GBE

Comparative Genomics Provide Insight Into the Evolution of European *Aphanomyces euteiches* Strains

Carol Kälin [1,†], Edoardo Piombo [1,†], David Manyara [1], Sofia Dubusc[1], Mukesh Dubey [1], Magnus Karlsson [1,*]

[1]Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Uppsala, Sweden

[†]These authors contributed equally.

*Corresponding author: E-mail: magnus.karlsson@slu.se.

Accepted: March 03, 2026

Abstract

Aphanomyces euteiches is an oomycete pathogen causing root rot in pea and other legumes, leading to devastating yield losses in pea production. Previously, three genetic groups of *A. euteiches* were identified in Europe based on simple sequence repeat markers. In this study, we determined the genome sequences of 68 European *A. euteiches* strains and performed a comparative genomic analysis with the aim to investigate the population genetic structure of *A. euteiches*, to delineate species boundaries and to identify gene families evolving under selection for gene gains or losses. Population genetic analysis based on genome-wide single nucleotide polymorphisms identified three genetic groups in Europe, where strains from Italy formed a distinct and genetically isolated group. Genealogical concordance phylogenetic species recognition analysis indicated that strains from Italy may represent a cryptic species. *Aphanomyces* strains differed in virulence on susceptible and partially resistant pea host genotypes, but virulence did not correlate with genetic groups. Analysis of virulence-related gene family evolution revealed a significant expansion of gene families acting in the detoxification of plant-derived secondary metabolites in plant-pathogenic *Aphanomyces* species. Moreover, the carbohydrate esterase family 1 (CE1) was significantly expanded in *A. euteiches* but not in the cryptic species represented by Italian strains. Modular structure analyses revealed exclusive presence of feruloyl esterase domains in CE1 homologs of plant-pathogenic *Aphanomyces* species, indicating a role in plant cell wall degradation.

Key words: *Aphanomyces euteiches*, carbohydrate esterase, comparative genomics, pea, phylogenetic analysis, root rot.

Significance

European strains of the plant pathogen *Aphanomyces euteiches* have previously been reported to cluster in three genetic groups on a north-to-south gradient, with the most southern group being genetically distinct. Using population genomics, we confirmed the presence of three genetic groups in Europe with strains from Italy being genetically distinct. A comparative genomics approach found genetic evidence to view the southern group as a separate species and identified virulence-related gene families evolving differently in the *Aphanomyces* genus. Our results contribute to a better understanding of the genetic diversity and the genetic determinants underlying virulence in *A. euteiches,* which in turn is important for development and implementation of efficient disease control measures in agricultural production.

Introduction

Oomycete pathogens have gained increasing recognition as global threats to animal and plant welfare over the last years as they can negatively impact ecosystems and global food security (Erwin and Ribeiro 1996, Kamoun 2003). These filamentous microorganisms were previously assigned to the kingdom of fungi but differ in both morphology and physiology and are now classified as Stramenopiles, being closer related to brown algae and diatoms (Latijnhouwers et al. 2003, Derevnina et al. 2016). The most prominent genus of plant-pathogenic oomycetes, *Phytophthora,* comprises members that cause severe damage in crops and ecosystems (Erwin and Ribeiro 1996, Kamoun 2006). Among them, *Phytophthora infestans* is well known and studied as the causal agent of potato and tomato blight, which led to the Irish potato famine in the 1840s (Haas et al. 2009, Yoshida et al. 2013).

Although less known, the oomycete genus *Aphanomyces* comprises plant parasitic, animal parasitic and saprotrophic members (Diéguez-Uribeondo et al. 2009). *Aphanomyces astaci* and *Aphanomyces invadans*, causal agents of the crayfish plague and epizootic ulcerative syndrome in fish, respectively, pose a great threat to freshwater wildlife and aquaculture industry (Iberahim et al. 2018, Martín-Torrijos et al. 2021). Within the plant-pathogenic *Aphanomyces* species, *Aphanomyces cochlioides* causes seedling damping-off and root rot in sugar beet, spinach, cockscomb and other related species (Rossi et al. 2023), whereas *Aphanomyces euteiches* causes root rot in a broad range of legumes and is considered the major threat to pea production (Wu et al. 2018). Aphanomyces root rot (ARR) symptoms on pea include softening and browning of the roots, leading to subsequent reduction of root volume and function (Hughes and Grau 2013). The diploid *A. euteiches* has a homothallic (self-fertile) reproduction mode that allows for the formation of sexual oospores (Grünwald and Hoheisel 2006), while formation of sporangia and motile zoospores allows clonal reproduction. The highly resilient oospores can remain dormant in agricultural soils for more than ten years and serve as primary inoculum of a new infection cycle (Papavizas and Ayers 1974, Grünwald and Hoheisel 2006). ARR mitigation methods are limited where crop rotation and avoidance of highly infested fields are among the most effective measures. Pea cultivars with complete resistance to ARR are currently lacking but several genetic loci segregating with partial resistance to ARR have been identified and serve as a basis for ongoing breeding efforts (Pilet-Nayel et al. 2002, Pilet-Nayel et al. 2005, Hamon et al. 2011 , Hamon et al. 2013, Lavaud et al. 2015, Desgroux et al. 2016, Desgroux et al. 2018, Wu et al. 2021, Wu et al. 2022, Kälin et al. 2023). However, for developing long-term resistant pea cultivars, a better understanding of *A. euteiches* genetic diversity as well as virulence-related genes and their evolution is essential.

Recently, three genetically differentiated groups of *A. euteiches* were identified in Europe using simple sequence repeat (SSR) markers. The three groups included strains from six countries and clustered in a northeastern group including all strains from Finland and two from eastern Sweden, a central European group comprised of strains from Sweden, Norway, the United Kingdom, and France and a southern group consisting of five strains from Italy (Kälin et al. 2022). The Italian strains showed to be genetically very distinct from the other European *A. euteiches* strains, raising questions about its taxonomic status. Quillévéré-Hamard et al. (2018) reported a weak genetic differentiation between *A. euteiches* from Burgundy compared with the rest of northern France, including the presence of many heterozygous loci in the Burgundy population, based on SSR markers. This shows the potential of the primarily homothallic *A. euteiches* to occasionally outcross in agricultural fields, which was hypothesized to be promoted by a crop rotation including several different legume host species such as pea, common bean, vetch, and alfalfa. Furthermore, the Burgundy population of *A. euteiches* displayed a higher diversity in virulence on pea, compared with the *A. euteiches* population from northern France (Quillévéré-Hamard et al. 2018). Sexual recombination can increase the adaptability of an oomycete to its environment by increasing genetic diversity and thereby its potential to overcome plant resistance or develop fungicide resistance (Fry 2008, Brurberg et al. 2011).

Monitoring the genetic diversity of pathogen populations is therefore important for designing efficient and sustainable disease control measures. However, neutral markers such as SSRs are limited in their usefulness to link phenotypes with their genetic determinants. The study of virulence-related genes, however, can provide both phylogenetic information as well as insight into the mechanisms used by pathogens to infect and cause disease in plants. In plant-pathogenic fungi and oomycetes, this type of genes is often located in rapidly evolving genomic regions and follows a rapid birth and death type of evolution (Fouché et al. 2018). Oomycete examples of virulence factors include cellulose-binding proteins, small secreted effector proteins and DNA-damaging proteins, secreted by plant-pathogenic oomycetes during the early stages of infection (Gaulin et al. 2008, Ramirez-Garcés et al. 2016, Camborde et al. 2022, Kiselev et al. 2022). Another example is proteases secreted to the apoplast to inhibit host defenses

and degrade host components (Rocafort et al. 2020, Kiselev et al. 2023).

Comparative genomics can be used to increase our understanding of phylogenetic relationships between closely related taxa and to identify the genomic consequences resulting from different lifestyle strategies, such as plant parasitism, animal parasitism, and saprotrophism. In plant-pathogenic oomycetes, comparative studies on gene content revealed expansions of gene families encoding carbohydrate-active enzymes (CAZymes) targeting the plant cell wall and secreted effectors in *Phytophthora* species, while *Pythium* species exhibited an abundance of genes associated with signal transduction and proteolytic degradation (Adhikari et al. 2013). When it comes to *Aphanomyces*, the sequenced and annotated genomes of *A. euteiches* (Gaulin et al. 2018), *A. cochlioides* (Botkin et al. 2022), *A. invadans* (GenBank acc. no.: GCF_000520115.1), and *Aphanomyces stellatus* (https://www.polebio.lrsv.ups-tlse.fr/aphanoDB/) are available. Genomic analyses of the *A. euteiches* secretome found it to be adapted to host cell wall deconstruction and identified an abundance of small secreted proteins that are thought to facilitate plant infection (Gaulin et al. 2018, Camborde et al. 2022). The *A. cochlioides* genome entails putative RxLR and crinkler (CRN) effector proteins (Botkin et al. 2022), whereas none of these effector families were detected in *A. invadans* (Iberahim et al. 2018). Genomic studies focusing on virulence factors can contribute to the development of efficient plant protection strategies, including effector-guided identification of plant disease resistance genes (Vleeshouwers et al. 2008; Chepsergon et al. 2021). Pathogen virulence genes can also be directly targeted by small RNA-based host-induced gene silencing (Sanju et al. 2015), a technique that was applied on the *Avr3a* effector gene of *P. infestans*, achieving moderate resistance (Thakur et al. 2015).

In this study, we genome-sequenced 68 *A. euteiches* strains collected in different geographical areas in Europe. We hypothesized that (i) the genetic differentiation into groups observed with SSR markers can be confirmed with genome-wide genetic markers, (ii) that all strains represent the species *A. euteiches*, and (iii) virulence-related gene families display different evolutionary trajectories between *Aphanomyces* species. Our results confirmed the presence of three genetically distinct groups of *A. euteiches* in Europe. However, in contrast with our hypothesis, a genealogical concordance phylogenetic species recognition analysis suggested that Italian strains represent a hitherto undescribed species of *Aphanomyces*. Furthermore, we identified several significant expansions and contractions of virulence-related gene copy numbers between *Aphanomyces* species, providing evidence for host adaptation. Among them, a significant expansion of the carbohydrate esterase family 1 (CE1) in *A. euteiches* was specifically driven by higher feruloyl esterase paralog numbers.

Results

Genome Sequencing and Gene Content

Short-read Illumina genome sequencing of 68 *A. euteiches* strains resulted in between 10 and 45 million reads per sample, corresponding to a coverage between 26X and 117X for individual strains. The resulting assemblies had an N50 between 26 and 37 Kb, and total lengths between 49 and 58 Mb, with an average of 50.6 Mb. Differences in coverage did not negatively influence the total length of the assemblies (Supplementary File S1). All assemblies had a single copy of all Benchmarking Universal Single-Copy Orthologs (BUSCO) known for stramenopiles, indicating near complete genome gene content. On average, the assemblies were composed of 14.72% transposons and repetitive elements, with a minimum of 11.93% and a maximum of 15.49%. On average, 16,847 protein-coding genes were predicted per assembly, a number which was not influenced by differences in coverage (Supplementary File S1). On average, 724 of these genes (4.3%) were predicted to be secreted. Moreover, several proteases (560 to 657) and CAZymes (345 to 462) were predicted in the genomes, with the most numerous CAZyme class being glycoside hydrolases, with an average of 206 genes per genome (Supplementary File S2).

The comparability of the Illumina-based assemblies to the *A. euteiches* ATCC201684 PacBio reference genome was also assessed (Supplementary File S3). An average of 68.2% of the reference strain genome (44.9 to 50.1 Mb) was covered per strain assembly, leaving 23.0 Mb unmapped. Of these, 14.98 Mb were consistently unmapped across all strains. This consistently unmapped fraction was enriched in repetitive elements, with 11.87 Mb (79%) overlapping annotated repeats. In addition, 2.55 Mb overlapped annotated genes, and functional summaries of these genes were dominated by repeat and transposon-associated domains (including reverse transcriptase, integrase, RNase H, helicase and zinc-finger–like domains), as well as hypothetical and low-complexity proteins. Prediction of secreted proteins identified 1,856 candidate secreted genes genome-wide. Of these, 253 overlapped consistently unmapped regions, but together covered only 0.37 Mb (2.27%) of the consistently unmapped reference fraction, indicating limited enrichment of candidate secreted or effector-like genes within these regions. Strain-specific regions averaged 25.8 Mb (49% to 54% of assemblies), suggesting substantial lineage-specific content.

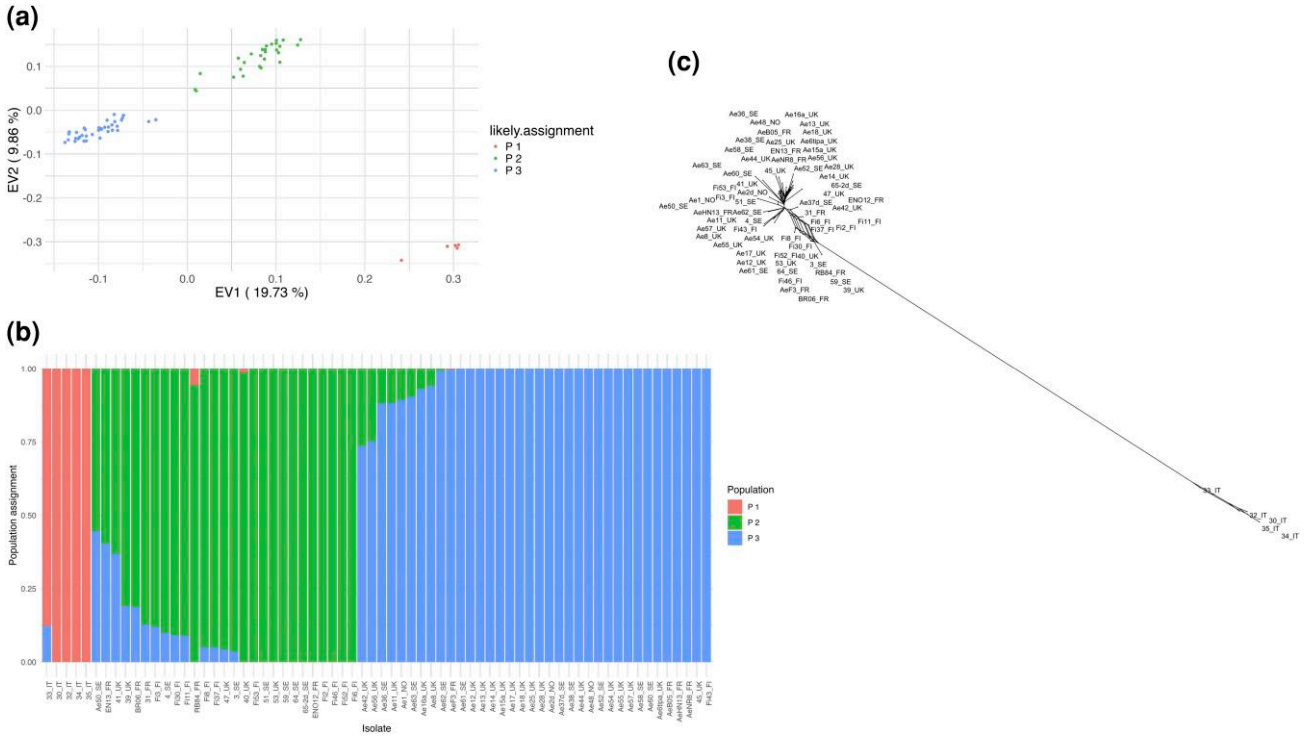

Fig. 1. Population structure analysis. a) Principal component analysis (PCA) plot of SNP data of 68 *Aphanomyces euteiches* strains compared to the reference genome ATCC201684 shows a clear clustering of strains from Italy in a separate population. Eigenvector 1 (19.73% of variance) is plotted on the x axis and eigenvector 2 (9.86% of variance) is plotted on the y axis. Points are colored according to the population to which the corresponding strains have been assigned. b) Population structure plot of all 68 *A. euteiches* strains with each vertical bar representing a strain colored according to the population to which it has been assigned. The clustering analysis supports three populations with the lowest cross-entropy at $K = 3$. Strains from Italy (Population 1) are indicated in orange. c) Phylogenetic network showing the relationship among all the 68 *A. euteiches* strains.

Population Structure Analysis Supports Three Genetically Distinct *A. euteiches* Populations

A total of 10,149 single nucleotide polymorphisms (SNPs) were identified in the genomes of the 68 *A. euteiches* strains, after filtering based on sequence depth and quality. From this main SNP dataset, a linkage disequilibrium (LD)-pruned SNP subset (1360 SNPs) was used in population structure analysis among the 68 *A. euteiches* strains using principal component analysis (PCA) and model-based clustering. The PCA revealed three distinct genetic clusters, with the five Italian strains forming a clearly separated group (Fig. 1a). Model-based clustering using sparse non-negative matrix factorization (sNMF) identified the lowest cross-entropy at K = 3, supporting the presence of three populations (Supplementary File S4). The Italian strains were assigned to a single population (Population 1, $n = 5$) with minimal admixture with the other two populations (Population 2, $n = 27$; Population 3, $n = 36$), consistent with their distinct clustering in the PCA (Fig. 1b; Supplementary File S4). A neighbor-net phylogenetic network further supported this structure (Fig. 1c). Strains within Population 2 and Population 3 clustered closely with some reticulation among them, while the Italian strains (Population 1) formed a clearly divergent clade (Fig. 1c).

Low Genetic Diversity Within *A. euteiches* Populations

Genetic diversity among the three *A. euteiches* populations was evaluated using the average nucleotide diversity (π), observed heterozygosity, the fixation index (FST), minor allele frequency spectra, and Tajima's D. Nucleotide diversity (π) varied significantly between populations ($P < 2.2 \times 10^{-16}$). Population 3 exhibited significantly lower genome-wide diversity (mean $\pi = 9.29 \times 10^{-5}$) compared with Population 1 (mean $\pi = 1.65 \times 10^{-4}$, $P < 2 \times 10^{-16}$) and Population 2 (mean $\pi = 1.68 \times 10^{-4}$, $P < 2 \times 10^{-16}$), while Populations 1 and 2 showed no significant difference ($P = 1$, Bonferroni-corrected pairwise t-tests). These overall low π values reflect low genetic variation within each population, consistent with the predominant homothallic reproductive mode of *A. euteiches*.

Observed genome-wide heterozygosity also varied significantly among the three populations ($P < 2 \times 10^{-16}$), with Population 1 exhibiting the highest mean heterozygosity (mean = 0.738, SD = 0.0612), followed by Population 2 (mean = 0.434, SD = 0.0545) and Population 3 (mean = 0.201, SD = 0.0422). Examination of individual-level heterozygosity distributions showed that the elevated heterozygosity in Population 1 was consistent across all five strains, with a narrow distribution and no extreme outliers (Fig. 2). In contrast, Populations 2 and 3 displayed lower heterozygosity and greater among-individual variation (Fig. 2). Analysis of heterozygosity in nonoverlapping 10 kb genomic windows further revealed that heterozygous sites were broadly distributed across the genome and not concentrated in extended contiguous regions. Across all populations, runs of elevated heterozygosity (heterozygosity ≥ 0.5) were short, typically spanning 30 to 50 kb. Notably, despite their high genome-wide heterozygosity, strains from Population 1 did not exhibit long heterozygous blocks, indicating that heterozygosity in this population is distributed across the genome rather than concentrated in large genomic segments.

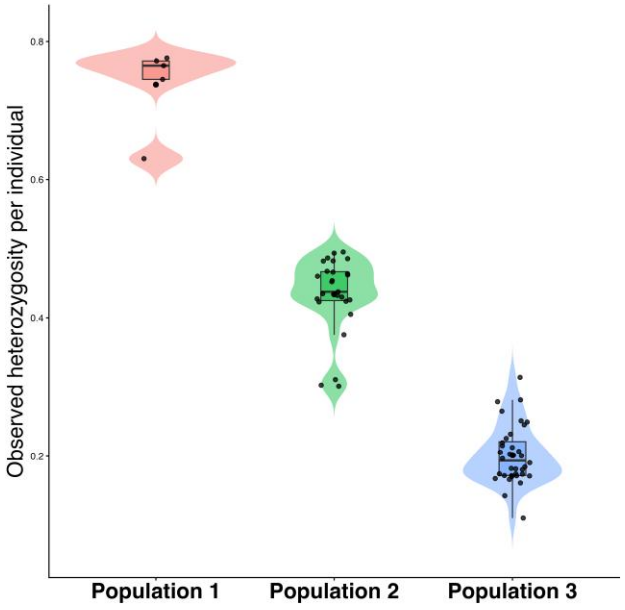

Fig. 2. Individual-level observed heterozygosity across *Aphanomyces euteiches* populations. Violin plots show the distribution of observed heterozygosity per individual strain within each population. Boxplots indicate the median and interquartile range, and points represent individual strains.

Genome-wide FST estimates revealed pronounced genetic differentiation across the three *A. euteiches* populations. The highest differentiation was observed between Population 1 and Population 3 (mean FST = 0.492, SD = 0.331, median = 0.584), indicating strong genetic isolation (Fig. 3a). Moderate differentiation was found between Population 1 and Population 2 (mean FST = 0.340, SD = 0.261, median = 0.342), suggesting

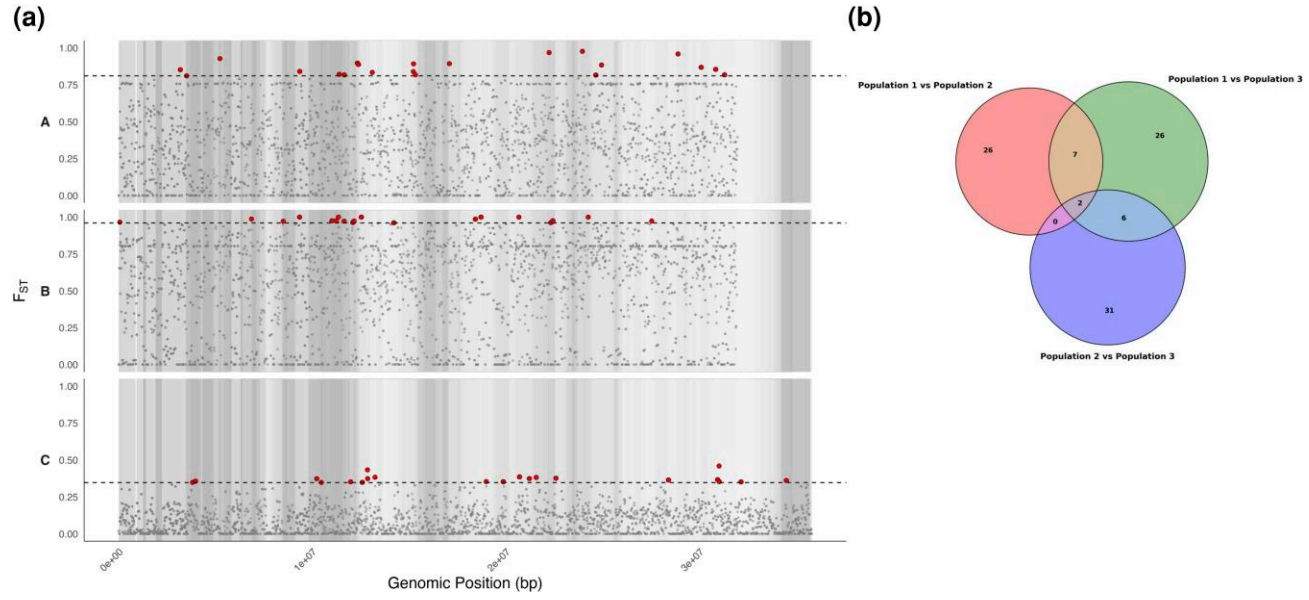

Fig. 3. Genetic differentiation analysis. a) Manhattan plots of genome-wide FST. a–c) Show FST values across genomic positions for pairwise populations comparisons: a) Population 1 versus. Population 2 (mean FST = 0.340), b) Population 1 versus. Population 3 (mean FST = 0.492), and c) Population 2 versus. Population 3 (mean FST = 0.081). Points represent FST in 10 Kb windows with a 5 Kb step. Red points indicate outliers above the 99th percentile threshold (dashed line), highlighting regions of significant genetic differentiation. Contigs are delineated by alternating gray and white backgrounds. b) Distribution of genes related to genetic differentiation across the three populations (genes located in genomic regions with highest 1% FST values).

significant genetic structure (Fig. 3a). In contrast, Populations 2 and 3 exhibited low genetic differentiation (Fig. 3a; mean FST = 0.081, SD = 0.088, median = 0.060), suggesting recent shared ancestry or ongoing gene flow. Using percentile-based outlier approach, 32, 31, and 34 genomic windows with unusually high FST values (above the 99th percentile of the genome-wide FST distribution) were identified between Population 1 and Population 2, Population 1 and Population 3, and Population 2 and Population 3, respectively (Fig. 3a). These highly differentiated windows overlapped with 35, 41, and 39 genes, respectively, defined as genes whose annotated mRNA features intersected with at least one outlier window and thus represent candidates for lineage-specific differentiation and potential local adaptation. Two genes overlapped high FST windows in all comparisons among the three populations, suggesting key roles in the genetic differentiation of the *A. euteiches* populations, while seven genes were common to Population 1-related comparisons, and 26 genes were unique to each of Population 2 and Population 3 (Fig. 3b; Supplementary File S5). In total, 98 genes, derived from 101 mRNA transcripts, overlapped high FST windows. Functional annotation of these genes revealed enrichment for domains such as serine proteases (53), zinc finger (31), RNA polymerase Rpb1 (9), P-type ATPase (7), and ankyrin repeat (6), suggesting roles in proteolysis, transcriptional regulation, ion transport, and protein interactions (Supplementary File S5). Thirty-five high FST windows (36% of all outlier windows) did not overlap with annotated mRNAs, genes, or CDS, suggesting divergence in intergenic or repetitive regions (Supplementary File S5).

Tajima's D varied significantly among the three *A. euteiches* populations ($P < 2 \times 10^{-16}$), supporting the inference of distinct demographic histories. Population 1 showed a significantly higher mean Tajima's D (1.10) than Population 2 (0.95, $P = 0.0003$) and Population 3 (−0.22, $P < 2 \times 10^{-16}$). Population 2 also differed significantly from Population 3 ($P < 2 \times 10^{-16}$). The positive Tajima's D values in Populations 1 and 2, together with minor allele frequency spectra enriched for intermediate and high-frequency variants (Fig. 4), suggest either balancing selection or recent population bottlenecks. Conversely, the negative Tajima's D in Population 3, combined with an excess of low-frequency alleles in its minor allele frequency spectrum (Fig. 4), points to a recent population expansion.

To further investigate the evolutionary dynamics of *A. euteiches*, LD decay was assessed in Populations 2 and 3, while Population 1 was excluded from the LD analysis due to its small sample size (five strains). LD decayed to 50% of its maximum value at approximately 3,150 bp in Population 2 and 2,950 bp in Population 3 (Fig. 5). The relatively short LD half-decay distances suggest that, despite a population structure shaped primarily by selfing, these *A. euteiches* populations may occasionally outcross and thus experience occasional recombination.

Italian *Aphanomyces* Strains Constitute a Separate Phylogenetic Species

The genetic separation of Italian strains compared with other European *A. euteiches* strains in the population genetic analyses prompted an investigation of species boundaries. A genealogical concordance phylogenetic species recognition analysis (GCPSR [Taylor et al. 2000]) was performed by comparing 94 BUSCO phylogenetic trees and locating points of incongruence. First, the phylogenetic analysis based on concatenated BUSCO alignments showed that the five Italian strains clustered as a separate clade, distinct from all other

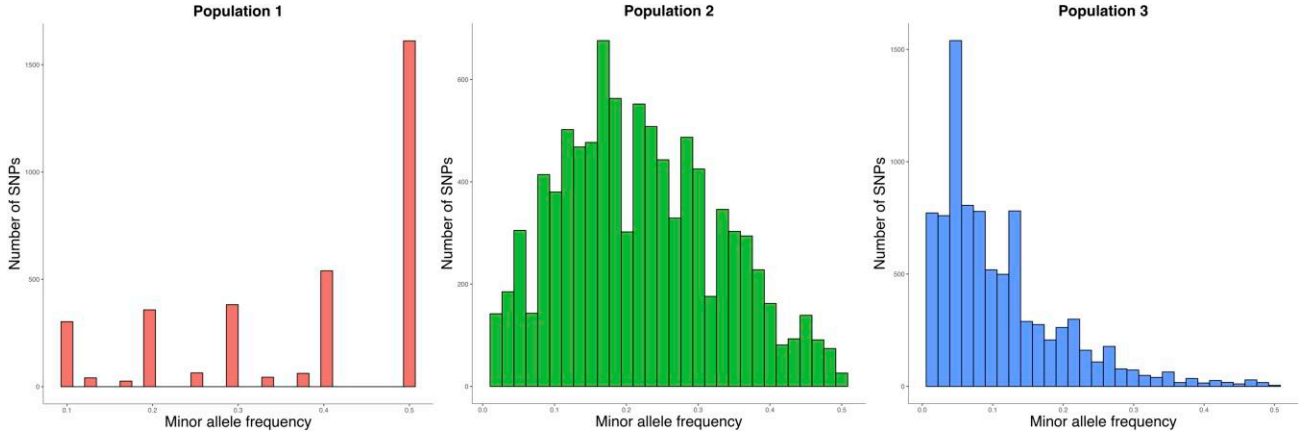

Fig. 4. Histograms of alternate allele frequency distribution for SNPs in the three *Aphanomyces euteiches* populations. Left corresponds to Population 1 (orange), middle corresponds to Population 2 (green), and right corresponds to Population 3 (blue).

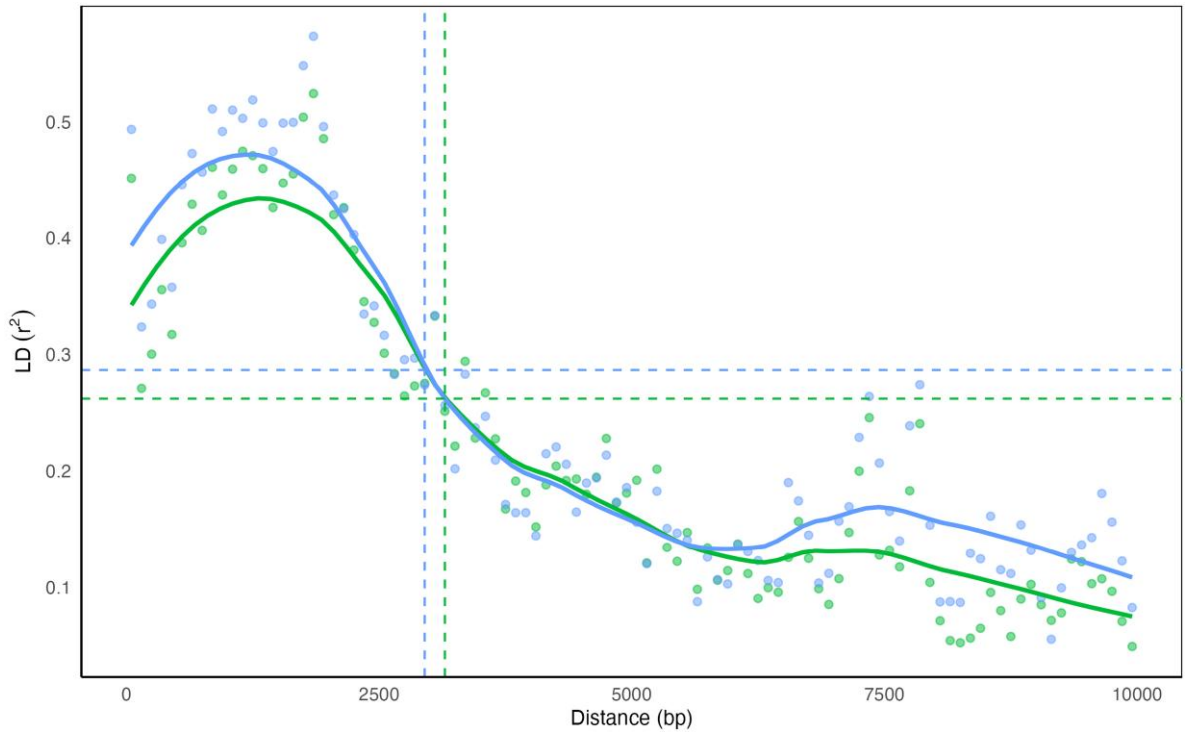

Fig. 5. Linkage disequilibrium (LD) decay in *Aphanomyces euteiches* Populations 2 and 3. The plot illustrates LD decay, measured as R^2, as a function of physical distance (bp) between loci for Population 2 (green) and Population 3 (blue). Points represent mean R^2 values in 100 bp bins, with LOESS-smoothed curves showing the decay trend. Dashed horizontal lines indicate half the maximum R^2 value for each population, and dashed vertical lines mark the estimated LD half-decay distances (Population 2: 3150 bp; Population 3: 2950 bp). Population 1 was excluded from the LD analysis due to its small sample size.

A. euteiches strains (Supplementary File S6). In 79 out of 94 phylogenetic trees of individual BUSCOs, the Italian strains formed a separate, well-supported (bootstrap values ≥ 73) clade (Supplementary File S6). In four of the remaining phylogenetic trees (15242at33634, 5074at33634, 25187at33634 and 26441at33634), individual Italian strains were intermingled with other European strains with 77% to 79% bootstrap support. The remaining eleven phylogenetic trees were inconclusive. When comparing these 79 trees, well-supported (bootstrap values ≥72) incongruence between the phylogenetic relationships between Italian strains was observed, consistent with incomplete lineage sorting or recombination, and that population 1 should be viewed as a separate phylogenetic species (PS1). Examples of phylogenetic trees showing incongruence between Italian strains are shown in Supplementary File S7. For Population 2 and Population 3 *A. euteiches* strains, limited genetic structure was observed with no correlation to geographic origin in BUSCO phylogenetic trees. Concordance factor analysis further supported the monophyly of the Italian strains. The branch subtending the Italian lineage showed a gene concordance factor (gCF) of 89.4% (84/94 gene trees) and a bootstrap support of 100, indicating strong locus-level agreement across BUSCO genes. In contrast, internal branches within the Italian clade exhibited low gene concordance (gCF = 9.6% to 12.8%), with the majority of loci supporting alternative topologies. The likelihood-based site concordance factor for the Italian stem branch was 14%, indicating limited site-level signal despite high gene-level concordance. A pairwise homoplasy index (PHI) test was performed on a concatenated BUSCO alignment subset including the five Italian strains, five representative non-Italian *A. euteiches* strains, and *A. euteiches* ATCC201684 as reference. The PHI test did not detect statistically significant evidence of recombination in this subset ($P = 0.3869$), suggesting that the observed discordance within the Italian clade is compatible with incomplete lineage sorting and/or recombination below the detection power of this test.

Aphanomyces Genetic Groups Does not Correlate With Virulence on pea

The genetic separation of *Aphanomyces* strains into three different genetic populations prompted investigations

into a possible correlation with phenotypic differences. Greenhouse pot assays to assess *Aphanomyces* virulence on two different pea genotypes differing in susceptibility was successful for 44 strains (Supplementary File S8). Statistical analyses revealed significant ($P < 0.001$) effects of *A. euteiches* strains and pea genotype, as well as their interaction on disease index (DI). None of the strains was significantly more virulent on the partially resistant genotype PI180693 compared with the susceptible Linnea. However, DI for 21 strains did not differ significantly ($P \geq 0.05$) between the two genotypes (Supplementary File S9).

To analyze differences in virulence between genetic groups, the entire data set was initially analyzed with a two-way ANOVA assessing the effect of pea genotype and genetic group on DI, as well as their interaction. Neither genetic group nor the interaction of genetic group and pea genotype were significant, whereas pea genotype showed to be highly significant ($P < 0.001$). In a second step, one-way ANOVAs (split by pea genotype) showed no significant differences in DI between genetic groups on either PI180693 or Linnea (Supplementary File S9).

Predicted Multidrug Resistance Transporters are Expanded in the Ancestor to Plant-Pathogenic *Aphanomyces* Species

To gain insight into the genomic consequences of speciation and adaptation in *Aphanomyces*, gene family expansions and contractions between the nonplant-pathogenic *A. invadans* and *A. stellatus*, and the plant-pathogenic *A. cochlioides*, *A. euteiches* and PS1 (Population 1) were analyzed with the program CAFE. Among the nonplant-pathogenic species, a total of 163 families were identified to evolve nonrandomly ($P < 0.05$). More specifically, there were 123 gene family expansions and four contractions in *A. stellatus*, while seven expansions and 29 contractions were detected in *A. invadans* (Fig. 6). Among them, metallopeptidase family M13, serine protease family S08, and serine carboxypeptidase family S28 underwent a significant ($P < 0.05$) contraction in *A. invadans* and an expansion in *A. stellatus* (Supplementary File S10). Furthermore, nine gene family expansions and ten contractions were detected on the phylogenetic branch representing the ancestor to the three plant-pathogenic *Aphanomyces* species (Table 1, Fig. 6). These included an expanded family of ATP-binding cassette (ABC) transporters (transporter classification database [TCDB] no. TC 3.A.1.201) predicted to be involved in multidrug resistance (N0.HOG0000102). Other expanded gene families were predicted to encode zinc-finger transcription factors (N0.HOG0000015 and N0.HOG0000368) and ricin B-like domain-containing glycoside hydrolases (N0.HOG0000468). Two oxidoreductase families (N0.HOG0000342 and N0.HOG0000589) and a metalloprotease family M28E were significantly ($P < 0.05$) contracted in the same branch (Supplementary File S10).

Predicted Multidrug Resistance Transporters are Further Expanded in the Ancestor to Pea-Infecting *Aphanomyces* Species

There were 19 significantly ($P < 0.001$) expanded and 1 contracted gene family associated with the branch representing the ancestor to the *A. euteiches* and PS1 species (Fig. 6). Among the expanded families, we identified three families of predicted cytochrome P450 monooxygenases (N0.HOG0000012, N0.HOG0000013 and N0.HOG0000076), one polyketide synthase family (N0.HOG0000218), one M20 peptidase family (N0.HOG0000139), one aspartic peptidase family (N0.HOG0000281), and two zinc-finger protein families (N0.HOG0000015 and N0.HOG0000955) (Supplementary File S10). Furthermore, four predicted transporter families were expanded: the multidrug resistance-associated ABC transporter family TC 3.A.1.208 (N0.HOG0000019), the pleiotropic drug resistance ABC transporter family TC 3.A.1.205 (N0.HOG0000023), the amino acid transporters TC 2.A.18.2 (N0.HOG0000043), and the major facilitator superfamily (MFS) transporter family TC 2.A.1.11 (N0.HOG0000159).

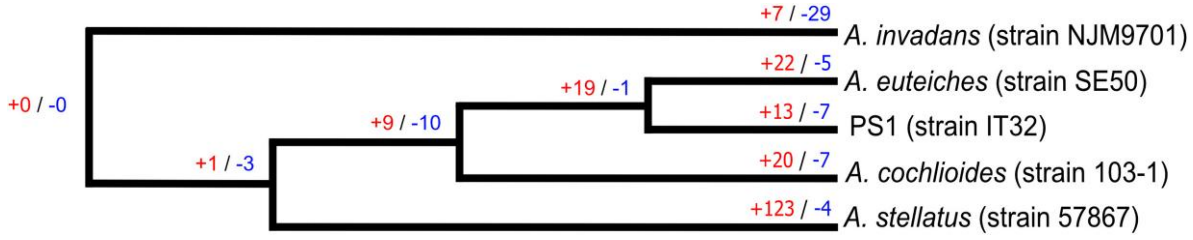

Fig. 6. Phylogenetic relationships between *Aphanomyces* species. The phylogenetic tree was generated in IQ-TREE based on concatenated BUSCOs and used for a gene family evolution analysis. The number of significant expansions (red) and contractions (blue) of orthogroups ($P < 0.001$), proteases, CAZymes, and Crinklers ($P < 0.05$) are indicated at each node and *Aphanomyces* species.

Table 1 Gene families evolving nonrandomly[a] in the ancestor to plant-pathogenic *Aphanomyces* species

Gene family no.	Gene count[b]	Gene change[c]	Annotation
N0.HOG0000015	13	+8	Zinc-finger protein
N0.HOG0000028	0	−2	Unknown function
N0.HOG0000102	10	+4	ABC multidrug resistance transporter
N0.HOG0000124	1	−3	Potassium channel protein
N0.HOG0000215	0	−2	Cellulose-binding protein
N0.HOG0000235	7	+4	Nucleotide sugar transferase
N0.HOG0000244	6	+3	Unknown function
N0.HOG0000297	5	+3	Unknown function
N0.HOG0000303	5	+3	Unknown function
N0.HOG0000342	0	−2	Ferredoxin reductase
N0.HOG0000368	5	+3	Zinc-finger protein
N0.HOG0000374	0	−3	Leucine-rich repeat protein
N0.HOG0000468	5	+3	Ricin B-like glycoside hydrolase
N0.HOG0000589	0	−2	Amine oxidase
N0.HOG0000638	0	−2	ATPase
N0.HOG0000965	0	−2	Unknown function
N0.HOG0001621	0	−2	Glutathione S-transferase
Protease_M28E	0	−2	Metalloprotease
Crinkler	121	+28	Crinkler

[a]Nonrandom evolution is defined by $P < 0.05$. [b]Predicted gene copy number in the ancestor to plant-pathogenic *Aphanomyces* species. [c]Predicted gene copy number change in comparison with the ancestral lineage.

A total of 35 gene family expansions and 12 contractions were detected in *A. euteiches* and PS1 (Fig. 6). Gene families expanded in PS1 (represented by strain IT32) included family N0.HOG0001960 putatively encoding GH81 endo-1,3-beta-glucanases and one M20 peptidase family (N0.HOG0000139). Expanded membrane transporter families included one folate/biopterin transporter family (N0.HOG0000038) of class TC 2.A.71.2, one sugar transporter family (N0.HOG0000284) class TC 2.A.1.1, while two ABC transporter families (N0.HOG0000033 and N0.HOG0000837) were significantly ($P < 0.001$) contracted. These two families were predicted to encode proteins involved in efflux of cholesterol and phospholipids (TC 3.A.1.211) and a family of pleiotropic drug resistance transporters (TC 3.A.1.205), respectively. Furthermore, two families of predicted oxidoreductases (N0.HOG0000210 and N0.HOG0000061) were expanded in PS1, while two others (N0.HOG0000181 and N0.HOG0000208) were contracted (Supplementary File S10). The expansion of M20 peptidases was also identified among protease gene families annotated by the MEROPS database, with a significant ($P < 0.05$) expansion of predicted M20D carboxypeptidases in PS1 (Supplementary File S10). Finally, one polyketide synthase family (N0.HOG0000208) and one glutathione S-transferase family (N0.HOG0000166) were lacking in PS1 and thereby contracted compared with the ancestral lineage.

Twenty-two gene families were identified to be significantly ($P < 0.001$) expanded in *A. euteiches* (represented by strain SE50), while five were contracted (Fig. 6). Expanded gene families included one orthogroup of predicted glutathione S-transferases (N0.HOG0000165), four families of predicted oxidoreductases (N0.HOG0000335, N0.HOG0000336, N0.HOG0000013 and N0.HOG0000208), two predicted sugar transporter families (TC 2.A.1.44. and TC 2.A.1.1. [N0.HOG0000453 and N0.HOG0001622]), one ABC transporter family (TC 3.A.1.211 [N0.HOG0000033]), and one gene family predicted to encode proteins containing ankyrin repeat motifs (N0.HOG0000177). Two predicted folate/biopterin transporter gene families of class TC 2.A.71.2 (N0.HOG0000038 and N0.HOG0000509) were contracted in *A. euteiches*.

Finally, we predicted an average of 168 crinkler effector genes per *Aphanomyces* genome (Supplementary File S10), based on local similarity to the CRN domain. The crinkler gene family was significantly ($P < 0.05$) expanded in the lineage representing the ancestor of plant-pathogenic species (Table 1), as well as in *A. euteiches* (SE50). Some CRN candidates were also predicted in the nonplant-pathogenic species *A. invadans* and *A. stellatus*, but these proteins were not homologous of any known CRN effectors. All *A. euteiches* strains were predicted to contain two secreted necrosis-inducing proteins, while four out of five strains of PS1 contained three genes (Supplementary File S10).

Predicted Secreted Cellulose-binding Feruloyl Esterases are Expanded in *A. euteiches*

Among CAZyme gene families, the carbohydrate esterase family 1 (CE1) was significantly ($P < 0.001$) expanded

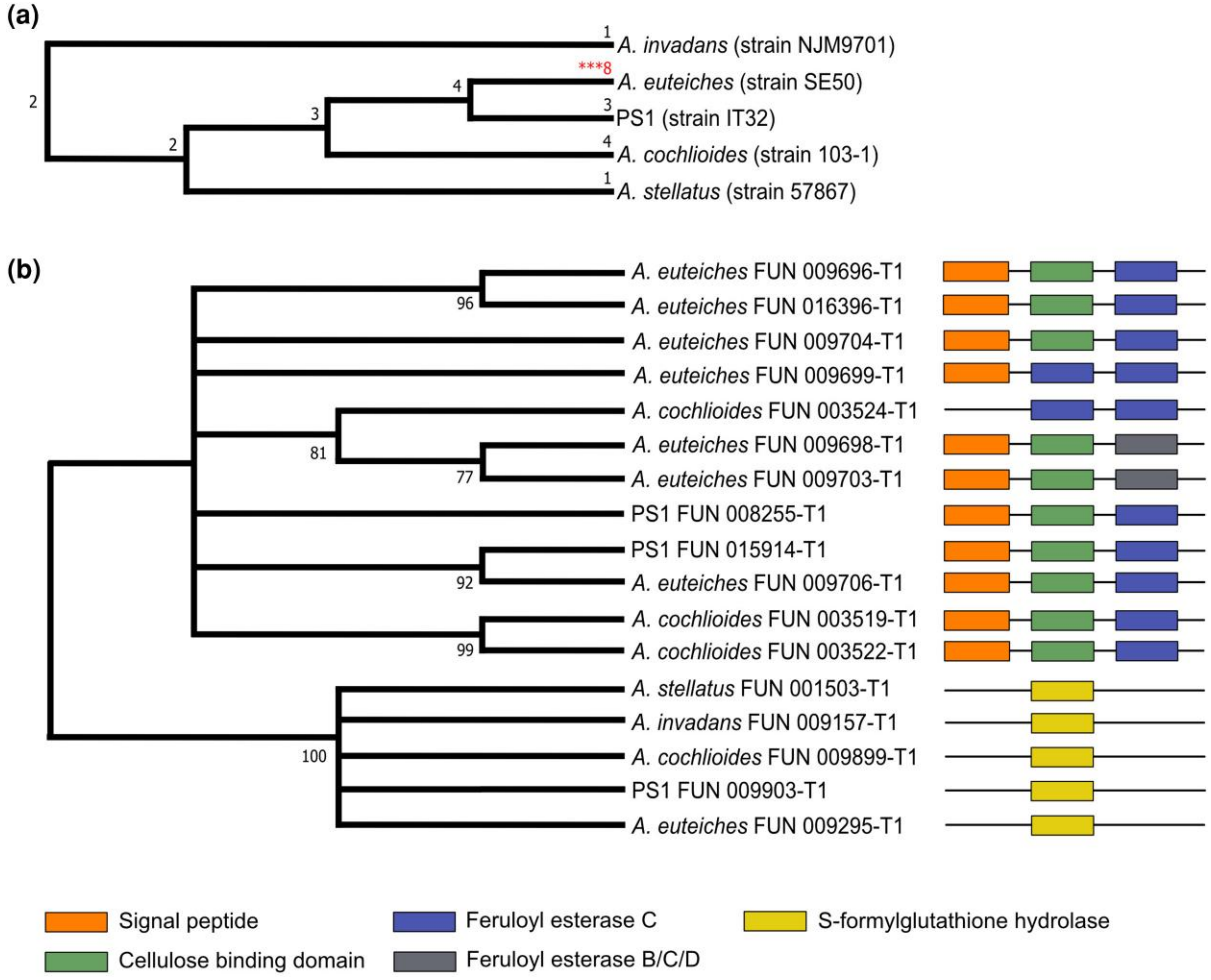

Fig. 7. Evolution of the *Aphanomyces* carbohydrate esterase family 1. Carbohydrate esterase family CE1 gene content is associated with lineages. The significant ($P < 0.001$) expansion of CE1 genes in *A. euteiches* strain SE50 is indicated with asterisks (a). Modular structure analysis of CE1 homologs in *A. invadans*, *A. stellatus*, *A. cochlioides*, *A. euteiches*, and phylogenetic species 1 (PS1) reveals one group of predicted S-formylglutathione hydrolases (EC 3.1.2.12, yellow) and one group of predicted C- or B/C/D-type feruloyl esterases (EC 3.1.1.73, blue and gray) (b).

in *A. euteiches* with eight (manually validated) genes, compared with three genes in PS1 (Fig. 7a). A phylogenetic analysis of all CE1 homologs in the five *Aphanomyces* strains identified two major groups (Fig. 7b). An analysis of the modular structure of the homologs revealed that one of the groups consisted of single orthologs of predicted S-formylglutathione hydrolases (EC 3.1.2.12), while the other group consisted of predicted C- or B/C/D-type feruloyl esterases (EC 3.1.1.73) including a secretion signal peptide and a cellulose-binding module (Fig. 7b). Members of this feruloyl esterase group were absent from the nonplant-pathogenic species, while it had two members in *A. cochlioides*, two members in PS1, and seven members in *A. euteiches*. Furthermore, high numbers of CE1 feruloyl esterases were not only present in strain SE50 (representing *A. euteiches* in the CAFE analysis) but also in all other *A. euteiches* strains from Population 2 and Population 3 (Supplementary File S11).

Discussion

In this study, we sequenced the genomes of 68 European *A. euteiches* strains using short-read Illumina technology. Although, on average, the relatively low N50 values of the resulting genomes indicate fragmented genomes unsuitable for comparisons of structural differences, the very high BUSCO recovery indicates that conserved gene space is largely captured within shared, mappable genomic regions, despite the presence of reference-specific and repeat-rich sequences. Our average genome size of 50.6 Mb falls well in the range of estimated oomycete genome sizes that span from as low as 37 Mb (*Pythium sylvaticum* and *Albugo*

laibachii) up to 280 Mb in some *Phytophthora* species (Judelson 2012). Our genome assemblies returned similar values for GC content, length, and protein-coding genes in all the considered strains. In comparison with the previously published reference genome of *A. euteiches*, the average length of our genomes was about 23 Mb shorter than the assembly of ATCC201684 (Kiselev et al. 2022). This difference is largely attributable to the improved resolution of repetitive regions by long-read PacBio sequencing, as nearly 80% of unmapped reference regions were composed of repetitive elements. While a subset of annotated genes overlapped these regions, functional characterization revealed enrichment for transposon- and repeat-associated domains. Prediction of the reference secretome further showed that only a small fraction of the unmapped reference regions overlapped candidate secreted proteins, indicating that most of the unmapped regions reflect repeat-dominated genomic compartments. However, we also identified additional contributions from reference-specific and lineage-specific sequences not captured by our Illumina assemblies, suggesting a certain level of genome plasticity.

Our population genetic analyses based on genome-wide SNPs confirm the division of European *A. euteiches* strains into three different genetic groups, as previously reported by Kälin et al. (2022) based on SSR markers. Strains from Italy form clearly a distinct genetic group separated from other *A. euteiches* strains. The previously reported clustering of the remaining *A. euteiches* strains into two separate genetic groups in central and northern Europe (Kälin et al. 2022) was also recovered using genome-wide SNPs. However, the distinct geographic separation between the groups, between the west and east parts of Sweden reported by Kälin et al. (2022), was not as clear when using genome-wide markers. This suggests ongoing movement and gene flow between the two groups. Initial population genetic studies on *A. euteiches* in the US had reported no population structure on a regional level (Malvick and Percich 1998, Malvick et al. 2009). In France, no genetic substructure among *A. euteiches* strains was found using sequence-related amplified polymorphism (SRAP) makers but *A. euteiches* strains originating from North America could be divided into three populations (Le May et al. 2018). A later and more extensive study of the *A. euteiches* genetic diversity in France using SSR markers revealed two main genetic clusters (Quillévéré-Hamard et al. 2018). The French strains used by Quillévéré-Hamard et al. (2018) are part of either Population 2 or Population 3 in our analysis although without any clear correlation with geographic origin.

The three genetic groups identified in the current work differ from each other in several population genetic characteristics, suggesting different demographic and evolutionary histories. The previously mentioned ongoing gene flow between Groups 2 and 3 is also supported by the low genetic divergence between Groups 2 and 3 based on FST in our analysis. Based on negative Tajima's D and the excess of rare variants in the minor allele frequency spectrum, we can speculate that Population 3 have undergone a population expansion perhaps driven by commercial pea production. In contrast, the positive Tajima's D and enrichment for high- and intermediate-frequency minor alleles in Groups 1 and 2 may suggest that they are the result from bottleneck events during geographic expansions to the south and north of Europe, respectively. The relatively rapid decay of LD observed in our genome-wide SNP data suggests that occasional recombination occurs in *A. euteiches* despite a predominantly homothallic mode of reproduction. This contrasts with the strong multilocus LD reported for North American populations, where genetic structure appeared to be largely shaped by clonal reproduction (Grünwald and Hoheisel 2006). Consistent with this view, the elevated levels of heterozygosity observed in Populations 1 and 2 are in line with previous reports suggesting that *A. euteiches* may occasionally outcross under agricultural field conditions (Quillévéré-Hamard et al. 2018). However, a more detailed examination of individual-level and genomic heterozygosity patterns refines this interpretation. Group 1 strains exhibited uniformly high heterozygosity across individuals, but heterozygous sites were broadly distributed across the genome rather than concentrated in long contiguous tracts. This pattern is not what would be expected if elevated heterozygosity were driven primarily by recent episodic outcrossing, which should generate extended heterozygous haplotype blocks (Pool and Nielsen 2009; Barton 2010; Satou et al. 2015). Together, these results suggest that while *A. euteiches* has the capacity to recombine and outcross, elevated heterozygosity does not necessarily reflect recent or frequent sexual reproduction but may instead result from older recombination events or the long-term maintenance of heterozygosity within lineages (Nordborg 2000; Charlesworth and Charlesworth 2010; Teterina et al. 2023).

In our study, the distinct genetic separation between the Italian strains that constitute group 1 and all other *A. euteiches* strains from Europe raises the question about the existence of a cryptic species of *Aphanomyces* in northern Italy. When applying a GCPSR analysis, Group 1 strains form a well-supported group, consistently separated from Groups 2 and 3, based on a high number of different loci, but with an incongruent structure within Group 1. This incongruent structure may be interpreted as signs of occasional outcrossing and recombination between Group 1 strains,

which is also reported in *A. euteiches* strains in the Burgundy region in France (Quillévéré-Hamard et al. 2018). However, in our BUSCO-based subset including Italian and representative non-Italian isolates, the PHI test did not detect statistically significant recombination, suggesting that the observed gene-tree discordance within Group 1 can be explained by incomplete lineage sorting and/or recombination at levels not detectable with the present dataset. Still, based on phylogenetic recognition, Group 1 may be viewed as a separate species, tentatively referred to as PS1. The use of GCPSR in the identification of cryptic species has been applied to several fungal species complexes. For example, the determination of species boundaries in the morphological species *Fusarium oxysporum* (Laurence et al. 2014) and in the identification of a novel *Blastomyces gilchristii* species (Brown et al. 2013) followed a GCPSR approach. Yet, one confounding factor in the interpretation of our results is the low number of strains representing Group 1, and the lack of sampled strains from the geographic area between group 1 and groups 2/3 (southern France, Switzerland, Germany). Whereas the genetic profile of the *A. euteiches* group from the Burgundy region was hypothesized to be linked with possible outcrossing between strains adapted to different legume hosts in the fields cropping history (Quillévéré-Hamard et al. 2018), we are lacking information about the cropping history of our sampled fields. The lack of difference in virulence on a susceptible and a partially resistant pea genotype between PS1 and *A. euteiches* genetic Groups 2 and 3 indicate that adaptation to pea infection is unlikely to be driving the observed genetic differentiation. A formal description of PS1 as a separate species requires additional sampling and phenotyping to address the issue with low number of representative strains and geographic sampling bias.

None of the *A. euteiches* strains is more virulent on the partially resistant pea genotype PI180693 compared with the susceptible Linnea. However, the high number of strains that displayed comparable disease indices on both hosts highlights the importance of identifying alternative resistance sources to achieve robust resistance that is not easily overcome by pathogen populations. Although we observe a large variation in the virulence phenotypes of *A. euteiches* strains, there is no clear connection to the genetic differentiation between Groups 2 and 3. Similarly, previous studies on the association of *A. euteiches* genotypic diversity and race phenotype or virulence have remained inconclusive (Malvick and Grau 2001, Malvick et al. 2009, Quillévéré-Hamard et al. 2018, Kälin et al. 2022).

In order to get a deeper insight into the evolutionary adaptations associated with different *Aphanomyces* species, a comparative analysis of gene content was performed, and CAFE analysis highlighted how a high number of gene families evolved nonrandomly for gene gains or losses during the evolution of the *Aphanomyces* genus. Various protease families underwent significant contractions in *A. invadans* but expansions in *A. stellatus*. Even though proteases, together with CAZymes, are reported to account for up to one-third of the *A. euteiches* secretome (Camborde et al. 2022), only two protease families (M20D and M28E) were found to evolve nonrandomly in plant-pathogenic species.

An orthogroup of 3.A.1.201 ABC transporters was among the gene families expanded in the plant-pathogenic species. This TCDB subfamily consists of multidrug resistance exporters (Saier et al. 2021), and its expansion in plant pathogens could be a way to adapt to the production of antimicrobial secondary metabolites, which are known defense mechanisms of many plant hosts (Ito et al. 2007, Huffaker et al. 2011, Zaynab et al. 2018). Interestingly, our data shows simultaneous contraction of two ABC transporter families in PS1 and an expansion in *A. euteiches*, a possible indication for different evolutionary trajectories between species driven by ecological niche differences. Next to the activation of ABC transporters, defense mechanisms in pathogenic microbes during plant infection include the activation of cytochrome P450 monooxygenases and glutathione S-transferases (Wink et al. 2012). Our results indeed showed significant expansions of cytochrome P450 monooxygenases in the pea pathogens *A. euteiches* and PS1. Further, glutathione S-transferases have been shown to act as fungal virulence factors in the detoxification of plant-derived reactive oxygen species and secondary metabolites such as isothiocyanates (Rahmanpour et al. 2009, Calmes et al. 2015, Gullner et al. 2018). These enzymes were reported to act in similar roles in the oomycete pathogen *P. infestans* (Bryant et al. 2006). Similarly to what was observed among ABC transporter families, we detected simultaneous contractions of two glutathione-S-transferase orthogroups in PS1 and the expansion of one of them in *A. euteiches*. On the other hand, we observe an expansion of an orthogroup of endo-1,3-beta-glucanases exclusively in PS1, which points to an increased ability to degrade oomycete cell walls that are mainly composed of β-1,3-glucan and cellulose (Bartnicki-Garcia 1968).

In the ancestral lineage of plant-pathogenic *Aphanomyces* species and in *A. euteiches* we observed an expansion of the CRN effector family. This confirms previous reports on the exclusiveness of CRNs to phytopathogenic oomycete species (Schornack et al. 2010, Amaro et al. 2017, Gaulin et al. 2018). The *A. euteiches* candidate effector AeCRN13 was described to induce host susceptibility to *Phytophthora capsici* when

transiently expressed in *Nicotiana benthamiana* (Ramirez-Garcés et al. 2016). Further characterization of the role of CRNs during infection and investigating possible adaptations to their target hosts could provide valuable insights into the mechanisms of *A. euteiches* pathogenicity and virulence.

In *A. euteiches*, the CE1 CAZyme family was significantly expanded in gene copy number. CEs act in the assembly, modification and breakdown of carbohydrates and glycoconjugates and are classified into 20 families in the CAZy database (Cantarel et al. 2009, Nakamura et al. 2017). A domain analysis of the predicted *Aphanomyces* members of the CE1 family revealed two major subgroups, one constituted of S-formylglutathione hydrolases, present in one copy in all strains, and one comprising feruloyl esterases C and B/C/D. This latter group is present only in plant-pathogenic species and is expanded in strain SE50 and other *A. euteiches* strains of the genetic groups 2 and 3, but not in PS1. Ferulate is normally involved in cross-linking hemicellulose polymers, lignin, arabinans and galactans, thereby having an important function in maintaining the integrity of the plant cell wall (Fry 1982, Ralph et al. 1995, Ralph et al. 1998, Caffall and Mohnen 2009). We suggest that feruloyl esterases could have a role in undermining this defense to facilitate infection and make plant cell wall components accessible to *A. euteiches*. However, additional functional studies, such as the production of feruloyl esterase deletion mutants or heterologous expression, are needed to establish causality between CE1 feruloyl esterases and virulence in *Aphanomyces*.

Materials and Methods

Aphanomyces DNA Extraction and Genome Sequencing

All *A. euteiches* strains used in this study were isolated from infected pea roots of the commercial cultivar Linnea that has been classified as highly susceptible to ARR (Kälin et al. 2022, Kälin et al. 2023). The presence and identification of *A. euteiches* were assessed by visual and microscopical inspection of diseased Linnea roots. Strains were grown in a variety of media: glucose peptone broth (glucose 5 g/L, peptone 20 g/L), vegetable medium (granini juice 20%, filtered, 0.3 g/L CaCO3), malt peptone broth (malt 30 g/L, peptone 5 g/L), Czapek-Dox ([CD]; sucrose 30 g/L, sodium nitrate 2 g/L, dipotassium phosphate 1 g/L, magnesium sulfate 0.5 g/L, potassium chloride 0.5 g/L, ferrous sulfate 0.01 g/L) or glucose peptone broth ([GPB]; glucose 5 g/L, peptone 20 g/L) for 5 to 7 d, at room temperature on a shaker (120 rpm, Supplementary File S9). Mycelia were vacuum filtered using filter paper (grade 1003; Ahlstrom Munksjö, Helsingfors, Finland) and freeze-dried prior to further processing. The variation in growth and harvested biomass of the *A. euteiches* strains posed a problem in the extraction of sufficient high-quality DNA for sequencing. Therefore, several growth media and DNA extraction methods were applied throughout the process (Supplementary File S12). DNA extractions were performed using either Genomic-tips (Qiagen, Hilden, Germany) with 20 µg, 100 µg or 500 µg tips, 3% hexadecyltrimethylammonium bromide (CTAB) extractions (Nygren et al. 2008) or NucleoBond High Molecular Weight DNA kit (Takara Bio, Kusatsu, Japan). A minimum of 2.5 µg of DNA in 10 mM Tris buffer (pH = 8) was used for sequencing. Genome sequencing was performed by the SNP&SEQ Technology Platform in Uppsala (Sweden). Sequencing libraries were prepared from 1 µg DNA using the TruSeq PCR-free DNA sample preparation kit (Illumina Inc., CA), targeting an insert size of 350 bp. Libraries were sequenced paired-end with 150 bp read length on an Illumina NovaSeq 6000 instrument.

Quality Control, Genome Assemblies and Annotation

Adapters were removed and reads cleaned using bbduk v. 38.9 (Bushnell 2018), with options "ktrim = r k = 23 mink = 11 hdist = 1 tpe tbo qtrim = r trimq = 10 maq = 10," and the success of the operation was determined with FastQC v. 0.11.9 (Andrews 2010). Afterward, genomes were assembled using SPAdes v. 3.15.0 (Bankevich et al. 2012) with options "–careful" and "–cov-cutoff auto," and gap closing was performed with ABySS-Sealer v. 2.2.5 (Paulino et al. 2015) with options "-k64 -k80 -k96 -k112 -k128." The quality of the resulting assemblies was determined with QUAST v. 2.3 (Gurevich et al. 2013) with default parameters, while their completeness was estimated using BUSCO v. 5.1.2 (Simão et al. 2015) using the "stramenopiles_odb10" dataset. Scaffolds smaller than 500 bases were removed, and after that the repetitive elements were identified with RepeatModeler v. 2.0.1 with option "-LTRStruct" and masked in the genomes using RepeatMasker v. 4.1.1 (Flynn et al. 2020) with default parameters.

Gene prediction was conducted using funannotate v. 1.8.15 (https://github.com/nextgenusfs/funannotate/tree/master). The software was first run on the *A. euteiches* ATCC201684 PacBio assembly, downloaded from AphanoDB (https://www.polebio.lrsv.ups-tlse.fr/aphanoDB/), with options "–optimize_augustus –organism other," to optimize parameters for the species. The whole Swiss-Prot database was used as protein evidence, while the expressed sequence tag sequences of the species available on AphanoDB (https://www.polebio.lrsv.ups-tlse.fr/aphanoDB/) were used as transcript evidence. Sequence reads mapping to the reference genome of

A. euteiches in a previous work (Kälin et al. 2024) were assembled using Trinity v. 2.11 (Grabherr et al. 2011) with default parameters and were used as transcript evidence as well. The same protein and transcript evidence, as well as the parameters file optimized for *A. euteiches* ATCC201684, were used to predict genes in all 68 strains. Following that, the predicted proteomes were analyzed using InterProScan v. 5.48 to 83 (Jones et al. 2014) and eggNOG version 5.0.2 (Huerta-Cepas et al. 2019) with default parameters, while the secretomes were predicted using the method described in Piombo et al. (2023). Finally, the proteomes were annotated with "funannotate annotate" (https://github.com/nextgenusfs/funannotate/tree/master), predicting CAZymes, proteases and gene ontology (GO) terms, as well as integrating the information obtained from InterProScan, eggNOG and the secretome prediction. The results of the gene prediction were converted to embl format with EMBLmyGFF3 v. 2.3 (Norling et al. 2018) and were submitted to the European Nucleotide Archive (ENA) with webin-cli v. 6.4.0 (https://github.com/enasequence/webin-cli). The same approach for transposon detection, gene prediction and gene annotation was used on the genomes of *A. cochlioides* (GCA_019828595.1), *A. invadans* (GCF_000520115.1) and *A. stellatus* (https://www.polebio.lrsv.ups-tlse.fr/aphanoDB/).

To assess the comparability of Illumina assemblies to the *A. euteiches* ATCC201684 PacBio reference genome, all 68 assemblies were aligned to the reference using NUCmer from MUMmer v. 4.0.0rc1 (Marçais et al. 2018) with a minimum match length of 50 bp and minimum cluster length of 200 bp. Alignments were then filtered for ≥ 95% identity and ≥ 500 bp length and covered regions were merged across all 68 strains. Reference coverage, unmapped regions, and strain-specific sequences were quantified using BEDTools v.2.31.1 (Quinlan 2014). Strain-specific sequences were defined as Illumina assembly sequences that did not align to the *A. euteiches* ATCC201684 PacBio reference genome under the applied alignment thresholds (at ≥ 95% identity, and ≥500 bp length), based on query-coordinate complements of NUCmer alignments. Reference coverage and unmapped regions were quantified on reference genome coordinates, whereas strain-specific sequences were quantified per assembly. Repeat content was evaluated by intersecting unmapped regions with the reference repeat annotation file, while functional and predicted secreted gene content were assessed by overlapping unmapped regions with the reference annotations. To summarize genome-wide alignment patterns across all 68 strains, three representative isolates spanning the observed range of reference coverage (low, median, and high) were selected, and whole-genome dot plots were generated to visualize the alignment of each of these representative Illumina assemblies against the PacBio reference genome individually. In addition, genome-wide reference coverage across all strains was visualized using a heatmap based on consecutive 10 kb reference bins, classifying bins as unmapped when ≥ 50% of bases in a bin lacked alignment in a given strain. To further assess the functional composition of the unmapped reference regions, the reference proteome was screened for candidate secreted proteins using SignalP v. 6.0 (Teufel et al. 2022), and proteins with multiple predicted transmembrane helices were excluded using TMHMM v. 2.0 (Krogh et al. 2001). Overlap between candidate secreted genes and unmapped reference regions was quantified using BEDTools v.2.31.1 (Quinlan 2014).

Analysis of Population Genetic Structure

Cleaned reads were mapped to the *A. euteiches* ATCC201684 PacBio assembly, downloaded from AphanoDB, using bowtie2 v. 2.4.2 (Langmead and Salzberg 2012) with the option "–very-sensitive." Duplicated reads were marked with Picard Mark Duplicates v. 2.18.29 (http://broadinstitute.github.io/picard/), BAM files were sorted and indexed using samtools v. 1.11 (Danecek et al. 2021), and SNP calling was performed with bcftools v. 1.11 (Danecek et al. 2021). Subsequently, VCFtools v.0.1.17 (Danecek et al. 2011) was used to obtain information about site-mean-depth and to filter SNPs using the options "–max-missing 0.5 –min-meanDP 5 –hwe 0.05 –minQ 500 –minDP 5 –maf 0.05." The exact scripts are available as a Snakemake workflow in Supplementary File S13. For the population structure analysis, a LD-pruned SNP dataset was generated from the main SNP dataset using PLINK v.1.90b7 (Purcell et al. 2007), applying a window size of 50 SNPs, step size of 5 SNPs, and an LD threshold of 0.5. Only SNPs without missing data were retained. Statistical phasing of SNPs was not performed, as *A. euteiches* is diploid and all downstream analyses focused on allele frequencies and individual genotypes, rather than on haplotype structure.

Population structure among the 68 *A. euteiches* strains was investigated using PCA in the R package "SNPRelate" v.1.4.0 (Zheng et al. 2012) and model-based clustering with the LEA package v.3.18.0 (Frichot and François 2015). PCA was conducted using the "snpgdsPCA" function in R v.4.4.2 (R Core Team 2025), and the first two eigenvectors were plotted. Ancestry proportions and admixture levels were inferred using sNMF with 20 replicates for K = 1–10; the lowest cross-entropy was observed at K = 3 (Frichot et al. 2014). To visualize reticulate relationships, a

neighbor-net phylogenetic network was constructed using SplitsTree v.4.19.2 (Huson and Bryant 2006).

Genetic Diversity Analysis

Nucleotide diversity (π) was calculated in nonoverlapping 10 kb windows for each *A. euteiches* population using VCFtools v.0.1.17 (Danecek et al. 2011) using the "–window-pi" option. Statistical comparisons among populations were conducted using analysis of variance (ANOVA) followed by Bonferroni-corrected pairwise t-tests in R v.4.2.3 (R Core Team 2025). Observed genome-wide heterozygosity was estimated per individual strain within populations using the "–het function" in VCFtools v.0.1.17 (Danecek et al. 2011). Summary statistics (mean, SD, minimum, and maximum) were calculated per population in R v.4.2.3 (R Core Team 2025), and individual-level heterozygosity distributions were visualized to assess within-population variation. To examine the genomic distribution of heterozygous sites, heterozygosity was further quantified in nonoverlapping genomic windows. For each strain, windowed heterozygosity was calculated in 10 kb windows as the proportion of heterozygous SNPs among all callable SNP sites per window. To assess potential clustering of heterozygous sites, contiguous runs of windows with elevated heterozygosity (heterozygosity ≥ 0.5) were identified for each strain, and the maximum length of such runs was recorded as a proxy for the size of heterozygous tracts.

Genome-wide genetic differentiation between populations was assessed using Weir and Cockerham's windowed FST, calculated in 10 kb sliding windows with a 5 kb step size in VCFtools v.0.1.17 (Danecek et al. 2011). Pairwise FST was calculated in sliding windows of 10 kb (step size 5 kb) between each population pair (population 1 vs. population 2, population 1 vs. population 3, population 2 vs. population 3). The mean, median, and SD of pairwise FST values were calculated in R v.4.2.3 (R Core Team 2025). To identify genomic regions exhibiting unusually high levels of differentiation, putative outlier windows were defined using a percentile-based threshold, with windows exceeding the 99th percentile of the genome-wide FST distribution in each pairwise comparison. Genes were associated with FST outliers if any annotated mRNA feature overlapped a 10 kb outlier FST window. Overlapping mRNA features were mapped to their parent gene IDs using annotations from the *A. euteiches* ATCC201684 reference genome, enabling identification of unique genes and their associated functional domains.

To investigate demographic signatures, Tajima's D was calculated for each population in nonoverlapping 10 kb windows with VCFtools v.0.1.17 (Danecek et al. 2011). Between population differences were tested using ANOVA and Bonferroni-corrected pairwise t-tests in R v.4.4.2 (R Core Team 2025). Minor allele frequency distribution was also generated for each population using VCFtools v.0.1.17 (Danecek et al. 2011) and visualized with ggplot2 in R v.4.4.2 (R Core Team 2025).

Linkage disequilibrium was assessed for each population using the –geno-r2 function implemented in VCFtools v.0.1.17 (Danecek et al. 2011) for unphased genotypes, calculating R^2 between SNP pairs within 10 kb windows. LD decay was determined in R v.4.2.3 (R Core Team 2025) by binning R^2 values in 100 bp intervals, fitting a LOESS curve (span = 0.5), and estimating the half-decay distance (the point where R^2 declines to half its maximum). Confidence intervals (95%) were derived via 1000 bootstrap resamples. Half-decay distances were compared across populations using ANOVA and Bonferroni-corrected pairwise t-tests.

Genealogical Concordance Phylogenetic Species Recognition Analysis

The Stramenopiles set of BUSCOs was used to identify single-copy BUSCOs in the 68 *A. euteiches* strains, as well as in the genomes of *A. cochlioides* (GCA_019828595.1), *A. invadans* (GCF_000520115.1) and *A. stellatus* (https://www.polebio.lrsv.ups-tlse.fr/aphanoDB/). Alignments of individual genes were done using mafft v. 7.453 (Katoh and Standley 2013) with options "–adjustdirectionaccurately –maxiterate 1000 –localpair." Phylogenetic trees were constructed using IQ-TREE v. 2.1.3 (Minh et al. 2020) including ModelFinder (Kalyaanamoorthy et al. 2017) with options "-m MFP -b 1000 –safe." A concatenated BUSCO alignment was used to infer a species tree under maximum likelihood in IQ-TREE. To quantify genealogical concordance among loci, gene concordance factors (gCF) were calculated using IQ-TREE by mapping the 94 individual BUSCO gene trees onto the concatenated reference tree. Likelihood-based site concordance factors (sCFl) were estimated with 1,000 quartet replicates "–scfl 1000" to quantify site-level support for internal branches. To assess potential recombination within the Italian lineage, a subset of the concatenated BUSCO alignment comprising the five Italian strains, five representative non-Italian *A. euteiches* strains (UK11, UK12, SE61, NO1, FI3), and *A. euteiches* ATCC201684 was analyzed using the pairwise homoplasy index (PHI) test for recombination implemented in SplitsTree v.4.19.2 (Huson and Bryant 2006), with significance evaluated at $P < 0.05$. Phylogenetic trees were visualized and condensed in MEGA-X v. 10.0.5 (Tamura et al. 2021) with a bootstrap cutoff value of 70%.

Analysis of Gene Family Evolution

The genomes of strains *A. cochlioides* (AC 103-1), *A. invadans* (GCF_000520115.1), *A. stellatus* 57867, *A. euteiches* SE50, and IT32 (here representing PS1) were used for an analysis of gene family evolution. Phylogenetic relationships between the selected strains were determined by selecting BUSCOs present in exactly one copy in all strains. The selected genes were concatenated and separated by a string of 20 Ns, aligned with mafft v. 7.453 (Katoh and Standley 2013) with options "–maxiterate 1000 –localpair" and used to construct a phylogenetic tree with IQ-TREE v. 2.1.3 (Minh et al. 2020) using ModelFinder (Kalyaanamoorthy et al. 2017) with options "-m MFP -b 1000 –safe."

Gene families were constructed by comparing the predicted proteomes of the five strains with Orthofinder v. 2.5.5 (Emms and Kelly 2019), and the resulting phylogenetic hierarchical orthogroups were used as gene families for an analysis using CAFE v. 5.0 (Mendes et al. 2021), with options "-e -P 0.001." Additional gene families of CAZymes and proteases were predicted by using the CAZy (Cantarel et al. 2009) and MEROPS (Rawlings et al. 2012) databases, respectively, with options "-e -P 0.05." Significant expansions and contractions in gene families were identified using CafeMiner (https://github.com/EdoardoPiombo/CafeMiner) and visualized with CafePlotter (https://github.com/moshi4/CafePlotter).

Putative CRN candidates were identified starting from CRN and CRN-like proteins identified in *A. euteiches* reported by Gaulin et al. (2018). Predicted proteomes were compared with both CRNs and CRN-like proteins using BLASTp v. 2.11.0 (Altschul et al. 1990) with a minimum identity of 60% and a minimum query coverage of 50%. Moreover, two hidden Markov Model (HMM) profiles were generated using hmmbuild v. 3.3.1 (Prakash et al. 2017) with default parameters: one profile was generated from the CRN sequences detected by Gaulin et al. (2018), while the second one used both the CRN and the CRN-like sequences of the same publication. Following this step, the proteomes of the strains were analyzed using hmmsearch v. 3.3.1 (Prakash et al. 2017) with default parameters for both HMM profiles. Any protein passing the BLAST filter, or presenting one of the two HMM profiles, was considered a CRN candidate. Necrosis-inducing proteins were predicted with InterProScan. Any protein presenting the "Necrosis-inducing protein" InterProScan identifier (IPR008701) was considered a NIP.

Predicted proteins of the CE1 family were aligned with MUSCLE (Edgar 2004) in MEGA v.11.0.10 (Tamura et al. 2021), and a maximum likelihood tree was constructed following the WAG + G model with 1,000 bootstrap repetitions and pairwise deletion of gaps and eventually condensed at a bootstrap cutoff value of 70%. The modular structure of CE1 proteins were predicted using InterProScan v. 5.48-83 (Jones et al. 2014).

Pathogen Virulence Assays on Pea Plants

Pea genotypes Linnea (susceptible) and PI180693 (partially resistant) were used for controlled infections with *A. euteiches* strains. Pea seeds were surface sterilized as described by Kälin et al. (2022), air-dried and pre-germinated on 0.8% water agar for 4 days at 20 °C. The *A. euteiches* strains were grown on corn meal agar ([CMA] BD Biosciences, San Jose, CA) plates for 3 weeks at 20 °C prior to inoculation. Three biological replicates in $9 \times 9 \times 9$ cm square plastic pots were filled with vermiculite (Sibelco, Antwerpen, Belgium) and four technical replicates (seedlings) were planted in holes of approx. 4 cm in depth and 1 cm in diameter and inoculated with one 10 mm diameter *A. euteiches* mycelium agar plug per seedling. Pea seedlings inoculated with mycelium-free CMA were included as negative controls. Pots inoculated with different *A. euteiches* strains and mock controls were kept on separate trays to prevent cross-contamination during the growth period. The experiment was conducted in a greenhouse with following conditions: 15 °C to 30 °C, 0 to 300 µ einstein per m^2/s. Sixteen days after inoculation, the roots of pea seedlings were harvested under running water and visually scored for signs of infection compared with the noninfected controls, using the scoring scale for DI as described in Kälin et al. (2022).

Statistical Analyses

Data from virulence assays on pea were tested for normality, and mock treatments were excluded from further analyses to approach normal distribution. The aov function in R (package stats v. 4.1.0, R Core Team 2021) was used to perform two-way analyses of variance (ANOVA) for the effect of *A. euteiches* strain and pea genotype on DI, as well as for their interaction, and for the effect of genetic group and pea genotype and their interaction on DI, followed by pairwise comparisons on ANOVA residuals using the pairs function. To determine the effect of genetic group (1, 2, and 3) on DI in Linnea and PI180693, two one-way ANOVAs were performed on the data set split by pea genotype. Tukey's "Honest Significant Difference" method (package stats v. 4.5.0, R Core Team 2021) was used on one-way ANOVA residuals to compare strain virulence between pea genotypes.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Agnese Kolodinska Brantestam and Anna-Kerstin Arvidsson at Findus Sverige AB for providing the *A. euteiches* strains and pea seeds, Katarina Ihrmark for her help with DNA extractions, and Anna Berlin and Sidhant Chaudhary for assisting with data analysis. We further acknowledge Christophe Le May, Anne Moussart, and Marie-Laure Pilet-Nayel at INRAE and Terres Inovia, France, for providing strains of *A. euteiches*. We thank Elodie Gaulin and Bernard Dumas at the Université de Toulouse, CNRS, UPS, France, for discussions.

Author Contributions

C.K., E.P., M.D., and M.K. planned and designed the experiment. S.D. and M.D. performed the lab work, while C.K., E.P., and D.M. analyzed the data. All authors read and approved the manuscript.

Funding

This work was funded by SLU Grogrund. DNA sequencing was performed by the SNP&SEQ Technology Platform in Uppsalaas part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The platform is supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

Data Availability

Raw sequences and annotated genomes are available on ENA under bioproject PRJEB64336.

Literature Cited

Adhikari BN, et al. Comparative genomics reveals insight into virulence strategies of plant pathogenic oomycetes. PLoS One. 2013:8:e75072. https://doi.org/10.1371/journal.pone.0075072.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990:215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Amaro TMMM, Thilliez GJA, Motion GB, Huitema E. A perspective on CRN proteins in the genomics age: evolution, classification, delivery and function revisited. Front Plant Sci. 2017:8:99. https://doi.org/10.3389/fpls.2017.00099.

Andrews S. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute; 2010.

Bankevich A, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012:19:455–477. https://doi.org/10.1089/cmb.2012.0021.

Bartnicki-Garcia S. Cell wall chemistry, morphogenesis, and taxonomy of fungi. Annu Rev Microbiol. 1968:22:87–108. https://doi.org/10.1146/annurev.mi.22.100168.000511.

Barton N. Understanding adaptation in large populations. PLoS Genet. 2010:6:e1000987. https://doi.org/10.1371/journal.pgen.1000987.

Botkin J, Chanda AK, Martin FN, Hirsch CD. A reference genome sequence resource for the sugar beet root rot pathogen *Aphanomyces cochlioides*. Mol Plant Microbe Interact. 2022:35:706–710. https://doi.org/10.1094/MPMI-11-21-0277-A.

Brown EM, et al. Phylogenetic analysis reveals a cryptic species *Blastomyces gilchristii*, sp. nov. within the human pathogenic fungus *Blastomyces dermatitidis*. PLoS One. 2013:8:e59237. https://doi.org/10.1371/journal.pone.0059237.

Brurberg MB, et al. Genetic analysis of *Phytophthora infestans* populations in the nordic European countries reveals high genetic variability. Fungal Biol. 2011:115:335–342. https://doi.org/10.1016/j.funbio.2011.01.003.

Bryant D, Cummins I, Dixon DP, Edwards R. Cloning and characterization of a theta class glutathione transferase from the potato pathogen *Phytophthora infestans*. Phytochemistry. 2006:67:1427–1434. https://doi.org/10.1016/j.phytochem.2006.05.012.

Bushnell B. BBTools: a suite of fast, multithreaded bioinformatics tools designed for analysis of DNA and RNA sequence data. Joint Genome Institute; 2018.

Caffall KH, Mohnen D. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. Carbohydr Res. 2009:344:1879–1900. https://doi.org/10.1016/j.carres.2009.05.021.

Calmes B, et al. Characterization of glutathione transferases involved in the pathogenicity of *Alternaria brassicicola*. BMC Microbiol. 2015:15:123. https://doi.org/10.1186/s12866-015-0462-0.

Camborde L, et al. An oomycete effector targets a plant RNA helicase involved in root development and defense. New Phytol. 2022:233:2232–2248. https://doi.org/10.1111/nph.17918.

Cantarel BL, et al. The carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. 2009:37:D233–D238. https://doi.org/10.1093/nar/gkn663.

Charlesworth B, Charlesworth D. Elements of evolutionary genetics. Roberts and Company Publishers; 2010.

Chepsergon J, Motaung T, Noyungayo Moleleki L. "Core" RxLR effectors in phytopathogenic oomycetes: a promising way to breeding for durable resistance in plants? Virulence. 2021:12:1921–1935. https://www.tandfonline.com/doi/full/10.1080/21505594.2021.1948277.

Danecek P, et al. The variant call format and VCFtools. Bioinformatics. 2011:27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Danecek P, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021:10:giab008. https://doi.org/10.1093/gigascience/giab008.

Derevnina L, et al. Emerging oomycete threats to plants and animals. Philos Trans R Soc Lond B Biol Sci. 2016:371:20150459. https://doi.org/10.1098/rstb.2015.0459.

Desgroux A, et al. Genome-wide association mapping of partial resistance to *Aphanomyces euteiches* in pea. BMC Genomics. 2016:17:124. https://doi.org/10.1186/s12864-016-2429-4.

Desgroux A, et al. Comparative genome-wide-association mapping identifies common loci controlling root system architecture and resistance to *Aphanomyces euteiches* in pea. Front Plant Sci. 2018:8:2195. https://doi.org/10.3389/fpls.2017.02195.
Diéguez-Uribeondo J, et al. Phylogenetic relationships among plant and animal parasites, and saprotrophs in aphanomyces (oomycetes). Fungal Genet Biol. 2009:46:365–376. https://doi.org/10.1016/j.fgb.2009.02.004.
Edgar R. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004:32:1792–1797. https://doi.org/10.1093/nar/gkh340.
Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019:20:238. https://doi.org/10.1186/s13059-019-1832-y.
Erwin DC, Ribeiro OK. Phytophthora diseases worldwide. American Phytopathological Society (APS Press); 1996.
Flynn JM, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020:117:9451–9457. https://doi.org/10.1073/pnas.1921046117.
Fouché S, Plissonneau C, Croll D. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. Curr Opin Microbiol. 2018:46:34–42. https://doi.org/10.1016/j.mib.2018.01.020.
Frichot E, François O. LEA: an R package for landscape and ecological association studies. Methods Ecol Evol. 2015:6:925–929. https://doi.org/10.1111/2041-210X.12382.
Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. Fast and efficient estimation of individual ancestry coefficients. Genetics. 2014:196:973–983. https://doi.org/10.1534/genetics.113.160572.
Fry SC. Phenolic components of the primary cell wall. Feruloylated disaccharides of d-galactose and l-arabinose from spinach polysaccharide. Biochem J. 1982:203:493–504. https://doi.org/10.1042/bj2030493.
Fry W. *Phytophthora infestans*: the plant (and *R* gene) destroyer. Mol Plant Pathol. 2008:9:385–402. https://doi.org/10.1111/j.1364-3703.2007.00465.x.
Gaulin E, et al. Transcriptome of *Aphanomyces euteiches*: new oomycete putative pathogenicity factors and metabolic pathways. PLoS One. 2008:3:e1723. https://doi.org/10.1371/journal.pone.0001723.
Gaulin E, et al. Genomics analysis of *Aphanomyces* spp. identifies a new class of oomycete effector associated with host adaptation. BMC Biol. 2018:16:43. https://doi.org/10.1186/s12915-018-0508-5.
Grabherr MG, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011:29:644–652. https://doi.org/10.1038/nbt.1883.
Grünwald NJ, Hoheisel G-A. Hierarchical analysis of diversity, selfing, and genetic differentiation in populations of the oomycete *Aphanomyces euteiches*. Phytopathology. 2006:96:1134–1141. https://doi.org/10.1094/PHYTO-96-1134.
Gullner G, Komives T, Király L, Schröder P. Glutathione S-transferase enzymes in plant-pathogen interactions. Front Plant Sci. 2018:9:1836. https://doi.org/10.3389/fpls.2018.01836.
Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013:29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
Haas BJ, et al. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nat Biotechnol. 2009:461:393–398. https://doi.org/10.1038/nature08358.
Hamon C, et al. New consistent QTL in pea associated with partial resistance to *Aphanomyces euteiches* in multiple French and American environments. Theor Appl Genet. 2011:123:261–281. https://doi.org/10.1007/s00122-011-1582-z.
Hamon C, et al. QTL meta-analysis provides a comprehensive view of loci controlling partial resistance to *Aphanomyces euteiches* in four sources of resistance in pea. BMC Plant Biol. 2013:13:45. https://doi.org/10.1186/1471-2229-13-45.
Huerta-Cepas J, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019:47:D309–D314. https://doi.org/10.1093/nar/gky1085.
Huffaker A, et al. Novel acidic sesquiterpenoids constitute a dominant class of pathogen-induced phytoalexins in maize. Plant Physiol. 2011:156:2082–2097. https://doi.org/10.1104/pp.111.179457.
Hughes TJ, Grau CR. Aphanomyces root rot or common root rot of legumes. The Plant Health Instructor; 2013.
Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006:23:254–267. https://doi.org/10.1093/molbev/msj030.
Iberahim NA, Trusch F, van West P. *Aphanomyces invadans*, the causal agent of epizootic ulcerative syndrome, is a global threat to wild and farmed fish. Fungal Biol Rev. 2018:32:118–130. https://doi.org/10.1016/j.fbr.2018.05.002.
Ito S-i, et al. α-Tomatine, the major saponin in tomato, induces programmed cell death mediated by reactive oxygen species in the fungal pathogen *Fusarium oxysporum*. FEBS Lett. 2007:581:3217–3222. https://doi.org/10.1016/j.febslet.2007.06.010.
Jones P, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014:30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031.
Judelson HS. Dynamics and innovations within oomycete genomes: insights into biology, pathology, and evolution. Eukaryot Cell. 2012:11:1304–1312. https://doi.org/10.1128/EC.00155-12.
Kälin C, et al. Genetic diversity of the pea root pathogen *Aphanomyces euteiches* in Europe. Plant Pathol. 2022:71:1570–1578. https://doi.org/10.1111/ppa.13598.
Kälin C, et al. Evaluation of pea genotype PI180693 partial resistance towards aphanomyces root rot in commercial pea breeding. Front Plant Sci. 2023:14:1114408. https://doi.org/10.3389/fpls.2023.1114408.
Kälin C, et al. Transcriptomic analysis identifies candidate genes for aphanomyces root rot disease resistance in pea. BMC Plant Biol. 2024:24:144. https://doi.org/10.1186/s12870-024-04817-y.
Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017:14:587–589. https://doi.org/10.1038/nmeth.4285.
Kamoun S. Molecular genetics of pathogenic oomycetes. Eukaryot Cell. 2003:2:191–199. https://doi.org/10.1128/EC.2.2.191-199.2003.
Kamoun S. A catalogue of the effector secretome of plant pathogenic oomycetes. Annu Rev Phytopathol. 2006:44:41–60. https://doi.org/10.1146/annurev.phyto.44.070505.143436.
Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013:30:772–780. https://doi.org/10.1093/molbev/mst010.
Kiselev A, et al. The root pathogen *Aphanomyces euteiches* secretes modular proteases in pea apoplast during host infection. Front Plant Sci. 2023:14:1140101. https://doi.org/10.3389/fpls.2023.1140101.

Kiselev A, San Clemente H, Camborde L, Dumas B, Gaulin E. A comprehensive assessment of the secretome responsible for host adaptation of the legume root pathogen *Aphanomyces euteiches*. J Fungi. 2022:8:88. https://doi.org/10.3390/jof8010088.

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J Mol Biol. 2001:305:567–580. https://doi.org/10.1006/jmbi.2000.4315.

Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012:9:357–359. https://doi.org/10.1038/nmeth.1923.

Latijnhouwers M, de Wit PJGM, Govers F. Oomycetes and fungi: similar weaponry to attack plants. Trends Microbiol. 2003:11:462–469. https://doi.org/10.1016/j.tim.2003.08.002.

Laurence MH, Summerell BA, Burgess LW, Liew ECY. Genealogical concordance phylogenetic species recognition in the *Fusarium oxysporum* species complex. Fungal Biol. 2014:118:374–384. https://doi.org/10.1016/j.funbio.2014.02.002.

Lavaud C, et al. Validation of QTL for resistance to *Aphanomyces euteiches* in different pea genetic backgrounds using near-isogenic lines. Theor Appl Genet. 2015:128:2273–2288. https://doi.org/10.1007/s00122-015-2583-0.

Le May C, et al. Genetic structure of *Aphanomyces euteiches* populations sampled from United States and France pea nurseries. Eur J Plant Pathol. 2018:150:275–286. https://doi.org/10.1007/s10658-017-1274-x.

Malvick D, Grau C. Characteristics and frequency of *Aphanomyces euteiches* races 1 and 2 associated with alfalfa in the midwestern United States. Plant Dis. 2001:85:740–744. https://doi.org/10.1094/PDIS.2001.85.7.740.

Malvick DK, Grünwald NJ, Dyer AT. Population structure, races, and host range of *Aphanomyces euteiches* from alfalfa production fields in the central USA. Eur J Plant Pathol. 2009:123:171. https://doi.org/10.1007/s10658-008-9354-6.

Malvick DK, Percich JA. Genotypic and pathogenic diversity among pea-infecting strains of *Aphanomyces euteiches* from the central and western United States. Phytopathology. 1998:88:915–921. https://doi.org/10.1094/PHYTO.1998.88.9.915.

Marçais G, et al. MUMmer4: a fast and versatile genome alignment system. PLoS Comput Biol. 2018:14:e1005944–e1005944. https://doi.org/10.1371/journal.pcbi.1005944.

Martín-Torrijos L, et al. Tracing the origin of the crayfish plague pathogen, *Aphanomyces astaci*, to the southeastern United States. Sci Rep. 2021:11:9332. https://doi.org/10.1038/s41598-021-88704-8.

Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics. 2021:36:5516–5518. https://doi.org/10.1093/bioinformatics/btaa1022.

Minh BQ, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020:37:1530–1534. https://doi.org/10.1093/molbev/msaa015.

Nakamura AM, Nascimento AS, Polikarpov I. Structural diversity of carbohydrate esterases. Biotechnol Res Int. 2017:1:35–51. https://doi.org/10.1016/j.biori.2017.02.001.

Nordborg M. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. Genetics. 2000:154:923–929. https://doi.org/10.1093/genetics/154.2.923.

Norling M, Jareborg N, Dainat J. EMBLmyGFF3: a converter facilitating genome annotation submission to European nucleotide archive. BMC Res Notes. 2018:11:584. https://doi.org/10.1186/s13104-018-3686-x.

Nygren CMR, et al. Growth on nitrate and occurrence of nitrate reductase-encoding genes in a phylogenetically diverse range of ectomycorrhizal fungi. New Phytol. 2008:180:875–889. https://doi.org/10.1111/j.1469-8137.2008.02618.x.

Papavizas GC, Ayers WA. *Aphanomyces* species and their root diseases in pea and sugarbeet. Technical Bull. 1974:1485. U.S. Dept. Agriculture.

Paulino D, et al. Sealer: a scalable gap-closing application for finishing draft genomes. BMC Bioinformatics. 2015:16:230. https://doi.org/10.1186/s12859-015-0663-4.

Pilet-Nayel L, et al. Quantitative trait loci for partial resistance to aphanomyces root rot in pea. Theor Appl Genet. 2002:106:28–39. https://doi.org/10.1007/s00122-002-0985-2.

Pilet-Nayel ML, et al. Consistent quantitative trait loci in pea for partial resistance to *Aphanomyces euteiches* isolates from the United States and France. Phytopathology. 2005:95:1287–1293. https://doi.org/10.1094/PHYTO-95-1287.

Piombo E, Guaschino M, Jensen DF, Karlsson M, Dubey M. Insights into the ecological generalist lifestyle of *Clonostachys* fungi through analysis of their predicted secretomes. Front Microbiol. 2023:14:1112673. https://doi.org/10.3389/fmicb.2023.1112673.

Pool JE, Nielsen R. Inference of historical changes in migration rate from the lengths of migrant tracts. Genetics. 2009:181:711–719. https://doi.org/10.1534/genetics.108.098095.

Prakash A, Jeffryes M, Bateman A, Finn RD. The HMMER web server for protein sequence similarity search. Curr Protoc Bioinformatics. 2017:60:3.15.1–3.15.23. https://doi.org/10.1002/cpbi.40.

Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007:81:559–575. https://doi.org/10.1086/519795.

Quillévéré-Hamard A, et al. Genetic and pathogenicity diversity of *Aphanomyces euteiches* populations from pea-growing regions in France. Front Plant Sci. 2018:9:1673. https://doi.org/10.3389/fpls.2018.01673.

Quinlan AR. BEDTools: the Swiss-army tool for genome feature analysis. Curr Protoc Bioinformatics. 2014:47:11.12.11–11.12.34. https://doi.org/10.1002/0471250953.bi1112s47.

Rahmanpour S, Backhouse D, Nonhebel HM. Induced tolerance of *Sclerotinia sclerotiorum* to isothiocyanates and toxic volatiles from *Brassica* species. Plant Pathol. 2009:58:479–486. https://doi.org/10.1111/j.1365-3059.2008.02015.x.

Ralph J, et al. Cell wall cross-linking in grasses by ferulates and diferulates. Lignin and lignan biosynthesis. American Chemical Society; 1998. p. 209–236.

Ralph J, Grabber JH, Hatfield RD. Lignin-ferulate cross-links in grasses: active incorporation of ferulate polysaccharide esters into ryegrass lignins. Carbohydr Res. 1995:275:167–178. https://doi.org/10.1016/0008-6215(95)00237-N.

Ramirez-Garcés D, et al. CRN13 candidate effectors from plant and animal eukaryotic pathogens are DNA-binding proteins which trigger host DNA damage response. New Phytol. 2016:210:602–617. https://doi.org/10.1111/nph.13774.

Rawlings ND, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 2012:40:D343–D350. https://doi.org/10.1093/nar/gkr987.

R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2021.

R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2025. https://www.r-project.org.
Rocafort M, Fudal I, Mesarich CH. Apoplastic effector proteins of plant-associated fungi and oomycetes. Curr Opin Plant Biol. 2020:56:9–19. https://doi.org/10.1016/j.pbi.2020.02.004.
Rossi V, et al. Rapid detection and quantification of *Aphanomyces cochlioides* in sugar beet. J Plant Pathol. 2023:105:1581–1591. https://doi.org/10.1007/s42161-023-01490-2.
Saier MH, et al. The transporter classification database (TCDB): 2021 update. Nucleic Acids Res. 2021:49:D461–D467. https://doi.org/10.1093/nar/gkaa1004.
Sanju S, et al. Host-mediated gene silencing of a single effector gene from the potato pathogen Phytophthora infestans imparts partial resistance to late blight disease. Funct Integr Genomics. 2015:15:697–706. https://link.springer.com/article/10.1007/s10142-015-0446-z.
Satou Y, et al. Sustained heterozygosity across a self-incompatibility locus in an inbred ascidian. Mol Biol Evol. 2015:32:81–90. https://doi.org/10.1093/molbev/msu268.
Schornack S, et al. Ancient class of translocated oomycete effectors targets the host nucleus. Proc Natl Acad Sci U S A. 2010:107:17421–17426. https://doi.org/10.1073/pnas.1008491107.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015:31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021:38:3022–3027. https://doi.org/10.1093/molbev/msab120.
Taylor JW, et al. Phylogenetic species recognition and species concepts in fungi. Fungal Genet Biol. 2000:31:21–32. https://doi.org/10.1006/fgbi.2000.1228.
Teterina AA, et al. Genomic diversity landscapes in outcrossing and selfing *Caenorhabditis* nematodes. PLoS Genet. 2023:19:e1010879. https://doi.org/10.1371/journal.pgen.1010879.
Teufel F, et al. Signalp 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022:40:1023–1025. https://doi.org/10.1038/s41587-021-01156-3.
Thakur A, et al. Artifical microRNA mediated gene silencing of Phytophthora infestans single effector Avr3a gene imparts moderate type of late blight resistance in potato. Plant Pathol J. 2015:14:1–12. https://scialert.net/abstract/?doi=ppj.2015.1.12.
Vleeshouwers V, et al. Effector genomics accelerates discovery and functional profiling of potato disease resistance and Phytophthora infestans avirulence genes. PLoS One. 2008:3:e2875. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0002875.
Wink M, Ashour ML, El-Readi MZ. Secondary metabolites from plants inhibiting ABC transporters and reversing resistance of cancer cells and microbes to cytotoxic and antimicrobial agents. Front Microbiol. 2012:3:130. https://doi.org/10.3389/fmicb.2012.00130.
Wu L, et al. *Aphanomyces euteiches*: a threat to Canadian field pea production. Engineering. 2018:4:542–551. https://doi.org/10.1016/j.eng.2018.07.006.
Wu L, et al. Mapping QTL associated with partial resistance to aphanomyces root rot in pea (*Pisum sativum* L.) using a 13.2 K SNP array and SSR markers. Theor Appl Genet. 2021:134:2965–2990. https://doi.org/10.1007/s00122-021-03871-6.
Wu L, Fredua-Agyeman R, Strelkov SE, Chang K-F, Hwang S-F. Identification of novel genes associated with partial resistance to aphanomyces root rot in field pea by BSR-seq analysis. Int J Mol Sci. 2022:23:9744. https://doi.org/10.3390/ijms23179744.
Yoshida K, et al. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. Elife. 2013:2:e00731. https://doi.org/10.7554/eLife.00731.
Zaynab M, et al. Role of secondary metabolites in plant defense against pathogens. Microb Pathog. 2018:124:198–202. https://doi.org/10.1016/j.micpath.2018.08.034.
Zheng X, et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012:28:3326–3328. https://doi.org/10.1093/bioinformatics/bts606.

Associate editor: Rebecca Zufall