

RESEARCH ARTICLE

Pangenome graph analysis reveals extensive effector copy-number variation in spinach downy mildew

Petros Skiadas^{1,2}, Sofía Riera Vidal¹, Joris Dommissie¹, Melanie N. Mendel^{2,3}, Joyce Elberse², Guido Van den Ackerveken², Ronnie de Jonge^{3,4}, Michael F. Seidl^{1*}

1 Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, The Netherlands, **2** Translational Plant Biology, Utrecht University, Utrecht, The Netherlands, **3** Plant-Microbe Interactions, Utrecht University, Utrecht, The Netherlands, **4** AI Technology for Life, Department of Information and Computing Sciences, Department of Biology, Utrecht University, Utrecht, The Netherlands

* m.f.seidl@uu.nl



OPEN ACCESS

Citation: Skiadas P, Riera Vidal S, Dommissie J, Mendel MN, Elberse J, Van den Ackerveken G, et al. (2024) Pangenome graph analysis reveals extensive effector copy-number variation in spinach downy mildew. *PLoS Genet* 20(10): e1011452. <https://doi.org/10.1371/journal.pgen.1011452>

Editor: Amelia E. Barber, Friedrich Schiller University Jena: Friedrich-Schiller-Universität Jena, GERMANY

Received: June 3, 2024

Accepted: October 7, 2024

Published: October 25, 2024

Copyright: © 2024 Skiadas et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: For this study, we sequenced isolates of the 19 denominated races of *P. effusa*. Cultures of these can be requested for research from Naktuinbouw, the Netherlands (Correll et al., 2015) [134] (S3 Table). The raw sequence data and genome assemblies and annotations generated in this study are available on NCBI under BioProject PRJNA772192. The genome assemblies, gene and repeat annotation, repeat library, effector clustering, gene variation,

Abstract

Plant pathogens adapt at speeds that challenge contemporary disease management strategies like the deployment of disease resistance genes. The strong evolutionary pressure to adapt, shapes pathogens' genomes, and comparative genomics has been instrumental in characterizing this process. With the aim to capture genomic variation at high resolution and study the processes contributing to adaptation, we here leverage an innovative, multi-genome method to construct and annotate the first pangenome graph of an oomycete plant pathogen. We expand on this approach by analysing the graph and creating synteny based single-copy orthogroups for all genes. We generated telomere-to-telomere genome assemblies of six genetically diverse isolates of the oomycete pathogen *Peronospora effusa*, the economically most important disease in cultivated spinach worldwide. The pangenome graph demonstrates that *P. effusa* genomes are highly conserved, both in chromosomal structure and gene content, and revealed the continued activity of transposable elements which are directly responsible for 80% of the observed variation between the isolates. While most genes are generally conserved, virulence related genes are highly variable between the isolates. Most of the variation is found in large gene clusters resulting from extensive copy-number expansion. Pangenome graph-based discovery can thus be effectively used to capture genomic variation at exceptional resolution, thereby providing a framework to study the biology and evolution of plant pathogens.

Author summary

Plant pathogens are known to evolve rapidly and overcome disease resistance of newly introduced crop varieties. This swift adaptation is visible in the genomes of these pathogens, which can be highly variable. Such genomic variation cannot be captured with contemporary comparative genomic methods that rely on a single reference genome or focus solely on protein coding genes. To overcome these limitations and compare multiple

and pangenome graph are available on zenodo with DOI 13270715. The code written for the creation of synteny-based orthogroups in this study is available on GitHub https://github.com/TeamMGE/Skiadas2024_pangenome.

Funding: This research was financially supported by the TopSector TKI Horticulture and Starting Materials, the Netherlands, through the project LWV19284. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

genomes in a robust and scalable method, we constructed the first pangenome graph for an oomycete filamentous plant pathogen with six telomere-to-telomere genome assemblies of *Peronospora effusa*. This high-resolution pangenomic framework enabled detailed comparisons of the genomes at any level, from the nucleotide to the chromosome, and for any subset of protein-coding genes or transposable elements, to discover novel biology and potential mechanisms for the rapid evolution of this devastating pathogen.

Introduction

Filamentous plant pathogens, such as fungi and oomycetes, cause devastating diseases on crop plants, resulting in severe damage in agriculture and natural ecosystems worldwide [1–3]. Disease management strategies mainly depend on chemical pesticides and host resistances [4–6]. However, filamentous plant pathogens can rapidly overcome plant resistances and develop pesticide tolerances [4,7], especially since resistant crop cultivars have often been deployed in large monocultures [8]. These agricultural practices exert strong selective pressure on pathogens, leading to rapid diversification of pathogen populations [9,10].

To establish a successful infection, pathogens need to create an environment that supports host colonization, for instance by circumventing or suppressing host immune responses [2,11]. To this end, pathogens secrete so-called ‘effector’ proteins that often play roles in deregulating the host immune system or in manipulating host physiology to release nutrients [12]. Effectors can, however, also be recognized by plant immune receptors [13]. Effector recognition often triggers hypersensitive defence responses, a process called effector-triggered immunity [14]. In turn, this strong selection pressure imposed by the host immune system drives the emergence of novel or the diversification of existing effector repertoires to re-establish successful host colonization [2], leading to ongoing co-evolutionary arms races between pathogens and their hosts [10].

Co-evolution between pathogens and their hosts shapes pathogens’ genomes, and genomic research over the last decades has started to uncover genomic signatures of molecular processes that fuel rapid effector diversification [15]. Comparative genomics of the oomycete plant pathogen *Phytophthora infestans* and related sister species have revealed a genome organization characterized by gene-dense genomic compartments that are largely conserved and by gene-sparse genomic compartments that are rich in transposable elements (TEs) and contain effectors and other virulence-related genes [16]. These gene-sparse compartments are often highly variable between species and strains of the same species with an overabundance of structural variation as well as effectors and virulence-related genes that often evolve under positive selection [17]. While this structure is not universal to all filamentous plant pathogens, it has been widely observed in many phylogenetically distant oomycete and fungal pathogens plant pathogens [3,18–22]. This led to the emergence of the two-speed genome model, in which genome organization is thought to facilitate the rapid diversification of virulence-related genes [18,22–24], thereby enabling pathogens to quickly adapt to new or changing environments and to overcome host defences. However, this model does not offer a universal explanation for the location of all effector genes, and it is yet unclear how this genome structure has evolved [22].

Thus far, research into the genomic diversity of specific filamentous plant pathogens has often focused on the direct comparison of genomes from multiple strains of a single species to a common, single reference genome [25]. However, comparisons to single-reference genomes can only capture a small fraction of the overall species-wide diversity, especially when genomes

vary in chromosomal organization as well as gene and repeat content [26,27]. The fraction of the genome that is common to all strains within a species has been termed the ‘core’ genome and the variable fraction is often referred to the ‘accessory’ genome [28,29]. Filamentous plant pathogens often show extensive presence/absence variation between strains, leading to considerable diversity in their accessory regions and thus potential functions [29]. Importantly, effectors and other virulence-related genes are often enriched in the accessory genome [29], and single reference approaches therefore hamper the identification and subsequent characterization of pathogen effector repertoires.

To overcome potential biases introduced by single reference genomes and to better represent and study the accessory genomes of filamentous plant pathogens, genomes from multiple strains can be jointly analysed to create a pangenome [27]. A pangenome can be defined as the total number of unique genes discovered among genomes of the same species [30–32]. However, this definition ignores the contribution of TEs and other non-coding sequences to genome diversity and adaptation [29,33–35], which is particularly relevant for TE-rich accessory regions in filamentous plant pathogens [36]. To include these regions, a pangenome needs to be defined as the total genome content of a species, thus creating a sequence-resolved pangenome. This pangenome can be represented by a variation graph where each node in the graph corresponds to a nucleotide sequence and edges between nodes indicate the direction and therefore the ‘path’ of sequences through the graph [31]. Consequently, genetic variation such as single nucleotide and structural variation between strains appear as ‘bubbles’ of alternative paths through the graph [31].

The oomycete *Peronospora effusa* causes downy mildew, the economically most important disease of cultivated spinach worldwide [37,38]. This pathogen has been traditionally managed by the extensive deployment of genetic disease resistances [39,40]. However, *P. effusa* rapidly breaks resistances of newly introduced varieties [38]. To date, 19 *P. effusa* races have been denominated based on their capacity to break spinach resistances [37,41,42]. *P. effusa* can reproduce both sexually and asexually and thus new races can emerge both from sexual and asexual recombination [43,44]. *P. effusa* is one of few oomycetes with a publicly available chromosome-level reference genome [36,45,46]. The *P. effusa* genome is relatively small compared with other oomycetes and organised in 17 chromosomes [45]. The number of chromosomes and their overall collinearity is largely conserved with closely related oomycetes, pointing to a conserved genome organisation in Peronosporales [45,47]. The *P. effusa* genome is predicted to be composed of about 53.7% TEs and around 9,000 protein-coding genes. Of these genes, 300 are predicted to encode for secreted effectors that are enriched in TE-rich genomic regions, thus *P. effusa* seems to follow the two-speed genome model [45]. Two main families of cytoplasmic effectors have been characterized in oomycetes thus far, the RXLR and Crinkler (or CRN) effector families [16,48]. These families of effectors are characterized by the presence of conserved motifs at the N-terminus downstream of the signal peptide, which has been hypothesized to contribute to effector translocation into the host cell [20,49,50]. The C-terminal regions of these effectors vary significantly, and it is responsible for the effector function in the plant cell [50–52]. Diversification of effector gene repertoires most likely drives the evolution of *P. effusa* and the rapid breakdown of host resistances [53]. However, comparison of the recently published chromosome-level genome assembly of *P. effusa* strain UA202013 with closely related and likely clonally evolving isolates from races 12, 13, and 14 reported limited genomic diversity [45]. To date, however, chromosome-level genome assemblies of multiple *P. effusa* strains are still lacking and thus the evolutionary processes that might contribute to emergence of novel races are largely unknown.

Here, we leveraged a pangenome graph-based approach to compare and reannotate genomes and expanded on it by creating synteny based single-copy orthogroups for all genes.

This approach enables a comprehensive, reference-free analysis of chromosome-level genome assemblies of multiple genetically diverse *P. effusa* isolates. We show that the chromosome structure and most protein-coding genes are largely conserved. In contrast, most genomic variation is caused by TEs, but also effector and other virulence-related gene repertoires are highly dynamic. These genes are often located in clusters of highly similar gene copies, which reveals their recent copy-number changes and is the leading cause for their variation between the isolates. Our pangenome approach provides a framework to compare isolates of microbial pathogens accurately and efficiently, which will be essential to understand how these are able to rapidly break host resistances in the future.

Results

Chromosome-level genome assemblies of six *Peronospora effusa* isolates reveal a highly conserved genome structure

To fully capture the genomic variation of spinach downy mildew *P. effusa* by constructing a pangenome graph, we first sought to create gapless chromosome-level genome assemblies of a suite of diverse *P. effusa* isolates by a combination of Nanopore long-reads and short Hi-C paired-end reads. We selected *P. effusa* isolates that belong to six denominated races with the aim to capture large parts of the genomic variation present between all thus far sequenced *P. effusa* isolates [44]. We selected one isolate for each of the three phylogenetic clusters we have previously identified (isolates classified as *Pe1*, *Pe5*, and *Pe14*), and three that most likely emerged via sexual reproduction between isolates of those phylogenetic clusters (isolates classified as *Pe4*, *Pe11*, and *Pe16*) [44] (S1 Fig).

For all six isolates, we used Nanopore, Hi-C, and Illumina sequencing data to create gapless, haploid assemblies with 17 chromosomes, either with or without an additional 18th chromosome, with overall genome sizes ranging between 57.8 and 60.5 Mb (Figs 1A, S2 and S3, S1 and S2 Tables, S1 Note). The 17 assembled chromosomes have telomeric repeats with the repeat motif 'TTTAGGG' on both ends, suggesting that we successfully obtained complete and chromosome-level genome assemblies for all six isolates. The assembly completeness, evaluated by Benchmarking Universal Single-Copy Ortholog (BUSCO) analysis using the Stramenopiles database (v. odb10), revealed a 99% BUSCO completeness score, indicating that we successfully captured the protein-coding regions.

Hi-C data provides chromatin contact information, based on which, one can extract details about the 3D organisation of chromosomes in the nucleus (Torres *et al.*, 2023). The Hi-C heatmaps of *P. effusa* show increased interaction frequencies between all chromosomes with a characteristic pattern that matches the Rab1 chromosome configuration [54,55]; all the telomeres co-localize within the nucleus, shown as dots in the heatmap, and similarly all centromeres co-localize, indicated by the characteristic 'x' pattern in the Hi-C heatmap (Figs 1C and S4). The Rab1 configuration has thus far only been found in species without the Condensin II complex, which organises the nucleus into chromosomal territories [54,55], and based on sequence similarity searches, Condensin II subunits seem to be absent in *P. effusa*. By using the centromeric interactions visible in the Hi-C heatmap, we identified the approximate locations of centromeres for each chromosome (S2 Table). These centromeric interactions are approximately 250 to 300 kb long, which is in line with centromeric regions (211 to 356 kb in size) measured in *Phytophthora sojae* [56]. Importantly, centromeric regions, except for chromosomes 1 and 15, are enriched for copies of a *Copia*-like LTR transposon (Fig 1A), which has been previously identified to occur at centromeres in different oomycetes [56].

The previously published genome assembly of the *P. effusa* isolate UA-202013 has 17 chromosomes [45], and is comparable in total genome size as well as gene and repeat content to

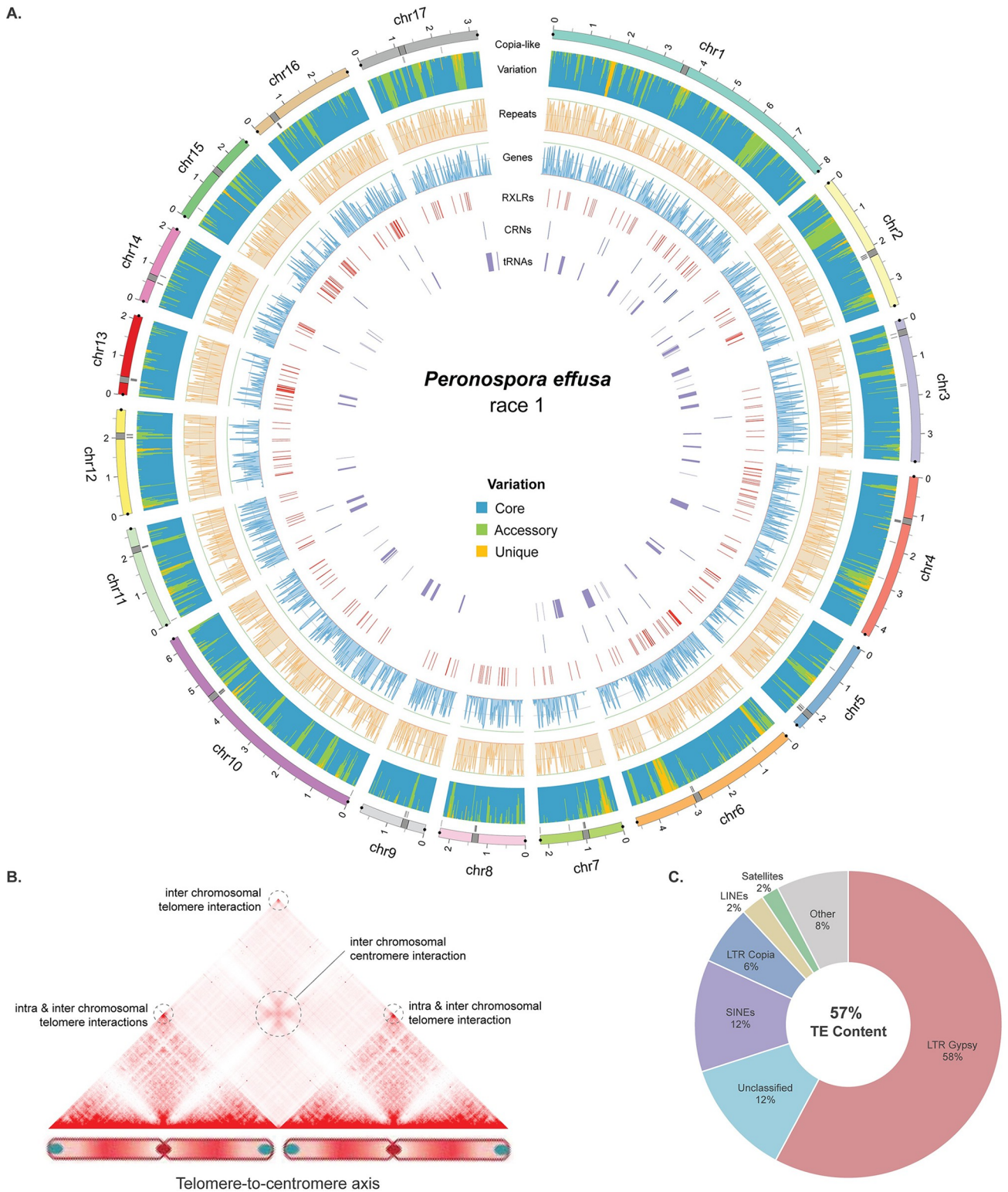


Fig 1. Chromosome-level genome assembly of *Peronospora effusa* race 1 (*Pe1*). A. The genome assembly of *Peronospora effusa* race 1 (*Pe1*) is visualised in a circular plot, individual tracks, starting from the outside to the inside: i) 17 chromosomes are shown in different colours, a grey rectangle points the location of the centromeres, grey lines underneath each chromosome show the location of the centromere-specific *Copia*-like element, which has been previously observed to be enriched in centromeres [56], and dots at either side of the chromosomes highlight the presence of telomeric repeats. ii) Stacked bar plot shows the genetic variation in the six *P. effusa* isolates, with blue representing conserved regions, green regions that are present in at

least two isolates, and orange unique regions present only in *Pe1*. iii) Line graph shows the coverage of repeat (orange) and gene (blue) content summarized in non-overlapping 20 kb windows. iv) Lines indicate the position of RXLR (red), CRN (green), and tRNA (purple) genes. **B.** Hi-C heatmap displays the spatial interactions between chromosomes in the nucleus. Genome-wide Hi-C data are aggregated over two chromosomes to highlight inter- and intra-chromosomal telomeric interactions and inter-chromosomal centromeric interactions. These interaction patterns suggest that *P. effusa* chromosomes are organized within the nucleus in a Rab1 chromosome configuration [59,60] (S3 Fig). **C.** Transposable element (TE) content in the *Pe1* genome assembly is shown separated over different TE families as percentage of the total length of TEs annotated.

<https://doi.org/10.1371/journal.pgen.1011452.g001>

the here assembled *P. effusa* isolates. Since the overall chromosome organization of oomycetes is thought to be highly conserved [36], we sought to investigate the co-linearity between our *Pe1* isolate, the UA-202013 isolate, and chromosome-level genome assemblies from *Bremia lactucae*, *Peronosclerospora sorghi*, and *Phytophthora infestans* [45–47,57]. To this end, we performed whole-genome alignments of these genomes based on sequence similarity and on relative position of protein-coding genes (S5A Fig). The *Pe1* and UA-202013 *P. effusa* isolates are completely collinear, and between species the chromosome structure remains largely conserved apart from few chromosome fusions and inversions of large chromosomal regions. We similarly observed that our six *P. effusa* isolates are nearly completely co-linear (S5B Fig). Thus, the chromosomal organization between different oomycetes and especially between isolates of the same species are highly conserved [58]. In *P. effusa*, although there are no cross chromosomal rearrangements, we nevertheless observed two large structural rearrangements (inversions), one in *Pe5* at the beginning of chromosome 17 and one in *Pe11* at the end of chromosome 1 (S5B Fig), indicating that intrachromosomal rearrangements have nevertheless occurred.

Pangenome graph-based comparison of six *Peronospora effusa* genomes reveals overall high conservation but highly variable TE and effector genes

To enable the comparisons between the *P. effusa* chromosome-level genome assemblies and to limit possible reference biases introduced by comparisons to a single reference genome, we created pangenome graphs based on Minigraph-Cactus[31]. The conserved structure of the chromosomes in *P. effusa* (S5B Fig) enabled us to generate a separate pangenome graph for each chromosome and then merge them for further analysis. The final pangenome graph has 1,948,729 nodes and 2,635,318 edges, which is only a small fraction of all theoretically possible connections between the nodes, indicating that these graphs are simple in structure. Most nodes have two connections, indicating a mostly linear graph, while only 30 nodes have a degree of eight or higher (S6A Fig). The total size of the graph is 76.6 Mb, 32% larger than the average *P. effusa* genome assembly (58 Mb), suggesting that while most of the genome is conserved, there are significant strain-specific genomic regions (Fig 2A).

To generate a homogenous framework to study the genetic differences between the six *P. effusa* isolates, we created a common TE library, and we exploited the pangenome graph to perform a joined structural annotation for the protein-coding genes on the six genome assemblies (Note S2). These methods ensure that the TE annotations are based on shared TE families and that genes that are present in identical regions in the pangenome graph will be annotated consistently between the isolates (e.g. identical intron-exon structure and ORF position) (Fig 2B steps 1–3). This resulted in the annotation of 9,916 to 10,364 protein-coding genes (on average 67% being functionally annotated with interproscan v. 5.52–86.0) for each isolate, adding between 18 to 459 genes that were missed by the individual gene annotation of each isolate. These genes were then assigned into groups based on their position on the pangenome graph, thus creating 12,739 single-copy gene orthogroups (Fig 2B step 4, S2 Note). To identify effector candidates, we searched the predicted protein-coding genes and additional open reading frames for those encoding proteins with a predicted signal peptide and for RXLR and CRN

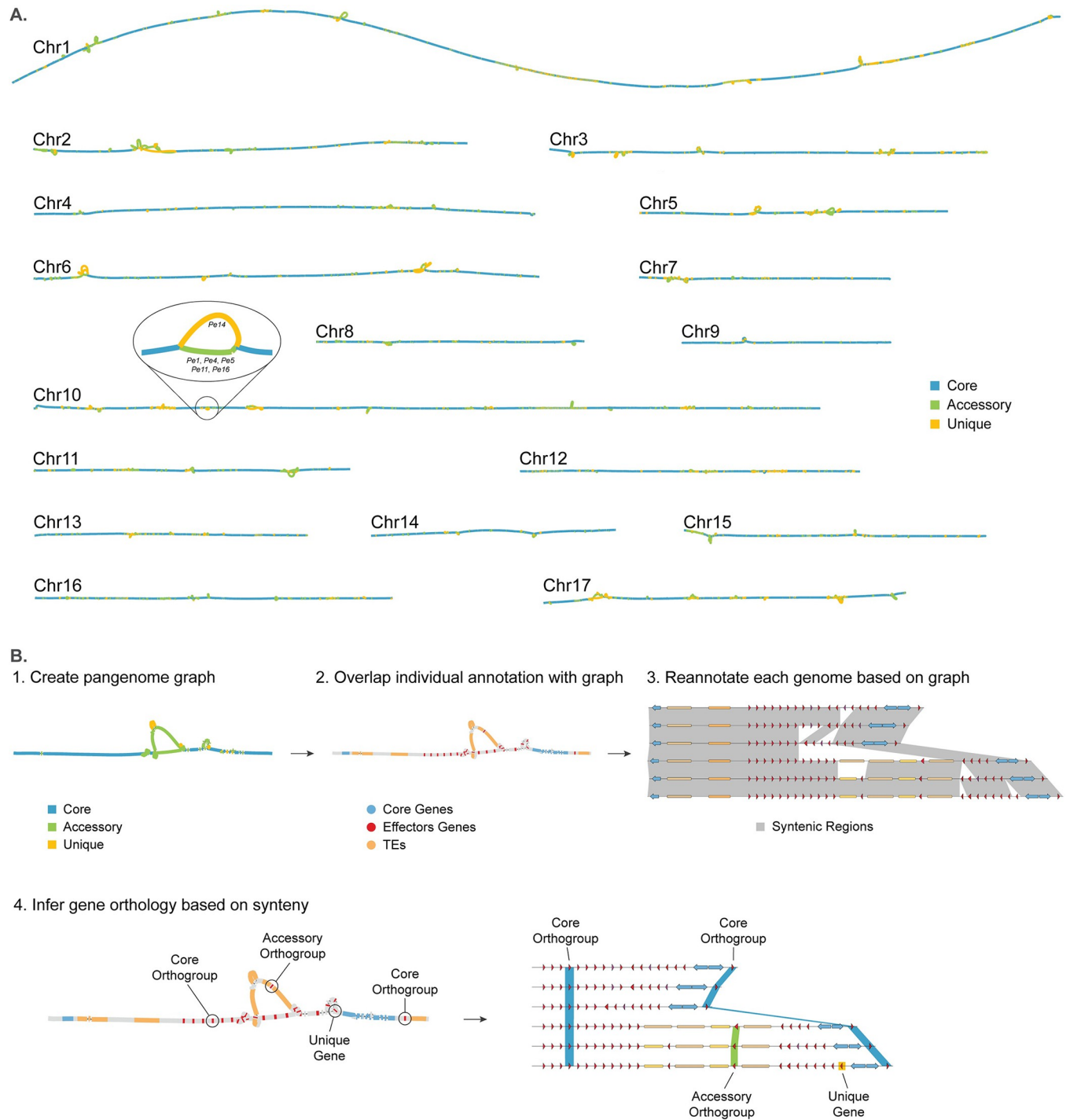


Fig 2. Reference-free genome annotation and whole-genome comparisons of six *Peronospora effusa* isolates with pangenome graphs. **A.** Pangenome graphs of the 17 chromosomes present in all *P. effusa* isolates. As an example of the graph structure, a variable region of chromosome 10 is highlighted, indicating the corresponding isolates for each accessory (green) and unique (orange) region in the graph. **B.** Step-by-step description of the pangenome graph analysis to reannotate and compare all protein-coding genes, including effector candidates. 1. The genome assemblies are used to create and annotate (determine the variation of each region) the pangenome graph; 2. Gene annotations are projected onto the pangenome graph; 3. All genome assemblies are re-annotated based on their alignment in the graph and associated gene annotation; 4. The overlap of genes on the graph is used to create orthologous gene groups based on synteny.

<https://doi.org/10.1371/journal.pgen.1011452.g002>

amino acid motifs. We identified 351 to 443 putative RXLR and 38 to 68 putative CRN effectors per isolate (Fig 1A and S2 Table).

To uncover the genomic variation between the six *P. effusa* isolates, we parsed the 17 pangenome graphs and inferred that 60% (around 45.6 Mb) of the *P. effusa* pangenome is conserved, 24% (18.6 Mb) is found in two or more isolates, and 16% (12.3 Mb) is unique for single isolates (Figs 2A, 3A, and S6C). Most of the observed variation originates from variants longer than 50 bp (87.7% of total variation size, 27.1 Mb), but single nucleotide variants (SNVs) are most frequent (76.5% of total variation number, 1.15 million) (Figs 3A and S6B). As expected, protein-coding regions are generally highly conserved (89.0%) compared with regions annotated as TEs (48.7%). This variation in TEs is mostly caused by large sequence variants (>1 bp) that are more abundant in TE regions (42%, 277,151) than in protein-coding regions (5%, 31,786), suggesting that large variation most often occurs in TEs or directly caused by TE activity.

Genetic variation in the pangenome graph can be specific to only a single isolate or shared between multiple isolates (S6D Fig). We observed that 20.2% (329,387) of the variation in TE regions is unique to only one isolate, while only 12.6% (30,948) of the variation in protein-coding regions is unique, suggesting that the addition of more diverse *P. effusa* isolates would differentially impact the number of observed genes and transposons. We therefore sought to visualize the variation discovered on the pangenome graph as a saturation plot for the whole pangenome (Fig 3B). The pangenome is clearly open (not saturated) for TEs and explains 80% of the observed variation between the *P. effusa* isolates. The TE content in *P. effusa* is highly dynamic and shaped by continuous TE expansions and deletions. To investigate the timeline of repeat expansion, we calculated the divergence of individual TE copies to their TE consensus sequence (Kimura distance), which uncovered a recent expansion of LTR *Gypsy* and *Copia* elements, as well as an older expansion of LTR *Gypsy* elements (Figs 3C and S7). The largest LTR *Gypsy* family has around 250 copies per isolate and is found in all chromosomes. These copies have a Kimura distance of only 0.002 and 60% occur within accessory or unique regions, suggesting that this TE family remained highly active after the divergence of the different *P. effusa* isolates. In contrast to the open TE pangenome, the pangenome for genes is nearly closed, demonstrating that by analysing only six diverse *P. effusa* isolates we have successfully captured most protein-coding genes in the population (Fig 3B). Effector genes though are much more variable than the rest of the genes, due to extensive copy-number variation. Since we have defined single copy orthogroups, copy-number variation is accounted for in saturation plots, revealing an open pangenome for effector genes (Fig 3B and S2 Note).

Chromosomes do not only differ in size, but also in the proportion of conserved regions, as well as repeat and gene content. We used these characteristics to study individual chromosomes with principal component analyses (PCA), which revealed four distinct groups of chromosomes (Fig 3D). The first group consists of highly conserved chromosomes (69–87% core), which includes the most repeat poor (52–62%) and smallest (1.7–2.3 Mb) chromosomes; chromosome 4 is an outlier based on chromosome size (4.3 Mb) and repeat content (70%) because it contains an older repeat expansion that affected its size but is shared between all isolates (S8 Fig; Panel Pe5_Chr4). The second group is characterized by a high repeat content (68–70%). The third group consists of the least conserved chromosomes (45–60% core), with some of the largest (3.2–4.6 Mb) chromosomes. Lastly, the fourth group consists of the two largest chromosomes 1 (8.1 Mb) and 10 (6.3 Mb), which clearly grouped apart of the other chromosomes even though these chromosomes are similarly conserved to most chromosomes, and they have an average repeat content (chromosome 1: 54% core, 61% repeats; chromosome 10: 58% core, 59% repeats). Interestingly, while most *P. effusa* chromosomes are completely syntenic with the chromosomes of *Bremia lactucae*, this is not the case for chromosome 10 [45], which

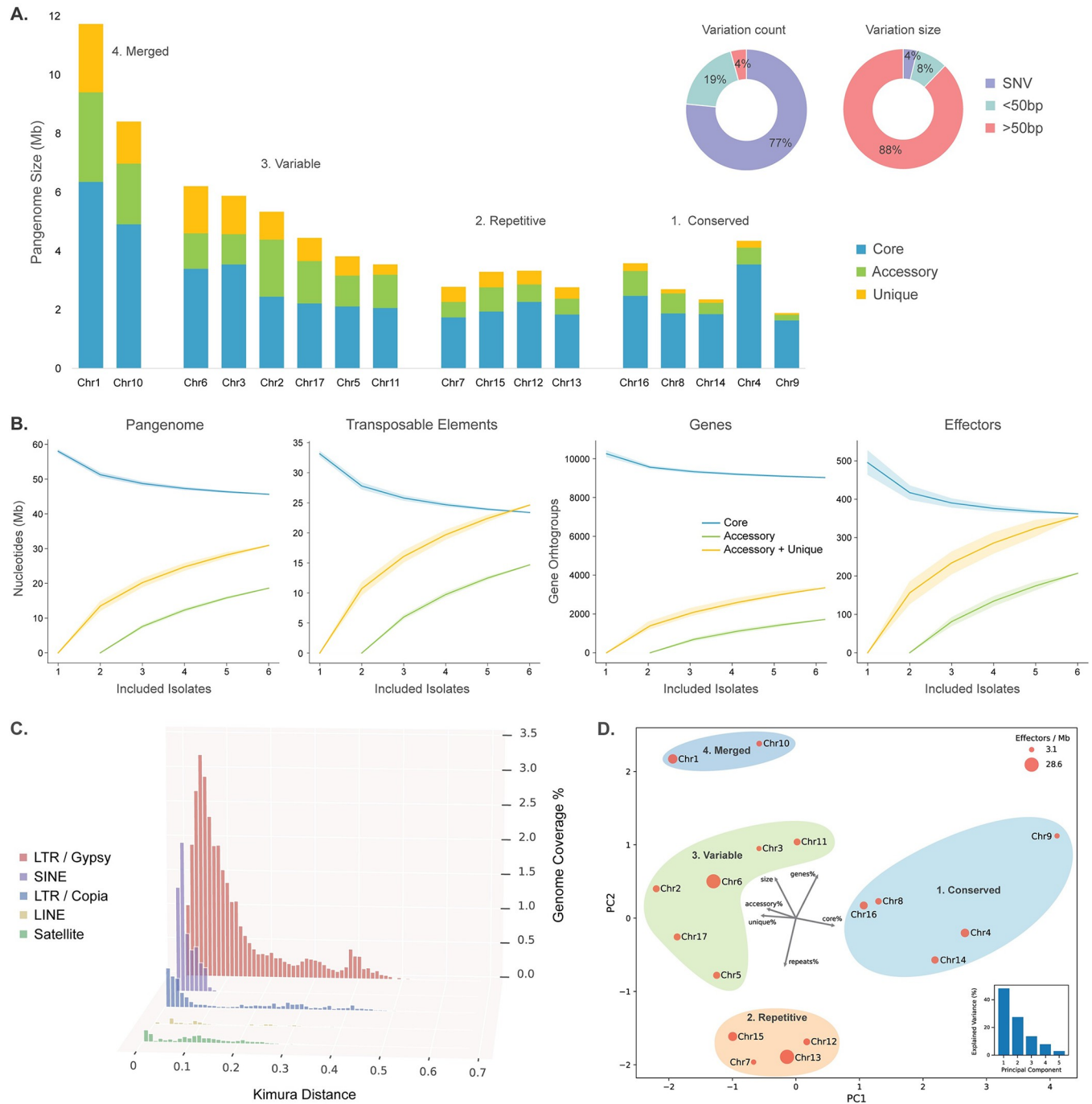


Fig 3. Pangenome graph based comparison of six *Peronospora effusa* isolates. **A.** The 17 chromosomes of *P. effusa* pangenome variation size represented in a stacked bar plot. The variation observed in the pangenome graph analysis for count and total size represented in two pie charts. **B.** Saturation plots based on the pangenome graph at the nucleotide level for the whole genome and transposable elements, and at the gene level for all genes and effector candidates. **C.** The transposable element landscape based on Kimura distances for the five largest TE superfamilies for *Pe1* are shown in order of coverage of the genome. Kimura distance is the measure of divergence between individual TE copies and the corresponding TE consensus sequence [61], i.e., the lower the Kimura distance, the more similar the copy is to the consensus and thus the more recently it was most likely copied. **D.** Principal component analyses of the 17 *P. effusa* chromosomes based on their percentage of genes, repeats, core, accessory, and unique region and their total size. The size of each point corresponds to the number of effector genes per Mb on each chromosome. The chromosomes are clustered in four distinct groups.

<https://doi.org/10.1371/journal.pgen.1011452.g003>

suggests that chromosome 10 and most likely also chromosome 1 might be the result of a fusion of two smaller ancestry chromosomes, similar to other chromosome fusion events that have been observed in Peronosporaceae [45,47] (S5A Fig).

***Peronospora effusa* has a variable and highly repetitive accessory chromosome**

The genome assembly of *P. effusa* isolate UA-202013 as well as the here assembled isolates *Pe1* and *Pe14* have 17 chromosomes (Fig 1) [45]. However, we also assembled an additional complete chromosome in *P. effusa* isolate *Pe5*, named chromosome 18. This chromosome is similar to other chromosomes as it is 2.1 Mb long, has full diploid coverage, has a centromere based on the Hi-C data, and is flanked by telomeric repeats (S3 Fig). In contrast to other chromosomes, however, it is mostly composed of TEs of the LINE and LTR superfamily (87% TE content compared with 54% genome-wide average), has higher GC content (54% compared with an average of 48%), and the two chromosomal arms are the reverse complement of each other (Fig 4A). Moreover, unlike other chromosomes, this one shares almost no sequence similarity with any other chromosome, since all the annotated LINE, satellite, and unknown repeats are unique to chromosome 18. Only four significant matches were found with other chromosomes (0.48, 0.80, 1.21, and 1.6 Mb), corresponding to copies of two LTR-*Gypsy* families, which had a recent expansion and are among the most abundant TE families in *P. effusa* (Fig 4C). Near the telomeres where the two LINE/L1 clusters are located, we annotated 18 putative protein-coding genes, nine in each of the chromosome arms. These sequences lack similarity to any other predicted proteins of *P. effusa* or any other known proteins from public databases. Since these are also not expressed, we concluded that these are most likely non-functional. Notably, we also assembled additional contigs in *Pe4*, *Pe11*, and *Pe16* that share sequence similarity with chromosome 18 of *Pe5*. However, these are 30–36% shorter, have only a single telomeric repeat, and miss half of a chromosomal arm that is unique for *Pe5* as shown in the pangenome graph (Fig 4B). We similarly observed protein-coding genes annotated in each of these contigs, yet we could not identify sequence similarity between different isolates, further corroborating that these predicted protein-coding genes are likely non-functional.

To better understand the occurrence and possible origin of this accessory chromosome, we used publicly available short-read data from whole-genome resequencing experiments for in total 32 *P. effusa* isolates [42,62,63]. Based on read mapping to the *Pe5* chromosome 18, we identified similar accessory chromosomes in nine *P. effusa* isolates, including sequencing data of three independent isolates that are classified as race 13 (*Pe13*, *R13*, and *Pfs13*) [42,62,63]. In eight *P. effusa* isolates, such as *Pe4*, *Pe11*, and *Pe16*, chromosome 18 is partially present (40–60% of the entire chromosome is covered), and in two isolates we could find only traces of the genetic material assigned to *Pe5* chromosome 18 (8–10% coverage) (Fig 4C). Interestingly, based on the known relationship between the isolates [44] (Fig 4D), we could not identify any traces of chromosome 18 in data from isolates of clusters i and iii as well as in *Pe10*, *Pe17*, and *NL-05*, suggesting that chromosome 18 was likely present in isolates that form cluster ii and that this chromosome was subsequently lost or degraded in a subset of isolates, possibly due to recombination with isolates from cluster i and iii (Fig 4D). Our analysis shows that chromosome 18 is also present in the closely related beet downy mildew *Peronospora farinosa* f. sp. *betae* (Fig 4C). The phylogeny suggests a closer relation of *P. farinosa* f. sp. *betae* to *P. effusa* cluster i isolates, for which chromosome 18 is absent (Fig 4D). These observations indicate that chromosome 18 was either present in their last common ancestor and has been recently lost in many of the *P. effusa* isolates, or gained from cluster ii after the differentiation of these *P. effusa* isolates.

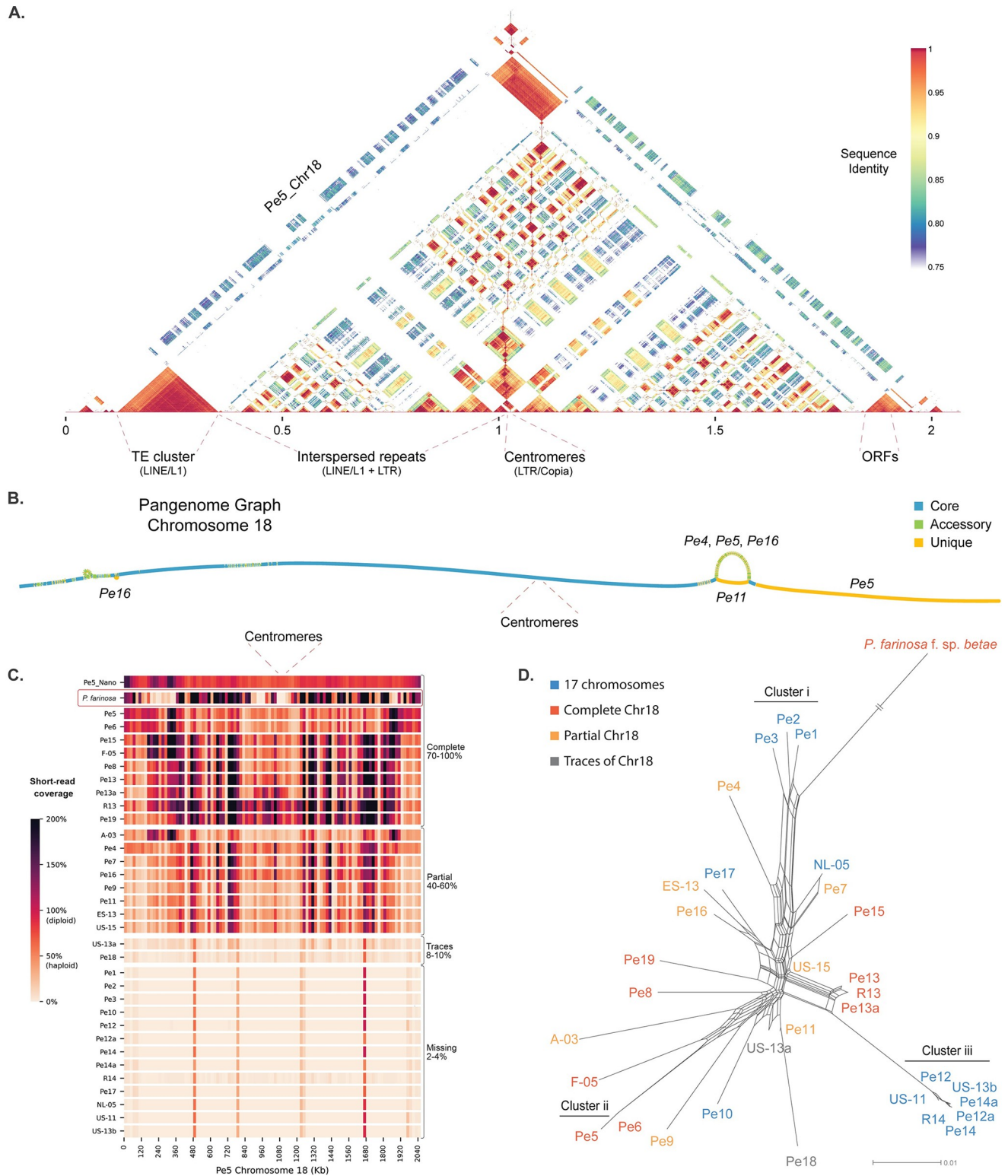


Fig 4. A highly repetitive chromosome is present in a subset of *Peronospora effusa* isolates. **A.** Self-alignment of *Pe5* chromosome 18 in 1.5 kb windows shows the repetitive nature of the accessory chromosome and highlights the two chromosomal arms that are reverse complimentary to each other. **B.** Pangenome graph of accessory chromosome 18 in *Pe4*, *Pe5*, *Pe11*, and *Pe16*. **C.** Presence of chromosome 18 in 32 *P. effusa* isolates and the outgroup *P. farinosa* f. sp. *betae* (indicated with red box) based on the short-read coverage compared with the *Pe5* chromosome assembly. When the coverage is 100% of the average coverage over the whole genome, it indicates that the chromosome is diploid. The top row displays the coverage of the Nanopore long reads

of *Pe5*, while the remaining rows show coverage based on Illumina short-read data for each analysed *P. effusa* isolate. **D.** The Presence of chromosome 18, as determined in C, is shown in the context of the *P. effusa* phylogeny. Similarity between *P. effusa* isolates and the outgroup *P. farinosa* f. sp. *betae* is represented in a neighbor-net phylogenetic network; the branch lengths are proportional to the calculated number of substitutions per site, with the exception of the outgroup where the branch is truncated.

<https://doi.org/10.1371/journal.pgen.1011452.g004>

Extensive variation in repeat-rich regions is caused by rapid repeat expansion and contraction

The most extensive variation in the pangenome graph occurs at gene clusters formed by genes annotated as tRNA genes (Fig 1A), which appear in the pangenome graph as large 'bubbles' of unique and accessory regions (Figs 2A and 5A). These clusters are found in multiple locations across all 17 core chromosomes (33–39 per isolate) and contain the vast majority of all tRNA genes annotated within a genome (around 99.6% of the more than 6,000 annotated in each *P. effusa* isolate). Individual clusters differ in size and can contain between ten to 970 tRNA genes. Notably, the tRNA genes of each cluster are nearly identical (>99.99% identity), while they share less similarity (<70% identity) with other tRNA genes identified outside the clusters. In between the tRNA genes, we discovered conserved regions with a highly similar open reading frames (ORFs; >98% identity), but single nucleotide deletions cause frame shifts changing the size of the predicted encoded proteins (Fig 5B). Interestingly, some of these ORFs are annotated as complete genes of around 300 nucleotides that potentially encode a protein with a signal peptide and an RXLR amino acid motif that is also found in effector candidates.

One of the biggest tRNA clusters can be found on chromosome 6, where a cluster of Arginine tRNAs was initially annotated. In *Pe1*, 570 tRNAs (72 bp in length) in intervals of 645 bp were identified in a single 412 kb region (*Pe1*_Chr6:3.2–3.6 Mb) (Fig 5B). The alignment of all copies of this 717 bp repeated sequence revealed that the tRNA gene is 100% identical while the remainder of the sequences is generally less conserved (94.9–98.9% identity). Within the cluster, we observed multiple subgroups of repeats (50–200 repeats). While within each of these groups individual copies are almost identical (>99% identity), indicating that these expanded recently, we observed significant differences between copies of different subgroups (94–98% identity) and these subgroups are separated by even less conserved sequences (80–90% identity) (Fig 5B). Moreover, in *Pe1*, we identified two cases where the repeated sequence is reversed (Fig 5B ii. and iii.). These inversions along with the changes in sequence identity are indicative for multiple and separate expansions of this 717 bp sequence within this cluster, which is unique to this location in chromosome 6.

In the other five isolates, the same repeated sequence as in *Pe1* can also be found, and in the pangenome graph we observed that the first 80 kb of this region is highly conserved between all isolates. In contrast, however, the remaining region is highly variable with different expansion patterns and large differences in the size of the cluster, ranging from 412 kb in *Pe1* to only 129 kb in *Pe14*. According to the pangenome graph, isolates *Pe16*, *Pe11*, and *Pe4* follow a similar path through the graph as *Pe1*, although *Pe1* has additional unique regions (Fig 5A). These three isolates have a similar but shorter path compared with *Pe1*, including the two inversions (Fig 5B). In contrast, *Pe5* has a unique path, starting from point (v.), which is caused by an expansion that is different compared with other isolates. Similarly, *Pe4* appears as a much shorter unique path, starting from the point of inversion (ii.), because the expansion in *Pe14* is much shorter and the inversions observed in other isolates are not present. These extensive differences between the *P. effusa* isolates suggests that these repetitive elements are continuously expanding and contracting.

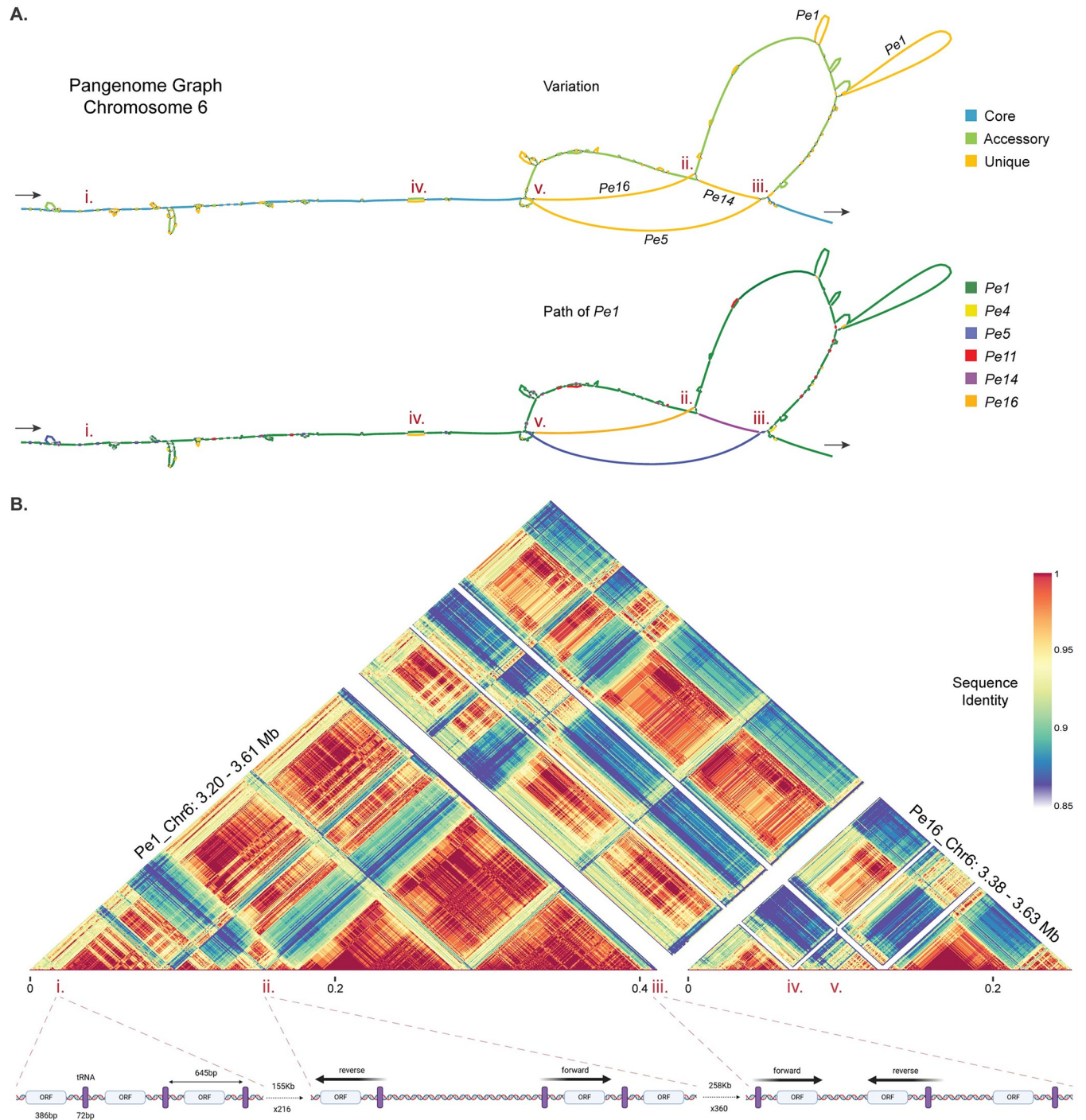


Fig 5. The rapid expansion of SINE-like elements causes extensive genomic variation between *P. effusa* isolates. **A.** Part of chromosome 6, 0.5 Mb in size, is visualised by two pangenome graphs, with the start and end of the graph indicated with arrows. The first graph represents the variation between the isolates and shows core (blue), accessory (green), and unique (orange) regions. In large unique regions, the name of isolate with the specific unique region is indicated. The second graph highlights the path of *Pe1* (dark green), and the unique regions of other isolates are coloured (*Pe4*: yellow, *Pe5*: blue, *Pe11*: red, *Pe14*: purple, *Pe16*: orange). Regions of interest in A and B are highlighted by roman numerals (i-v). **B.** The corresponding region of *Pe1* (Chr6: 3.20–3.61 Mb) and *Pe16* (Chr6: 3.38–3.63 Mb) is aligned in 717-bp windows, which matches the size of the identified SINE-like repetitive element. The alignment is visualised by a heatmap with the colour representing the sequence identity of each alignment from 85% (blue) to 100% (red). A schematic of the observed structure of this cluster based on the annotation of tRNA genes and ORFs is depicted below the alignment and the corresponding regions in the heatmap are indicated with red dashed lines.

<https://doi.org/10.1371/journal.pgen.1011452.g005>

The repetitive tRNA sequences and the coupled ORFs that we uncovered in the *P. effusa* genomes resemble short interspersed nuclear elements (SINEs) that have been characterised in mammals, plants, insects, and in *Phytophthora infestans* where 15 families have been characterised with up to 2,000 copies [64,65]. Their activity and number of families varies greatly with up to a million copies found in the human genome and from 22 families in Amaranthaceae to 200 families in Metazoa [65]. Notably, the rapid and continuous expansion that we observed here resembles the activity of TEs rather than tRNAs and protein-coding genes. SINEs are Class I TEs, propagating by copy-paste mechanism, although they do not encode any proteins and thus are dependent on other TEs for their expansion [66]. Most SINEs are characterised by the presence of a tRNA-like sequence at the 5' terminal region, by a central conserved region, and by a relatively short sequence of 200–700 bp [66]. The taxonomic distribution of a specific SINE family is often clade-specific and does not expand to more distantly related taxonomic groups [65], thus we were not able to match the sequences found in *P. effusa* with any known SINE sequences. Based on these observations, we therefore propose to classify these sequences as SINE-like. To annotate them throughout the genomes, we extracted consensus sequences from each cluster, in total 563 ranging from 284 to 1,070 bp in size, and added them to our *P. effusa* common TE library. The TE annotation with these sequences revealed that 5.4 to 5.9% of the genome assembly in each of the six *P. effusa* strains is composed by these SINE-like sequences, which have undergone recent expansions (Fig 3C and S2 Table), thus explaining the observed differences between the isolates. Although, SINEs have not yet been characterised in their oomycetes genome assemblies, a similarly large amount of tRNA genes was recently reported in the chromosome-level genome of *Phytophthora infestans* [46], suggesting that SINE-like elements are commonly present and particularly active in Peronosporaceae.

Variation in virulence-related genes originate from changes in gene copy-number in gene clusters

By plotting all genes of each isolate in their chromosome position, we observed that most of the gene variation between isolates is concentrated in few, highly variable genomic regions (Fig 6A, clustering of green and yellow bars). To quantify this observation, we searched for genes that share the same functional annotation and are located next to each other in the genome. Of the 1,393 functional groups with at least two genes in *Pe1*, 155 occur in at least one cluster. Additionally, of the 59 functional groups with at least 20 genes, 13 have at least 15% of their genes in a cluster. These observations suggest that physical clustering of functionally related genes is a common phenomenon.

By using the synteny-based, single copy orthogroups on the pangenome graph, we calculated the ratio of substitution rates at non-synonymous and synonymous sites (dN/dS). Overall, genes are highly conserved with 73% being core and 72% having dN/dS values lower than one, indicating that they evolve under negative selection (average dN/dS 0.83). In contrast, genes encoding RXLR and CRN effector candidates show extensive variation with only 50.5% found to be core and less than half (46%) evolve under negative selection and 54% show signs of positive selection (average dN/dS 1.49) (Fig 6B). Interestingly, effectors that are not part of physically co-localizing gene clusters show a distribution of dN/dS values that is comparable to the rest of the genes, and 67% evolve under negative selection (average dN/dS 0.98). In contrast, the dN/dS values of effectors that are clustered in the genome have a bimodal distribution and only 46% evolve under negative selection (average dN/dS 1.81). Thus, most of the variation observed in effectors originates from the presence-absence of genes located in distinct gene clusters (Figs 6C and S9). For example, we observed an effector gene cluster at the beginning of chromosome 6 where *Pe5* has a unique expansion with 45 effectors, with most genes

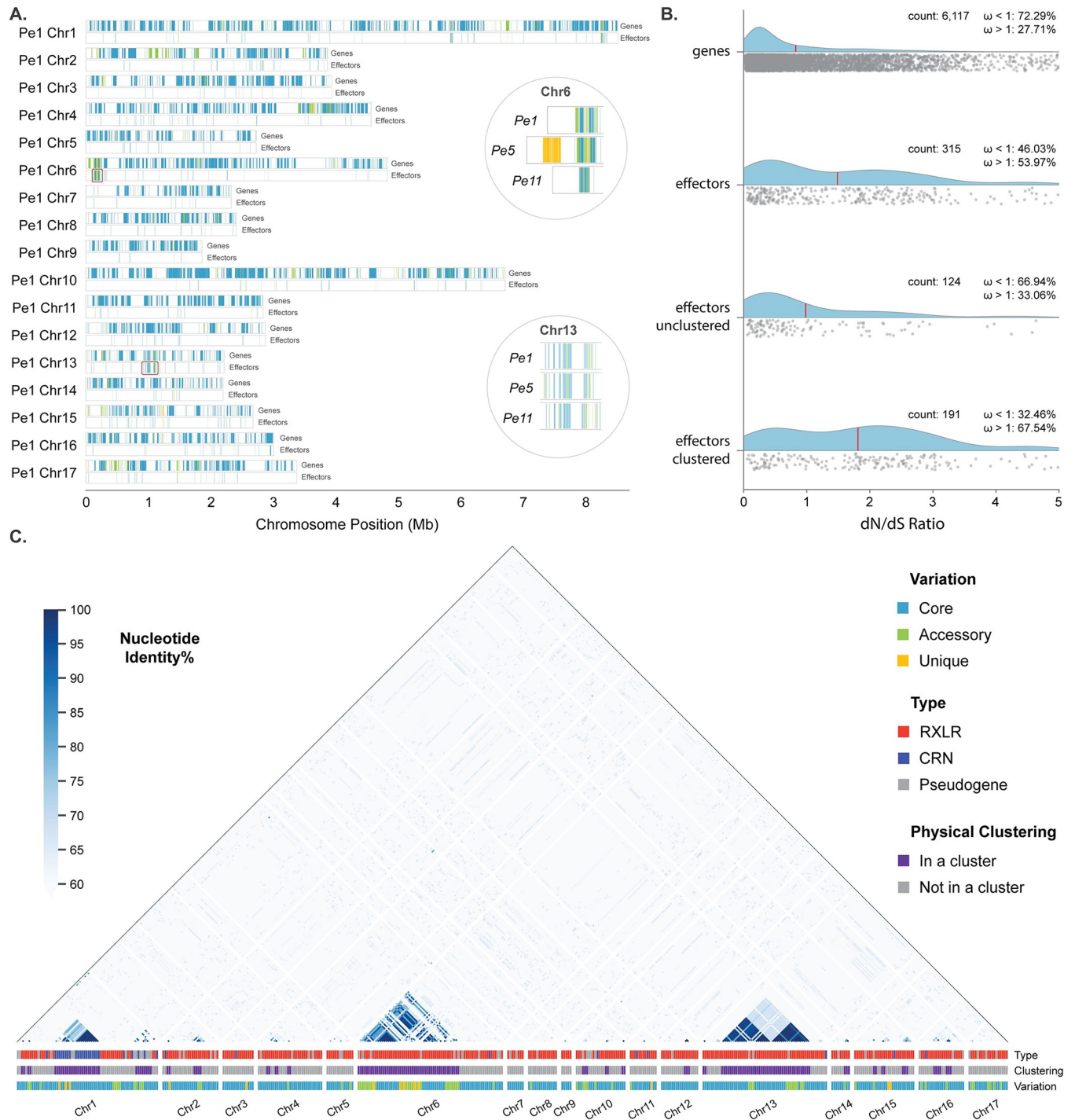


Fig 6. Effector variation is caused by gene copy-number changes in effector gene clusters. **A.** Gene positions in each chromosome of *Pe1* for genes and genes encoding (RXLR and CRN) effector candidates are coloured based on their conservation, blue for core, green for accessory, and orange for unique genes. Two effector clusters are highlighted on chromosome 6 and 13, and their comparisons between *Pe1*, *Pe5*, and *Pe11* is shown. **B.** Gene evolution measured for each orthogroup with dN/dS ratio (ω) shown for all genes, all effectors as well unclustered and clustered effectors. The average value is indicated by a red line. **C.** Heatmap of the nucleotide identity of the 445 annotated candidate effector genes in *Pe1* highlighting the high sequence similarity between clustered effectors. The type of the encoded effector, whether the gene belongs in a cluster, and the variation of the gene based on the pangenome graph is indicated below the heatmap.

<https://doi.org/10.1371/journal.pgen.1011452.g006>

having elevated dN/dS values (average dN/dS 2.15). We also observed two less variable effector gene clusters in the middle of chromosome 13 where *Pe11* and *Pe14* share nine accessory effectors and *Pe14* has two unique effectors (average dN/dS 1.31 and 0.57) (Fig 6A).

Similar to effectors, other genes encoding proteins with functions commonly associated with virulence show variation between isolates that is driven by gene presence-absence changes specifically in gene clusters. Most notably, we identified 21 genes encoding necrosis-inducing proteins (62% core), which are known to induce plant cell death or suppress plant immune responses, but in oomycetes are known to be nontoxic [67–69]. Moreover, we discovered 22 genes related to glycoside hydrolase family 12 in a single cluster in the genome (45% core), known to be present in many bacteria and fungi [70], and 17 papain family cysteine protease genes (76% core) that are known to be involved in virulence or defence, amongst many other functions [71]. When considering genes that share the same functional annotation, those that are physically clustered in the genome are much more similar to each other than genes outside clusters (68% vs 32% protein identity). Around half of the clustered genes are variable between the isolates (52%), while most of the genes outside clusters are conserved (94%) (Figs 6C and S9B). These observations suggest that genes with the same function that are physically clustered in the genome are the result of a gene copy number expansion. As a result, the presence-absence variation we observed in these clusters can be characterised as changes in gene copy-number, which is the main driver of variation.

Changes in gene copy-number in effector clusters are associated with TE activity

The largest group of genes associated with virulence, RXLR effectors, are highly variable between the isolates, which is mostly caused by gene copy-number variation within physical gene clusters. Of the total 601 RXLR orthogroups in the pangenome 47% (283) occur in clusters, but 66% (192) of the 292 variable RXLR orthogroups are found in clusters of two or more effectors. For each isolate, we observed 15 clusters with between three to five gene copies and three RXLR effector clusters with 14 or more copies, one in the beginning of chromosome 6 (Pe1_Chr6: 71–238 kb) with 30 (*Pe11*) to 111 (*Pe5*) copies, and two more at the centre of chromosome 13 (Pe1_Chr13: 908–1076 kb) that encompasses 24 (*Pe1*) to 26 (*Pe16*) copies and 14 (*Pe5*) to 23 (*Pe14*) copies, respectively. These clusters consist of almost identical sequences (>96% protein identity of genes within a cluster), indicating recent gene expansion within these clusters.

The two effector clusters in chromosome 13 are present in an otherwise highly conserved region (Fig 7A). Variation between *P. effusa* isolates is primarily caused by single gene deletions or insertions (Fig 7B and S10). In addition to copy-number variation, gene sequences also differ due to small sequence changes, and we observed ten examples of pseudogenization caused by single nucleotide deletions and subsequent frame shifts. Moreover, all the 114 effector genes found in the second RXLR cluster on chromosome 13 have non-synonymous substitutions in 13 sites, seven that are shared (found in more than six genes), and six that are found in only one or two genes. This variation creates 17 distinct variants of the RXLR protein encoded by this cluster (Fig 7C). Most of this variation is found in the N-terminus and only two are found in the C-terminus, which is assumed to be the functional domain that exerts its activity inside of the plant cell [72]. Despite extensive copy-number differences, only two out of the 23 RXLR effector orthogroups in this cluster have dN/dS values > 1 (average dN/dS = 0.57). When we consider all effector genes in this cluster, only 14 out of the 114 effectors show positive selection (dN/dS > 1) and the sequence of most orthologous effectors remains conserved between the isolates. As a result, the expansion of effector copy-number is the leading cause of variation in this cluster.

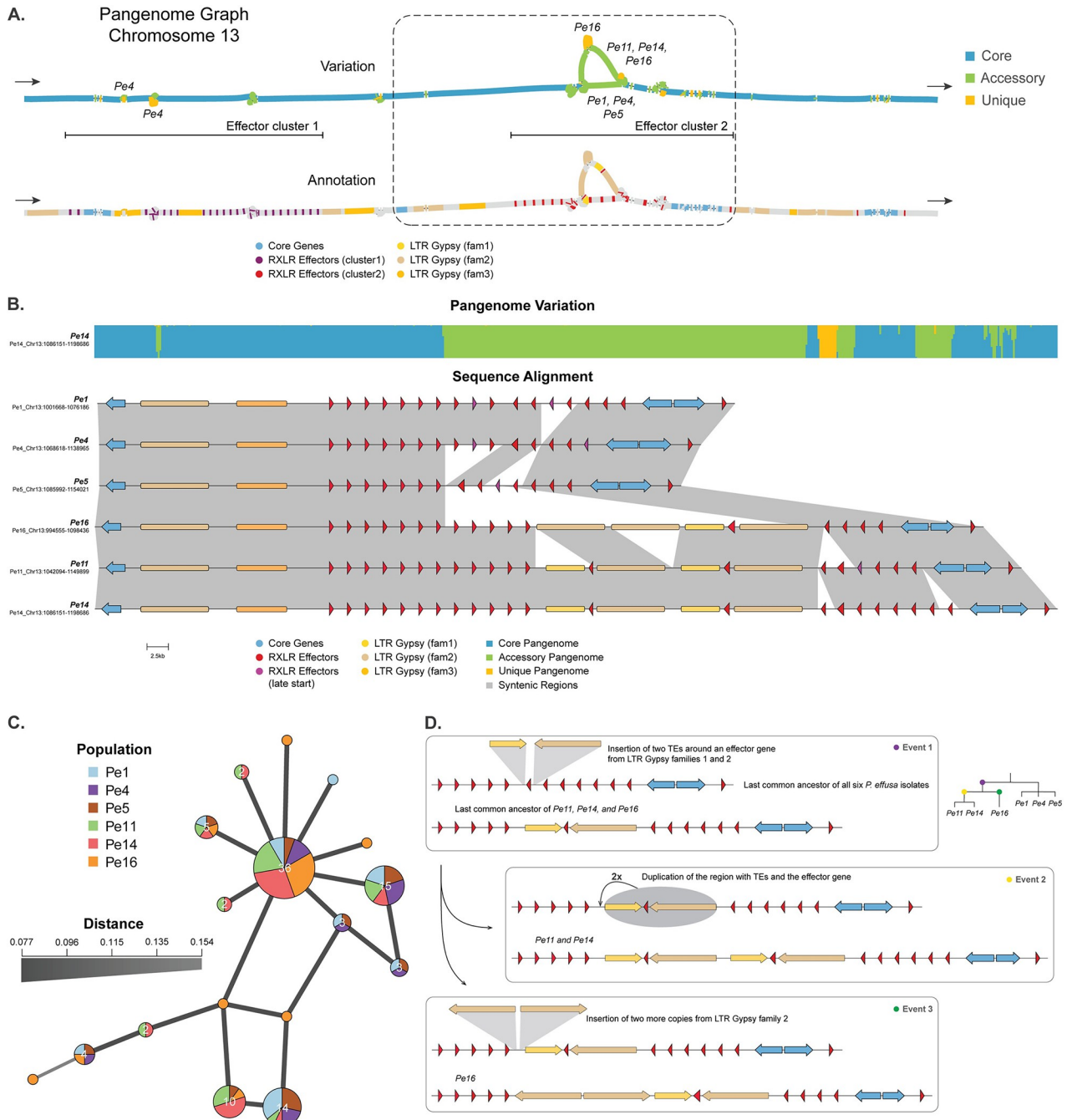


Fig 7. Insertions of transposable elements causes variation in effector clusters and are associated with changes in effector copy-numbers. **A.** Part of chromosome 13, 0.2 Mb in size, is visualised by two pangenome graphs, with the start and end of the graph indicated with arrows. The first graph shows the variation between *P. effusa* isolates with core (blue), accessory (green), and unique (orange) regions. For large accessory and unique regions, the isolates that have these regions are indicated. The second graph visualises the effector genes belonging to the first (purple) and second (red) effector clusters on the graph. The highlighted area is expanded in **B.** Alignment of corresponding chromosome 13 region for the six *P. effusa* isolates containing the second effector cluster. Effector genes (red), pseudogenes (purple), and core genes (blue) are indicated as arrows, and LTR Gypsy repeats (fam1: yellow, fam2: brown and, fam3: orange) are indicated as boxes. The pangenome variation for this region is visualised in a stacked bar plot (top track; core blue, accessory green, and unique orange), and syntenic regions between *P. effusa* isolates are connected with grey ribbons (bottom). **C.** Genotype network indicating the genetic variation between the RXLR effector proteins in the cluster. Each node is an effector genotype present in the second cluster on chromosome 13 and each edge is the protein sequence distance between the nodes. The size of the nodes represents the total number of effectors, which is indicated in each node, and the colours of the nodes represent how each genotype is divided between the isolates. The colour and the width of the edges represent the distance between each node. The length of each edge is not informative and purely aids visualization. **D.** Proposed sequence of events that generated the genetic variation observed in the effector cluster across the six *P. effusa* isolates.

<https://doi.org/10.1371/journal.pgen.1011452.g007>

Larger sequence variation in RXLR clusters is caused by the insertion of TEs, shown as alternative paths in the pangenome graph (Fig 7A). In the first RXLR cluster, we discovered a deletion of an LTR *Gypsy* element and an RXLR gene expansion that occurred specifically in *Pe4*. The second RXLR cluster has extensive variation caused by insertion of LTR *Gypsy* elements in *Pe16*, *Pe11*, and *Pe14* that coincides with copy-number expansion of the RXLR genes. To explain the evolution of this region, we propose the following sequence of events based on the analyses of the pangenome graph. First, in the last common ancestor of isolates *Pe11*, *Pe14*, and *Pe16* two LTR *Gypsy* families were inserted around an effector gene. Second, in the last common ancestor of *Pe11* and *Pe14* the region of the two LTR *Gypsies* surrounding the effector gene were duplicated, thereby creating a new copy of the gene, unique for these two isolates. Third, in the ancestor *Pe16* two more LTR *Gypsy* copies were inserted upstream of the previous event (Fig 7D). The new effector copies in *Pe11* and *Pe14* are identical in sequence and share unique deletions of ten codons in centre of the gene sequence, and a unique nucleotide deletion at the 3' end that causes an early stop. The phylogeny of all the effector genes in the cluster shows that the subgroup of effectors, with downstream effector genes in *Pe11* and *Pe14*, is the origin of these new effector copies (S11 Fig). These observations suggest that the activity of two LTR *Gypsy* families in this region has created the conditions for the expansion of the effectors in *Pe14* and *Pe16*, possibly by homologous recombination. Although this event explains only a part of the overall gene copy-number variation observed in this cluster, it represents a recent example of the TE insertion and gene duplication in these variable gene regions.

Discussion

Despite extensive efforts in disease management and control, filamentous plant pathogens are rapidly evolving and continue to successfully infect their hosts [2]. Pathogen adaptation is thought to be driven by few highly variable genomic regions that harbour most of the known variation in a species [18,22–24]. Traditional comparative genomics methods, based on single reference genome sequences and short-read sequencing data, cannot fully reconstruct these regions, and thus fail to capture their variation and evolution [25]. Here we generated, to our knowledge, the first pangenome graph for an oomycete by incorporating six genetically diverse *P. effusa* isolates to improve structural annotation and to perform in-depth, genome-wide comparisons of their genes, TEs, and effectors. Our analysis revealed a highly conserved genome structure, with 80% of the observed variation being associated with TEs. Genes are typically conserved, but virulence-related genes are highly variable. This variation is mainly caused by gene copy-number changes that occur inside large gene clusters, which appear to be the major mode of genome evolution in *P. effusa*. By constructing a pangenome graph, we've gained invaluable insights into the evolution of this important spinach pathogen. This approach can be applied to many more filamentous plant pathogens and thus presents an exciting new route to study the full extent of pathogen adaptation by precisely documenting genomic variation.

The use of pangenome graphs to compare multiple genomes has recently gained momentum in multiple organisms, including bacteria, animals, fungi, and human [27,73–75]. The alignment of multiple genomes of individuals or strains in a pangenome graph enables researchers to construct a common structural annotation, thus consistently discovering genes for each genome and reducing potential variation that could have been merely caused by different technical interpretations in the separate structural annotation of the individual genomes [76]. For example, we identified in total 1,842 genes that would have been missed in the individual annotations (Note S2). We also demonstrate that the pangenome graph can be used to directly determine the synteny of genes and to create synteny-anchored single-copy orthogroups. By utilising these pangenome based methods of structural annotation, the full

extent of gene variation between the genomes can be discovered, including gene copy-number variations.

Large-scale chromosomal rearrangements between isolates are commonly observed in filamentous fungi such as in *Verticillium dahliae* and *Magnaporthe oryzae*, and these have been proposed to be an important mechanism leading to genetic variation [3,35,77–82]. In contrast, chromosome structure is highly conserved in *P. effusa*, even though we selected phylogenetically distant isolates for our analyses. The conserved chromosome structure is largely shared with other species within the Peronosporaceae such as *Bremia lactucae* and *Peronosclerospora sorghi*, with the exception of few chromosomal fusions [36]. Our analysis indicates that the chromosomes 1 and 10 might also be the result of a fusion of two smaller chromosomes, showing that chromosome fusions indeed do occur in Peronosporaceae. Nevertheless, they show a remarkable chromosomal stability that could result from sexual reproduction that is an essential part of the yearly life cycle of most oomycetes [43,44].

In contrast to the overall conserved chromosome structure in Peronosporaceae (S5A Fig), we discovered the presence of an accessory chromosome in 19 out of 32 *P. effusa* isolates that does not correspond with any of the chromosomes in *B. lactucae* or *P. sorghi* [45,47]. The accessory chromosome is highly repetitive, and the activity of LINES and LTRs appears to underlie the variation observed between *P. effusa* isolates (S8 Fig; Panel Pe5_Chr18). Despite their activity, none of the TE families found on chromosome 18 can be found on any other chromosome, except for four LTR elements located on the chromosomal arms from two LTR families, which experienced recent expansion following the differentiation of the *P. effusa* isolates. Consequently, it seems more likely that this chromosome was recently acquired from a close relative, possibly with a horizontal transfer from *P. farinosa* f. sp. *betae*, rather than deleted from multiple isolates. Accessory chromosomes in asexually reproducing filamentous fungi, like *Fusarium oxysporum*, are often associated with virulence [2,3,77], but since we were not able to identify protein-coding genes on chromosome 18, its physiological significance remains uncertain. Additionally, we observed degradation of this chromosome in many isolates that likely emerged from sexual reproduction between *P. effusa* isolates (Fig 4D) [44], suggesting that sexual reproduction could be contributing to its degradation.

By creating chromosome-level genome assemblies based on long-read data, we were able to discover the TE content and its variation in *P. effusa*. Incomplete genome assemblies typically miss regions of recent TE activity [19], which for example explains the discrepancy in genome sizes compared with previous *P. effusa* reference genomes (28 Mb larger) [42,63]. Our pangenomic comparison reveals that the TE expansion in *P. effusa* is, in many cases, more recent than the differentiation of the here analysed isolates. Most notably, SINE-like sequences have recently expanded and contribute to the extensive variation between the *P. effusa* isolates. We additionally observed variation in nanopore coverage in SINE-like cluster in *Pe1* and *Pe5*, which indicates high heterozygous variation in contrast to the rest of the genome (drops and peaks in coverage) or repeat collapse (peaks in coverage) (S3 Fig, *Pe1* chr6 and chr11, *Pe5* chr2, chr3, and chr6). TE expansion in other oomycete genomes has been thought to play a similar role in their genome expansion and variation [16,24,35]. Notably, while other TE families have been expanding and recent copies can be found on multiple chromosomes, each SINE-like family in *P. effusa* is specific to a single chromosomal location, and expansion creates large repetitive regions of almost identical sequences. Since each of these elements are exclusive to a specific region of a chromosome, it is likely that they are not able to mobilise. As a result, the observed variation can be attributed to non-allelic homologous recombination. Additionally, it is worth noting that these elements are commonly annotated as tRNA. For example, the recent annotation of more than 7,000 tRNA genes in *Phytophthora infestans* suggests that these SINE-like elements are active and play an important role in oomycete genome evolution [46].

The pangenome graph uncovered extensive and recent changes in gene copy-number, especially for RXLR and CRN effectors as well as for other virulence-related genes. While the molecular mechanism has yet to be discovered, recent TE activity is often observed close to genes that display gene copy-number variation. This phenomenon has also been described in *Phytophthora*, where effector gene copy-number variation was shown for RXLR genes *Avr1a* and *Avr3a* in *Phytophthora ramorum* [83] and RXLR and CRN genes in *Phytophthora sojae* [84], and these copy-number changes are thought to impact pathogen fitness [83,84]. It is also well established that NLP genes have recently expanded in oomycetes by gene duplications and that these genes are organised in clusters [68]. In *P. effusa* 15/18 NLPs are organised in two clusters in which all the variation between NLPs can be observed. Interestingly, when comparing the orthologous genes, we observed that clustered effectors more often evolve under positive selection than unclustered effectors, suggesting that locally copied effectors rapidly generate novel genotypes. Thus, dynamic gene clusters, embedded in otherwise highly conserved chromosomes, evolve by extensive copy-number changes and serve as cradles for genomic variation in oomycetes. Given the frequency and the emergence of novel *P. effusa* races, copy-number variation of virulence genes and diversification of novel gene copies, likely plays an important role in the adaptation of *P. effusa*. Adding more and more closely related *P. effusa* isolates to the pangenome will therefore be essential to associate effector variation with virulence and the capacity to break deployed spinach resistance traits.

Pangenome graphs will most likely transform the field of comparative genomics, resulting in more accurate and in-depth analysis of the variation of a large collection of diverse genomes [25,29]. Especially chromosome-level genome assemblies will be essential to achieve the necessary resolution to uncover and study highly variable regions that are enriched for virulence-related genes [3,36]. Beyond expanding the analysis by adding more isolates, a step further would be to create a pangenome graph that incorporates fully phased diploid or polyploid genome assemblies, thus uncovering possible variation and recombination between haplotypes [85]. For example, oomycetes are mostly diploids, but there are various reports of whole-genome duplication in *Phytophthora betacei*, and aneuploidy in *Phytophthora capsici* and in *Phytophthora cinnamomi* [86–90]. We therefore anticipate that pangenomic approaches will be instrumental to uncover the full extent of genomic variation in filamentous plant pathogens [27,91], which, in turn, will unveil new theories about their emergence and evolution, impacting our ability to predict and manage plant diseases.

Materials and methods

Peronospora effusa infection on soil-grown spinach and spore isolation

Spinach plants were sown in potting soil (Primasta, the Netherlands) and kept under long-day conditions (16-hour light, 21 °C). Two to three weeks after germination, the spinach plants were inoculated with *P. effusa* by spraying them with *P. effusa* spores suspended in water using a spray gun. Following inoculation, we placed the plants under 9-hour light and 16 °C, and the lids of the plastic trays were sprayed with water and covered to keep the plants humid. After 24 hours, the vents on the lids were opened. The lids of the boxes were again sprayed with water and the vents were closed 7–10 days after inoculation, creating a humid environment that promotes the sporulation of *P. effusa*.

To harvest *P. effusa* spores for Oxford Nanopore sequencing, we collected leaves with sporulating *P. effusa* from spinach plants and placed them in a glass bottle with tap water. The spores were brought into suspension by shaking the bottle vigorously. Soil and other large contaminants were removed by filtering the spore suspension over a 50- μ m nylon mesh filter (Merck Millipore, USA). To remove small biological contaminants, the remaining filtrate was filtered 11- μ m nylon mesh filter (Merck Millipore, USA) using the Merck™ All-Glass Filter

Holder (47 mm) and a vacuum pump, resulting in the spores which remaining on top of the filter and contaminants washing through the filter. The spores were washed several times, scraped off the filter, and kept at -80°C .

High-molecular weight DNA extraction protocol

To isolate high-molecular weight (HMW) DNA, the collected *P. effusa* spores were ground to a fine powder in liquid nitrogen together with 0.17–0.18mm glass beads. The ground spores were washed in cold Sorbitol solution (100 mM Tris-HCl pH 8.0; 5 mM EDTA pH 8.0; 0.35 M Sorbitol, 1% PVP-40, 1% β -mercaptoethanol, pH 8.0). The tissue was lysed by incubation in extraction buffer (1.25 M NaCl, 200 mM Tris.HCl, pH 8.5, 25 mM EDTA, pH 8.0, 3% CTAB, 2% PVP-40, 1% β -mercaptoethanol) and incubated with proteinase K and RNase A for 60 minutes at 65°C , and throughout the incubation mixed by gentle inversion. The samples were centrifuged to pellet and remove the debris. HMW DNA was further purified by phenol/chloroform/IAA and chloroform/IAA extraction, another RNase treatment, extraction with phenol/chloroform/IAA and chloroform/IAA and isopropanol precipitation. DNA concentration and integrity were determined using Nanodrop, Qubit, and Tapestation.

Genome sequencing using Oxford Nanopore

We obtained long-read sequencing data for six *P. effusa* isolates (*Pe1*, *Pe4*, *Pe5*, *Pe11*, *Pe14* and *Pe16*) with Oxford Nanopore sequencing technology (Oxford Nanopore, UK) at the USEQ sequencing facility (the Netherlands). The ligation-based sequencing kit from Oxford Nanopore was used for library preparation (ONT—SQK-LSK109; Oxford Nanopore, UK) following the manufacturer's protocol. We used a Nanopore MinION flowcell (R10) for real-time sequencing and base-calling of the raw sequencing data was performed using Guppy (version 4.4.2; default settings). The raw long-read sequencing data were checked for contamination using Kraken2 (version 2.0.9; default settings) [92].

Genome sequence using Hi-C

We obtained Hi-C sequencing data for six *P. effusa* isolates (*Pe1*, *Pe4*, *Pe11*, *Pe14*, and *Pe16*). Spores were crosslinked in formaldehyde 1% and incubated at room temperature for 15 minutes with periodic mixing. Glycine was added to a final concentration of 125 mM, followed by incubation for another 15 minutes. The spores were pelleted by centrifugation and washed with Milli-Q water. Glass beads were added, and the samples were vortexed. Crosslinked samples were pelleted by centrifugation, kept at minus 80 and send to Phase genomics (USA) for sequencing.

Transcriptome sequencing using RNAseq

RNA was extracted from spores or infected spinach leaves with a Kingfisher System using the MaxMag Plant RNA isolation Kit. The sequencing was done at USEQ (the Netherlands), using the Truseq RNA stranded polyA library prep and the samples were sequenced on the Next-Seq500 platform with 2 x 75 bp mid output (210 M clusters).

Genome assembly

We used the long-read Oxford Nanopore sequencing data to produce chromosome-level genome assemblies for six *P. effusa* isolates. The reads were corrected, trimmed, and assembled using Canu (version 2.3) [93] with the following command:

```
canu -nanopore  $\{\text{input\_reads}\}$  genomeSize = 58M -d  $\{\text{output\_dir}\}$  -p  $\{\text{sample\_name}\}$   
corOutCoverage = 40 mhapMemory = 100g corMhapFilterThreshold = 0.000000002
```

```
mhapBlockSize = 500 ovlMerThreshold = 500 corMhapOptions = "--threshold 0.80--num-hashes 512--num-min-matches 3--ordered-sketch-size 1000--ordered-kmer-size 14--min-olap-length 800--repeat-idf-scale 50"
```

To remove contigs that come from possible contaminants, uncollapsed haplotypes, and other assembly artefacts, we run a pipeline of publicly available tools to filter and curate draft genome assemblies. First, uncollapsed contigs were removed with Purge Haplotigs (version 1.1.1; -a 90) [94]. The taxonomy of the remaining contigs was determined with CAT contigs (version 5.2; default settings) [95] and was visualised together with the Nanopore read coverage and QC content with Blobtools (version 2.3.3; default settings). The contigs classified as contamination were removed, while contigs classified as oomycete, Peronosporaceae, *Peronospora*, and the unclassified contigs were retained. The mitochondrial contigs were removed based on their characteristically different GC content (22% GC of the mitochondrion vs 44–52% of the nuclear genome for *P. effusa*). Finally, genome assembly metrics were measured with QUAST (version 5.0.2) [96].

Scaffolding assemblies with Hi-C data and closing gaps

The curated assemblies were further scaffolded to full chromosomes by using Hi-C short-read data that were aligned to the reference assemblies with Juicer (version 1.6; default settings) [97]. Based on this alignment, the pairwise Hi-C contacts along the contigs were visualised on a heatmap generated with 3D-dna (version 180922, -r 0 -e) [98]. Based on the heatmap, the contigs were manually scaffolded to 17 or 18 chromosomes using Juicebox (Normalisation: balanced, Resolution: 50Kb) [99].

Gaps of the scaffolded chromosomes were closed with FinisherSC (version 2.1; default settings) [100]. The scaffolded assemblies were corrected for single nucleotide polymorphisms using Illumina short-reads with four rounds of Pilon (version 1.23;—diploid,—fixbases) [101].

Transposable element annotation

To create a combined transposable element (TE) library for all *P. effusa* isolates, we used EarlGrey (version 2.0) [102], with dependencies RepeatMasker (version 4.1.2) [103], RepeatModeler (version 2.0.2) [104], and DFAM (version 3.6) [105]. First, the EarlGrey pipeline was run on the genome of *Pe1* with option '-r 2759', specifying the subset of eukaryotic sequences in the DFAM database as reference TE library. The TE library produced was then expanded by using it as an input for the EarlGrey pipeline, rather than the DFAM library, and recursively running it with additional *P. effusa* isolates (S13 Fig). The here identified SINE-like clusters were initially inconsistently annotated by EarlGrey. To remove these consensus sequences, TE consensus sequences with hits to tRNAs of 0.001 e-value or lower, at least 30 matching positions and a coverage and identity of at least 80% were discarded. To ensure that duplicated genes are not annotated as TEs, the final combined TE library was then filtered for TE families that overlapped with gene annotations that have also RNAseq coverage on the *P. effusa* assembly, thus removing 21 consensus sequences from the library. To complete the TE library, we manually added 563 SINE-like consensus sequences from the *Pe1* genome. This combined and filtered TE library was then used to annotate and soft-mask the genomes using RepeatMasker (version 4.1.2 -e rmbblast -xsmall -s -nolow).

Genome annotation

The soft-masked genomes and the RNAseq short-read data were used for structural gene prediction and functional annotation with the funannotate pipeline (version 1.8.7) [106] with the following commands:

```

funannotate train -i ${ref_fasta} -o ${output_dir}—cpus ${threads}—memory 150G -l
${R1} -r ${R2}—species "Peronospora effusa"—isolate ${sample}—stranded no—jaccard_clip
—max_intronlen 600
funannotate predict -i ${ref_masked} -o ${output_annotate}—cpus ${threads} -d ${db}—
name ${sample}—species "Peronospora effusa"—isolate ${sample}—max_intronlen 600—bus-
co_db stramenopiles_odb10—organism other
funannotate update -i ${output_dir}—cpus ${threads}—max_intronlen 600—alt_tran-
scripts 0.3
funannotate iprscan -i ${output_dir} -m local—cpus ${threads}—iprscan_path ${iprscan}
funannotate annotate -i ${output_dir}—cpus ${threads} -d ${db}—busco_db
stramenopiles_odb10

```

The assembled and structurally annotated genomes were visualised with Circos (v. 0.69–8) [107] (Fig 1A).

Secretome and effector prediction

In addition to genes annotated by funannotate, we also extracted all open reading frames longer than 70 amino acids encoded in the repeat-masked genome using ORFfinder (v. 0.4.3, -ml 210 -s 0). We filtered ORFs when they overlapped with the annotated genes using bedtools intersect (version 2.30.0, default settings) [108].

For the prediction of the secretome of *P. effusa* races, we used the Predictor pipeline (v. 1.2.6, default settings) [109]. We extracted the list of all sequences predicted to contain a signal peptide by at least one of the tools included in the pipeline (SignalP3.0, SignalP4.0, SignalP5.0, SignalP6.0, DeepSig, and Phobius) [110–115]. From those, the ones containing multiple transmembrane domains according to Phobius or TMHMM [116] were excluded from the secretome.

The secreted proteins were then screened to detect the presence of the conserved motifs described in RXLR and Crinkler oomycete effectors [48]. To perform regular expression searches, the EffectR package for R [117] was used to detect the following patterns in the first 100 amino acids after the signal peptide: i) a divergent version of the RXLR-EER complete motif ([RQGH]XL[RQK]-[ED][ED][RK] [16], and ii) canonical and divergent versions of the RXLR motif alone (RXLR or [RQGH]XL[RQK]). For the Crinklers, we used regular expressions to detect the occurrence of canonical LFLAK, a degenerated version with maximum two allowed changes in the motif (established as L[FYRL][LKF][ATVRK][KRN]) [118], and the HVLV motif.

We also performed sequence profile searches using HMMER v3.3 [119]. To create the sequence profile for the WY domain [120], the sequences described were aligned using MAFFT v7.453 (ginsi) [121], and an HMM profile was built from the alignment using hmmbuild [119]. This profile was used to screen the secreted proteins using hhmsearch (—incE 10). A similar approach was followed to create profiles to search for the RXLR(-EER) motifs using the 253 reviewed RXLR effectors deposited at the UniProt database (extracted with the search string “family:rxlr effector family” AND reviewed:yes” in January 2022) [122]. Three different profiles were built from the proteins annotated to contain a RXLR motif, an EER motif, or both RXLR-EER. To search for Crinklers, we used an HMM profiles for the LFLAK and the DWL domains described by Armitage *et al.* (2018).

The list of RXLR candidates includes all secreted proteins displaying a canonical or divergent RXLR-EER motif, the canonical RXLR, the WY domain, or a divergent RXLR, in case the protein was also found by least one of the Uniprot HMM profiles. For Crinklers, all proteins identified with the LFLAK/DWL regular expressions or HMM profiles were included in the list of candidate effectors.

Pangenome graphs and common annotation

A pangenome graph was built per chromosome and then merged in a final pangenome graph following the Minigraph-Cactus Pangenome Pipeline, HPRC Graph (step-by-step): Splitting by Chromosome (version 2.6.4,—filter 0—vcf full—gfa full) [31]. This results in one, unfiltered pangenome graph that was used in the downstream analysis. Pangenome graphs from minigraph were visualized with bandage (version 0.8.1) [123].

The hal output of the pangenome graph, the gene annotation, the annotated protein sequence, and the RNAseq coverage for each *P. effusa* isolate were used as input to collectively reannotate genes with the Comparative-Annotation-Toolkit (version 2.2.1,—augustus—augustus-cgp—assembly-hub—filter-overlapping-genes) [76]. Genes that were not characterized as protein-coding or had no alternative source transcripts were removed from the annotation (gene_biotype = protein_coding).

Synteny-based gene orthogroups

The pangenome graph was annotated with the genes of each isolate, by translating the gene locations in each genome to their corresponding locations in the pangenome graph. To create gene orthogroups based on their synteny on the pangenome graph, we computed the overlaps between gene annotations in the pangenome per chromosome. This information was then used to group genes into orthogroups using a 0.6 distance cutoff (this cutoff can be adjusted by the user). In total, this resulted in 12,379 single-copy orthogroups, which we characterised as core, accessory, or unique based on the number of isolates represented in each orthogroup. The code for this computational pipeline is available on GitHub.

We then calculated the percentage of orthogroups with gene presence-absence variation for each functional group as the sum of unique and accessory orthogroups. When gene presence-absence variation is found in gene clusters of recently copied genes, we consider this as strong evidence for gene copy-number variation between the isolates.

Saturation plots

Saturation plots are created to visualise the variation between the isolates on the nucleotide (to compare the whole genome or TEs) or orthogroup level (to compare all genes or any subset of genes). All possible comparisons between isolates are calculated for combination of one to all six isolates. For each comparison, each nucleotide or single copy orthogroup was characterised based on the number of isolates represented, as core (all isolates in the comparison), unique (only one isolate for comparisons of two isolates or more), or accessory. The line for each category (core, accessory, and the sum of accessory and unique) is drawn on the mean for each combination (one to six) and the range of all calculations are shown by the shadow behind each line. The code for creating these plots is part of the computational pipeline that is available on GitHub.

Gene evolution

The protein sequences of each single copy orthogroup was aligned with MAFFT (version 7.453,—auto—anysymbol) [121]. This alignment was used to guide the nucleotide sequence alignment with PhyKIT thread_dna (version 1.11.7, default settings) [124]. To calculate the dN/dS ratio, we used the python package dnDs (version 2.1). The calculation was made pairwise for each orthogroup, excluding those that would result in a division by zero or with dN/dS score higher than five to avoid saturation of substitutions [125].

Splits tree

We performed variant calling, using the genome assembly of *Pe1* as a reference, and the short-read data of 32 *P. effusa* isolates. The short reads were aligned using bwa-mem2 (version 2.2.1, default settings) [126] and a joint VCF file was generated with both variant and invariant sites with GATK (version 4.4.0.0, GenotypeGVCFs -all-sites) [127]. The single nucleotide variants were transformed into a distance matrix with PGDSpider (version 2.1.1.5) [128], which was then used to construct a decomposition network using the Neighbor-Net algorithm with SplitsTree (version 4.17.0) [129]. We calculated the branch confidence of the network using 1,000 bootstrap replicates.

Effector genotype network

We collected from all six *P. effusa* isolates the effector protein sequences from the effector cluster in chromosome 13 and we performed variant calling using GATK (version 4.4.0.0). The variant sites were used to create a genotype network using the R package clusterPoppr (version 2.9.4, default settings) [130].

Visualisations

Using ModDotPlot (version 0.7.2) [131], we visualised the repeat in chromosome 18 (-id 75 -k 11 -r 3000) and the repetitive region in chromosome 6 (-id 85 -k 11 -r 500). Clinker (version 0.0.28) [132] was used to visualise the effector cluster on chromosome 13, by aligning the gene protein sequences of six *P. effusa* isolates in that region. All other visualisations were created in python using matplotlib and seaborn [133].

Supporting information

S1 Fig. Six isolates were selected to capture the genomic variation of *Peronospora effusa*. Neighbor-net phylogenetic network of *P. effusa* isolates that cover all known isolates based on a distance matrix of genome-wide nucleotide differences, containing 260,616 biallelic sites. The branch lengths are proportional to the calculated number of substitutions per site. The parallel edges connecting different isolates indicate conflicting phylogenetic signals. The six selected isolates are from six distinct races and are indicated with red circles.
(TIF)

S2 Fig. Nanopore sequencing for *Pe1* isolate resulted in high quality long-reads. A. Weighted histogram of read lengths after log transformation. B. Phylogenetic classification of reads based on Kraken2 (Wood et al., 2019) and plotted with MultiQC (Ewels et al., 2016). All the reads classified as oomycote belong to *P. effusa*.
(TIF)

S3 Fig. The genome assemblies of *Peronospora effusa* race 1 (*Pe1*) and race 5 (*Pe5*) are visualised in a circular plot. Individual tracks, starting from the outside to the inside: i) 17/18 chromosomes are shown in different colours, a grey rectangle point the location of the centromeres. ii) Line graph shows the coverage of repeat (orange) and gene (blue) content summarized in non-overlapping 20 kb windows. iii) Lines indicate the position of RXLR (red), CRN (green), and tRNA (black) genes. iv) Inverted line graph shows the nanopore average coverage of nanopore reads to the genome assembly. Coverage of 1 indicates a diploid coverage, 0.5 haploid coverage, and coverage higher than one indicates an underrepresentation of a repetitive region.
(TIF)

S4 Fig. Genome-wide Hi-C contact map for the 18 chromosomes of the *Pe5* assembly. Hi-C heatmap displays the spatial interactions between chromosomes in the nucleus. Chromosome boundaries are indicated with blue lines. The observed interactions indicate a Rabl chromatin configuration with telomeric and centromeric regions inside the nucleus.
(TIF)

S5 Fig. Whole-genome alignments of genomes based on sequence similarity and on relative position of protein-coding genes. A. Comparison of six oomycete chromosome-level genome assemblies revealing conserved chromosome structure with a few chromosomal fusions. B. Comparison of our six chromosome-level genome assemblies for *P. effusa* revealing highly conserved chromosome structure with two rearrangements in chromosome in *Pe11* and chromosome 17 in *Pe5*.
(TIF)

S6 Fig. The structure and variation of the pangenome graph. A. Histogram of the node degree, i.e. the number of connections, of each node of the pangenome graph. B. Histogram of the sequence length of each node of the pangenome graph up to a max length of 5 kb. C. Bar plot of the percentage of the genome size that is core, accessory, or unique for each isolate and the pangenome. D. Bar plot of the total size of pangenome graph nodes that belong from one to six isolates.
(TIF)

S7 Fig. The transposable element landscape based on Kimura distances for *Pe1* and *Pe5*. Kimura distance is the measure of divergence between individual TE copies and the corresponding TE consensus sequence (Kimura, 1980), i.e., the lower the Kimura distance, the more similar the copy is to the consensus and thus the more recent it was most likely copied.
(TIF)

S8 Fig. The transposable element landscape based on Kimura distances split up for each chromosome of *Pe5*. Kimura distance is the measure of divergence between individual TE copies and the corresponding TE consensus sequence (Kimura, 1980), i.e., the lower the Kimura distance, the more similar the copy is to the consensus and thus the more recent it was most likely copied.
(TIF)

S9 Fig. Virulence-related proteins are encoded by highly variable clusters of recently copied genes. A. Heatmap of the average protein identity for each cluster and the average protein identity for all unclustered proteins. B. Heatmap of the percentage of proteins that are core or variable for each protein cluster and for all unclustered proteins.
(TIF)

S10 Fig. Alignment of the corresponding chromosome 13 region for six *P. effusa* isolates containing the first effector cluster. A. Part of chromosome 13, 0.2 Mb in size, is visualised by two pangenome graphs, with the start and end of the graph indicated with arrows. The first graph shows the variation between *P. effusa* isolates with core (blue), accessory (green), and unique (orange) regions. For large accessory and unique regions, the isolates that have these regions are indicated. The second graph visualises the effector genes belonging to the first (purple) and second (red) effector clusters on the graph. The highlighted area is expanded in B. Effector genes (red), pseudogenes (purple), and core genes (blue) are indicated as arrows, and LTR Gypsy repeats (fam3: orange, fam2: brown) are indicated as boxes. The pangenome variation for this region is visualised in a stacked bar plot (core blue, accessory green, and unique orange) and syntenic regions between *P. effusa* isolates are connected with grey

ribbons.
(TIF)

S11 Fig. Effector gene phylogeny reveals the origin of the duplicated genes. Nucleotide phylogeny of the 114 RXLR effector genes and five pseudogenes for the six *P. effusa* isolates in the second cluster of chromosome 13. The neighbour joining tree was built with IQ-TREE (v. 1.6.12) (Minh et al., 2020) and visualised with iTOL (v. 6.9) (Letunic & Bork, 2024). The location on the tree of the two genes downstream of the duplicated genes is indicated with arrows.
(TIF)

S12 Fig. K-mer analysis confirms the completeness of the genome assembly in *Peronospora effusa* race 1 (*Pe1*). The short-read data of *Pe1* were split into K-mers (size 27b) and their frequency is plotted using KAT (v. 2.4.2) (Mapleson et al., 2017). The K-mers are mapped to the genome assembly of *Pe1* and the colour represents the number of copies that they are represented in the assembly.
(TIF)

S13 Fig. Pangenome graph based on the joint analyses of the whole genome assembly of all six *P. effusa* isolates. The pangenome graph was produced using the full genome assemblies of our six *P. effusa* isolates in a joint analysis, rather than generating separate pangenome graphs for each chromosome individually, which were then merged. Nodes are randomly coloured.
(TIF)

S14 Fig. Sequence of publicly available tools used for the separate structural genome annotation of the six *Peronospora effusa* isolates. These structural annotations were then overlapped with the pangenome graph for reference-free genome annotation and whole-genome comparisons (Fig 2).
(TIF)

S1 Table. A. Nanopore sequencing data used to assemble the genomes of *Peronospora effusa*. B. Hi-C sequencing data used to assemble the genomes of *Peronospora effusa*.
(XLSX)

S2 Table. A. Overview of the chromosome level genome assemblies of *Peronospora effusa*. B. Overview of the genome annotation of *Peronospora effusa*. C. Overview of the repeat annotation of *Peronospora effusa*. D. Overview of the tRNA annotation of *Peronospora effusa*.
(XLSX)

S3 Table. Details about the origin of the here sequenced *Peronospora effusa* isolates.
(XLSX)

S1 Note. Chromosome-level genome assemblies for six *Peronospora effusa* isolates.
(DOCX)

S2 Note. Pangenome graph of six *Peronospora effusa* isolates enables comprehensive study of genome variation.
(DOCX)

Acknowledgments

We acknowledge the Utrecht Sequencing Facility (USEQ) for providing sequencing service and data. USEQ is subsidized by the University Medical Center Utrecht and The Netherlands X-omics Initiative (NWO project 184.034.019).

Author Contributions

Conceptualization: Guido Van den Ackerveken, Ronnie de Jonge, Michael F. Seidl.

Data curation: Petros Skiadas, Sofía Riera Vidal, Joris Dommissse.

Formal analysis: Petros Skiadas, Sofía Riera Vidal, Joris Dommissse, Melanie N. Mendel, Joyce Elberse.

Funding acquisition: Guido Van den Ackerveken, Ronnie de Jonge, Michael F. Seidl.

Investigation: Petros Skiadas.

Methodology: Sofía Riera Vidal, Joris Dommissse, Melanie N. Mendel, Joyce Elberse, Michael F. Seidl.

Project administration: Guido Van den Ackerveken, Ronnie de Jonge, Michael F. Seidl.

Software: Petros Skiadas, Sofía Riera Vidal, Joris Dommissse.

Supervision: Guido Van den Ackerveken, Ronnie de Jonge, Michael F. Seidl.

Visualization: Petros Skiadas.

Writing – original draft: Petros Skiadas, Michael F. Seidl.

Writing – review & editing: Melanie N. Mendel, Guido Van den Ackerveken, Ronnie de Jonge, Michael F. Seidl.

References

1. McMullan M, Gardiner A, Bailey K, Kemen E, Ward BJ, Cevik V, et al. Evidence for suppression of immunity as a driver for genomic introgressions and host range expansion in races of *Albugo candida*, a generalist parasite. *Elife*. 2015; 2015: e04550. <https://doi.org/10.7554/eLife.04550> PMID: 25723966
2. Hartmann FE, Sánchez-Vallet A, McDonald BA, Croll D. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *The ISME Journal* 2017 11:5. 2017; 11: 1189–1204. <https://doi.org/10.1038/ismej.2016.196> PMID: 28117833
3. Westerhoven AC van, Aguilera-Galvez C, Nakasato-Tagami G, Shi-Kunne X, Parte EM de la, Chavarro-Carrero E, et al. Segmental duplications drive the evolution of accessory regions in a major crop pathogen. *New Phytologist*. 2024; 6: 42163. <https://doi.org/10.1111/NPH.19604> PMID: 38402521
4. Bourguet D, Delmotte F, Franck P, Guillemaud T, Reboud X, Vacher C, et al. Combining selective pressures to enhance the durability of disease resistance genes. *Front Plant Sci*. 2016; 7: 225758. <https://doi.org/10.3389/fpls.2016.01916> PMID: 28066472
5. Zaccaron AZ, Stergiopoulos I. Analysis of five near-complete genome assemblies of the tomato pathogen *Cladosporium fulvum* uncovers additional accessory chromosomes and structural variations induced by transposable elements effecting the loss of avirulence genes. *BMC Biology* 2024 22:1. 2024; 22: 1–21. <https://doi.org/10.1186/S12915-024-01818-Z> PMID: 38281938
6. Fayyaz A, Robinson G, Chang PL, Bekele D, Yimer S, Carrasquilla-Garcia N, et al. Hiding in plain sight: Genome-wide recombination and a dynamic accessory genome drive diversity in *Fusarium oxysporum* f.sp. *ciceris*. *Proc Natl Acad Sci U S A*. 2023; 120: e2220570120. https://doi.org/10.1073/PNAS.2220570120/SUPPL_FILE/PNAS.2220570120.SD09.XLSX
7. Corkley I, Fraaije B, Hawkins N. Fungicide resistance management: Maximizing the effective life of plant protection products. *Plant Pathol*. 2022; 71: 150–169. <https://doi.org/10.1111/PPA.13467>
8. Miller ME, Nazareno ES, Rottschaefer SM, Riddle J, Santos Pereira D Dos, Li F, et al. Increased virulence of *Puccinia coronata* f. sp. *avenae* populations through allele frequency changes at multiple putative Avr loci. *PLoS Genet*. 2020; 16: 1–30. <https://doi.org/10.1371/journal.pgen.1009291> PMID: 33370783
9. Mohd-Assaad N, McDonald BA, Croll D. The emergence of the multi-species NIP1 effector in *Rhynchosporium* was accompanied by high rates of gene duplications and losses. *Environ Microbiol*. 2019; 21: 2677–2695. <https://doi.org/10.1111/1462-2920.14583> PMID: 30838748

10. Möller M, Stukenbrock EH. Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology* 2017 15:12. 2017; 15: 756–771. <https://doi.org/10.1038/nrmicro.2017.76> PMID: 28781365
11. Rovenich H, Boshoven JC, Thomma BPHJ. Filamentous pathogen effector functions: of pathogens, hosts and microbiomes. *Curr Opin Plant Biol.* 2014; 20: 96–103. <https://doi.org/10.1016/j.pbi.2014.05.001> PMID: 24879450
12. Judelson HS, Ah-Fong AM V. Update on Plant-Oomycete Interactions Exchanges at the Plant-Oomycete Interface That Influence Disease 1[OPEN]. *Plant Physiol.* 2019; 179: 1198–1211. <https://doi.org/10.1104/pp.18.00979> PMID: 30538168
13. Cook DE, Mesarich CH, Thomma BPHJ. Understanding Plant Immunity as a Surveillance System to Detect Invasion. *Annual Review of Phytopathology.* Annual Reviews; 2015. pp. 541–563. <https://doi.org/10.1146/annurev-phyto-080614-120114> PMID: 26047564
14. Sharpee WC, Dean RA. Form and function of fungal and oomycete effectors. *Fungal Biol Rev.* 2016; 30: 62–73. <https://doi.org/10.1016/j.fbr.2016.04.001>
15. Ye W, Wang Q, Tripathy S, Zhang M, Vetukuri RR. Editorial: Genomics and Effectomics of Filamentous Plant Pathogens. *Front Genet.* 2021; 12: 648690. <https://doi.org/10.3389/fgene.2021.648690> PMID: 33613652
16. Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, et al. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature.* 2009; 461: 393–398. <https://doi.org/10.1038/nature08358> PMID: 19741609
17. Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, et al. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science (1979).* 2010; 330: 1540–1543. <https://doi.org/10.1126/science.1193070> PMID: 21148391
18. Dong S, Raffaele S, Kamoun S. The two-speed genomes of filamentous pathogens: Waltz with plants. *Current Opinion in Genetics and Development.* Elsevier Current Trends; 2015. pp. 57–65. <https://doi.org/10.1016/j.gde.2015.09.001> PMID: 26451981
19. Thomma BPHJ, Seidl MF, Shi-Kunne X, Cook DE, Bolton MD, van Kan JAL, et al. Mind the gap; seven reasons to close fragmented genome assemblies. *Fungal Genetics and Biology.* 2016; 90: 24–30. <https://doi.org/10.1016/j.fgb.2015.08.010> PMID: 26342853
20. Ingram TW, Oh Y, Adhikari TB, Louws FJ, Dean RA. Comparative genome analyses of 18 *Verticillium dahliae* tomato isolates reveals phylogenetic and race specific signatures. *Front Microbiol.* 2020; 11: 3078. <https://doi.org/10.3389/fmicb.2020.573755> PMID: 33329432
21. Knaus BJ, Tabima JF, Shakya SK, Judelson HS, Grünwald NJ. Genome-wide increased copy number is associated with emergence of dominant clones of the Irish potato famine pathogen *Phytophthora infestans*. *mBio.* 2020; 11: 1–13. <https://doi.org/10.1128/mBio.00326-20> PMID: 32576669
22. Torres DE, Oggenfuss U, Croll D, Seidl MF. Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biol Rev.* 2020; 34: 136–143. <https://doi.org/10.1016/J.FBR.2020.07.001>
23. Frantzeskakis L, Kusch S, Panstruga R. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Mol Plant Pathol.* 2019; 20: 3–7. <https://doi.org/10.1111/mpp.12738> PMID: 30557450
24. Raffaele S, Kamoun S. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology* 2012 10:6. 2012; 10: 417–430. <https://doi.org/10.1038/nrmicro2790> PMID: 22565130
25. Everhart S, Gambhir N, Stam R. Population genomics of filamentous plant pathogens—A brief overview of research questions, approaches, and pitfalls. *Phytopathology.* 2021; 111: 12–22. <https://doi.org/10.1094/PHYTO-11-20-0527-FI> PMID: 33337245
26. Della Coletta R, Qiu Y, Ou S, Hufford MB, Hirsch CN. How the pan-genome is changing crop genomics and improvement. *Genome Biol.* 2021; 22: 1–19. <https://doi.org/10.1186/s13059-020-02224-8> PMID: 33397434
27. Garcia JF, Morales-Cruz A, Cochetel N, Minio A, Figueroa-Balderas R, Rolshausen PE, et al. Comparative Pangenomic Insights into the Distinct Evolution of Virulence Factors Among Grapevine Trunk Pathogens. *Molecular Plant-Microbe Interactions.* 2024; 37: 127–142. <https://doi.org/10.1094/MPMI-09-23-0129-R> PMID: 37934016
28. McCarthy CGP, Fitzpatrick DA. Pan-genome analyses of model fungal species. *Microb Genom.* 2019; 5. <https://doi.org/10.1099/mgen.0.000243> PMID: 30714895
29. Badet T, Croll D. The rise and fall of genes: origins and functions of plant pathogen pangenomes. *Curr Opin Plant Biol.* 2020; 56: 65–73. <https://doi.org/10.1016/j.pbi.2020.04.009> PMID: 32480355

30. Badet T, Oggenfuss U, Abraham L, McDonald BA, Croll D. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. *BMC Biology* 2020 18:1. 2020; 18: 1–18. <https://doi.org/10.1186/S12915-020-0744-3> PMID: 32046716
31. Hickey G, Monlong J, Ebler J, Novak AM, Eizenga JM, Gao Y, et al. Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nature Biotechnology* 2023. 2023; 1–11. <https://doi.org/10.1038/s41587-023-01793-w> PMID: 37165083
32. McInerney JO, McNally A, O'Connell MJ. Why prokaryotes have pangenomes. *Nat Microbiol.* 2017; 2. <https://doi.org/10.1038/nmicrobiol.2017.40> PMID: 28350002
33. Torres DE, Thomma BPHJ, Seidl MF. Transposable Elements Contribute to Genome Dynamics and Gene Expression Variation in the Fungal Plant Pathogen *Verticillium dahliae*. *Genome Biol Evol.* 2021; 13: 1–19. <https://doi.org/10.1093/gbe/evab135> PMID: 34100895
34. Fedoroff N V. Transposable elements, epigenetics, and genome evolution. *Science* (1979). 2012; 338: 758–767. https://doi.org/10.1126/SCIENCE.338.6108.758/ASSET/4D4639DF-CF55-418E-9680-EFB443D9A854/ASSETS/GRAPHIC/338_758_F9.JPEG
35. Faino L, Seidl MF, Shi-Kunne X, Pauper M, Van Den Berg GCM, Wittenberg AHJ, et al. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res.* 2016; 26: 1091–1100. <https://doi.org/10.1101/gr.204974.116> PMID: 27325116
36. Fletcher K, Michelmore R. Genome-Enabled Insights into Downy Mildew Biology and Evolution. 2023; 61. <https://doi.org/10.1146/annurev-phyto-021622-103440> PMID: 37268005
37. Lyon R, Correll J, Feng C, Bluhm B, Shrestha S, Shi A, et al. Population structure of *Peronospora effusa* in the southwestern United States. *PLoS One.* 2016; 11: e0148385. <https://doi.org/10.1371/journal.pone.0148385> PMID: 26828428
38. Ribera A, Bai Y, Wolters AMA, van Treuren R, Kik C. A review on the genetic resources, domestication and breeding history of spinach (*Spinacia oleracea* L.). *Euphytica.* Springer; 2020. pp. 1–21. <https://doi.org/10.1007/s10681-020-02585-y>
39. Kandel SL, Hulse-Kemp AM, Stoffel K, Koike ST, Shi A, Mou B, et al. Transcriptional analyses of differential cultivars during resistant and susceptible interactions with *Peronospora effusa*, the causal agent of spinach downy mildew. *Sci Rep.* 2020; 10: 1–13. <https://doi.org/10.1038/s41598-020-63668-3> PMID: 32317662
40. Koike S, Smith R, Schulbach K. Resistant cultivars, fungicides combat downy mildew of spinach. *Calif Agric* (Berkeley). 1992; 46: 29–30. <https://doi.org/10.3733/ca.v046n02p29>
41. Feng C, Saito K, Liu B, Manley A, Kammeijer K, Mauzey SJ, et al. New races and novel strains of the spinach downy mildew pathogen *Peronospora effusa*. *Plant Dis.* 2018; 102: 613–618. <https://doi.org/10.1094/PDIS-05-17-0781-RE> PMID: 30673485
42. Klein J, Neilen M, van Verk M, Dutilh BE, van den Ackerveken G. Genome reconstruction of the non-culturable spinach downy mildew *Peronospora effusa* by metagenome filtering. *PLoS ONE.* 2020. <https://doi.org/10.1371/journal.pone.0225808> PMID: 32396560
43. Feng C, Lamour K, Dhillon BDS, Villarreal-Zaballos MI, Castroagudin VL, Bluhm BH, et al. Genetic diversity of the spinach downy mildew pathogen based on hierarchical sampling. *bioRxiv.* 2020; 2020.02.18.953661. <https://doi.org/10.1101/2020.02.18.953661>
44. Skiadas P, Klein J, Quiroz-Monnens T, Elberse J, de Jonge R, Van den Ackerveken G, et al. Sexual reproduction contributes to the evolution of resistance-breaking isolates of the spinach pathogen *Peronospora effusa*. *Environ Microbiol.* 2022; 24: 1622–1637. <https://doi.org/10.1111/1462-2920.15944> PMID: 35191594
45. Fletcher K, Shin O-H, Clark KJ, Feng C, Putman AI, Correll JC, et al. Ancestral Chromosomes for Family Peronosporaceae Inferred from a Telomere-to-Telomere Genome Assembly of *Peronospora effusa*. 2022 [cited 24 May 2022]. <https://doi.org/10.1094/MPMI-09-21-0227-R> PMID: 35226812
46. Matson MEH, Liang Q, Lonardi S, Judelson HS. Karyotype variation, spontaneous genome rearrangements affecting chemical insensitivity, and expression level polymorphisms in the plant pathogen *Phytophthora infestans* revealed using its first chromosome-scale assembly. *PLoS Pathog.* 2022; 18: e1010869. <https://doi.org/10.1371/JOURNAL.PPAT.1010869> PMID: 36215336
47. Fletcher K, Martin F, Isakeit T, Cavanaugh K, Magill C, Michelmore R. The genome of the oomycete *Peronosclerospora sorghi*, a cosmopolitan pathogen of maize and sorghum, is inflated with dispersed pseudogenes. *G3 Genes|Genomes|Genetics.* 2023; 13: 340. <https://doi.org/10.1093/G3JOURNAL/JKAC340> PMID: 36592124
48. Saraiva M, Ściślak ME, Ascurra YT, Ferrando TM, Zic N, Henard C, et al. The molecular dialog between oomycete effectors and their plant and animal hosts. *Fungal Biol Rev.* 2022 [cited 1 Nov 2022]. <https://doi.org/10.1016/J.FBR.2022.10.002>

49. Whisson SC, Boevink PC, Moleleki L, Avrova AO, Morales JG, Gilroy EM, et al. A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* 2007 450:7166. 2007; 450: 115–118. <https://doi.org/10.1038/nature06203> PMID: 17914356
50. Wood KJ, Nur M, Gil J, Fletcher K, Lakeman K, Gann D, et al. Effector prediction and characterization in the oomycete pathogen *Bremia lactucae* reveal host-recognized WY domain proteins that lack the canonical RXLR motif. *PLoS Pathog.* 2020; 16: e1009012. <https://doi.org/10.1371/JOURNAL.PPAT.1009012> PMID: 33104763
51. Dou D, Kale SD, Wang X, Chen Y, Wang Q, Wang X, et al. Conserved C-Terminal Motifs Required for Avirulence and Suppression of Cell Death by *Phytophthora sojae* effector Avr1b. *Plant Cell.* 2008; 20: 1118–1133. <https://doi.org/10.1105/TPC.107.057067> PMID: 18390593
52. Jiang RHY, Tyler BM. Mechanisms and Evolution of Virulence in Oomycetes. 2011; 50: 295–318. <https://doi.org/10.1146/ANNUREV-PHYTO-081211-172912> PMID: 22920560
53. van Kogelenberg M, Clark AR, Jenkins Z, Morgan T, Anandan A, Sawyer GM, et al. Diverse phenotypic consequences of mutations affecting the C-terminus of FLNA. *J Mol Med.* 2015; 93: 773–782. <https://doi.org/10.1007/s00109-015-1261-7> PMID: 25686753
54. Hoencamp C, Dudchenko O, Elbatsh AMO, Brahmachari S, Raaijmakers JA, Schaik T van, et al. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science (1979).* 2021; 372: 28. <https://doi.org/10.1126/SCIENCE.ABE2218> PMID: 34045355
55. Torres DE, Reckard AT, Klocko AD, Seidl MF. Nuclear genome organization in fungi: From gene folding to Rab1 chromosomes. *FEMS Microbiology Reviews.* Oxford Academic; 2023. pp. 1–22. <https://doi.org/10.1093/femsre/fuad021> PMID: 37197899
56. Fang Y, Coelho MA, Shu H, Schotanus K, Thimmappa BC, Yadav V, et al. Long transposon-rich centromeres in an oomycete reveal divergence of centromere features in Stramenopila-Alveolata-Rhizaria lineages. *PLoS Genet.* 2020; 16: 1–30. <https://doi.org/10.1371/journal.pgen.1008646> PMID: 32150559
57. Fletcher K, Gil J, Bertier LD, Kenefick A, Wood KJ, Zhang L, et al. Genomic signatures of heterokaryosis in the oomycete pathogen *Bremia lactucae*. *Nat Commun.* 2019; 10: 1–13. <https://doi.org/10.1038/s41467-019-10550-0> PMID: 31201315
58. Judelson HS. Sexual Reproduction in Plant Pathogenic Oomycetes: Biology and Impact on Disease. *Sex in Fungi.* 2014; 445–458. <https://doi.org/10.1128/9781555815837.CH27>
59. Hoencamp C, Dudchenko O, Elbatsh AMO, Brahmachari S, Raaijmakers JA, Schaik T van, et al. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science (1979).* 2021; 372: 28. <https://doi.org/10.1126/SCIENCE.ABE2218> PMID: 34045355
60. Torres DE, Reckard AT, Klocko AD, Seidl MF. Nuclear genome organization in fungi: From gene folding to Rab1 chromosomes. *FEMS Microbiology Reviews.* Oxford Academic; 2023. pp. 1–22. <https://doi.org/10.1093/femsre/fuad021> PMID: 37197899
61. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol.* 1980; 16: 111–120. <https://doi.org/10.1007/BF01731581> PMID: 7463489
62. Feng C, Lamour KH, Bluhm BH, Sharma S, Shrestha S, Dhilon BDS, et al. Genome sequences of three races of *peronospora effusa*: A resource for studying the evolution of the spinach downy mildew pathogen. *Molecular Plant-Microbe Interactions.* 2018; 31: 1230–1231. <https://doi.org/10.1094/MPMI-04-18-0085-A> PMID: 29944056
63. Fletcher K, Klosterman SJ, Derevnina L, Martin F, Bertier LD, Koike S, et al. Comparative genomics of downy mildews reveals potential adaptations to biotrophy. *BMC Genomics.* 2018; 19: 8–10. <https://doi.org/10.1186/s12864-018-5214-8> PMID: 30486780
64. Whisson SC, Avrova AO, Lavrova O, Pritchard L. Families of short interspersed elements in the genome of the oomycete plant pathogen, *Phytophthora infestans*. *Fungal Genetics and Biology.* 2005; 42: 351–365. <https://doi.org/10.1016/j.fgb.2005.01.004> PMID: 15749054
65. Han G, Zhang N, Jiang H, Meng X, Qian K, Zheng Y, et al. Diversity of short interspersed nuclear elements (SINEs) in lepidopteran insects and evidence of horizontal SINE transfer between baculovirus and lepidopteran hosts. *BMC Genomics.* 2021; 22: 1–16. <https://doi.org/10.1186/S12864-021-07543-Z/FIGS/7>
66. Kanhayuwa L, Coutts RHA. Short Interspersed Nuclear Element (SINE) Sequences in the Genome of the Human Pathogenic Fungus *Aspergillus fumigatus* Af293. *PLoS One.* 2016; 11. <https://doi.org/10.1371/JOURNAL.PONE.0163215> PMID: 27736869
67. Feng BZ, Zhu XP, Fu L, Lv RF, Storey D, Tooley P, et al. Characterization of necrosis-inducing NLP proteins in *Phytophthora capsici*. *BMC Plant Biol.* 2014; 14: 126. <https://doi.org/10.1186/1471-2229-14-126> PMID: 24886309

68. Seidl MF, Van Den Ackerveken G. Activity and Phylogenetics of the Broadly Occurring Family of Microbial Nep1-Like Proteins. 2019; 57: 367–386. <https://doi.org/10.1146/annurev-phyto-082718-100054> PMID: 31283435
69. Cabral A, Oome S, Sander N, K fner I, N rnberger T, Van Den Ackerveken G. Nontoxic Nep1-like proteins of the downy mildew pathogen *Hyaloperonospora arabidopsidis*: repression of necrosis-inducing activity by a surface-exposed region. *Mol Plant Microbe Interact*. 2012; 25: 697–708. <https://doi.org/10.1094/MPMI-10-11-0269> PMID: 22235872
70. Zhu Z, Qu J, Yu L, Jiang X, Liu G, Wang L, et al. Three glycoside hydrolase family 12 enzymes display diversity in substrate specificities and synergistic action between each other. *Mol Biol Rep*. 2019; 46: 5443–5454. <https://doi.org/10.1007/s11033-019-04999-x> PMID: 31359382
71. Ozhelvacı F, Steczkiewicz K. Identification and classification of papain-like cysteine proteinases. *Journal of Biological Chemistry*. 2023; 299: 104801. <https://doi.org/10.1016/j.jbc.2023.104801> PMID: 37164157
72. Zheng X, McLellan H, Fraiture M, Liu X, Boevink PC, Gilroy EM, et al. Functionally Redundant RXLR Effectors from *Phytophthora infestans* Act at Different Steps to Suppress Early flg22-Triggered Immunity. *PLoS Pathog*. 2014; 10: e1004057. <https://doi.org/10.1371/JOURNAL.PPAT.1004057> PMID: 24763622
73. Yang Z, Guarracino A, Biggs PJ, Black MA, Ismail N, Wold JR, et al. Pangenome graphs in infectious disease: a comprehensive genetic variation analysis of *Neisseria meningitidis* leveraging Oxford Nanopore long reads. *Front Genet*. 2023; 14: 1225248. <https://doi.org/10.3389/FGENE.2023.1225248/BIBTEX>
74. Rice ES, Alberdi A, Alfieri J, Athrey G, Balacco JR, Bardou P, et al. A pangenome graph reference of 30 chicken genomes allows genotyping of large and complex structural variants. *BMC Biol*. 2023; 21: 1–17. <https://doi.org/10.1186/S12915-023-01758-0/FIGS/5>
75. Gao Y, Yang X, Chen H, Tan X, Yang Z, Deng L, et al. A pangenome reference of 36 Chinese populations. *Nature* 2023 619:7968. 2023; 619: 112–121. <https://doi.org/10.1038/s41586-023-06173-7> PMID: 37316654
76. Fiddes IT, Armstrong J, Diekhans M, Nachtweide S, Kronenberg ZN, Underwood JG, et al. Comparative Annotation Toolkit (CAT)—simultaneous clade and personal genome annotation. *Genome Res*. 2018; 28: 1029–1038. <https://doi.org/10.1101/GR.233460.117> PMID: 29884752
77. Van Dam P, Fokkens L, Ayukawa Y, Van Der Gragt M, Ter Horst A, Brankovics B, et al. A mobile pathogenicity chromosome in *Fusarium oxysporum* for infection of multiple cucurbit species. *Scientific Reports* 2017 7:1. 2017; 7: 1–15. <https://doi.org/10.1038/s41598-017-07995-y> PMID: 28831051
78. Hoogendoorn K, Barra L, Waalwijk C, Dickschat JS, van der Lee TAJ, Medema MH. Evolution and diversity of biosynthetic gene clusters in *Fusarium*. *Front Microbiol*. 2018; 9: 1–12. <https://doi.org/10.3389/fmicb.2018.01158> PMID: 29922257
79. Menardo F, Praz CR, Wicker T, Keller B. Rapid turnover of effectors in grass powdery mildew (*Blumeria graminis*). *BMC Evol Biol*. 2017; 17: 1–14. <https://doi.org/10.1186/s12862-017-1064-2> PMID: 29089018
80. Seidl MF, Thomma BPHJ. Sex or no sex: Evolutionary adaptation occurs regardless. *BioEssays*. 2014; 36: 335–345. <https://doi.org/10.1002/bies.201300155> PMID: 24531982
81. Langner T, Harant A, Gomez-Luciano LB, Shrestha RK, Malmgren A, Latorre SM, et al. Genomic rearrangements generate hypervariable mini-chromosomes in host-specific isolates of the blast fungus. *PLoS Genet*. 2021; 17: e1009386. <https://doi.org/10.1371/journal.pgen.1009386> PMID: 33591993
82. De Jonge R, Van Esse HP, Maruthachalam K, Bolton MD, Santhanam P, Saber MK, et al. Tomato immune receptor *Ve1* recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proc Natl Acad Sci U S A*. 2012; 109: 5110–5115. <https://doi.org/10.1073/pnas.1119623109> PMID: 22416119
83. Mathu Malar C, Yuzon JD, Das S, Das A, Panda A, Ghosh S, et al. Haplotype-phased genome assembly of virulent *Phytophthora ramorum* isolate ND886 facilitated by long-read sequencing reveals effector polymorphisms and copy number variation. *Molecular Plant-Microbe Interactions*. 2019; 32: 1047–1060. <https://doi.org/10.1094/MPMI-08-18-0222-R> PMID: 30794480
84. Qutob D, Tedman-Jones J, Dong S, Kuflu K, Pham H, Wang Y, et al. Copy Number Variation and Transcriptional Polymorphisms of *Phytophthora sojae* RXLR Effector Genes Avr1a and Avr3a. *PLoS One*. 2009; 4: e5066. <https://doi.org/10.1371/JOURNAL.PONE.0005066> PMID: 19343173
85. Henningsen EC, Lewis D, Nazareno E, Huang Y-F, Steffenson BJ, Boesen B, et al. A high-resolution haplotype pangenome uncovers somatic hybridization, recombination and intercontinental migration in oat crown rust. *bioRxiv*. 2024; 2024.03.27.583983. <https://doi.org/10.1101/2024.03.27.583983>

86. Engelbrecht J, Duong TA, Prabhu SA, Seedat M, van den Berg N. Genome of the destructive oomycete *Phytophthora cinnamomi* provides insights into its pathogenicity and adaptive potential. *BMC Genomics*. 2021; 22: 1–15. <https://doi.org/10.1186/S12864-021-07552-Y/FIGS/4>
87. Kasuga T, Bui M, Bernhardt E, Swiecki T, Aram K, Cano LM, et al. Host-induced aneuploidy and phenotypic diversification in the Sudden Oak Death pathogen *Phytophthora ramorum*. *BMC Genomics*. 2016; 17: 1–17. <https://doi.org/10.1186/S12864-016-2717-Z/FIGS/5>
88. Hu J, Shrestha S, Zhou Y, Mudge J, Liu X, Lamour K. Dynamic Extreme Aneuploidy (DEA) in the vegetable pathogen *Phytophthora capsici* and the potential for rapid asexual evolution. *PLoS One*. 2020; 15: e0227250. <https://doi.org/10.1371/JOURNAL.PONE.0227250> PMID: 31910244
89. Seidl MF, Van Den Ackerveken G, Govers F, Snel B. Reconstruction of Oomycete Genome Evolution Identifies Differences in Evolutionary Trajectories Leading to Present-Day Large Gene Families. *Genome Biol Evol*. 2012; 4: 199–211. <https://doi.org/10.1093/gbe/evs003> PMID: 22230142
90. Ayala-Usma DA, Cárdenas M, Guyot R, Mares MC De, Bernal A, Muñoz AR, et al. A whole genome duplication drives the genome evolution of *Phytophthora betacei*, a closely related species to *Phytophthora infestans*. *BMC Genomics*. 2021; 22: 1–21. <https://doi.org/10.1186/S12864-021-08079-Y/FIGS/8>
91. Sirén J, Monlong J, Chang X, Novak AM, Eizenga JM, Markello C, et al. Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. *Science (1979)*. 2021; 374. <https://doi.org/10.1126/science.abg8871> PMID: 34914532
92. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. 2019; 20: 1–13. <https://doi.org/10.1186/S13059-019-1891-0/FIGS/2>
93. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 2017; 27: 722–736. <https://doi.org/10.1101/gr.215087.116> PMID: 28298431
94. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*. 2018; 19: 1–10. <https://doi.org/10.1186/s12859-018-2485-7> PMID: 30497373
95. Von Meijenfeldt FAB, Arkhipova K, Cambuy DD, Coutinho FH, Dutilh BE. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biol*. 2019; 20: 1–14. <https://doi.org/10.1186/s13059-019-1817-x> PMID: 31640809
96. Mikheenko A, Pribelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUASt-LG. *Bioinformatics*. 2018; 34: i142–i150. <https://doi.org/10.1093/bioinformatics/bty266> PMID: 29949969
97. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst*. 2016; 3: 95–98. <https://doi.org/10.1016/j.cels.2016.07.002> PMID: 27467249
98. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science (1979)*. 2017; 356: 92–95. <https://doi.org/10.1126/science.aal3327> PMID: 28336562
99. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst*. 2016; 3: 99–101. <https://doi.org/10.1016/j.cels.2015.07.012> PMID: 27467250
100. Lam KK, Labutti K, Khalak A, Tse D. FinisherSC: a repeat-aware tool for upgrading de novo assembly using long reads. *Bioinformatics*. 2015; 31: 3207–3209. <https://doi.org/10.1093/bioinformatics/btv280> PMID: 26040454
101. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014; 9. <https://doi.org/10.1371/journal.pone.0112963> PMID: 25409509
102. Baril T, Galbraith J, Hayward A. Earl Grey: a fully automated user-friendly transposable element annotation and analysis pipeline. *bioRxiv*. 2024; 2022.06.30.498289. <https://doi.org/10.1101/2022.06.30.498289>
103. Smit A, Hubley R, Green P. RepeatMasker Open-4.0. 2013. Available from: <http://www.repeatmasker.org>
104. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, 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> PMID: 32300014
105. Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob DNA*. 2021; 12: 1–14. <https://doi.org/10.1186/S13100-020-00230-Y/FIGS/8>

106. Palmer J. funannotate. 2017. Available from: <https://github.com/nextgenusfs/funannotate>
107. Krzywinski M, Ave W. Genome Visualization with Circos and Hive Plots. 2015.
108. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26: 841–842. <https://doi.org/10.1093/bioinformatics/btq033> PMID: 20110278
109. Jones DAB, Rozano L, Debler JW, Mancera RL, Moolhuijzen PM, Hane JK. An automated and combinatorial method for the predictive ranking of candidate effector proteins of fungal plant pathogens. *Scientific Reports* 2021 11:1. 2021; 11: 1–13. <https://doi.org/10.1038/s41598-021-99363-0> PMID: 34611252
110. Bendtsen JD, Nielsen H, Von Heijne G, Brunak S. Improved Prediction of Signal Peptides: SignalP 3.0. *J Mol Biol*. 2004; 340: 783–795. <https://doi.org/10.1016/j.jmb.2004.05.028> PMID: 15223320
111. Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* 2011 8:10. 2011; 8: 785–786. <https://doi.org/10.1038/nmeth.1701> PMID: 21959131
112. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology* 2019 37:4. 2019; 37: 420–423. <https://doi.org/10.1038/s41587-019-0036-z> PMID: 30778233
113. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology* 2022 40:7. 2022; 40: 1023–1025. <https://doi.org/10.1038/s41587-021-01156-3> PMID: 34980915
114. Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res*. 2007; 35: W429–W432. <https://doi.org/10.1093/nar/gkm256> PMID: 17483518
115. Savojardo C, Martelli PL, Fariselli P, Casadio R. DeepSig: deep learning improves signal peptide detection in proteins. *Bioinformatics*. 2018; 34: 1690–1696. <https://doi.org/10.1093/bioinformatics/btx818> PMID: 29280997
116. 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> PMID: 11152613
117. Tabima JF, Grünwald NJ. EffectR: An expandable R package to predict candidate RXLR and CRN effectors in oomycetes using motif searches. *Molecular Plant-Microbe Interactions*. 2019; 32: 1067–1076. <https://doi.org/10.1094/MPMI-10-18-0279-TA> PMID: 30951442
118. Zhao S, Shang X, Bi W, Yu X, Liu D, Kang Z, et al. Genome-Wide Identification of Effector Candidates With Conserved Motifs From the Wheat Leaf Rust Fungus *Puccinia triticina*. *Front Microbiol*. 2020; 11: 534830. <https://doi.org/10.3389/FMICB.2020.01188/BIBTEX>
119. Eddy SR. HMMER: biosequence analysis using profile hidden Markov models. In: <http://hmmer.org/>. Aug 2023.
120. Boutemy LS, King SRF, Win J, Hughes RK, Clarke TA, Blumenschein TMA, et al. Structures of *Phytophthora* RXLR Effector Proteins: A CONSERVED BUT ADAPTABLE FOLD UNDERPINS FUNCTIONAL DIVERSITY *. *Journal of Biological Chemistry*. 2011; 286: 35834–35842. <https://doi.org/10.1074/JBC.M111.262303> PMID: 21813644
121. 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> PMID: 23329690
122. Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, et al. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021; 49: D480–D489. <https://doi.org/10.1093/nar/gkaa1100> PMID: 33237286
123. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*. 2015; 31: 3350. <https://doi.org/10.1093/bioinformatics/btv383> PMID: 26099265
124. Steenwyk JL, Buida TJ, Labella AL, Li Y, Shen XX, Rokas A. PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. *Bioinformatics*. 2021; 37: 2325–2331. <https://doi.org/10.1093/bioinformatics/btab096> PMID: 33560364
125. De La Torre AR, Li Z, Van De Peer Y, Ingvarsson PK. Contrasting Rates of Molecular Evolution and Patterns of Selection among Gymnosperms and Flowering Plants. *Mol Biol Evol*. 2017; 34: 1363–1377. <https://doi.org/10.1093/molbev/msx069> PMID: 28333233
126. Md V, Misra S, Li H, Aluru S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. *Proceedings—2019 IEEE 33rd International Parallel and Distributed Processing Symposium, IPDPS 2019*. 2019; 314–324. <https://doi.org/10.1109/IPDPS.2019.00041>

127. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Auwera GA Van der, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*. 2018; 201178. <https://doi.org/10.1101/201178>
128. Lischer HEL, Excoffier L. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*. 2012; 28: 298–299. <https://doi.org/10.1093/bioinformatics/btr642> PMID: 22110245
129. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*. Oxford Academic; 2006. pp. 254–267. <https://doi.org/10.1093/molbev/msj030> PMID: 16221896
130. Kamvar ZN, Tabima JF, unwald NJ. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*. 2014; 2014: 1–14. <https://doi.org/10.7717/peerj.281> PMID: 24688859
131. Sweeten AP, Schatz MC, Phillippy AM. ModDotPlot—Rapid and interactive visualization of complex repeats. *bioRxiv*. 2024; 2024.04.15.589623. <https://doi.org/10.1101/2024.04.15.589623> PMID: 38712106
132. Gilchrist CLM, Chooi Y-H. clinker & clustermap.js: automatic generation of gene cluster comparison Figs. *Bioinformatics*. 2021; 37: 2473–2475. <https://doi.org/10.1093/bioinformatics/btab007> PMID: 33459763
133. Waskom ML. seaborn: statistical data visualization. *J Open Source Softw*. 2021; 6: 3021. <https://doi.org/10.21105/JOSS.03021>
134. Correll J, Toit L du, Koike S, Ettekoven K van. Guidelines for Spinach Downy Mildew: *Peronospora farinosa* f. sp. *spinaciae* (Pfs). 2015 [cited 18 Jan 2022]. Available from: https://moam.info/peronospora-farinosa-f-sp-spinaciae-cppi_597ef8431723dd68e375d99d.html