PLOS GENETICS

RESEARCH ARTICLE

Population genomics of the grapevine pathogen *Eutypa lata* reveals evidence for population expansion and intraspecific differences in secondary metabolite gene clusters

Cristobal A. Onetto[1], **Mark R. Sosnowski**[2,3], **Steven Van Den Heuvel**[1], **Anthony R. Borneman**[1,3]*

1 The Australian Wine Research Institute, Adelaide, Australia, **2** South Australian Research and Development Institute, Adelaide, Australia, **3** School of Wine, Food and Agriculture, The University of Adelaide, Adelaide, Australia

* anthony.borneman@awri.com.au

Citation: Onetto CA, Sosnowski MR, Van Den Heuvel S, Borneman AR (2022) Population genomics of the grapevine pathogen *Eutypa lata* reveals evidence for population expansion and intraspecific differences in secondary metabolite gene clusters. PLoS Genet 18(4): e1010153. https://doi.org/10.1371/journal.pgen.1010153

Editor: Geraldine Butler, University College Dublin, IRELAND

Received: December 21, 2021

Accepted: March 17, 2022

Published: April 1, 2022

Data Availability Statement: The sequencing data and genome assemblies included in this study are publicly available at NCBI under BioProject PRJNA787025. The genome assembly and annotations of isolate TAS7 are available in NCBI under assembly accession JAKUDF000000000. Annotations for the remaining 39 isolates are available in the supplementary data.

Funding: This work was supported by Australian grapegrowers and winemakers through their

Abstract

Eutypa dieback of grapevine is an important disease caused by the generalist Ascomycete fungus *Eutypa lata*. Despite the relevance of this species to the global wine industry, its genomic diversity remains unknown, with only a single publicly available genome assembly. Whole-genome sequencing and comparative genomics was performed on forty Australian *E. lata* isolates to understand the genome evolution, adaptation, population size and structure of these isolates. Phylogenetic and linkage disequilibrium decay analyses provided evidence of extensive gene flow through sexual recombination between isolates obtained from different geographic locations and hosts. Investigation of the genetic diversity of these isolates suggested rapid population expansion, likely as a consequence of the recent growth of the Australian wine industry. Genomic regions affected by selective sweeps were shown to be enriched for genes associated with secondary metabolite clusters and included genes encoding proteins with a role in nutrient acquisition, degradation of host cell wall and metal and drug resistance, suggesting recent adaptation to both abiotic factors and potentially host genotypes. Genome synteny analysis using long-read genome assemblies showed significant intraspecific genomic plasticity with extensive chromosomal rearrangements impacting the secondary metabolite production potential of this species. Finally, k-mer based GWAS analysis identified a potential locus associated with mycelia recovery in canes of *Vitis vinifera* that will require further investigations.

Author summary

Eutypa dieback of grapevine, caused by the Ascomycete fungus *Eutypa lata*, is responsible for significant economic losses to the wine industry. Despite the worldwide prevalence of this pathogen, its genomic diversity remains unknown, with only a single publicly

investment body Wine Australia, with matching funds from the Australian Government. The funders 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.

available genome assembly. This knowledge gap was addressed by performing whole-genome sequencing of 40 *E. lata* isolates sourced from different hosts and geographical locations around Australia. Investigation of the genetic diversity of this population showed a high degree of gene-flow and sexual recombination as well as demographic expansion. Through the inspection of signatures of selective sweeps, repeat-mediated chromosomal rearrangements, and pan-genomic elements, it was shown that this species has a highly dynamic secondary metabolite production potential that could have important implications for its pathogenicity and lifestyle. In addition, application of a k-mer based GWAS methodology, identified a potential locus associated with the growth of this species within canes of *Vitis vinifera*.

1. Introduction

Eutypa dieback of grapevines is responsible for significant economic losses to the wine industry worldwide [1–3]. The fungal disease is caused by the Ascomycete *Eutypa lata* [1], which can affect a wide variety of woody plant species including grapevine, apricot, cherry, olive, peach and walnut [1–4]. Disease is generally spread by wind-dispersed ascospores infecting fresh pruning wounds. After landing on a suitable wound, *E. lata* ascospores rapidly germinate and begin colonizing the xylem vessels of the host [1]. The mycelium slowly spreads through the wood tissue, colonizing spurs, cordon and trunk, and eventually causing death of the grapevine [5].

Common symptoms of infection include wood necrosis observed as wedge-shaped cankers, leaf chlorosis and tattering of leaf margins, and stunting of shoots [1,6,7]. Foliar symptoms are thought to be caused by acetylenic phenol and chromene metabolites, which are produced by *E. lata* in the infected wood and then translocated to the foliage via the plant vascular system [8–10].

Variability in pathogenesis and disease susceptibility has been recorded between both strains of *E. lata* [6] and cultivars of *V. vinifera* [11], which are attributed to differences in the production of secondary metabolites [8] and xylem morphology and lignin composition of the wood, respectively [12,13]. Furthermore, analysis of the single draft genome available for this species revealed a large diversity of plant cell wall degrading enzymes and secondary metabolite clusters [14,15] that may explain the diversity of hosts infected by *E. lata*.

E. lata is considered a generalist fungus, regularly reported in continents where grapevines and specific *Prunus* species are cultivated, including Europe, North America, Australia, and South Africa [16]. Microsatellite based investigations of the genetic diversity of the *E. lata* population demonstrated high levels of gene flow between isolates and a lack of association between specific genotypic groups and either host or geographic location [17–19]. This lack of population structure has been hypothesised to arise from regular genetic reshuffling through sexual recombination and frequent immigration among hosts and geographic locations by spore dispersal and human-mediated transport of infected material [17,18].

In Australia, *E. lata* has been recognised as a major agricultural pathogen for over 60 years [20] and is suggested to have been introduced into the country from Europe through the transport of infected plant material [18]. Despite the relevance of this species to the Australian wine industry, the genomic diversity of the Australian *E. lata* population remains unknown. While microsatellite-based investigations have shed light on the global genetic diversity of this species [17,18], this technique has well documented limitations for the study of population genetics [21]. Additional genome sequencing efforts are necessary to understand the genome evolution

of this species, including adaptation, population size and structure as well as the genetic determinants of pathogenicity. In this study we address this knowledge gap by performing whole-genome sequencing and population genomic analyses of 40 *E. lata* isolates obtained from different hosts and geographical locations around Australia, including four isolates sequenced using long-read nanopore technology.

2. Material and methods

2.1. DNA extraction and genome sequencing of *E. lata* isolates

Forty *E. lata* isolates were obtained from the South Australian Research and Development Institute (SARDI) collection (S1 Table) from which 35 have been previously published and phenotyped [6,7,22,23]. Details on isolation source are available in S1 Table. For short-read shotgun DNA sequencing, isolates were grown on Potato Dextrose (PD) broth (Sigma, Australia) for 10 days at 22˚C after which samples were pelleted by centrifugation. DNA extraction from pellets was performed using a MagAttract Microbial DNA Kit (Qiagen, Australia) and a Precellys Evolution Homogenizer (Bertin, France) (30 s, 4500 rpm). Library preparation and sequencing was performed in the Ramaciotti Centre for Genomics (University of New South Wales, Sydney, Australia). Sequencing libraries were prepared using the Illumina DNA library kit and sequenced with an Illumina NovaSeq 6000 using 2 x 150 bp chemistry on a SP flowcell.

For long-read sequencing using nanopore technology, mycelium pellets were thoroughly squeeze-dried using a paper mesh, transferred into a mortar, and then frozen by directly pouring liquid nitrogen on top of the samples. Samples were then ground into a fine powder and transferred into 1.5 mL tubes already placed in dry ice. High molecular weight DNA was extracted directly from these tubes using a Gentra Puregene Yeast/Bact DNA extraction kit (Qiagen, Australia). Sequencing libraries were prepared using the SQK-LSK109 and EXP-NBD104 kits. Fast5 files were base called and demultiplexed using Guppy v. 4.5.3 (Oxford Nanopore Technologies, Oxford, UK) using the 'hac' model with a minimum quality score filtering of 7.

2.2. Genome assembly and annotation

Short-read assemblies of 36 *E. lata* isolates were performed with Spades v. 3.15.2 [24] and then filtered for contigs with a minimum size of 500 bp (S2 Table). Long-read assemblies of isolates TAS7, MA101, 511–17 and B003 (S2 Table) were assembled using Canu v. 2.1.1 [25] and Flye v. 2.8.3 [26]. Both assemblies were then combined using quickmerge v. 0.3 [27] to improve contiguity. Reads were then mapped back to assemblies and low coverage contigs were tagged and removed with Tapestry v. 1.0.0 [28]. Finally, assemblies were polished three times to correct for SNPs and indels using short-reads and Pilon v. 1.24 [29].

Gene prediction of genome assemblies was performed following the funannotate pipeline v. 1.8.7 [30], including Genemark-ES v. 4.68 [31], SNAP [32], Augustus v. 3.3.3 [33] and Glimmerhmm v. 3.0.4 [34] annotations trained using BUSCO v. 2 [35]. Repeats were identified using RepeatMasker v. 4.1.1 [36] and LTRharvest [37]. Functional annotations were performed using the UniProt database (2021_02), Interproscan 5 [38], Pfam v. 34.0 [39], antiSMASH v. 5.0 [40], SignalP v. 4.1 [41] and dbCAN v. 9.0 [42]. Assembly and annotation statistics for all isolates are available in S2 Table.

2.3. Genome synteny analysis

Prior to alignment, contigs from each assembly were reordered based on the reference assembly using Mauve v. 2.4.0 [43]. Alignments were performed between the long-read assemblies

of isolates TAS7, MA101, 511–17 and B003 using Nucmer v. 3.1 [44] and coordinates obtained with the show-coords function. Synteny breakpoints associated with chromosomal rearrangements were queried against the predicted repeats using BEDTools v. 2.30.0 [45] and only repeats with a distance of < 2 kb from the synteny breakpoint were retained and manually inspected. Synteny breakpoints located within 50 kb of contig ends were masked to avoid highly repetitive telomeric repeat regions. For visualisation of the alignments, closely adjacent synteny blocks were smoothed and plotted using Circos v. 0.69 [46].

2.4. Phylogenetic and population genetics analyses

Illumina reads from 40 isolates (S1 Table) were mapped to the long-read genome assembly of *E. lata* strain TAS7 using Bowtie2 v. 2.3.4 [47] keeping only reads with a minimum mapping quality of 30. Duplicate reads were removed with Picard v. 2.18.4 [48]. Pileups were generated using SAMtools v. 1.8 [49] and variants were called using VarScan v. 2.3.9 [50] with a minimum variant allele frequency threshold of 0.4 and minimum supporting reads at position cut-off of 10. Consensus calls were obtained for all positions where SNPs were found in any of the samples and merged using BcfTools v. 1.8 [51], keeping only homozygous SNP calls.

A maximum-likelihood phylogenetic tree was constructed using IQ-TREE v. 2.1.2 [52] with the GTR+ASC model using the consensus SNP matrix filtered by calls present in all samples and with a minor allele frequency ratio (MAF) of 0.05. Linkage disequilibrium decay was estimated by calculating the correlation coefficient (r^2) between pairs of loci present in the largest contig with a maximum distance of 100 kbp using vcftools v. 0.1.17 [53]. Values were averaged in 1 bp windows and plotted as a function of distance using R [54] and the ggplot2 package [55]. Tajima's *D* was calculated in 100 kbp windows across all the genome using vcftools v. 0.1.17 [53] and the MAF unfiltered SNP matrix. Plotting of the MAF spectrum was performed in R [54] using the ggplot2 package [55] and the MAF unfiltered SNP matrix.

Pan-genome analysis was performed with predicted coding DNA sequences (CDS) of the short-read assemblies using the GET-HOMOLOGUES-EST pipeline [56] and OrthoMCL [57].

Screen for selective sweeps across the genome was performed using RAiSD v. 2.9 [58] and the μ statistics. The μ statistic relies on multiple signatures of a selective sweep via the enumeration of SNP vectors, including expected reduction of variation in the region of a sweep, shifts in site frequency spectrum (SFS) and emergence of localized LD patterns on each side of the beneficial mutation [58]. Prior to running the pipeline, missing regions in all samples as well as telomeric and centromeric repeats were masked to avoid inflated μ scores. The top 0.01% scored windows were then selected and genes withing these regions were extracted using BEDTools v. 2.30.0 [45].

2.5. Genome wide association analysis (GWAS)

GWAS analyses was performed using the kmerGWAS pipeline v. 0.2 [59]. The k-mer database was built with 2 x 150 bp Illumina reads using KMC v. 3 [60] with a k-mer size of 31 bp. The kinship matrix was calculated using EMMA [61] with a minor allele frequency of 0.02 and a minor allele count of 5. GWAS was performed using the previously published phenotype data of mycelium recovery for 25 *E. lata* isolates [6]. In Sosnowski, Lardner [6], this phenotype was determined by drilling *E. lata* containing agar plugs into the apex of rootlings of *V. vinifera*. Mycelial spread was assessed 24 months after inoculation by isolation at 5 to 10 mm intervals and confirmed using specific DNA markers. Long-nanopore reads from isolate B003 were queried for the presence of the top significantly associated k-mers and then mapped back to the long-read genome assembly.

2.6. Meta-transcriptomic analysis

Previously published RNA-seq data [62] of grapevines showing symptoms of Eutypa dieback (samples ED1-8) were combined and mapped using the splice-aware mapping software STAR v. 2.7.9a [63] to a multi-species closed reference genome described by Morales-Cruz, Allenbeck [62] with the replacement of *E. lata* isolate UCREL1 with the long-read assembly of *E. lata* isolate TAS7. High quality mapping reads to *E. lata* were retained and gene counts were performed using featureCounts v.2.0.0 [64]. Counts were normalised as transcripts per million (TPM) as previously described by Wagner, Kin [65].

3. Results and discussion

The genetic diversity of an Australian *E. lata* population was investigated through whole genome sequencing and SNP-based analyses. Forty isolates, sourced from different hosts and grape-growing geographical locations, were subjected to short-read sequencing (S1 Table). In addition to short-read sequencing, four of these isolates (S2 Table) were also sequenced using nanopore long-read technology to obtain contiguous *de novo* genome assemblies for SNP-based analyses and investigation of the presence of genomic architectural rearrangements.

To better understand the phylogeny of the Australian *E. lata* population, SNPs were identified between the forty isolates. After filtering, 740,941 SNPs were retained and used for phylogenetic reconstruction (Fig 1A). The absence of heterozygous SNPs confirmed all isolates were either haploids or homozygous diploids. Phylogenetic analysis showed no evidence of clonality (Fig 1A). Two sets of isolates sourced from the regions of McLaren Vale (isolates MA089, SAPN01, MA101, MA176) and Clare Valley (C001, C002 and C003) grouped into clades based on geographic location (Fig 1A). From these, isolates labelled MA were obtained from different grapevines within a single vineyard and isolates C from three grapevine varieties (Cabernet Sauvignon, Chardonnay and Riesling) located in the Clare Valley. Isolates MA076 and MA002 did not form a clade with isolates MA089, MA101 and MA176 obtained from the same McLaren Vale vineyard (Fig 1A). The phylogenetic relationships between the McLaren Vale isolates MA089, SAPN01, MA101 and MA176, as well as the Clare Valley isolates C001, C002 and C003 suggested the presence of local recombination within specimens infecting the same vineyard. In agreement with previous population studies of *E. lata* [17–19], no clear grouping was observed based on either geographic location or host for the remaining isolates investigated (Fig 1A). These observations provide further evidence of gene flow between isolates infecting different hosts as previously reported by Travadon and Baumgartner [17].

Genetic reshuffling via sexual recombination and frequent gene flow between hosts has been suggested to prevent geographic and host differentiation in this species [17,18]. To estimate the degree of sexual recombination within the *E. lata* population, linkage disequilibrium (LD) was calculated for all biallelic SNPs present across the largest assembled chromosome (6.7 Mbp) of isolate TAS7 (Fig 1B). In recombining genomes, a decrease in linkage between two loci is expected as a function of the distance separating them. LD decay calculations using the squared correlation coefficient decreased to half of its maximum value between 100–150 bp (Fig 1B). Higher LD decays have been reported for well-studied sexually reproducing species such as *Neurospora crassa* (780 bp) [66]. Furthermore, LD decay distance as low as 110 bp has been observed in the obligately outcrossing mushroom *Schizophyllum commune* [67,68]. The short LD decay values observed for *E. lata* fall within the levels of sexually reproducing species and indicate that sexual reproduction and ascospore dispersal is the likely method of propagation for this population.

The lack of canonical mating-type regions has been reported for several members of the Xylariales [69] including *E. lata* strain UCREL1, where only a putative *MAT1-2-1* gene was

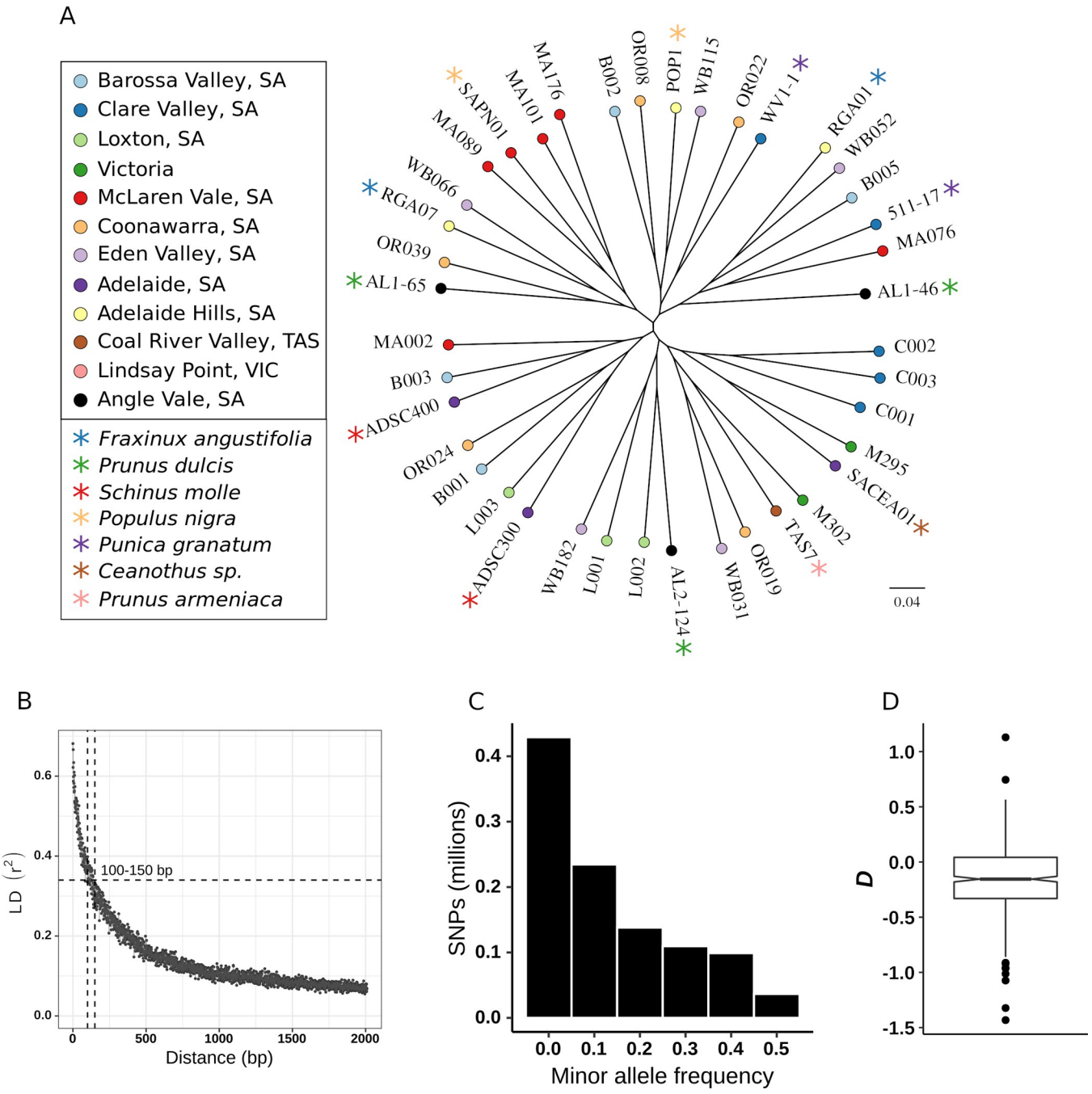

Fig 1. Genetic diversity within 40 Australian *Eutypa lata* isolates. A. Maximum-likelihood phylogenetic tree constructed from 740,941 single nucleotide polymorphisms (SNPs). Isolation location and host are indicated in the legend. Isolates without an asterisk (∗) were isolated from *Vitis vinifera*. B. Linkage disequilibrium (LD) decay on all bi-allelic SNPs present in the largest contig (6.7 Mbp). Data is presented as the squared correlation coefficient (r^2) between pairs of SNPs averaged in 1 bp windows. Dotted lines indicate half of the maximum observed value and the corresponding physical distance. C. Minor allele frequency spectrum of SNPs present in all samples. D. Boxplot of Tajima's *D* distribution calculated per 100 kbp windows.

https://doi.org/10.1371/journal.pgen.1010153.g001

annotated. Furthermore, this *MAT1-2-1* homolog was found to be unlinked to the orthologs of *sla2* and *apn2*, two genes that are normally linked to mating-type genes in ascomycetes species [70,71]. Examination of putative mating loci across the forty isolates in this study showed that all isolates contained a mating locus arrangement identical to UCREL1, with a putative *MAT1-*

2-1 gene unlinked to *sla2* and *apn2* (S3 Table). Mating type loci are known to be areas of recombination suppression in many fungal species [72]. LD was therefore investigated in all regions where a putative *MAT1-2-1* gene was found, as well as the locus of genes *sla2* and *apn2*. None of the regions showed signatures of high LD suggesting these regions do not determine the mating type of this species.

Inspection of the topology of the phylogenetic network showed the presence of a star-like tree characterized by long external branches (Fig 1A). Star-like phylogenies are consistent with a rapid expansion of a population following a bottleneck [73,74]. To further investigate the possibility of a recent bottleneck within the *E. lata* population, the minor allele frequency spectrum (MAF) (Fig 1C) and Tajima's *D* statistic (Fig 1D) were estimated across this set of isolates. The MAF spectrum displayed a distribution towards rare alleles (Fig 1C), without a spectrum distortion that would be indicative of a very recent population bottleneck [75], while Tajima's *D* statistic showed a tendency towards negative values (Fig 1D). Both measures are consistent with a rapid expansion of the Australian *E. lata* population. Establishment of *E. lata* in Australia has been previously proposed to have occurred through small founder population (s) [18]. The whole-genome population data is consistent with this theory, while further suggesting that the *E. lata* population is continuing to expand in Australia. It can be hypothesised that the recent significant growth experienced by the Australian wine industry has contributed to the rapid propagation of this species since its initial introduction into the country more than a century ago [1].

The highly contiguous genome assemblies produced using long-read data allowed for the investigation of signatures of selective sweeps across the *E. lata* population. To avoid interferences of background selection and potential false positives arising from demographic history, a stringent percentile threshold of 99.99% was applied for the μ statistic (see Materials and Methods). After filtering, 53 genomic regions were identified as being under selection, which ranged from 3.6 to 40.7 kb in length and contained a total of 240 ORFs (Fig 2, S4 Table). The complement of ORFs located within these selective sweep regions were enriched (p-value: 7.74e-10) for genes predicted to comprise members of several secondary metabolite clusters (S4 Table, Fig 2). The role of secondary metabolites in the lifecycle of *E. lata* is poorly understood, however these clusters likely have diverse roles including pathogenicity, the inhibition of competing microorganisms, host adaptation and dealing with environmental stressors, as previously reported in other fungal species [76].

Acetylenic phenols and chromene secondary metabolites have been proposed as the main phytotoxins produced by *E. lata* [9], however the specific biosynthetic clusters that are involved in the production of these compounds have not been experimentally confirmed. Recently, a putative cluster responsible for the synthesis of compounds, such as eutypine, that contain a 1,3-enyne moiety was identified and confirmed in an *Aspergillus* sp. [77]. The genome of *E. lata* strain UCREL1 contains a similar putative gene cluster, which has been previously suggested as being responsible for the production of eutypine [77]. It was possible to identify all nine genes that belong to this cluster in each of the forty isolates sequenced in this study (Fig 2). However, this cluster was not located within any region proposed to have undergone a selective sweep (Fig 2), indicating that the eutypinic cluster was not under recent positive selection.

Regions affected by selective sweeps also encoded proteins playing a role in transmembrane transport (Zn/Fe and amino acid permeases, drug and metal resistance and sugar and vitamin MFS transporters), degradation of host cell wall (CAZymes GH28, GH55, GH93, AA3 and AA7), protein degradation (MEROPS S08A, S10, S12 and M20D) and regulatory pathways, amongst others (S4 Table). The broad range of gene functions suggests that the *E. lata* population is adapting to both abiotic factors and potentially host genotypes. By using previously

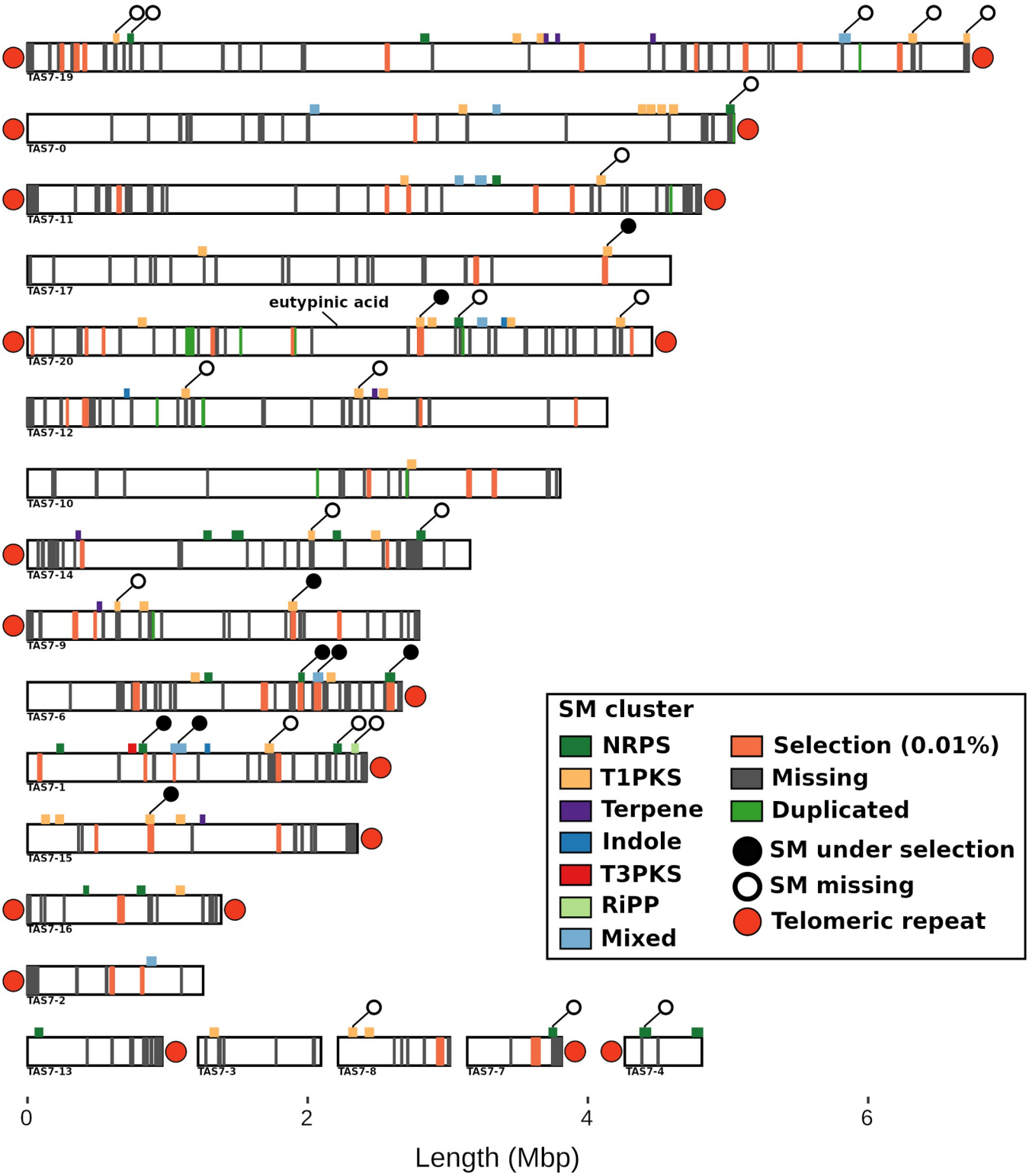

Fig 2. Karyotype representation of the genome assembly of *Eutypa lata* isolate TAS7. Predicted secondary metabolite (SM) clusters are shown above each contig with type of cluster indicated in the legend. Regions marked inside the contigs correspond to the following: (orange) top 0.01% μ scores representing signatures of selective sweep within the Australian population, (grey) missing genomic regions and (green) duplicated regions that overlap with CDSs present in at least one isolate. SM clusters under positive selection and missing in at least one sample are tagged with filled and empty circles, respectively. The genomic location of the previously reported eutypinic acid cluster [77] is marked.

https://doi.org/10.1371/journal.pgen.1010153.g002

published meta-transcriptomic data [62] and a multi-species reference genome (see Materials and Methods) including the assembly of isolate TAS7, 166 of these genes were shown to be expressed in-planta (TPM > 20) (S4 Table), suggesting they play a role during the symptomatic growth of this species in *V*. *vinifera*. Comparative analysis with established *E*. *lata* populations from other continental regions such as Europe and California will help to elucidate if there is divergence in selection due to environmental factors and diversity of hosts.

Comparison of the high-quality genome assemblies revealed significant intraspecific genomic plasticity. Synteny analyses showed the presence of extensive intra- and inter-chromosomal rearrangements, including large translocations, duplications and deletions of gene encoding regions (S1 Fig).

The genome architecture of fungal plant pathogens is known to be highly diverse [78] with intraspecific differences in genome size and architecture previously reported for several filamentous Ascomycetes [79,80]. Genomic loci that are rich in DNA repeats and/or transposable elements have been shown to frequently coincide with these breakpoints in genomic synteny [80–83], and are thought to evolve at faster rates, as well as harbour virulence-related genes [78].

To investigate the involvement of repetitive elements in the observed structural rearrangements between isolates of *E*. *lata*, the location of repetitive loci was assessed relative to breakpoints in synteny between the four isolates. Inspection of individual breakpoints in synteny between isolates 511-17/MA101 (S1A Fig) and TAS7/MA101 (S1B Fig) showed that 19 of 32 (511–17) and 27 of 41 (TAS7) (S1A Fig), as well as 21 of 27 (TAS7) and 18 of 25 (MA101) (S1B Fig) synteny breakpoints, respectively, were associated with repeat regions and LTR retrotransposons, suggesting that these elements are also providing the means for alterations in genome architecture in this species.

While most of the genetic diversity of *E*. *lata* is thought to arise from sexual recombination, these structural rearrangements may play a crucial role in the diverse pathogenicity reported [6] through gain and loss of virulence related genes [80]. Chromosomal rearrangements have also been linked with suppressed recombination [84], however, the short LD-decay values estimated in this population (Fig 1B) suggest these rearrangements are not largely affecting recombination.

To investigate how these genomic architectural differences impact the gene content of this population, the extent of the core and pan-genome was estimated using the predicted ORFs of all forty isolates. Pan-genome analysis indicated that 85% of predicted ORFs in TAS7 are shared across the population, with the *E*. *lata* core (ORFs present in all isolates) genome comprising 12533 ORFs. The soft-core genome (ORFs present in 95% of isolates), which allows for the presence of potentially missing or fragmented genes, was composed of 13184 ORFs (S2 Fig). From the pan-genome, the shell (3 to 37 isolates) and cloud ($\leq$ 2 isolates) genomes were composed of 4178 and 2331 ORFs, respectively (S2 Fig).

Genes within the shell and cloud represent a flexible genome that may reflect the individual lifestyle and adaptation of these isolates and captures sequences affected by structural variations. Enrichment analysis of Pfam domains associated with these isolate-specific genes showed significantly enriched domains in both the shell and cloud genomes that were associated with secondary metabolism, including polyketide synthases and cytochrome P450s (S5 Table), suggesting variability in the secondary metabolite production potential between these isolates.

While the abundance of specific classes of predicted secondary metabolite clusters across the four long-read assemblies were similar (S6 Table), reference-based mapping (against TAS7) of 39 isolates showed that only 62 out of the 82 predicted clusters observed in TAS7 were shared across these forty isolates (Fig 2).

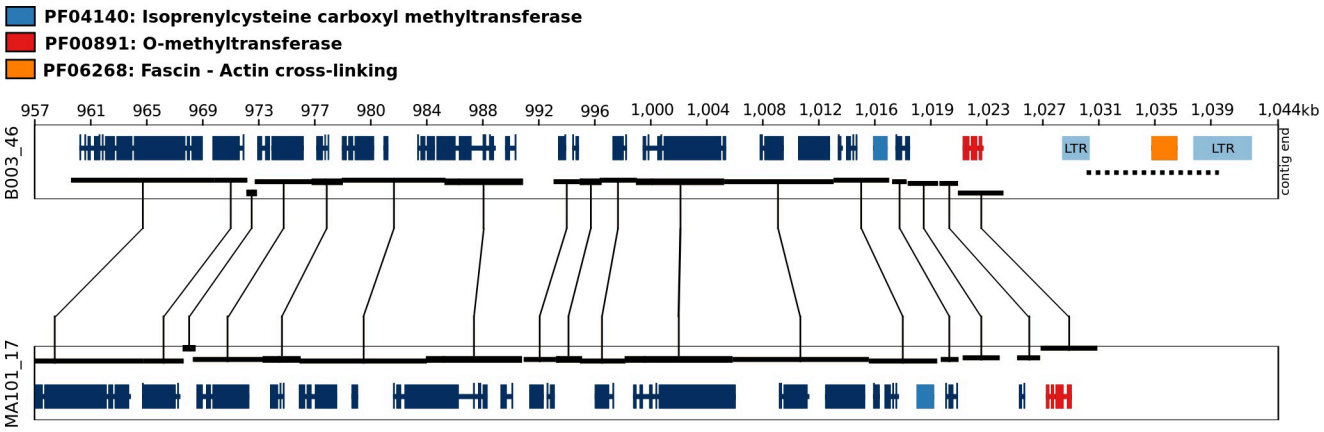

Fig 3. Genomic location of the 20 most significant k-mers associated with previously published phenotype data of 25 *Eutypa lata* isolates [6]. Synteny blocks between isolates B003 (k-mer containing isolate) and MA101 (absence of k-mers) are represented with black linked boxes. The mapping genomic location of long-reads containing the 20 most significant k-mers (p-value 6.65e-7) in the genome of isolate B003 is denoted with a dotted line. Pfam domains for neighbouring genes are indicated in the legend. The presence/absence of these kmers and the phenotype data are available in S7 Table.

https://doi.org/10.1371/journal.pgen.1010153.g003

Previous phenotypic screens of *E. lata* have demonstrated significant variability in the production of key metabolites between isolates [8,85,86], as well as significant variation in levels of pathogenicity [6,87–89]. When combined with the results of the pan-genome and reference-based analyses, it is evident that the metabolite production potential within *E. lata* is highly variable and could explain the broad differences in pathogenicity reported in several studies [6,87–89]. Future screens using untargeted metabolomics and the development of novel metabolomic methods [90] will be needed to confirm the diversity of metabolites produced, and how these compounds correlate with differences in pathogenicity and lifestyle of this species.

Variations in mycelium growth has been previously reported for 25 of the *E. lata* isolates included in this study [6]. Mycelial growth and colonization within the woody tissues of *V. vinifera* is considered a determinant of disease severity, with important implications for disease management, often requiring removal of tissue to avoid complete colonization and death of the infected grapevine. To investigate potential associations between genomic elements and the reported phenotypes, a k-mer-based GWAS methodology was applied [59]. Short-read sequences were used to assemble a k-mer database that was subsequently correlated against previously published phenotypic data concerning mycelial recovery of the sequenced isolates of *E. lata* within canes of *V. vinifera* [6]. The k-mer based methodology allowed for reference-free association of a broad range of genetic variants, including structural variations, that are not usually observed with conventional SNP-based GWAS methodologies.

Genomic locations that could be potentially linked to differences in mycelial recovery were defined by selecting the 20 most significantly associated k-mers, followed by mapping the k-mer associated long-reads back to the genome assembly of the isolate (Fig 3, S7 Table). Inspection of k-mer associated mapping locations showed that all reads were localised to a single genomic region that was predicted to encode a protein with an actin cross-linking domain (Fig 3), suggesting a possible role in actin-crosslinking and hyphal growth. In isolate B003, this gene is flanked by predicted LTR elements (Fig 3). Inspection of this region in isolate MA101 (non-k-mer containing isolate, S4 Table) showed a complete absence of the region, including the predicted ORF and flanking LTR elements (Fig 3). Due to the presence of repeat elements and the observed relationship between these elements and genomic synteny breakpoints (S1 Fig), it is likely that this region has been subjected to LTR-induced structural rearrangement in several of the isolates investigated. These preliminary results will need to be confirmed using a

larger phenotype sample size as commonly employed in standard GWAS analyses. Genetic transformation of *E. lata* using previously reported protoplast-based methods [91] will also be required to confirm the functionality of this genomic locus.

4. Conclusions

In this study we performed genome sequencing and comparative analyses of forty *E. lata* isolates sourced from diverse grape-growing regions across Australia, representing the first whole-genome population study of this important agricultural pathogen. The genetic diversity of this population showed a high degree of gene-flow and sexual recombination between isolates sourced from different hosts and geographic locations as well as signs of recent demographic expansion. Inspection of signatures of selective sweeps, repeat-mediated chromosomal rearrangements and pan-genomic elements revealed a highly dynamic secondary metabolite production potential that could have important implications for the pathogenicity and lifestyle of this species. K-mer based GWAS analysis identified a locus associated with mycelia recovery in canes of *V. vinifera* that will require further investigations. This study also provides a publicly available dataset of sequencing reads and genome assemblies including four isolates assembled using long-read technology. These genomic resources will be of high importance for future investigations aiming to understand the physiology, pathogenesis, and global genetic diversity of *E. lata*.

Supporting information

S1 Fig. Synteny between the genome assemblies of four *Eutypa lata* isolates represented as circos plots. Synteny blocks are coloured based on the reference sequence and query assembly is represented in grey. Query contigs have been ordered based on synteny length to reference. Contig IDs are labelled A) Genome synteny between *E. lata* isolate MA101 and isolate 511–17 and B) *E. lata* isolate MA101 and isolate TAS7. Repeat elements neighbouring (< 2 kb) synteny breakpoints are shown, excluding repeats located in contig ends. C) Genome synteny between isolates TAS7, B003, MA101 and 511–17.
(TIF)

S2 Fig. Gene-based pan-genome analysis of forty *Eutypa lata* isolates.
(TIFF)

S1 Table. Isolation details of the 40 *Eutypa lata* isolates included in this study.
(XLSX)

S2 Table. Genome assembly and annotation statistics for the 40 *Eutypa lata* isolates included in this study.
(XLSX)

S3 Table. Locus of putative *MAT1-2-1*, *sla2* and *apn2* genes in the 40 isolates investigated.
(XLSX)

S4 Table. Functional annotation and expression of genes located within regions under selective sweep.
(XLSX)

S5 Table. Enriched Pfam domains in shell and cloud genomes.
(XLSX)

S6 Table. Predicted secondary metabolite gene clusters in the long-read assembly of four *Eutypa lata* isolates.
(XLSX)

S7 Table. Presence/absence of the top significant k-mer and mycelium recovery phenotype for 25 *Eutypa lata* isolates.
(XLSX)

S1 Data. Annotation files for 39 *Eutypa lata* isolates in GFF3 format.
(ZIP)

S2 Data. Numerical data underlying graphs presented in Fig 1.
(XLSX)

Author Contributions

Conceptualization: Cristobal A. Onetto, Mark R. Sosnowski, Anthony R. Borneman.

Data curation: Cristobal A. Onetto.

Formal analysis: Cristobal A. Onetto.

Investigation: Cristobal A. Onetto, Steven Van Den Heuvel.

Methodology: Cristobal A. Onetto.

Project administration: Mark R. Sosnowski, Anthony R. Borneman.

Supervision: Anthony R. Borneman.

Writing – original draft: Cristobal A. Onetto.

Writing – review & editing: Mark R. Sosnowski, Steven Van Den Heuvel, Anthony R. Borneman.

References

1. Carter MV. The status of Eutypa lata as a pathogen: CAB International; 1991.
2. Siebert J. Eutypa: the economic toll on vineyards. Wines Vines. 2001; 82:50–6.
3. Wicks T, Davies K. The effect of Eutypa on grapevine yield. Australian Grapegrower and Winemaker. 1999:15–7.
4. Munkvold G, Marois J. Eutypa dieback of sweet cherry and occurrence of Eutypa lata perithecia in the central valley of California. Plant Dis. 1994; 78(2):200–7.
5. Pascoe I. Grapevine trunk diseases-black goo decline, esca, Eutypa dieback and others. Australian Grapegrower and Winemaker. 1999; 429:27–8.
6. Sosnowski MR, Lardner R, Wicks TJ, Scott ES. The influence of grapevine cultivar and isolate of *Eutypa lata* on wood and foliar symptoms. Plant Dis. 2007; 91(8):924–31. https://doi.org/10.1094/PDIS-91-8-0924 PMID: 30780424
7. Pitt WM, Trouillas FP, Gubler WD, Savocchia S, Sosnowski MR. Pathogenicity of Diatrypaceous fungi on grapevines in Australia. Plant Dis. 2013; 97(6):749–56. https://doi.org/10.1094/PDIS-10-12-0954-RE PMID: 30722588
8. Mahoney N, Lardner R, Molyneux RJ, Scott ES, Smith LR, Schoch TK. Phenolic and heterocyclic metabolite profiles of the grapevine pathogen *Eutypa lata*. Phytochemistry. 2003; 64(2):475–84. https://doi.org/10.1016/s0031-9422(03)00337-6 PMID: 12943765
9. Smith LR, Mahoney N, Molyneux RJ. Synthesis and structure−phytotoxicity relationships of acetylenic phenols and chromene metabolites, and their analogues, from the grapevine pathogen *Eutypa lata*. J Nat Prod. 2003; 66(2):169–76. https://doi.org/10.1021/np020415t PMID: 12608846

10. Tey-Rulh P, Philippe I, Renaud J-M, Tsoupras G, de Angelis P, Fallot J, et al. Eutypine, a phytotoxin produced by *Eutypa lata* the causal agent of dying-arm disease of grapevine. Phytochemistry. 1991; 30 (2):471–3. https://doi.org/10.1016/0031-9422(91)83707-R.

11. Sosnowski MR, Ayres MR, McCarthy MG, Scott ES. Winegrape cultivars (*Vitis vinifera*) vary in susceptibility to the grapevine trunk pathogens *Eutypa lata* and *Diplodia seriata*. Aust J Grape Wine Res. 2021; 28:166–74. https://doi.org/10.1111/ajgw.12531.c

12. Rolshausen PE, Greve LC, Labavitch JM, Mahoney NE, Molyneux RJ, Gubler WD. Pathogenesis of *Eutypa lata* in grapevine: identification of virulence factors and biochemical characterization of cordon dieback. Phytopathology. 2008; 98(2):222–9. https://doi.org/10.1094/PHYTO-98-2-0222 PMID: 18943199

13. Hamblin J. Factors affecting grapevine susceptibility to Eutypa dieback. University of Adelaide, Adelaide. 2015.

14. Blanco-Ulate B, Rolshausen PE, Cantu D. Draft genome sequence of the grapevine dieback fungus *Eutypa lata* UCR-EL1. Genome Announcements. 2013; 1(3):e00228–13. https://doi.org/10.1128/genomeA.00228-13 PMID: 23723393

15. Morales-Cruz A, Amrine KCH, Blanco-Ulate B, Lawrence DP, Travadon R, Rolshausen PE, et al. Distinctive expansion of gene families associated with plant cell wall degradation, secondary metabolism, and nutrient uptake in the genomes of grapevine trunk pathogens. BMC Genomics. 2015; 16(1):469. https://doi.org/10.1186/s12864-015-1624-z PMID: 26084502

16. Carter M, Bolay A, Rappaz F. An annotated host list and bibliography of Eutypa armeniacae: Station fédérale de recherches agronomiques de Changins; 1983.

17. Travadon R, Baumgartner K. Molecular polymorphism and phenotypic diversity in the Eutypa Dieback pathogen *Eutypa lata*. Phytopathology. 2015; 105(2):255–64. https://doi.org/10.1094/PHYTO-04-14-0117-R PMID: 25084304.

18. Travadon R, Baumgartner K, Rolshausen PE, Gubler WD, Sosnowski MR, Lecomte P, et al. Genetic structure of the fungal grapevine pathogen *Eutypa lata* from four continents. Plant Pathol. 2012; 61 (1):85–95. https://doi.org/10.1111/j.1365-3059.2011.02496.x

19. Lardner R, Scott ES, Stummer BE. Genetic variation in Australian isolates of the grapevine pathogen *Eutypa lata*. Australas Plant Pathol. 2007; 36(2):149–56. https://doi.org/10.1071/AP06092

20. Wicks T, Hall B. Eutypa dieback, a serious disease. The Australian Grapegrower and Winemaker. 1997; 405:61–2.

21. Putman AI, Carbone I. Challenges in analysis and interpretation of microsatellite data for population genetic studies. Ecol Evol. 2014; 4(22):4399–428. https://doi.org/10.1002/ece3.1305 PMID: 25540699

22. Sosnowski MR, Dinh SQ, Sandow MJ. First Report of *Eutypa lata* Causing Dieback and Wood Canker of Pomegranate (*Punica granatum*) in Australia. Plant Dis. 2020; 104(2):568–. https://doi.org/10.1094/pdis-09-19-1912-pdn

23. Trouillas FP, Pitt WM, Sosnowski MR, Huang R, Peduto F, Loschiavo A, et al. Taxonomy and DNA phylogeny of Diatrypaceae associated with *Vitis vinifera* and other woody plants in Australia. Fungal Diversity. 2011; 49(1):203–23. https://doi.org/10.1007/s13225-011-0094-0

24. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012; 19(5):455–77. https://doi.org/10.1089/cmb.2012.0021 PMID: 22506599

25. 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. https://doi.org/10.1101/gr.215087.116 PMID: 28298431

26. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019; 37(5):540–6. https://doi.org/10.1038/s41587-019-0072-8 PMID: 30936562

27. Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res. 2016; 44(19):e147–e. https://doi.org/10.1093/nar/gkw654 PMID: 27458204

28. Davey JW, Davis SJ, Mottram JC, Ashton PD. Tapestry: validate and edit small eukaryotic genome assemblies with long reads. bioRxiv. 2020. https://doi.org/10.1101/2020.04.24.059402

29. 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 (11):e112963. https://doi.org/10.1371/journal.pone.0112963 PMID: 25409509

30. Palmer J, Stajich J. Funannotate: eukaryotic genome annotation pipeline. Zenodo 2017; 10.

31. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 2008; 18(12):1979–90. https://doi.org/10.1101/gr.081612.108 PMID: 18757608

32. Korf IJBB. Gene finding in novel genomes. 2004; 5(1):59. https://doi.org/10.1186/1471-2105-5-59 PMID: 15144565

33. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics. 2003. https://doi.org/10.1093/bioinformatics/btg1080 PMID: 14534192

34. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 1999; 27(23):4636–41. https://doi.org/10.1093/nar/27.23.4636 PMID: 10556321

35. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of Eukaryotic, Prokaryotic, and viral Genomes. Mol Biol Evol. 2021. https://doi.org/10.1093/molbev/msab199 PMID: 34320186

36. Smit A, Hubley R, Green P. RepeatMasker Open-4.0. 2013–2015. 2015.

37. Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinf. 2008; 9(1):18. https://doi.org/10.1186/1471-2105-9-18 PMID: 18194517

38. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014; 30(9):1236–40. https://doi.org/10.1093/bioinformatics/btu031 PMID: 24451626

39. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar Gustavo A, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2020. https://doi.org/10.1093/nar/gkaa913 PMID: 33125078

40. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 2019. https://doi.org/10.1093/nar/gkz310 PMID: 31032519

41. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011; 8:785. https://doi.org/10.1038/nmeth.1701 PMID: 21959131

42. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2012. https://doi.org/10.1093/nar/gks479 PMID: 22645317

43. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004; 14(7):1394–403. https://doi.org/10.1101/gr.2289704 PMID: 15231754

44. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004. https://doi.org/10.1186/gb-2004-5-2-r12 PMID: 14759262

45. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010. https://doi.org/10.1093/bioinformatics/btq033 PMID: 20110278

46. Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009. https://doi.org/10.1101/gr.092759.109 PMID: 19541911

47. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012. https://doi.org/10.1038/nmeth.1923 PMID: 22388286

48. Toolkit P. Picard toolkit. Broad Institute, Github Repository. 2019.

49. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011. https://doi.org/10.1093/bioinformatics/btr509 PMID: 21903627

50. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012. https://doi.org/10.1101/gr.129684.111 PMID: 22300766

51. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021; 10(2). https://doi.org/10.1093/gigascience/giab008 PMID: 33590861

52. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2014; 32(1):268–74. https://doi.org/10.1093/molbev/msu300 PMID: 25371430

53. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011; 27(15):2156–8. https://doi.org/10.1093/bioinformatics/btr330 PMID: 21653522

54. Team RC. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2013. Document freely available on the internet at: http://wwwr-projectorg. 2015.

55. Wickham H. ggplot2: elegant graphics for data analysis: Springer; 2016.

56. Contreras-Moreira B, Cantalapiedra CP, García-Pereira MJ, Gordon SP, Vogel JP, Igartua E, et al. Analysis of plant pan-genomes and transcriptomes with GET_HOMOLOGUES-EST, a clustering solution for sequences of the same species. Frontiers in Plant Science. 2017. https://doi.org/10.3389/fpls.2017.00184 PMID: 28261241

57. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003. https://doi.org/10.1101/gr.1224503 PMID: 12952885

58. Alachiotis N, Pavlidis P. RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. Comm Biol. 2018. https://doi.org/10.1038/s42003-018-0085-8 PMID: 30271960

59. Voichek Y, Weigel D. Identifying genetic variants underlying phenotypic variation in plants without complete genomes. Nat Genet. 2020; 52(5):534–40. https://doi.org/10.1038/s41588-020-0612-7 PMID: 32284578

60. Kokot M, Długosz M, Deorowicz S. KMC 3: counting and manipulating k-mer statistics. Bioinformatics. 2017. https://doi.org/10.1093/bioinformatics/btx304 PMID: 28472236

61. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, et al. Efficient control of population structure in model organism association mapping. Genetics. 2008; 178(3):1709–23. https://doi.org/10.1534/genetics.107.080101 PMID: 18385116

62. Morales-Cruz A, Allenbeck G, Figueroa-Balderas R, Ashworth VE, Lawrence DP, Travadon R, et al. Closed-reference metatranscriptomics enables in planta profiling of putative virulence activities in the grapevine trunk disease complex. Mol Plant Pathol. 2018; 19(2):490–503. https://doi.org/10.1111/mpp.12544 PMID: 28218463

63. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2012; 29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635 PMID: 23104886

64. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2013; 30(7):923–30. https://doi.org/10.1093/bioinformatics/btt656 PMID: 24227677

65. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. 2012; 131(4):281–5. https://doi.org/10.1007/s12064-012-0162-3 PMID: 22872506

66. Ellison CE, Hall C, Kowbel D, Welch J, Brem RB, Glass NL, et al. Population genomics and local adaptation in wild isolates of a model microbial eukaryote. Pro Natl Acad Sci. 2011. https://doi.org/10.1073/pnas.1014971108 PMID: 21282627

67. Baranova MA, Logacheva MD, Penin AA, Seplyarskiy VB, Safonova YY, Naumenko SA, et al. Extraordinary genetic diversity in a wood decay mushroom. Mol Biol Evol. 2015. https://doi.org/10.1093/molbev/msv153 PMID: 26163667

68. Nieuwenhuis BPS, James TY. The frequency of sex in fungi. Philos Trans R Soc, B. 2016. https://doi.org/10.1098/rstb.2015.0540 PMID: 27619703

69. Robinson AJ, Natvig DO. Diverse members of the Xylariales lack canonical mating-type regions. Fungal Genet Biol. 2019; 122:47–52. https://doi.org/10.1016/j.fgb.2018.12.004 PMID: 30557613

70. Butler G, Kenny C, Fagan A, Kurischko C, Gaillardin C, Wolfe KH. Evolution of the MAT locus and its Ho endonuclease in yeast species. Proc Natl Acad Sci U S A. 2004; 101(6):1632–7. https://doi.org/10.1073/pnas.0304170101 PMID: 14745027

71. Gioti A, Mushegian AA, Strandberg R, Stajich JE, Johannesson H. Unidirectional evolutionary transitions in fungal mating systems and the role of transposable elements. Mol Biol Evol. 2012; 29(10):3215–26. https://doi.org/10.1093/molbev/mss132 PMID: 22593224

72. Idnurm A, Hood ME, Johannesson H, Giraud T. Contrasted patterns in mating-type chromosomes in fungi: hotspots versus coldspots of recombination. Fungal Biol Rev. 2015; 29(3):220–9. https://doi.org/10.1016/j.fbr.2015.06.001.

73. Exposito-Alonso M, Becker C, Schuenemann VJ, Reiter E, Setzer C, Slovak R, et al. The rate and potential relevance of new mutations in a colonizing plant lineage. PLoS Genet. 2018. https://doi.org/10.1371/journal.pgen.1007155 PMID: 29432421

74. Latorre SM, Reyes-Avila CS, Malmgren A, Win J, Kamoun S, Burbano HA. Differential loss of effector genes in three recently expanded pandemic clonal lineages of the rice blast fungus. BMC Biol. 2020; 18(1):88. https://doi.org/10.1186/s12915-020-00818-z PMID: 32677941

75. Luikart G, Allendorf F, Cornuet J-M, Sherwin W. Distortion of allele frequency distributions provides a test for recent population bottlenecks. J Hered. 1998; 89(3):238–47. https://doi.org/10.1093/jhered/89.3.238 PMID: 9656466

76. Howlett BJ. Secondary metabolite toxins and nutrition of plant pathogenic fungi. Curr Opin Plant Biol. 2006. https://doi.org/10.1016/j.pbi.2006.05.004 PMID: 16713733

77. Chen Y-R, Naresh A, Liang S-Y, Lin C-H, Chein R-J, Lin H-C. Discovery of a dual function Cytochrome P450 that catalyzes enyne formation in cyclohexanoid terpenoid biosynthesis. Angew Chem, Int Ed. 2020; 59(32):13537–41. https://doi.org/10.1002/anie.202004435 PMID: 32343875

78. Möller M, Stukenbrock EH. Evolution and genome architecture in fungal plant pathogens. Nat Rev Microbiol. 2017; 15(12):756–71. https://doi.org/10.1038/nrmicro.2017.76 PMID: 28781365

79. Moolhuijzen P, See PT, Hane JK, Shi G, Liu Z, Oliver RP, et al. Comparative genomics of the wheat fungal pathogen *Pyrenophora tritici-repentis* reveals chromosomal variations and genome plasticity. BMC Genomics. 2018; 19(1):279. https://doi.org/10.1186/s12864-018-4680-3 PMID: 29685100

80. de Jonge R, Bolton MD, Kombrink A, van den Berg GCM, Yadeta KA, Thomma BPHJ. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res. 2013; 23(8):1271–82. https://doi.org/10.1101/gr.152660.112 PMID: 23685541

81. 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(7262):393–8. https://doi.org/10.1038/nature08358 PMID: 19741609

82. Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, et al. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science. 2010; 330(6010):1543–6. https://doi.org/10.1126/science.1194573 PMID: 21148392

83. Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, et al. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nat Comm. 2011; 2(1):202. https://doi.org/10.1038/ncomms1189 PMID: 21326234

84. Sun Y, Svedberg J, Hiltunen M, Corcoran P, Johannesson H. Large-scale suppression of recombination predates genomic rearrangements in *Neurospora tetrasperma*. Nat Comm. 2017; 8(1):1140. https://doi.org/10.1038/s41467-017-01317-6 PMID: 29074958

85. Lardner R, MAhoney N, Zanker TP, Molyneux RJ, Scott ES. Secondary metabolite production by the fungal pathogen *Eutypa lata*: Analysis of extracts from grapevine cultures and detection of those metabolites in planta. Aust J Grape Wine Res. 2006; 12(2):107–14. https://doi.org/10.1111/j.1755-0238.2006.tb00049.x.

86. Rolshausen PE, Mahoney NE, Molyneux RJ, Gubler WD. A Reassessment of the species concept in *Eutypa lata*, the causal agent of Eutypa dieback of grapevine. Phytopathology. 2006; 96(4):369–77. https://doi.org/10.1094/PHYTO-96-0369 PMID: 18943418

87. Péros JP, Jamaux-DesprÉAux I, Berger G, Gerba D. The potential importance of diversity in *Eutypa lata* and co-colonising fungi in explaining variation in development of grapevine dieback. Mycol Res. 1999; 103(11):1385–90. https://doi.org/10.1017/S0953756299008291.

88. Péros J-P, Berger G, Lahogue F. Variation in pathogenicity and genetic structure in the *Eutypa lata* population of a single vineyard. Phytopathology. 1997; 87(8):799–806. https://doi.org/10.1094/PHYTO.1997.87.8.799 PMID: 18945047

89. Péros J-P, Berger G. Genetic structure and variation in aggressiveness in european and Australian populations of the grapevine dieback fungus, *Eutypa lata*. Eur J Plant Pathol. 2003; 109(9):909–19. https://doi.org/10.1023/B:EJPP.0000003648.10264.62

90. Aron AT, Petras D, Schmid R, Gauglitz JM, Büttel I, Antelo L, et al. Native mass spectrometry-based metabolomics identifies metal-binding compounds. Nat Chem. 2021. https://doi.org/10.1038/s41557-021-00803-1 PMID: 34795435

91. Bradshaw R, Duan G, Long P. Transformation of fungal grapevine trunk disease pathogens with the green fluorescent protein gene [*Vitis vinifera* L.]. Phytopathol Mediterranea. 2005.