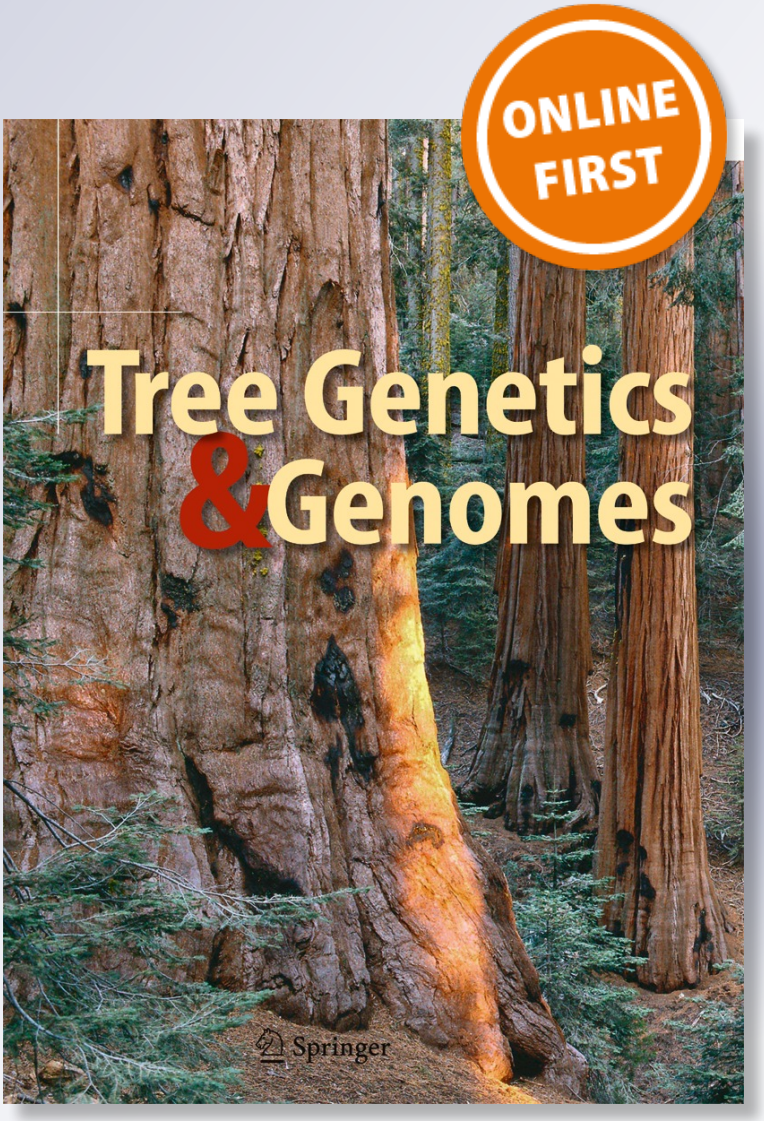
Dual RNA-seq of the plant pathogen Phytophthora ramorum *and its tanoak host*

Katherine J. Hayden, Matteo Garbelotto, Brian J. Knaus, Richard C. Cronn, Hardeep Rai & Jessica W. Wright

Tree Genetics & Genomes

ISSN 1614-2942

Tree Genetics & Genomes
DOI 10.1007/s11295-014-0698-0



Springer

Your article is protected by copyright and all rights are held exclusively by Springer-Verlag Berlin Heidelberg (outside the USA). This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com”.

Springer

Author's personal copy

Tree Genetics & Genomes
DOI 10.1007/s11295-014-0698-0

ORIGINAL PAPER

Dual RNA-seq of the plant pathogen *Phytophthora ramorum* and its tanoak host

Katherine J. Hayden · Matteo Garbelotto · Brian J. Knaus · Richard C. Cronn · Hardeep Rai · Jessica W. Wright

Received: 20 August 2012 / Revised: 28 November 2013 / Accepted: 27 December 2013
© Springer-Verlag Berlin Heidelberg (outside the USA) 2014

Abstract Emergent diseases are an ever-increasing threat to forests and forest ecosystems and necessitate the development of research tools for species that often may have few pre-existing resources. We sequenced the mRNA expressed by the sudden oak death pathogen *Phytophthora ramorum* and its most susceptible forest host, tanoak, within the same tissue at two time points after inoculation, and in uninfected tanoak controls. Using the *P. ramorum* genome to differentiate host and pathogen transcripts, we detected more than 850 *P. ramorum* transcripts at 5 days post-inoculation and a concurrent upregulation of host genes usually associated with pathogenicity. At 1 day, in contrast, we did not detect pathogen expression or significant enrichment of functional categories of host transcripts relative to controls, highlighting the importance of sequencing depth for *in planta* studies of host–pathogen interactions. This study highlights processes in molecular host–pathogen interactions in forest trees and provides a first reference transcriptome for tanoak, allowing the preliminary identification of disease-related genes in this study and facilitating future work for this and other members of the family Fagaceae.

Keywords Hemibiotroph · Host–pathogen interactions · Oomycete · Pathogenesis · Sudden oak death

Communicated by J. Wegrzyn

Electronic supplementary material The online version of this article (doi:10.1007/s11295-014-0698-0) contains supplementary material, which is available to authorized users.

K. J. Hayden (✉) · M. Garbelotto
Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA
e-mail: katherine.hayden@nancy.inra.fr

B. J. Knaus · R. C. Cronn
Pacific Northwest Research Station, USDA-Forest Service, Corvallis, OR, USA

H. Rai
Utah State University, Logan, UT, USA

J. W. Wright
Pacific Southwest Research Station, USDA-Forest Service, Davis, CA, USA

K. J. Hayden
Interactions Arbres-Microorganismes, UMR1136, Université de Lorraine, 54500 Vandoeuvre-lès-Nancy, France

Present Address:
K. J. Hayden
Interactions Arbres-Microorganismes, UMR1136, INRA, 54280 Champenoux, France

Introduction

Emergent diseases are an increasing threat to forests worldwide (Brasier 2008; Burdon et al. 2006; Loo 2009). Mass die-offs of foundation species have cascading ecosystem effects stemming from irreversible changes in community structure (Ellison et al. 2005), including a global impact due to the atmospheric carbon emitted by dying trees (Breshears and Allen 2002; Kurz et al. 2008). Sudden oak death (SOD, causal agent *Phytophthora ramorum*) is one disease with such far-reaching impacts. Tanoak (*Notholithocarpus densiflorus*, formerly *Lithocarpus densiflorus*), a monotypic genus in the family Fagaceae (Manos et al. 2009), is the only species in US forests that has been found to be susceptible to both disease forms: sudden oak death, a fatal canker known for the destruction it has caused in US forests but which does not support sporulation, and ramorum blight, a leaf and twig blight which has minimal fitness effects on its primary hosts, yet is responsible for the pathogen's spread (Davidson et al. 2003, 2008). Tanoak is endemic to California and coastal

Published online: 26 February 2014

Author's personal copy

Oregon and holds approximately 68 Tg of carbon across its range, with highest densities in areas that are at high risk for sudden oak death (Lamsal et al. 2011). As the primary masting tree in many coastal evergreen forests, tanoak is truly a foundational species, providing food for humans and wildlife (Bowcutt 2011; Meyers et al. 2006), and being one of the primary ectomycorrhizal hosts in redwood and Douglas-fir forest systems (Bergemann and Garbelotto 2006). Tanoaks are highly susceptible to SOD, and the decline or extirpation of large size classes is predicted in areas that favor disease (Cobb et al. 2012; Ramage et al. 2011; Waring and O'Hara 2008). However, individual trees have been reported to survive 10 years or longer in heavily infested sites (McPherson et al. 2010), and there are indications that resistance may play a role in tanoak disease dynamics (Hayden et al. 2011, 2013).

The understanding of plant host–pathogen relationships at a genomic scale has been largely limited to herbaceous hosts and particularly to model systems. Nonetheless, a long history of research in forest species has done much to advance the understanding of the molecular genetics of disease resistance in trees. This has included the discovery of resistance genes (reviewed by Ersoz et al. 2010), and genomic studies of chestnut (*Castanea* spp.) including those focused on disease resistance (Barakat et al. 2009; Wheeler and Sederoff 2009). Genome-wide association studies have lent insights to adaptive traits in *Pinus* species (Eckert et al. 2009; Lepoittevin et al. 2011), but due to the stature, long generation times, and large genomes of forest trees, the genomic study of non-commercial species—for which populations of clonal replicates or defined phenotypic lines are unavailable—has been especially limited (Parchman et al. 2010; Rampant et al. 2011).

The advent of high-throughput sequencing coupled with advances in bioinformatic technologies has paved the way for genomic studies in systems that were previously intractable to research (Ekblom and Galindo 2011; Neale and Kremer 2011). One such approach is RNA-seq, the sequencing of cDNA libraries from individuals experiencing various physiological challenges, with no prior knowledge of the genes' identity or sequence (Mortazavi et al. 2008). Typically, methods employing RNA-seq have relied on reference genomes to assemble and quantify the millions of small sequence fragments produced by high-throughput methods, but bioinformatic advances now allow the de novo assembly of an entire transcriptome (Grabherr et al. 2011; Robertson et al. 2010).

The complete genome of *P. ramorum* has been sequenced and annotated (Tyler et al. 2006), and this has added much to the understanding of oomycete evolution and pathogenicity (e.g., Jiang et al. 2008; Kasuga et al. 2012; Seidl et al. 2011; Win et al. 2012). As we report here, the genome sequence also enables bioinformatic tools to separate pathogen transcripts from the host when both have been concurrently sequenced in infected plant tissue. Subtraction of human-genomic sequences from pathogens has already been used to separate and even diagnose microbial sequences from infected tissue (Isakov et al. 2011; Weber et al. 2002; Xu et al. 2003). These methods have recently been applied to tomatoes (Li et al. 2012), for which a reference genome is also available. Simultaneous transcriptional profiling of a plant host–pathogen interaction in systems with both the host and pathogen references available has been accomplished, first using SuperSAGE (serial analysis of gene expression) (Matsumura et al. 2003) and more recently using RNA-seq (Kawahara et al. 2012; Petre et al. 2012). Expression analysis after the generation of de novo reference transcriptomes for both host and pathogen has been reported in a turfgrass pathosystem (Orshinsky et al. 2012). The work we report here combines approaches, using reference-based mapping to subtract pathogen sequences, followed by de novo assembly of a the transcriptome of the non-model forest-tree host.

The precise mode of infection of tanoaks by *P. ramorum* in nature has not been characterized, but it has been observed that disease in large trees follows a pattern of aerial spread of sporangia in rain splash or rare wind events, and frequently begins as leaf and twig infections in the crown, with trunk cankers following (Hansen et al. 2008). Tanoak mortality likely results from girdling by cambial cankers (Rizzo et al. 2002), but mycelia and extensive tylosis formation have also been observed in the trees' xylem, suggesting that vascular obstruction also plays an important role in disease etiology (Parke et al. 2007; Stamm and Parke 2013). In laboratory studies of *P. ramorum*, infection most often occurs within 48 h from the release of swimming zoospores from the asexual sporangia (Moralejo and Descals 2011). *Phytophthora* spp. are hemibiotrophic, living in necrotic tissue following the first stages of infection. Effectors, large families of secreted proteins (Hardham and Cahill 2010; Kamoun 2006), are thought to have a dual role in hemibiotrophs, initially helping to evade plant defenses, but in later stages inducing programmed plant cell death and subsequent necrosis (Jiang et al. 2006; Kleemann et al. 2012). The chemical application of members of the elicitin family of effectors has been shown to independently cause photosynthetic declines in tanoak (Manter et al. 2007).

We undertook a first transcriptomic study of tanoak infected with *P. ramorum*, with the aim to provide a reference transcriptome for tanoak. We sequenced tissue from infected and non-inoculated control leaves at two times post-inoculation, in order to provide guidance for future experimental methods while making preliminary identification of disease-related genes in this pathosystem.

Author's personal copy

Materials and methods

Experimental treatments

Pathogen isolate

Within North American forests, *P. ramorum* populations appear to be comprised of relatively few genotypes, which all were clonally derived from a single founder individual (Croucher et al. 2013; Mascheretti et al. 2008) within the North American 1 (NA1) lineage (Ivors et al. 2006). NA1 isolate Pr52 (CBS110537; ATCC MYA-2436, taken from the Garbelotto Lab culture collection and originally obtained from D. Rizzo, University of California, Davis) has been frequently used in laboratory inoculations (Anacker et al. 2008; Garbelotto et al. 2009; Hayden et al. 2011; Tooley et al. 2009; Tooley et al. 2004). Comparative studies have reported this strain to be highly pathogenic on leaf and canker hosts (Hüberli and Garbelotto 2011).

To induce sporangia formation, 1×1 cm pieces V8 agar (Erwin and Ribeiro 1996) colonized by the Pr52 were flooded with sterile 1 % soil extract and incubated at 18 °C for 2 days, followed by a 20-min cold shock at 4 °C to induce zoospore release. To prevent premature encystment by zoospores, all plastic and glassware to come in contact with zoospores were soaked in 10 % hydrochloric acid and rinsed in deionized water prior to use.

All pathogen propagation and plant inoculations were completed under permit and according to conditions set by the United States Department of Agriculture Animal and Plant Health Inspection Service, and the California Department of Food and Agriculture.

Tanoak cuttings

Because tanoak is a non-commercial species, there are no established stocks of clonal propagates or breeding lines. A 4-year-old potted *N. densiflorus* sapling grown in a lath house in Berkeley, CA from wild-collected seed was selected for this first study, based on its moderate susceptibility to detached leaf inoculation assays (Lundquist and Hayden, unpublished data). Twelve cut branches, approximately 20–30 cm long with at least three leaf nodes, were moved from a lath house to randomized positions in a growth chamber with a 12 h, 20 °C day, and 13 °C night, in individual sterilized 500 mL bottles of deionized water, and were misted with tap water in 15-min intervals twice daily.

After equilibrating for 24 h in the growth chamber, a 1-cm length of petiole and midrib on each of three leaves per cutting was wounded by lightly scraping with a sterile scalpel. Immediately thereafter, 100 μL of zoospore suspension averaging 10^5 spores per milliliter (inoculation treatments) or sterile 1 % soil extract (controls) was dropped on the leaf axil, which was surrounded by a cup made from Parafilm (SPI Supplies) to prevent runoff of the suspension.

Leaves were harvested two times after inoculation. Plants' transcriptional responses to inoculation with pathogens have been documented as early as 1 h post-inoculation (Norelli et al. 2009; Orłowska et al. 2011), but disease progression is generally at first quite slow in tanoaks, and visible lesions usually do not develop earlier than 3–5 days post-inoculation (personal observation, K. Hayden). The accumulation of *P. ramorum* DNA in inoculated *Umbellularia californica* leaf tissue has been shown to increase after 48 h post-inoculation, corresponding with the appearance of lesions (Hayden et al. 2006). At 24 h (1 day) and 120 h (5 days) after inoculation, a 3 cm by 1 cm strip of tissue, including both petiole/midrib and leaf blade, was excised and flash-frozen from each leaf on three cuttings per inoculation treatment and time period. There were thus four sample units: 1 day inoculated (I1), 1 day control (C1), 5 days inoculated (I5), and 5 days control (C5) (Fig. 1). At 1 day no lesions were visible, and at 5 days faint, continuous lesions were visible in inoculated leaves.

RNA isolation, library preparation, and high-throughput sequencing

Prior to extraction, all of the excised, frozen tissue corresponding to a treatment group was pooled into a single, 75–100-mg sample. Bulk RNA was extracted using a chloroform–isoamyl alcohol extraction and lithium chloride precipitation (Chang et al. 2007; Le Provost et al. 2007). The terminal DNAse treatment and column cleanup steps described by Le Provost et al. were omitted to save time and costs because the Illumina RNA-seq protocol includes purification of poly-adenylated RNA, removing DNA. The purity and concentration of RNA extracts were characterized with the Bioanalyzer 2100 platform (Agilent Technologies, Santa Clara, CA).

Samples were diluted to contain a total of 5 μg ribonucleic acids in 50 μL of nuclease-free water. Poly-adenylated RNA was purified and enriched, and unstranded libraries were prepared as described by Illumina (San Diego, CA) GAII protocols (part no. 1004898 Rev. A), except that Agencourt AmPure XP kits (Beckman Coulter, Brea, CA) were used for PCR purification steps.

Single-end, non-indexed 80-bp fragment sequencing was performed at the Oregon State University Center for Genome Research and Biocomputing (CGRB), using Illumina GAII equipment and software pipeline version 1.2. Each flowcell included a PhiX control lane.

There was a single sample and flowcell lane for each treatment group. Groups I1, I5, and C1 were sequenced in the same flowcell. Sequencing failed in the C5 lane in the

Author's personal copy

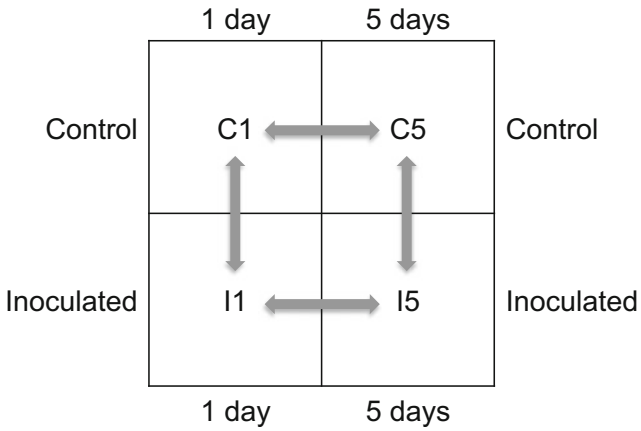


Fig 1 Experimental scheme. Pairwise comparisons for differential expression analysis are indicated by *gray arrows*; *horizontal arrows* represent comparisons within treatment, and *vertical arrows* represent those within time

shared flowcell, so C5 data were generated on a different flowcell, in the same facility and with the same protocols.

Sequence data were initially processed using the Illumina pipeline, followed by filtering for primer dimers using custom scripts (brianknaus.com). The reads were further filtered for cumulative quality scores, with an error probability limit P=0.05, using CLC Genomics software (www.clcbio.com), and for a minimum 40-bp fragment size after quality filtering.

Differentiation of host and pathogen sequences

Filtered reads were mapped to the *P. ramorum* genome (Tyler et al. 2006) using two degrees of stringency: first to the unfiltered genome model ramorum1 using CLC Genomics software with permissive mapping parameters of 50 % length fraction and 80 % minimum similarity, in order to conservatively remove pathogen sequence, and secondly, to characterize the observed *P. ramorum* transcription, sequences were mapped to the genome and transcriptome together using the combined programs Bowtie2 and TopHat2 (Langmead et al. 2009; Trapnell et al. 2009), with the call: tophat2 –solexa-quals –library-type fr-unstranded -i 30 -I 5000 -G [path to Phyra1_1.gff] [path to ramorum1 indices] [data file in fastq format].

The option –i 30 sets the minimum intron length to 30 bp, and -I sets the maximum distance between splice junctions to 5,000. These lengths are smaller than defaults for mammalian systems, in accordance with the relative paucity and smaller size of oomycete introns (Kamoun 2003; Shen et al. 2011). After mapping, isoforms were finalized and compared to the reference transcripts using the program Cufflinks (Trapnell et al. 2010, 2012).

P. ramorum functional annotation

Mappings to the *P. ramorum* Phyra1.1 model were first annotated using the reference transcript model annotation (Tyler et al. 2006). Secondly, we performed a local blastx search of translated nucleotides against custom databases of *Phytophthora* species RXLR avirulence homologs (Jiang et al. 2008), cellulose-binding and elicitin-like proteins (Sierra et al. 2010), and crinkler and necrosis-inducing (CRN) proteins (Haas et al. 2009; Tyler et al. 2006), as well as against the PHI-base pathogen–host interaction database release 3.2 (Winnenburg et al. 2006). All searches were implemented with a minimum e-value cutoff of 10^{-10}.

Assembly and validation of de novo tanoak transcriptome

The reads that were not mapped to the *P. ramorum* genome model ramorum1 by the least stringent mapping procedure were pooled together to create a de novo transcriptome. The pooled reads were assembled using the program Trinity (v20110519, Grabherr et al. 2011), which uses three software modules to refine an initial transcriptome assembly. All three steps were performed on the CGRB compute cluster, with the third "Butterfly" step using a highly parallel step via a custom script and the Sun Grid Engine queuing system; contigs with fewer than 300 bp were excluded from the data.

To confirm that *P. ramorum* sequences had been fully removed from the dataset and to identify sequences corresponding to endophytes or other microbial associates, the tanoak transcriptome was queried against the NCBI comprehensive non-redundant (nr) database using a translated-nucleotide-to-translated-nucleotide (tblastx) search (Altschul 1990). An additional blastn search was performed against the mitochondrial (mt) databases (www.ncbi.nlm.nih.gov). In both cases, an expected value of less than 10^{-10} was used to call matches.

Tanoak mapping, normalization, and statistical testing

Quantification of sequenced mRNA poses a special problem: alternative splicing can cause reads from even unique gene regions to map equally well to multiple transcripts (reviewed by Pachter 2011). For this reason, we first mapped each lane individually to the reference using the short-read aligner Bowtie (Langmead et al. 2009) v0.12.7 and converted the output in Sequence Alignment/Map (SAM) format to the binary BAM format via SAMtools (Li et al. 2009). We then used eXpress (v1.0.0; Roberts and Pachter 2013) to resolve multiple and other ambiguous mappings, using parameters as recommended. This program uses an expectation-maximization algorithm to assign mappings probabilistically,

Author's personal copy

but unlike other programs with similar functionality, eXpress does not require an annotated genome.

Analysis of RNA-seq expression data is confounded by differences in sequence counts among lanes, as well as by biases for sequence length and GC-content (Bullard et al. 2010; Hansen et al. 2010; Oshlack and Wakefield 2009). We therefore performed three normalizations on the sets of transcript counts. First, initial transcript-length and GC-content normalization was performed in eXpress, generating effective counts rather than total or unique counts. Second, full-quantile normalization for transcript-level GC-content and length biases was performed within lanes using the R package EDASeq (Risso et al. 2011). Finally, full-quantile normalization in EDASeq was again used to correct for among-lane differences in GC-content and library size.

Differences in normalized counts among treatment groups were tested in 4 pairwise comparisons (Fig. 1). Within time, we compaired Inoculated vs Control at 1 day (I1 vs C1); and Inoculated vs Control at 5 days (I5 vs C5). Within treatment, comparisons were 5 days vs 1 day, Inoculated (I5 vs I1); and 5 days vs 1 day, Control (C5 vs C1).

Seven thousand seven hundred thirty-two transcripts with an experiment-wide mean effective count less than ten across lanes were assumed to be non-informative and were excluded prior to pair-wise comparisons, leaving 40,656 transcripts for analysis.

Pairwise differences in normalized counts were tested using the negative binomial test for experiments in the R package DESeq (v1.8.3; Anders and Huber 2010), using pooled dispersion estimates with only the fitted values shared across genes as recommended by the authors for samples without replication. The significance threshold was set at FDR$<$0.05 after correction for multiple comparisons (Benjamini and Hochberg 1995).

Tanoak functional annotation

Similar to those mapping to *P. ramorum*, tanoak transcripts were annotated by a Blast2GO search of the NCBI nr database, with a minimum e-value of 1×10^{-10}, followed by searches against custom databases. The Plant Resistance Genes database (Sanseverino et al. 2010), which consists of resistance genes derived from reference genes, from NCBI Unigene and from predictions based on domain functional prediction, and the PHI-base interaction database (Winnenburg et al. 2006) were used to annotate resistance genes and other pathogen–host interaction genes within the tanoak transcripts. Finally, the transcriptome sequences were searched against a set of sequences related to disease response in *Castanea mollissima* and *Castanea dentata* infected with the chestnut blight fungus, *Cryphonectria parasitica* (Barakat et al. 2009).

We tested for over-representation of protein categories within each pairwise comparison of time and treatment (Fig. 1) using Fisher's exact test within the Blast2GO platform (Conesa et al. 2005). Sequences with greatest similarity to taxa other than plants were excluded from the data before testing.

Results

Concurrent sequencing of *P. ramorum* and tanoak transcriptomes

We used short-read high-throughput sequencing of tanoak cuttings, with and without inoculation by the sudden oak death pathogen *P. ramorum*, to (1) provide the first concurrent *in planta* transcriptomes of *P. ramorum* and a forest host and (2) to get a first portrait of the gene sets involved in the interaction. A total of 52 Gbp of sequence data were produced, ranging from 331 to 2,186 Mbp per sequence lane (Table 1).

Differentiation of host and pathogen sequences

P. ramorum was convincingly detected in the I5 sample by both the most permissive mapping method (CLC Genomics for removal of putative pathogen sequences), with 4.5 % mapped by the least stringent criteria, and the most stringent (Bowtie/Tophat/Cufflinks for *P. ramorum* transcript characterization, with 1.1 % mapped) (Table 2). The percentages of reads mapped to *P. ramorum* were not distinguishable among I1, C1, and C5 groups and ranged from 0.1 to 1.4 %, depending on the mapping method (Table 2). The GC-content of mapped reads was highest in lane I5, and at 0.58 was equivalent to the average GC-content of transcribed sequences described in *P. ramorum* (Jiang and Govers 2006) and in other *Phytophthora* species (Kamoun 2003).

Sequential mapping of the I5 sample to the *P. ramorum* genome and transcriptome model references identified 858 unique transcripts, from 729 unique reference transcripts (Supplementary Table 1). The majority of these, 522, either completely matched or were contained within an annotated intron chain. Two hundred twenty-five overlapped annotated exons and an additional 21 shared at least one splice junction with the reference. This last group may represent improvements to current gene models or may be novel isoforms.

P. ramorum functional annotation

Only sample I5 returned more than negligible blast hits to pathogenicity-related databases. In the control and 1-day samples, none of the sequences that were mapped to the *P. ramorum* genome were matched to genes in the *Phytopthora*-specific pathogenicity databases and no more

Author's personal copy

Table 1 Summary statistics for RNA-seq of tanoak and *P. ramorum* (I), and mock-inoculated tanoak controls (C), including mean read length after trimming for quality

Identifier	Treatment	Time post-inoculation	Raw reads	Mean read length (bp)
I1	Inoculated	1 day	4,420,578	75
I5	Inoculated	5 days	10,353,182	70
C1	Control	1 day	20,238,150	75
C5	Control	5 day	28,324,515	77

than three were matched to the more general PHI-base pathogen proteins. The very few transcripts mapped to *P. ramorum* in sample I1, however, were significantly longer than those mapped from the control lanes (P=0.007, $F_{2,22}$=20.6), while there was no significant difference in transcript length within treatment groups. In contrast, in *P. ramorum*-mapped transcripts in sample I5, there were matches to at least 30 CBEL-type proteins (Tyler et al. 2006; Sierra et al. 2010) and 6 Avr-homologs (Jiang et al. 2008) and to a single CRN (Tyler et al. 2006; Haas et al. 2009) (Supplementary Table 2).

Assembly and validation of de novo tanoak transcriptome

The Trinity assembly of all reads that did not map to the *P. ramorum* reference produced a de novo transcriptome consisting of 48,388 unique sequences in 34,642 isogroups, with $N50$=1,289. The vast majority (more than 29,000) of the 31,894 that were matched to sequences in the NCBI nr database with taxonomic information were best matched to a sequence within the plants, Viridiplantae (Fig. 2). The next-most represented group was fungi (primarily Ascomycota). The third most abundant taxonomic category was Arthropoda, of which 607 out of 644 transcripts matched aphid sequences. These aphid sequences were likely due to the presence of aphids in the growth chamber. A total of six transcripts matched mitochondrial sequences. Very few (three transcripts) included Stramenopiles among the best matches, and only a single transcript returned a best hit to a *Phytophthora*: a *Phytophthora* sojae hypothetical protein. The dearth of matches to close relatives of *P. ramorum* suggests that differences in gene expression among inoculated and non-inoculated samples are not due to presence of perhaps novel *P. ramorum* transcripts, which failed to map to the reference, but rather a response of the plant itself. It is possible that there may be systematic differences in the presence or magnitude of expression of transcripts originating from microbial or insect associates in inoculated tissue relative to controls. However, if such systematic differences exist, most are likely to be erased with the exclusion of transcripts which were best matched to taxa outside plants.

Tanoak mapping, normalization, and statistical testing

Sequenced library sizes were dramatically different among treatments/lanes, both in raw reads and in transcript-level counts (Table 1, Supplementary Fig. 1). Post-normalization transcript-level count and GC-content distributions were comparable among lanes (Supplementary Figs. 1 and 2).

The largest differences in expression were observed in inoculated samples at 5 days post-inoculation compared to 1 day post-inoculation (within treatment: I5 vs. I1, Fig. 3), in which 1,040 transcripts were significantly differentially expressed. This finding contrasts with 169 differentially expressed transcripts in the control (samples C5 vs. C1). Within sampling time, there were 245 transcripts significantly differentially expressed between inoculated and control at 1 day post-inoculation and 271 at 5 days post-inoculation (Fig. 3). Results from unreplicated comparisons should be viewed as preliminary identification of differentially expressed sequences (Bullard et al. 2010; Pachter 2011) to be used in guiding further study, but in need of confirmation through replicated experiments.

Tanoak functional annotation

From the de novo tanoak transcriptome, consisting of the 48,388 unique sequences assembled from all reads that were

Table 2 Summary statistics for mapping to the *P. ramorum* reference via different protocols, and to the de novo tanoak transcriptome

Identifier	*P. ramorum* genome[a] (lax mapping)	*P. ramorum* genome and transcript model[a] (stringent mapping)			De novo tanoak transcriptome
	Mapped (%)	Mapped (%)	Transcripts	GC-content	Mapped reads
I1	1.2	0.1	8	0.51	0.5×10^7
I5	4.5	1.1	858	0.58	1.4×10^7
C1	1.4	0.1	12	0.51	2.5×10^7
C5	0.9	0.1	13	0.51	3.5×10^7

[a] Tyler et al. 2006

Author's personal copy

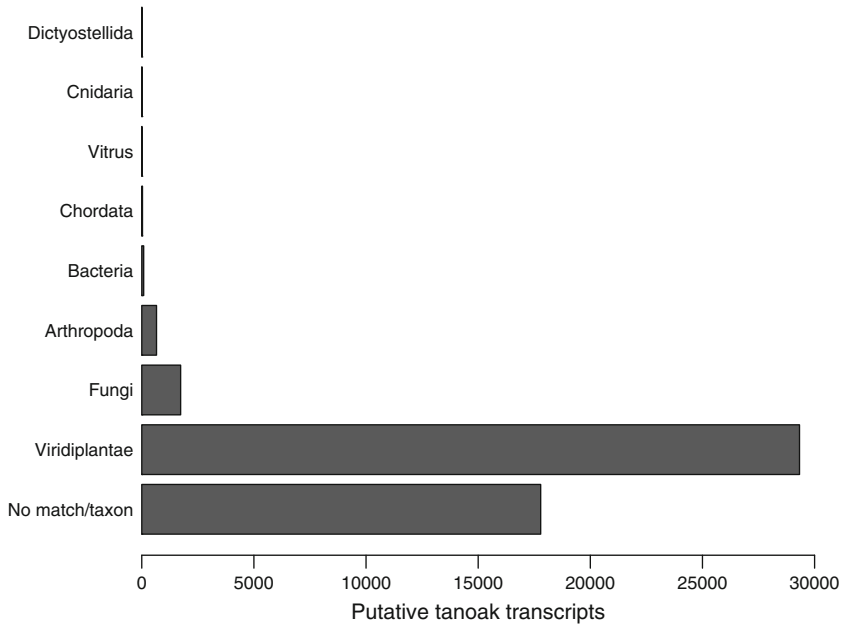


Fig. 2 Taxonomic characterization of the 48,388 contigs in the de novo transcriptome by matching to the NCBI nr database; broad taxonomic group name is followed by the number of transcripts with a top match to this group. Seven taxa with fewer than ten matching transcripts are not labeled; these include stramenopiles, to which three transcripts matched

not mapped to *P. ramorum*, 34,378 were matched to the NCBI nr database. Of these matches, 24,351 were assigned protein functional annotations. Relatively few matches to LRR-domain and other curated plant resistance genes (Sanseverino et al. 2010) were observed, 1,059 within the entire tanoak assembly, and not more than 6 were represented in any differentially expressed subset. There was more similarity to *C. mollissima* and *C. dentata* disease-related transcripts, with some 7,051 sequence matches overall.

Enrichment tests of Gene Ontology (Ashburner et al. 2000) terms among sequences differentially expressed in pairwise comparisons (Fig. 1) lend preliminary insight to changes in

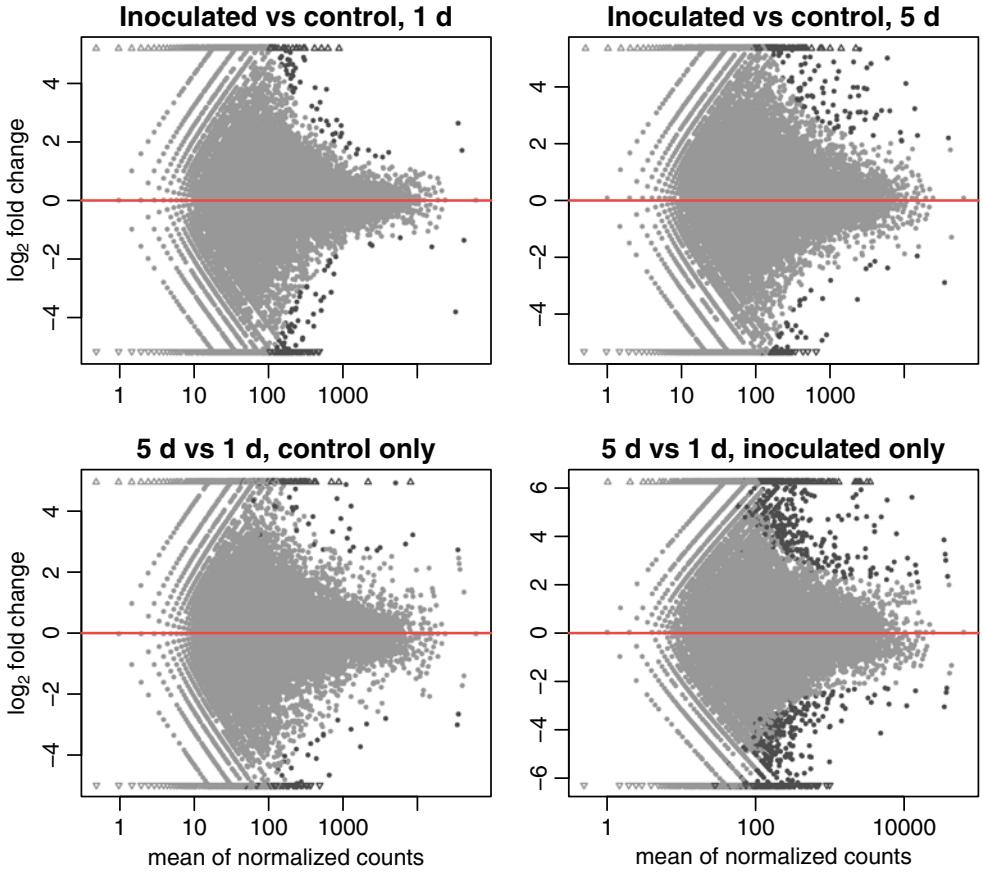


Fig. 3 Pairwise comparisons of expression of tanoak transcripts. Transcripts with statistically significant differential expression at an FDR-adjusted $P<0.05$ are represented with *dark gray dots*. Transcripts with zero expression in one condition are represented by *open triangles*

Author's personal copy

branch physiology over time (Table 3, C5 vs C1, I5 vs I1) and inoculation status (I5 vs C5, I1 vs C1). The need for verification with further replicates is highlighted by low sequence yield in sample I1 (Table 1, Supplementary Fig 1). Housekeeping functions were included among enriched terms in comparisons to I1 (Table 3), rendering these comparisons useful only as supporting results of the others and as a starting point for investigations of this pathosystem. Regardless of inoculation status and consistent with senescence, there was a decrease in functions, processes, and cellular components related to photosynthesis and thylakoid membranes, and an increase in ion binding and myrcene and terpene synthase functions in the cut branches over time. At 5 days, there were significant increases in protein functions associated with disease and defense response in inoculated branches relative to control (Table 3). Specifically, processes and functions involving the binding, metabolism, and catabolism of chitins, aminoglycans, and polysaccharides were increased relative to both the control and to sample I1. There was also marginally significant enrichment of monooxygenase and oxidoreductase activity in inoculated tissue relative to control at 5 days.

Discussion

We provide the first transcriptome for tanoak, a species that is unique and valuable in its own right, and has relationships to economically important Fagaceae, and identify sequences that may be important in the tanoak-*P. ramorum* pathosystem. Emergences of newly discovered (Brasier et al. 2005; Donahoo et al. 2006; Duran et al. 2008; Hansen et al. 2003; Reeser et al. 2007; Scott et al. 2009) and newly hybridized (Brasier et al. 1999) *Phytophthora* spp. in forests have accelerated in recent years. The techniques we present may be applied to these systems for genomic study of new hosts and pathogens simultaneously—a similar technique could be applied to use homology to *P. ramorum*, *P. sojae*, and *Phytophthora infestans* to sort host and pathogen sequences even for *Phytophthora* spp. for which genomes have not been sequenced.

We assembled the first transcriptome for tanoak; this will now be available for use for future research on this species, or use for comparative genomics in *Quercus* and *Castanea*, or other hardwoods. The assembly of more than 48,000 contigs is reasonably in line with the 69,000 contigs in an assembly of multiple *Quercus robur* EST libraries (Ueno et al. 2010), and the 28,000 and 40,000 assembled from 454 sequences from the *C. dentata* and *C. mollissima* transcriptomes (Barakat et al. 2009). Although there is no doubt that more sequencing at greater depths and of different tissues will reveal new sequences and isoforms, this transcriptome provides a base upon which further research may be built.

P. ramorum transcripts were directly sequenced and subtracted from host transcripts in inoculated leaves, 5 days post-inoculation. For concurrent sequencing of host and pathogen together, the 5 days time proved to be better, as after only 1 day, the amount of mapped *P. ramorum* sequence was not different than that of uninoculated controls. Some amount of mapping is expected due to chance alone even when no *P. ramorum* is present, due to highly conserved regions or to similarity with transcripts of endophytes or of other microbial associates that are inevitably sequenced along with plant tissues.

It was not clear what true proportion of transcripts in infected tanoak tissue should be expected to have been of pathogen origin. Approximately 5 % of the DNA in leaf lesions on California bay laurel (*U. californica*) leaves has been reported to be from *P. ramorum* (Hayden et al. 2006). The 4.5 % of transcripts in I5 mapped to the *P. ramorum* with permissive mapping parameters approached this figure, but may have included some false matches. More stringent mapping found considerably less pathogen, but this result likely included false negatives. The true proportional representation of pathogen transcripts in tanoak leaves 5 days post-inoculation was likely somewhat less than that observed in genomic DNA in bay laurel. Differences in pathogen growth rates in the host substrates, relative differences in transcriptome and genome sizes, and the relative difficulty of extracting of *P. ramorum* nucleic acids from tanoak (Hayden et al. 2006) might all contribute to the detection of fewer pathogen transcripts. Nonetheless, the GC-content of sequences mapped to the *P. ramorum* genome in I5 was identical to that reported in the literature for *Phytophthora* coding sequences (Jiang and Govers 2006; Kamoun 2003), further evidence that these sequences were of pathogen origin. The de novo transcriptome of tanoak was assembled from reads that were unmapped under permissive settings, in order to err on the side of removing *P. ramorum* from the dataset.

The pathogenicity-related transcripts attributed to *P. ramorum* included elicitins, lectins, pectinesterases, and cellulose-binding proteins, all of which are associated with necrosis or breakdown of host cells, as well as 6 Avr-homologs, which are associated both with infection and with the switch from biotrophy to necrotrophy (Hein et al. 2009; Lee and Rose 2010). Dual RNA-seq of host and pathogen by nature is challenged by the very small percentage of pathogen mRNA within the total RNA in infected plant tissue (Westermann et al. 2012). The number of transcripts mapped to *P. ramorum* was slightly more than 5 % of the 15,743 gene models in the *P. ramorum* reference; deeper sequencing by spreading sequencing across flow cell lanes and/or altering inoculation and RNA extraction methods to increase representation of pathogen mRNA will surely identify more pathogen transcripts.

Author's personal copy

Table 3 Gene ontology annotations that were significantly enriched (Fisher's exact test, within-comparison FDR<0.10) in sets of differentially expressed of transcripts

	GO term	ID	%Ref	I5 vs C5	I5 vs I1	C5 vs C1	I1 vs C1
MF	(R)-limonene synthase activity	GO:0034002	0.00	–	0.7 (6.38E−02)	–	–
	Carbon-oxygen lyase activity	GO:0016835	0.74	–	3.5 (1.74E−02)	13.3 (1.67E−03)	–
	Carbon-oxygen lyase activity, acting on phosphates	GO:0016838	0.13	3.8 (9.83E−02)	2.8 (4.30E−06)	8.9 (1.67E−03)	–
	Carbon-oxygen lyase activity, acting on polysaccharides	GO:0016837	0.03	–	–	4.4 (5.33E−02)	-
	Cation binding	GO:0043169	14.32	–	–	35.6 (7.84E−02)	–
	Chitin binding	GO:0008061	0.07	3.8 (3.02E−02)	1.8 (6.78E−04)	–	–
	Chitinase activity	GO:0004568	0.10	6.4 (7.90E−05)	2.5 (1.59E−05)	–	–
	Electron carrier activity	GO:0009055	1.66	10.3 (3.95E−02)	–	*−18.5 (3.71E−02)*	–
	Endopeptidase inhibitor activity	GO:0004866	0.03	–	1.8 (9.71E−05)	–	–
	Endopeptidase regulator activity	GO:0061135	0.03	–	2.1 (4.65E−06)	–	–
	Enzyme inhibitor activity	GO:0004857	0.19	–	2.1 (8.70E−03)	–	–
	Heme binding	GO:0020037	1.36	–	–	11.1 (8.71E−02)	–
	Hydrolase activity, hydrolyzing O-glycosyl compounds	GO:0004553	1.70	–	4.9 (7.49E−02)	–	–
	Ion binding	GO:0043167	14.33	–	–	35.6 (7.84E−02)	–
	Lyase activity	GO:0016829	2.04	–	6.7 (2.41E−03)	17.8 (3.09E−03)	–
	Magnesium ion binding	GO:0000287	0.72	–	3.5 (1.29E−02)	8.9 (7.84E−02)	–
	Monooxygenase activity	GO:0004497	1.20	9.0 (3.95E−02)	–	–	–
	Myrcene synthase activity	GO:0050551	0.05	–	2.1 (9.35E−06)	6.7 (3.09E−03)	–
	NADH dehydrogenase (quinone) activity	GO:0050136	0.35	–	–	8.9 (1.32E−02)	11.5 (5.51E−02)
	NADH dehydrogenase (ubiquinone) activity	GO:0008137	0.35	–	–	8.9 (1.32E−02)	11.5 (5.51E−02)
	NADH dehydrogenase activity	GO:0003954	0.39	–	–	8.9 (1.86E−02)	11.5 (7.09E−02)
	Oxidoreductase activity	GO:0016491	10.89	–	–	33.3 (2.61E−02)	42.3 (2.72E−02)
	Oxidoreductase activity, acting on NADH or NADPH	GO:0016651	0.77	–	–	8.9 (9.70E−02)	–
	Oxidoreductase activity, acting on NADH or NADPH, quinone or similar compound as acceptor	GO:0016655	0.47	–	–	8.9 (3.00E−02)	–
	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	GO:0016705	1.86	10.3 (7.32E−02)	–	–	–
	Pattern binding	GO:0001871	0.16	–	1.8 (2.98E−02)	–	–
	Pectate lyase activity	GO:0030570	0.03	–	–	4.4 (5.33E−02)	–
	Peptidase inhibitor activity	GO:0030414	0.05	–	1.8 (2.74E−04)	–	–
	Peptidase regulator activity	GO:0061134	0.05	–	2.1 (1.33E−05)	–	–
	Polysaccharide binding	GO:0030247	0.16	–	1.8 (2.98E−02)	–	–
	Secologanin synthase activity	GO:0050616	0.05	–	–	4.4 (7.36E−02)	–
	Serine-type endopeptidase inhibitor activity	GO:0004867	0.02	–	1.1 (1.39E−02)	–	–
	Structural constituent of cuticle	GO:0042302	0.03	–	1.1 (4.63E−02)	–	–
	Structural constituent of ribosome	GO:0003735	2.38	–	12.7 (2.69E−12)	–	–
	Structural molecule activity	GO:0005198	2.98	–	14.5 (1.69E−12)	–	–
	Terpene synthase activity	GO:0010333	0.09	–	2.1 (1.23E−04)	6.7 (9.44E−03)	–
	Tetrapyrrole binding	GO:0046906	1.42	9.0 (7.84E−02)	–	11.1 (9.86E−02)	–
BP	Aminoglycan catabolic process	GO:0006026	0.09	6.4 (6.61E−05)	2.5 (8.64E−06)	–	–
	Aminoglycan metabolic process	GO:0006022	0.14	6.4 (2.24E−04)	2.5 (1.12E−04)	–	–
	ATP synthesis coupled electron transport	GO:0042773	0.23	–	–	6.7 (5.84E−02)	–
	Carbohydrate catabolic process	GO:0016052	0.98	–	3.5 (9.24E−02)	–	–
	Cell killing	GO:0001906	0.02	2.6 (8.30E−02)	–	–	–

Author's personal copy

Table 3 (continued)

	GO term	ID	%Ref	I5 vs C5	I5 vs I1	C5 vs C1	I1 vs C1
BP	Cell wall macromolecule catabolic process	GO:0016998	0.16	–	1.8 (3.23E−02)	–	–
	Cell wall organization or biogenesis	GO:0071554	1.13	–	3.9 (7.55E−02)	–	–
	Cell–cell signaling	GO:0007267	0.08	–	1.4 (3.23E−02)	–	–
	Cellular biosynthetic process	GO:0044249	17.68	–	26.1 (5.07E−02)	–	–
	Cellular macromolecule biosynthetic process	GO:0034645	11.89	–	19.1 (6.00E−02)	–	–
	Cellular respiration	GO:0045333	0.68	–	–	13.3 (1.67E−03)	19.2 (2.48E−03)
	Chitin catabolic process	GO:0006032	0.09	6.4 (6.61E−05)	2.5 (8.64E−06)	–	–
	Chitin metabolic process	GO:0006030	0.12	6.4 (1.37E−04)	2.5 (4.61E−05)	–	–
	Electron transport chain	GO:0022900	0.74	–	–	13.3 (1.67E−03)	19.2 (2.48E−03)
	Energy derivation by oxidation of organic compounds	GO:0015980	0.73	–	–	13.3 (1.67E−03)	19.2 (2.48E−03)
	Gene expression	GO:0010467	11.58	–	19.4 (2.24E−02)	–	–
	Generation of precursor metabolites and energy	GO:0006091	1.86	–	–	*−18.5 (5.56E−02)*	19.2 (5.51E−02)
	Killing of cells of other organism	GO:0031640	0.01	2.6 (7.21E−02)	–	–	–
	Lateral inhibition	GO:0046331	0.00	–	0.7 (6.38E−02)	–	–
	Macromolecule biosynthetic process	GO:0009059	11.92	–	19.1 (6.21E−02)	–	–
	Nervous system development	GO:0007399	0.03	–	1.1 (4.63E−02)	–	–
	Neurogenesis	GO:0022008	0.02	–	1.1 (1.99E−02)	–	–
	Oxidation-reduction process	GO:0055114	9.65	–	–	31.1 (2.71E−02)	42.3 (1.37E−02)
	Oxidative phosphorylation	GO:0006119	0.23	–	–	6.7 (5.84E−02)	–
	Photosynthesis	GO:0015979	0.83	–	–	*−22.2 (6.41E−05)*	*−13.6 (1.36E−02)*
	Photosynthesis, light reaction	GO:0019684	0.47	–	–	*−14.8 (5.49E−03)*	–
	Polysaccharide catabolic process	GO:0000272	0.25	6.4 (2.51E−03)	2.5 (3.10E−03)	–	–
	Regulation of endopeptidase activity	GO:0052548	0.00	–	0.7 (6.38E−02)	–	–
	Regulation of peptidase activity	GO:0052547	0.02	–	1.1 (1.99E−02)	–	–
	Respiratory electron transport chain	GO:0022904	0.29	–	–	8.9 (8.94E−03)	15.4 (2.48E−03)
	Translation	GO:0006412	4.17	–	16.3 (1.37E−11)	–	–
	Translational elongation	GO:0006414	0.43	–	2.5 (5.60E−02)	–	–
CC	Chloroplast	GO:0009507	9.56	–	–	*−37.0 (5.28E−02)*	–
	Chloroplast part	GO:0044434	4.67	–	–	*−29.6 (1.39E−02)*	–
	Chloroplast thylakoid	GO:0009534	1.63	–	–	*−29.6 (7.90E−06)*	*−13.6 (6.38E−02)*
	Chloroplast thylakoid membrane	GO:0009535	1.25	–	–	*−29.6 (1.58E−06)*	*−13.6 (3.35E−02)*
	Eukaryotic translation elongation factor 1 complex	GO:0005853	0.05	–	1.1 (7.67E−02)	–	–
	Extracellular region	GO:0005576	1.14	–	6.7 (1.65E−06)	–	–
	Intracellular non-membrane-bounded organelle	GO:0043232	5.68	–	17.7 (1.78E−09)	–	–
	Light-harvesting complex	GO:0030076	0.03	–	–	*−7.4 (2.17E−02)*	–
	Lipid particle	GO:0005811	0.05	–	1.1 (9.24E−02)	–	–
	Macromolecular complex	GO:0032991	10.43	–	21.6 (2.00E−05)	–	–
	Mitochondrial envelope	GO:0005740	1.32	–	–	11.1 (7.84E−02)	19.2 (1.56E−02)
	Mitochondrial inner membrane	GO:0005743	1.03	–	–	11.1 (4.26E−02)	19.2 (8.09E−03)
	Mitochondrial membrane	GO:0031966	1.23	–	–	11.1 (7.10E−02)	19.2 (1.39E−02)
	Mitochondrial part	GO:0044429	1.61	–	–	–	19.2 (3.32E−02)
	Non-membrane-bounded organelle	GO:0043228	5.68	–	17.7 (1.78E−09)	–	–
	Organelle inner membrane	GO:0019866	1.27	–	–	11.1 (7.61E−02)	19.2 (1.45E−02)
	Organelle subcompartment	GO:0031984	1.66	–	–	*−29.6 (7.90E−06)*	*−13.6 (6.38E−02)*
	Photosynthetic membrane	GO:0034357	1.46	–	–	*−33.3 (4.05E−07)*	*−13.6 (5.25E−02)*
	Photosystem	GO:0009521	0.23	–	–	–	*−9.1 (1.40E−02)*

Author's personal copy

Table 3 (continued)

	GO term	ID	%Ref	I5 vs C5	I5 vs I1	C5 vs C1	I1 vs C1
CC	Plastid part	GO:0044435	4.79	–	–	*–29.6 (1.53E–02)*	–
	Plastid thylakoid	GO:0031976	1.63	–	–	*−29.6 (7.90E−06)*	*−13.6 (6.38E−02)*
	Plastid thylakoid membrane	GO:0055035	1.25	–	–	*–29.6 (1.58E–06)*	*−13.6 (3.35E−02)*
	Proton-transporting two-sector ATPase complex	GO:0016469	0.43	–	2.5 (5.77E−02)	–	–
	Respiratory chain	GO:0070469	0.44	–	–	11.1 (2.17E−03)	15.4 (7.82E−03)
	Ribonucleoprotein complex	GO:0030529	3.47	–	14.1 (2.05E−10)	–	–
	Ribosome	GO:0005840	2.75	–	13.8 (1.76E−12)	–	–
	Thylakoid	GO:0009579	2.02	–	–	*−33.3 (2.43E−06)*	–
	Thylakoid membrane	GO:0042651	1.35	–	–	*−33.3 (4.05E−07)*	*−13.6 (4.08E−02)*
	Thylakoid part	GO:0044436	1.62	–	–	*−33.3 (6.86E−07)*	*−13.6 (6.38E−02)*

The percentage of sequences with the given annotation is given for the entire de novo reference and for each differentially expressed set, with FDR in parentheses. Enrichment tests were performed separately for sets with increased or decreased expression; negative entries, in italics, indicate pairs where expression was lower in the first term

MF molecular function, *BP* biological processes, *CC* cellular component

P. ramorum transcripts matching cellulose-binding, elecitin, and lectin-like proteins were detected in greater numbers, with searches revealing 30 transcripts matching genes with those annotations in *P. ramorum* (Tyler et al. 2006) and *P. infestans* databases (Sierra et al. 2010). However, only one matched to CRN (crinkler and necrosis-inducing) genes, matching a pseudogene originally identified in *P. infestans* (Haas et al. 2009).

Tanoak disease response transcripts were better characterized than *Phytophthora* transcripts. More than 1,000 R-gene homologs were discovered within the transcriptome, presenting a starting point to examine tanoak–*P. ramorum* molecular interactions in greater detail. The 160 tanoak transcripts that were differentially expressed after inoculation and were matched to chestnut genes associated with disease response to a fungal canker (Barakat et al. 2009) likewise point to regions for further study. The pathogenesis-related families of chitinases, glycosyl hydrolases, and thaumatin proteins (Van Loon and Van Strien 1999) were especially well represented. Expression of these gene families has been shown to increase over the course of infection of grape vine in susceptible interactions with powdery mildew, but not in resistant interactions (Fung et al. 2008). Nonetheless, in early infection stages they are more highly expressed in incompatible interactions of rice with rice blast fungus than compatible interactions (Kawahara et al. 2012). Among the matches to chestnut sequences were two dirigent-like disease resistance genes with increased expression after inoculation. Dirigent-like genes are associated with suberization and the synthesis of lignin and phytoalexins (Davin and Lewis 2000). Xylem obstructions caused by tyloses, suberized bubbles of parenchymal tissue, have been implicated in the loss of hydrolic conductivity and rapid wilting response that can accompany *P. ramorum* infection in tanoaks (Collins et al. 2009; Parke et al. 2007) and have been shown to increase in proportion to the relative production of pathogen elicitins (Manter et al. 2007). Rapid tylosis production may present a barrier to pathogen movement through vessels and, like other defensive responses, is also associated with pathogenicity. In either case, these genes may present useful targets of study as biomarkers for infection, disease response, or disease tolerance.

Preliminary analysis of differential expression over time after inoculation identified sequences involved in tanoak disease response; these sequences can help target research of disease interactions in forest trees. Functions for polysaccharide and aminoglycan binding and metabolis/catabolis were enriched in tanoak sequences with increased expression 5 days post-inoculation, as were monooxygenase and electron carrying functions, consistent with signal transduction and cell killing activity. These stand in comparison to the reduced expression in terms related to photosynthesis in non-inoculated branches over time, perhaps demonstrating that the observed disease response was in spite of overall stress and reduction in activity in cut branches. Despite the problems inherent with comparisons of single technical replicates—and in particular, the confounding of sequencing depth with treatment effects—there was demonstrable enrichment of plant expression of proteins associated with necrosis and disease response in susceptible hosts 5 days post-inoculation, at the same time that *P. ramorum* was expressing a variety of cellulose-binding and lectin-like proteins and a few avirulence homologs. Together, they portray an interaction of a susceptible host and hemibiotrophic pathogen during its necrotrophic phase.

With sufficient pathogen growth and sequencing depth, it is possible to identify interacting pathogenicity and resistance

Author's personal copy

processes for pathosystems with a distinct paucity of host genomic resources—and the same could be accomplished for interactions of a well-annotated host with a new pathogen. As there is no expectation that disease introductions will decelerate in coming years, this type of research will become increasingly important.

Acknowledgments We are grateful for funding provided by the Western Forest Transcriptome Survey of the USDA-Forest Service: Pacific Northwest Research Station, Pacific Southwest Research Station, and Rocky Mountain Research Stations; the Gordon and Betty Moore Foundation; and the National Science Foundation Ecology of Infectious Diseases Program. We thank Abdelali Barakat, Kenan Celtik, Tara Jennings, Alex Lundquist, Barb Rotz, and Chris Sullivan for generous assistance.

Data archiving statement This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GAOS00000000. The version described in this paper is the first version, GAOS01000000.

References

Altschul S (1990) Basic local alignment search tool. J Mol Biol 215:403–410. doi:10.1006/jmbi.1990.9999

Anacker B, Rank N, Hüberli D et al (2008) Susceptibility to *Phytophthora ramorum* in a key infectious host: landscape variation in host genotype, host phenotype, and environmental factors. New Phytol 177:756–766. doi:10.1111/j.1469-8137.2007.02297.x

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106. doi:10.1186/gb-2010-11-10-r106

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene Ontology: tool for the unification of biology. Nat Genet 25:25–29. doi:10.1038/75556

Barakat A, DiLoreto D, Zhang Y et al (2009) Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection. Bmc Plant Biol 9:51. doi:10.1186/1471-2229-9-51

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300

Bergemann SE, Garbelotto M (2006) High diversity of fungi recovered from the roots of mature tanoak (*Lithocarpus densiflorus*) in northern California. Can J Bot 84:1380–1394. doi:10.1139/b06-097

Bowcutt F (2011) Tanoak target: the rise and fall of herbicide use on a common native tree. Environ Hist 16:197–225. doi:10.1093/envhis/emr032

Brasier C (2008) The biosecurity threat to the UK and global environment from international trade in plants. Plant Pathol 57:792–808. doi:10.1111/j.1365-3059.2008.01886.x

Brasier CM, Cooke DEL, Duncan JM (1999) Origin of a new *Phytophthora* pathogen through interspecific hybridization. Proc Natl Acad Sci U S A 96:5878–5883. doi:10.1073/pnas.96.10.5878

Brasier C, Beales PA, Kirk SA et al (2005) *Phytophthora kernoviae* sp. nov., an invasive pathogen causing bleeding stem lesions on forest trees and foliar necrosis of ornamentals in the UK. Mycological Res 109:853–859. doi:10.1017/S0953756205003357

Breshears DD, Allen CD (2002) The importance of rapid, disturbance-induced losses in carbon management and sequestration. Global Ecol Biogeography 11:1–5. doi:10.1046/j.1466-822X.2002.00274.x

Bullard J, Purdom E, Hansen K, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 11:94. doi:10.1186/1471-2105-11-94

Burdon J, Thrall P, Ericson L (2006) The current and future dynamics of disease in plant communities. Annu Rev Phytopathol 44:19–39. doi:10.1146/annurev.phyto.43.040204.140238

Chang S, Puryear J, Cairney J (2007) A simple and efficient method for isolating RNA from pine trees. Plant Mol Biol Rep 11:113–116

Cobb RC, Filipe JAN, Meentemeyer RK et al (2012) Ecosystem transformation by emerging infectious disease: loss of large tanoak from California forests. J Ecology 100(3):712–722. doi:10.1111/j.1365-2745.2012.01960.x

Collins B, Parke J, Lachenbruch B, Hansen E (2009) The effects of *Phytophthora ramorum* infection on hydraulic conductivity and tylosis formation in tanoak sapwood. Can J Forest Res 39:1766–1776. doi:10.1139/X09-097

Conesa A, Gotz S, Garcia-Gomez JM et al (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676. doi:10.1093/bioinformatics/bti610

Croucher PJP, Mascheretti S, Garbelotto M (2013) Combining field epidemiological information and genetic data to comprehensively reconstruct the invasion history and the microevolution of the sudden oak death agent *Phytophthora ramorum* (Stramenopila: Oomycetes) in California. Biol Invasions 15:2281–2297. doi:10.1007/s10530-013-0453-8

Davidson J, Werres S, Garbelotto M, et al. (2003) Sudden Oak Death and associated diseases caused by *Phytophthora ramorum*. PHP. doi:10.1094/PHP-2003-0707-01-DG

Davidson J, Patterson H, Rizzo D (2008) Sources of inoculum for *Phytophthora ramorum* in a redwood forest. Phytopathology 98:860–866. doi:10.1094/PHYTO-98-8-0860

Davin LB, Lewis NG (2000) Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis. Plant Physiol 123:453–462. doi:10.1104/pp.123.2.453

Donahoo R, Blomquist CL, Thomas SL et al (2006) *Phytophthora foliorum* sp nov., a new species causing leaf blight of azalea. Mycol Res 110:1309–1322. doi:10.1016/j.mycres.2006.07.017

Duran A, Gryzenhout M, Slippers B et al (2008) *Phytophthora pinifolia* sp nov associated with a serious needle disease of *Pinus radiata* in Chile. Plant Pathol 57:715–727. doi:10.1111/j.1365-3059.2008.01893.x

Eckert A, Bower A, Wegrzyn J et al (2009) Asssociation genetics of coastal Douglas-fir (*Pseudotsuga menziesu* var. *menziesii*, Pinaceae). I. cold-hardiness related traits. Genetics 182:1289–1302. doi:10.1534/genetics.108.102350

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107:1–15. doi:10.1038/hdy.2010.152

Ellison A, Bank M, Clinton B et al (2005) Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. Front Ecol Environ 3:479–486

Ersoz ES, Wright MH, González-Martínez SC et al (2010) Evolution of disease response genes in Loblolly Pine: insights from candidate genes. PLoS ONE 5:e14234. doi:10.1371/journal.pone.0014234

Erwin DC, Ribeiro OK (1996) *Phytophthora* diseases worldwide. APS, St. Paul

Fung RWM, Gonzalo M, Fekete C, Kovacs LG, He Y, Marsh E, McIntyre LM, Schachtman DP, Qiu W (2008) Powdery mildew induces defense-oriented reprogramming of the transcriptome in a susceptible but not in a resistant grapevine. Plant Physiol 146:236–249. doi:10.1104/pp. 107.108712

Garbelotto M, Harnik T, Schmidt D (2009) Efficacy of phosphonic acid, metalaxyl-M and copper hydroxide against *Phytophthora ramorum*

Author's personal copy

in vitro and *in planta*. Plant Pathol 58:111–119. doi:10.1111/j.1365-3059.2008.01894.x

Grabherr MG, Haas BJ, Yassour M et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotech 29:644–652. doi:10.1038/nbt.1883

Haas BJ, Kamoun S, Zody MC et al (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature 461:393–398. doi:10.1038/nature08358

Hansen E, Reeser P, Davidson J et al (2003) *Phytophthora nemorosa*, a new species causing cankers and leaf blight of forest trees in California and Oregon, USA. Mycotaxon 88:129–138

Hansen E, Kanaskie A, Prospero S et al (2008) Epidemiology of *Phytophthora ramorum* in Oregon tanoak forests. Can J Forest Res 38:1133–1143. doi:10.1139/X07-217

Hansen K, Brenner S, Dudoit S (2010) Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res 38:1–7. doi:10.1093/nar/gkq224

Hardham AR, Cahill DM (2010) The role of oomycete effectors in plant–pathogen interactions. Funct Plant Biol 37:919–925. doi:10.1071/FP10073

Hayden K, Ivors K, Wilkinson C, Garbelotto M (2006) TaqMan chemistry for *Phytophthora ramorum* detection and quantification, with a comparison of diagnostic methods. Phytopathology 96:846–854. doi:10.1094/PHYTO-96-0846

Hayden KJ, Nettel A, Dodd RS, Garbelotto M (2011) Will all the trees fall? Variable resistance to an introduced forest disease in a highly susceptible host. For Ecol Manage 261:1781–1791. doi:10.1016/j.foreco.2011.01.042

Hayden KJ, Garbelotto M, Dodd R, Wright JW (2013) Scaling up from greenhouse resistance to fitness in the field for a host of an emerging forest disease. Evol App. doi:10.1111/eva.12080

Hein I, Gilroy E, Armstrong M, Birch P (2009) The zig-zag-zig in oomycete-plant interactions. Mol Plant Pathol 10:547–562

Hüberli D, Garbelotto M (2011) *Phytophthora ramorum* is a generalist plant pathogen with differences in virulence between isolates from infectious and dead-end hosts. Forest Pathol 42:8–13. doi:10.1111/j.1439-0329.2011.00715.x

Isakov O, Modai S, Shomron N (2011) Pathogen detection using short-RNA deep sequencing subtraction and assembly. Bioinformatics 27:2027–2030. doi:10.1093/bioinformatics/btr349

Ivors K, Garbelotto M, Vries I et al (2006) Microsatellite markers identify three lineages of *Phytophthora ramorum* in US nurseries, yet single lineages in US forest and European nursery populations. Molecular Ecology 15:1493–1505. doi:10.1111/j.1365-294X.2006.02864.x

Jiang R, Govers F (2006) Nonneutral GC3 and retroelement codon mimicry in *Phytophthora*. J Mol Evol 63:458–472. doi:10.1007/s00239-005-0211-3

Jiang RHY, Tyler BM, Whisson SC et al (2006) Ancient origin of elicitin gene clusters in *Phytophthora* genomes. Mol Biol Evol 23:338–351. doi:10.1093/molbev/msj039

Jiang R, Tripathy S, Govers F, Tyler B (2008) RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with more than 700 members. Proc Natl Acad Sci U S A 105:4874–4879. doi:10.1073/pnas.0709303105

Kamoun S (2003) Molecular genetics of pathogenic Oomycetes. Eukaryotic Cell 2:191–199. doi:10.1128/EC.2.2.191-199.2003

Kamoun S (2006) A catalogue of the effector secretome of plant pathogenic Oomycetes. Annu Rev Phytopathol 44:41–60. doi:10.1146/annurev.phyto.44.070505.143436

Kasuga T, Kozanitas M, Bui M et al (2012) Phenotypic diversification is associated with host-induced transposon derepression in the sudden oak death pathogen *Phytophthora ramorum*. PLoS ONE 7:e34728. doi:10.1371/journal.pone.0034728

Kawahara Y, Oono Y, Kanamori H et al (2012) Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. PLoS ONE 7:e49423. doi:10.1371/journal.pone.0049423

Kleemann J, Rincon-Rivera LJ, Takahara H et al (2012) Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Collectrotrichum higginsianum*. PLoS Pathog 8:e1002643. doi:10.1371/journal.ppat.1002643

Kurz WA, Stinson G, Rampley GJ et al (2008) Risk of natural disturbances makes future contribution of Canada's forests to the global carbon cycle highly uncertain. Proc Natl Acad Sci U S A 105:1551–1555. doi:10.1073/pnas.0708133105

Lamsal S, Cobb RC, Hall Cushman J et al (2011) Spatial estimation of the density and carbon content of host populations for *Phytophthora ramorum* in California and Oregon. For Ecol Manage 262:989–998. doi:10.1016/j.foreco.2011.05.033

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. doi:10.1186/gb-2009-10-3-r25

Lee S-J, Rose JK (2010) Mediation of the transition from biotrophy to necrotrophy in hemibiotrophic plant pathogens by secreted effector proteins. Plant Signal Behav 5:769–772

Le Provost G, Herrera R, Paiva J et al (2007) A micromethod for high throughput RNA extraction in forest trees. Biol Res 40:291–297

Lepoittevin C, Harvengt L, Plomion C, Garnier-Géré P (2011) Association mapping for growth, straightness and wood chemistry traits in the *Pinus pinaster* Aquitaine breeding population. Tree Genet Genomes 8:113–126. doi:10.1007/s11295-011-0426-y

Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. doi:10.1093/bioinformatics/btp352

Li R, Gao S, Hernandez AG et al (2012) Deep sequencing of small RNAs in tomato for virus and viroid identification and strain differentiation. PLoS One. doi:10.1371/journal.pone.0037127

Loo J (2009) Ecological impacts of non-indigenous invasive fungi as forest pathogens. Biol Invasions 11:81–96. doi:10.1007/s10530-008-9321-3

Manos P, Cannon C, Oh S (2009) Phylogenetic relationships and taxonomic status of the paleoendemic fagaceae of Western North America: recognition of a new genus, *Notholithocarpus*. Madrono 55:181–190

Manter DK, Kelsey RG, Karchesy JJ (2007) Photosynthetic declines in *Phytophthora ramorum*-infected plants develop prior to water stress and in response to exogenous application of elicitins. Phytopathology 97:850–856. doi:10.1094/PHYTO-97-7-0850

Mascheretti S, Croucher P, Vettraino A et al (2008) Reconstruction of the Sudden Oak Death epidemic in California through microsatellite analysis of the pathogen *Phytophthora ramorum*. Mol Ecol 17:2755–2768. doi:10.1111/j.1365-294X.2008.03773.x

Matsumura H, Reich S, Ito A et al (2003) Gene expression analysis of plant host–pathogen interactions by SuperSAGE. Proc Natl Acad Sci U S A 100:15718–15723. doi:10.1073/pnas.2536670100

McPherson B, Mori S, Wood D et al (2010) Responses of oaks and tanoaks to the sudden oak death pathogen after 8y of monitoring in two coastal California forests. Ecol Manage 259:2248–2255. doi:10.1016/j.foreco.2010.02.020

Meyers KJ, Swiecki TJ, Mitchell AE (2006) Understanding the native Californian diet: identification of condensed and hydrolyzable tannins in tanoak acorns (*Lithocarpus densiflorus*). J Agric Food Chem 54:7686–7691. doi:10.1021/jf061264t

Moralejo E, Descals E (2011) Diplanetism and microcyclic sporulation in *Phytophthora ramorum*. Forest Pathol 41:349–354. doi:10.1111/j.1439-0329.2010.00674.x

Mortazavi A, Williams B, McCue K et al (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621–628. doi:10.1038/nmeth.1226

Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. Nat Rev Genet 12:111–122. doi:10.1038/nrg2931

Norelli J, Farrell R, Bassett C et al (2009) Rapid transcriptional response of apple to fire blight disease revealed by cDNA suppression

Author's personal copy

subtractive hybridization analysis. Tree Genet Genomes 5:27–40. doi:10.1007/s11295-008-0164-y

Orłowska E, Fiil A, Kirk H-G et al (2011) Differential gene induction in resistant and susceptible potato cultivars at early stages of infection by *Phytophthora infestans*. Plant Cell Rep 31:187–203. doi:10.1007/s00299-011-1155-2

Orshinsky AM, Hu J, Opiyo SO et al (2012) RNA-seq analysis of the *Sclerotinia homoeocarpa*—creeping bentgrass pathosystem. PLoS ONE 7:e41150. doi:10.1371/journal.pone.0041150

Oshlack A, Wakefield MJ (2009) Transcript length bias in RNA-seq data confounds systems biology. Biol Direct 4:14. doi:10.1186/1745-6150-4-14

Pachter L (2011) Models for transcript quantification from RNA-Seq. arXiv.org arXiv:1104.3889v2 [q-bio.GN].

Parchman TL, Geist KS, Grahnen JA et al (2010) Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. BMC Genomics 11:180. doi:10.1186/1471-2164-11-180

Parke J, Oh E, Voelker S et al (2007) *Phytophthora ramorum* colonizes tanoak xylem and is associated with reduced stem water transport. Phytopathology 97:1558–1567. doi:10.1094/PHYTO-97-12-1558

Petre B, Morin E, Tisserant E et al (2012) RNA-seq of early-infected poplar leaves by the rust pathogen *Melampsora larici-populina* uncovers PtSultr3;5, a fungal-induced host sulfate transporter. PLoS ONE 7:e44408. doi:10.1371/journal.pone.0044408

Ramage BS, O'Hara KL, Forrestel AB (2011) Forest transformation resulting from an exotic pathogen: regeneration and tanoak mortality in coast redwood stands affected by sudden oak death. Can J Forest Res 41:763–772. doi:10.1139/x11-020

Rampant PF, Lesur I, Boussardon C et al (2011) Analysis of BAC end sequences in oak, a keystone forest tree species, providing insight into the composition of its genome. BMC Genomics 12:292. doi:10.1186/1471-2164-12-292

Reeser PW, Hansen EM, Sutton W (2007) *Phytophthora siskiyouensis*, a new species from soil, water, myrtlewood (*Umbellularia californica*) and tanoak (*Lithocarpus densiflorus*) in southwestern Oregon. Mycologia 99:639–643. doi:10.3852/mycologia.99.5.639

Risso D, Schwartz K, Sherlock G, Dudoit S (2011) GC-Content normalization for RNA-seq data. BMC Bioinformatics 12:480. doi:10.1186/1471-2105-12-480

Rizzo D, Garbelotto M, Davidson J et al (2002) *Phytophthora ramorum* as the cause of extensive mortality of *Quercus* spp. and *Lithocarpus densiflorus* in California. Plant Dis 86:205–214

Roberts A, Pachter L (2013) Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Meth 10:71–73. doi:10.1038/nmeth.2251

Robertson G, Schein J, Chiu R et al (2010) De novo assembly and analysis of RNA-seq data. Nat Methods 7:909–912. doi:10.1038/nmeth.1517

Sanseverino W, Roma G, De Simone M et al (2010) PRGdb: a bioinformatics platform for plant resistance gene analysis. Nucleic Acids Res 38:D814–D821. doi:10.1093/nar/gkp978

Scott PM, Burgess TI, Barber PA et al (2009) *Phytophthora multivora* sp nov., a new species recovered from declining Eucalyptus, Banksia, Agonis and other plant species in Western Australia. Persoonia 22: 1–13. doi:10.3767/003158509X415450

Seidl MF, Van den Ackerveken G, Govers F, Snel B (2011) A domain-centric analysis of Oomycete plant pathogen genomes reveals unique protein organization. Plant Physiol 155:628–644. doi:10.1104/pp. 110.167841

Shen D, Ye W, Dong S, Wang Y, Dou D (2011) Characterization of intronic structures and alternative splicing in *Phytophthora sojae* by comparative analysis of expressed sequence tags and genomic sequences. Can J Microbiol 57:84–90. doi:10.1139/W10-103

Sierra R, Rodriguez-R LM, Chaves D et al (2010) Discovery of *Phytophthora infestans* genes expressed in planta through mining of cDNA libraries. PLOS ONE 5(3):e9847. doi:10.1371/journal.pone.0009847

Stamm E, Parke J (2013) The effect of *Phytophthora ramorum* on the physiology and xylem function of young tanoak trees. Proceedings of the Fifth Sudden Oak Death Science Symposium, June 19–22, 2012, Petaluma, CA, USA GTR PSWXXX

Tooley P, Kyde K, Englander L (2004) Susceptibility of selected ericaceous ornamental host species to *Phytophthora ramorum*. Plant Dis 88:993–999

Tooley P, Browning M, Kyde K, Berner D (2009) Effect of temperature and moisture period on infection of Rhododendron "Cunningham's White" by *Phytophthora ramorum*. Phytopathology 99:1045–1052. doi:10.1094/PHYTO-99-9-1045

Trapnell C, Pachter L, Salzberg S (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111. doi:10.1093/bioinformatics/btp120

Trapnell C, Williams BA, Pertea G et al (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28: 511–515. doi:10.1038/nbt.1621

Trapnell C, Roberts A, Goff L et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protocols 7:562–578. doi:10.1038/nprot.2012.016

Tyler B, Tripathy S, Zhang X et al (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313:1261–1266. doi:10.1126/science.1128796

Ueno S, Provost GL, Léger V, et al (2010) Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. BMC Genomics 11:650. doi:10.1186/1471-2164-11-650

Van Loon LC, Van Strien EA (1999) The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. Physiol Mol Plant Pathol 55:85–97. doi:10.1006/pmpp.1999.0213

Waring K, O'Hara K (2008) Redwood/tanoak stand development and response to tanoak mortality caused by *Phytophthora ramorum*. Ecol Manage 255:2650–2658. doi:10.1016/j.foreco.2008.01.025

Weber G, Shendure J, Tanenbaum DM et al (2002) Identification of foreign gene sequences by transcript filtering against the human genome. Nat Genet 30:141–142. doi:10.1038/ng818

Westermann AJ, Gorski SA, Vogel J (2012) Dual RNA-seq of pathogen and host. Nat Rev Micro 10:618–630. doi:10.1038/nrmicro2852

Wheeler N, Sederoff R (2009) Role of genomics in the potential restoration of the American chestnut. Tree Genet Genomes 5:181–187. doi:10.1007/s11295-008-0180-y

Win J, Krasileva KV, Kamoun S et al (2012) Sequence divergent RXLR effectors share a structural fold conserved across plant pathogenic Oomycete species. PLoS Pathog 8:e1002400. doi:10.1371/journal.ppat.1002400

Winnenburg R, Baldwin TK, Urban M et al (2006) PHI-base: a new database for pathogen host interactions. Nucleic Acids Res 34: D459–D464. doi:10.1093/nar/gkj047

Xu Y, Stange-Thomann N, Weber G et al (2003) Pathogen discovery from human tissue by sequence-based computational subtraction. Genomics 81:329–335. doi:10.1016/S0888-7543(02)00043-5