg	-	-	-	-	0	1.00
TN-RC09-1-24 x TN-TTU05-A34	45	BC x F1	Nanking, Gideon	-10.1	9.7	fg	58.7	0.93	99.8	0.50	79.3	0.71
D1-2-4 x op	14	BC x op	Clapper	-10.5	17.4	deg	85.5	0.80	-	-	-	-
VA-LCPT-LCU4/2-231 x op	41	BC x op	Clapper	-19.1	10.1	g	80.5	-	-	-	-	-

Table S3. Orthogroups (og) with copy number expansion and significantly greater induced expression responses in resistant *C. mollissima* as compared with susceptible *C. dentata*. Standardized expression, measured as transcripts per million (tpm), was summed across genes within orthogroups. The top 20 of 128 orthogroups with functional annotations and significantly greater total expression in *C. mollissima* versus *C. dentata* are shown (*t*-test Benjamini-Hochberg FDR-adjusted $P < 0.05$). For a complete list of 128 orthogroups, see Supplementary Data 17.

og	mollissima copy n	dentata copy n	dentata tpm 3d	mollissima tpm 3d	fdr 3d	dentata tpm 10d	mollissima tpm 10d	fdr 10d	Arabidopsis annotation
19839	2	1	0.1	1543	0.001	0.1	3288	0	drought-induced 21
18446	2	1	1105	2277	0.009	1619	3122	0.139	Histone superfamily
9412	2.25	1	313	2268	0.029	150	735	0.005	beta-1,3-glucanase 1
17925	5.5	0	0	1347	0.01	0	444	0	elicitor-activated gene 3-2
17994	2	1	39	700	0	33	1051	0	conserved peptide upstream open reading frame 5
16264	2	0	0	387	0.007	0	807	0.002	RmlC-like cupins superfamily
18406	1	0	0	648	0.008	0	506	0.001	hydroxyproline-rich glycoprotein
18407	2.25	1	105	1028	0.011	746	782	0.961	hydroxyproline-rich glycoprotein
15927	2	1	13	531	0	13	422	0	Tetratricopeptide repeat (TPR)-like
19006	7	5	104	379	0.034	199	739	0.028	GAST1 homolog
20122	2	1	19	404	0.001	26	413	0	Cellulase (glycosyl hydrolase family 5)
18248	2	1	36	329	0	45	329	0	extra-large G-protein 1
17236	3	2	0	291	0.001	0	285	0	PF13668 - Ferritin-like domain
18260	2	1	7	413	0.035	6	164	0.011	Uncharacterised protein family (UPF0497)
16310	2	1	36	254	0	80	330	0.019	Protein of unknown function (DUF1644)
17849	2	1	101	242	0.016	40	294	0.001	alpha/beta-Hydrolases
18414	2	1	0.2	207	0.007	3	167	0.001	hydroxyproline-rich glycoprotein
16267	2	1	68	235	0	55	236	0.001	PF10291 - Muniscin C-terminal mu homology domain
15933	2	1	4	204	0	6	116	0	DNAJ heat shock N-terminal domain

Table S4. Metabolites are constitutively present at greater concentrations (µg/g dry weight) in the stems of resistant *C. mollissima* (*n* = 8) relative to susceptible *C. dentata* (*n* = 11) at Benjamini-Hochberg FDR adjusted P (P.adj) < 0.05. Differences with (*C. mollissima* x *C. dentata)* × *C. dentata* BC_1 hybrids (*n* = 11) are also included. The concentrations of the six metabolites that differ the most between species were determined with authentic standards, whereas concentrations of all other metabolites are reported as sorbitol equivalents. Metabolites in italics are unknowns designated by their retention times (min) and key mass-to-charge ratios.

Metabolite	*C. dentata* mean ± sem	*C. mollissima* mean ± sem	Cmol v. Cden fold change	P.adj	F1 mean ± sem	F1 v. Cden fold change	P.adj
homogentisic acid-2-O-glucoside	1.7 ± 0.2	386.2 ± 109.1	227.17	0.005	295.5 ± 115.7	173.82	0.133
lupeol	12.2 ± 1.8	273.4 ± 70.7	22.41	0	91.0 ± 46.6	7.46	0.283
betulin	6.6 ± 0.8	144.6 ± 38.5	21.91	0.005	69.5 ± 27.1	10.53	0.168
homogentisic acid	0.1 ± 0	1.8 ± 0.5	18.00	0.005	0.9 ± 0.3	9.00	0.109
erythrodiol	0.6 ± 0.1	10.8 ± 3.8	18.00	0.021	5.7 ± 2.6	9.50	0.207
B-phenyllactic acid	2.9 ± 1.9	26.4 ± 7.2	9.10	0.009	18.7 ± 6.6	6.45	0.168
muco-inositol	333.2 ± 72.7	1099.4 ± 167.5	3.3	0	770 ± 203	2.31	0.207
shikimic acid	125.9 ± 11	403.5 ± 71.1	3.2	0	445.4 ± 191	3.54	0.283
ursolic acid	8.4 ± 0.9	23.9 ± 3.1	2.83	0	13.2 ± 2.6	1.57	0.266
Gallyl alcohol glucoside	72 ± 15.8	203.1 ± 28.1	2.82	0	177 ± 36.9	2.46	0.131
16.23 267 355 369	18.5 ± 3.9	50.3 ± 4.6	2.72	0	50.9 ± 14.1	2.76	0.181
15.57 457 575 487	115.2 ± 10	291.9 ± 72.4	2.53	0.034	238 ± 66.2	2.07	0.266
10.99 334 256	15.1 ± 2.7	36.1 ± 8.1	2.39	0.038	32.6 ± 5.7	2.15	0.109
10.99 285 375 300	2 ± 0.3	4.2 ± 0.9	2.15	0.036	3.9 ± 0.6	1.97	0.109
11.05 267 356 400	158.9 ± 25.6	340.9 ± 76	2.15	0.049	343.5 ± 78.5	2.16	0.181
homogentisaldehyde	80.3 ± 11.1	168.5 ± 18	2.1	0	139.3 ± 21	1.74	0.137
14.48 356 341 283 glycoside	6.5 ± 0.7	13 ± 1	1.99	0	13 ± 1.3	1.99	0
9.87 356 253 268	6.7 ± 0.6	13.1 ± 1.7	1.95	0.005	13.5 ± 2.2	2.02	0.109
coumaroyl-4-O-quinic acid	0.7 ± 0.1	1.3 ± 0.1	1.9	0.013	1.1 ± 0.2	1.6	0.25
9.96 143 272 156	70.4 ± 10.4	133.5 ± 9.1	1.9	0	112.2 ± 12.1	1.59	0.131
11.50 363 273 183	5.9 ± 0.8	10.8 ± 1.9	1.84	0.043	8 ± 1.6	1.36	0.475
homogentisaldehyde glucoside	41.7 ± 6.5	74.7 ± 11	1.79	0.039	69.4 ± 6.9	1.66	0.109
ethyl-phosphate	118 ± 7.2	205.9 ± 19.7	1.74	0	160 ± 12.9	1.36	0.109
galactinol	28.3 ± 3.5	48 ± 7.4	1.7	0.044	47.9 ± 6.8	1.69	0.133
15.71 297 guaiacyl lignan	1279.8 ± 157.7	1960.1 ± 154.4	1.53	0.028	1450.1 ± 201.7	1.13	0.74
11.93 389 491	186 ± 16.8	275 ± 23.7	1.48	0.021	213.9 ± 23.4	1.15	0.583
1,6-anhydroglucose	148 ± 10.5	214.7 ± 16.8	1.45	0.013	201.7 ± 14.9	1.36	0.109
gallic acid	392.3 ± 21.7	557.3 ± 23.2	1.42	0	513.3 ± 52.7	1.31	0.204
10.51 306 288 217	13.6 ± 0.9	19 ± 2.1	1.4	0.044	24.6 ± 6.6	1.81	0.283
10.91 450 dehydro-sugar	10.5 ± 0.6	13.1 ± 0.9	1.25	0.049	14.7 ± 1.4	1.4	0.109
myo-inositol	1521.4 ± 95.7	969.7 ± 86.8	0.64	0.005	1123.5 ± 136.5	0.74	0.159
glucose-6-phosphate	6.2 ± 0.6	3.9 ± 0.4	0.63	0.039	4.7 ± 0.7	0.76	0.35
phenylalanine	4 ± 0.4	2.3 ± 0.3	0.58	0.02	3.8 ± 0.4	0.95	0.898
12.49 539 359 204	38.8 ± 2.8	22.6 ± 2.1	0.58	0	34.8 ± 2.8	0.9	0.553
malic acid	1104.8 ± 26.6	615.2 ± 40	0.56	0	798.8 ± 64.1	0.72	0
erythronic acid	3.4 ± 0.4	1.8 ± 0.2	0.54	0.032	2.9 ± 0.4	0.86	0.655
glutamic acid	69.8 ± 8.7	35.6 ± 6.7	0.51	0.032	83.5 ± 18.9	1.2	0.74
threonic acid	33.9 ± 5	11.6 ± 1.6	0.34	0.009	24.1 ± 4.5	0.71	0.359
16.10 456 369 327	1.4 ± 0.3	0.4 ± 0.1	0.3	0.021	0.9 ± 0.2	0.61	0.266
salicylic acid	25.9 ± 4.1	6.2 ± 0.6	0.24	0.005	28.3 ± 8.7	1.09	0.912
salicylic acid-2-O-glucoside	71.8 ± 13.7	11.7 ± 0.8	0.16	0.009	71.5 ± 27.4	1	1

Table S5. Comparisons of *Cr. parastica* growth on agar supplemented with metabolites that are present at higher concentrations in *C. mollissima* bark as compared with *C. dentata*. Unpaired *t*-tests were performed at each time point between the treatment groups and their corresponding solvent controls. Significance is denoted as: ****$P < 0.0001$; ***$P < 0.001$; **$P < 0.01$; *p < 0.05; ns = not significant; NA = not available due to identical values between treatment and control. Figure S11 shows comparisons of fungal growth at each time point.

	Day 2		Day 4		Day 6		Day 8		Day 10	
Treatment	*P*	sig.	*P*	sig.	*P*	sig.	*P*	sig.	*P*	sig.
Betulin (0.1 mg/ml)	NA	NA	0.2567	ns	0.6432	ns	0.7998	ns	0.6708	ns
Erythrodiol (0.1 mg/ml)	0.4284	ns	0.021	*	0.0023	**	0.0114	*	0.0845	ns
Lupeol (0.1 mg/ml)	<0.0001	****	<0.0001	****	<0.0001	****	<0.0001	****	0.0004	***
Homogentisic acid (1.0 mg/ml)	0.003	**	0.0017	**	<0.0001	****	0.0167	*	0.661	ns
Homogentisic acid (0.1 mg/ml)	0.2555	ns	0.5712	ns	0.9581	ns	0.9894	ns	0.8132	ns
β-Phenyllactic acid (0.1 mg/ml)	0.0048	**	<0.0001	****	<0.0001	****	<0.0001	****	<0.0001	****
Phaseoloidim (0.1 mg/ml)	0.5671	NA	0.8309	ns	0.3962	ns	NA	NA	NA	NA

Table S6. Genomic prediction accuracy with decreasing numbers of markers for *C. dentata* hybrids with > 70% *C. dentata* ancestry. Five-fold cross-validation was used to estimate the mean and standard deviation (SD) of predictive ability (correlation between genomic estimated breeding values and phenotypes) and prediction accuracy (correlation between genomic estimated breeding values and true breeding values). Prediction accuracy was calculated by dividing the predictive ability by the square root of each trait's heritability. To test the effect of marker density on genomic prediction, the complete set of 93,332 SNPs was pruned to lower densities by removing markers with decreasing linkage disequilibrium within 1000-marker windows. The accuracy of genomic prediction was compared with that of prediction from pedigree relationships alone. Prediction accuracies were compared for inoculated trees (blight n = 3,143, root rot survival n = 434, and tree height n = 956) versus a larger population of trees that were inoculated and naturally infected (blight n = 3,288, root rot survival n = 434, and tree height n = 1,028).

Trait	N markers	Predictive ability ± SD (inoculated)	Prediction accuracy ± SD (inoculated)	Predictive ability ± SD (inoc. + naturally infected)	Prediction accuracy ± SD (inoc. + naturally infected)
Blight resistance index	25676	0.377 ± 0.028	0.608 ± 0.046	0.388 ± 0.026	0.657 ± 0.044
	10522	0.374 ± 0.027	0.604 ± 0.044	0.387 ± 0.027	0.655 ± 0.047
	5835	0.369 ± 0.024	0.594 ± 0.039	0.379 ± 0.027	0.641 ± 0.047
	3353	0.341 ± 0.033	0.549 ± 0.053	0.362 ± 0.031	0.613 ± 0.053
	1087	0.329 ± 0.027	0.531 ± 0.043	0.333 ± 0.030	0.564 ± 0.051
	328	0.271 ± 0.057	0.437 ± 0.091	0.269 ± 0.022	0.456 ± 0.037
	0 (pedigree)	0.314 ± 0.049	0.506 ± 0.079	0.314 ± 0.032	0.532 ± 0.055
Root rot survival	25676	0.485 ± 0.059	0.617 ± 0.075	-	-
	10522	0.488 ± 0.065	0.62 ± 0.082	-	-
	5835	0.477 ± 0.063	0.606 ± 0.08	-	-
	3353	0.455 ± 0.041	0.578 ± 0.052	-	-
	1087	0.391 ± 0.09	0.497 ± 0.114	-	-
	328	0.399 ± 0.059	0.507 ± 0.075	-	-
	0 (pedigree)	0.416 ± 0.058	0.529 ± 0.073	-	-
Tree height	25676	0.465 ± 0.057	0.574 ± 0.071	0.455 ± 0.021	0.559 ± 0.025
	10522	0.451 ± 0.055	0.556 ± 0.068	0.450 ± 0.035	0.553 ± 0.04
	5835	0.465 ± 0.057	0.574 ± 0.07	0.471 ± 0.026	0.579 ± 0.032
	3353	0.423 ± 0.065	0.522 ± 0.08	0.454 ± 0.023	0.558 ± 0.027
	1087	0.313 ± 0.081	0.386 ± 0.099	0.339 ± 0.021	0.417 ± 0.026
	328	0.275 ± 0.114	0.339 ± 0.14	0.250 ± 0.035	0.307 ± 0.043
	0 (pedigree)	0.434 ± 0.092	0.536 ± 0.113	0.432 ± 0.036	0.532 ± 0.044

DESCRIPTION OF SUPPLEMENTARY DATA

Accessible on Dryad: https://doi.org/10.5061/dryad.4xgxd25mj

Supplementary Data 1 | Blight and growth phenotypes for 5.5k trees

Supplementary Data 2 | Genotyping data at ~93k SNPs from 5,003 trees in variant calling format (VCF)

Supplementary Data 3 | Hybrid ancestry estimates for ~5k genotyped trees

Supplementary Data 4 | Genetic maps for the 'GMBig x Horn' *C. dentata* full sib family

Supplementary Data 5 | Genome ('Ellis-1') coordinates for *C. dentata* and *C. mollissima* ancestry in hybrids

Supplementary Data 6 | Pedigree for trees phenotyped in this study

Supplementary Data 7 | Seedling blight resistance data for ~1k backcross progeny

Supplementary Data 8 | Seedling blight resistance data for LSAC progeny

Supplementary Data 9 | Blight resistance and height data for ~500 'Darling 54' progeny

Supplementary Data 10 | A bed file with orthogroup and tandem array information for each gene

Supplementary Data 11 | Pan-genome sets for C. dentata and C. mollissima genomes in 'Ellis' coordinates

Supplementary Data 12 | Syntenic block breakpoints

Supplementary Data 13 | Chestnut blight timecourse RNA-seq metadata

Supplementary Data 14 | Chestnut blight timecourse RNA-seq count matrix for *C. dentata*

Supplementary Data 15 | Chestnut blight RNA-seq count matrix for *C. mollissima*

Supplementary Data 16 | Species- and allele-specific gene expression responses to *Cr. parasitica*.

Supplementary Data 17 | Blight-responsive orthogroups with higher copy number and expression upregulation in resistant *C. mollissima* v. susceptible *C. dentata*

Supplementary Data 18 | KEGG pathway enrichment among single-copy orthologs with species-specific and common expression responses to chestnut blight fungal inoculation

Supplementary Data 19 | Secondary metabolites present at higher concentrations in the stems of resistant *C. mollissima* v. susceptible *C. dentata*

Supplementary Data 20 | *Cryphonectria parasitica* fungal growth in the presence of metabolites that are at higher concentrations in resistant *C. mollissima* stems v. susceptible *C. dentata* stems

Supplementary Data 21 | Root rot resistance phenotypes for ~27 k open-pollinated progeny of *C. dentata* backcross hybrids, resistant *C. mollissima,* and susceptible *C. dentata* controls

Supplementary Data 22 | Family means for root resistance used for genetic analyses

Supplementary Data 23 | *C. dentata* v. *C. mollissima* ancestry calls used for QTL mapping

Supplementary Data 24 | Genetic map used for QTL mapping

Supplementary Data 25 | Effect of inheriting a *C. mollissima* allele on blight resistance and root rot survival across 641 loci used in QTL analysis

Supplementary Data 26 | Coordinates and effect sizes for blight and root rot QTL intervals and GWAS peaks

Supplementary Data 27 | *P*-values from genome-wide association studies of blight and root rot resistance

Supplementary Data 28 | Breeding values for blight and root rot resistance estimated with single-step GBLUP

Supplementary Data 29 | Genomic estimated breeding values for blight resistance, root rot resistance, and height growth from simulation of progeny genotypes