Title

Closed-reference metatranscriptomics enables in planta profiling of putative virulence activities in the grapevine trunk disease complex

Journal

Molecular Plant Pathology, 19(2)

ISSN

1464-6722

Authors

Morales-Cruz, Abraham
Allenbeck, Gabrielle
Figueroa-Balderas, Rosa
et al.

Publication Date

2018-02-01

DOI

10.1111/mpp.12544

Molecular Plant Pathology

MOLECULAR PLANT PATHOLOGY DOI: 10.1111/mpp.12544

Technical advance

Closed-reference metatranscriptomics enables *in planta* profiling of putative virulence activities in the grapevine trunk disease complex

ABRAHAM MORALES-CRUZ[1], GABRIELLE ALLENBECK[1], ROSA FIGUEROA-BALDERAS[1], VANESSA E. ASHWORTH[2], DANIEL P. LAWRENCE[3], RENAUD TRAVADON[3], RHONDA J. SMITH[4], KENDRA BAUMGARTNER[5], PHILIPPE E. ROLSHAUSEN[2] AND DARIO CANTU[1,*]

[1]*Department of Viticulture and Enology, University of California Davis, Davis, CA 95616, USA*
[2]*Department of Botany and Plant Sciences, University of California Riverside, Riverside, CA 92521, USA*
[3]*Department of Plant Pathology, University of California Davis, Davis, CA 95616, USA*
[4]*University of California Cooperative Extension, Sonoma County, Santa Rosa, CA 95403, USA*
[5]*United States Department of Agriculture - Agricultural Research Service, Crops Pathology and Genetics Research Unit, Davis, CA 95616, USA*

SUMMARY

Grapevines, like other perennial crops, are affected by so-called 'trunk diseases', which damage the trunk and other woody tissues. Mature grapevines typically contract more than one trunk disease and often multiple grapevine trunk pathogens (GTPs) are recovered from infected tissues. The co-existence of different GTP species in complex and dynamic microbial communities complicates the study of the molecular mechanisms underlying disease development, especially under vineyard conditions. The objective of this study was to develop and optimize a community-level transcriptomics (i.e. metatranscriptomics) approach that could monitor simultaneously the virulence activities of multiple GTPs *in planta*. The availability of annotated genomes for the most relevant co-infecting GTPs in diseased grapevine wood provided the unprecedented opportunity to generate a multi-species reference for the mapping and quantification of DNA and RNA sequencing reads. We first evaluated popular sequence read mappers using permutations of multiple simulated datasets. Alignment parameters of the selected mapper were optimized to increase the specificity and sensitivity for its application to metagenomics and metatranscriptomics analyses. Initial testing on grapevine wood experimentally inoculated with individual GTPs confirmed the validity of the method. Using naturally infected field samples expressing a variety of trunk disease symptoms, we show that our approach provides quantitative assessments of species composition, as well as genome-wide transcriptional profiling of potential virulence factors, namely cell wall degradation, secondary metabolism and nutrient uptake for all co-infecting GTPs.

* *Correspondence*: Email: dacantu@ucdavis.edu

Keywords: Botryosphaeria dieback, Esca, Eutypa dieback, metagenomics, next-generation sequencing, Phomopsis dieback.

INTRODUCTION

When interacting with their host, plant pathogens do not exist in isolation, but are part of complex and dynamic microbial communities (Allen and Banfield, 2005). Such communities may comprise multiple pathogenic species and other microorganisms with negative, neutral or beneficial interactions whilst colonizing the same plant organ (Fitt *et al.*, 2006; Turner *et al.*, 2013). Under field conditions, a combination of microbe–microbe and microbe–plant interactions contributes to the development and severity of a disease (Lamichhane and Venturi, 2015). Direct and potentially synergistic relations may develop between the co-infecting microbes, including inter-species signalling (Hogan, 2006; Hosni *et al.*, 2011) and metabolic exchange and complementarity (Ponomarova and Patil, 2015; Zelezniak *et al.*, 2015). Indirectly, by interfering with host immune responses (Nomura *et al.*, 2005), by modifying the physicochemical characteristics of the host environment or by killing host cells, pathogens can facilitate host colonization by other microbes, which collectively may exacerbate disease symptoms (Rowe *et al.*, 1985).

Complex diseases that affect woody structures of perennial plants, such as grapevine trunk diseases, are often the result of infection by multiple pathogens, which simultaneously or sequentially colonize the host tissue (Hiscox *et al.*, 2015). All permanent structures of a grapevine can be infected by different fungi that cause distinct trunk diseases (Bertsch *et al.*, 2013). Among these, Botryosphaeria dieback, Esca, Eutypa dieback and Phomopsis dieback are the most common (Bertsch *et al.*, 2013). Trunk

pathogens colonize the permanent woody structures of the vine mainly through wounds (Rolshausen *et al.*, 2010). As they spread through the host via a combination of cell wall degradation, toxin secretion and necrosis, damage to the vascular tissues causes the progressive reduction in water and nutrient translocation, and eventually leads to the death of the infected organs and, possibly, of the entire vine (Pouzoulet *et al.*, 2014). Disease symptoms in organs distal to the infected tissue may also develop as a result of secondary metabolites (i.e. phytotoxins) secreted by the pathogens, as in the case of Eutypa dieback and Esca (Andolfi *et al.*, 2011).

The majority of studies on host–plant interactions in the trunk disease complex rely on inoculations with a single pathogen (e.g. Camps *et al.*, 2010; Czemmel *et al.*, 2015). Nonetheless, grapevine wood is typically colonized by trunk pathogens and other wood-colonizing fungi, some of which are saprophytes or endophytes (Bruez *et al.*, 2014; Luque *et al.*, 2009; Péros *et al.*, 1999; Travadon *et al.*, 2016; Úrbez-Torres *et al.*, 2006). Such fungi are thought to interact with each other, possibly influencing disease symptoms and severity (Pierron *et al.*, 2016; Sparapano *et al.*, 2001; Whitelaw-Weckert *et al.*, 2013). For example, co-inoculations of the two Esca pathogens, *Phaeomoniella chlamydospora* and *Phaeoacremonium minimum*, resulted in more severe grapevine decline than did single-species inoculations. A similar additive interaction was observed in co-inoculations of *Ilyonectria macrodidyma* (one of the causal agents of Black leg) and *Diplodia seriata* (one of the causal agents of Botryosphaeria dieback) (Whitelaw-Weckert *et al.*, 2013). A negative interaction was reported between *Phaeoa. minimum* and *Fomitiporia punctata*, an Esca secondary pathogen (Sparapano *et al.*, 2001). These few examples of controlled inoculations help us understand a few idealized scenarios, but no study has attempted so far to dissect the *in planta* activities of the trunk pathogen community under vineyard conditions.

In recent years, next-generation sequencing (NGS)-based approaches have enabled the profiling of microbial communities on an unprecedented scale (Vernikos *et al.*, 2015). Methods based on massively parallel sequencing of short DNA fragments amplified by polymerase chain reaction (PCR) have become broadly applied (i.e. DNA barcoding; Schoch *et al.*, 2012). Genome shotgun sequencing of complex biological samples (i.e. metagenomics) has recently emerged as a more effective approach that overcomes some of the limitations of DNA barcoding (Tyson *et al.*, 2004). Metatranscriptomics, the shotgun sequencing of the community mRNAs, presents an even greater improvement for microbial ecology studies. Unlike methods that target microbial DNA, which cannot differentiate between viable and dead microorganisms, metatranscriptomics targets the metabolically active fraction of the microbiome (Bridge and Spooner, 2001; Fraissinet-Tachet *et al.*, 2013; Kuske *et al.*, 2015).

To date, none of these NGS-based approaches has been used to study the activities of pathogen communities associated with trunk diseases. A key impediment to the application of these methods *in planta* is the low abundance of microbial DNA and RNA relative to host nucleic acids; this results in low sequencing coverage for the entire microbiome, often less than 1% of the total sequencing output (Blanco-Ulate *et al.*, 2013d, 2015; Jones *et al.*, 2014). Insufficient sequencing depth limits the application of metagenomics and metatranscriptomics methods that rely on the *de novo* assembly of the metagenome and metatranscriptome (Scholz *et al.*, 2012). The implementation of methods that rely on sequence alignment to references comprising all species potentially associated with a sample is often problematic because of the scarce genomic information available for the species under analysis (Filippidou *et al.*, 2015). To overcome this limitation, we sequenced and assembled the genomes of the most common grapevine trunk pathogens (GTPs): *Eutypa lata*, *Neofusicoccum parvum*, *Dip. seriata*, *Phaeoa. minimum*, *Phaeom. chlamydospora* and *Diaporthe ampelina* (Blanco-Ulate *et al.*, 2013a,b,c; Morales-Cruz *et al.*, 2015). The predicted proteomes of each genome were annotated to obtain a comprehensive catalogue of all potential virulence functions associated with cell wall degradation, secondary metabolism and nutrient uptake (Morales-Cruz *et al.*, 2015).

Here, we utilized all the available genomic references for the most common GTPs to develop a metagenomics and metatranscriptomics approach that relies entirely on NGS read mapping onto a multi-species closed reference. In this work, we: (i) developed and optimized a bioinformatics pipeline to precisely align short reads from a biologically complex sample to a reference that comprises multiple species; (ii) tested the bioinformatics pipeline in controlled experiments with artificial inoculation of grape woody stems using single GTP species; (iii) applied the multi-species reference approach to vineyard samples from naturally infected vines showing a variety of trunk disease symptoms; (iv) compared the multi-species reference approach with the *de novo* assembly of metatranscriptomics data; and (v) analysed the expression of putative virulence functions at the GTP community level. Our results demonstrate that mapping-based metatranscriptomics can profile the *in planta* expression of thousands of putative virulence factors of multiple pathogenic species, thereby enabling the study of complex diseases under field conditions at the molecular level.

RESULTS AND DISCUSSION

Optimization of a multi-species closed-reference read mapping protocol using simulated datasets

To develop a reference-based metagenomics and metatranscriptomics approach for the *in planta* detection and quantification of trunk pathogens, we first assessed the specificity and sensitivity

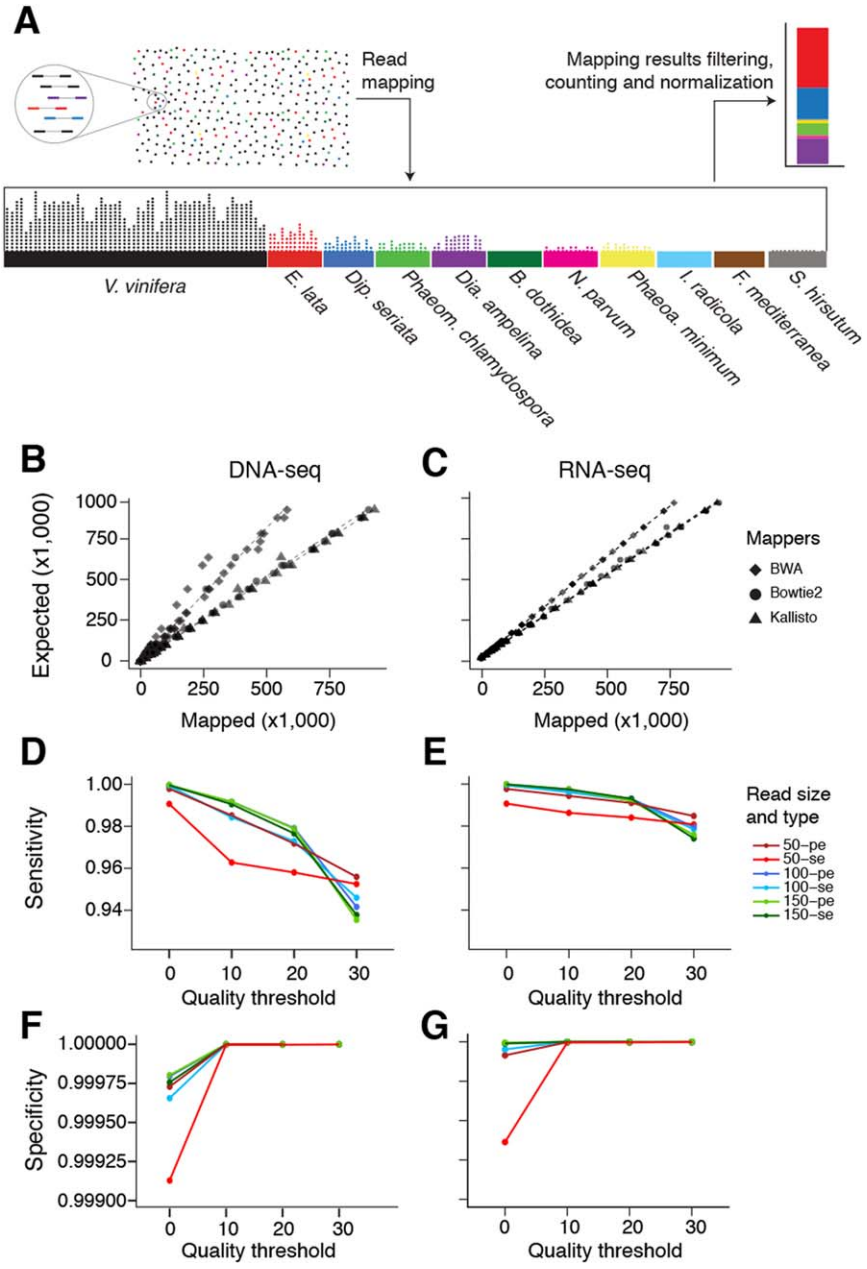

Fig. 1 Bioinformatics approach and optimization. (A) Schematic representation of the approach used to quantify sequencing reads for each co-infecting grapevine trunk pathogen species. Dots represent paired-end reads in a given sample, whereas colours indicate the species from which the reads were derived. Genomic references were combined into a single reference for read mapping. After mapping, the reads were filtered based on mapping quality, counted and normalized. Correlations of the number of simulated short reads with three different short-read mappers and the expected number of reads based on (B) DNA sequencing (DNA-seq) and (C) RNA sequencing (RNA-seq) data. Evaluation of sensitivity (D, E) and specificity (F, G) when using Bowtie2 with different sequencing parameters and mapping quality thresholds. pe, paired-end read; se, single-end read.

of popular mapping software using simulated datasets. Unlike traditional plant–microbe assays involving a single pathogen, for which RNA-sequencing (RNA-seq) methods are established and widely used (e.g. Blanco-Ulate *et al.*, 2013d, 2015; Jones *et al.*, 2014), we had to test software performance with short reads from complex biological samples aligned to a reference that comprises multiple species. The genome sequences of grape and the 10 species most commonly associated with grapevine trunk diseases (Table S1, see Supporting Information) were concatenated into a single multi-species genome reference (total size, 952 Mb; Blanco-Ulate *et al.*, 2013a, b, c; Floudas *et al.*, 2012; Jaillon *et al.*, 2007; Morales-Cruz *et al.*, 2015; Nordberg *et al.*, 2014) to map the metagenomic reads (Fig. 1A). Similarly, transcriptomes of the same species were concatenated to create a multi-species reference of 154 998 protein-coding sequences to map the metatranscriptomic reads. To assess mapping specificity and sensitivity, we generated synthetic Illumina reads from the genomes and transcriptomes of the same organisms as were used as references for read mapping employing ART, a NGS read simulator (Huang *et al.*, 2012). We generated a total of 40 independent metagenomics and metatranscriptomics simulated datasets by combining variable proportions of the simulated reads from *Vitis vinifera*

(80%–100% of the total reads per sample) and from the 10 fungal species (0%–20% of the total reads per sample; Table S2, see Supporting Information). Proportions between grape and fungal reads were chosen to reflect the composition of real infected grape wood samples (see below).

The reads of each simulated sample were mapped using three popular short-read aligners: BWA (Li and Durbin, 2009), Bowtie2 (Langmead and Salzberg, 2012) and Kallisto (Bray *et al.*, 2016). Because the taxonomic composition of each of the 20 simulated samples was known, we could determine the sensitivity and specificity of each aligner by comparing the expected and observed counts of the mapped reads (Fig. 1B,C). The sensitivity (i.e. mapping rate) was high with both Bowtie2 and Kallisto, which aligned 93.5% $\pm$ 5.12% and 98.7% $\pm$ 1.58% of the expected reads, respectively (Fig. 1B). The mapping rate of BWA was significantly lower (69.89% $\pm$ 11.70%; $P = 0.0019$). Rates of non-specific mapping were remarkably low in all tests: Bowtie2 and BWA mapped to the wrong species zero and six (3.97×10^{-7}%) reads, respectively, whereas Kallisto failed to map onto the correct species 11 377 reads (3.12×10^{-4}%). Similar patterns of sensitivity and specificity were observed when the simulated metatranscriptomic samples were mapped onto the concatenated multi-species transcriptome (Fig. 1C). We selected Bowtie2 for further optimization because of its high sensitivity and better specificity compared with the other aligners.

Using the simulated datasets, we evaluated the impact on mapping sensitivity and specificity of: (i) increasing the threshold of the minimum mapping quality (Q0, Q10, Q20 and Q30); (ii) different read lengths (50, 100 and 150 bp); and (iii) different sequencing modes (single- vs. paired-end; Fig. 1D–G). The weakest sensitivity and specificity were achieved with 50-bp single-end reads and no mapping quality threshold. The maximum specificity was reached at Q10, with little improvement when more stringent parameters were applied. Nonetheless, sensitivity suffered when increasing quality thresholds were applied. We observed a decline of 2.00% $\pm$ 1.37% in sensitivity from Q20 to Q30 with only marginal improvement in specificity. Concerning the impact of read length and sequencing mode, the highest specificity was achieved with 150-bp paired-end reads, but with little difference with other combinations. Based on these results, all the analyses described below were carried out using Bowtie2 and a mapping quality cut-off of Q20.

Controlled inoculations of single pathogen isolates confirm the specificity of the multi-species closed-reference mapping approach

To test the selected and optimized mapping method on infected grapevine samples, we collected and sequenced RNA from vines artificially inoculated with individual GTPs. The woody stems of 12 vines were inoculated with mycelial plugs of either *N. parvum* (isolate UCD646So) or *Phaeoa. minimum* (isolate UCR-PA7). From each plant, a stem sample at the inoculation site ($\pm$2 cm from the point of inoculation) was processed for RNA extraction and sequencing. In all samples, the inoculated species were detected by metatranscriptomics analysis as the most abundant, confirming the specificity of mapping to a multi-species reference (Fig. 2A,B). On average, 90.74% $\pm$ 3.90% and 66.61% $\pm$ 15.83% of the total fungal reads (i.e. non-*V. vinifera* reads) mapped on *N. parvum* and *Phaeoa. minimum*, respectively. Although it was not part of the inoculations, the Esca pathogen *Phaeom. chlamydospora* was detected in all assayed plants, albeit at a much lower level than the inoculated pathogens (Fig. 2A,B). On average, *Phaeom. chlamydospora* had 15.9 and 4.46 times fewer reads than the most abundant species in the samples inoculated with *Phaeoa. minimum* and *N. parvum*, respectively. Given that *Phaeom. chlamydospora* has been isolated from asymptomatic grapevine propagation materials (Eskalen *et al.*, 2007; Ferreira *et al.*, 1999; Mugnai *et al.*, 1999), we assume that it was present before inoculation. This hypothesis is consistent with its detection in all 12 samples and with little variation in read counts (5562.08 $\pm$ 711.74 reads).

To further assess the specificity when reads from multiple species were mapped, we concatenated the reads obtained from both *N. parvum* and *Phaeoa. minimum* inoculations, and mapped the combined reads to the multi-species reference. The simultaneous mapping of the combined dataset led to identical results as when reads were mapped separately ($R = 1.00$; Fig. 2C,D). Both results, the quantitative determination of the inoculated species as the most abundant species and the absolute specificity when mapping reads from multiple species, confirmed that the optimized method is able to assign the reads to the correct species in real experimental data.

Application of the multi-species closed-reference mapping approach to field samples

With the objective of studying GTPs under vineyard conditions, we then applied the multi-species mapping approach to samples taken from mature vines (>8 years old) showing a variety of the most common symptoms associated with trunk diseases (Table S3, see Supporting Information). We collected 28 wood samples from distinct plants with the following combinations of symptoms: Eutypa dieback foliar and wood symptoms (eight samples; Fig. 3A,C), Esca foliar and wood symptoms (eight samples; Fig. 3B,C), wood symptoms and no foliar symptoms (six samples; Fig. 3C) and apparently healthy plants with no foliar or wood symptoms (six samples; Fig. 3D). For samples with wood symptoms, the tissue at the margin of the necrotic region was collected as shown in Fig. 3C. For samples from apparently healthy plants, tissue without disease symptoms was collected from similar locations in the wood as the diseased samples (Fig. 3D). DNA and

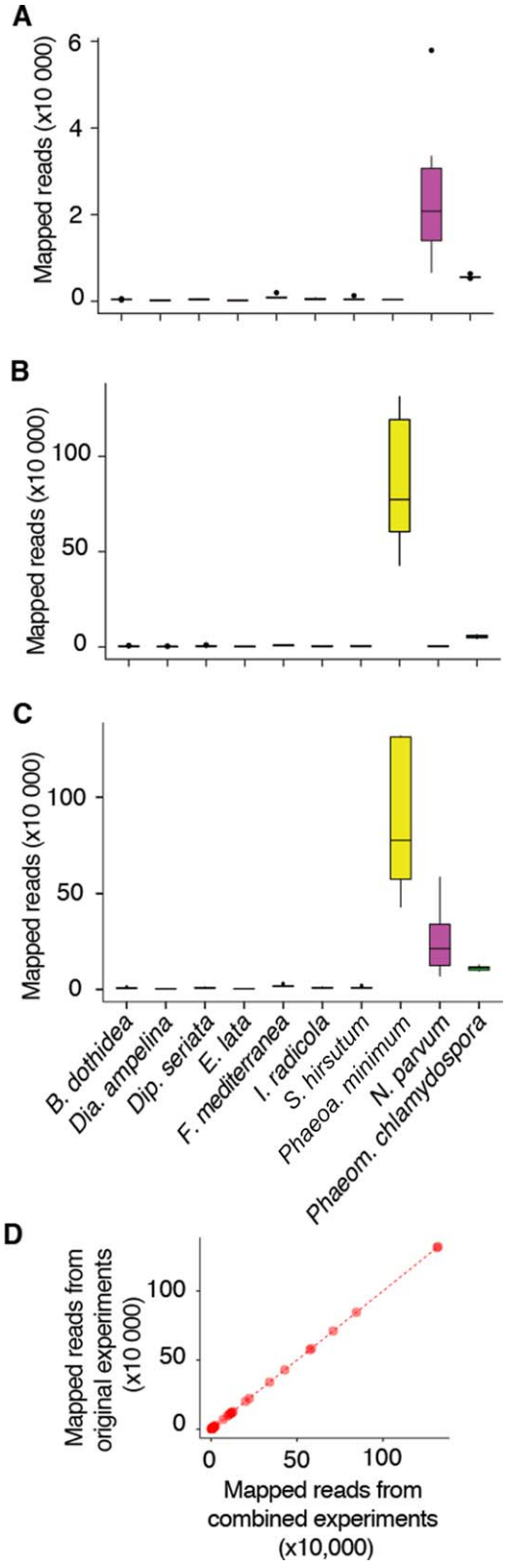

Fig. 2 Evaluation of reference-based metatranscriptomics using woody samples inoculated with selected wood pathogens. Grapevine woody stems were inoculated with either *Neofusicoccum parvum* (A) or *Phaeomoniella minimum* (B), and total RNA sequenced. Reads from both experiments were concatenated and mapped together to assess the specificity when multiple species were present in the dataset (C, D). Boxplots (A–C) show the distribution of read counts mapped onto the different pathogens on the multi-species reference. (D) Correlation between counts of mapped reads between individual and combined read sets.

RNA were extracted from each sample, and used for metagenomics (DNA sequencing, DNA-seq) and metatranscriptomics (RNA-seq) analysis. The same samples were also used for culture-based identification. Fungal colony purification led to the identification of up to two GTPs per sample (Fig. S1, see Supporting Information), which corresponded to the foliar symptoms. *Eutypa lata* was isolated from all vines with Eutypa dieback symptoms. *Phaeomoniella chlamydospora* and *Phaeoa. minimum* were recovered from 50% and 38%, respectively, of vines with Esca symptoms. From vines with no foliar symptoms, wood symptoms were associated with the presence of *Phaeom. chlamydospora*, *N. parvum*, *Dip. seriata* and *Dia. ampelina*, but not *E. lata*. *Phaeomoniella chlamydospora*, *N. parvum* and *Dip. seriata* were also recovered from vines with Eutypa dieback and Esca symptoms.

For metagenomics and metatranscriptomics analysis of the 28 samples, we generated an average of 57.5 ± 16.5 million DNA-seq and RNA-seq reads per sample, of which 52.0 ± 14.0 million reads were retained after quality and size trimming (Tables S4 and S5, see Supporting Information). A high proportion of the reads were mapped to the multi-species reference (83.17% ± 11.35% of DNA-seq reads and 87.80% ± 4.70% of RNA-seq reads). The majority of the reads were mapped onto *V. vinifera* (95.72% ± 4.45% and 99.99% ± 0.01% from plants with and without foliar symptoms, respectively). A greater fraction of reads mapped onto the GTP references when samples were taken from plants with foliar symptoms (4.28% ± 4.45%) compared with those without (0.01% ± 0.01%; Fig. 4A). The reads that did not map to the multi-species reference may derive from organisms that are not included in the reference, as suggested by the taxonomic analysis of *de novo* assembled contigs described below. To quantify the taxonomic composition of the samples, we removed the reads that mapped onto the grape genome and normalized the counts to account for both uneven sequencing throughput and differences in genome or transcriptome sizes between species. Species composition profiles were then constructed using the normalized counts (Fig. 4B,C). The detected species and their relative abundances matched what we expected based on the visible foliar symptoms and what we know about their aetiology. *Eutypa lata* was the predominant species in the samples from vines with Eutypa dieback symptoms with 83.05% ± 10.68% of the reads assigned to the GTP. In all of

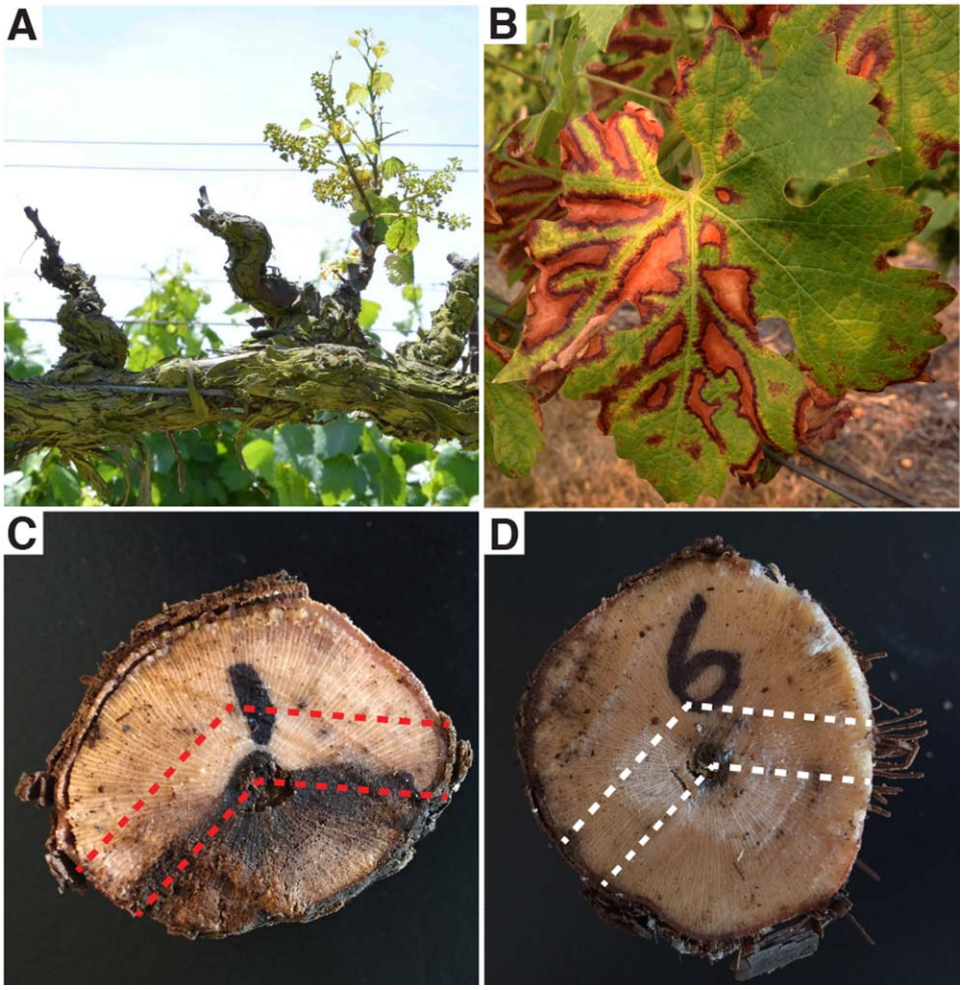

Fig. 3 Symptoms of trunk disease infections that were employed to classify the field samples used in this study. Foliar symptoms of Eutypa dieback (A) and Esca (B), wood cankers from vines without any foliar symptoms (C) and apparently healthy (i.e. non-symptomatic) wood (D). In (C) and (D), broken lines depict the areas in which RNA and DNA were extracted. In wood from symptomatic plants, tissue was collected at the margin of the cankers (red; C). In apparently healthy vines, tissue was collected in similar locations as in diseased vines (white, D).

these samples, however, *Phaeom. chlamydospora* (7.49% ± 8.47%) and *Dip. seriata* (6.69% ± 3.93%) were also detected at high levels. Overall, the results of metagenomics and metatranscriptomics revealed a greater species complexity than suggested by fungal isolation. This may be a result of the fact that culture-based identification can favour fungi that grow rapidly (Bridge and Spooner, 2001). The advantage of applying a sequencing method *in planta* was particularly evident in samples from vines with Esca foliar and wood symptoms and vines with no foliar symptoms. In the former, both Esca pathogens were consistently detected as the predominant species in all samples, together with smaller amounts of *Dia. ampelina*. Vines with wood symptoms and no foliar symptoms had by far the most complex and variable composition, ranging from mostly *Dia. ampelina* to mostly *Phaeom. chlamydospora*, or different proportions of these two species in combination with *Dip. seriata*. Much lower fungal counts, mostly associated with *Phaeom. chlamydospora* and *Phaeoa. minimum*, were detected in apparently healthy (i.e. no foliar or wood symptoms of trunk diseases) plants. Although both of these species have been reported from apparently healthy plants (Bruez *et al.*, 2014; Casieri *et al.*, 2009; Hofstetter *et al.*, 2012; Pancher *et al.*, 2012), the quantitative nature of the mapping method determined that the symptomatic samples had close to 197 times more reads, on average, than the apparently healthy plants (Fig. 4A). These results clearly show that, even if potentially pathogenic species can be detected in healthy samples, their abundance is significantly greater in plants with disease symptoms. Furthermore, this suggests that pathogen detection methods aiming to differentiate between the early and late stages of infection should be quantitative.

Species profiles determined by metagenomics showed a significant linear correlation with the results of metatranscriptomics ($R = 0.84$; $P = 2.2e^{-16}$; Fig. 4F). This correlation is an independent validation of the relative amount of each species using the two different methods. The two methods in fact do not only involve two different types of nucleic acid, but also different extraction methods, library preparation and mapping references. Some of the differences between the results of DNA-seq and

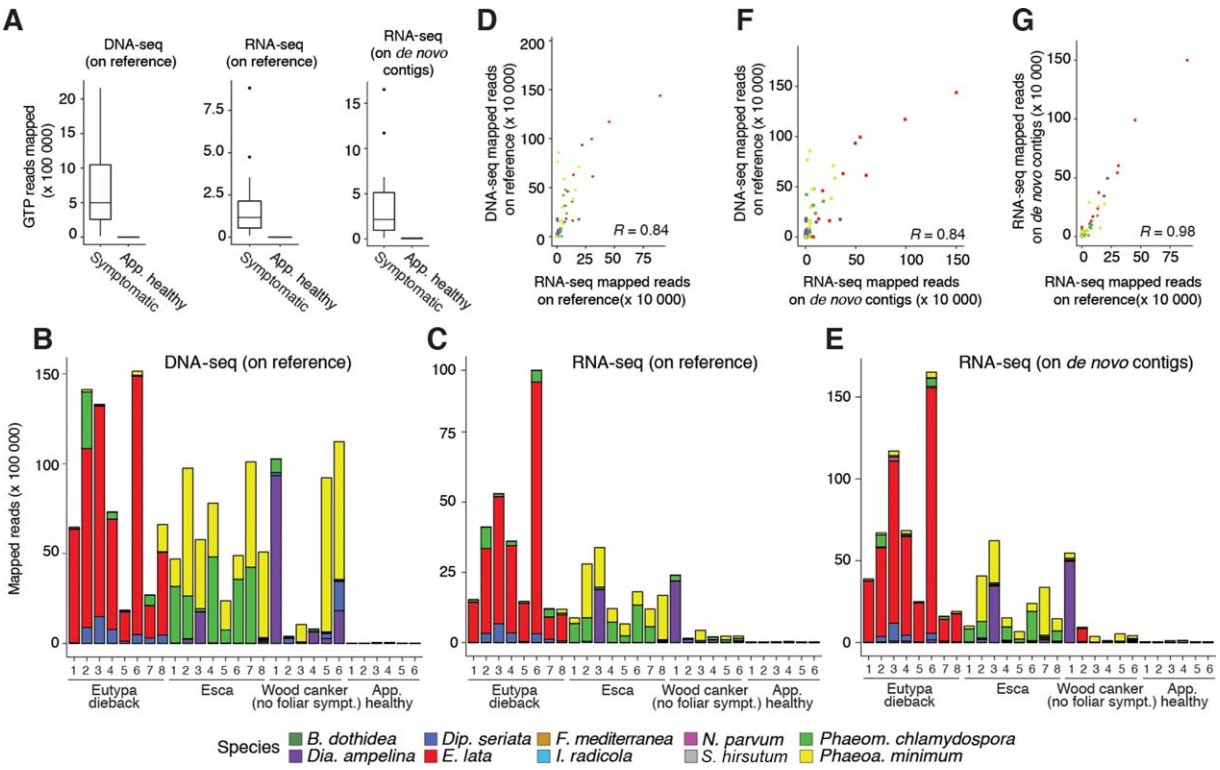

Fig. 4 Metagenomics and metatranscriptomics analyses of field samples. (A) Boxplots show the distribution of counts of mapped reads onto the multi-species genome and transcriptome or *de novo* assembled contigs. Stacked bars show the counts of mapped reads per trunk pathogen species in the reference from DNA sequencing (DNA-seq) (B) and RNA sequencing (RNA-seq) (C) analyses. (D) Scatterplot shows the correlation of counts per species determined using DNA-seq and RNA-seq. (E) Stacked bars show the counts of mapped reads on *de novo* assembled contigs assigned to the different trunk pathogen species. (F, G) Scatterplots show the correlation of counts per species determined using reference-based approaches (DNA-seq in F, RNA-seq in G) and mapping on *de novo* assembled contigs.

RNA-seq mapping values per species may be caused by the different metabolic activities of the fungi present in the sample at the time of collection. This was particularly evident in the samples with wood symptoms and no foliar symptoms, in which fungal DNA and RNA quantities showed a lower correlation.

Metatranscriptomics based on *de novo* assembly validates the multi-species closed-reference mapping approach

A potential limitation of the reference-based approach described above is that mapping is restricted to a closed reference. This may force reads to align to the concatenated transcriptomes even if they derive from species not included in the multi-species reference. To evaluate the impact of restricting the mapping of the metatranscriptomics reads to a closed reference, we compared the results of the multi-species closed-reference approach with results of read mapping onto *de novo* assembled contigs (Fig. S2, see Supporting Information). Transcripts of the 28 samples were assembled *de novo* using the quality-filtered RNA-seq reads. As a pre-filtering step, we excluded from the assembly process the reads that mapped to the transcriptome of *V. vinifera*. A total of 662 million reads that did not map to the *V. vinifera* reference were assembled into 2.3 million contigs using MEGAHIT (Li *et al.*, 2015). An average of 80.9 $\pm$ 18.3 million contigs per sample were obtained. We then assigned contigs to taxonomic groups by homology to the National Center for Biotechnology Information (NCBI) protein RefSeq database using BLASTX (e-value $< 1e^{-6}$). To avoid false taxonomic assignments as a result of missing information in the database, we limited the taxonomic assignments to the genus level. Despite the pre-filtering steps, an average of 85.01% $\pm$ 11.87% of the contigs were still assigned to *V. vinifera* proteins in RefSeq. A total of 167 fungal genera were detected across the 28 samples (Table S6, see Supporting Information), including nine of the ten genera of the multi-species reference (all except *Botryosphaeria*). Asymptomatic samples showed significantly smaller amounts of fungal contigs (0.16% $\pm$ 0.08% of the total contigs) compared with the symptomatic samples (10.99% $\pm$ 6.50% of the total contigs; $P = 1.82e^{-6}$). On average, 46.64% $\pm$ 25.99% of the fungal contigs were assigned to the GTP present in the closed reference. Overall, the assembled GTP transcripts were shorter (663.35 $\pm$ 27.50 bp) than the

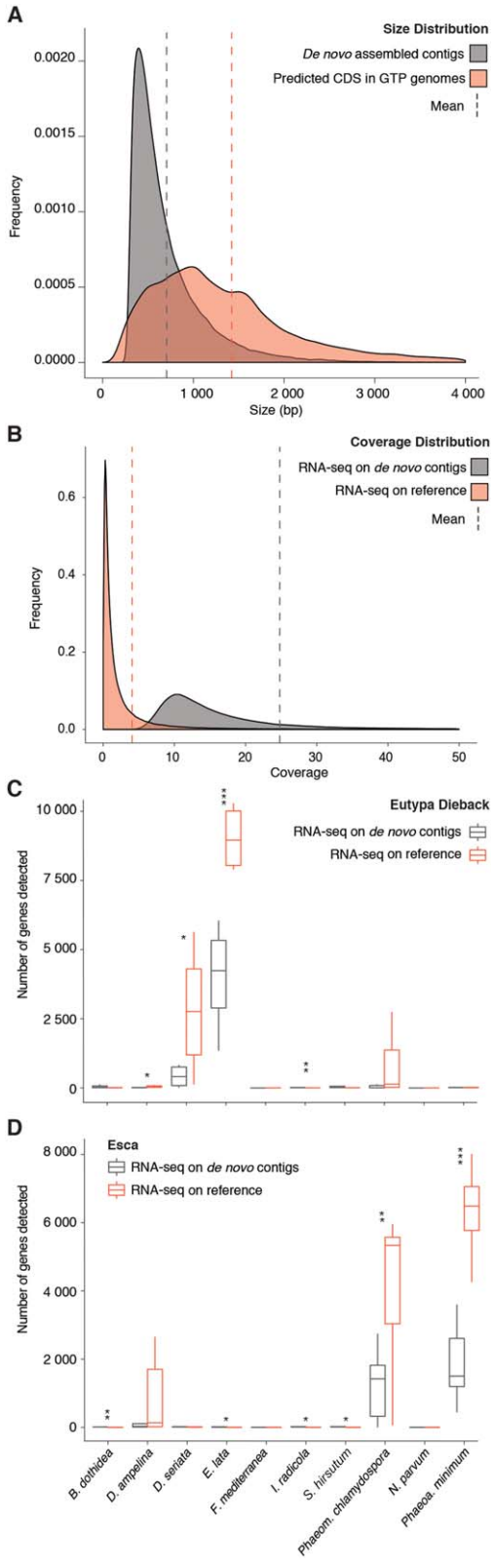

Fig. 5 Evaluation of the metatranscriptomics using the reference-based method or *de novo* assembly. Size (A) and RNA-seq read coverage (B) distribution of contigs *de novo* assembled and the protein-coding sequences (CDS) predicted from the genomic references. (C, D) Boxplots summarizing the distribution of the number of detected protein-coding genes assigned to each pathogen species in samples with Eutypa dieback (C) and Esca (D) symptoms. Asterisks indicate significant differences in means: $*P \leq 0.05$; $**P \leq 0.01$; $***P \leq 0.001$.

predicted coding sequence (CDS) in the GTP genomes (1422.30 ± 1076.37 bp), indicating fragmentation and potential redundancy of the fungal transcripts in *de novo* assembled contigs (Fig. 5A). The rest of the contigs (53.36% ± 25.99% on average per sample) were assigned to 158 non-GTP taxa (Table S6 and Fig. S3, see Supporting Information).

The abundance (i.e. expression level) of each *de novo* assembled contig was measured by mapping the metatranscriptomic reads onto the contigs. An average of 0.57 ± 0.78 million reads per sample were mapped to *de novo* contigs assigned to GTPs. The median mapping coverage per contig was 13.8× (Fig. 5B), approximately three-fold higher than the median coverage obtained by the closed-reference-based approach. The overall higher coverage of mapping on *de novo* assembled contigs suggests that transcripts with higher expression levels were preferentially assembled by MEGAHIT. This hypothesis is also supported by the greater number of genes per sample detected by the closed-reference-based approach (8711.46 ± 6055.12 CDS/sample) compared with the number of mapped *de novo* contigs (7522.36 ± 6889.68 contigs/sample). The smaller number of contigs compared with the total CDS used in the reference-based approach may have also contributed to increase the median coverage. To estimate the difference in the number of genes detected by the two metatranscriptomics approaches, we assigned homology between the *de novo* assembled contigs and the predicted CDS that composed the multi-species reference using BLASTN (e-value < 1e-6; Fig. 5). In most cases, the reference-based approach detected a significantly larger number of CDS than the *de novo* assembly approach (Fig. 5). Based on these results, we can conclude that the *de novo* assembly approach yielded a narrower representation of the metatranscriptome limited to those transcripts with sufficient coverage to allow assembly.

The taxonomic profiles obtained using the *de novo* assembly were very similar to those obtained using the reference-based metatranscriptomic ($R = 0.98$; $P = 2.2e^{-16}$; Fig. 4C,E,G) and metagenomic ($R = 0.84$; $P = 2.2e^{-16}$; Fig. 4B,F) data. These results confirmed that the multi-species reference-based approach did not lead to false taxonomic assignments, at least at the genus level, because of the closed reference. On average, 53.33% ± 19.58% of the reads mapped to contigs from GTPs. *Phaeomoniella*, *Eutypa*, *Diaporthe* and *Phaeoacremonium* were often the predominant genera (Fig. S3). For example, the genus

Phaeomoniella accounted for up to 93.1% of reads in sample AH2, *Eutypa* for 64.9% in sample ED7, *Diaporthe* for 60.6% in sample WC1 and *Phaeoacremonium* for 51.6% in sample WC3. Most of the remaining 158 non-GTP genera contributed individually to less than 5% of the total number of reads per sample and collectively to 35.21% ± 12.55%. The few genera that contributed to more than 5% in at least one sample (Fig. S3) included known plant pathogens or endophytes, such as *Aspergillus*, *Pestalotiopsis*, *Leptosphaeria*, *Stagonospora*, *Bipolaris*, *Pyrenophora* and *Setosphaeria* (Eyal, 1999; Howlett *et al.*, 2001; Kumar *et al.*, 2002; Lamari and Bernier, 1989; Li and Strobel, 2001; Munkvold and Marois, 1995; Nierman *et al.*, 2005; Perkins and Pedersen, 1987), as well as other ubiquitous fungi (Corte *et al.*, 2015; Feng *et al.*, 2014; Longcore *et al.*, 1999). Overall, the greater contribution of GTP genera to the total read count per sample suggests a dominant role of these genera in the fungal community.

In conclusion, although the *de novo* assembly approach validated the results of the reference-based approach, it provided a narrower and more fragmented representation of the GTP community's metatranscriptome. These limitations, combined with a more complex pipeline (Fig. S2), longer processing time and more intensive computational requirements, provides further support to the reference-based approach as the more effective method for profiling transcriptional activities *in planta* of a fungal community of known composition.

Multi-species closed-reference-based metatranscriptomics allows the *in planta* profiling of virulence function expression

Phytotoxic metabolites and cell wall-degrading enzymes are considered to be key pathogenicity and virulence factors underlying trunk disease development (Andolfi *et al.*, 2011; Bertsch *et al.*, 2013; Rolshausen *et al.*, 2008). We have previously annotated all predicted protein-coding genes of all GTP genomes included in the present work (Morales-Cruz *et al.*, 2015). Functional annotations were assigned based on the presence of conserved domains, as well as homology to proteins from relevant specialized databases focused on potential virulence factors involved in primary and secondary cell wall decomposition (i.e. carbohydrate active enzymes, CAZymes), secondary metabolism (i.e. cytochrome P450s, biosynthetic gene clusters) and nutrient uptake (i.e. transporters). The first study that profiled the transcriptional activities of these virulence factors during experimental infections with *N. parvum* of grapevine woody stems revealed that physically clustered genes coding for putative virulence functions share common regulatory sequences and are induced depending on the substrate or stage of plant infection (Massonnet *et al.*, 2016). In this work, we utilized the optimized reference-based metatranscriptomics approach to obtain information on the *in planta* expression of virulence genes when multiple GTPs co-infect the same host under vineyard conditions.

We first extracted the normalized mapping counts of all genes assigned to any of the five broad categories of putative virulence factors (CAZymes, cytochrome P450s, peroxidases, genes belonging to biosynthetic gene clusters and transporters; Table S7, see Supporting Information). A total of 530 functions were identified from 12 951 different genes across all samples and GTP species. CAZymes and transporters were the most expressed functions, followed by secondary metabolism, cytochrome P450s and peroxidases in all pathogens (Fig. 6A). Expression differences between different potential virulence factors became clear when we analysed the expression patterns of specific functions. A principal component analysis (PCA) based on the expression data of each virulence function clearly separated samples based on the type of disease symptoms with which they were associated (Fig. 6B). Separations by the two principal components were confirmed by a partial least squares discriminant analysis (PLS-DA, $R^2Y = 0.988$, $Q^2Y = 0.903$, Fig. S4, see Supporting Information). These results suggest that pathogen species associated with the same disease, even if phylogenetically distant, as in the case of Esca pathogens, activate similar virulence functions at comparable expression levels.

To explore some of the functions most distinctively expressed between samples with Esca and Eutypa dieback symptoms, we extracted the major loadings separating the two groups based on the PLS-DA. For Eutypa dieback samples, the major loadings comprised genes potentially associated with cell wall degradation, secondary metabolism and toxin secretion (Fig. 6C): a Versatile Peroxidase (VP), two CAZymes (GT74 and CMB23), five cytochrome P450s and two transporters. VPs have been shown to degrade lignin in other fungi (Camarero *et al.*, 1999; Ruiz-Dueñas *et al.*, 2009). GT74 and CMB23 CAZymes have α-1,2-L-fucosyltransferase- and mannan-binding functions, and therefore may potentially be involved in host cell wall manipulation (Perrin *et al.*, 1999; Stoll *et al.*, 2000). The identified cytochrome P450s included CYP560, CYP551 and CYP619, whose activities have been associated with fungal secondary metabolism (Artigot *et al.*, 2009; Moktali, 2013). In *Aspergillus clavatus*, CYP619 has been shown to participate in the biosynthesis of the toxin patulin (Artigot *et al.*, 2009). One of the two transporters identified belonged to the Autotransporter-1 Family (AT-1, code 1.B.12), which may facilitate the secretion of toxins, as seen in bacteria (Tang and Saier, 2014).

Among the functions that more strongly contributed to separate Esca samples, we also found interesting putative virulence functions. These included the following: a transporter from the family Immunophilin-like Prolyl:Peptidyl Isomerase Regulator (I-PPI, code 9.A.14), which has been extensively studied for its role in the exclusion of antifungal drugs in yeast (Arevalo-Rodriguez *et al.*, 2004); CYP542, which has been associated with secondary metabolism (Moktali, 2013); and CYP655, similar to the

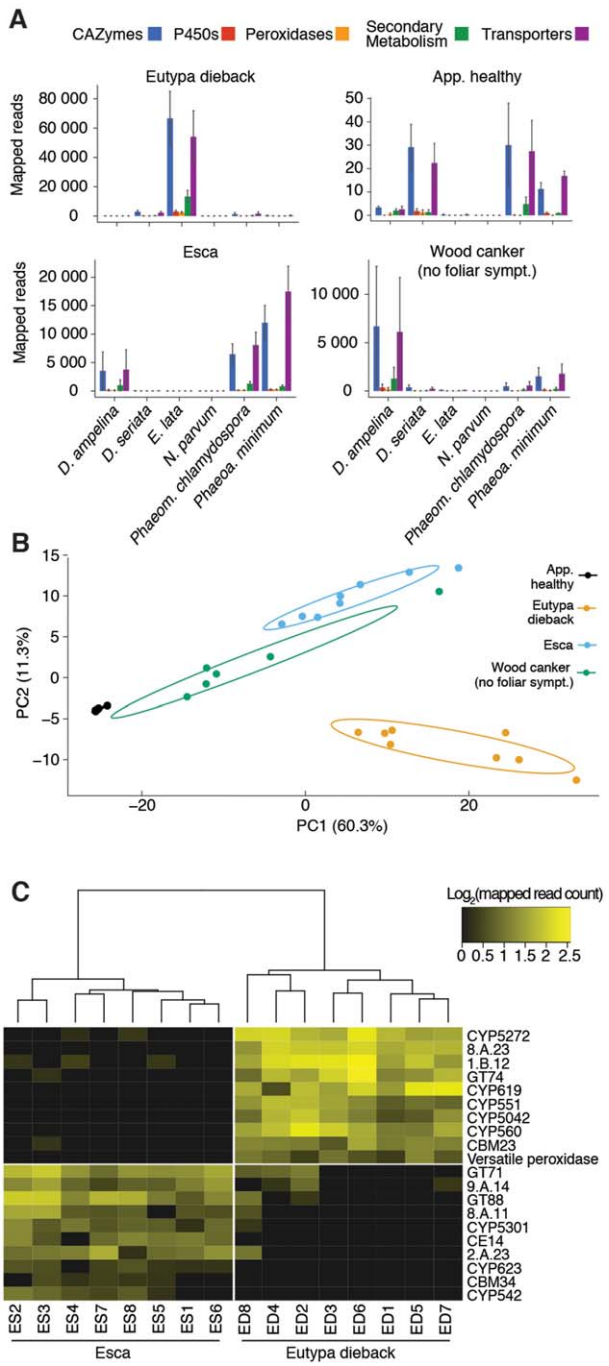

Fig. 6 Virulence function expression based on metatranscriptomics analysis. (A) Barplots showing the expression levels measured as normalized read counts of genes grouped by broad functional categories with potential virulence activity for each trunk pathogen species. Principal Component Analysis using as input the sum of reads for each samples that mapped on all genes the share the same Pfam annotation (B). (C) Heatmap showing the expression levels of specific functions identified as major PLS-DA loading separating samples with Esca and Eutypa dieback symptoms.

polyketide synthase (PKS)-NRPS hybrid responsible for the biosynthesis of the tenellin toxin in *Beauveria bassiana* (Xiao *et al.*, 2012). CAZymes GT71 and CBM34 with potential α-mannosyltransferase and granular starch-binding activities, respectively, were also characteristic of the Esca samples, and may be associated with the degradation of host cell wall or storage polysaccharides (Crucello *et al.*, 2015; Machovič and Janeček, 2006).

CONCLUSIONS

In this study, we have shown that existing mapping software can be optimized and applied to study fungal communities associated with grapevine trunk diseases. We have demonstrated that a closed reference that includes the most important species associated with a complex biological system can effectively overcome the limitation of the use of *de novo* assembled contigs and take advantage of the existing gene models and functional annotations. This multi-species reference can be expanded as more genomes of fungi inhabiting grapevines are sequenced. The highly specific and quantitative nature of the mapping approach can find useful applications for disease diagnostics in both production vineyards and nurseries. This metatranscriptomics approach can not only be applied to field samples, as shown here, but also to controlled co-inoculations with different combinations of multiple pathogens under the same environmental conditions to determine the patterns of gene expression as disease symptoms develop in the different organs. Community-level transcriptional analysis integrated with chemical analysis of secreted toxins will help to determine the relative contribution of each co-infecting agent to the development of different types of trunk disease.

EXPERIMENTAL PROCEDURES

Optimization of read mapping onto a multi-species closed reference

The genome and transcriptome references available for grapevine trunk-associated fungi were concatenated and used as a multi-species reference. This reference was used as input in the ART simulator (Huang *et al.*, 2012) to generate *in silico* reads with the characteristics of reads produced by an Illumina (San Diego, CA, USA) HiSeq 2500 sequencer. From the simulated reads, 20 synthetic samples were produced with different proportions of each species to test the mapping method under different scenarios. Reads were mapped with Bowtie2, BWA and Kallisto with different size lengths (50, 100 and 150 bp) and mapping quality filters (Q0, Q10, Q20, Q30). Each combination was evaluated by sensitivity and specificity values.

Controlled inoculation of *Phaeoa. minimum* and *N. parvum*

In March 2014, 1-year-old dormant cuttings of *V. vinifera* 'Cabernet Sauvignon' clone 29 and 'Merlot' clone 15 were cut to uniform length

(~10 cm) containing two nodes. Cuttings were surface sterilized in 1% sodium hypochlorite for 15 min, soaked in water overnight and stored in hermetic plastic bags in a cold room (2 °C). In April 2014, on the day of inoculations, each cutting was wounded at approximately 3 cm below the uppermost node with a 3-mm cork borer. A 3-mm mycelial plug from a 3-day-old culture of *N. parvum* isolate UCD646So or from a 10-day-old culture of *Phaeoa. minimum* (syn. *Togninia minima*) isolate UCR-PA7 was inserted into the wound, and sealed with vaseline (Unilever, London, UK) and Parafilm (Bemis Co., Neenah, WI, USA). Non-inoculated controls were wounded and 'mock-inoculated' with an agar plug from a sterile Petri plate of potato dextrose agar (PDA; Difco Laboratories, Detroit, MI, USA). After inoculation, cuttings were submerged in melted paraffin wax within 4 cm of the roots to prevent moisture loss as roots and shoots formed. Cuttings were potted in a mixture of perlite and vermiculite (1 : 1) in aerated plant bands ($5 \times 5 \times 20$ cm^3; Monarch Manufacturing Inc., Salida, CO, USA) held in plastic trays with a partially opened bottom allowing appropriate aeration and drainage. Inoculated plants were supplied with bottom heat to enhance the rooting process by placing plastic containers onto heating pads; the minimal rooting temperature was maintained above 24 °C at night. Plants were incubated in the glasshouse [natural sunlight photoperiod, 25 ± 1 °C (day), 18 ± 3 °C (night)]. Two replicate experiments were conducted over two successive weeks in two distinct glasshouses. In each experiment, plants were arranged in a completely randomized design, with six inoculated and six mock-inoculated plants per cultivar.

Because the pathogens colonize wood at different rates, infection was confirmed at 2 and 10 months post-inoculation for *N. parvum* and *Phaeoa. minimum*, respectively, by observation of an internal lesion (i.e. wood discoloration surrounding the inoculation site). First, the green shoots, roots and bark of each plant were removed and discarded, and the woody stems were surface sterilized in 1% sodium hypochlorite for 2 min and rinsed with deionized water. Then, the shoot was cut longitudinally to reveal the lesion. To confirm that the pathogens were responsible for the lesion's presence in inoculated plants, and its absence from non-inoculated plants, pathogen recovery was attempted by cutting 10 pieces ($2 \times 5 \times 5$ mm^3) of wood from the distal margin of the lesions, followed by surface disinfestation in 0.6% sodium hypochlorite (pH 7.2) for 30 s, rinsed twice for 30 s in sterile deionized water, plated on PDA amended with tetracycline (5 mg/mL) and incubated in the dark at 22 °C for 14–21 days. Wood cylinders approximately 2 cm in length were collected with flame-sterilized forceps immediately beneath the inoculation site, placed in liquid nitrogen and stored at −80 °C until processing for nucleic acid extraction.

Symptomatic and non-symptomatic sample collection from mature vines

In 2014 and 2015, 28 wood samples were collected from vineyards affected by trunk diseases. One wood sample was collected per vine, so that a total of 28 vines were sampled. The vineyards were located in different California grape production areas and included wine grapes cv. Pinot Meunier (14 samples) located in Sonoma County, and table grapes cvs. Flame (seven samples), Dovine (one sample), Thompson Seedless (three samples) and Crimson (three samples) located in Fresno County. Vines were selected based on wood and/or foliar disease symptom expression, as described previously (Gubler *et al.*, 2015; Rolshausen *et al.*, 2015). Each collected sample was cut into two wood pieces: one half was used for the culture-based fungal identification, whereas the other half was submerged in liquid nitrogen and stored at −80 °C for nucleic acid extraction.

Culture-based identification method

Fungal isolates were recovered from woody necrotic areas (i.e. showing signs of streaking, canker or browning) on PDA amended with tetracycline (100 ppm), with two plates per sample. Wood chips (approximately $3 \times 3 \times 3$ mm^3 in size) were removed from the necrotic areas with a sterile blade, disinfested in 10% bleach (sodium hypochlorite) for 2 min and rinsed twice in distilled water for 2 min. Plates were incubated at room temperature in the dark and inspected several times per week for 2 weeks. Fungal isolates with culture morphologies typical of *Botryosphaeria*, *Diaporthe*, *Diplodia, Eutypa*, *Neofusicoccum, Phaeoacremonium* and *Phaeomoniella* were hyphal tip purified and transferred to PDA. We recovered a total of six isolates from wood samples that were further identified by ITS rDNA sequencing. DNA was extracted from mycelium scraped from the surface of 14-day-old cultures grown at room temperature (DNeasy® Plant Kit; Qiagen, Valencia, CA, USA), following the manufacturer's instructions. The nuclear loci rDNA Internal Transcribed Spacers (ITS1/5.8S/ITS2) were amplified using PCR primers ITS1 and ITS4 (White *et al.*, 1990). PCR was performed with the following cycling parameters: one cycle at 94 °C for 5 min; 35 cycles at 94 °C for 1 min, 58 °C for 1 min and 72 °C for 1.5 min; and a final elongation step at 72 °C for 5 min. PCR products were sequenced in both forward and reverse directions (Genomic Core Sequencing Facility, University of California, Riverside, CA, USA). BLASTn searches of GenBank identified homologous sequences with high identity.

Nucleic acid extraction and sequencing library preparation

Frozen wood pieces were ground to a fine powder using a TissueLyser II (Qiagen) in stainless steel jars frozen in liquid nitrogen. Part of the ground tissue was employed to extract DNA using a modified version of the protocol described in Stoffel *et al.* (2012) with 200 mg of initial ground tissue. The DNA was assayed for concentration, purity and integrity with Qubit (Life Technologies, Eugene, OR, USA), Nanodrop 2000 (Thermo Scientific, Waltham, MA, USA) and an agarose 1X gel, respectively. High-quality DNA was fragmented by sonication using a Bioruptor Diagenode (Denville, NJ, USA). Fragmented DNA was employed as template for library preparation using a KAPA Biosystems Illumina (Boston, MA, USA) kit. Size selection after ligating adapters was performed with an eGel system (Invitrogen, Carlsbad, CA, USA). Final libraries were sequenced in an Illumina HiSeq3000 as 150 bp paired ends. Another aliquot of ground wood tissue was employed to extract RNA using the protocol described in Blanco-Ulate *et al.* (2013d). RNA concentration and purity were evaluated similarly to that of DNA, and integrity was evaluated with a 2% agarose gel. RNA was used as input for the Illumina TruSeq Kit for library preparation. Final libraries were sequenced in an Illumina HiSeq3000 as 150 bp paired ends.

De novo assembly of metatranscriptomic data and taxonomic assignment

Quality filtered reads of the metatranscriptomic data were mapped to the *V. vinifera* genome reference using Bowtie2 (as described above). The reads that did not map to the *V. vinifera* genome were used as input for the assembler MEGAHIT (Li *et al.*, 2015). The software was run in the meta-sensitive mode and only contigs larger than 300 bp were retained from the contigs generated. The reads used to generate the contigs were mapped with Bowtie2 onto the assembled contigs to determine the contig expression levels. Contigs with counts of <10 reads and/or that aligned (e-value $< 10^{-6}$) to *V. vinifera* were removed. The RefSeq protein database of fungi and *V. vinifera* was used as database to assign taxonomic memberships to the *de novo* assembled contigs using BLASTX (e-value $< 10^{-6}$). The taxonomy of the protein with the best hit of each contig was extracted and used to assign a genus to the aligned contig. Finally, the number of reads per taxonomic unit was used to create an abundance profile based on the number of reads.

ACKNOWLEDGEMENTS

This work was supported by the American Vineyard Foundation (grants 2014-1798 and 2015-1798) and the USDA, National Institute of Food and Agriculture, Specialty Crop Research Initiative (grant 2012-51181-19954). A.M.-C. was also partially supported by The Horace O. Lanza Scholarship, The Pearl & Albert J. Winkler Scholarship in Viticulture and the André Tchelistcheff and Dr Richard Peterson Scholarship.

REFERENCES

Allen, E.E. and Banfield, J.F. (2005) Community genomics in microbial ecology and evolution. *Nat. Rev. Microbiol.* **3**, 489–498.

Andolfi, A., Mugnai, L., Luque, J., Surico, G., Cimmino, A. and Evidente, A. (2011) Phytotoxins produced by fungi associated with grapevine trunk diseases. *Toxins*, **3**, 1569–1605.

Arevalo-Rodriguez, M., Wu, X., Hanes, S.D. and Heitman, J. (2004) Prolyl isomerases in yeast. *Front Biosci.* **9**, 2420–2446.

Artigot, M.P., Loiseau, N., Laffitte, J., Mas-Reguieg, L., Tadrist, S., Oswald, I.P. and Puel, O. (2009) Molecular cloning and functional characterization of two CYP619 cytochrome P450s involved in biosynthesis of patulin in *Aspergillus clavatus*. *Microbiology*, **155**, 1738–1747.

Bertsch, C., Ramírez-Suero, M., Magnin-Robert, M., Larignon, P., Chong, J., Abou-Mansour, E., Spagnolo, A., Clément, C. and Fontaine, F. (2013) Grapevine trunk diseases: complex and still poorly understood. *Plant Pathol.* **62**, 243–265.

Blanco-Ulate, B., Rolshausen, P. and Cantu, D. (2013a) Draft genome sequence of *Neofusicoccum parvum* isolate UCR-NP2, a fungal vascular pathogen associated with grapevine cankers. *Genome Announc.* **1**, e00339–13.

Blanco-Ulate, B., Rolshausen, P. and Cantu, D. (2013b) Draft genome sequence of the ascomycete *Phaeoacremonium aleophilum* strain UCR-PA7, a causal agent of the esca disease complex in grapevines. *Genome Announc.* **1**, e00390–13.

Blanco-Ulate, B., Rolshausen, P.E. and Cantu, D. (2013c) Draft genome sequence of the grapevine dieback fungus *Eutypa lata* UCR-EL1. *Genome Announc.* **1**, e00228–13.

Blanco-Ulate, B., Vincenti, E., Powell, A.L. and Cantu, D. (2013d) Tomato transcriptome and mutant analyses suggest a role for plant stress hormones in the interaction between fruit and *Botrytis cinerea*. *Front. Plant Sci.* **4**, 142.

Blanco-Ulate, B., Amrine, K.C., Collins, T.S., Rivero, R.M., Vicente, A.R., Morales-Cruz, A., Doyle, C.L., Ye, Z., Allen, G., Heymann, H., Ebeler, S.E. and Cantu, D. (2015) Developmental and metabolic plasticity of white-skinned grape berries in response to *Botrytis cinerea* during noble rot. *Plant Physiol.* **169**, 2422–2443.

Bray, N.L., Pimentel, H., Melsted, P. and Pachter, L. (2016) Near-optimal probabilistic RNA-Seq quantification. *Nat. Biotechnol.* **34**, 525–527.

Bridge, P. and Spooner, B. (2001) Soil fungi: diversity and detection. *Plant Soil*, **232**, 147–154.

Bruez, E., Vallance, J., Gerbore, J., Lecomte, P., Da Costa, J.-P., Guerin-Dubrana, L. and Rey, P. (2014) Analyses of the temporal dynamics of fungal communities colonizing the healthy wood tissues of esca leaf-symptomatic and asymptomatic vines. *PloS One*, **9**, e95928.

Camarero, S., Sarkar, S., Ruiz-Dueñas, F.J., Martıńez, M.J., and Martıńez, Á.T. (1999) Description of a versatile peroxidase involved in the natural degradation of lignin that has both manganese peroxidase and lignin peroxidase substrate interaction sites. *J. Biol. Chem.* **274**, 10 324–10 330.

Camps, C., Kappel, C., Lecomte, P., Léon, C., Gomès, E., Coutos-Thévenot, P. and Delrot, S. (2010) A transcriptomic study of grapevine (*Vitis vinifera* cv. Cabernet-Sauvignon) interaction with the vascular ascomycete fungus *Eutypa lata*. *J. Exp. Bot.* **61**, 1719–1737.

Casieri, L., Hofstetter, V., Viret, O. and Gindro, K. (2009) Fungal communities living in the wood of different cultivars of young *Vitis vinifera* plants. *Phytopathol. Mediterr.* **48**, 73–83.

Corte, L., di Cagno, R., Groenewald, M., Roscini, L., Colabella, C., Gobbetti, M. and Cardinali, G. (2015) Phenotypic and molecular diversity of *Meyerozyma guilliermondii* strains isolated from food and other environmental niches, hints for an incipient speciation. *Food Microbiol.* **48**, 206–215.

Crucello, A., Sforça, D.A., Horta, M.A.C., dos Santos, C.A., Viana, A.J.C., Beloti, L.L., de Toledo, M.A., Vincentz, M., Kuroshu, R.M. and de Souza, A.P. (2015) Analysis of genomic regions of *Trichoderma harzianum* IOC-3844 related to biomass degradation. *PloS One*, **10**, e0122122.

Czemmel, S., Galarneau, E.R., Travadon, R., McElrone, A.J., Cramer, G.R. and Baumgartner, K. (2015) Genes expressed in grapevine leaves reveal latent wood infection by the fungal pathogen *Neofusicoccum parvum*. *PloS One*, **10**, e0121828.

Eskalen, A., Feliciano, A. and Gubler, W. (2007) Susceptibility of grapevine pruning wounds and symptom development in response to infection by *Phaeoacremonium aleophilum* and *Phaeomoniella chlamydospora*. *Plant Dis.* **91**, 1100–1104.

Eyal, Z. (1999) The *Septoria tritici* and *Stagonospora nodorum* blotch diseases of wheat. *Eur. J. Plant Pathol.* **105**, 629–641.

Feng, P., Lu, Q., Najafzadeh, M., van den Ende, A.G., Sun, J., Li, R., Xi, L., Vicente, V.A., Lai, W., Lu, C. and De Hoog, G.S. (2014) *Cyphellophora* and its relatives in *Phialophora*: biodiversity and possible role in human infection. *Fungal Divers.* **65**, 17–45.

Ferreira, J., Van Wyk, P. and Calitz, F. (1999) Slow dieback of grapevine in South Africa: stress-related predisposition of young vines for infection by *Phaeoacremonium chlamydosporum*. *S. Afr. J. Enol. Viticult.* **20**, 43–46.

Filippidou, S., Junier, T., Wunderlin, T., Lo, C.-C., Li, P.-E., Chain, P.S. and Junier, P. (2015) Under-detection of endospore-forming firmicutes in metagenomic data. *Comput. Struct. Biotechnol. J.* **13**, 299–306.

Fitt, B.D., Huang, Y.-J., van den Bosch, F. and West, J.S. (2006) Coexistence of related pathogen species on arable crops in space and time. *Phytopathology*, **44**, 163–182.

Floudas, D., Binder, M., Riley, R., Barry, K., Blanchette, R.A., Henrissat, B., Martinez, A.T., Otillar, R., Spatafora, J.W., Yadav, J.S., Aerts, A., Benoit, I., Boyd, A., Carlson, A., Copeland, A., Coutinho, P.M., de Vries, R.P., Ferreira, P., Findley, K., Foster, B., Gaskell, J., Glotzer, D., Gorecki, P., Heitman, J., Hesse, C., Hori, C., Igarashi, K., Jurgens, J.A., Kallen, N., Kersten, P., Kohler, A., Kues, U., Kumar, T.K., Kuo, A., LaButti, K., Larrondo, L.F., Lindquist, E., Ling, A., Lombard, V., Lucas, S., Lundell, T., Martin, R., McLaughlin, D.J., Morgenstern, I., Morin, E., Murat, C., Nagy, L.G., Nolan, M., Ohm, R.A., Patyshakuliyeva, A., Rokas, A., Ruiz-Duenas, F.J., Sabat, G., Salamov, A., Samejima, M., Schmutz, J., Slot, J.C., St John, F., Stenlid, J., Sun, H., Sun, S., Syed, K., Tsang, A., Wiebenga, A., Young, D., Pisabarro, A., Eastwood, D.C., Martin, F., Cullen, D., Grigoriev, I.V. and Hibbett, D.S. (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science*, **336**, 1715–1719.

Fraissinet-Tachet, L., Marmeisse, R., Zinger, L. and Luis, P. (2013) Metatranscriptomics of soil eukaryotic communities, in The Ecological Genomics of Fungi (Martin, F. ed.), Hoboken, NJ: John Wiley & Sons, Inc, doi: 10.1002/9781118735893.ch14.

Gubler, W.D., Mugnai, L. and Surico, G. (2015) Esca, Petri and grapevine leaf stripe diseases. In *Compendium of Grape Diseases, Disorders and Pests*, 2nd edn (Wilcox, W.F., Gubler, W.D. and Uyemoto, J.K., eds), pp. 52–57. St. Paul, MN: APS Press.

Hiscox, J., Savoury, M., Müller, C.T., Lindahl, B.D., Rogers, H.J. and Boddy, L. (2015) Priority effects during fungal community establishment in beech wood. *ISME J.* **9**, 2246–2260.

Hofstetter, V., Buyck, B., Croll, D., Viret, O., Couloux, A. and Gindro, K. (2012) What if esca disease of grapevine were not a fungal disease? *Fungal Divers.* **54**, 51–67.

Hogan, D.A. (2006) Talking to themselves: autoregulation and quorum sensing in fungi. *Eukaryot. Cell*, **5**, 613–619.

Hosni, T., Moretti, C., Devescovi, G., Suarez-Moreno, Z.R., Fatmi, M.B., Guarnaccia, C., Pongor, S., Onofri, A., Buonaurio, R. and Venturi, V. (2011) Sharing of quorum-sensing signals and role of interspecies communities in a bacterial plant disease. *ISME J.* **5**, 1857–1870.

Howlett, B.J., Idnurm, A. and Pedras, M.S.C. (2001) *Leptosphaeria maculans*, the causal agent of blackleg disease of Brassicas. *Fungal Genet. Biol.* **33**, 1–14.

Huang, W., Li, L., Myers, J.R. and Marth, G.T. (2012) ART: a next-generation sequencing read simulator. *Bioinformatics*, **28**, 593–594.

Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyère, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro, C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pè, M.E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.F., Weissenbach, J., Quétier, F., Wincker, P. and French-Italian Public Consortium for Grapevine Genome Characterization (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**, 463–467.

Jones, L., Riaz, S., Morales-Cruz, A., Amrine, K.C., McGuire, B., Gubler, W.D., Walker, M.A. and Cantu, D. (2014) Adaptive genomic structural variation in the grape powdery mildew pathogen, *Erysiphe necator*. *BMC Genomics*, **15**, 1081.

Kumar, J., Schäfer, P., Hückelhoven, R., Langen, G., Baltruschat, H., Stein, E., Nagarajan, S. and Kogel, K.H. (2002) *Bipolaris sorokiniana*, a cereal pathogen of global concern: cytological and molecular approaches towards better control. *Mol. Plant Pathol.* **3**, 185–195.

Kuske, C.R., Hesse, C.N., Challacombe, J.F., Cullen, D., Herr, J.R., Mueller, R.C., Tsang, A. and Vilgalys, R. (2015) Prospects and challenges for fungal metatranscriptomics of complex communities. *Fungal Ecol.* **14**, 133–137.

Lamari, L. and Bernier, C. (1989) Toxin of *Pyrenophora tritici-repentis*: host-specificity, significance in disease, and inheritance of host reaction. *Phytopathology*, **79**, 740–744.

Lamichhane, J.R. and Venturi, V. (2015) Synergisms between microbial pathogens in plant disease complexes: a growing trend. *Front. Plant Sci.* **6**, 385.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.

Li, D., Liu, C.-M., Luo, R., Sadakane, K. and Lam, T.-W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, **31**, 1674–1676.

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

Li, J.Y. and Strobel, G.A. (2001) Jesterone and hydroxy-jesterone antioomycete cyclohexenone epoxides from the endophytic fungus *Pestalotiopsis jesteri*. *Phytochemistry*, **57**, 261–265.

Longcore, J.E., Pessier, A.P. and Nichols, D.K. (1999) *Batrachochytrium dendrobatidis* gen. et sp. nov., a chytrid pathogenic to amphibians. *Mycologia*, **91**, 219–227.

Luque, J., Martos, S., Aroca, A., Raposo, R. and Garcia-Figueres, F. (2009) Symptoms and fungi associated with declining mature grapevine plants in northeast Spain. *J. Plant Pathol.* **91**, 381–390.

Machovič, M. and Janeček, Š. (2006) Starch-binding domains in the post-genome era. *Cell. Mol. Life Sci. CMLS*, **63**, 2710–2724.

Massonnet, M., Morales-Cruz, A., Figueroa-Balderas, R., Lawrence, D.P., Baumgartner, K. and Cantu, D. (2016) Condition-dependent co-regulation of genomic clusters of virulence factors in the grapevine trunk pathogen *Neofusicoccum parvum*. *Mol. Plant Pathol.*

Moktali, V. (2013) *A Comparative Genomic Investigation of Niche Adaptation in Fungi*. PhD Thesis. The Pennsylvania State University, University Park, PA, USA.

Morales-Cruz, A., Amrine, K.C., Blanco-Ulate, B., Lawrence, D.P., Travadon, R., Rolshausen, P.E., Baumgartner, K. and Cantu, D. (2015) 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*, **16**, 469.

Mugnai, L., Graniti, A. and Surico, G. (1999) Esca (black measles) and brown wood-streaking: two old and elusive diseases of grapevines. *Plant Dis.* **83**, 404–418.

Munkvold, G. and Marois, J. (1995) Factors associated with variation in susceptibility of grapevine pruning wounds to infection by *Eutypa lata*. *Phytopathology*, **85**, 249–256.

Nierman, W.C., Pain, A., Anderson, M.J., Wortman, J.R., Kim, H.S., Arroyo, J., Berriman, M., Abe, K., Archer, D.B., Bermejo, C., Bennett, J., Bowyer, P., Chen, D., Collins, M., Coulsen, R., Davies, R., Dyer, P.S., Farman, M., Fedorova, N., Fedorova, N., Feldblyum, T.V., Fischer, R., Fosker, N., Fraser, A., García, J.L., García, M.J., Goble, A., Goldman, G.H., Gomi, K., Griffith-Jones, S., Gwilliam, R., Haas, B., Haas, H., Harris, D., Horiuchi, H., Huang, J., Humphray, S., Jiménez, J., Keller, N., Khouri, H., Kitamoto, K., Kobayashi, T., Konzack, S., Kulkarni, R., Kumagai, T., Lafon, A., Latgé, J.P., Li, W., Lord, A., Lu, C., Majoros, W.H., May, G.S., Miller, B.L., Mohamoud, Y., Molina, M., Monod, M., Mouyna, I., Mulligan, S., Murphy, L., O'Neil, S., Paulsen, I., Peñalva, M.A., Pertea, M., Price, C., Pritchard, B.L., Quail, M.A., Rabbinowitsch, E., Rawlins, N., Rajandream, M.A., Reichard, U., Renauld, H., Robson, G.D., Rodriguez de Córdoba, S., Rodríguez-Peña, J.M., Ronning, C.M., Rutter, S., Salzberg, S.L., Sanchez, M., Sánchez-Ferrero, J.C., Saunders, D., Seeger, K., Squares, R., Squares, S., Takeuchi, M., Tekaia, F., Turner, G., Vazquez de Aldana, C.R., Weidman, J., White, O., Woodward, J., Yu, J.H., Fraser, C., Galagan, J.E., Asai, K., Machida, M., Hall, N., Barrell, B. and Denning, D.W. (2005) Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature*, **438**, 1151–1156.

Nomura, K., Melotto, M. and He, S.-Y. (2005) Suppression of host defense in compatible plant–*Pseudomonas syringae* interactions. *Curr. Opin. Plant Biol.* **8**, 361–368.

Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., Smirnova, T., Grigoriev, I.V. and Dubchak, I. (2014) The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res.* **42**, D26–D31.

Pancher, M., Ceol, M., Corneo, P.E., Longa, C.M.O., Yousaf, S., Pertot, I. and Campisano, A. (2012) Fungal endophytic communities in grapevines (*Vitis vinifera* L.) respond to crop management. *Appl. Environ. Microbiol.* **78**, 4308–4317.

Perkins, J. and Pedersen, W. (1987) Disease development and yield losses associated with northern leaf blight on corn. *Plant Dis.* **71**, 940–943.

Péros, J.-P., Jamaux-Despréaux, I., Berger, G. and Gerba, D. (1999) The potential importance of diversity in *Eutypa lata* and co-colonising fungi in explaining variation in development of grapevine dieback. *Mycol. Res.* **103**, 1385–1390.

Perrin, R.M., DeRocher, A.E., Bar-Peled, M., Zeng, W., Norambuena, L., Orellana, A., Raikhel, N.V. and Keegstra, K. (1999) Xyloglucan fucosyltransferase, an enzyme involved in plant cell wall biosynthesis. *Science*, **284**, 1976–1979.

Pierron, R.J., Pouzoulet, J., Couderc, C., Judic, E., Compant, S. and Jacques, A. (2016) Variations in early response of grapevine wood depending on wound and inoculation combinations with *Phaeoacremonium aleophilum* and *Phaeomoniella chlamydospora*. *Front. Plant Sci.* **7**, 268.

Ponomarova, O. and Patil, K.R. (2015) Metabolic interactions in microbial communities: untangling the Gordian knot. *Curr. Opin. Microbiol.* **27**, 37–44.

Pouzoulet, J., Pivovaroff, A.L., Santiago, L.S. and Rolshausen, P.E. (2014) Can vessel dimension explain tolerance toward fungal vascular wilt diseases in woody plants? Lessons from Dutch elm disease and esca disease in grapevine. *Front. Plant Sci.* **5**, 253.

Rolshausen, P., Greve, L., Labavitch, J., Mahoney, N., Molyneux, R. and Gubler, W. (2008) Pathogenesis of *Eutypa lata* in grapevine: identification of virulence factors and biochemical characterization of cordon dieback. *Phytopathology*, **98**, 222–229.

Rolshausen, P.E., Úrbez-Torres, J.R., Rooney-Latham, S., Eskalen, A., Smith, R.J. and Gubler, W.D. (2010) Evaluation of pruning wound susceptibility and protection against fungi associated with grapevine trunk diseases. *Am. J. Enol. Viticult.* **61**, 113–119.

Rolshausen, P.E., Sosnowski, M., Trouillas, F.P. and Gubler, W.D. (2015) Eutypa dieback. In *Compendium of Grape Diseases, Disorders and Pests*, 2nd edn (Wilcox, W.F., Gubler, W.D. and Uyemoto, J.K., eds), pp. 57–61. St. Paul, MN: APS Press.

Rowe, R., Riedel, R. and Martin, M. (1985) Synergistic interactions between *Verticillium dahliae* and *Pratylenchus penetrans* in potato early dying disease. *Phytopathology*, **75**, 412–418.

Ruiz-Dueñas, F.J., Morales, M., García, E., Miki, Y., Martínez, M.J. and Martínez, A.T. (2009) Substrate oxidation sites in versatile peroxidase and other basidiomycete peroxidases. *J. Exp. Bot.* **60**, 441–452.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W. and Fungal Barcoding Consortium.; Fungal Barcoding Consortium Author List (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proc. Natl. Acad. Sci.* **109**, 6241–6246.

Scholz, M.B., Lo, C.-C. and Chain, P.S. (2012) Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Curr. Opin. Biotechnol.* **23**, 9–15.

Sparapano, L., Bruno, G. and Graniti, A. (2001) Three-year observation of grapevines cross-inoculated with esca-associated fungi. *Phytopathol. Mediterr.* **40**, 376–386.

Stoffel, K., van Leeuwen, H., Kozik, A., Caldwell, D., Ashrafi, H., Cui, X., Tan, X., Hill, T., Reyes-Chin-Wo, S., Truco, M.J., Michelmore, R.W. and Van Deynze, A. (2012) Development and application of a 6.5 million feature Affymetrix Genechip® for massively parallel discovery of single position polymorphisms in lettuce (*Lactuca* spp.). *BMC Genomics*, **13**, 185.

Stoll, D., Boraston, A., Stålbrand, H., McLean, B.W., Kilburn, D.G. and Warren, R.A.J. (2000) Mannanase Man26A from *Cellulomonas fimi* has a mannan-binding module. *FEMS Microbiol. Lett.* **183**, 265–269.

Tang, F. and Saier, M.H. (2014) Transport proteins promoting *Escherichia coli* pathogenesis. *Microb. Pathog.* **71**, 41–55.

Travadon, R., Lecomte, P., Diarra, B., Lawrence, D.P., Renault, D., Ojeda, H., Rey, P. and Baumgartner, K. (2016) Grapevine pruning systems and cultivars influence the diversity of wood-colonizing fungi. *Fungal Ecol.* **24**, 82–93.

Turner, T.R., James, E.K. and Poole, P.S. (2013) The plant microbiome. *Genome Biol.* **14**, 1.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S. and Banfield, J.F. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, **428**, 37–43.

Úrbez-Torres, J., Leavitt, G., Voegel, T. and Gubler, W. (2006) Identification and distribution of *Botryosphaeria* spp. associated with grapevine cankers in California. *Plant Dis.* **90**, 1490–1503.

Vernikos, G., Medini, D., Riley, D.R. and Tettelin, H. (2015) Ten years of pan-genome analyses. *Curr. Opin. Microbiol.* **23**, 148–154.

White, T.J., Bruns, T., Lee, S. and Taylor, J. (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protoc.* **18**, 315–322.

Whitelaw-Weckert, M., Rahman, L., Appleby, L., Hall, A., Clark, A., Waite, H. and Hardie, W.J. (2013) Co-infection by *Botryosphaeriaceae* and *Ilyonectria* spp. fungi during propagation causes decline of young grafted grapevines. *Plant Pathol.* **62**, 1226–1237.

Xiao, G., Ying, S.-H., Zheng, P., Wang, Z.-L., Zhang, S., Xie, X.-Q., Shang, Y., St Leger, R.J., Zhao, G.P., Wang, C. and Feng, M.G. (2012) Genomic perspectives on the evolution of fungal entomopathogenicity in *Beauveria bassiana*. *Sci. Rep.* **2**, 483.

Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D.R., Bork, P. and Patil, K.R. (2015) Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl. Acad. Sci.* **112**, 6449–6454.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website.

Table S1 Genomes and transcriptomes used to generate the multi-species reference.

Table S2 Composition of simulated samples used for the testing and optimization of read mapping.

Table S3 Description of the field samples used in the study including the culture-based diagnosis of the associated trunk diseases.

Table S4 DNA-seq and mapping metrics of the field samples.

Table S5 RNA-seq and mapping metrics of the field samples.

Table S6 Fungal genera detected in each sample based on homology to peptides in the National Center for Biotechnology Information (NCBI) protein RefSeq database (BLASTX; e-value $<$ 1e-6).

Table S7 Normalized read counts of RNA-seq per function detected in each sample.

Fig. S1 Taxonomy of the grapevine trunk pathogens isolated from the field samples.

Fig. S2 Pipeline used for taxonomic assignment of *de novo* assembled RNA-seq contigs.

Fig. S3 Barplot showing the abundance of the genera detected by mapping RNA-seq data onto the *de novo* assembled contigs.

Fig. S4 Description of the partial least squares discriminant analysis (PLS-DA) model using the R package 'ropls'.