

DISSERTATION

TRANSCRIPTOME AND SECRETOME OF TWO *PYTHIUM* SPECIES  
DURING INFECTION AND SAPROPHYTIC GROWTH

Submitted by

Jorge Rafael Ibarra Caballero

Department of Bioagricultural Sciences and Pest Management

In partial fulfillment of the requirements  
for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2016

Doctoral Committee:

Advisor: Ned A. Tisserat

Jan Leach  
Ann Hess  
Courtney Jahn

Copyright by Jorge Rafael Ibarra Caballero 2016

All Rights Reserved

## ABSTRACT

### TRANSCRIPTOME AND SECRETOME OF TWO *PYTHIUM* SPECIES DURING INFECTION AND SAPROPHYTIC GROWTH

In the first part of this dissertation, I describe how we obtained and analyzed the full complement of transcripts –**the transcriptome**- and the set of secreted proteins –**the secretome**- of *Pythium irregulare* and *Pythium iwayamai* isolates when they were infecting plant hosts and when they were growing saprophytically. Additionally, these two treatments were performed at two different temperatures (4° and 19°C). The assembled transcriptomes were annotated, and a closer analysis of the expression profiles of transcripts coding for pathogenicity-related proteins is shown. Secreted proteins were semi-quantified and their likely functions were determined based on the annotation of the corresponding transcripts.

In the second part of this document, I include four appendices. Each one is about a different project that I contributed to during my stay at the Department of Bioagricultural Sciences and Pest Management. An article has been published for each project.

For the study included in appendix 1, I used basic plant pathology and microbiology techniques to fulfill Koch's postulates for a disease affecting Turkish filbert trees. I also used some molecular techniques to aid in the identification of the isolated bacteria.

The work presented in appendix 2 was a follow-up of the previous one. We sequenced the genome of *Xanthomonas arboricola* pv. *Corylina* isolated from Turkish filbert leaves. Then, we

assembled a draft genome of the bacterium. The assembled genome was annotated and the robustness of the assembly was verified.

Appendix 3 includes the report about the genome assemblies of 3 different isolates of *Lonsdalea quercina* subsp. *quercina*. As in the previous work, the genomes were annotated and the robustness of the assemblies was verified.

In appendix 4, the report of a study on the population structure of *Geosmithia morbida* is presented. This fungus and the walnut twig beetle *Pityophthorus juglandis* are associated with the Thousand Cankers disease of walnut and wingnut trees. There is big concern because the disease has moved from the western United States to the native range of a highly susceptible walnut species in the eastern USA. I contributed to this research in the analysis of the data, the statistical analyses and the writing of the paper.

## ACKNOWLEDGMENTS

I would like to acknowledge and thank my advisor Ned Tisserat for all his help and support, as well as my committee members Jan Leach, Courtney Jahn and Ann Hess.

My brother René and my sister Cecilia were very important to me during all these years. All my gratitude to them.

Also many thanks to my lab mates Marcelo, Megan and specially Rachael and Emily, for all their companionship, help and support during the good and the very difficult times.

Finally, I won't be here without the love of my daughter Luci: she is the inspiration of my life.

## TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iv
TRANSCRIPTOME AND SECRETOME OF TWO <i>PYTHIUM</i> SPECIES DURING INFECTION AND SAPROPHYTIC GROWTH.....	1
SUMMARY.....	1
INTRODUCTION.....	2
MATERIAL AND METHODS.....	4
RESULTS AND DISCUSSION.....	10
CONCLUSIONS.....	46
SUPPLEMENTAL FILES, FIGURES AND TABLES.....	48
REFERENCES.....	55
APPENDIX 1. Leaf Spotting of Turkish Filbert in Colorado Caused by <i>Xanthomonas arboricola</i> pv. <i>corylina</i> and <i>Pseudomonas syringae</i> pv. <i>syringae</i> .....	62
APPENDIX 2. Genome Sequence of <i>Xanthomonas arboricola</i> pv. <i>Corylina</i> , Isolated from Turkish Filbert in Colorado.....	67
APPENDIX 3. Genome Sequences of Strain ATCC 29281 and Pin and Northern Red Oak Isolates of <i>Lonsdalea quercina</i> subsp. <i>quercina</i> .....	71
APPENDIX 4. Population Structure of <i>Geosmithia morbida</i> , the Causal Agent of Thousand Cankers Disease of Walnut Trees in the United States.....	75

# TRANSCRIPTOME AND SECRETOME OF TWO *PYTHIUM* SPECIES DURING INFECTION AND SAPROPHYTIC GROWTH

## SUMMARY

We obtained and analysed the transcriptome and the secretome of *Pythium irregulare* and *Pythium iwayamai* isolates when they were infecting creeping bentgrass and when they were growing saprophytically in liquid medium. These two treatments were performed at two different temperatures (4° and 19°C). The assembled transcriptomes were annotated, and a closer analysis of the expression profiles of transcripts coding for pathogenicity-related proteins is shown. Secreted proteins were semi-quantified and their likely functions were determined based on the annotation of the corresponding transcripts.

Our results confirm but also expand previous work made on the subject using bioinformatic predictive tools. RxLR/Q effectors and cutinases were absent, and a low complement of pectinases was observed. Proteases, glycosidases, and ABC transporters showed the largest numbers of differentially expressed pathogenicity-related protein transcripts comparing the infection and saprophytic treatments. We found numerous metalloproteases that were not described previously in *Pythium* species. Significant differences on protease transcript expression were also observed comparing the two different temperatures in *P. iwayamai*.

Proteases, glycosidases and elicitors were the three main classes of proteins secreted by both *Pythium* species. Proteases appear to be the main class of enzymes that are used to break through the plant cell walls. Among glycosidases, only a single representative of family 7, with cellulose degrading activity, was found secreted by *Pythium irregulare* isolates when infecting plants; *Pythium iwayamai* isolates secreted family 7 and family 3, with cellulose degrading

activity also. Elicitins were secreted in much larger abundance in the saprophytic state, but their function is still unclear.

We did not find a correlation between transcript level or differential expression and protein secretion. We also find a very low correlation between the presence of signal peptide sequences for secretion in the transcripts and the actual secreted proteins. Those that had the signal could be secreted at a different time, but many without the signal were secreted, confirming other reports. These results point to the usefulness of the proteomics analysis to study the direct interaction of plant pathogens with their hosts.

## INTRODUCTION

The oomycetes are a group of organisms belonging to the phylum Heterokonta or Stramenopila (Lévesque 2011, Schroeder *et al* 2013). They include very important plant pathogens including the genera *Phytophthora*, *Pythium*, *Plasmopara* and *Peronospora* (Kamoun 2003). It has been suggested that Fungi- and oomycete-caused diseases are the biggest threat to global food production (Bebber and Gurr 2015, Fisher *et al* 2012).

Although not as devastating as their two very famous relatives: *Phytophthora infestans* and *Plasmopara viticola*, *Pythium* species are present in almost every soil constantly producing root and foliar diseases (Kamoun 2003, Bala *et al* 2010). Currently, more than 140 species are included in the genus *Pythium* (Martin 2000, Robideau *et al* 2011). Some *Pythium* species are able to grow and infect plants at temperatures just above freezing and have been found even in Antarctica surviving the freezing conditions in the form of resistant structures called oospores (Tojo and Newsham 2012; Bridge, Newsham and Denton 2008). Among these low temperature *Pythium* species is *P. iwayamai*, which causes a snow mold disease of several grass species

(Matsumoto 2009, Lipps and Bruehl 1978; Hoshino, Xiao and Tkachenko 2009). Another *Pythium* species, *P. irregulare* (probably a complex of species or a single variable species (Spies *et al* 2011)) is not normally considered a snow mold pathogen, but it has been shown being able to grow and produce disease at cold temperatures and often produces more disease at lower temperatures than other *Pythium* species (Hendrix and Campbell 1973, Cantrell and Dowler 1971, Stovold 1974).

*Pythium* species are necrotrophs and they promptly kill cells of their plant hosts after entering their roots or leaves. They penetrate host tissues mainly through stomata and wounds, as they are not able to produce enough pressure to break through intact epidermal cells (McDonald *et al* 2002). Once inside, they need to overcome plant cell defenses and breach cell walls to reach the cytoplasm. They can do that by deploying the effector molecules directed against the plant defenses. **Effectors, broadly defined, include enzymes that can degrade plant cell wall constituents, as well as other molecules that disable or reduce plant defense responses** (Kamoun 2006). A number of studies have examined oomycete effectors, primarily *Phytophthora* effectors, (Kamoun 2006, Judelson 2012, Ellis *et al* 2009, Haas *et al* 2009, Jiang and Tyler 2012) and these results have sometimes been extrapolated to infer effectors for oomycetes in general. But genomic and some transcriptomic studies have shown that some well-characterized *Phytophthora* effectors are absent or under-represented in *Pythium* (Lévesque *et al* 2010, Adhikari *et al* 2013). *Phytophthora*, a hemi-biotroph, first establishes a close interaction with plants cells and does not kill them until late during the disease cycle. In contrast, *Pythium* is a necrotroph that kills host cells at the onset of infection. Thus, one might expect a different effector repertory.

During the current genomics era, many studies on oomycetes gene repertoires have been done using gene prediction softwares (Zerillo *et al* 2013, Wawra *et al* 2012, Adhikari *et al* 2013). Expression support has been obtained for some *Pythium* species but it is limited for genes expressed during the infection process (Adhikari *et al* 2013). Gene products that are secreted by oomycetes have also been predicted, mainly through the use of “secretion signal domains” finding softwares (Kamoun 2006, Raffaele *et al* 2010, Bendtsen 2004), but it has been shown in some cases that the correlation is low between the predicted and the actual proteins that are secreted (Soanes, Richards and Talbot 2007, Antelmann *et al* 2001, Meinken and Min 2012). The objective of our study was to examine and compare the transcriptome of *P. iwayamai* and *P. irregulare* isolates during saprophytic growth in a liquid medium and during an early pathogenic phase when these oomycetes were infecting creeping bentgrass (*Agrostis stolonifera* L.). Specifically, we were interested in expression of pathogenicity-related proteins during the infection process. We broadened the transcriptome analysis by comparing the transcript expression of pathogenicity-related proteins at two different temperatures (4°C and 19°C). Additionally, we made proteomic profiles of the proteins secreted by these isolates during saprophytic growth and early infection and compared them and to the corresponding transcriptomes. Our findings, and their correspondence with the predicted expressed and secreted proteins, are discussed.

## **MATERIALS AND METHODS**

Two isolates each of *Pythium iwayamai* S.Ito (P174a or CBS 132417, and P191 or CBS 132418) and *Pythium irregulare* Buisman (P174b or CBS 139189 and P176 or CBS 250.28), were used in this study. Isolates P174a and b were collected from annual bluegrass (*Poa annua* L.) on a golf

course fairway exhibiting snow mold symptoms near Aspen, CO in 2009. P191 was collected from an annual bluegrass golf fairway near Steamboat Springs, CO in 2011. Isolate P176 was collected from a diseased bean (*Phaseolus vulgaris* L) in The Netherlands in 1928 (Van Der Plaats-Niterink 1981). Isolates were stored at 4°C in sterile water. To prepare inoculum, isolates were first grown for one week at 25°C (P174b and P176) or 4°C (P174a and P191) in potato dextrose agar (PDA) (Difco, Maryland, CN 213400) and then stored at 4°C for two weeks. For each isolate, two, 1-cm-diameter, agar plugs containing *Pythium* mycelium were put in flasks filled with 100 ml of potato dextrose broth (PDB) (Difco CN 254920) and incubated at 12°C with agitation (80 rpm) for at least 4 days. Mycelium from three flasks was removed from the broth and placed in a Petri dish filled with sterile, distilled water. The agar plug was carefully removed using a sterile scalpel and the mycelium was divided into approximately equal volumes (Fig.1.1a) that was later added to each of the treatment vials.

The *Pythium* mycelium was exposed to four treatments. To observe potential protein expression differences during the saprophytic and pathogenic phases of the oomycete's growth, mycelium was placed in vials containing just a growth medium ("in medium" or MED) or in vials containing the medium and creeping bentgrass (*Agrostis stolonifera* L cv. 'Penncross') seedlings (designated 'with plants' or WP). Then, to observe the effects of temperature on transcript expression, the four *Pythium* isolates growing in MED and WP were exposed to two different temperatures (4°C and 19°C) for a total of 4 treatment combinations.

In the WP treatment, the annual bluegrass seed was disinfested in a 0.5% sodium hypochlorite solution with constant agitation for 10 minutes and rinsed with sterile distilled water. Seed was placed on a sterile Whatman #2 filter paper. Excess water was removed by vacuum filtering and the seed was then air dried overnight. Seed aliquots (0.2g) were placed in polystyrene square

Petri plates (Carolina, North Carolina, CN 741470) containing 6 ml of 15% water agar (Difco) amended with 0.08% of Miracle-Gro plant food (Scotts Miracle-Gro, Ohio, CN 1001123), and the following antibiotics per liter: 0.125g ampicillin (Sigma, Missouri, CN A0166), 200  $\mu$ l piramicin solution 2.5% (MP Biochemicals, California, CN 156254), 500  $\mu$ l rifampicin solution (0.01g rifampicin Sigma, CN R7382-5G dissolved in 1 ml DMSO Sigma, CN D5879) and 1 ml PCNB solution (2g Pentachloronitrobenzene Sigma, CN P3395 dissolved in 400 ml 95% ethanol ACROS, Pennsylvania, CN 61509-0020) (Fig.1.1b). Seed was incubated at 25°C with 11 hours of light. Seed germination occurred after 2-3 days. After 14 days the temperature was lowered to 18°C for three additional days. For inoculation, plates with growing plants were placed horizontally and 15 ml of a minimal medium containing 3g NaNO<sub>3</sub> (Fisher, Pennsylvania, S343-500), 1g K<sub>2</sub>HPO<sub>4</sub> (Fisher P-288), 5g MgSO<sub>4</sub> (Mallinckrodt, Missouri, 6070-3), 0.5g KCl (Mallinckrodt 6858), 0.01 mg FeSO<sub>4</sub>·7H<sub>2</sub>O (JT Baker, Pennsylvania, 2070-01), and 0.16 g NH<sub>4</sub>Cl (Mallinckrodt 3384) per L (adjusted to pH 6.0) was added to the plate. *Pythium* mycelium was placed close to the tip of the leaves (Fig.1.1c).

For the MED treatments mycelium was placed in Petri dishes in which 15 ml the minimal medium plus 0.5 g asparagine (JT Baker B583-05) and 7 g dextrose (Difco, 0155-17-4) had been added. The addition of the sugar and amino acid was necessary because preliminary experiments indicated the *Pythium* isolates grew poorly in the minimal medium only. Eight replicate plates for each treatment were prepared. For each isolate, four plates were incubated at 19°C and another set of four plates at 4°C in the dark. The WP treatments were checked twice daily for any evidence of mycelium growing on of the leaf surface.



Figure 1.1. Experimental setting. (a) Mycelium grown in PDB were rinsed in water and divided in two similar volumes for each treatment. (b) Creeping bentgrass seeds were sown in square Petri dishes containing agar. (c) The mycelium was placed close to the leaf tips.

A sample was collected from treatment vials at each temperature when hyphae were observed growing from approximately 50% of the leaves (Fig.1.2a and 1.2b) and when no symptoms (chlorosis or necrosis) were observed in the leaves. For RNA extraction, 0.1g of leaves+mycelium (WP) or 0.1g of mycelium alone (MED) were collected, immediately frozen in

liquid nitrogen, and stored at -80°C until processed. For proteomics analysis, 1 ml aliquots of the liquid medium **from the same plates** were removed and stored at -80°C until analyzed.

RNA from two replicates (two different plates) of each treatment combination was extracted using a combination of trizol (Ambion, California, CN 15596026) and filtration columns (as described in [http://www.untergasser.com/lab/protocols/rna\\_prep\\_comb\\_trizol\\_v1\\_0.htm](http://www.untergasser.com/lab/protocols/rna_prep_comb_trizol_v1_0.htm), with modifications): to each sample, two steel beads (2.3 mm diameter, BioSpec, Oklahoma, CN 11079123ss) and 1 ml of trizol were added and then shaken two times for 25 sec at 4.0 m/sec in a FastPrep FP120 homogenizer (Thermo Savant, California). Samples were centrifuged at 12,000xg for 10 min at 4°C. 200 µl of chloroform (Fisher, CN C606-1) were added to the supernatant. The samples were shaken by hand vigorously and then left to settle for 3 min. The samples were centrifuged again at 12,000xg for 15 min at 4°C. To the supernatant (400-450 µl), 400 µl of 96% ethanol were added, mixed well by inverting several times, and kept at room temperature for 10 min. RNA was then purified using column-tubes and reagents from a “High pure RNA isolation kit” (Roche, Indiana, CN 11828665001) after the lysis step. mRNA-Seq libraries preparation and sequencing were performed at the University of Southern California Epigenome Center (Los Angeles CA). The 75 cycle single-read runs were executed in an Illumina NexSeq 550 system.

Transcriptome analysis, using the two replicates of each treatment combination, was performed using the Tuxedo protocol (Trapnell *et al* 2010, Trapnell *et al* 2012, Trapnell *et al* 2013) with Bowtie2 version 2.2.4, TopHat v. 2.0.13, Cufflinks v. 2.2.1 and CummeRbund v. 2.7.2. Reads were mapped to the corresponding *P. iwayamai* and *P. irregulare* genome contigs (<http://pythium.plantbiology.msu.edu>) with a minimum intron length setting of 5 bp and a maximum of 5,000 bp. For the transcript abundance estimation, the maximum intron length was

also set to 5,000 bp; the upper quartile normalization option in Cufflinks was used to “improve robustness of differential expression calls for less abundant genes and transcripts” (<http://cole-trapnell-lab.github.io/cufflinks/cufflinks/index.html>). In the final transcript abundance comparisons, at the “gene” level, only those transcripts with a fragments per kilobase of transcript per million mapped reads (fpkm) value larger than 1.0 in at least one of the conditions were considered. Transcript sequences were obtained from the Cufflinks output using the `fastaFromBed` utility in BEDtools (Quinlan and Hall 2010). Transcript sequences were blasted to the NCBI nr database and to the Uniprot TrEMBL database using the NCBI BLAST+ suite v. 2.2.30. Blastx results (top hits only) were imported into Blast2GO v. 2.8 (Götz *et al* 2008) to obtain Gene Ontology (GO) and Interpro annotation. For the “pathogenicity related” group and for proteins found in the proteomics analysis, GO terms without a corresponding Interpro signature were verified using the NCBI Conserved Domain Search tool (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

Protein LC-MS/MS from 3 replicates (3 different plates) was performed at the Colorado State University Proteomics and Metabolomics Facility (Fort Collins, CO). Protein samples were in-solution trypsin digested and then analyzed using a Thermo Scientific Orbitrap Velos MS coupled with nanoHPLC instrument. Each sample (30 µg) was injected twice for the LC-MS/MS. Results from the two injections were combined together and considered as a single replicate.

Proteomics analysis, using the three replicates of each treatment combination, was conducted following Vaudel *et al* tutorial (Vaudel *et al* 2014). Raw output files from the mass spectrometer were converted to mgf files with MSConverter (Chambers *et al* 2012). Databases were generated by performing 6-frame translations of the transcriptomes obtained for the corresponding species,

using EMBOSS v. 6.6.0 (Rice, Longden and Bleasby 2000). Spectra were searched against the corresponding database with SearGUI 1.23.3 (Vaudel et al 2011) selecting X!Tandem (Craig and Beavis 2004), MS-GF+ (Kim and Pevzner 2014) and OMSSA (Gerr *et al* 2004) as searching engines. Identified peptides and proteins were obtained using PeptideShaker v. 0.37.3 (Vaudel *et al* 2015). Only “validated” proteins and only those proteins present in at least two out of the three replicates were considered as present in the corresponding samples. Replicates were normalized by Total Spectral Count Normalization (Gokce *et al* 2011), and then corrected for dilution of the original samples. Equality of variances was tested using Levene’s test, and T-test were made comparing WP and MED treatments. P-values were corrected for multiple comparisons by Benjamini & Hochberg method.

## **RESULTS AND DISCUSSION**

### **Infection process and sampling**

Mycelium of both *P. iwayamai* isolates was observed on approximately 50% of the creeping bentgrass leaf blades at 6 and 10 days after inoculation at 19° C and 4° C, respectively (Fig. 1.2a and b) and samples for RNA sequencing and proteomics analysis were collected at this time. No leaf necrosis was apparent. Both isolates continued to colonize and form oospores in the plant tissue (Figs. 1.2c and d) resulting in extensive chlorosis and necrosis and by approximately 21 days at 19° C and 35 days at 4° C respectively, the plants were dead (Figs. 1.3a and b). Similarly, mycelium of *P. irregulare* isolates was observed on approximately 50% of leaf blades five and 13 days after inoculation at 19° C and 4° C, respectively, and samples were collected at this time. Plants were killed after approximately 20 days at 19° C by both isolates and 35 days at 4° C by

isolate 174b (Figs. 1.3a and b). *P. irregulare* isolate 176, originally isolated from bean roots, did not cause chlorosis or necrosis to the bentgrass at 4° C, even after 7 weeks incubation.

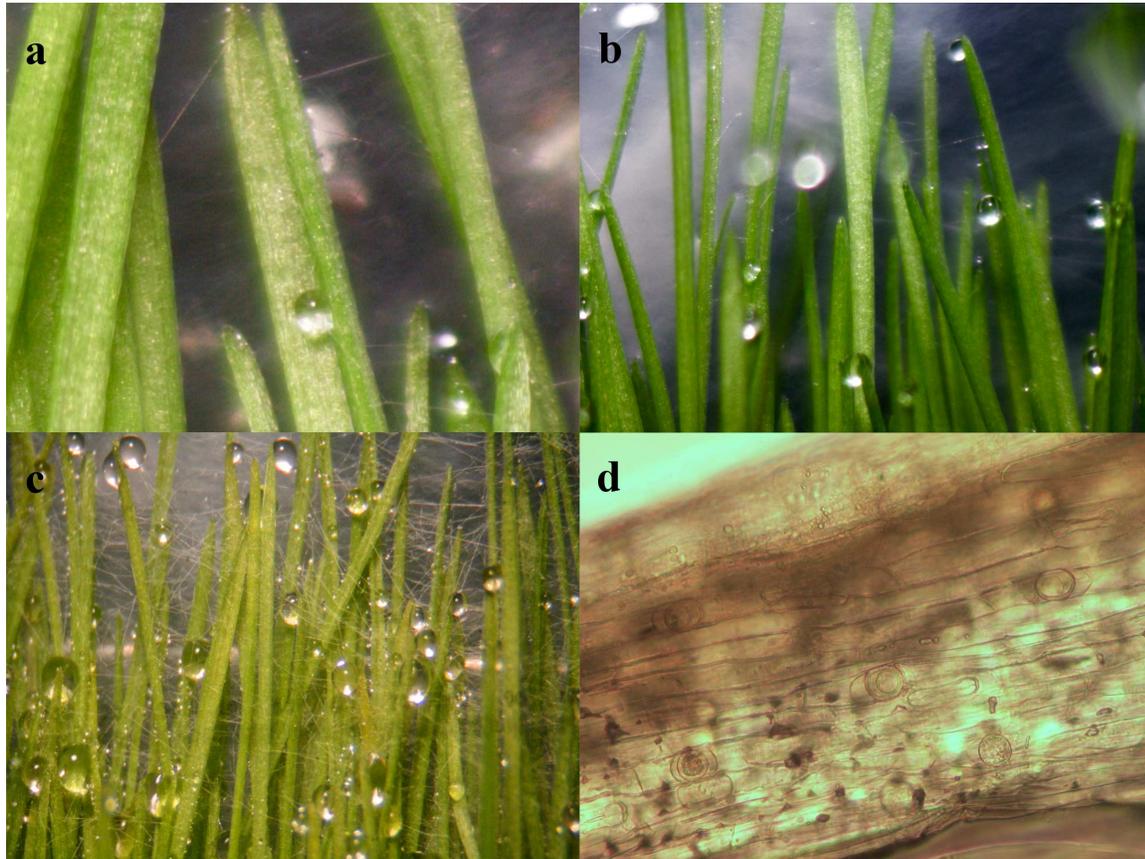


Figure 1.2. Sampling for transcriptome and secretome analysis occurred when approximately 50% of the asymptomatic leaves exhibited hyphal growth (a and b). Later, abundant hyphae colonized the leaves causing chlorosis and necrosis (c), and oospores eventually formed inside the leaf cells (d).

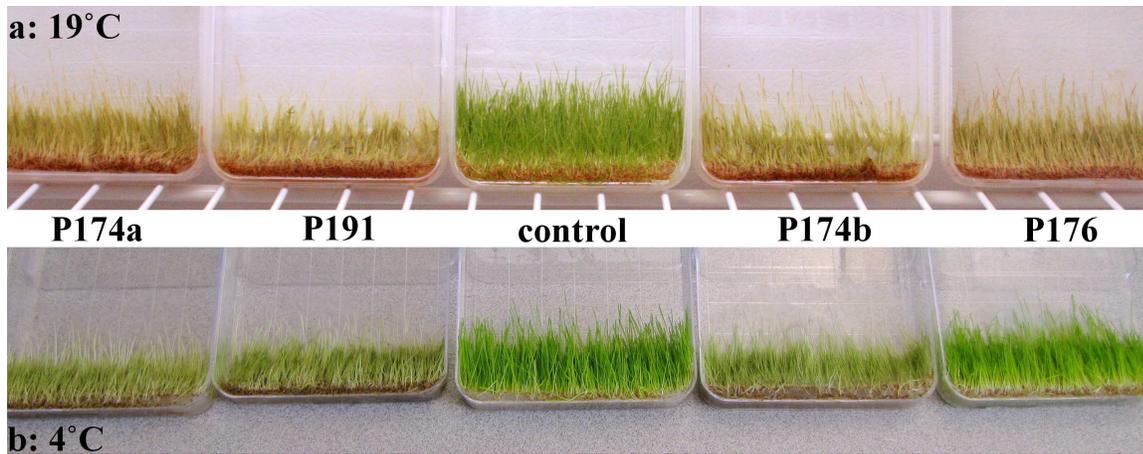


Figure 1.3. Disease severity in creeping bentgrass following inoculation with *P. iwayamai* isolates 174a and 191 and *P. irregulare* isolates 174b and 176 at 4°C (bottom) and 19°C (top). All isolates except *P. irregulare* 176 at 4°C eventually killed the grass. Non-inoculated plants (middle containers) remained asymptomatic.

### Transcriptome analysis overview

More than 20 million reads, with an average read length of 74 bp, were obtained for each *Pythium* isolate and treatment combination (Table S1). This represented a total of 13,216 transcripts corresponding to 11,767 different genes (1,449 transcripts were isoforms) in *P. irregulare* and 17,219 transcripts corresponding to 15,873 genes (with 1,346 isoforms) in *P. iwayamai* based on Cufflinks transcriptome assembly (Supplemental files 1 and 2). Transcripts were matched to both the NCBI nr and the Uniprot TrEMBL databases by BLASTx, and while the resulting number of hits and distribution of the top hits were similar for the two *Pythium* spp., they did vary between databases (Fig. S1a-d). Most of the hits in both databases were to related oomycete (*Phytophthora*, *Pythium*, *Albugo*, and *Hyaloperonospora*) genes. The number of hits for *P. iwayamai* and *P. irregulare* were highest to *Phytophthora* spp. in the NCBI nr database, but to the more closely related *P. ultimum* in the UNIPROT TrEMBL database, probably reflecting the greater number of *P. ultimum* transcripts annotated in the latter database.

Along with statistical data and BlastX information, Gene Ontology and InterproScan terms were also obtained (Tables S2-17), but GO terms were actually associated to the Blastx hits and not to the sequences *per se*. Even with low e-values, the percent similarity for most of the transcripts to the top hit was <95% (Tables S2-17). Therefore, for a better assignment of a possible function of the transcript-coded or detected proteins, we relied on results of InterproScan sequence searches that depended on the actual sequence of the transcript or protein. When needed to confirm or assign a function, results obtained using the NCBI Conserved Domains Search Tool were also added to the annotation of selected transcripts (transcripts coding for pathogenicity-related proteins, differentially expressed transcripts and transcripts coding for detected secreted proteins).

The consistency of transcript outputs among treatments was assessed by looking at the distribution of fragments per kilobase of transcript per million mapped reads (fpkm). The fpkm values for each treatment combination were similar in density plots (Figs. S2a and b) and boxplot graphs (Figs. S2c and d), indicating that transcript outputs behaved similarly. Scatterplots of all transcript fpkm values indicated there was less dispersion between the two isolates of the same species when grown in the WP relative to the MED treatments at both temperatures (Fig. S3). Furthermore, volcano plots indicated there were more differentially expressed genes between isolates of the same species in MED compared to WP treatments at both temperatures (Fig. S4). Thus, the presence of the living host tended to reduce differences in transcript expression levels between the two species isolates.

## **Expression of pathogenicity-related protein transcripts**

We focused our analysis on those transcripts that coded for putative pathogenicity-related proteins as defined by Adhikari *et al* (2013), as well as metallopeptidases, other peptidases, and peroxidases that might be involved in the infection process (Table 1.1).

Proteases were the most represented pathogenicity-related protein transcripts, followed by ABC transporters and glycoside hydrolases. The numbers of different transcripts in both species for aspartic proteases, pectate lyases, lipases, phospholipases, peptidase inhibitors, cytochrome P450, necrosis inducing proteins, elicitors, and crinkler proteins, were similar to or slightly smaller than the number of predicted genes reported by Adhikari *et al* (2013). A larger number of cysteine peptidases and lectin-like transcripts were expressed in both *Pythium* species whereas serine peptidases, glycoside hydrolases and ABC transporters were expressed in similar or smaller numbers in *P. irregulare* but in larger numbers in *P. iwayamai* compared to their predicted genes (Adhikari *et al* 2013). The larger number of transcribed vs. predicted genes could have resulted from a single gene producing multiple transcripts or because of the difficulty in predicting genes with many or short exons, long introns, or unorthodox starting or splicing sites (Minoche *et al* 2015).

We checked whether transcripts for each pathogenicity-related protein category were differentially expressed ( $FDR \leq 0.01$ ) during plant infection (WP) compared to saprophytic growth (MED) at both 4° and 19°C (Table 1.2). Most of the putative pathogenicity-related protein transcripts did not show differential expression, suggesting that they are transcribed constitutively, and at similar levels, regardless of the presence of plants. Therefore, many of the corresponding proteins may not be directly involved in pathogenesis but instead act as house-keeping enzymes or degrading proteins present in the surrounding medium.

In general, when differential expression was observed, there were more pathogenicity-related protein transcripts differentially expressed when isolates were infecting plants relative to saprophytic growth (WP>MED, Table 1.2). It would be expected that the proteins coded by these transcripts were involved in the interaction of *Pythium* mycelium with the host. Nevertheless some pathogenicity-related transcripts in certain protein categories were differentially expressed when mycelium was growing saprophytically (MED>WP), sometimes more than when infecting plants. This phenomenon also was observed for a number of transcripts that were not considered pathogenicity-related (Tables S6-9 and S14-17). Therefore, there appear to be two different sets of transcripts coding for different proteins with the same biochemical function (for example, protein cleavage) but possibly with different biological function; one set being differentially expressed during the infection process and the other during saprophytic growth.

The lack of cutinase candidate transcripts in both species is in agreement with their absence in the *Pythium* genome as reported by Adhikari *et al* (2013). These results support the hypothesis of Zerillo *et al* (2013) and others (Lévesque *et al* 2010) that globose sporangial *Pythium* species, including *P. irregulare* and *P. iwayamai*, are less effective at degrading suberized roots, leaves and stems than other plant pathogenic oomycetes that produce cutinases; nevertheless, they can penetrate leaves possibly through stomata.

Pectate lyase/pectin esterases (PL/PE) are involved plant cell wall and middle lamella maceration and are important during plant infection by certain oomycetes (Marín-Rodríguez, Orchard and Seymour 2002, Jia *et al* 2009). However, Zerillo *et al* (2013) found that *P. iwayamai* and *P. irregulare* had fewer annotated PL/PE genes than other *Pythium* and *Phytophthora* species, and concluded that these species likely did not completely saccharify this complex sugar. We also observed only a small number of PL/PE transcripts, with only three at 4°C and one at 19° in

isolate 174b and one in isolate 176 at both temperatures, that were differentially expressed in *P. irregulare* in the presence of plants and just one in *P. iwayamai*, in the saprophytic treatment (Table 1.2). Furthermore, no PL/PE proteins were detected among secreted proteins (Tables S18-S21). This is in agreement with our observation that there was not extensive leaf maceration caused by these two species in our experiments (Fig.1.3).

We did not find RxLR/Q effector transcripts in *P. irregulare* or *P. iwayamai*. A large number of probable RxLR/Q effector genes are present in *Phytophthora*, but not *Pythium* genomes (Adhikari *et al* 2013). Some RxLR proteins have been associated with suppression of plant defense responses by biotrophic or hemi-biotrophic oomycetes (Kamoun 2006, Win *et al* 2012) and their absence in the two *Pythium* species studied here is consistent with their necrotrophic life style. As necrotrophs, they do not establish a complex relationship with the host as do biotrophs but rather kill host cells to gain access to their nutrient content.

Table 1.1. Number of transcripts coding for pathogenicity-related proteins. For some categories, isoforms were detected and the total numbers including them are shown in parenthesis. The identification of Crinkler-coding transcripts was based only on BlastX results because no characteristic domain or other signature for them was found in the Interpro nor Conserved Domains databases.

Protein	Number of transcripts	
	<i>Pythium irregulare</i>	<i>Pythium iwayamai</i>
Aspartic peptidase	22 (23) <sup>x</sup>	29 (30)
Cysteine peptidase	64 (78)	77 (97)
Serine peptidase	103 (111)	142 (153)
Metallo peptidase	70 (86)	92 (106)
Other peptidases	15 (16)	13 (16)
Total proteases	274 (314)	353 (402)
Glycoside hydrolase	89 (103)	114 (130)
Peptate lyase /pectin esterase	11	9 (12) <sup>y</sup>
Lipase	22 (25)	24 (28)
Phospholipase	26 (28)	27 (28)
Peptidase inhibitor	13	15
Cytochrome P450	38	52 (55)
ABC transporter	118 (124)	160 (164)
Necrosis inducing protein	1	3
Elicitin-like protein	37	29
Lectin-like protein	48 (56)	54 (61)
Crinkler	3	7
Peroxidase/Catalase/Super oxide dismutase	24 (25)	22 (25)
RXLR/Q effector candidate	0	0
Cutinase	0	0

<sup>x</sup>Numbers in parentheses indicate the total number of transcripts including isoforms.

<sup>y</sup>Includes the polysaccharide lyase rhamnogalacturonase.

Table 1.2. Number of differentially expressed (FDR≤0.01) transcripts in the different categories of pathogenicity related proteins, for *Pythium irregulare* isolates a) 174b and 176, and *P. iwayamai* isolates b) 174a and 191, at 4°C and 19°C.

Protein	Number of differentially expressed transcripts							
	<i>Pythium irregulare</i> 174b				<i>Pythium irregulare</i> 176			
	4°C		19°C		4°C		19°C	
	WP>MED <sup>x</sup>	MED>WP	WP>MED	MED>WP	WP>MED	MED>WP	WP>MED	MED>WP
Protease	21	2	30	4	19	1	16	3
Glycoside hydrolase	11	4	6	5	8	1	4	1
Pectate lyase / pectin esterase	3	0	1	0	1	0	1	0
Lipase	2	0	1	2	3	0	0	0
Phospholipase	2	0	1	0	3	0	1	0
Peptidase inhibitor	0	0	0	1	3	2	0	0
Cytochrome P450	5	0	3	0	5	0	0	0
ABC transporter	7	1	8	2	7	7	9	6
Necrosis inducing protein	1	0	0	0	0	0	0	0
Elicitin-like protein	5	6	4	3	4	1	3	2
Lectin-like protein	3	0	1	1	3	0	4	0
Crinkler	0	0	1	0	0	0	0	0
Peroxidase / Catalase / Super oxide dismutase	3	1	2	0	0	1	4	0

Protein	Number of differentially expressed transcripts							
	<i>Pythium iwayamai</i> 174a				<i>Pythium iwayamai</i> 191			
	4°C		19°C		4°C		19°C	
	WP>MED <sup>x</sup>	MED>WP	WP>MED	MED>WP	WP>MED	MED>WP	WP>MED	MED>WP
Protease	8	22	39	13	7	7	33	11
Glycoside hydrolase	9	3	7	5	7	4	13	2
Pectate lyase / pectin esterase	0	1	0	0	0	0	0	0
Lipase	1	0	1	0	1	0	0	0
Phospholipase	1	3	1	1	2	1	1	0
Peptidase inhibitor	1	0	0	2	0	0	0	0
Cytochrome P450	0	3	0	1	3	5	2	1
ABC transporter	3	10	7	2	6	1	7	8
Necrosis inducing protein	0	0	0	0	0	0	0	0
Elicitin-like protein	2	1	3	0	2	1	2	0
Lectin-like protein	3	2	1	2	2	2	1	4
Crinkler	1	0	0	0	2	0	0	0
Peroxidase / Catalase / Super oxide dismutase	2	0	2	3	3	2	1	1

<sup>x</sup> WP>MED represents overexpression of transcripts when mycelium was grown with plants relative to saprophytic growth in medium only; MED>WP represents the opposite overexpression of transcripts.

The largest numbers of differentially expressed pathogenicity-related transcripts were found in proteases, glycoside hydrolases and ABC transporters, possibly reflecting the fact that those were the most represented categories. Proteases and glycosidases along with elicitors were the most abundant of secreted proteins as well (Fig.1.16, Tables S18-S21). These pathogenicity-related proteins are discussed in more detail below.

## PROTEASES

Certain proteases may be involved in the degradation of plant cell wall proteins (Cassab 1998, Li *et al* 2010) and facilitate infection by *Pythium*. These and other degraded proteins can also be used for nutrition (Davis *et al* 2005). A total of 314 and 404 protease transcripts (including isoforms) were found in *P. irregulare* and *P. iwayamai*, respectively (Table 1.1). The overall expression profiles for these proteases were generally similar between isolates and species (Figs. 1.4a-h), with most transcripts having a low fpkm value. However, there were differences in expression levels of several transcripts at the two temperatures tested. Three transcripts in *P. irregulare* isolate 174b had a higher expression (FDR>0.01, data not shown) at 19°C whereas 9 transcripts were higher at 4°C. Four transcripts in *P. irregulare* isolate 176, were higher at 19°C and only one at 4°C. In contrast, numerous transcripts in *P. iwayamai* had different (FDR>0.01) levels of expression. There were 48 transcripts in isolate 174a with higher expression at 19°C and six at 4°C. Similarly, there were 33 transcripts in isolate 191 with higher expression levels at 19°C versus seven at 4°C (data not shown). A considerable number of differentially expressed transcripts was also observed in *P. iwayamai* isolates at 19°C compared to 4°C (Fig. 1.5). A possible explanation for this phenomenon is that *P. iwayamai* is a psychrophilic pathogen that goes dormant as temperatures increase (Matsumoto 2009). We observed that our *P. iwayamai* isolates stopped growing and lost viability above 20°C (data not shown). Thus, these isolates

likely were under stress at 19°C even though they were still capable of causing disease. It is possible that *P. iwayamai* proteases that are functional at a low temperature lose activity as temperatures increase because of thermal denaturation (Hoshino, Xiao and Tkachenko 2009; Feller 2013) while another set of proteases become overexpressed (because we did not gather secretome data for *P. iwayamai* at 19°C, we don't know whether these proteases were actually secreted). *P. irregulare*, on the other hand, grows and infects plants over a wide temperature range (Cantrell and Dowler 1971) and the protease expression profiles of our isolates appeared not to be as affected by temperature.

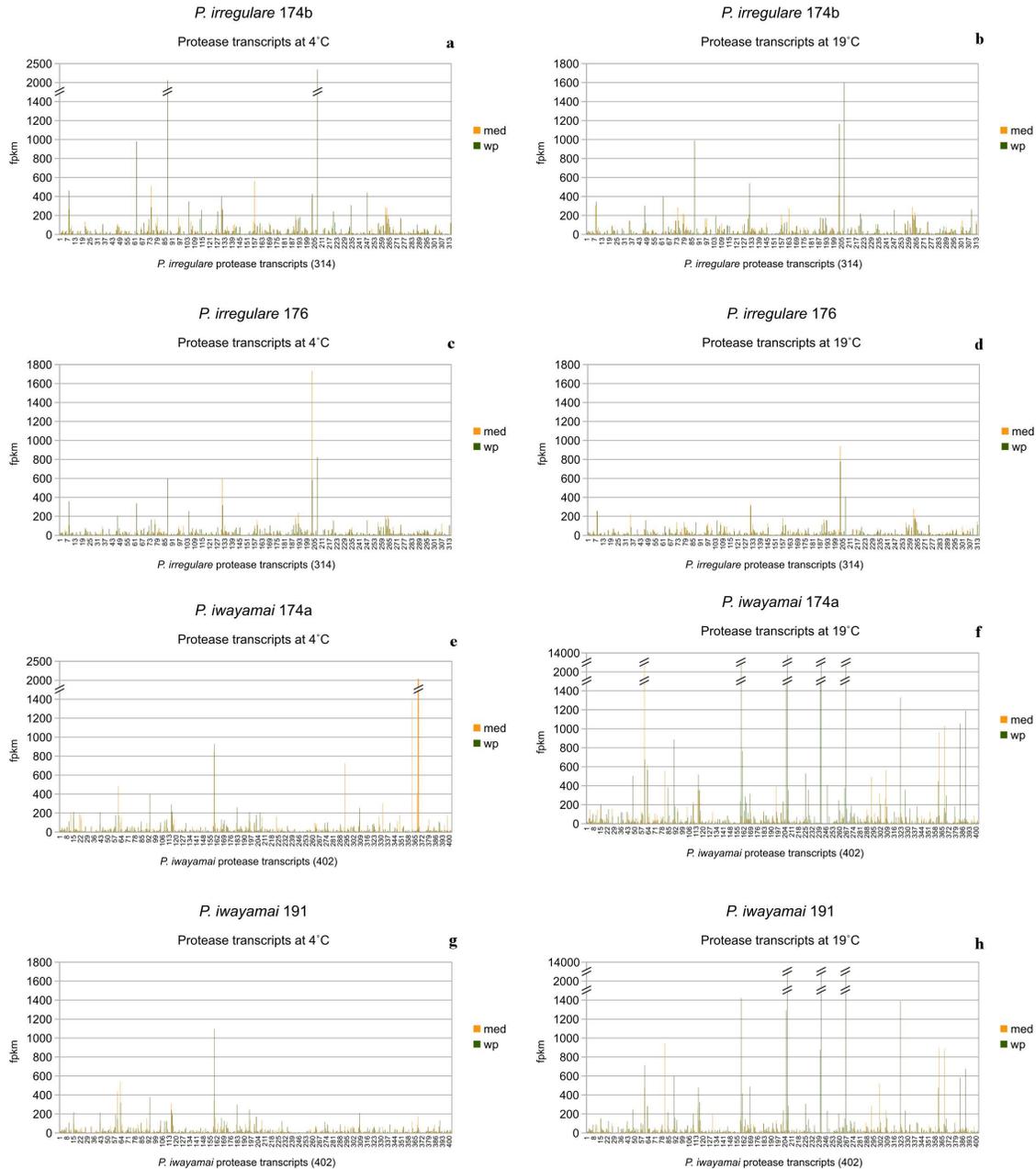


Figure 1.4. Expression level (fpkm) of all protease transcripts in *P. irregulare* (a-d) and *P. iwayamai* (e-h). Note the change in scale at the top of the graphs. Numbers on the x-axis in graphs of the same species correspond to the same transcript.

Of the protease transcripts that were differentially expressed ( $FDR \leq 0.01$ ), most were in the WP compared to the MED treatment at the two temperatures tested (Table 1.2a-d, Figs. 1.5a-h), except in *P. iwayamai* 174a at 4°C where 8 and 22 proteases were differentially expressed in the

WP and MED treatments, respectively. Proteases coded by some of these transcripts, particularly those differentially expressed in the presence of plants, may be involved in the infection process.

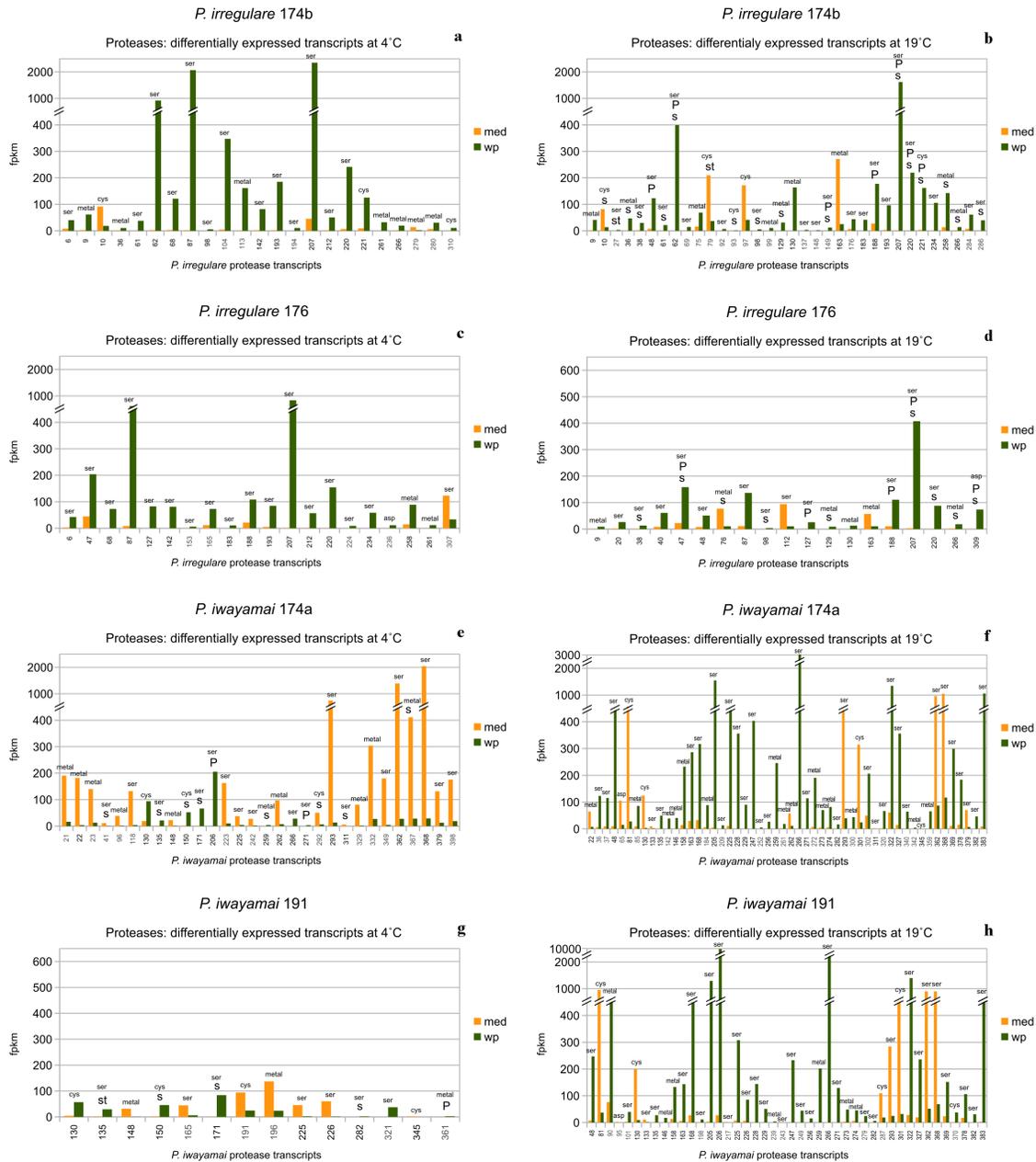


Figure 1.5. Differentially expressed transcripts (FDR $\leq$ 0.01) coding for **proteases** in *P. irregulare* (a-d) and *P. iwayamai* (e-h), at 4°C and 19°C. Bars were labeled according to the protease they coded for: ser(ine), metal(lo), cys(tiene) or asp(artic). Bars were labeled with a “P” on top when the corresponding protein was detected in the proteomics analysis (performed at 19°C for *P. irregulare* and at 4°C for *P. iwayamai*). In these same graphs, an “s” was added when the corresponding transcript contained a signal peptide sequence and “st” when they contained a

signal peptide and transmembrane domain sequence. The same numbers on the x-axis represent the same transcripts **for isolates of the same species**.

The majority of differentially expressed protease transcripts in both *Pythium* species were for serine peptidases (Fig. 1.5). These differentially expressed serine peptidase transcripts were present in a higher proportion compared to the total number of proteases expressed (Table 1.1). This is similar to what has been observed in other filamentous fungi (Olivieri *et al* 2002, Davis *et al* 2005, Nair and Geethu 2015, Muszewska *et al* 2011). Serine proteases are considered important virulence factors in Fungi and oomycetes (Olivieri *et al* 2002, Davis *et al* 2005, Nair and Geethu 2015). Additionally, the two *P. iwayamai* but not *P. irregulare* isolates also secreted a number of metallopeptidases (Fig. 1.7c and d). Metalloproteases have been shown to play a role in the virulence of several plant pathogens (Dow, Davies and Daniels, 1998; Davis *et al* 2005; Meijer *et al* 2014; Jashni *et al* 2015), and they could have a similar role in *P. iwayamai* pathogenicity.

Approximately 45% of the differentially expressed protease transcripts in *P. irregulare* isolates included a signal peptide sequence and for 10 of them there were corresponding secreted proteins (Figs. 1.5b and d). In contrast, fewer differentially expressed protease transcripts included a signal peptide sequence in *P. iwayamai* isolates, and none of the corresponding proteases were secreted (Figs. 1.5e and g).

The relative abundance of all transcripts that coded for secreted proteases, regardless of their statistical significance, were compared (Fig. 1.6). Transcript expression levels within each *P. irregulare* isolate were similar with the exception of high levels of expression in transcripts 62 and 87 in isolate 174b (Figs. 1.6a and b). Overall, transcript expression was lower in *P. iwayamai*, but relative expression levels between the two *P. iwayamai* isolates were also similar with the exception of transcript 206 (Figs. 1.6c and d). Approximately half of the transcripts in *P.*

*irregulare* had a signal peptide sequence including one (149) with a transmembrane region sequence (Fig. 1.6b). Few *P. iwayamai* transcripts included a signal peptide sequence, and in disagreement with what is expected for transcripts of secreted proteins (Meiken and Min 2012, Raffaele *et al* 2010).

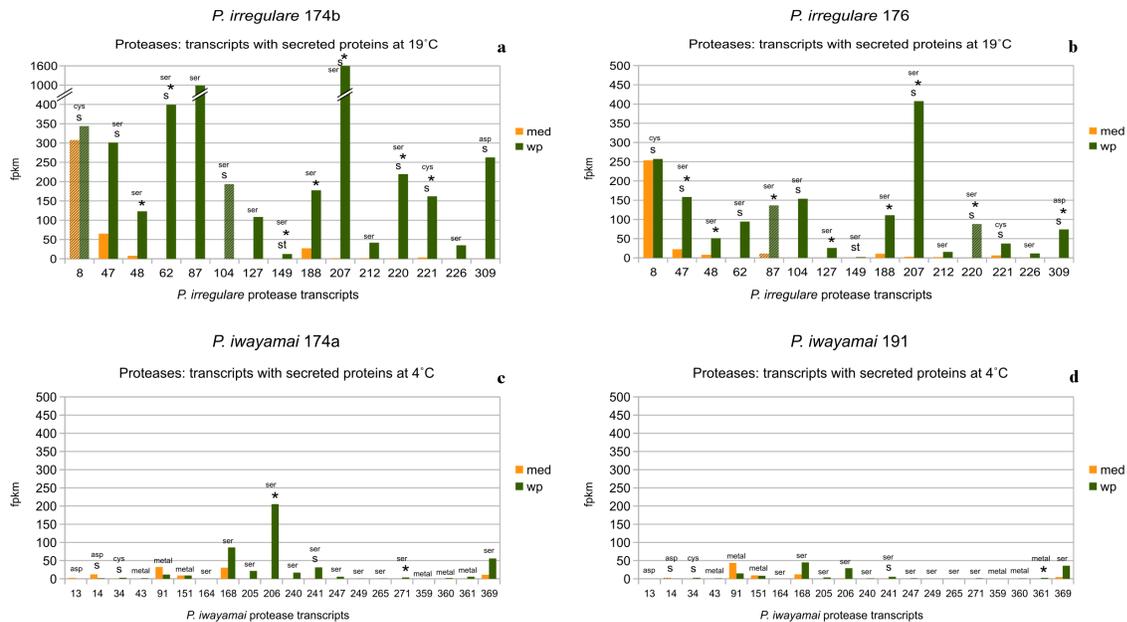


Figure 1.6. Transcripts coding for secreted proteases in *P. irregulare* (a,b) and *P. iwayamai* (c,d). Note the change in scale at the top of isolate 174b graph (a). Bars included a diagonal pattern when the corresponding protein was detected only in one isolate of the same species. An “s” was added above bars when the corresponding transcript contained a signal peptide sequence and “st” when they also contained a signal peptide and transmembrane domain sequence. Those bars labeled with a \* had significantly different fpkm number ( $FDR \leq 0.01$ ). The same numbers on the x-axis represent the same transcripts **for isolates of the same species**.

A proteomics analysis was performed for each isolate of *P. irregulare* (19°C only) and *P. iwayamai* (4°C only), respectively. The proteases exhibited the largest diversity of secreted proteins in all the pathogenicity-related protein categories (Fig.1.16, Tables S18-21). However, many of the protease transcripts identified in the transcriptomes (Fig. 1.4, Table 1.1) were not secreted. Furthermore, those that were secreted often had large variations in their spectral count values among the three replicate samples, making treatment comparisons difficult. Nevertheless,

the overall pattern of spectral counts was similar in both species, despite the fact that the transcripts levels were higher in *P. irregulare* (Fig. 1.6).

Some transcripts that were differentially expressed during the pathogenic phase (WP>MED) also had a corresponding protein secreted (labeled with a 'w' in Fig. 1.7). However, there was no correlation between transcript differential expression in WP and secretion of the coded protein. For example, while protease transcripts 47, 48, 188 and 220 in *P. irregulare* were differentially expressed in the WP treatment at 19° C surprisingly they were secreted in higher numbers while growing saprophytically (MED) (Figs. 1.7a and b); it appears that these transcribed protease transcripts are only translated or the corresponding proteins only secreted when no living host cells are present. Only one secreted protein was coded by a differentially expressed transcript in the two *P. iwayamai* isolates (Figs. 1.7c and d).

Many of the secreted proteases were present in both isolates of the same species; however a few were unique. For example, more different types of proteases were secreted in *P. irregulare* isolate 174b compared to 176, and more of these proteases were secreted in the saprophytic (MED) growth phase. The numbers of secreted proteins were similar in both isolates of *P. iwayamai* with higher proportions secreted in the pathogenic phase (WP). Some of the proteases included a signal peptide, with only one having a signal peptide+transmembrane region (labeled 's' and 'st' respectively in Fig. 1.7), in their corresponding transcripts, but many of them did not.

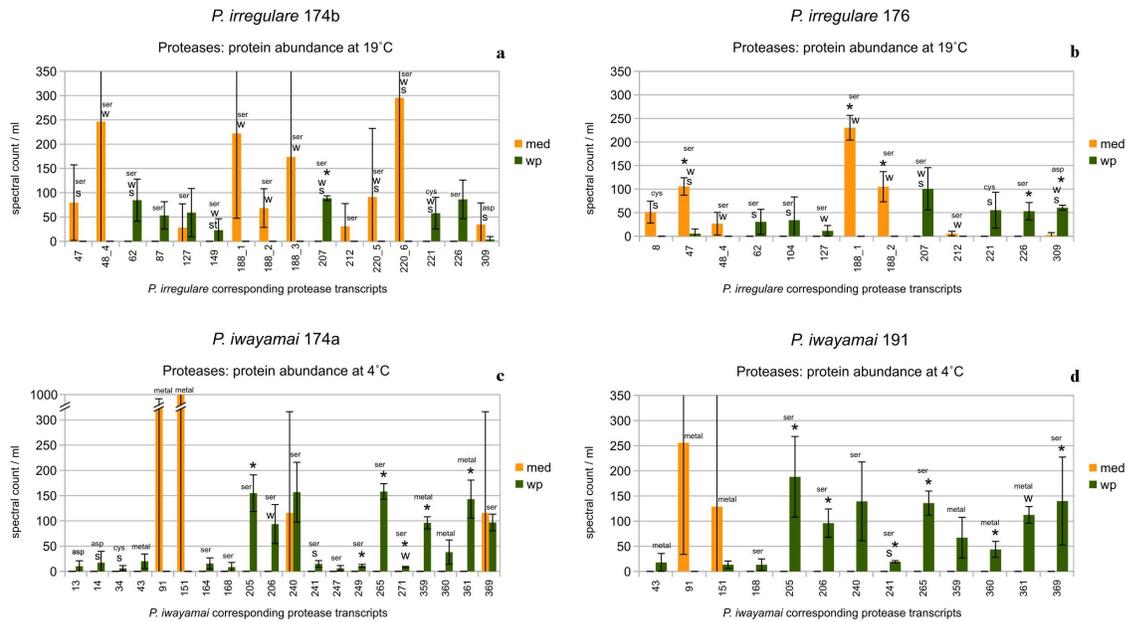


Figure 1.7. Proteases detected in the extracellular medium when *P. irregulare* (a-b) or *P. iwayamai* (c-d) were grown saprophytically (MED) or in the presence of plants (WP). Numbers on the x-axis represent the same protein in graphs **for isolates of the same species and correspond to numbers in the transcriptome graphs**. Those numbers with an additional number after an underscore correspond to different reading frames of the same transcript. Bars were labeled with a “w” if the corresponding transcripts were significantly more abundant in WP compared to MED. They were also labeled with a “s” or with “st” if a signal peptide or a signal peptide+transmembrane region were coded in the corresponding transcripts. Those \* labeled bars had significantly different spectral count values (BH adjusted p-value  $\leq 0.05$ ).

Secreted proteases have been described for many plant pathogens, including bacteria (Dow, Davies and Daniels, 1998), fungi (Di Pietro, González Roncero and Ruiz Roldán, 2009), and oomycetes (Weiland 2004, Davis *et al* 2005, Meijer *et al* 2014). Davis *et al* (2005) found that three *Pythium* species grew in a protein-rich medium but not in one containing only polysaccharides purified from plant cell walls, indicating these oomycetes were capable of degrading and utilizing proteins but not complex carbohydrates as a food source. We found that proteases were the major class of secreted proteins in both *Pythium* species and in both growing conditions (WP and MED). These proteases appear to be important by helping *Pythium* at least partially degrade cell walls during the infection process (Cassab 1998, Li *et al* 2010). Their

expression in both the WP and MED also suggests they play a role in degrading extracellular proteins for nutrition.

## GLYCOSIDASES

Glycoside hydrolases are important in degradation of plant cell walls (Walton 1994; Kubicek, Starr and Glass 2014). *Pythium* species, including the two studied here, contain a limited set of these enzymes that facilitate hyphal penetration, but they do not have an extensive arsenal of glycosidases as do other oomycetes (e.g. *Phytophthora spp.*) (Zerillo *et al* 2013) that allows for complete disintegration of the plant cell wall.

The total number of glycoside hydrolase transcripts was similar in *P. irregulare* and *P. iwayamai* (103 vs. 132 respectively, including isoforms) and to the gene candidates for each species (134 and 124 respectively) (Table 1.1) as described by Zerillo *et al* (2013). These hydrolases were further segregated into corresponding CAZy (GH) families (Lombard *et al* 2013) to obtain information of their probable enzymatic function (Table 1.3).

Table 1.3. Number of transcripts coding for glycoside hydrolases and their assignment to CAZy families.

CAZy family	Number of transcripts	
	<i>P. irregulare</i>	<i>P. iwayamai</i>
GH1	8	15
GH2	8	11
GH3	9	8
GH5	13	13
GH6	4	4
GH7	2	4
GH 9&16	7	10
GH13	1	2
GH15	2	1
GH17	10	12
GH18	2	1
GH28	0	1
GH30	6	7
GH31	5	6
GH32	1	1
GH35	1	1
GH37	0	1
GH38	1	0
GH47	2	4
GH63	0	2
GH72	5	4
GH81	1	1
GH98	0	1
Not assigned	1	5
Total	89	114

The number of transcripts for each GH family was again similar to the number of candidate genes reported by Zerillo *et al* (2013). The GH families 1, 3, 5, with high numbers of transcripts (Table 1.3), could be involved in plant cell wall cellulose degradation whereas hydrolases in GH families 2 and 9&16, also present in high numbers, can also act on other plant cell wall polysaccharides. The GH families 6 and 7, with smaller numbers of transcripts, are possibly directed to plant cell wall cellulose degradation (Zerillo *et al* 2013). As stated above, a number of transcripts corresponded to CAZy families 9 and 16. They had the Interpro domain IPR005200

described as “a family of beta-1,3(4)-glucanases” assigned, incorrectly we believe, to family 8 (<http://www.ebi.ac.uk/interpro/entry/IPR005200>). Members of that family only show endo- $\beta$ -1,3-glucanase activity, while members of both families 9 and 16 show  $\beta$ -1,3(4)-glucanase activity. One of the substrates for enzymes with that “dual” activity are cereal D-glucans. The high transcript numbers in GH17 likely are associated with  $\beta$ -1,3 and  $\beta$ -1,6 glucan metabolism in the *Pythium* cell wall (Zerillo *et al* 2013).

A few annotated genes found by Zerillo *et al* (2013) were absent in the transcriptomes. The transcripts of one predicted gene each in families GH78, 85, 89 and 123 in both *Pythium* species and four annotated genes each in *P. irregulare* and two and three genes respectively in *P. iwayamai* for families GH61 and 109 were absent. We also did not find transcripts for predicted genes in families GH28, 37, (two genes each) and 63 (one gene) in *P. irregulare* or for GH38 (two genes) in *P. iwayamai*. Some of these underrepresented glycosidase families in the *Pythium* genomes are involved in pectin degradation (GH28 and 78) or possibly in chitin degradation (GH85) (Zerillo *et al* 2013). The GH61 enzymes are monooxygenases now classified as AA9 ([http://www.cazypedia.org/index.php/Auxiliary\\_Activity\\_Family\\_9](http://www.cazypedia.org/index.php/Auxiliary_Activity_Family_9)) with cellulose as their main substrate. One possible explanation for the absence of transcripts is that these families of genes may not have been expressed under the conditions and at the time when our samples were taken. A unique transcript in family GH98 was not annotated by Zerillo *et al* (2013) but was found in *P. iwayamai*. This enzyme acts on arabinogalactans frequently found attached to proteins. The other notable difference was the presence of 8 and 11 different transcripts coding for GH 2 enzymes, with  $\beta$ -mannosidase activity, in *P. irregulare* and *P. iwayamai* respectively, whereas only two genes were annotated in their genomes. Polysaccharides containing mannose are found in seeds and in primary and secondary plant cell walls (Melton *et al* 2009).

As was the case with proteases, the overall expression of glycosidase transcripts was similar between isolates and species (Fig. 1.8). Most of the transcripts had a low fpkm value, with three (12, 88 and 115) showing a relatively high number primarily in *P. iwayamai* at 4°C WP. Two transcripts (12 and 88) were differentially expressed WP in isolate 174a, but only one (88) in isolate 191. They both coded for a GH family 1, and the corresponding protein was not secreted (Fig. 1.9e and g). They were probably involved in the maintenance/growth of the *Pythium* cell wall.

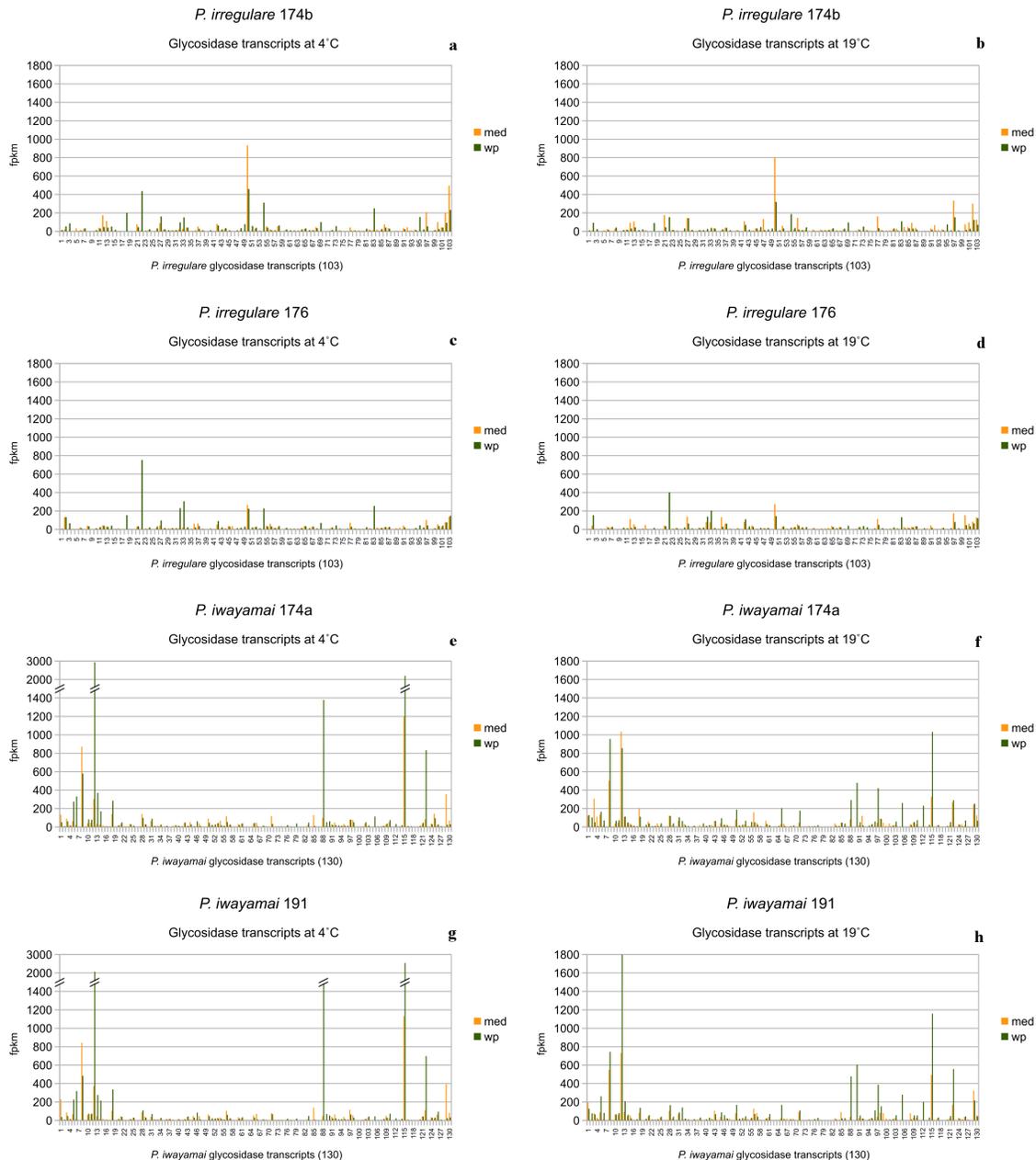


Fig. 1.8. Expression level (fpkm) of all glycosidase transcripts in *P. irregulare* (a-d) and *P. iwayamai* (e-h). Note the change in scale at the top of some the graphs. Numbers on the x-axis **between isolates of the same species** represent the same transcript.

A smaller subset of glycoside hydrolase transcripts exhibited differential expression between the infectious and saprophytic growth phases (Table 1.2, Figs. 1.9a-h), particularly in *P. irregulare* isolate 176. Overall, more differentially expressed transcripts occurred in the WP treatment in the two *Pythium* species and at both temperatures. Four GH 1 transcripts in *P. irregulare* and 18

in *P. iwayamai* were differentially expressed in the WP but not in the MED treatment. The same was observed for a smaller number of GH 9&16 transcripts (one in *P. irregulare* and 3 in *P. iwayamai*) and for one GH18 transcript (chitinase activity) in *P. iwayamai*. This suggests that these transcripts are involved in the response of *Pythium* to the presence of host plants. However, in contrast to the protease transcripts, only a few of the differentially expressed glycosidase transcripts in *P. irregulare* coded for secreted proteins and these did not include a signal peptide sequence (Figs. 1.9b and d). None of the secreted glycosidases were coded by differentially expressed transcripts in *P. iwayami* isolates (Figs. 1.9e and g). This suggests that these glycosidases could be secreted at a different time, or they could remain inside the *Pythium* hyphae.

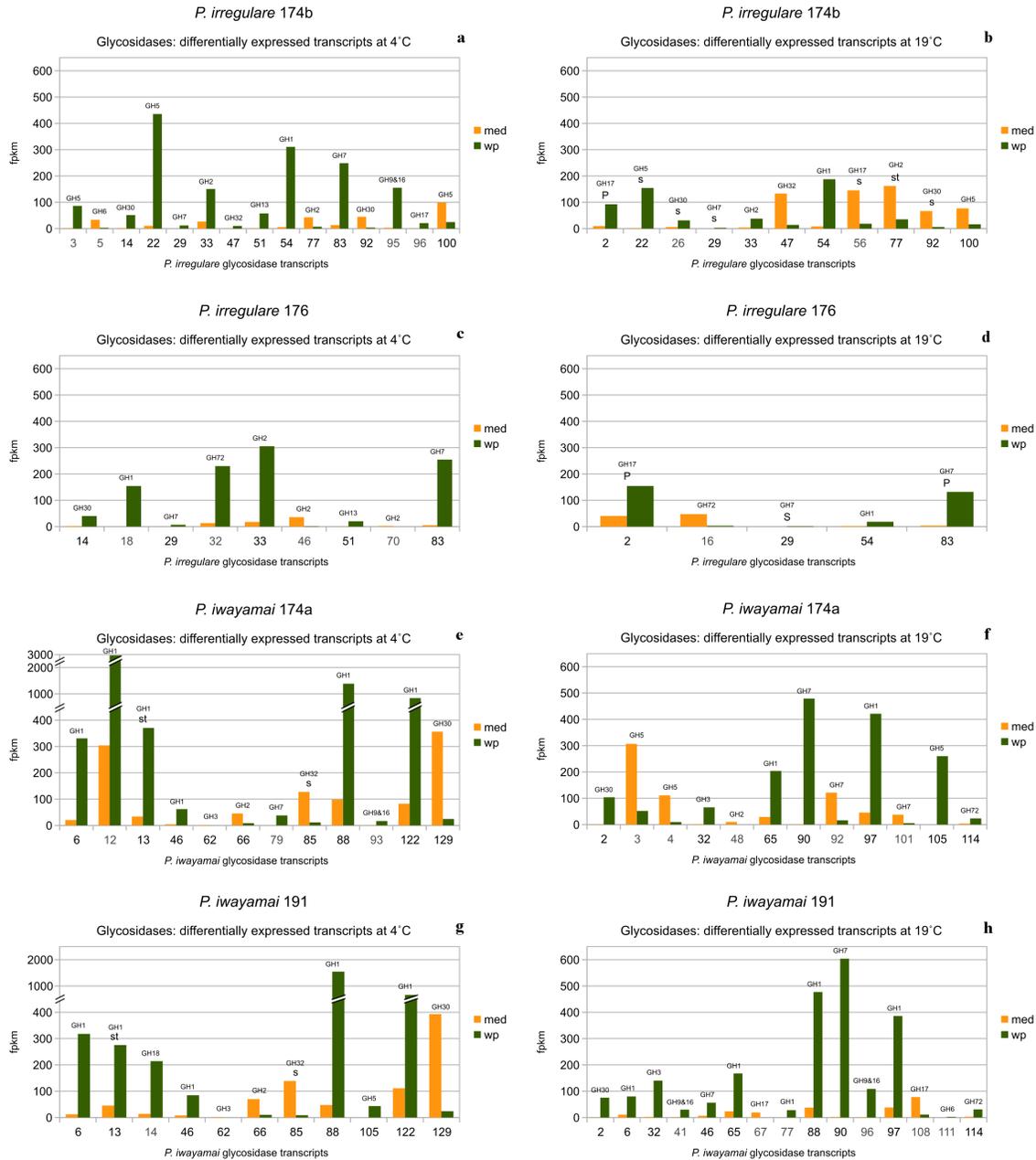


Figure 1.9. Differentially expressed transcripts ( $FDR \leq 0.01$ ) coding for **glycosidases** in *P. irregulare* (a-d) and *P. iwayamai* (e-h), at 4°C and 19°C. Bars were labeled with a “P” on top when the corresponding protein was detected in the proteomics analysis, performed at 19°C for *P. irregulare* and at 4°C for *P. iwayamai*. An “s” was added when the corresponding transcript contained a signal peptide sequence and “st” when they contained signal peptide and transmembrane domain sequences. Numbers on the x-axis represent the same transcripts for isolates of the same species.

Among all the transcripts that coded for secreted glycosidases, only a few of them included a signal peptide sequence (Fig. 1.10).

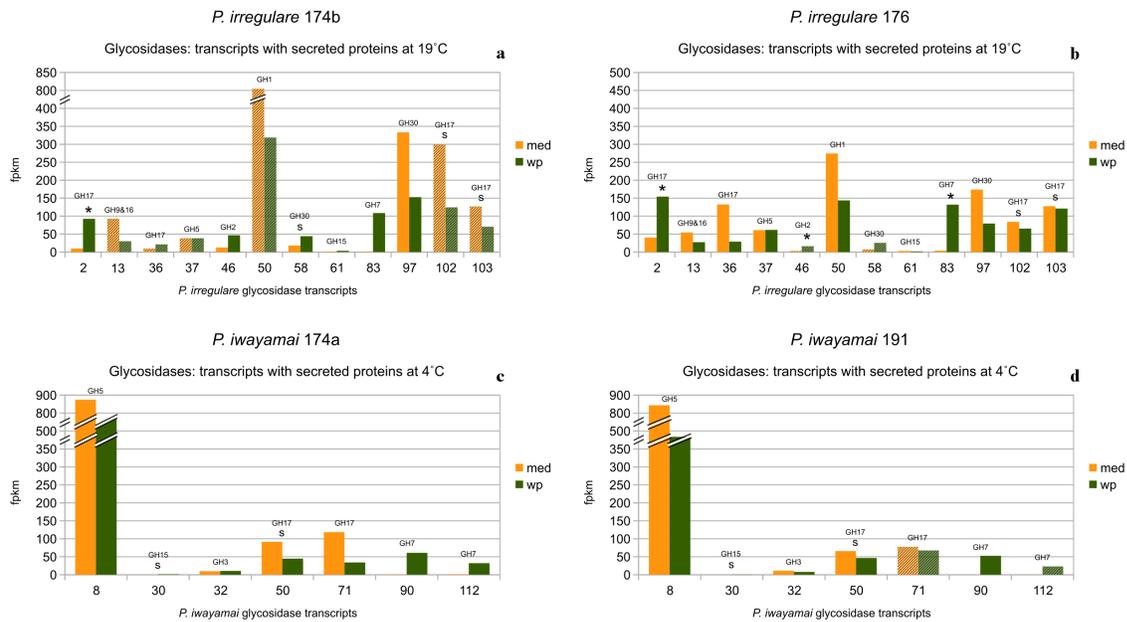


Figure 1.10. Transcripts coding for secreted glycosidases in *P. irregulare* (a,b) and *P. iwayamai* (c,d). Note the change in scale at the top of the graphs. Bars included a diagonal pattern when the corresponding protein was detected only in one isolate of the same species. An “s” was added when the corresponding transcript contained a signal peptide sequence. Those bars labeled with a \* had significantly different fpkm value ( $FDR \leq 0.01$ ). Numbers on the x-axis represent the same transcripts **for isolates of the same species**.

Most of the secreted glycosidase proteins in *P. irregulare* were only found in the MED treatment (Figs. 1.11a and b). These proteins corresponded to GH 17 (four), 30 (two), and 1, 2, 5, 15 ( $\alpha$ -amylase), and 9&16 (one each). Only one *P. irregulare* glycosidase was secreted in the WP treatment (number 83 in Figs. 1.11a-b) and it corresponded to GH 7 (directed against cellulose). The two *P. iwayamai* isolates exhibited a glycosidase protein secretion pattern that was distinct from *P. irregulare* (Figs. 1.11c and d) where a more diverse group of secreted glycosidases were observed in the WP compared to the MED treatment. Glycosidases 90 and 112 in WP belonged to GH 7. The other glycosidases in WP belonged in GH 3, 5 and 15. The two different glycosidases in MED belonged to GH 17.

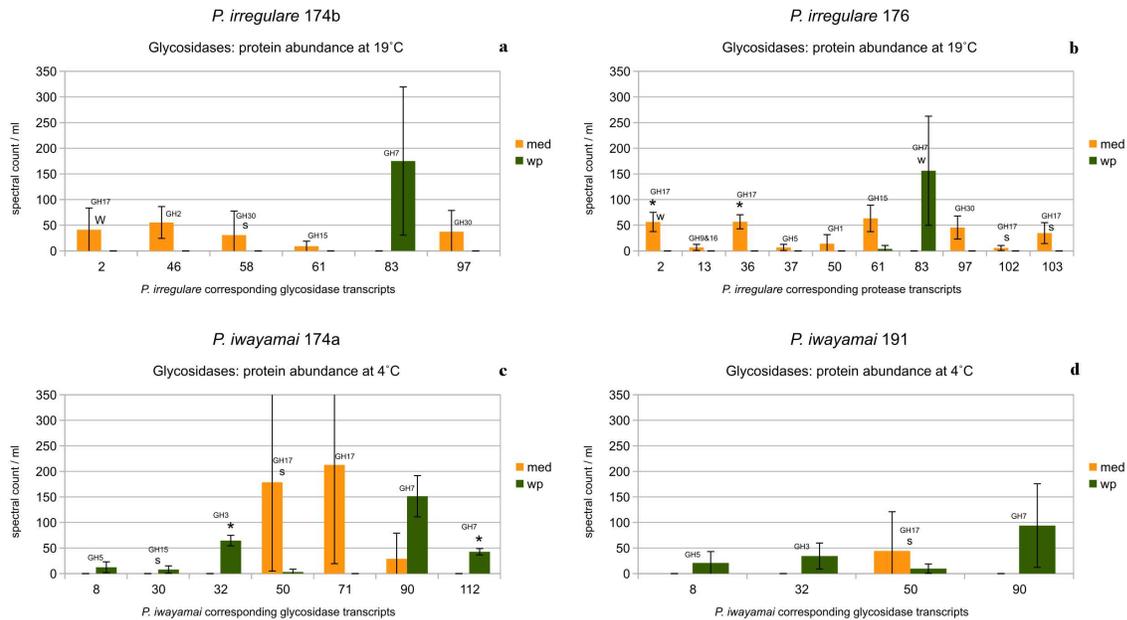


Figure 1.11. Glycosidase proteins detected in the extracellular medium when *P. irregulare* (a-b) or *P. iwayamai* (c-d) were grown saprophytically (MED) or in the presence of plants (WP). Numbers on the x-axis represent the same protein **between isolates of the same species and correspond to numbers in the transcriptome graphs**. Bars were labeled with a “w” if the corresponding transcripts were significantly more abundant in WP compared to MED. They were also labeled with an “s” if a signal peptide was coded in the corresponding transcripts. Those \* labeled bars had significantly different spectral count values (BH adjusted p-value  $\leq 0.05$ ).

Our results support many of the conclusions of Zerillo *et al* (2013) concerning the role of glycosidases in plant infection by *Pythium* species. They annotated certain genes coding for GH families (GH 1, 3, 5, 6 and 7) that could be secreted and suggested that they could act on extracellular (i.e. plant) cellulose. We observed in both *P. irregulare* and *P. iwayamai* that glycosidases belonging to GH 7 were secreted, and likely helping to degrade plant cell walls, only when they were in the presence of plants. Similarly, *P. iwayamai* glycosidases in GH 3 were secreted only in the presence of plants. The role of glycosidases in GH 5 in plant cellulose degradation is not as clear because they were secreted in the WP treatment in *P. iwayamai*, but only in the MED treatments in *P. irregulare* (Fig. 1.11). Curiously glycosidases belonging to GH 6 were not detected among secreted proteins in either species, and those belonging to GH 1 were

found only in *P. irregulare* isolate 176, in MED. The only glycosidases consistently found in MED treatments in both species belonged to GH 17 (Fig. 1.11), and were probably associated with maintenance of the *Pythium* cell wall (Zerillo *et al* 2013).

## ELICITINS

Elicitins and elicitin-like proteins were first described in *Phytophthora* and later in *Pythium* species (Ponchet *et al* 1999). They function as lipid carriers, mainly for sterols (Blein *et al* 2002), and can cause a hypersensitive response and necrosis in a small group of plants that have the appropriate receptors to detect them (Vlesshouwer *et al* 2006). It is not clear if they have another role in the infection process in other plant species (Blein *et al* 2002). Recently, elicitins have been used to try to increase the resistance of potato to the late blight disease (Du *et al* 2015). In our studies, both *P. irregulare* and *P. iwayamai* exhibited very high elicitin transcript expression levels in both WP and MED treatments (Fig. 1.12, Tables S2-17). Their overall expression profiles were similar within each species, with some transcripts having a small fpkm value while others having a very high number, especially in *P. iwayamai* (Fig. 1.12).

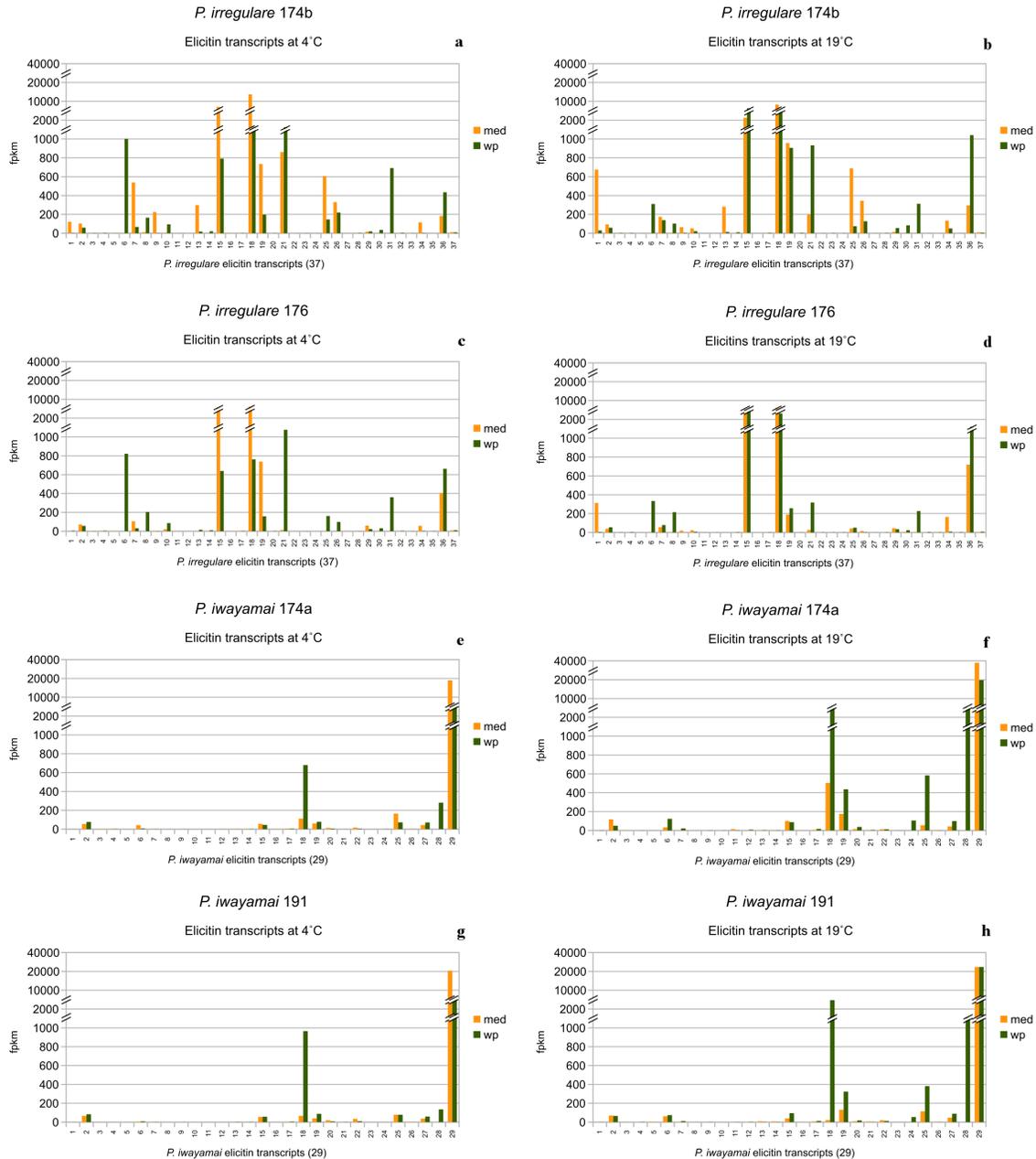


Figure. 1.12. Expression level (fpkm) of all elicitor transcripts in *P. irregulare* (a-d) and *P. iwayamai* (e-h). Numbers in the x-axis in graphs of the same species represent the same transcript.

Expression levels for three elicitor transcripts in both species were high (>1,000 fpkm), with one higher than 10,000 fpkm in *P. iwayamai*. Twelve transcripts were differentially expressed in *P. irregulare* isolate 174b, five in isolate 176, and only three in *P. iwayamai* (Fig. 1.13). There was

no clear trend in transcript expression; some were overexpressed in MED whereas others were differentially expressed in WP. As with protease and glycosidase transcripts, transcripts associated with secreted elicitors often did not contain a signal peptide and one included a transmembrane region sequence in *P. irregulare* (Fig. 1.14)

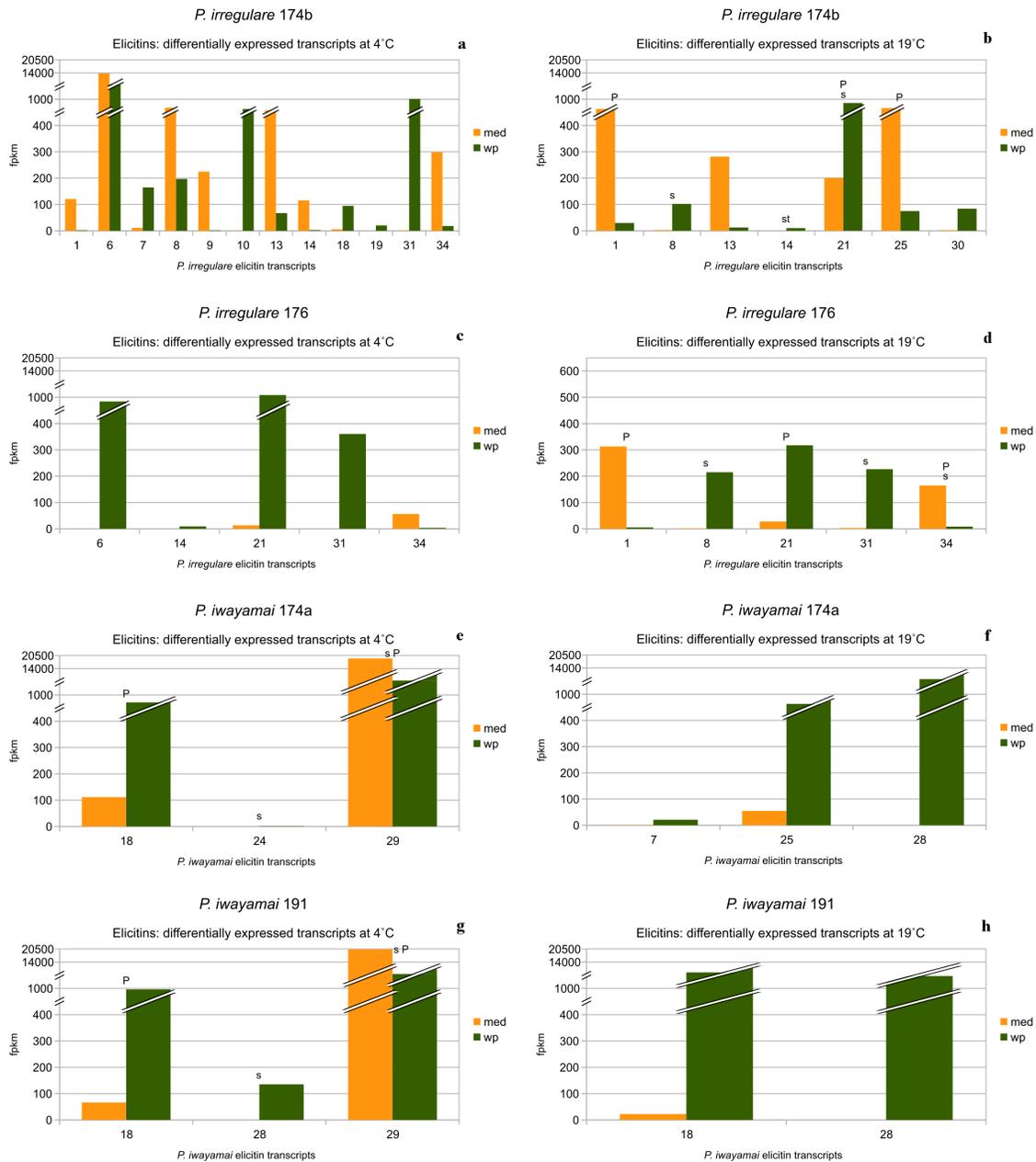


Figure 1.13. Differentially expressed transcripts (FDR<0.01) coding for **elicitors** in *P. irregulare* (a-d) and *P. iwayamai* (e-h), at 4°C and 19°C. Bars were labeled with a “P” on top when the

corresponding protein was detected in the proteomics analysis, performed at 19°C for *P. irregulare* and at 4°C for *P. iwayamai*. An “s” was placed above bars when the corresponding transcript contained a signal peptide sequence and “st” when they contained signal peptide and transmembrane domain sequences. Numbers in the x-axis in graphs of the same species represent the same transcript.

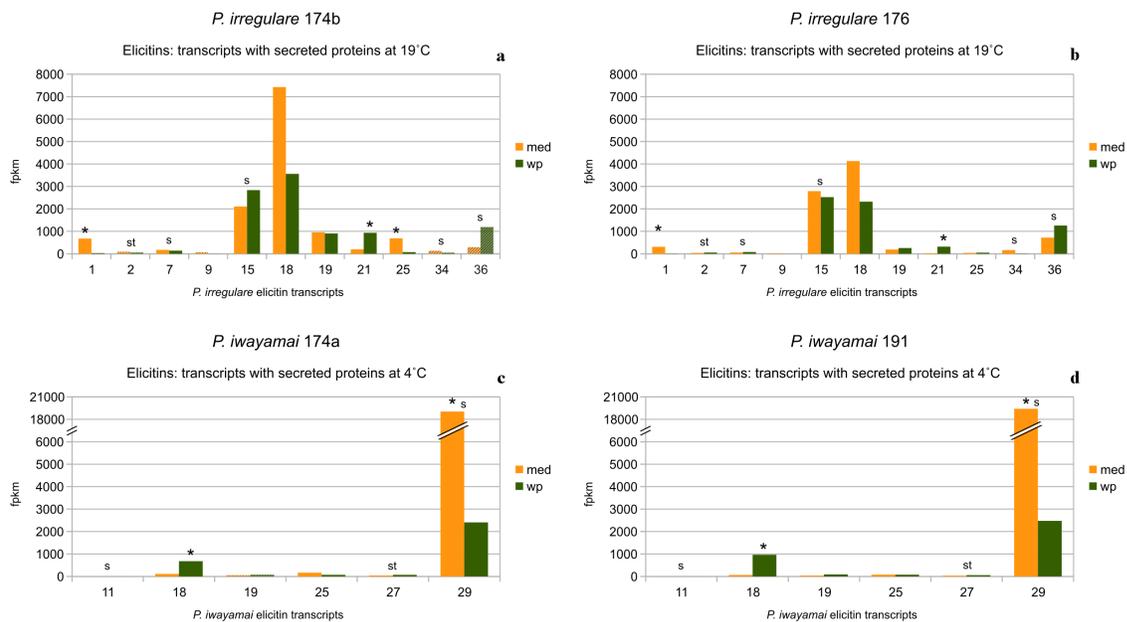


Figure 1.14. Transcripts coding for secreted elicitors in *P. irregulare* (a,b) and *P. iwayamai* (c,d). Note the change in scale at the top of the graphs. Bars included a diagonal pattern when the corresponding protein was detected only in one isolate of the same species. An “s” was added when the corresponding transcript contained a signal peptide sequence and “st” when they contained signal peptide and transmembrane domain sequences. Those bars labeled with a \* had significantly different fPKM number ( $FDR \leq 0.01$ ). In graphs of the same species numbers in the x-axis were assigned to the same transcripts.

Elicitins were also the most abundant of secreted proteins in the four *Pythium* isolates (Fig. 1.14, Tables S18-21). Secretion profiles between isolates of the same species, but not between species, were similar. In *P. irregulare* a more diverse set of elicitors, and with higher spectral counts, was detected in the MED compared to WP treatments (Figs. 1.15a and b, Tables S18 and S19). Some of the proteins corresponded with differentially expressed transcripts (Nos. 1, 21, 34) but not

those with the highest spectral counts (Nos. 15, 18, 19). In *P. iwayamai* only one elicitin (No. 29) was detected in MED but it had a very high spectral count (Figs. 1.15c and d). This secreted elicitin also corresponded to a differentially expressed transcript in the MED treatment. Similarly, one secreted elicitin (18), but with a much lower spectral count, exhibited a differentially expressed transcript in the WP treatment.

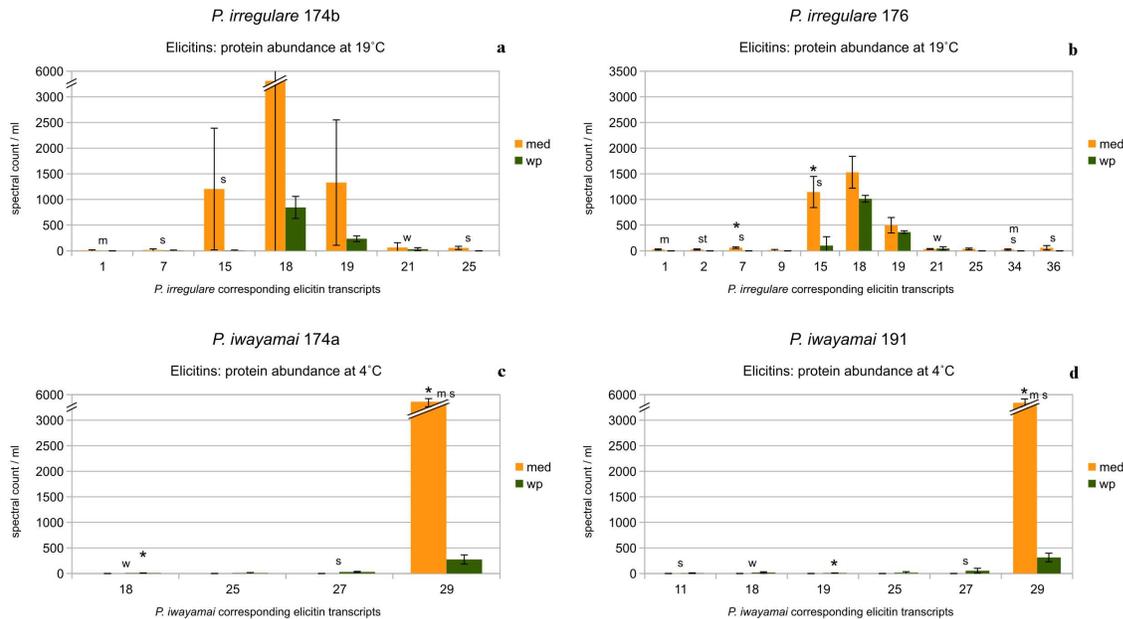


Figure 1.15. Elicitin proteins detected in the extracellular medium when *P. irregulare* (a-b) or *P. iwayamai* (c-d) were grown in medium only (MED) or in the medium with plants (WP). Numbers on the x-axis represent the same protein in graphs of isolates of the same species. Bars were labeled with a “w” or a “m” if the corresponding transcripts were significantly overexpressed with-plants or in-medium respectively. They were also labeled with a “s” or with “st” if a signal peptide or a signal peptide+transmembrane region were coded in the corresponding transcripts. \* labeled bars have significantly different spectral count values (BH adjusted p-value  $\leq 0.05$ ).

In our study, both elicitins with the higher spectral counts in *P. irregulare* (No. 18, Table S7 and S9) and *P. iwayamai* (No. 29, Table S14 and S16) were similar to sylvaticin, an elicitin with unknown function produced by *P. sylvaticum* (Lascombe *et al* 2007). As previously stated, the role of elicitins in the infection process by oomycetes is not clear (Akino, Takemoto and Hosaka 2014). They have been shown to be important for infection of woody plants by some

*Phytophthora* species (Oßwald *et al* 2014). However, they have also been detected in high abundance in liquid cultures of *Phytophthora* in the absence of plants (Ponchet *et al* 1999, Meijer *et al* 2014) and recently in the predicted secretome of the saprophytic oomycete *Thraustotheca clavata* (Misner *et al* 2014). This suggests a more saprophytic-directed function as lipid carriers (Blein *et al* 2002). They may also act as pathogen-associated molecular patterns that elicit a hypersensitive response and diminish the pathogen virulence, but only in those plants that are able to detect them (Colas *et al* 2001).

### OTHER SECRETED PROTEINS

We found evidence (presence in at least two out of three replicates) of 112 and 77 different proteins secreted by *P. irregulare* and by *P. iwayamai* isolates, respectively. As previously discussed the most abundant secreted proteins were proteases, followed by glycoside hydrolases and elicitors (Fig. 1.16, Tables S18-S21).

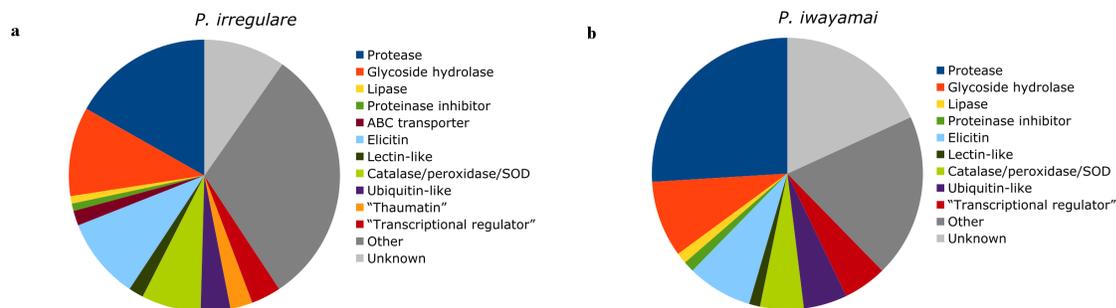


Fig. 1.16. Pie chart of proteins categories detected among secreted proteins in *P. irregulare* (a) and *P. iwayamai* (b). The “Other” category comprises proteins with a variety of functions not considered pathogenicity-related.

Other pathogenicity-related proteins observed in the transcriptome were not secreted or were found just in one out of three replicates. But other proteins, not considered pathogenicity-related, were consistently secreted in one or both species (Fig.1.16, Tables S18-23). Ubiquitin (both species) and thaumatin-like proteins (only in *P. irregulare*) were primarily secreted in the MED treatments. Ubiquitin was reported to be secreted by *Phytophthora infestans* growing in a liquid

medium (Huet, Sallé-Tourne and Pernollet 1994), but its role outside the cell is unknown.

Thaumatococcus-like proteins have long been considered plant pathogenesis-related proteins (Vigers *et al* 1992) but they also have numerous other functions in a number of organisms (Liu, Sturrock and Ekramoddoullah 2010) including antifreeze activity in winter rye (Hon *et al* 1995).

However, it is unlikely that they were associated with antifreeze activity in our experiments because they were secreted by *P. irregulare* at 19°C but not by *P. iwayamai* at 4°C. Endo- $\beta$ -1,3-glucanase activity also has been reported for thaumatococcus-like proteins (Grenier, Potvin and Asselin 2000) and they could be acting on modification of the *P. irregulare* cell wall.

Proteins that contain the domain PHA03307, described in a herpes virus transcriptional regulator ICP4, were also found in both *Pythium* species; primarily in the MED treatment in *P. irregulare* but mostly in the WP treatment in *P. iwayamai* (Tables S18-S21). The function of these proteins is unknown; interestingly, this group of protein induces ubiquitin genes in virus-infected cells (Latchman, Estridge and Kemp, 1987).

A number of other non pathogenicity-related proteins were secreted (Tables S22 and S23).

Transaldolase type 1, part of the pentose phosphate pathway, was secreted only in MED in both *Pythium* species; it is probably acting on glucose, but we have no explanation why it occurred outside the *Pythium* cells. Some secreted proteins with an EGF-like, cysteine-rich domain, have been reported in the *Phytophthora infestans* secretome, and possibly could be novel effectors (Raffaele *et al* 2010). These proteins were also secreted in our *Pythium* isolates, but primarily in the MED treatment and thus not supporting the idea they acting as effectors. Several other proteins including adenosine deaminase, glutathione-S-transferase were secreted in at least some of the isolates.

There were a number of ‘unknown’ proteins, some of which were secreted in all isolates, which we were unable to assign a clear function; (Tables S22 and S23). In *P. irregulare* they were mostly found in the MED treatments, whereas in *P. iwayamai* some very similar proteins as well as other different types of proteins were secreted mainly in the WP treatments. A protein of unknown function (DUF3984 family) was found in *P. irregulare* in the MED treatment, but two proteins of the same family and another from a different one (DUF1645) were found mainly with-plants in *P. iwayamai*. Some of the proteins found only or mostly in the WP treatments may be good candidates to be studied as possible novel *Pythium* effectors.

We did not observe differential transcript expression (4°C versus 19°C) that could be coding for antifreeze proteins in *P. iwayamai* (data not shown), and there was also no evidence of secreted antifreeze proteins (Tables S20 and S21). We did not grow *P. iwayamai* at the lower temperature (0°C) that is normally associated with disease development on grasses under snow cover (Blunt, Koski and Tisserat 2013), and it is possible that these proteins were not expressed in our low temperature (4°C) treatments. Nevertheless, our results corroborate those of Hoshino, Xiao and Tkachenko (2009) who found no antifreeze protein activity in *P. iwayamai*. Recently *P. iwayamai* and *P. paddicum*, another snow mold oomycete, were shown to survive exposure to -20°C when they were inside, but not outside, plant tissue (Murakami *et al* 2015) suggesting that they use plant compounds as cryoprotectants.

#### GENERAL COMMENTS ABOUT SECRETED PROTEINS

An unexpected result was the overall greater diversity of secreted proteins in the same class (i.e. different proteases, glycosidases, etc.) and/or with a higher spectral count when isolates were grown in the MED versus the WP treatment (Tables S18-S21, Figs. 1.7, 1.11, 1.15). As was the case with transcript expression, it appeared that a different set of proteins within each protein

class (same biochemical function) were secreted in each of the two treatments; one set of proteins was more abundant when *Pythium* was growing in the presence of plants whereas the other set was more abundant in their absence (MED) regardless of temperature. This could have resulted in lower than expected secretion levels for specific proteins under certain conditions. Severino *et al* (2014) observed that a number of proteins including NPL-effectors, Avr1b\_1, glycoside hydrolases and pectate lyases, were secreted in lower amounts when *Phytophthora plurivola* mycelium was exposed to root exudates of a highly susceptible host compared to mycelium grown in the absence of root exudates (Severino *et al* 2014).

What is the function of all those proteins that are released to the media by these *Pythium* species in the absence of a host? It is likely that the MED treatment, containing available carbon and nitrogen sources, mimics dying plant cells leaking nutrients and scavenger proteins are secreted by *Pythium* mycelium to take advantage of those nutrient sources. For example, specific proteases (Fig. 1.7), glycosidases (Fig.1.11) and elicitors (Fig. 1.15) in the MED treatment might be associated with utilization of plant cytoplasm-released proteins, glycosides and lipids as a food source. In contrast secreted enzymes in the WP treatment during the initial phase of plant infection and colonization (as was sampled in these experiments) could be directed more towards plant cell wall and membrane degradation. This hypothesis is in part supported by the almost exclusive presence of enzymes belonging to the GH 7 family of glycosidases, and that only function as cellulases (Zerillo *et al* 2013), in the WP treatment. If it is true that MED treatment resembles late stages of the *Pythium* infection process when nutrients are being released by dying cells, the same proteins should also be secreted in advance stages of disease in the WP treatments. However, we did not collect samples at these dates to confirm this hypothesis.

## PLANT PROTEINS

It was not our objective to study the transcripts and secreted proteins associated with the diseased creeping bentgrass. But it is worth mentioning that only one secreted plant protein was consistently detected during infection by both *Pythium* species, with an Interpro domain IPR000877 corresponding to a family of serine protease inhibitors similar to that found in barley (Stevens *et al* 1996). Serine proteases were the main class of proteases secreted by both *Pythium* species and it is likely that secretion of a serine protease inhibitor was a plant defense response. Nevertheless, a number of different serine proteases as well as some cysteine proteases and metalloproteases were also secreted by *P. irregulare* and *P. iwayamai* (Fig. 1.7) making a defense based solely on a single serine protease inhibitor ineffective.

A few ubiquitin-like proteins and those with unknown function were detected in some treatments (results not shown), but their overall numbers were low. It is likely that plant cells were still intact during the initial infection period that we sampled, and that a large number of plant proteins had not yet begun to leak from the diseased tissue. It is also possible that most secreted plant proteins, with the exception of the one protease inhibitor, were rapidly degraded by *Pythium* proteases. Alternatively, some plant proteins may have been secreted below the level of detection by the MS system that we used.

## CONCLUSIONS

*Pythium irregulare* and *P. iwayamai*, in agreement with their necrotrophic life style, did not express effectors (i.e. RXLR/Q effectors) that have been described in more complex hemibiotroph *Phytophthora* (Adhikari *et al* 2013). Instead, they primarily expressed and secreted a repertoire of proteases and glycosidases that partially degraded plant cell walls. They did not

express, or at least secrete, cutinases, pectin lyases and other enzymes necessary to completely macerate plant cell walls. Many proteins with similar biochemical function were secreted in the absence of plants, indicating that not all “pathogenicity-related proteins” were actually involved in pathogenicity. This is especially true for elicitors that were transcribed and secreted in great abundance when growing saprophytically.

*Pythium irregulare* and *Pythium iwayamai* belong to two distinct but closely related phylogenetic clades that are characterized by having globose sporangia (Lévesque and De Cock 2004). Overall, their transcriptomes and secretomes were similar during saprophytic and parasitic growth as well as growth at the two temperatures tested. This suggests that these species primarily rely on a core set of enzymes that are expressed regardless of growing conditions, and that other *Pythium* species with globose sporangia may behave similarly.

We did not find any correlation between the presence of a signal peptide and the absence of transmembrane region in transcripts to the expected presence of secreted proteins (Raffaele *et al* 2010, Meiken and Min 2012). Also, there was no correlation between transcript levels and the corresponding levels of secreted proteins. There are limitations in the computer-based prediction of secreted proteins, including those without a signal peptide or those with heterogeneity in the signal peptide sequence when present (Severino *et al* 2014, Meijer *et al* 2014). Another intrinsic limitation of the prediction approach is the difficulty to determine, once a set of proteins with signal peptide has been identified, which ones will be secreted, or at what stage of the infection process they will be secreted. Our results seem to confirm that, and point to the usefulness of the proteomics analysis to confirm secretion of candidate proteins based on predictive softwares.

## SUPPLEMENTAL FILES, FIGURES AND TABLES

Fig.S1

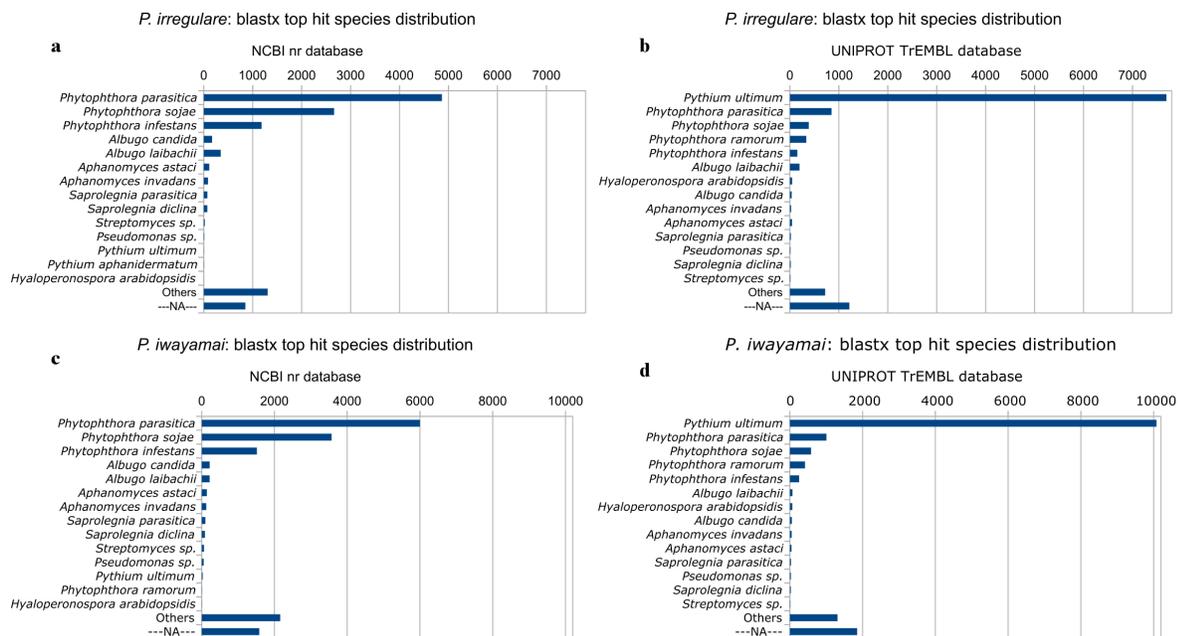


Figure S1. Blastx top hit species distribution of *P. irregulare* (a and b) and *P. iwayamai* (c and d) transcripts. *Phytophthora parasitica* predominates when the search was done against the NCBI nr database; *Pythium ultimum* had the biggest number of hits when using the Uniprot TrEMBL database.

Fig.S2

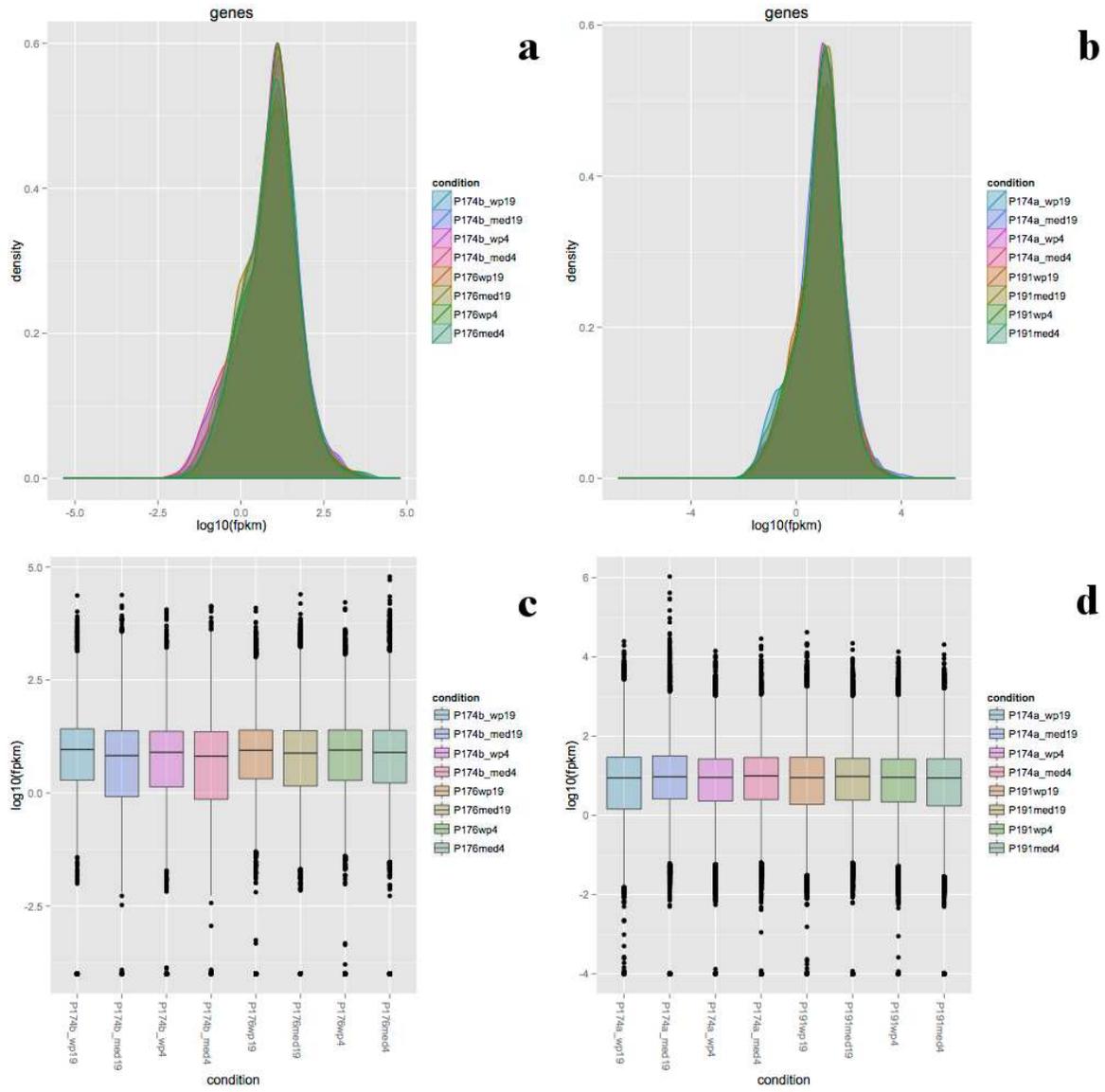


Figure S2. Density and boxplots of all transcripts fpkm.

Fig.S3

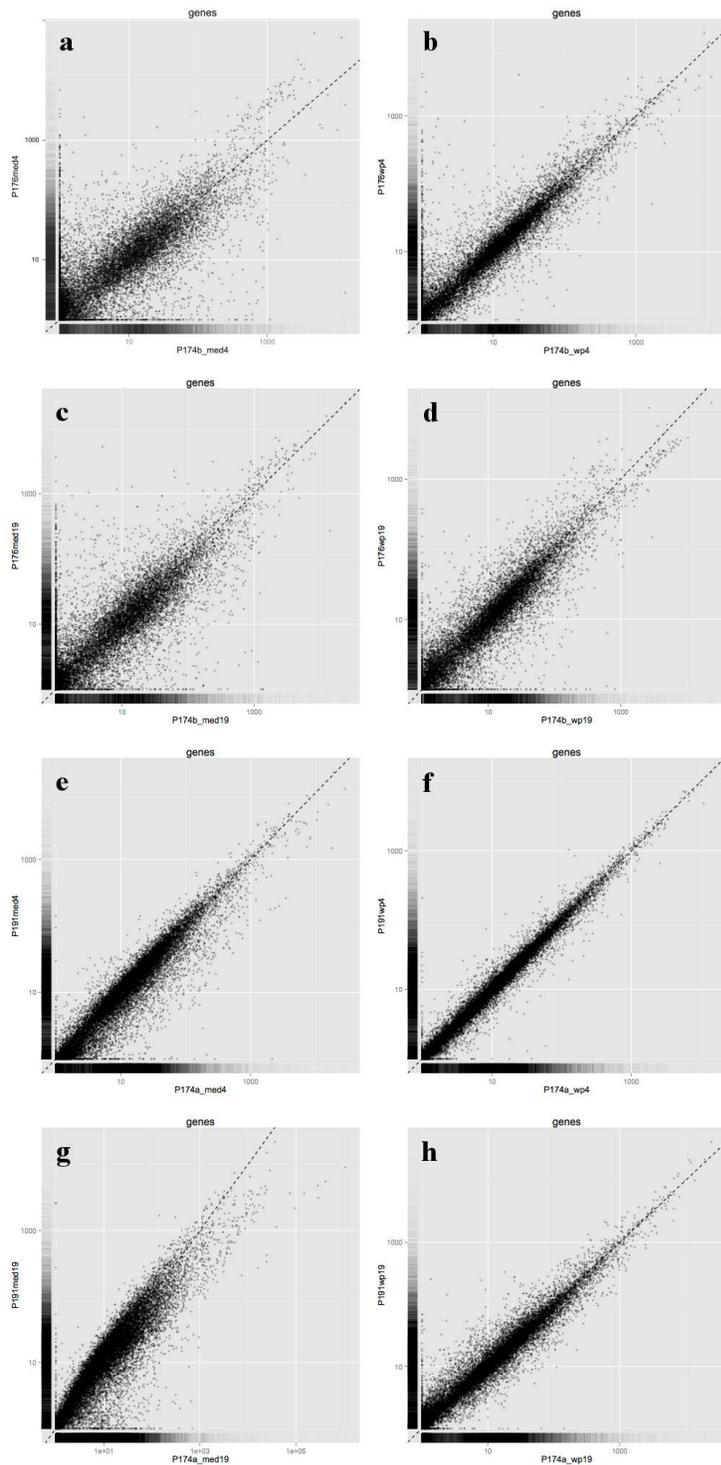


Figure S3. Scatter plots of all transcripts fpkm of the two isolates of each species, when they were with-plants (wp) or in-medium (med) at the two different temperatures tested (4° and 19°C). Plots (a) to (d) correspond to *P. irregulare* isolates; (e) to (h) correspond to *P. iwayamai* isolates. Less dispersion is observed when comparing both isolates with-plants.

Fig.S4

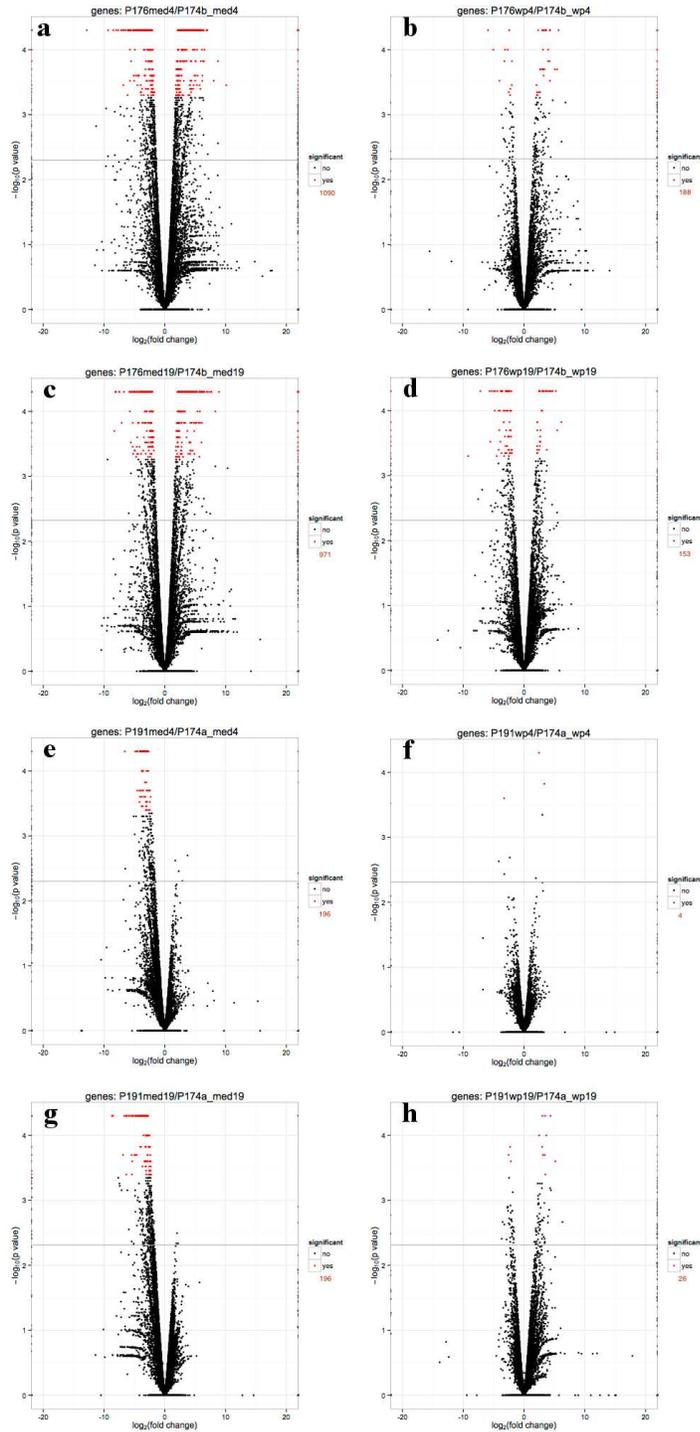


Figure S4. Volcano plots showing the amount of transcripts with significantly different level of expression (red dots,  $FDR \leq 0.01$ ). A horizontal line was drawn at the  $-\log_{10}(p \text{ value})$  corresponding to a FDR of 0.05. The actual number of red dots is shown at the right. (a)-(d): *P. irregulare* isolates; (e)-(h): *P. iwayamai* isolates.

ST1\_RNASeq

Supplemental Table 1. Number of total reads per each individual sample, average of two replicates (rep) for each condition and temperature, and coverage. “WP”: with-plants; “MED”: in-medium.

<i>Pythium</i> sp	Isolate	Treatment	Total reads		Average reads	Transcriptome coverage*
			Rep 1	Rep 2		
<i>P. irregulare</i>	174b	WP 19°C	29606357	30453051	30029704	76.6X
<i>P. irregulare</i>	176	WP 19°C	35805864	28579935	32192900	82.2X
<i>P. irregulare</i>	174b	MED 19°C	32722424	27615439	30168932	77.0X
<i>P. irregulare</i>	176	MED 19°C	44358488	20095394	32226941	82.3X
<i>P. irregulare</i>	174b	WP 4°C	30660819	28202042	29431431	75.1X
<i>P. irregulare</i>	176	WP 4°C	31772625	30972209	31372417	80.1X
<i>P. irregulare</i>	174b	MED 4°C	30290218	28199778	29244998	74.6X
<i>P. irregulare</i>	176	MED 4°C	28663759	38975106	33819433	86.3X
<i>P. iwayamai</i>	174a	WP 19°C	29328809	28999708	29164259	74.4X
<i>P. iwayamai</i>	191	WP 19°C	36907063	31127387	34017225	86.8X
<i>P. iwayamai</i>	174a	MED 19°C	29694431	36395101	33044766	81.2X
<i>P. iwayamai</i>	191	MED 19°C	37868663	38203148	38035906	93.5X
<i>P. iwayamai</i>	174a	WP 4°C	29831062	32311053	31071058	79.3X
<i>P. iwayamai</i>	191	WP 4°C	35266362	29769426	32517894	83.0X
<i>P. iwayamai</i>	174a	MED 4°C	28758325	29555202	29156764	71.7X
<i>P. iwayamai</i>	191	MED 4°C	23323160	26867592	25095376	61.7X

\*Total transcriptome size were for *P. iwayamai* 30,105,478 bp and 28,993,015 bp for *P. irregulare*. Average read length was 74 bp.

The content of the next files is too big to be included here. They are also included in the submission of this Dissertation to ProQuest/UMI

SFile1\_Pir\_transcripts\_seqs.fasta

Sfile2\_Pir\_transcripts\_seqs.fasta

ST2\_P174b\_med4\_P176med4.csv  
ST3\_P174b\_wp4\_P176wp4.csv  
ST4\_P174b\_med19\_P176med19.csv  
ST5\_P174b\_wp19\_P176wp19.csv  
ST6\_P174b\_wp4\_med4.csv  
ST7\_P174b\_wp19\_med19.csv  
ST8\_P176\_wp4\_med4.csv  
ST9\_P176wp19\_med19.csv  
ST10\_P174a\_med4\_P191med4.csv  
ST11\_P174a\_wp4\_P191wp4.csv  
ST12\_P174a\_med19\_P191med19.csv  
ST13\_P174a\_wp19\_P191wp19.csv  
ST14\_P174a\_wp4\_med4.csv  
ST15\_P174a\_wp19\_med19.csv  
ST16\_P191\_wp4\_med4.csv  
ST17\_P191\_wp19\_med19.csv  
ST18\_P174b\_proteins\_table.csv  
ST19\_P176\_proteins\_table.csv  
ST20\_P174a\_proteins\_table.csv

ST21\_P191\_proteins\_table.csv

ST22\_Pir\_other\_proteins.csv

ST23\_Piw\_other\_proteins.csv

## REFERENCES

- Adhikari, B.N., *et al.* 2013. Comparative genomics reveals insight into virulence strategies of plant pathogenic oomycetes. *PloS ONE* 8(10):e75072.
- Akino, S., Takemoto, D., and Hosake, K. 2014. *Phytophthora infestans*: a review of past and current studies on potato late blight. *Journal of General Plant Pathology* 80:24-37.
- Antelmann, H. *et al.* 2001. A proteomic view of genome-based signal peptide predictions. *Genome Research* 11(9):1484-1502.
- Bala, K., *et al.* 2010. Taxonomy, DNA barcoding and phylogeny of three new species of *Pythium* from Canada. *Personia* 25:22-31.
- Bebber, D.P. and Gurr, S.J. 2015. Crop-destroying fungal and oomycete pathogens challenge food security. *Fungal Genetics and Biology* 74:62-64.
- Bendtsen, J.D. *et al.* 2004. Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* 340:783-795.
- Blein, J., *et al.* 2002. From elicitors to lipid-transfer proteins: a new insight in cell signalling involved in plant defense mechanisms. *Trends in Plant Science* 7(7):293-296.
- Blunt, T., Koski T., and Tisserat, N. 2013. Effect of snow removal on *Typhula* blight development at high elevation golf courses in Colorado. Online. *Plant Health Progress* doi:10.1094/PHP-2013-0821-02-RS.
- Bridge, P.D., Newsham, K.K. and Denton, G.J. 2008. Snow mould caused by a *Pythium sp.*: a potential vascular plant pathogen in the maritime Antarctic. *Plant Pathology* 57:1066-1072.
- Cantrell, H.F., and Dowler, W.M. 1971. Effects of temperature and pH on growth and composition of *Pythium irregulare* and *Pythium vexans*. *Mycologia* 63(1):31-37.
- Cassab, G.I. 1998. Plant cell wall proteins. *Annual Review of Plant Physiology* 49:281-309.
- Chambers, M.C. *et al.* 2012. A cross-platform for mass spectrometry and proteomics. *Nature Biotechnology* 30:918-920.
- Colas, V., *et al.* 2001. Elicitor genes expressed in vitro by certain tobacco isolates of *Phytophthora parasitica* are down regulated during compatible interactions. *Molecular Plant-Microbe interactions* 14(3):326-335.
- Craig, R. and Beavis, R.C. 2004. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 20:1466-1467.

- Davis, D.J., *et al.* 2005. Relationship between temperature optima and secreted protease activities of three *Pythium* species and pathogenicity toward plant and animal host. *Mycological Research* 110:96-103.
- Di Pietro, A., González Roncero, M.I., and Ruiz Roldán, M.C. 2009. From tools of survival to weapons of destruction: the role of cell wall-degrading enzymes in plant infection, in Deising, H. (Ed). *The Mycota V.* 2<sup>nd</sup> ed. Springer-Verlag Berlin Heidelberg.
- Dow, J.M., Davies, H.A., and Daniels, M.J. 1998. A metalloprotease from *Xanthomonas campestris* that specifically degrades proline/hydroxyproline rich glycoproteins of the plant extracellular matrix. *Molecular Plant-Microbe Interactions* 11(11):1085-1093.
- Du, J., *et al.* 2015. Elicitin recognition confers enhanced resistance to *Phytophthora infestans* in potato. *Nature Plants* 1:15034.
- Ellis, J.G., *et al.* 2009. Recent progress in discovery and functional analysis of effector proteins of fungal and oomycete plant pathogens. *Current Opinion in Plant Biology* 12:399-405.
- Feller, G. 2013. Psychrophilic enzymes: from folding to function and biotechnology. *Scientifica* 2013: Article ID 512840. doi:10.1155/2013/512840.
- Fisher, M.C., *et al.* 2012. Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484:186-194.
- Geer, L.Y., *et al.* 2004. Open mass spectrometry search algorithm. *Journal of Proteome Research* 3:958-964.
- Gokce, E., *et al.* 2011. Evaluation of normalization methods on GeLC-MS/MS label-free spectral counting data to correct for variation during proteomic workflows. *Journal of the American Society for Mass Spectrometry* 22:2199-2208.
- Götz, S., *et al.* 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* 36(10):3420-3435.
- Grenier, J., Potvin, C., and Asselin, A. 2000. Some fungi express  $\beta$ -1,31-glucanases similar to thaumatin-like proteins. *Mycologia* 92(5):841-848.
- Hass, B.J., *et al.* 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461:383-398.
- Hendrix, F.F., and Campbell, A. 1973. Pythiums as plant pathogens. *Annual Review of Phytopathology* 11:77-98.
- Hon, W., *et al.* 1995. Antifreeze proteins in winter rye are similar to pathogenesis-related proteins. *Plant Physiology* 109:879-889.

Hoshino, T., Xiao, N., and Tkachenko, O.B. 2009. Cold adaptation in the phytopathogenic fungi causing snow molds. *Mycoscience* 50:26-38.

Huet, J., Sallé-Tourne, M., and Pernollet, J. 1994. Amino acid sequence and toxicity of the  $\alpha$  elicitor secreted with ubiquitin by *Phytophthora infestans*. *Molecular Plant-Microbe Interactions* 7(2):302-304.

Jashni M. K., *et al.* 2015. Synergistic action of a metalloprotease and a serine protease from *Fusarium oxysporum* f. sp. *lycopersici* cleaves chitin-binding tomato chitinases, reduces their antifungal activity, and enhances fungal virulence. *Molecular Plant-Microbe Interactions* 26(9):996-1008.

Jia, Y.J., *et al.* 2009. Polygalacturonase, pectate lyase and pectin methylesterase activity in pathogenic strains of *Phytophthora capsici* incubated under different conditions. *Journal of Phytopathology* 157:585-591.

Jiang, R.H.Y., and Tyler, B.M. 2012. Mechanisms and evolution of virulence in oomycetes. *Annual Review of Phytopathology* 50:295-318.

Judelson, H.S. 2012. Dynamics and innovations within oomycete genomes: insights into biology, pathology, and evolution. *Eukaryotic Cell* 11(11):1304-1312.

Kamoun, S. 2003. Molecular genetics of pathogenic oomycetes. *Eukaryotic Cell* 2(2):191-199.

Kamoun, S. 2006. A catalogue of the effector secretome of plant pathogenic oomycetes. *Annual Review of Phytopathology* 44:41-60.

Kim, S. And Pevzner P.A. 2014. MS-GF + makes progress towards a universal database search tool for proteomics. *Nature Communications* 5:5227-5236.

Kubicek, C.P., Starr, T.L., and Glass, N.L. 2014. Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annual Review of Phytopathology* 52:427-451.

Lascombe, M., *et al.* 2007. Structure of sylvaticin, a new  $\alpha$ -elicitor-like protein from *Pythium sylvaticum*. *Acta Crystallography D* 63:1102-1108.

Latchman, D.S., Estridge, J.K., and Kemp, L.M. 1987. Transcriptional induction of the ubiquitin gene during herpes simplex virus infection is dependent upon the viral immediate-early protein ICP4. *Nucleic Acid Research* 15(18):7283-7293.

Lévesque, C.A. 2011. Fifty years of oomycetes-from consolidation to evolutionary and genomic exploration. *Fungal Diversity* 50:35-46.

Lévesque, C.A., and De Cock, A.W.A.M. 2004. Molecular phylogeny and taxonomy of the genus *Pythium*. *Mycological Research* 108(12): 1363-1383.

Lévesque, C.A., *et al.* 2010. Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biology* 11:R73.

Li, J., *et al.* New insights into the evolution of subtilisin-like serine protease genes in Pezizomycotina. *BMC Evolutionary Biology* 10: 68-81.

Lipps, P.E. and Bruehl, G.W. 1978. Snow rot of winter wheat in Washington. *Phytopathology* 68:1120-1127.

Liu, J., Sturrock, R., and Ekramoddoullah, A. 2010. The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. *Plant Cell Reports* 29:419-436.

Lombard, V., *et al.* 2013. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acid Research* 42:D490-D495.

MacDonald, E., *et al.* 2002. Biochemical interaction between hyphae of two *Pythium* species (Oomycota) and host tissues. *Fungal Genetics and Biology* 37:245-249.

Marín-Rodríguez, M.C., Orchard, J., and Seymour, G.B. 2002. Pectate lyases, cell wall degradation and fruit softening. *Journal of Experimental Botany* 53(377):2115-2119.

Martin, F.N. 2000. Phylogenetic relationship among some *Pythium* species inferred from sequence analysis of the mitochondrial encoded cytochrome oxidase II gene. *Mycologia* 92(4):711-727.

Matsumoto, N. 2009. Snow molds: a group of fungi that prevail under snow. *Microbes and Enviroments* 24(1):14-20.

Meijer, H.J.G., *et al.* 2014. Profiling the secretome and extracellular proteome of the potato late blight pathogen *Pytophthora infestans*. *Molecular and Cellular Proteomics* 13: 2101-2113.

Meiken, J. and Min, X.J. 2012. Computational prediction of protein subcellular locations in eukaryotes: an experience report. *Computational Molecular Biology* 2(1):1-7.

Melton, L.D., *et al.* 2009. Mannans in primary and secondary plant cell walls. *New Zealand Journal of Forestry Science* 39:153-160.

Minoche, A.E., *et al.* 2015. Exploiting single-molecule transcript sequencing for eukaryotic gene prediction. *Genome Biology* 16:184-196.

Misner, I., *et al.* 2014. The secreted proteins of *Achlya hypogyna* and *Thraustotheca clavata* identify the ancestral oomycete secretome and reveal gene acquisitions by horizontal gene transfer. *Gene Biology and Evolution* 7(1):120-135.

- Murakami, R., *et al.* 2015. Surviving freezing in plant tissues by oomycetous snow molds. *Cryobiology* 70:208-210.
- Muszevska, A., *et al.* 2011. Independent Subtilases Expansions in Fungi Associated with Animals. *Molecular Biology and Evolution* 28(12):3395–3404.
- Nair, R.A., and Geethu, Ch. 2015. Purification and characterization of secretory serine protease from necrotrophic oomycete, *Pythium myriotylum* Dreschler. *World Journal of Microbiology and Biotechnology* 31:85-94.
- Olivieri, F., *et al.* 2002. Characterization of an extracellular serine protease of *Fusarium eumartii* and its action on pathogenesis related proteins. *European Journal of Plant Pathology* 108: 63–72.
- Obwald, W., *et al.* 2014. Strategies of attack and defense in woody plant-Phytophthora interactions. *Forest Pathology* 44:169-190.
- Ponchet, M., *et al.* 1999. Are elicitors cryptograms in plant-Oomycete communications?. *Cellular and Molecular Life Sciences* 56:1020-1047.
- Quinlan, A.R. and Hall, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841-842.
- Raffaele, S., *et al.* 2010. Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*. *BMC Genomics* 11:637.
- Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* 16(6):276-277.
- Robideau, G.P., *et al.* 2011. DNA barcoding of oomycetes with cytochrome c oxydase subunit I and internal transcribed spacer. *Molecular Ecology Resources* 11:1002-1001.
- Severino, V., *et al.* 2014. Molecular profiling of the *Phytophthora plurivora* secretome: a step towards understanding the cross-talk between plant pathogenic oomycetes and their host. *PLoS ONE*: 9(11):e112317.
- Shroeder, K.L., *et al.* 2013. Molecular detection and quantification of *Pythium* species: evolving taxonomy, new tools, and challenges. *Plant Disease* 97(1):4-20.
- Soanes, D.M., Richards, T.A., and Talbot, N.J. 2007. Insights from sequencing fungal and oomycetes genomes: what can we learn about plant disease and the evolution of pathogenicity. *The Plant Cell* 19(11):3318-3326.
- Spies, C.F.J., *et al.* 2011. Molecular analysis of *Pythium irregulare* isolates from grapevines in South Africa suggest a single variable species. *Fungal Biology* 115:1210-1224.

- Stevens, C., *et al.* 1996. Defense-related gene activation during an incompatible interaction between *Stagonospora (Septoria) nodorum* and barley (*Hordeum vulgare* L.) coleoptile cells. *Plant Molecular Biology* 31:741-749.
- Stovold, G.E. 1974. Root rot caused by *Pythium irregulare* Bulsman, an important factor in the decline of established subterranean clover pastures. *Australian Journal of Agricultural Research* 25:537-548.
- Tojo, M. and Newsham, K.K. 2012. Snow moulds in polar environments. *Fungal Ecology* 5(4):395-402.
- Trapnell, C., *et al.* 2010. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* 28:511-515.
- Trapnell, C., *et al.* 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* 7(3):562-578.
- Trapnell, C., *et al.* 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology* 31(1):46-54.
- Van Der Plaats-Niterink, J. 1981. Monograph of the genus *Pythium*. *Studies in Mycology* No. 21, Centraalbureau voor Schimmelcultures, Baarn, 244 pp. ISSN 0166-0616
- Vaudel, M., *et al.* 2011. SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics* 11, 996- 999 .
- Vaudel, M., *et al.* 2014. Shedding light on black boxes in protein identification. *Proteomics* 14:1001-1005.
- Vaudel, M., *et al.* 2015. PeptideShaker enables reanalysis of MS\_derived proteomics data sets. *Nature Biotechnology* 33(1):22-24.
- Vigers, A., *et al.* 1992. Thaumatin-like pathogenesis-related proteins are antifungal. *Plant Science* 83:155\_161.
- Vleeshouwers, V.G.A.A., *et al.* 2006. Agrobacterium-based high-throughput screening reveals specific recognition of INF elicitors in *Solanum*. *Molecular Plant Pathology* 7(6):499-510.
- Walton, J.D., 1994. Deconstructing the cell wall. *Plant Physiology* 104:1113-1118.
- Wawra, S., *et al.* 2012. Secretion, delivery and function of oomycete effector proteins. *Current Opinion in Microbiology* 15:685-691.
- Weiland, J.J. 2004. Production of protease isozymes by *Aphanomyces cochlioides* and *Aphanomyces euteiches*. *Physiological and Molecular Plant Pathology* 65:225-233.

Win, J., *et al.* 2012. Sequence divergent RXLR effector share a structural fold conserved across plant pathogenic oomycete species. *PloS Pathogens* 8(1):e1002400.

Zerillo, M.M., *et al.* 2013. Carbohydrate-active enzymes in *Pythium* and their role in plant cell wall and storage polysaccharide degradation. *PloS ONE* 8(9):e72572.

## APPENDIX 1.

### **Leaf Spotting of Turkish Filbert in Colorado Caused by *Xanthomonas arboricola* pv. *corylina* and *Pseudomonas syringae* pv. *syringae***

Jorge Ibarra, Graduate Student, and Jacob Snelling, Research Associate, Department of Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, CO 80523-1177; Kathleen Alexander, City Forester, City of Boulder Forestry, 5200 Pearl Street, Boulder, CO 80301-2437; and Ned Tisserat, Professor, Department of Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, CO 80523-1177

© 2012 Plant Management Network. Accepted for publication 7 March 2012. Published 17 May 2012.

Corresponding author: Ned Tisserat. [ned.tisserat@colostate.edu](mailto:ned.tisserat@colostate.edu)

**Included here with permission of the owner of the copyright.**

Ibarra, J., Snelling, J., Alexander, K., and Tissera, N. 2012. Leaf spotting of Turkish filbert in Colorado caused by *Xanthomonas arboricola* pv. *corylina* and *Pseudomonas syringae* pv. *syringae*. Online. Plant Health Progress doi:10.1094/PHP-2012-0517-01-BR.

Turkish filbert (*Corylus colurna* L.) is a medium-sized tree native to southwest Asia and southeast Europe. It is well adapted to urban settings, although its use as an ornamental in the United States has been limited (1). In 2010 and 2011, Turkish filberts at a site in Boulder, CO, USA, exhibited leaf spotting, partial defoliation, and shoot dieback (Fig. A1.1). Small, black angular leaf spots formed along leaf margins and veins resulting in leaf crinkling. Tan to brown lesions with a translucent sheen dotted the bark on twigs whereas shepherd's crooks formed on blighted succulent shoots. Bacterial streaming was consistently observed in necrotic tissue.

Approximately 5 to 10% of the canopy in each tree was affected. Trees were well maintained and did not appear to be suffering from drought or other stresses that may have contributed to the disease. However, spring weather in both years was cooler and wetter than normal.



Fig. A1.1. Symptoms caused by *Xanthomonas arboricola* pv. *corylina* in Turkish filbert include: (A) crown thinning resulting from defoliation and shoot blighting; (B) small, black angular spots concentrated on the leaf veins and margins; (C) tan twig lesions; and (D) shepherd's crook formation in blighted shoots. *Pseudomonas syringae* pv. *syringae* was also isolated from leaf

spots.

Two different bacteria were cultured on nutrient agar from leaf spots; the first isolate had smooth, yellow colonies whereas the second was translucent and exhibited irregular colony margins. Only bacteria with yellow colonies were isolated from blighted shoots. The yellow bacterium had sequence similarities of 99% to the 16S rDNA and *gyrB* (partial sequence only) respectively, of *Xanthomonas arboricola* pv. *corylina* ([blast.ncbi.nlm.nih.gov/Blast.cgi](http://blast.ncbi.nlm.nih.gov/Blast.cgi)). The translucent bacterium had sequence similarities of 99% and 98% to the 16S rDNA and to *hrpL*, respectively, of *Pseudomonas syringae* pv. *syringae*. Furthermore, a 752-bp fragment was amplified using primers specific to the syringomycin gene, *syrB* (Fig. A1.2). This distinguished our isolate from *Pseudomonas syringae* pv. *coryli*, which lacks *syrB* and is associated with leaf spotting and blighting of European filbert (*C. avellana*) in Italy (4).

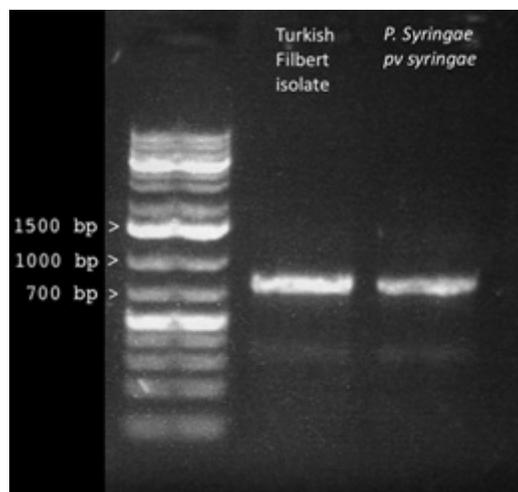


Fig. A1.2. Amplification of the syringomycin gene, *syrB*, from a *Pseudomonas syringae* pv. *syringae* isolate collected from Turkish filbert in Colorado (lane 2) and an isolate from bean (lane 3). *Pseudomonas syringae* pv. *coryli* lacks *syrB* and has only been found in European filbert in Italy.

The upper and lower leaf surfaces of four, one-year-old Turkish filbert trees were brushed to runoff with bacterial suspensions (107 cfu/ml) of *X. arboricola* pv. *corylina*, *P. syringae* pv.

*syringae*, or a mixture (1:1 v/v) of both species. Leaves of an additional four trees were brushed with distilled water. Trees were placed in a mist chamber for three days and then transferred to a greenhouse. After five days, a few angular, necrotic spots developed in leaves inoculated with one or both bacteria, but not in leaves treated with water (Fig. A1.3). No additional spotting was observed for the next month, but by seven weeks post-inoculation the number of leaf spots had increased to approximately the same severity in trees inoculated both with *X. arboricola* pv. *corylina* and with the mixture of *X. arboricola* pv. *corylina* and *P. syringae* pv. *syringae*, but not in those inoculated with *P. syringae* pv. *syringae* alone (Fig. A1.3). The two bacteria consistently were isolated from lesions. In the case of the mixed inoculation, both *X. arboricola* pv. *corylina* and *P. syringae* pv. *syringae* were recovered from the same lesions.

*Xanthomonas arboricola* pv. *corylina* is an important pathogen of European filbert in nut production areas (2,3). Turkish filbert has been listed as a host of *X. arboricola* pv. *corylina* in Europe (3), and Miller et al. (2) successfully reproduced symptoms in this species following artificial inoculations in Oregon in 1949. However, we were unable to find any published reports of natural infection of Turkish filbert in the United States. Thus, North American arborists should be aware of potential damage caused by this bacterium to this host. We did not reproduce shoot blight by artificial inoculation, but the dieback we observed in Turkish filbert is consistent with symptoms caused by *X. arboricola* pv. *corylina* in European filbert. We also documented that *P. syringae* pv. *syringae* can contribute to leaf spotting in Turkish filbert.

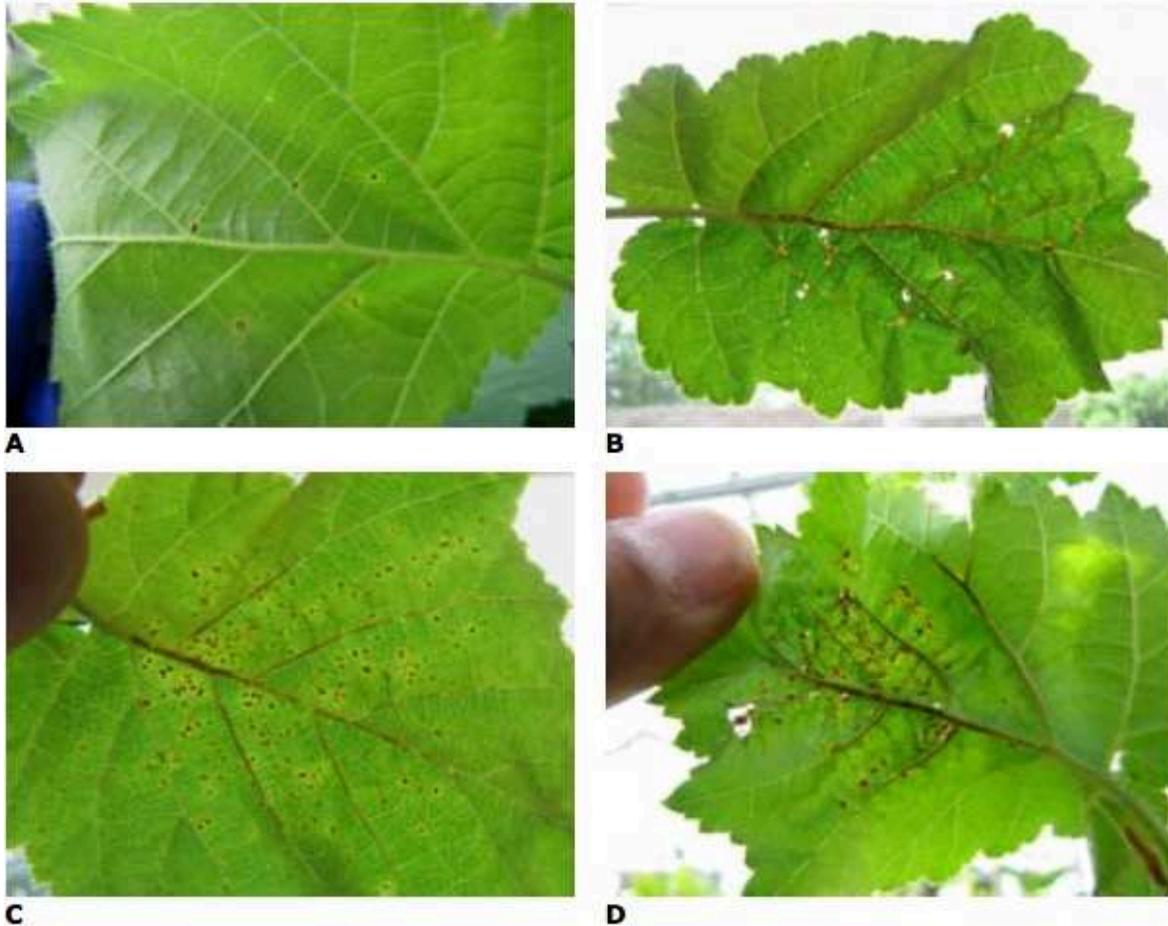


Fig.A1.3. Leaf spot symptoms in Turkish filbert: five days post inoculation with (A) a mixture of *Xanthomonas arboricola* pv. *corylina* and *Pseudomonas syringae* pv. *syringae*; and seven weeks post inoculation with (B) *P. syringae* pv. *syringae*, (C) *X. arboricola* pv. *corylina*, and (D) a mixture of the two bacteria.

## LITERATURE CITED

1. Dirr, M. A. 1998. Manual of Woody Landscape Plants: Their Identification, Ornamental Characteristics, Culture, Propagation and Uses, 5th Edn. Stipes Publ., Champaign, IL.
2. Miller, P. W., Bollen, W. B., and Simmons, J. E. 1949. Filbert bacteriosis and its control. Agric. Exp. Stn. Tech. Bull. 16. Oregon State College, Corvallis, WA.
3. OEPP/EPPO. 1986. Data sheets on quarantine organisms: No. 134, *Xanthomonas campestris* pv. *corylina* (Miller et al. 1940) Dye 1978. OEPP/EPPO Bulletin 16:13- 16.
4. Scortichini, M., Rossi, M. P., Loreti, S., Bosco, A., Fiori, M., Jackson, R. W., Stead, D. E., Aspin, A., Marchesi, U., Zini, M., and Janse, J. D. 2005. *Pseudomonas syringae* pv. *coryli*, the causal agent of bacterial twig dieback of *Corylus avellana*. Phytopathology 95:1316-1324.

## APPENDIX 2.

### **Genome Sequence of *Xanthomonas arboricola* pv. *Corylina*, Isolated from Turkish Filbert in Colorado**

Jorge Ibarra Caballero<sup>a</sup>, Marcelo M. Zerillo<sup>a</sup>, Jacob Snelling<sup>a\*</sup>, Christina Boucher<sup>b</sup>, Ned Tisserat<sup>a</sup>  
Department of Bioagricultural Science and Pest Management, Colorado State University, Fort Collins, Colorado, USA<sup>a</sup>; Department of Computer Science, Colorado State University, Fort Collins, Colorado, USA<sup>b</sup>

\* Present address: Jacob Snelling, Oregon State University, Corvallis, Oregon, USA. J.I.C. and M.M.Z. contributed equally to this work.

**Previously, we reported the isolation of a bacterium producing leaf spots in Turkish filbert. Here, we present the draft genome assembly of the bacterium identified as *Xanthomonas arboricola* pv. *corylina*. To our knowledge, this is the first published genome of this pathovar of *X. arboricola*.**

**Received** 2April2013. **Accepted** 8April2013. **Published** 23May2013.

**Citation** Ibarra Caballero J, Zerillo MM, Snelling J, Boucher C, Tisserat N. 2013. Genome sequence of *Xanthomonas arboricola* pv. *corylina*, isolated from Turkish filbert in Colorado. *Genome Announc.* 1(3):e00246-13. doi:10.1128/genomeA.00246-13. **Copyright** © 2013 Ibarra Caballero et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license.

Address correspondence to Ned Tisserat, [ned.tisserat@colostate.edu](mailto:ned.tisserat@colostate.edu).

*Xanthomonas arboricola* pv. *corylina* is a pathogen of *Corylus avellana* L. and it also infects

other species (1). In a previous report (2), we identified *X. arboricola* pv. *corylina* from leaf spots of *Corylus colurna* L. Here, we sequenced the genome of *X. arboricola* pv. *corylina* isolate NCCB100457 with 100 cycles of paired- end reads using an Illumina HiSeq sequencer at the USC Epi- genomic Center. More than 13.3 million 100-base-long reads were produced for each end. We performed two genome assem- blies, one using the A5 pipeline (3) and another using SPAdes v.2.3 (4). The A5 assembly has a total length of 5,227,695 bp consisting of 48 scaffolds (281 contigs), 43 of which were >500 bp, with the longest being 840,880 bp; the N50 is 263,170 bp and the G+C content is 65.46%. The SPAdes assembly has a total length of 5,398,516 bp placed in 810 nodes, 182 of which were >500 bp, with the longest being 263,848 bp; the N50 is 77,809 bp and the G+C content is 65.35%. We annotated the assembled genomes using the RAST server (5) and detected 4,452 and 4,500 coding sequences representing 445 and 449 “subsystems” for the A5 and SPAdes assemblies, respectively. Those numbers are similar to those in *Xanthomonas axonopodis* pv. *citri* (5,274,174 bp and 4,489 coding sequences represented in 461 subsystems), which is the most closely related organism present in the RAST database, based on nucleotide similarity. To further check the robustness of the NCCB100457 genome assemblies, we performed a BLASTn (6) search against 12 genes of *X. arboricola* pv. *corylina* that are commonly used as markers and that are present in the GenBank database. The genes *acnB*, *dnaK*, *fstZ*, *fyuA*, *gapA*, *qumA*, *rpoB*, and *rpoD* were detected and have 100% identity with other isolates from the same pathovar, while *fstX*, *groEL*, and 16S rRNA genes have 99% identity and *gyrB* has 98% identity.

We confirmed the presence of all 11 *hrp2* type 3 secretion system (T3SS) genes that are ubiquitous to all pathovars of *X. arboricola*; we also detected 20 out of 21 corresponding effector protein (T3E) genes that are present in all other *X. arboricola* pv. *corylina* isolates, including

ATCC 19313, collected from *Corylus maxima* in the United States (7); the only exception is *avrBs3*, which is absent in NCCB100457. A PCR with *avrBs3*-specific primers (7) using the genomic DNA of NCCB100457 did not result in amplification, whereas amplicons were detected in the positive controls *Xanthomonas oryzae* pv. *oryzae* PXO99A and *X. oryzae* pv. *oryzicola* BLS256. The gene *xopH*, which is present in most *X. arboricola* pv. *corylina* isolates but not in ATCC 19313, is also absent in our isolate. We did not detect 31 T3E genes in the NCCB100457 genome that are also absent in other *X. arboricola* pv. *corylina* genomes (7). However, we detected a putative *avr* similar to *hpoG1* from *Xanthomonas campestris* pv. *vasculorum* (NCBI accession no. ZP\_06487712.1) that was not reported before in *X. arboricola* pv. *corylina*. Differences in the arsenal of secretion systems and effectors can account for the pathogenicity and host specificity in pathogenic bacteria (8), including in isolate NCCB100457.

**Nucleotide sequence accession numbers.** This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/ GenBank under the accession no. APMC00000000. The version described in this paper is the first version, accession no. APMC01000000.

## ACKNOWLEDGMENTS

We thank Jan Leach for providing the DNA from *X. oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola*.

No external funding was used to support this research.

## REFERENCES

1. EPPO. 2004. Diagnostic protocols for regulated pests: PM7/22, *Xanthomonas arboricola* pv. *corylina*. Bull. OEPP/EPPO Bull. 34:155–157.

2. Ibarra J, Snelling J, Alexander K, Tisserat N. May 2012. Leaf spotting of Turkish filbert in Colorado caused by *Xanthomonas arboricola* pv. *corylina* and *Pseudomonas syringae* pv. *syringae*. Plant Health Prog. <http://www.plantmanagementnetwork.org/php/elements/sum.aspx?id=10321&photo=5671>.
3. Tritt A, Eisen JA, Facciotti MT, Darling AE. 2012. An integrated pipeline for *de novo* assembly of microbial genomes. PLoS One 7:e42304.
4. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455–477.
5. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics 9:75.
6. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.
7. Hajri A, Pothier JF, Fischer-Le Saux M, Bonneau S, Poussier S, Boureau T, Duffy B, Manceau C. 2012. Type three effector gene distribution and sequence analysis provide new insights into the pathogenicity of plant-pathogenic *Xanthomonas arboricola*. Appl. Environ. Microbiol. 78: 371–384.
8. Dean P. 2011. Functional domains and motifs of bacterial type III effector proteins and their roles in infection. FEMS Microbiol. Rev. 35:1100–1125.

### APPENDIX 3.

#### **Genome Sequences of Strain ATCC 29281 and Pin and Northern Red Oak Isolates of *Lonsdalea quercina* subsp. *quercina***

Jorge Ibarra Caballero<sup>a</sup>, Marcelo M. Zerillo<sup>a</sup>, Jacob Snelling<sup>a</sup>, Whitney Cranshaw<sup>a</sup>, Christina Boucher<sup>b</sup>, Ned Tisserat<sup>a</sup>

Department of Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, Colorado, USA<sup>a</sup>; Department of Computer Science, Colorado State University, Fort Collins, Colorado, USA<sup>b</sup>

**Two bacteria identified as *Lonsdalea quercina* subsp. *quercina* were isolated from oak trees showing symptoms of drippy blight. Here, we present their draft genome assemblies, as well as that of the type strain of this species. To our knowledge, these are the first published genome sequences of this subspecies of *Lonsdalea quercina*.**

**Received** 21May2014. **Accepted** 27May2014. **Published** 12June2014.

**Citation** Ibarra Caballero J, Zerillo MM, Snelling J, Cranshaw W, Boucher C, Tisserat N. 2014. Genome sequences of strain ATCC 29281 and pin and northern red oak isolates of *Lonsdalea quercina* subsp. *quercina*. *Genome Announc.* 2(3):e00584-14. doi:10.1128/genomeA.00584-14.

**Copyright** © 2014 Ibarra Caballero et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license.

Address correspondence to Ned Tisserat, ned.tisserat@colostate.edu.

*Lonsdalea quercina* (Hildebrand and Schroth 1967) Brady et al. 2012 (1), first described as *Erwinia quercina* (2) and later as *Brenneria quercina* (3), is associated with bacterial gummosis and twig necrosis of several oak (*Quercus*) species (drippy blight disease) (2, 4). We recovered two isolates, NCCB100489 and NCCB100490, in Colorado from a symptomatic pin oak (*Quercus palustris* Muench) and a northern red oak (*Quercus rubra* L.), respectively, which exhibited small cankers and gummosis at the Kermes scale (*Allokermes galliformis* Riley) feeding sites (5). We sequenced the genomes of these isolates and the genome of the *L. quercina* subsp. *quercina* type strain, ATCC 29281, which was originally collected from a diseased California live oak (*Quercus agrifolia* Nee) in California (2). Sequencing with 100 cycles of paired-end reads using the Illumina HiSeq sequencer at the University of Southern California (USC) Epigenome Center yielded more than 22.7 (NCCB100489), 23.8 (NCCB100490), and 32.9 (ATCC 29281) million 100-base-long reads for each paired end of each one of the three isolates. The assemblies were performed using Mira 4 Orc4 (6).

We obtained total assembly lengths of 3,848,371 bp, 3,847,844 bp, and 3,850,073 bp, contig numbers (and largest contig sizes in base pairs) of 35 (707,372), 35 (512,358), and 37 (557,855), *N*50 sizes of 334,484 bp, 279,081 bp, and 282,816 bp, and G+C contents of 55.6, 55.6, and 55.1% for strains NCCB100489, NCCB100490, and ATCC 29281, respectively. We annotated the assembled genomes using the RAST server (7) and detected 3,398 (NCCB100489), 3,400 (NCCB100490), and 3,460 (ATCC 29281) coding sequences representing 475, 475, and 472 subsystems, respectively. To further check the robustness of the assemblies, we performed a BLASTn (8) search for genes from *L. quercina* isolates that had previously been deposited in the GenBank database. In the ATCC 29281 genome assembly, the 16S rRNA gene was found to be 99% similar to the same gene reported in *L. quercina* subsp. *quercina* strain LMG 2725, while

the *atpD*, *gapA*, *gapDH*, *infB*, *recA*, and *rpoB* genes were 100% identical and *dnaJ*, *gyrB*, and *ompA* were 99%, 99%, and 97% similar, respectively, to the reported sequences in ATCC 29281 deposited into the GenBank by other groups. All these mentioned genes were identical to each other in the NCCB100489 and NCCB100490 genome assemblies, but these isolates had lower similarities to the corresponding sequences of 16S rRNA (99%), *atpD* (95%), *gapA* (97%), *gapDH* (98%), *infB* (96%), *recA* (97%), *rpoB* (98%), *dnaJ* (95%), *gyrB* (98%), and *ompA* (93%) in ATCC 29281. Nevertheless, a multilocus (16S, *gyrB*, *atpD*, and *infB*) phylogenetic tree based on the Bayesian inference (9) grouped NCCB100489, NCCB100490, and ATCC 29281 in an *L. quercina* subsp. *quercina* cluster, distinct from *L. quercina* subsp. *iberica* and *L. quercina* subsp. *britannica* isolates (data not shown). Biological testing is needed to explore the pathogenicity and host range differences of these and other *L. quercina* strains.

**Nucleotide sequence accession numbers.** These whole genome shotgun projects have been deposited at DDBJ/EMBL/ GenBank under accession no. JIBO00000000, JIBP00000000, and JIBQ00000000. The versions described in this paper are the first versions, JIBO01000000, JIBP01000000, and JIBQ01000000.

## ACKNOWLEDGMENT

No external funding was used to support this research.

## REFERENCES

1. Brady CL, Cleenwerck I, Denman S, Venter SN, Rodríguez-Palenzuela P, Coutinho T, De Vos P. 2012. Proposal to reclassify *Brenneria quercina* (Hildebrand and Schroth 1967) Hauben et al. 1999 into a new genus, *Lonsdalea* gen. nov., as *Lonsdalea quercina* comb. nov., descriptions of *Lonsdalea quercina* subsp. *quercina* comb. nov., *Lonsdalea quercina* subsp. *iberica* subsp. nov. and *Lonsdalea quercina* subsp. *britannica* subsp. nov., emendation of the description of the genus *Brenneria*, reclassification of *Dickeya dieffenbachiae* as *Dickeya*

- dadantii* subsp. *dieffenbachiae* comb. nov., and emendation of the description of *Dickeya dadantii*. Int. J. Syst. Evol. Microbiol. 62:1592–1602. <http://dx.doi.org/10.1099/ijs.0.035055-0>.
2. Hildebrand DC, Schroth MN. 1967. A new species of *Erwinia* causing the drippy nut disease of live oaks. *Phytopathology* 57:250–253.
  3. Hauben L, Moore ER, Vauterin L, Steenackers M, Mergaert J, Verdonck L, Swings J. 1998. Phylogenetic position of phytopathogens within the *Enterobacteriaceae*. *Syst. Appl. Microbiol.* 21:384–397. [http://dx.doi.org/10.1016/S0723-2020\(98\)80048-9](http://dx.doi.org/10.1016/S0723-2020(98)80048-9).
  4. Tóth T, Lakatos T, Koltay A. 2013. *Lonsdalea quercina* subsp. *populi* subsp. nov., isolated from bark canker or poplar trees. *Int. J. Syst. Evol. Microbiol.* 63:2309–2313. <http://dx.doi.org/10.1099/ijs.0.042911-0>.
  5. Snelling J, Tisserat NA, Cranshaw W. 2011. Kermes scale (*Allokermes* sp.) and the drippy nut pathogen (*Brenneria quercina*) associated with a decline of red oak species in Colorado. *Phytopathology* 101:S168.
  6. Chevreur B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information, p 45–56. *In* Computer science and biology. Proceedings of the German Conference on Bioinformatics, GCB '99. GCB, Hannover, Germany.
  7. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75. <http://dx.doi.org/10.1186/1471-2164-9-75>.
  8. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402. <http://dx.doi.org/10.1093/nar/25.17.3389>.
  9. Huelsenbeck JP, Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755. <http://dx.doi.org/10.1093/bioinformatics/17.8.754>.

## APPENDIX 4.

### **Population Structure of *Geosmithia morbida*, the Causal Agent of Thousand Cankers Disease of Walnut Trees in the United States**

Marcelo M. Zerillo<sup>1\*</sup>, Jorge Ibarra Caballero<sup>1</sup>, Keith Woeste<sup>2</sup>, Andrew D. Graves<sup>3</sup>, Colleen Hartel<sup>4</sup>, Jay W. Pscheidt<sup>5</sup>, Jadelys Tonos<sup>4</sup>, Kirk Broders<sup>6</sup>, Whitney Cranshaw<sup>1</sup>, Steven J. Seybold<sup>7</sup>, Ned Tisserat<sup>1</sup>

**1** Department of Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, Colorado, United States of America, **2** USDA Forest Service Hardwood Tree Improvement and Regeneration Center, Department of Forestry and Natural Resources, Purdue University, West Lafayette, Indiana, United States of America, **3** USDA Forest Service, Forest Health Protection, Albuquerque, New Mexico, United States of America, **4** Department of Forestry and Natural Resources, Purdue University, West Lafayette, Indiana, United States of America, **5** Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon, United States of America, **6** Department of Biological Sciences, University of New Hampshire, Durham, New Hampshire, United States of America, **7** USDA Forest Service, Pacific Southwest Research Station, Davis, California, United States of America

#### **Abstract**

The ascomycete *Geosmithia morbida* and the walnut twig beetle *Pityophthorus juglandis* are associated with thousand cankers disease of *Juglans* (walnut) and *Pterocarya* (wingnut). The disease was first reported in the western United States (USA) on several *Juglans* species, but has been found more recently in the eastern USA in the native range of the highly susceptible

*Juglans nigra*. We performed a comprehensive population genetic study of 209 *G. morbida* isolates collected from *Juglans* and *Pterocarya* from 17 geographic regions distributed across 12 U.S. states. The study was based on sequence typing of 27 single nucleotide polymorphisms from three genomic regions and genotyping with ten microsatellite primer pairs. Using multilocus sequence-typing data, 197 *G. morbida* isolates were placed into one of 57 haplotypes. In some instances, multiple haplotypes were recovered from isolates collected on the same tree. Twenty-four of the haplotypes (42%) were recovered from more than one isolate; the two most frequently occurring haplotypes (H02 and H03) represented 36% of all isolates. These two haplotypes were abundant in California, but were not recovered from Arizona or New Mexico. *G. morbida* population structure was best explained by four genetically distinct groups that clustered into three geographic regions. Most of the haplotypes isolated from the native range of *J. major* (Arizona and New Mexico) were found in those states only or present in distinct genetic clusters. There was no evidence of sexual reproduction or genetic recombination in any population. The scattered distribution of the genetic clusters indicated that *G. morbida* was likely disseminated to different regions at several times and from several sources. The large number of haplotypes observed and the genetic complexity of *G. morbida* indicate that it evolved in association with at least one *Juglans spp.* and the walnut twig beetle long before the first reports of the disease.

**Citation:** Zerillo MM, Ibarra Caballero J, Woeste K, Graves AD, Hartel C, et al. (2014) Population Structure of *Geosmithia morbida*, the Causal Agent of Thousand Cankers Disease of Walnut Trees in the United States. PLoS ONE 9(11): e112847.  
doi:10.1371/journal.pone.0112847

**Editor:** Ned B. Klopfenstein, USDA Forest Service - RMRS, United States of America

**Received** January 19, 2014; **Accepted** October 16, 2014; **Published** November 13, 2014

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

**Funding:** This research was supported by funds provided by the USDA NIFA Western Region IPM Center and Critical Issues-Emerging and New Plant and Animal Pests and Diseases grant programs, the USDA Forest Service Forest Health Monitoring Program (Detection Monitoring Grant No. INT-DM-09-01 and Evaluation Monitoring Grant No. INT-EM-B-11-03), and USDA Forest Service Special Technology Development Grant R4-2011-01 (administered by A.S. Munson). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

\* Email: mzerillo@colostate.edu

☞ These authors contributed equally to this work.

## **Introduction**

*Juglans nigra* L. (Juglandaceae), commonly referred to as black walnut or eastern black walnut, is a native tree species of eastern North America (Figure A4.1A). Its wood is highly prized for use in cabinetry, gunstocks, veneer, and other finished wood products, and the nuts are an important nutritional source for wildlife [1–3]. This species was widely planted in the western United States (USA) as an ornamental and nut-bearing tree [1,2] during European colonization

and the subsequent development of rural and urban landscapes. However, black walnut does not constitute a major proportion of trees in the modern urban landscape in this region.

In the early 1990 s, widespread decline and death of *J. nigra* from an undetermined cause was observed in Oregon (OR), USA [4,5]. Similar mortality in Utah (UT) in the late 1990 s [4] and in New Mexico (NM) in 2001 [6] was attributed to drought and attack by the walnut twig beetle (WTB), *Pityophthorus juglandis* Blackman (Coleoptera: Scolytidae) [7,8]. Beginning in 2001, decline and mortality of trees was also noted in several cities in Colorado (CO) [9]. Symptoms included yellowing of foliage and thinning of the upper crown, followed by twig and branch dieback and eventually tree death [9]. In 2008 an undescribed species of the ascomycete *Geosmithia* associated with the WTB was shown to cause bark cankers around WTB galleries in dying *J. nigra* in CO [9], and the fungus was subsequently described as *Geosmithia morbida* M. Kolarik, E. Freeland, C. Utle, & N.Tisserat sp.nov. (Ascomycota: Hypocreales) [10]. Many other *Geosmithia* species are symbionts of bark beetles, but *G. morbida* was the first phytopathogenic species reported in this genus [10]. The disease in *J. nigra* was given the common name thousand cankers disease (TCD) because of the enormous number of coalescing cankers that are formed around WTB entrance holes and galleries when the beetle aggregates in the inner bark of severely affected trees [9]. The WTB and TCD have now been documented as causing *J. nigra* mortality throughout much of the western USA and more recently, in many locations

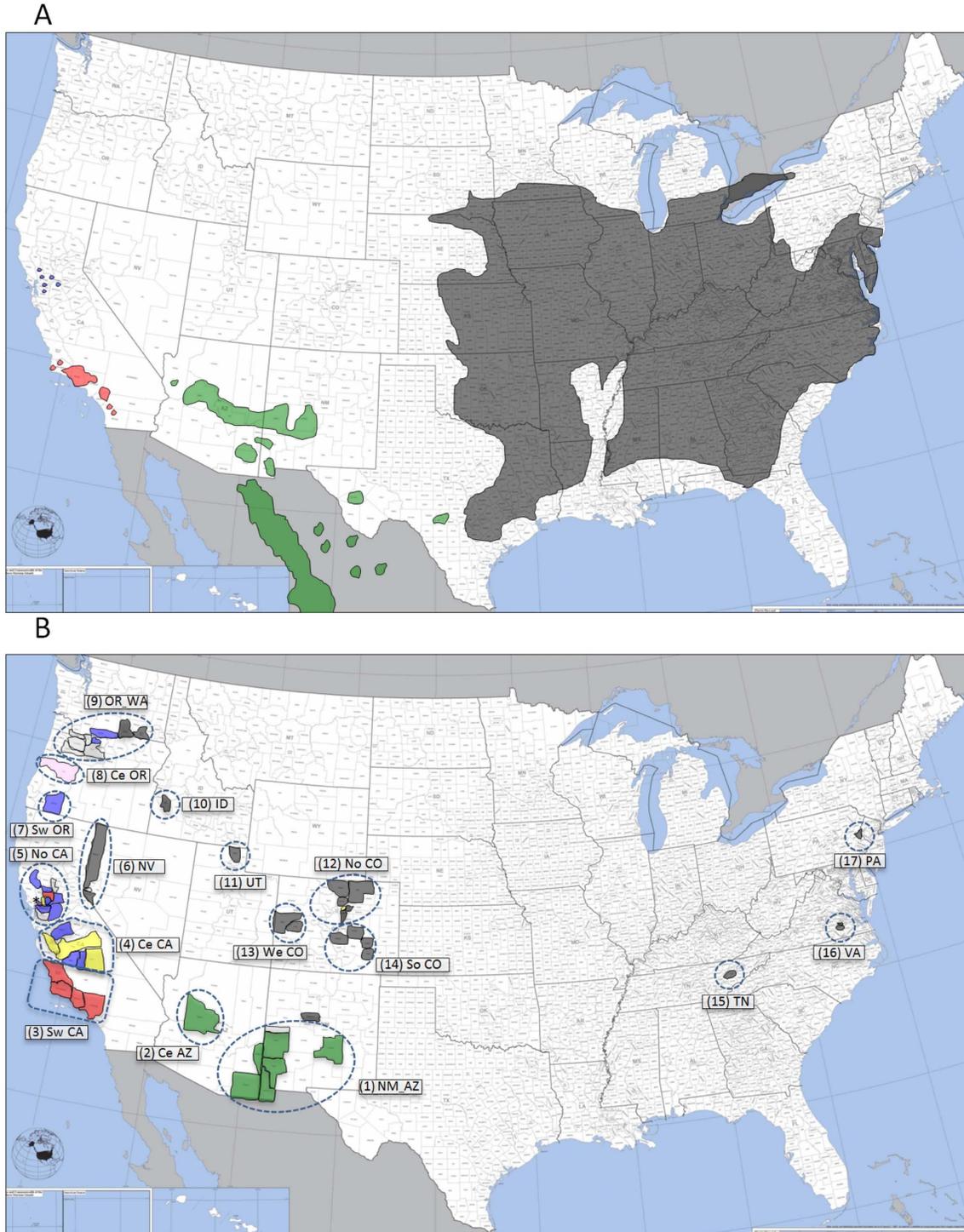


Figure A4.1. Distribution of some native species of *Juglans* in the United States (A) and sampling regions (17) of native and adventive *Juglans* for *Geosmithia morbida* (B). Regions/counties were color-coded according to the tree species with green = *J. major*; red = *J. californica*; yellow = *J. regia*; blue = *J. hindsii*; black = *J. nigra*; pink = *J. cinerea*; light gray = unidentified *Juglans* spp. or hybrids. (Figure adapted from U.S. Census Bureau and U.S. Geological Survey at [www.census.gov/](http://www.census.gov/)). doi:10.1371/journal.pone.0112847.g001

within the native range of this species [4,11–18]. In 2013, TCD was also confirmed in *J. nigra* planted in northeastern Italy [19].

The source of the current TCD epidemic remains unclear. There were no recorded collections of the WTB or *G. morbida* from *J. nigra* in the western USA prior to 1959 in southern California (CA); none prior to 1996 in UT [20]; and none in the native range of *J. nigra* until 2010 [11,12].

WTB was first collected in 1896 from *J. major* (Torr.) A. Heller (Arizona walnut) in NM [20,21], and it has since been collected widely throughout the northern native range of this species in Arizona (AZ), NM, and once in Chihuahua in Mexico (Figure A4.1A) [20,22].

*Geosmithia morbida* has also been isolated consistently from small, superficial cankers surrounding WTB galleries in native stands of *J. major* in AZ and NM, but the fungus does not cause extensive branch dieback or mortality in this species [20,23]. This has led to speculation that the origin of the WTB and *G. morbida* is from *J. major*, and that there was a host shift in the recent past by the beetle and its fungal symbiont to the more susceptible *J. nigra*. Another possible source of the WTB and *G. morbida* in the current epidemic are populations from Los Angeles County where WTB was collected in 1959 from *J. nigra* and from *J. californica* S. Watson (southern California black walnut) [7], the latter is a native tree in this region [24].

*Geosmithia morbida* has been recovered subsequently from declining and also relatively healthy appearing *J. californica* throughout its range in southern CA (Figure A4.1A) [25] and from many other species of *Juglans* and the closely related *Pterocarya* growing in arboreta and germplasm collections in CA (SJS, unpublished data).

The population structure of *G. morbida* in the USA has been partially characterized. Previous studies have indicated that the fungus is genetically complex [10,26]. Freeland and collaborators [26] identified 12 haplotypes from 145 *G. morbida* isolates collected in the western USA based

on rDNA ITS sequences, and 10 haplotypes from 56 isolates based on partial b-tubulin (BT) sequence comparisons. There were no correlations among haplotypes and the hosts or geographic regions from which the isolates were collected [26]. Representative ITS haplotypes recovered from different geographic regions and hosts were all pathogenic, and although there were slight differences in aggressiveness among isolates, canker sizes were not significantly different [26]. Freeland and collaborators [26] also showed that of four *G. morbida* isolates recovered from different cankers on the same tree, all had different di-locus haplotypes based on rDNA ITS and BT sequences. Hadziabdic and collaborators [27] used 15 polymorphic microsatellite loci to reveal high haploid genetic diversity in *G. morbida* isolates collected from the eastern USA and OR. They identified two genetic clusters that corresponded to isolates collected in: 1) OR and North Carolina (NC), and 2) Pennsylvania (PA) and Tennessee (TN). A sexual state for *G. morbida* has not been detected that could account for the observed genetic variability, but clonal organisms may have limited recombination events that can change population structure [28,29].

The main objective of our study was to determine the diversity and spatial pattern of *G. morbida* haplotypes collected widely and intensively from both native and introduced *Juglans* and *Pterocarya* hosts in various locations in the USA. *Pterocarya*, a non-native member of the Juglandaceae in the USA, was included because trees in this genus were recently reported with TCD symptoms [11]. We used multilocus sequence typing (MLST) and a microsatellite (or SSR, simple sequence repeats) analysis comprised of three genomic regions and ten markers, respectively, to observe the population structure of 209 *G. morbida* isolates collected from 17 different geographic regions (Figure A4.1B). Specifically we were interested in: i) comparing the population structure of *G. morbida* from disjunct geographic locations and ii) deducing the

potential source of the TCD outbreak and whether multiple introductions of the pathogen have occurred.

## **Materials and Methods**

### Fungal collection and isolation

Isolates of *G. morbida* were collected from 54 counties in 12 states in the USA (Table A4.1).

Samples were collected, transported, and shipped under the terms of USDA-APHIS permit P526P-11- 03416. No specific permissions were required for collection of isolates from any of the locations and the field studies did not involve endangered or protected species. It is important to clarify that “population” in the manuscript means a group of geographically co-located isolates. Because *G. morbida* is primarily asexual, possibly these isolates do not represent genetically interacting entities. For the purpose of analyses, isolates were grouped into one of 17 geographic regions separated by relatively large distances: (1) NM\_AZ (New Mexico and Arizona), (2) central AZ, (3) southwestern CA (California), (4) central CA, (5) northern CA, (6) NV (Nevada), (7) southwestern OR (Oregon), (8) central OR, (9) OR\_WA (northern OR and southern Washington), (10) ID (Idaho), (11) UT (Utah), (12) northern CO (Colorado), (13) western CO, (14) southern CO, (15) TN (Tennessee), (16) VA (Virginia), and (17) PA (Pennsylvania) (Figure A4.1B and Table A4.1). Isolates were cultured from the margins of stem or branch cankers surrounding *P. juglandis* galleries located in the phloem or bark of symptomatic *J. californica*, *J. cinerea* L. [30], *J. hindsii* (Jeps.) Jeps. ex R.E. Sm., *J. major*, *J. nigra*, and *J. regia* L., from hybrids of these species, or from undetermined species of *Juglans*, and from *Pterocarya stenoptera* C. DC. (Chinese wingnut) and *P. fraxinifolia* (Lam.) Spach (Caucasian wingnut) (Figure A4.1B and Table A4.1). Isolates from *J. californica*, *J. hindsii*, *J.*

*nigra*, and *J. regia* were collected in areas where TCD epidemics were in progress and where infected trees were readily apparent. Many *J. nigra* were large trees located in urban areas and often on private property, so sampling was restricted to trees with cankered branches that were easily accessible. Branch samples from *J. californica* and *J. hindsii* were collected from trees located in their native ranges (parks, National Forest lands, roadside riparian areas), or from germplasm collections. Samples from *J. hindsii* were also collected from roadside plantings in the Central Valley (CA) or in OR and WA. Samples from *J. regia* were collected primarily in Central Valley walnut orchards, with a few exceptions, Cache County, UT and Jefferson County, CO (Table A4.1). Samples from *J. major* were collected from trees located in their native range (Figure A4.1A), which were widely scattered, sparse and did not exhibit severe symptoms of TCD. Thus, for *J. major*, samples were collected arbitrarily from declining or asymptomatic trees that were identified during road surveys. Isolation from cankers was performed as described previously [9].

Table A4.1. Locations, hosts, haplotypes and genetic clusters based on the four-cluster-MLST-DAPC model, for *Geosmithia morbida* isolates.

Cluster	Haplotype	Isolate	GeoReg*	Year	State	County	Latitude	Longitude	Collector	Host
1	H01	1582 <sup>J</sup>	4_CeCA	2011	CA	Merced	37°18'55.3''	-120°31'38.2''	SJS/PLD	<i>J. hindsii</i>
1	H03	1217 <sup>**</sup> ☆	12_NoCO	2007	CO	Boulder	40°0'53''	-105°16'13''	NT	<i>J. nigra</i>
1	H03	1249	5_NoCA	2008	CA	Yolo	38°32'21.4''	-121°47'46.4''	SJS	<i>J. californica</i>
1	H03	1250	12_NoCO	2008	CO	Boulder	40°0'53''	-105°16'13''	NT	<i>J. nigra</i>
1	H03	1261	5_NoCA	2008	CA	San Joaquin	37°51'05.7''	-121°16'59.0''	SJS	<i>J. hindsii</i>
1	H03	1262	3_SwCA	2008	CA	Ventura	34°28'26.6''	-118°45'39.4''	SJS/TWC	<i>J. californica</i>
1	H03	1308 <sup>B</sup>	12_NoCO	2009	CO	Jefferson	39°45'30.42''	-105°04'49.8''	CU	<i>J. nigra</i>
1	H03	1387	5_NoCA	2009	CA	Solano	38°29'54.4''	-121°58'20.2''	SJS	<i>J. hindsii</i>
1	H03	1441	9_ORWA	2009	OR	Clackamas	45°11'24.0''	-122°12'36.0''	JP/CU	<i>Juglans sp.</i>
1	H03	1484 <sup>D</sup>	5_NoCA	2010	CA	Lake	39°10'21.4''	-122°54'32.2''	SJS	<i>J. hindsii</i>
1	H03	1489 <sup>D</sup>	5_NoCA	2010	CA	Lake	39°10'21.4''	-122°54'32.2''	SJS	<i>J. hindsii</i>
1	H03	1490 <sup>D</sup>	5_NoCA	2010	CA	Lake	39°10'21.4''	-122°54'32.2''	SJS	<i>J. hindsii</i>
1	H03	1491 <sup>D</sup>	5_NoCA	2010	CA	Lake	39°10'21.4''	-122°54'32.2''	SJS	<i>J. hindsii</i>
1	H03	1507 <sup>E</sup>	15_TN	2010	TN	Knox	36°01'45.82''	-83°55'28.61''	SS	<i>J. nigra</i>
1	H03	1515	4_CeCA	2010	CA	San Benito	34°36'28.5''	-120°21'08.5''	SJS	<i>J. regia</i>
1	H03	1520 <sup>G</sup>	4_CeCA	2010	CA	Fresno	36°46'03.1''	-119°56'53.0''	SJS/EF	<i>J. regia</i>
1	H03	1521 <sup>G</sup>	4_CeCA	2010	CA	Fresno	36°46'03.1''	-119°56'53.0''	SJS/EF	<i>J. regia</i>
1	H03	1524	15_TN	2010	TN	Knox	35°57'38.30''	-83°55'14.66''	WC	<i>J. nigra</i>

1	H03	1583 <sup>J</sup>	4_CeCA	2011	CA	Merced	37°18'55.3''	-120°31'38.2''	SJS/PLD	<i>J. hindsii</i>
1	H03	1659	17_PA	1011	PA	Bucks	39°57'08.41''	-75°09'49.64''	SJS	<i>J. nigra</i>
1	H03	1699	4_CeCA	2011	CA	Tulare	36°15'03.12''	-119°13'01.31''	SJS/EF	<i>J. hindsii</i>
1	H03	1822	5_NoCA	2012	CA	Solano	38°30'02.4''	-121°58'42.2''	SJS/PLD/SMH	<i>P. stenoptera</i>
1	H05	1236**	9_ORWA	2007	WA	Benton	46°12'24.48''	-119°46'08.12''	WC	<i>J. nigra</i>
1	H08	1505	15_TN	2010	TN	Knox	35°57'38.30''	-83°55'14.66''	SJS	<i>J. nigra</i>
1	H09	1274	12_NoCO	2008	CO	Jefferson	39°45'57.95''	-105°04'37.94''	CU	<i>J. nigra</i>
1	H09	1459	5_NoCA	2010	CA	Contra Costa	37°47'00.6''	-121°58'31.8''	SJS	<i>J. hindsii</i>
1	H09	1599 <sup>K</sup>	12_NoCO	2011	CO	Boulder	40°13'28.95''	-105°16'16.96''	NT	<i>J. nigra</i>
1	H09	1662 <sup>O</sup>	16_VA	2011	VA	Dinwiddie	37°15'44.8''	-77°24'43.0''	SJS	<i>J. nigra</i>
1	H09	1667 <sup>O</sup>	16_VA	2011	VA	Chesterfield	37°15'44.8''	-77°24'43.0''	SJS	<i>J. nigra</i>
1	H11	1233	5_NoCA	2008	CA	Yolo	38°32'50''	-121°47'52''	SJS	<i>J. hindsii</i>
1	H11	1272	3_SwCA	2008	CA	Ventura	34°28'26.6''	-118°45'39.4''	SJS/TWC	<i>J. californica</i>
1	H11	1503	none	2010	NM	Bernalillo	35°06'38.53''	-106°36'35.97''	SJS	<i>J. nigra</i>
1	H16	1239**	12_NoCO	2008	CO	Jefferson	39°50'11.95''	-105°02'13.94''	NT	<i>J. regia</i>
1	H16	1271	12_NoCO	2008	CO	Jefferson	39°45'19.95''	-105°13'15.96''	NT	<i>J. nigra</i>
1	H17	1573	1_NMAZ	2011	NM	Catron	33°37'10.7''	-108°53'39.1''	SJS/ADG	<i>J. major</i>
1	H18	1311 <sup>B</sup>	12_NoCO	2009	CO	Jefferson	39°45'30.42''	-105°04'49.8''	CU	<i>J. nigra</i>
1	H19	1547	12_NoCO	2010	CO	Denver	39°44'24''	-104°59'32''	NT	<i>J. nigra</i>
1	H19	1273	12_NoCO	2008	CO	Jefferson	39°45'30.42''	-105°13'15.95''	CU	<i>J. nigra</i>
1	H19	1312 <sup>B</sup>	12_NoCO	2008	CO	Jefferson	39°45'30.42''	-105°04'49.8''	CU	<i>J. nigra</i>

1	H19	1313 <sup>B</sup>	12_NoCO	2009	CO	Jefferson	39°45'30.42''	-105°04'49.8''	CU	<i>J. nigra</i>
1	H19	1596 <sup>K</sup>	12_NoCO	2011	CO	Boulder	40°13'28.95''	-105°16'16.96''	NT	<i>J. nigra</i>
1	H19	1597 <sup>K</sup>	12_NoCO	2011	CO	Boulder	40°13'28.95''	-105°16'16.96''	NT	<i>J. nigra</i>
1	H21	1598 <sup>K</sup>	12_NoCO	2011	CO	Boulder	40°13'28.95''	-105°16'16.96''	NT	<i>J. nigra</i>
1	H21	1453	1_NMAZ	2010	NM	Grant	32°37'18.7''	-108°24'25.2''	DL	<i>J. major</i>
1	H21	1540	1_NMAZ	2010	NM	Lincoln	33°44'24.0''	-105°27'36.0''	WC	<i>J. major</i>
1	H21	1558	1_NMAZ	2010	NM	Grant	32°43'48.0''	-108°22'48.0''	WC	<i>J. major</i>
1	H21	1559	1_NMAZ	2010	NM	Grant	32°43'48.0''	-108°22'48.0''	WC	<i>J. major</i>
1	H21	1670	1_NMAZ	2011	NM	Lincoln	33°06'22.5''	-105°48'17.8''	ADG	<i>J. major</i>
1	H22	1478 <sup>C</sup>	3_SwCA	2010	CA	Santa Barbara	34°36'28.5''	-120°21'08.5''	SJS	<i>J. californica</i>
1	H23	1309	11_UT	2009	UT	Cache	41°55'41.91''	-111°48'29.33''	NT	<i>J. nigra</i>
1	H25	1462	5_NoCA	2010	CA	Contra Costa	37°47'00.6''	-121°58'31.8''	SJS	<i>J. hindsii</i>
1	H25	1476 <sup>C</sup>	3_SwCA	2010	CA	Santa Barbara	34°36'28.5''	-120°21'08.5''	SJS	<i>J. californica</i>
1	H25	1518 <sup>F</sup>	4_CeCA	2010	CA	Fresno	36°46'03.1''	-119°56'53.0''	SJS/EF	<i>J. regia</i>
1	H26	1407	4_CeCA	2009	CA	Tulare	36°14'08.6''	-119°14'32.6''	SJS/ADG/EF	<i>J. regia</i>
1	H26	1461	5_NoCA	2010	CA	Contra Costa	37°47'00.6''	-121°58'31.8''	SJS	<i>J. hindsii</i>
1	H26	1486 <sup>D</sup>	5_NoCA	2010	CA	Lake	39°10'21.4''	-122°54'32.2''	SJS	<i>J. hindsii</i>
1	H27	1440	9_ORWA	2009	OR	Clackamas	45°11'24.0''	-122°12'36.0''	JP/CU	<i>Juglans sp.</i>
1	H27	1509 <sup>E</sup>	15_TN	2010	TN	Knox	36°01'45.82''	-83°55'28.61''	SJS	<i>J. nigra</i>
1	H29	1506 <sup>E</sup>	15_TN	2010	TN	Knox	36°01'45.82''	-83°55'28.61''	SJS	<i>J. nigra</i>
1	H35	1545	12_NoCO	2010	CO	Denver	39°44'24''	-104°59'32''	NT	<i>J. nigra</i>

1	H43	1572 <sup>H</sup>	1_NMAZ	2011	NM	Hidalgo	31°26'25.6''	-108°58'45.9''	SJS/ADG	<i>J. major</i>
1	H44	1314 <sup>B</sup>	12_NoCO	2009	CO	Jefferson	39°45'30.42''	-105°04'49.8''	CU	<i>J. nigra</i>
1	H56	1821 <sup>P</sup>	5_NoCA	2012	CA	Solano	38°19'10''	-121°55'27''	SJS/PLD/SMH	<i>P.stenoptera</i>
2	H02	<u>1224</u>	12_NoCO	2008	CO	Boulder	40°13'28''	-105°16'16''	NT	<i>J. nigra</i>
2	H02	<u>1225</u>	12_NoCO	2008	CO	Boulder	40°13'28''	-105°16'16''	NT	<i>J. nigra</i>
2	H02	<u>1227</u>	5_NoCA	2009	CA	Yolo	38°32'21.4''	-121°47'46.4''	SJS	<i>J.californica</i>
2	H02	<u>1245</u>	10_ID	2008	ID	Ada	43°35'56.76''	-116°12'52.58''	WC	<i>J. nigra</i>
2	H02	<u>1252</u>	3_SwCA	2008	CA	S Luis Obispo	35°03'08.0''	-119°54'33.0''	SJS/DLW	<i>J. hindsii</i> X <i>J. major</i>
2	H02	<u>1260</u>	9_ORWA	2008	OR	Hood River	45°42'24.1''	-121°31'18.1''	JP/CU	<i>Juglans sp.</i>
2	H02	<u>1263</u>	3_SwCA	2008	CA	Ventura	34°28'26.6''	-118°45'39.4''	SJS/TWC	<i>J.californica</i>
2	H02	<u>1264</u>	3_SwCA	2008	CA	Ventura	34°28'25.8''	-118°45'39.4''	SJS/TWC	<i>J.californica</i>
2	H02	<u>1267</u>	5_NoCA	2008	CA	San Joaquin	37°51'05.7''	-121°16'59.0''	SJS	<i>J. hindsii</i>
2	H02	<u>1268</u>	3_SwCA	2008	CA	Ventura	34°20'02.5''	-118°54'07.5''	SJS/TWC	<i>J.californica</i>
2	H02	<u>1270</u>	5_NoCA	2008	CA	San Joaquin	37°51'05.7''	-121°16'59.0''	SJS	<i>J. hindsii</i>
2	H02	<u>1275</u>	3_SwCA	2008	CA	Ventura	34°20'02.5''	-118°54'07.5''	SJS/TWC	<i>J.californica</i>
2	H02	<u>1279</u>	12_NoCO	2008	CO	Jefferson	39°45'30.42''	-105°04'49.8''	CU	<i>J. nigra</i>
2	H02	<u>1289</u>	5_NoCA	2009	CA	Yolo	38°32'21.4''	-121°47'46.4''	SJS	<i>J.californica</i>
2	H02	1315 <sup>B</sup>	12_NoCO	2009	CO	Jefferson	39°45'30.42''	-105°04'49.8''	CU	<i>J. nigra</i>
2	H02	1324	12_NoCO	2009	CO	Boulder	40°03'05''	-105°02'59''	CU	<i>J. nigra</i>
2	H02	1334	14_SoCO	2009	CO	Otero	38°03'09''	-103°43'12''	WC	<i>J. nigra</i>
2	H02	1335	14_SoCO	2009	CO	Crowley	38°09'55''	-103°56'46''	WC	<i>J. nigra</i>

2	H02	<u>1345</u>	9_ORWA	2009	OR	Hood River	45°42'24.1''	-121°31'18.1''	JP/CU	<i>Juglans sp.</i>
2	H02	<u>1352</u>	3_SwCA	2009	CA	Los Angeles	34°03'02.5''	-117°49'29.7''	SJS	<i>J.californica</i>
2	H02	<u>1355</u>	5_NoCA	2009	CA	Yolo	38°32'21.4''	-121°47'46.4''	SJS	<i>J. hindsii</i>
2	H02	<u>1365</u>	5_NoCA	2009	CA	Yolo	38°32'21.4''	-121°47'46.4''	SJS	<i>J. hindsii</i>
2	H02	<u>1381</u>	9_ORWA	2009	OR	Clackamas	45°11'24.0''	-122°12'36.0''	JP/CU	<i>Juglans sp.</i>
2	H02	<u>1385</u>	5_NoCA	2009	CA	Sutter	39°03'40.9''	-121°36'49.1''	SJS	<i>J. nigra</i> X <i>J. hindsii</i>
2	H02	1392	5_NoCA	2009	CA	Yolo	38°32'05.0''	-121°48'12.1''	SJS	<i>J. hindsii</i>
2	H02	<u>1403</u>	4_CeCA	2009	CA	Kings	36°19'41.0''	-119°32'37.4''	SJS/ADG	<i>J. hindsii</i>
2	H02	1408	12_NoCO	2009	CO	Boulder	40°13'28''	-105°16'16''	NT	<i>J. nigra</i>
2	H02	<u>1427</u>	5_NoCA	2009	CA	Sutter	39°03'40.9''	-121°36'49.1''	SJS	<i>J. nigra</i> X <i>J. hindsii</i>
2	H02	<u>1452</u>	5_NoCA	2009	CA	Yolo	38°32'21.4''	-121°58'14.2''	SJS	<i>J. hindsii</i>
2	H02	1481 <sup>C</sup>	3_SwCA	2010	CA	Santa Barbara	34°36'28.5''	-120°21'08.5''	SJS	<i>J.californica</i>
2	H02	<u>1482</u> <sup>C</sup>	3_SwCA	2010	CA	Santa Barbara	34°36'28.5''	-120°21'08.5''	SJS	<i>J.californica</i>
2	H02	<u>1485</u> <sup>D</sup>	5_NoCA	2010	CA	Lake	39°10'21.4''	-122°54'32.2''	SJS	<i>J. hindsii</i>
2	H02	1493	3_SwCA	2010	CA	Los Angeles	34°03'12''	-118°14'43''	WC	<i>J.californica</i>
2	H02	1514	4_CeCA	2009	CA	Tulare	36°14'08.6''	-119°14'32.6''	SJS/ADG/EF	<i>J. regia</i>
2	H02	<u>1523</u>	15_TN	2010	TN	Knox	35°55'30.05''	-83°59'19.85''	SJS/CL	<i>J. nigra</i>
2	H02	1525	15_TN	2010	TN	Knox	35°55'30.05''	-83°59'19.85''	SJS/CL	<i>J. nigra</i>
2	H02	1530	9_ORWA	2010	WA	Klickitat	45°44'49.9''	-120°26'16.2''	SJS/CL	<i>J. hindsii</i>
2	H02	<u>1535</u>	14_SoCO	2010	CO	Otero	38°03'9''	-103°43'12''	WC	<i>J. nigra</i>

2	H02	1581 <sup>J</sup>	4_CeCA	2011	CA	Merced	37°18'55.3''	-120°31'38.2''	SJS/PLD	<i>J. hindsii</i>
2	H02	1590	9_ORWA	2011	WA	Walla Walla	46°13'48.0''	-118°28'48.0''	JM	<i>J. nigra</i>
2	H02	1600 <sup>K</sup>	12_NoCO	2011	CO	Boulder	40°13'28.95''	-105°16'16.96''	NT	<i>J. nigra</i>
2	H02	1602 <sup>K</sup>	12_NoCO	2011	CO	Boulder	40°13'28.95''	-105°16'16.96''	NT	<i>J. nigra</i>
2	H02	1631 <sup>M</sup>	6_NV	2011	NV	Carson City	39°10'21.55''	-119°46'7.92''	ADG/TWC	<i>Juglans sp.</i>
2	H02	1633 <sup>M</sup>	6_NV	2011	NV	Carson City	39°10'21.55''	-119°46'7.92''	ADG/TWC	<i>Juglans sp.</i>
2	H02	1636 <sup>N</sup>	3_SwCA	2011	CA	Los Angeles	34°03'12''	-103°56'46''	WC	<i>J. californica</i>
2	H02	1660	16_VA	2011	VA	Dinwiddie	37°15'44.8''	-77°24'43.0''	SJS	<i>J. nigra</i>
2	H02	1665 <sup>O</sup>	16_VA	2011	VA	Chesterfield	37°15'44.8''	-77°24'43.0''	SJS	<i>J. nigra</i>
2	H02	1704	6_NV	2011	NV	Washoe	39°31'13.73''	-119°48'27.9''	SJS/PLD	<i>J. nigra</i>
2	H02	1830 <sup>P</sup>	35_NoCA	2012	CA	Solano	38°19'10''	-121°55'27''	SJS/PLD/	<i>P. fraxinifolia</i>
2	H02	2032	22_SwCA	2014	CA	Los Angeles	34°8.452'	-118°3.444'	KJG/JEH/SMH/	<i>P. fraxinifolia</i>
2	H04	1223	11_UT	2008	UT	Cache	41°55'1.74''	-111°48'48.8''	NT	<i>J. nigra</i>
2	H04	1266	3_SwCA	2008	CA	Ventura	34°25.548'	-119°05.354'	SJS/TWC	<i>J. californica</i>
2	H04	1338	9_ORWA	2009	OR	Marion	4°4'22.69''	-122°51'32.1''	JP/CU	<i>Juglans sp.</i>
2	H04	1368	13_WeCO	2009	CO	Mesa	39°3'49.93''	-108°33'2.3''	BH	<i>J. nigra</i>
2	H04	1380	5_NoCA	2009	CA	Sacramento	38°39.6624'	-121°28.723'	SJS	<i>J. hindsii</i>
2	H04	1432	9_ORWA	2009	OR	Wasco	45°09'36.0''	-121°09'36.0''	JP/CU	<i>Juglans sp.</i>
2	H04	1439	9_ORWA	2009	OR	Marion	45°8'32.17''	-122°51'32.1''	JP/CU	<i>Juglans sp.</i>
2	H04	1534	7_SwOR	2010	OR	Jackson	42°25'37.6''	-122°57'20.5''	SJS/CL	<i>J. hindsii</i>
2	H06	1366	1_NMAZ	2009	AZ	Yavapai	34°58'25.1''	-112°39'52.4''	WC	<i>J. major</i>

2	H07	<u>1428</u>	5_NoCA	2010	CA	Alameda	37°40'29.3''	-121°53'8.9''	SJS/PLD	<i>J. hindsii</i>
2	H10	<u>1258</u>	9_ORWA	2008	OR	Wasco	45°36'04.0''	-121°10'58.1''	MP	<i>Juglans sp.</i>
2	H10	<u>1276</u>	12_NoCO	2008	CO	Larimer	40°18'22''	-105°4'44.9''	NT	<i>J. nigra</i>
2	H10	<u>1346</u>	9_ORWA	2009	OR	Wasco	45°41'1.14''	-121°23'51.9''	JP/CU	<i>Juglans sp.</i>
2	H10	1601 <sup>K</sup>	12_NoCO	2011	CO	Boulder	40°13'28.95''	-105°16'16.96''	NT	<i>J. nigra</i>
2	H10	1672	15_TN	2011	TN	Knox	35°55'30.05''	-83°59'19.85''	SJS/CL	<i>J. nigra</i>
2	H12	1533	7_SwOR	2010	OR	Jackson	42°25'37.6''	-122°57'20.5''	SJS/CL	<i>J. hindsii</i>
2	H12	1584 <sup>J</sup>	4_CeCA	2010	CA	Merced	37°18'55.3''	-120°31'38.2''	SJS/PLD	<i>J. hindsii</i>
2	H13	<u>1393</u>	12_NoCO	2009	CO	Weld	40°18'9''	-105°4'51''	NT	<i>J. nigra</i>
2	H14	<u>1256</u> **	9_ORWA	2008	OR	Multnomah	45°30'42.48''	-122°40'32.2''	MP	<i>Juglans sp.</i>
2	H14	1634 <sup>N</sup>	3_SwCA	2011	CA	Los Angeles	34°3'12''	-103°56'46''	WC	<i>J. californica</i>
2	H15	1610	5_NoCA	2011	CA	Yolo	38°32'21.4''	-121°47'42.0''	SJS/CL	<i>J. major</i> X <i>J. hindsii</i>
2	H20	<u>1285</u>	13_WeCO	2008	CO	Delta	38°44'41.316''	-108°4'15.0''	BH	<i>J. nigra</i>
2	H20	1305 <sup>B</sup>	12_NoCO	2009	CO	Jefferson	39°45'30.42''	-105°04'49.8''	CU	<i>J. nigra</i>
2	H24	<u>1218</u> ** ☆	12_NoCO	2007	CO	Boulder	40°0'56''	-105°16'45''	NT	<i>J. nigra</i>
2	H24	<u>1259</u>	9_ORWA	2008	OR	Marion	45°08'49.9''	-122°51'29.9''	JP/CU	<i>Juglans sp.</i>
2	H24	<u>1318</u>	14_SoCO	2009	CO	Otero	38°03'11.56''	-103°42'52.7''	WC	<i>J. nigra</i>
2	H24	1404	4_CeCA	2009	CA	Tulare	36°14'08.6''	-119°14'32.6''	SJS/ADG/EF	<i>J. regia</i>
2	H24	1494	3_SwCA	2011	CA	Los Angeles	34°03'12''	-103°56'46''	WC	<i>J. californica</i>
2	H24	1532	9_ORWA	2010	OR	Wasco	45°36'29.6''	-121°07'40.3''	SJS/CL	<i>J. hindsii</i>
2	H28	<u>1409</u>	4_CeCA	2009	CA	Tulare	36°14'08.6''	-119°14'32.6''	SJS/ADG/EF	<i>J. regia</i>

2	H30	1269**	3_SwCA	2008	CA	Ventura	34°20'02.5''	-118°54'07.5''	SJS/TWC	<i>J.californica</i>
2	H30	<u>1350</u>	12_NoCO	2009	CO	Jefferson	39°49'12.0''	-105°06'40.0''	CU	<i>J. nigra</i>
2	H30	1430	9_ORWA	2009	OR	Marion	45°08'32.14''	-122°51'31.8''	JP/CU	<i>Juglans sp.</i>
2	H30	1519 <sup>G</sup>	4_CeCA	2010	CA	Fresno	36°46'03.1''	-119°56'53.0''	SJS/EF	<i>J. regia</i>
2	H30	<u>1522</u>	15_TN	2010	TN	Knox	35°57'37.76''	-83°55'15.2''	WC/NT	<i>J. nigra</i>
2	H30	1726	8_CeOR	2011	OR	Lane	43°57'00.0''	-122°52'48.0''	MS	<i>J. cinerea</i>
2	H30	1727	8_CeOR	2011	OR	Lane	43°57'00.0''	-122°52'48.0''	MS	<i>J. cinerea</i>
2	H31	<u>1483</u> <sup>D</sup>	5_NoCA	2010	CA	Lake	39°10'21.4''	-122°54'32.2''	SJS	<i>J. hindsii</i>
2	H34	<u>1247</u>	14_SoCO	2008	CO	El Paso	38°50'0.35''	-104°49'18.5''	NT	<i>J. nigra</i>
2	H42	1623 <sup>L</sup>	1_NMAZ	2011	AZ	Cochise	31°35'32.5''	-109°30'23.9''	TWC	<i>J. major</i>
2	H42	1624 <sup>L</sup>	1_NMAZ	2011	AZ	Cochise	31°35'32.5''	-109°30'23.9''	TWC	<i>J. major</i>
2	H42	1627 <sup>L</sup>	1_NMAZ	2011	AZ	Cochise	31°35'32.5''	-109°30'23.9''	TWC	<i>J. major</i>
2	H45	1626 <sup>L</sup>	1_NMAZ	2011	AZ	Cochise	31°35'32.5''	-109°30'23.9''	TWC	<i>J. major</i>
2	H46	1628 <sup>L</sup>	1_NMAZ	2011	AZ	Cochise	31°35'32.5''	-109°30'23.9''	TWC	<i>J. major</i>
2	H47	1402	9_ORWA	2009	OR	Wasco	45°41'1.14''	-121°23'51.9''	JP	<i>Juglans sp.</i>
2	H47	1585	4_CeCA	2011	CA	Tulare	36°07'51.1''	-119°08'16.9''	SJS/EF/PLD	<i>J. regia</i>
2	H47	1586	4_CeCA	2011	CA	Tulare	36°07'51.1''	-119°08'16.9''	SJS/EF/PLD	<i>J. regia</i>
2	H48	1306 <sup>B</sup>	12_NoCO	2009	CO	Jefferson	39°45'30.42''	-105°04'49.8''	CU	<i>J. nigra</i>
2	H49	1496	14_SoCO	2010	CO	Fremont	38°26'26.34''	-105°14'17.3''	WC	<i>J. nigra</i>
2	H51	1708	6_NV	2011	NV	Douglas	39°0'15.10''	-119°50'42.0''	SJS/PD/GD	<i>J. nigra</i>
2	H52	1625 <sup>L</sup>	1_NMAZ	2011	AZ	Cochise	31°35'32.5''	-109°30'23.9''	TWC	<i>J. major</i>

2	H55	1513	5_NoCA	2010	CA	Solano	38°16'12.0''	-121°56'24.0''	RMB	<i>J. regia</i>
2	H57	<u>1246</u>	14_SoCO	2008	CO	El Paso	38°50'0.35''	-104°49'18.5''	NT	<i>J. nigra</i>
3	H36	1569 <sup>H</sup>	1_NMAZ	2011	NM	Hidalgo	31°26'25.6''	-108°58'45.9''	SJS/ADG	<i>J. major</i>
3	H37	1576 <sup>I</sup>	1_NMAZ	2011	NM	Catron	33°37'10.7''	-108°53'39.1''	SJS/ADG	<i>J. major</i>
3	H38	<u>1301</u> <sup>B</sup>	12_NoCO	2009	CO	Jefferson	39°45'30.42''	-105°04'49.8''	CU	<i>J. nigra</i>
3	H38	<u>1302</u> <sup>A</sup>	2_CeAZ	2009	AZ	Yavapai	34°55'25.1''	-112°50'34.7''	WC	<i>J. major</i>
3	H38	<u>1321</u> <sup>B</sup>	12_NoCO	2009	CO	Jefferson	39°45'30.42''	-105°04'49.8''	CU	<i>J. nigra</i>
3	H38	1340	1_NMAZ	2009	NM	Grant	32°43'48.0''	-108°22'48.0''	WC	<i>J. major</i>
3	H38	1388	1_NMAZ	2009	AZ	Yavapai	34°55'25.1''	-112°50'34.7''	WC	<i>J. major</i>
3	H38	<u>1401</u>	1_NMAZ	2009	NM	Catron	33°37'10.7''	-108°53'39.1''	SJS/ADG	<i>J. major</i>
3	H38	1554	2_CeAZ	2010	AZ	Yavapai	34°41'42.62''	-112°09'44.4''	WC	<i>J. major</i>
3	H38	1556	1_NMAZ	2010	NM	Grant	32°38'17.9''	-108°06'18.4''	WC	<i>J. major</i>
3	H38	1574 <sup>I</sup>	1_NMAZ	2010	NM	Catron	33°37'10.7''	-108°53'39.1''	SJS/ADG	<i>J. major</i>
3	H39	1373	1_NMAZ	2010	NM	Grant	32°54'21.7''	-107°43'45.5''	DL	<i>J. major</i>
3	H39	1421	1_NMAZ	2010	NM	Grant	32°54'21.7''	-107°43'45.5''	DL	<i>J. major</i>
3	H39	1422	1_NMAZ	2010	NM	Grant	32°55'08.5''	-107°34'35.1''	DL	<i>J. major</i>
3	H39	1557	1_NMAZ	2010	NM	Grant	32°52'44.5''	-107°52'04.5''	WC	<i>J. major</i>
3	H40	<u>1248</u>	12_NoCO	2008	CO	Jefferson	39°45'30.42''	-105°13'15.95''	NT	<i>J. nigra</i>
3	H40	1322	2_CeAZ	2009	AZ	Yavapai	34°27'05.32''	-112°16'09.39''	WC	<i>J. major</i>
3	H40	1323 <sup>B</sup>	12_NoCO	2008	CO	Jefferson	39°45'30.42''	-105°04'49.8''	NT	<i>J. nigra</i>
3	H40	<u>1358</u>	2_CeAZ	2009	AZ	Yavapai	34°55'25.1''	-112°50'34.7''	WC	<i>J. major</i>

3	H40	1420	2_CeAZ	2009	AZ	Yavapai	34°55'25.1''	-112°50'34.7''	WC	<i>J. major</i>
3	H41	1571 <sup>H</sup>	1_NMAZ	2010	NM	Hidalgo	31°26'25.6''	-108°58'45.9''	SJS/ADG	<i>J. major</i>
3	H50	1644	15_TN	2011	TN	Knox	35°55'22.8''	-83°59'31.4''	SF	<i>J. nigra</i>
3	H53	1479 <sup>C</sup>	3_SwCA	2010	CA	Santa Barbara	34°36'28.5''	-120°21'08.5''	SJS	<i>J.californica</i>
3	H54	1570 <sup>H</sup>	1_NMAZ	2010	NM	Hidalgo	31°26'25.6''	-108°58'45.9''	SJS/ADG	<i>J. major</i>
4	H32	1299 <sup>**A</sup>	2_CeAZ	2009	AZ	Yavapai	34°55'25.1''	-112°50'34.7''	WC	<i>J. major</i>
4	H33	1234 <sup>**</sup>	2_CeAZ	2008	AZ	Yavapai	34°55'25.1''	-112°50'34.7''	WC	<i>J. major</i>
4	H33	1303 <sup>A</sup>	2_CeAZ	2009	AZ	Yavapai	34°55'25.1''	-112°50'34.7''	WC	<i>J. major</i>
4	H33	1307 <sup>A</sup>	2_CeAZ	2009	AZ	Yavapai	34°55'25.1''	-112°50'34.7''	WC	<i>J. major</i>
4	H33	1412	1_NMAZ	2009	AZ	Yavapai	34°55'25.1''	-112°50'34.7''	WC	<i>J. major</i>
4	H33	1550	2_CeAZ	2010	AZ	Yavapai	34°55'25.1''	-112°50'34.7''	WC	<i>J. major</i>
4	H33	1675	2_CeAZ	2010	AZ	Yavapai	34°55'25.1''	-112°50'34.7''	WC	<i>J. major</i>
-	-	1222	11_UT	2007	UT	Cache	41°55'1.74''	-111°48'48.81''	NT	<i>J. regia</i>
-	-	1228	5_NoCA	2008	CA	Yolo	38°32'21.4''	-121°47'46.4''	SJS	<i>J.californica</i>
-	-	1229	5_NoCA	2008	CA	Yolo	38°32'21.4''	-121°47'46.4''	SJS	<i>J.californica</i>
-	-	1347	OR_WA	2009	OR	Wasco	45°41'0.43''	-121°23'50.26''	JP	<i>Juglans sp.</i>
-	-	1348	3_SwCA	2009	CA	Los Angeles	34°30'48.2''	-118°37'04.6''	SJS	<i>J.californica</i>
-	-	1349	9_ORWA	2009	OR	Wasco	45°41'0.43''	-121°23'50.26''	JP	<i>Juglans sp.</i>
-	-	1359	13_WeCO	2009	CO	Mesa	39°03'49.93''	-108°33'2.33''	BH	<i>J. nigra</i>
-	-	1383	3_SwCA	2009	CA	Los Angeles	34°30'48.2''	-118°37'04.6''	SJS	<i>J.californica</i>
-	-	1405	4_CeCA	2009	CA	Kings	36°19'41.0''	-119°32'37.4''	SJS/ADG	<i>J. hindsii</i>

—	—	<u>1473</u> <sup>C</sup>	3_SwCA	2010	CA	Santa Barbara	34°36'28.5''	-120°21'08.5''	SJS	<i>J.californica</i>
—	—	<u>1528</u>	9_ORWA	2010	WA	Walla Walla	46°03'50.8''	-118°18'40.7''	SJS/CL	<i>J. nigra</i>
—	—	<u>1529</u>	9_ORWA	2010	WA	Walla Walla	46u03'50.8''	2118u18'40.7''	SJS/CL	<i>J. nigra</i>

\* Geographical regions are depicted in Fig. A4.1B. \*\* Isolates selected in the first trial of MLST analysis.

☆ Isolates deposited at Centraalbureau voor Schimmelcultures as CBS 124663 (1217) and CBS 124664 (1218). Isolates underlined were subjected to SSR analysis. <sup>A-P</sup>Common letters indicate isolates from different cankers of the same tree. Collectors are as follows: R. M. Bostock (RMB), T. W. Coleman (TWC), W. Cranshaw (WC), P.L. Dallara (PLD), G. Durham (GD), E. Fichtner (EF), S. Fraedrich (SF), A. D. Graves (ADG), K.J. Greby (KJG), B. Hammon (BH), J.E. Henrich (JEH), S.M. Hishinuma (SMH), D. Leatherman (DL), C. Leslie (CL), A. Liu (AL), J. McKenna (JM), L.M. Ohara (LMO), J. Pscheidt (JP), M. Putnam (MP), S. Schlarbaum (SS), S. J. Seybold (SJS), N. Tisserat (NT), C. Utley (CU), D.L. Wood (DLW). All isolations were made in the laboratory of NT with the exception of isolate 1513, which was made in the laboratory of RMB. doi:10.1371/journal.pone.0112847.t001

In order to obtain a haploid individual of *G. morbida*, all isolates were either single spored by serial dilution plating of conidia or by inducing the yeast phase by rapid shaking (250 rpm) of liquid culture, followed by serial dilution plating. A total of 209 haploid isolates were maintained on one-half strength potato dextrose agar (1/2 PDA, Difco Corp., Sparks, MD, USA) [10]. Of these isolates, 197 or 107 were assessed by MLST or SSR analysis, respectively, whereas 95 were tested by both methods (Table A4.1). In some cases, multiple *G. morbida* isolates were collected from different cankers on the same tree, or from more than one *Juglans* species from the same location (Table A4.1) to ascertain whether individual trees or adjacent trees contained more than one *G. morbida* haplotype.

#### DNA extraction

Single-spore isolates were grown on half-strength potato dextrose broth (Difco) for 3 days at 25°C with shaking (120 rpm). Mycelium was collected and lyophilized and DNA extraction was performed by using the Easy DNA Kit (Invitrogen, Carlsbad, CA, USA) or DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA), according to the manufacturer's recommendations. DNA concentrations and purities were estimated with a spectrophotometer (Thermo Scientific NanoDrop 1000, Loughborough, UK).

#### Isolation and identification of SSR sequences

DNA was extracted from two single-spore-derived cultures of isolates 1217 (CBS 124663) and 1218 (CBS 124664) (Table A4.1), known a priori to be characterized by different ITS sequences. DNA from both isolates was pooled and sent to the Cornell University Genomics Facility; enzymatically digested with HincII; ligated; and enriched for microsatellites by hybridization to probes containing random repeats. Enriched fragments were processed into a mate-pair library

and sequenced with 454 GS FLX (Roche, Indianapolis, IN, USA) sequencing technology. Sequences were assembled by using SeqMan Pro (Lasergene version 8.1.1; DNASTAR, Madison, WI, USA) yielding a total of 13,392 contigs and 21,737 singlets that were screened for the presence of microsatellite repeats by using EagleView software [31]. Primers were designed from a subset of selected contigs based on: presence of microsatellite repeats; putative presence of polymorphism within a contig; location of repeats within the sequence; and read quality. The software Primer 3 [32] was used to identify primer pairs from contigs. Primer pairs that produced a strong and consistent signal and produced polymorphic amplicons were selected and used to characterize DNA from designated isolates (Table A4.1, underlined isolates). Sequence data from loci used in this study were deposited in GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) (Table A4.S1). For all loci, polymerase chain reaction (PCR) protocol was 4 min denaturation at 94°C, followed by 25 cycles of 45 sec at 94°C, 45 sec at 56°C and 45 sec at 72°C, followed by 8 cycles of 10 sec at 94°C, 45 sec at 51°C and 45 sec at 72°C. At the end of the amplification, the samples were held at 72°C for 11 min and then at 4°C until they could be removed and stored at 220°C for dilution and mixing with ROX standards. Reactions included dNTP (0.3 mM each), bovine serum albumin (10 mg/ml) (New England Biolabs, Ipswich, MA, USA), in a final volume of 15 µl and all other reagents as described previously [33]. Microsatellite haplotypes were produced by using the three primer method [33], and alleles were binned by using Gene Mapper software (Applied Biosystems, South San Francisco, CA, USA). Quality control of microsatellite markers was guaranteed by genotyping the loci of all samples at least twice and independently. All genotyping plates contained DNA from three isolates (1217, 1218, and 1234), which were used as control (allelic standards).

Table A4.2. Primers tested on MLST analysis of *Geosmithia morbida* isolates.

Gene/region	Sequence (5' → 3')	Genome	Outcome
ITS	1F: CTTGGTCATTTAGAGGAAGTAA 4: TCCTCCGCTTATTGATATGC	Fungi	single band and SNPs
BT	1: AACATGCGTGAGATTGTAAGT Geo*: TCTCGACAAAACGTACCTCGT	<i>Fusarium</i>	single band and SNPs
FsACC	F: CTCGTGAGATCATGATCCAGT R: GTTGATAACAGCGGAGAGCT	<i>Fusarium</i>	multiple unspecific band
FsGPD	F: CATGTACGTCGTCGGTGTC R: CGCTTACTTGGAGGCATCG	<i>Fusarium</i>	multiple unspecific band
FsHMG	F: GGCAAGATTCCTGGTTACGC R: TTCATACCCATAGCGTCACC	<i>Fusarium</i>	did not amplify
FsICL	F: GGAGGTTGAGGCTGTCAAG R: GCTTGGTGAGCTTCATGACA	<i>Fusarium</i>	did not amplify
FsMPD	F: CGTCGAGAACACCATCACAAA R: ATGGGGGTTGCCAATTCGCT	<i>Fusarium</i>	did not amplify
FsSOD	F: TGGGACATCACCGGTAACGA R: CAGTCTTGAGAGACTCCTCG	<i>Fusarium</i>	did not amplify
FsTOP1	F: AGGAGCACATGACGACCAAG R: GATCCTGATCAGCCATGATC	<i>Fusarium</i>	multiple unspecific band
FsUGP1	F: CAGATGCGAAATGCTCTGAC R: AGGATATCGACGTTGTGGC	<i>Fusarium</i>	single band
Methionine aminopeptidase (MAP)	F: GCGAATAACGCTGCAATTCT R: AACCCGGAGTGACAACACTGAC	<i>G. morbida</i>	single band and SNPs

Ribosomal L18ae protein family	F: CTTGGTGTCTGCTTGGTGA R: ACCCCGAGAAGGTCAAGAAC	<i>G. morbida</i>	single band
Dolichyl-phosphate-mannose-protein mannosyltransferase	F: TCTTCTGGCTGTTTCATGACG R: CGAGGACACGGAAAATGAAG	<i>G. morbida</i>	single band
Amino acid permease	F: TATCAGCGCTTGCAAATACG R: GCAATCATGGAAATGTGTCG	<i>G. morbida</i>	single band
40S ribosomal protein S2	F: GCCCATCAAGGAGTACCAGA R: GACGTGTAGGCGTCTTCGAT	<i>G. morbida</i>	single band
Kinesin	F: GCTTCGCTACAGGTGAGTCC R: AGACTCCAGCGGTTGTCCTG	<i>G. morbida</i>	single band

\* BT Geo was designed based on *G. morbida* genome, inwardly oriented after amplification by using BT22 [36].  
doi:10.1371/journal.pone.0112847.t002

Locus/isolate combinations that produced inconsistent allele sizes or were null were considered as missing data. The ten amplicon sizes (corresponding to each SSR) of each isolate were entered into an Excel spreadsheet and compared visually (manually). Isolates sharing the same amplicon sizes were considered to be clonal.

#### Isolation and identification of MLST sequences

We attempted to amplify DNA from 16 genomic regions by using the following primer pairs: for housekeeping genes identified in the *G. morbida* genome; for those developed for *Fusarium solani* MLST analysis [34]; for the rRNA internal transcribed spacer (ITS) region [35]; and for the b-tubulin (BT) gene [36], to obtain a MLST-based analysis of *G. morbida* (Table A4.2).

Details on primer design for MLST sequences based on the *G. morbida* genome are provided in File S1.

Amplification of these 16 regions by PCR consisted of 10X Standard Taq Reaction Buffer (New England Biolabs), 0.2 U of Taq DNA Polymerase, 2 mM MgCl<sub>2</sub>, 0.2 mM of each dNTP, 0.25 mM of each primer, 20 ng of DNA, and sterile deionized water added for a final volume of 20 ml. Annealing temperatures were initially set 5°C below the lowest T<sub>m</sub> of the primer pairs.

Because some of the primers were not based on the *G. morbida* genome, we also performed gradient PCR with annealing temperatures 65°C of those set initially. Parameters were: 95°C for 5 min; followed by 35 cycles of 95°C for 30 sec, varying annealing temperatures based on the primer pair for 45 sec, 72°C for 1.5 min; and completed with 72°C for 5 min. The presence of amplicons was verified by using a Sub-Cell GT Cell (BioRad, Hercules, CA, USA) electrophoresis system. Amplified sequences that exhibited a single well-defined band were purified by using PureLink PCR Purification Kit (Invitrogen). Ten nanograms of DNA per 100

base pairs (bp) were combined with 10 pmol of primer and sequenced with the BigDye version 3.1 ready reaction kit (Applied Biosystems) on an ABI 3730 automated sequencer at the Proteomics and Metabolomics Facility at Colorado State University. Chromatograms were visualized by using Sequence Scanner Software 1.0 from Applied Biosystems (<http://www.appliedbiosystems.com/>) and bases were only accepted if the Phred quality score was equal to or greater than 20 [37]. Low quality reads and those containing overlapping sequences were either re-sequenced or sequenced with the primer flanking the opposite side. Sequences were edited and aligned by using BioEdit [38]. Only three of the 16 primer sets amplified genomic regions with SNPs. These resulted in 7, 13, and 7 polymorphisms in the ITS, BT, and methionine aminopeptidase (MAP) sequences (Figure A4.S1) of 197 *G. morbida* isolates, respectively (Table A4.2). PCR amplicons of the ITS, BT, and MAP region/genes were 565, 438, and 468 bp in length and the sequences were trimmed to 516, 363, and 365 bp, respectively, for MLST analysis. Sequences were deposited in GenBank with the following accession numbers: ITS, KJ148225 to KJ148419; BT, KJ148030 to KJ148224; and MAP, KJ148420 to KJ148614.

## Data Analyses

**Genetic structure and host specificity.** Haplotypes (isolates in the case of SSR data) were placed into genetic clusters based on the posterior probability of their allele frequencies by using STRUCTURE software [39] and based on multivariate methods by using Discriminant Analysis of Principal Components (DAPC) [40] in the Adegnet software package in R [41]. Haplotypes were assigned to the clusters based on independent analyses of the SSR and MLST data. SSR inputs for the analysis were the amplicon sizes of the ten microsatellite markers (Table A4.S1) from each isolate, whereas MLST inputs were each one of the polymorphic sites (SNPs and insertions/deletions) of the three genomic regions (Figure A4.S1). Even though some

polymorphic sites were consecutive and likely linked, each one of the 27 (as shown in results) was treated as a single locus (Figure A4.S1). The three MLST genomic regions were aligned by using BioEdit 7.2.4 [38]; concatenated (total length of 1,244 bp), and then combined into unique haplotypes (Table A4.S1A). Isolates containing identical haplotypes were considered as one, i.e., the dataset was clone- corrected.

A provisional genetic cluster assignment independent of geographical location was determined by using the Bayesian clustering algorithm implemented in the software STRUCTURE [39]. We used the admixture, correlated frequency model, and tested K (number of genetic clusters)=1 to 10. Parameters were estimated under the null hypothesis of panmixia where loci are at Hardy-Weinberg (H-W) equilibrium. Twenty independent runs were performed with a burn-in period and run length of 50,000 and 500,000 iterations, respectively. The optimal number of populations (K) was estimated by using Structure Harvester [42] according to the ad hoc statistic  $\Delta K$ , which is based on the rate of change in the log probability of the data between succeeding K values [43]. The assignment of each isolate/haplotype to a cluster was based on the quality threshold of q, which denotes the admixture proportion for each individual belonging to a cluster. In a situation where cluster x had the highest q-value for a determined individual, this individual was assigned to that cluster only if cluster x had a q-value of  $\geq 0.75$ , or if the q-value was  $\geq 0.5$  and twice the q-value of another cluster. Clones whose q-values did not fit these parameters were not assigned to any specific cluster.

Violations of the assumptions in STRUCTURE (panmixia, H- W equilibrium, linkage equilibrium in clonal subgroups) can produce incorrect assignments; therefore, the results from STRUCTURE were compared to a second genetic cluster assignment by using DAPC. DAPC optimizes variation among clusters to the detriment of variation within clusters and, contrary to

the results from STRUCTURE, is neutral to any *a priori* genetic hypothesis. The differentiation of individuals within and among populations was calculated by using Analysis of Molecular Variance (AMOVA) [44], which is based on a fixation index ( $F_{ST}$ ), using the Arlequin software package [45]. The resulting  $F_{ST}$  values were entered into a table and used to determine the most reliable genetic cluster model assignment for *G. morbida*.

Host specificity was also analyzed independently with both SSR and MLST data by using AMOVA. The analysis was only conducted for host species from which more than five *G. morbida*-isolates had been collected.  $F_{ST}$  P-values ( $\alpha=0.05$ ) were used to determine host specificity.

**Phylogenetic analyses, sexual reproduction/ recombination, neutrality, and linkage disequilibrium tests.** Due to the higher robustness of genetic clustering with the MLST data (see Results) and because more isolates were genotyped by using this approach, we performed most of the population genetic analyses with a focus on the MLST data.

Phylogenetic analysis showing the relatedness of *G. morbida* haplotypes and other *Geosmithia* species was conducted by using two concatenated genomic regions (ITS/BT). MAP was not included because of the absence of sequence data for other species in public databases. Fifty-five sequences were compared, six of which corresponded to other species of *Geosmithia* downloaded from GenBank (accessions are shown in Figure A4.2); the others corresponded to *G. morbida* haplotypes. Because MAP was not considered, some of the haplotypes sharing the same ITS and BT sequences were combined in a single leaf (e.g., H32 and H33, Figure A4.2). Sequences were aligned by using ClustalW [46] and trimmed in MEGA 5.05 [47], with lengths of 520 bp and 406 bp for ITS and BT, respectively. We performed two analyses: a Bayesian

inference of phylogeny with a variant of Markov chain Monte Carlo [48] in MrBayes v3.1.2 [49] by using a general time- reversible model with inverse-gamma rates of evolution for 1,500,000 generations and a burn-in of 0.25; and a maximum- likelihood analysis in PhyML 3.0 [50] (<http://www.atgc-montpellier.fr/phyml/>) with default values except for bootstrap of 1,000 replicates. The unrooted phylogenetic tree obtained in MrBayes was formatted in MEGA5 [47] and branches with bootstrap values equal or greater than 50% were shown. Bootstrap values greater than 500 obtained in maximum-likelihood analysis were also indicated in that tree.

The program GenClone 2.0 [51] was used to calculate  $p_{\text{gen}}$ , i.e., the likelihood that MLST haplotypes resulted from sexual reproduction. Neutrality of the polymorphic loci was tested based on coalescent simulations with the program DnaSP 5.10 [52]; 1,000 replicates were used to test Tajima's  $D$ ; Fu and Li's  $D$ ,  $D^*$ ,  $F$  and  $F^*$ ; and Fu's  $F$  neutrality statistics. MultiLocus 1.2 [53] was used to test linkage disequilibrium of alleles at different loci, and to test the null hypothesis of recombination. Indexes of association  $I_A$  and  $r$ -bar  $d$  were obtained with their corresponding probabilities by using 10,000 randomizations.

**Private alleles.** We used the most robust cluster assignment (four-cluster-MLST-DAPC model, as explained under “Genetic Structure”) to investigate private alleles in SSR analysis, i.e., alleles found in only one cluster or one group of interest. This analysis was performed manually.

**Geographical structure.** Genotype (or haplotype) diversity of each geographic region was also calculated by using DnaSP 5.10 [52], measuring the genetic variance of haplotypes located in the same region. Significance of the proposed geographic structures was subsequently tested with Hudson's Permtest, which computes the mean number of pairwise differences within ( $K^*_{ii}$  or  $K^*_{jj}$ ) and between groups ( $K^*_{ij}$ ) along with the probability that  $K^*_{ij}$  exceeds  $K^*_{ii}$  by chance [54]

(<http://wwwabi.snv.jussieu.fr/public/mpweb/>). Robustness of geographic clusters was calculated by using AMOVA in the Arlequin software package [44,45].

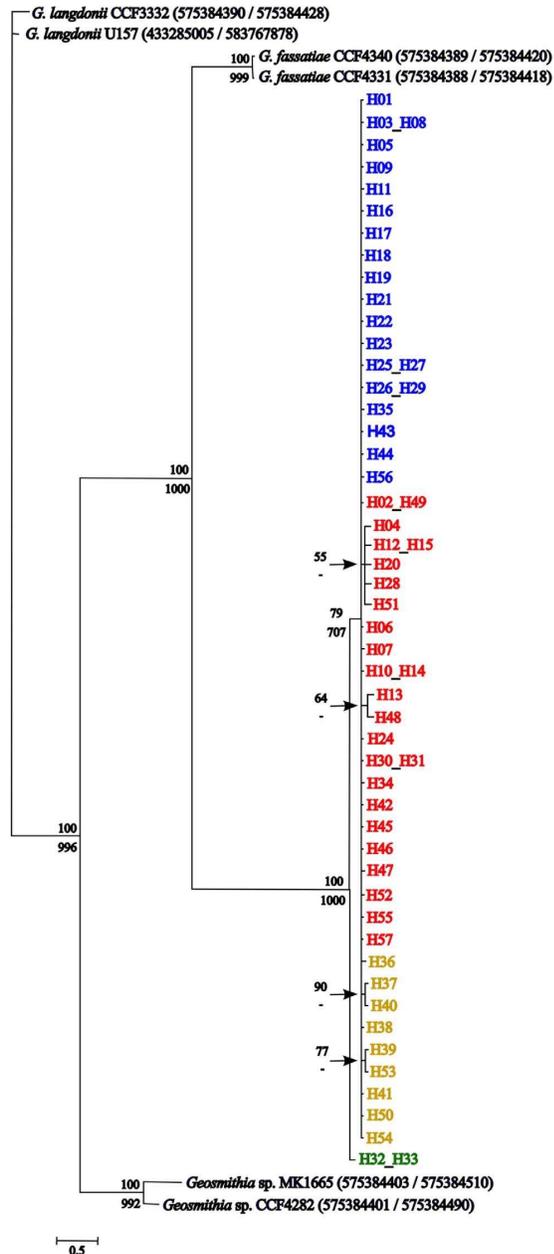


Figure A4.2. Unrooted phylogenetic tree of *Geosmithia* species based on ITS/BT sequences. A Bayesian analysis was performed for 1,500,000 generations by using a GTR-gamma distributed model of evolution (invariant sites). Bayesian percentages ( $\geq 50\%$ ) are depicted above each branch, and maximum likelihood bootstrap values ( $\geq 500$ ) obtained by using PhyML (default parameters) are shown below most branches. *Geosmithia morbida* haplotypes are color coded according to their genetic cluster assignment (four-cluster-MLST-DAPC model, as in Figure A4.3) and haplotypes sharing the same ITS and BT sequences are co-located. Leaves pertaining

to the same branch were arranged together according to their cluster assignment. GenBank accession numbers of other *Geosmithia* spp. are identified within parenthesis.

doi:10.1371/journal.pone.0112847.g002

## Results

### Isolation and identification of SSR sequences

Ten microsatellite (SSR) loci revealed a total of 59 alleles in *G. morbida* (Table A4.S1).

Amplicon sizes were compared, and two isolates (1348 and 1352) were identified as clonal because all but one amplicon from these isolates had the same size. Loci 2514 and 3416 failed to amplify in an unusually high percentage of samples (75% and 63%, respectively), from which other loci amplified consistently, which may indicate the presence of null alleles at those loci (Table A4.S2).

### Isolation and identification of MLST sequences

From the set of 16 pairs of primers tested with *G. morbida* DNA as template (Table A4.2) all eight based on the *Fusarium solani* genome [34] failed to amplify a single well-defined band. The six pairs of primers designed based on the *G. morbida* genome produced a single amplicon (Table A4.2), and one pair corresponded to a polymorphic region (MAP) when the sequences of eight experimental isolates were compared (Table A4.1, isolates with two asterisks). ITS and BT produced polymorphic sequences [55], as reported in other species of *Geosmithia* [56], and were therefore used in this population genetic study. BT was originally amplified by using a pair of primers based on *Fusarium* spp. genome [36], but after sequencing the corresponding fragment from *G. morbida* one primer was refined accordingly (Table A4.2). We did not detect sequences

with consistent noise or base call overlap, which would be an indication of different variants (alleles) of multi-copy genes (or genomic regions) within the same individual.

### Haplotype assignment

Using MLST data, 197 *G. morbida* isolates were placed into one of 57 haplotypes based on 11, 16, and 4 variants of ITS, BT, and MAP, respectively (Table A4.S2 and Table A4.1), according to the combination of 7, 13, and 7 polymorphic loci, respectively (Figure A4.S1). Multiple haplotypes were identified from isolates collected from different cankers on the same tree in cases from AZ, CA, CO, NM, TN, and VA (Table A4.1). Twenty-four of the haplotypes (42%) were represented by more than one isolate (Table A4.1). Overall, haplotypes H02 and H03 occurred the most frequently in the survey, representing 36% of all *G. morbida* isolates examined (Figure A4.3). They were detected in all geographic regions where more than 4 isolates were collected, except neither haplotype occurred in the native range of *J. major* (central AZ and NM\_AZ), and haplotype H03 did not occur in southern CO (Figure A4.3). Haplotype H02 was especially abundant in the three geographic regions in CA (southwestern, central, and northern). Haplotype H03 was also abundant in northern CA and central CA but less so in the native range of *J. californica* in southwestern CA. Most of the geographic regions shared haplotypes with other regions, with the exception of central AZ and NM\_AZ. In fact, most of the haplotypes present in those two regions were either exclusive or only shared between them, and all isolates from those locations were collected from *J. major* (Figure A4.3). The exceptions were haplotypes H38 and H40, which were also collected from a single county (Jefferson County) in northern CO, but from *J. nigra*. Isolate 1503 (Table A4.1) was not grouped with the other NM\_AZ isolates because it was collected from *J. nigra* in an urban planting in Albuquerque (NM) and carried a haplotype that was not found in the native range of *J. major*.

The SSR data was not organized into haplotypes because no repeated haplotypes were identified when the allele sizes of the ten loci were analyzed, with the exception of isolates 1348 and 1352, as described above.

#### Data analyses

**Genetic structure.** According to analysis with STRUCTURE, 107 *G. morbida* isolates were best organized into two clusters based on SSR allele frequency with a smaller  $\Delta K$ -peak at 4 (Figure A4.S2). Similarly, the clone-corrected MLST data from 57 haplotypes resulted in a  $\Delta K$  peak at 4 following analysis with STRUCTURE (Figure A4.S2).

Assignment of the SSR data to genetic clusters by using DAPC resulted in an optimal theoretical  $K = 6$  (Figure A4.4A). One cluster containing only three isolates, 1234, 1303, and 1307 (in green in Figure A4.4A) segregated from a group of five clusters. When the three distant isolates were removed and data re-analyzed, the remaining isolates were re-organized into three clusters (Figure A4.4B). Therefore, in further SSR-DAPC analysis we considered a  $K = 4$ , comprising the one cluster containing the three distant isolates (in green in Figure A4.4A) as well as the three additional clusters in Figure A4.4B. The fixation index of  $K = 4$  was similar to the six-clusters-SSR-DAPC model obtained originally (Table A4.3).

The MLST data analyzed with DAPC was first assigned to two clusters containing 55 and 2 haplotypes each (Figure A4.5A). When the distant cluster containing haplotypes H32 and H33 was excluded (in green in Figure A4.5A), the remaining 55 haplotypes segregated into three distinct clusters (Figure A4.5B). Fixation indices were similar for both two and four MLST-DAPC-clusters (Table A4.3). When the cluster assignment of the haplotypes on the four-cluster-MLST-STRUCTURE model was compared to the four-cluster-MLST-DAPC, they were mostly

correlated, despite some discrepancies (Figure A4.5A, green cluster; and Figure A4.5B, all clusters). The four-cluster-MLST-DAPC model had a high (>0.25) fixation index and it was the most informative when compared to other models of genetic and geographic (as described below) clustering of *G. morbida* isolates. Therefore, the four-cluster-MLST-DAPC model was used to assign haplotypes to clusters in further analyses (Table A4.1 and Figure A4.3).

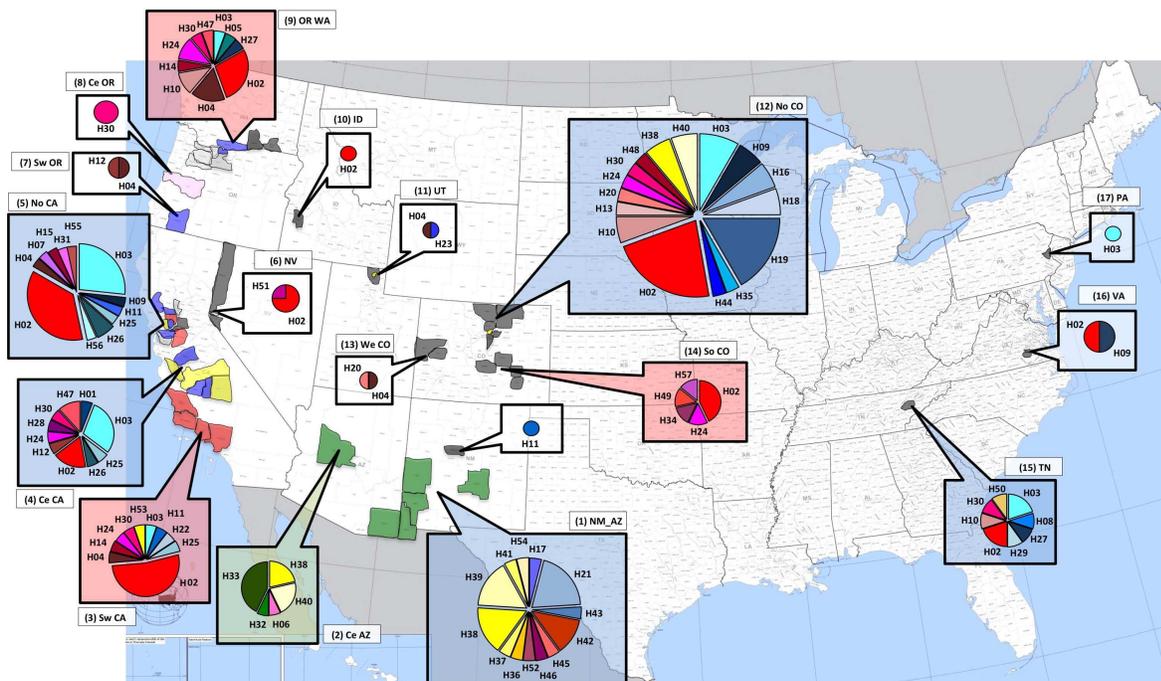


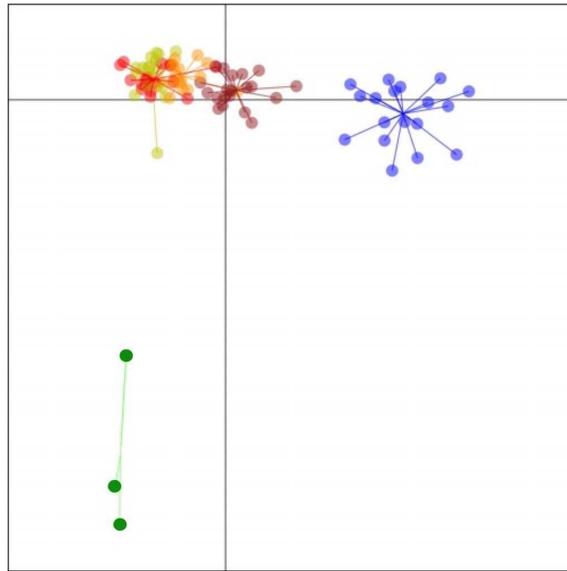
Figure A4.3. Distribution of 57 MLST-based *Geosmithia morbida* haplotypes in the United States. The size of wedges in each pie chart is proportional to the number of isolates. Haplotype colors relate to genetic clusters identified in the four-cluster-MLST-DAPC model where cluster 1 = shades of blue, cluster 2 = shades of red/brown, cluster 3 = shades of yellow, and cluster 4 = shades of green. Callouts are color-coded according to the three-region geographic-Hudson's Permtest model, where: 1) blue = NM\_AZ, central CA, northern CA, northern CO, and TN; 2) green = central AZ; and 3) red = southwestern CA, OR\_WA, and southern CO. Callouts in white indicate regions not assessed by using Hudson's Permtest. Counties are color-coded as in Figure A4.1B. (U.S. map adapted from US Census Bureau at <https://www.census.gov/>). doi:10.1371/journal.pone.0112847.g003

Examining the haplotype assignment, based on the four-cluster- MLST-DAPC model, cluster 1 (blue color in pie charts of Figure A4.3) contained 21 haplotypes (63 isolates) including the

abundant and widely distributed haplotype H03 (Table A4.1 and Figure A4.3). Cluster 1 was especially abundant in northern CO, northern CA, central CA and TN. Cluster 2 (red in pie charts, Figure A4.3) contained 25 haplotypes (103 isolates), including the abundant and widely distributed haplotype H02. Cluster 2 was especially abundant in the three regions in CA, OR\_WA, and in southern CO, where it accounted for more than half of the individuals (Table A4.1 and Figure A4.3). Cluster 3 (yellow/amber in pie charts, Figure A4.3) was represented by 9 haplotypes (24 isolates) and was abundant in central AZ and NM\_AZ, but not present in most of the other regions, except for two haplotypes in northern CO and one in southwestern CA and TN (Table A4.1 and Figure A4.3). Cluster 4 (green in pie charts, Figure A4.3) contained 2 haplotypes (7 isolates) that were only detected from *J. major* in central AZ (Table A4.1 and Figure A4.3).

We found correspondence in cluster assignment between MLST and SSR data by using DAPC (Figure A4.4A, A4.4B, A4.5A, and A4.5B). The outliers (isolates) 1234, 1303, and 1307 in SSR analysis (Figure A4.4A) corresponded to H33 (Figure A4.5A), which is also one of the outliers in MLST analysis. The single isolate (1299) pertaining to H32 (Figure A4.5A) was not assessed by the SSR method. Most of the isolates (75%) in SSR-cluster 1 corresponded to cluster 2 in MLST analysis, and most of the isolates (47%) in SSR-cluster 3 corresponded to cluster 3 in MLST analysis. The SSR-cluster 2 was comprised of 67% of MLST-cluster 2 and 33% of MLST-cluster 1 isolates.

A



B

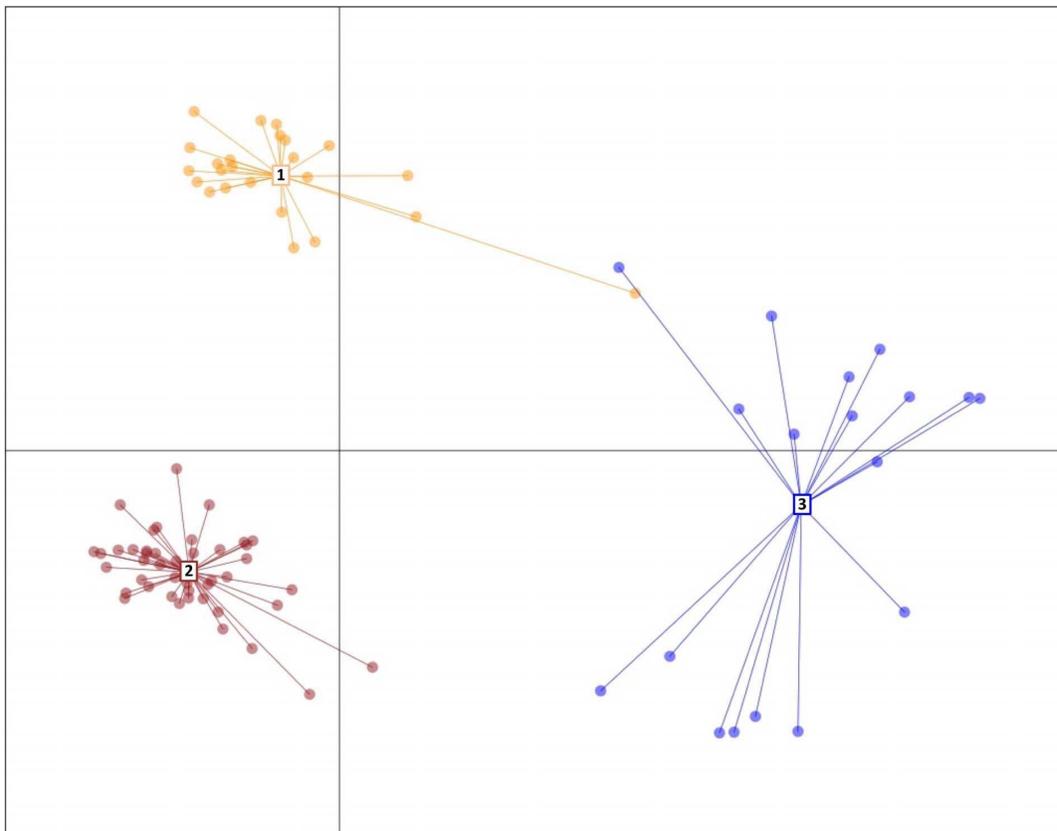


Figure A4.4. Coordinates of the SSR profile of 112 isolates of *G. morbida* based on DAPC analysis: of a theoretical  $K = 6$  (A); and the same, but excluding the distant isolates in green, which resulted in  $K = 3$  (B). doi:10.1371/journal.pone.0112847.g004

**Host specificity.** Based on pairwise  $F_{ST}$  analysis, the SSR and MLST data indicated that *J. major* isolates were different from *J. californica*, *J. hindsii*, *J. nigra* and *J. regia* isolates collectively, whereas *J. nigra* isolates were different from *J. californica* isolates (Table A4.S3).

**Phylogenetic analyses.** In a phylogenetic tree based on Bayesian inference with ITS and BT sequences (Figure A4.2), all 57 haplotypes of *G. morbida* (Table A4.S2) grouped in the same clade with strong bootstrap support, apart from several other *Geosmithia* species. Within the *G. morbida* clade, isolates in the haplotypes H32 and H33 formed their own sub-clade with strong bootstrap support and the remaining haplotypes clustered together into a second sub-clade.

**Sexual reproduction/recombination, neutrality, and linkage disequilibrium tests.**

Independent analyses of both the whole set of 57 haplotypes and haplotypes assigned to MLST-DAPC clusters showed that the likelihood of the observed haplotypes arising from sexual reproduction was extremely low ( $p_{\text{gen}} < 0.05$ ). In both the complete dataset and clone-corrected data, the  $r_d$  test rejected the null hypothesis of recombination in MLST loci ( $P < 0.001$ ) and we obtained similar results by using SSR data ( $P < 0.001$ ). The polymorphic MLST-loci were all selectively neutral, whereas polymorphisms in SSR loci were presumed neutral as they were extra-genic. Overall, the three genomic regions used in the MLST analysis were in linkage disequilibrium (index of association = 0.8005,  $P < 0.001$ ) and pairwise tests indicated the same (index of association = 0.8005,  $P = 0.029$ ).

**Private alleles.** Private alleles may indicate restrictions in dispersal or evolutionary isolation. The four-cluster-MLST-DAPC model was used to investigate private alleles in SSR analysis (Table A4.S2). In this regard, isolates in cluster 4, showed a high number of private alleles and included private alleles at seven of ten loci examined. Moreover, the most common allele at

locus 194 (270 bp), locus 2849 (310 bp and 316), and locus 4045 (184 bp) were completely absent from isolates of cluster 4, reinforcing their genetic distinctiveness. Isolates belonging to cluster 3 showed private alleles at SSR loci 223, 194, 1851, and 6823; cluster 2 had private alleles at 2849, 3416 and 7713, and cluster 1 at 4045, 6823 and 7713 (Table A4.S2).

**Geographical structure.** *Geosmithia morbida* isolates were placed initially into one of 17 geographic regions based on the proximity of the counties where the isolates were collected. These regions, while arbitrarily determined, were for the most part spatially isolated from one another (Figure A4.1B). Even though the fixation index ( $F_{ST} = 0.167$ ) (Table A4.3) was not low when the 17 regions were tested, the population structure in some regions was not different from others according to Hudson's Permtest [54]. Because of that, a series of Hudson's test-based pairwise comparisons of 8 and then 6 regions (regions that we compared are described in Table A4.3) that included more than four isolates indicated that the populations could be combined into three different ( $P, 0.001$ ) macro regions encompassing: 1) NM\_AZ, central CA, northern CA, northern CO, and TN; 2) central AZ and 3) southwestern CA, OR\_WA, and southern CO (depicted in callouts in Figure A4.3). Therefore the three-region geographic model was the most robust when compared to other geographic models. This conclusion is based on fixation indices ( $F_{ST}$ ) that identified the lowest within and highest among population variation (Table A4.3), and all pairwise  $F_{ST}$  values were significant (Table A4.4). The haplotype diversity was high across all three macro regions [ $0.70$  ( $SD \pm 0.02$ )]. The values obtained for the three regions individually were  $0.67$  ( $SD \pm 0.03$ ),  $0.77$  ( $SD \pm 0.08$ ), and  $0.54$  ( $SD \pm 0.08$ ) in: 1) NM\_AZ/Ce CA/No CA/No CO/TN; 2) Ce AZ; and 3) Sw CA/OR\_WA/So CO, respectively.

Table A4.3. *Geosmithia morbida* molecular variance determined by AMOVA of Bayesian (Structure), DAPC, and Hudson’s Permtest analyses.

Number of clusters	Fixation indices ( $F_{ST}$ )	Variation within/among
Genetic clusters (data/test)		
4 (MLST/Bayesian)	0.515	48.51/51.49
2 (MLST/Bayesian)	0.253	74.73/25.27
6 (SSR/DAPC)	0.425	57.50/42.50
4 (SSR/DAPC)	0.461	46.08/53.92
2 (MLST/DAPC)	0.739	26.13/73.87
4 (MLST/DAPC)	0.612	38.81/61.19
Geographic clusters		
17 regions	0.167	83.31/16.69
8 regions*	0.198	80.16/19.84
6 regions**	0.212	78.78/21.22
3 regions***	0.248	75.23/24.77

\* 1) central AZ; 2) NM\_AZ; 3) northern and central CA; 4) southwestern CA; 5) OR\_WA; 6) northern CO; 7) southern CO; and 8) TN. \*\* 1) central AZ; 2) NM\_AZ; 3) northern and central CA and northern CO; 4) southwestern CA; 5) OR\_WA and southern CO; and 6) TN. \*\*\* The three “macro” regions were: 1) NM\_AZ, central CA, northern CA, northern CO, and TN; 2) central AZ and 3) southwestern CA, OR\_WA, and southern CO. doi:10.1371/journal.pone.0112847.t003

## Discussion

### Isolates and genetic diversity

*Geosmithia morbida* was isolated consistently from cankers surrounding WTB galleries from all *Juglans* and *Pterocarya* showing TCD symptoms and in all regions that we surveyed (Fig.

A4.1A), with the exception of *J. major* in Texas where the WTB was not detected [23].

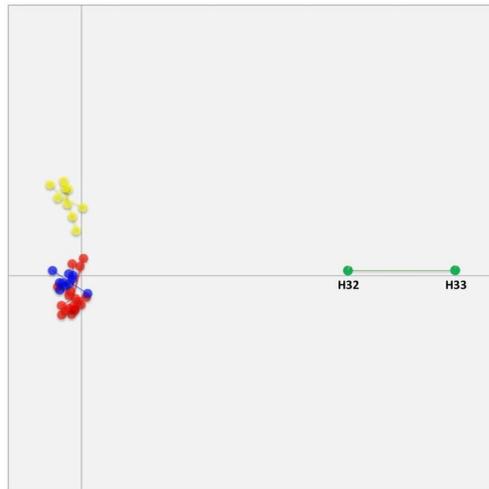
Considering the substrate specificity of bark beetles and *Geosmithia* associates to their respective plant hosts [56], and the highly diverse and complex genetic structure in *G. morbida* that we and others [10,26,27] have observed, it is unlikely that the current TCD epidemic was a result of a

new association between the WTB and *G. morbida*.

The number of polymorphic MLST loci described here and elsewhere [10] was not low, as we had first expected. However, it is unlikely that haplotypes are admixed (sexual recombinants) because no teleomorph has been observed and according to tests for sexual reproduction, neutrality, and disequilibrium, sexual recombination in *G. morbida* is nonexistent or infrequent. Considerable intra-species variation of the ITS rDNA region has been reported in other *Geosmithia* species [10,56], and this may be a common characteristic of this genus. Thus, this variability does not necessarily support the hypothesis that *G. morbida* is a species complex.

All isolates were identified initially as *G. morbida* based on the similarity of their morphological characteristics and their growth patterns and color in culture. However, the placement of haplotypes H32 and H33 in a group distinct from the other *G. morbida* haplotypes according to the ITS/BT-based phylogenetic tree and based on analyses of genetic and geographic cluster assignments could be viewed as evidence for a hypothetical cryptic species (Figure A4.2, 4A, and 5A). All isolates characterized by H32 and H33 were derived from *J. major*, whereas the other haplotypes were collected from many *Juglans* and *Pterocarya* species. However, in previous work, our team has found that isolate 1234 (characterized by H33 and first documented from *J. major* in Kolarčík et al. 2011 [11]) is indistinguishable from other *G. morbida* isolates in terms of its morphology and pathogenicity. A more thorough multi-gene phylogenetic analysis in the future including other species of *Geosmithia* and more isolates of *G. morbida* from *J. major* in the southwestern USA and Mexico might be necessary to investigate the potential for cryptic species related to *G. morbida*.

A



B

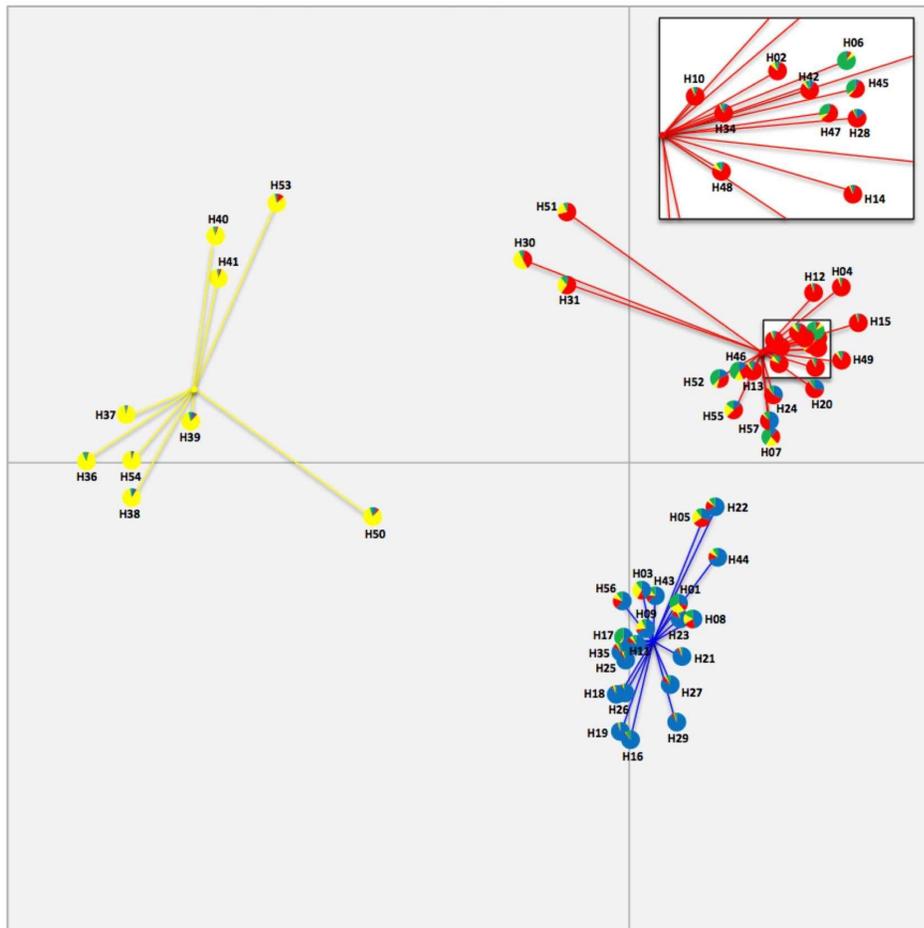


Figure A4.5. Coordinates of 57 (A) and 55 (B) haplotypes of *Geosmithia morbida* from the MLST-DAPC model. The most distant cluster (cluster 4 in green) comprised of haplotypes H32 and H33 is identified (A), as well as the coordinates of all remaining haplotypes when H32 and H33 were excluded (B). A comparison between the assignments of the MLST-DAPC and MLST-STRUCTURE models are shown in detail. Pie charts give the probability of assignment

of haplotypes to the four genetic clusters obtained in the four-clusters-MLST-STRUCTURE model. They are represented by colors, cluster 1 = blue, cluster 2 = red, cluster 3 = yellow and cluster 4 = green. Haplotypes in the box (in B) were amplified for better resolution.  
doi:10.1371/journal.pone.0112847.g005

#### Genetic, geographic, and host groupings

Our results indicated that the *G. morbida* genetic structure in the USA was best explained by four distinct genetic groups. Cluster assignment correspondence between MLST and SSR by using the DAPC method was moderate, probably due to some missing data in SSR (Table A4.S2). The most robust analysis was achieved by DAPC analysis with MLST data (Figure A4.5A, B and Table A4.3), with support of STRUCTURE. The four-cluster- MLST-DAPC model differed from the two clusters identified by Hadziabdic and collaborators [27], probably because we sampled more extensively and in areas not sampled in their study.

Haplotypes in genetic clusters 1 and 2 (blue and red) dominated in locations outside of AZ and NM (i.e., outside the native range of *J. major*); vice versa for clusters 3 and 4 (yellow and green) (Figure A4.3). The high proportion of widely-dispersed, closely-related *G. morbida* haplotypes supports the hypothesis of a recent invasion by at least three genetic groups into regions other than central AZ and NM\_AZ (Figure A4.3).

The haplotypes were further grouped into three diverse geographic groups (callouts in Figure A4.3), possibly indicating that regions within the same group exchanged or had a common source of haplotypes. A comparable genetic and geographic structure was observed for *Ophiognomonium clavignenti-juglandacearum*, an exotic canker pathogen of *J. cinerea* in North America [57], which consisted of three genetic clusters of haplotypes that were geographically disjunct. In the past, many epidemics caused by invasive plants, pathogens, or insects were

assumed to be initiated by a single introduction event followed by a radiation from the introduction point. However, several recent studies have found that this is likely not the case and that, in most invasion events, multiple introductions have occurred followed by migration and admixture of populations [58–61]. This may be particularly true for forest pathogens as damage caused by these organisms may go unnoticed and undetected for decades in more remote areas. This allows greater migration and genetic drift to occur between and among populations. Based on the amount of genetic diversity observed in the ITS, BT, and MAP regions and the fact that *G. morbida* is present but does not cause mortality in some *Juglans* species, it is likely that this fungus was present in at least one location in North America for decades, being moved virtually unnoticed amongst *Juglans* populations by the WTB. However, once mortality on *J. nigra* was observed out of its natural range and the fungus was detected and described for the first time, the full extent of the distribution of *G. morbida* was finally realized [4,9,11].

Although we found some evidence of host specialization, these data were strongly influenced by the uneven geographic distribution of the *Juglans* species, i.e., host and region were highly confounded. *Geosmithia morbida* isolates from *J. californica* and *J. major* were derived almost exclusively from trees in or near their native ranges, whereas isolates from other hosts were collected in multiple regions and from planted trees. Isolates from the Eurasian native, *J. regia*, were derived primarily from adventive orchards in CA and two outlying samples from CO and UT. Moreover, the number of isolates collected from the hosts was extremely unbalanced. Further studies that extend the work of Utley *et al.* [62] by evaluating virulence of selected isolates on various hosts are needed to confirm any evidence of host specialization.

Table A4.4. Pairwise  $F_{ST}$  values calculated for the three-population geographical model observed with Hudson's Permtest.

Geographic cluster	1	2	3
1	–	0.423*	0.109*
2	0.423*	–	0.491*

Significant ( $P < 0.05$ ) values are denoted by (\*). Geographical regions: 1) NM\_AZ/Ce CA/No CA/No CO/TN; 2) Ce AZ; and 3) Sw CA/OR\_WA/So CO. doi:10.1371/journal.pone.0112847.t004

### Introduction of multiple *G. morbida* haplotypes to geographically isolated areas

We recovered multiple *G. morbida* haplotypes representing different genetic clusters from different cankers on the same tree; their introduction could have occurred in different events or all at once from a single WTB-infested log. Even small infested logs harbor large numbers of beetles [63]. The WTB is not considered a strong flier [25] and even though *G. morbida* produces dry conidia, they are unlikely to be dispersed by the wind because the spores are only formed in beetle galleries and feeding sites [62]. Therefore, the most probable means of *G. morbida* introduction into geographically isolated areas is by importation of contaminated logs or wood products (with bark attached) [25]. The anthropogenic movement of wood containing many haplotypes would explain the complex mixture of *G. morbida* haplotypes observed at most locations.

The high proportions of haplotypes H02 and H03 in most of the regions may represent their relative abundance in the original or secondary invasive population; the frequency of haplotype reintroduction during subsequent dispersal events; or it could indicate some type of competitive advantage [26,55].

Possible origin of *G. morbida* isolates in the TCD epidemic: *J. major* native range as primary

source

We hypothesized that *G. morbida* was dispersed from the native range of *J. major*, where the WTB vector was first collected [7,21]. There is circumstantial evidence of an isolated introduction of *G. morbida* (H38 and H40) from central AZ into northern CO (yellow in pie charts in Figure A4.3), but these haplotypes represent a small fraction of the total number. Thus, the two haplotypes were probably part of an independent and secondary introduction of *G. morbida* into northern CO.

In general, we did not detect haplotypes collected from *J. major* in AZ and NM in other areas in the USA. Although we did not find shared haplotypes, the Hudson's Permtest [54] analysis indicated that the genetic makeup of haplotypes from NM\_AZ was not significantly different from those found in central CA, northern CA, northern CO, and TN (blue callouts in Figure A4.3). The high number of exclusive *G. morbida* haplotypes collected from *J. major* in central AZ and NM\_AZ (Figure A4.3) and the large number of private alleles, combined with the absence in this region of the two most frequently occurring haplotypes (H02 and H03), make it unlikely that these two regions are the direct geographic sources of *G. morbida* isolates associated with TCD outbreaks in the majority of locations. *Juglans major* has an extensive but non-contiguous range that extends south into Mexico [64], and our sampling was restricted to the extreme northern portion of the full range (Figure A4.1A) of this species. If *G. morbida* is present throughout the range of *J. major*, its population structure may be variable because of genetic isolation. Even within central AZ and NM\_AZ, we noted what appeared to be geographically isolated genetic clusters of *G. morbida* (Table A4.1). A more intensive survey in unsampled regions of *J. major* is warranted.

Possible origin of *G. morbida* isolates in the TCD epidemic: *J. californica* native range as primary source

Another possible source of *G. morbida* haplotypes causing the TCD epidemic could be southwestern CA. All *G. morbida* isolates from this region were collected from *J. californica*, a species that may be a native host of the WTB and its symbiont *G. morbida* [7]. Moreover, a very high proportion (89%) of the haplotypes identified in southwestern CA, in particular haplotypes H02 and H03, was also found in at least one other region (Figure A4.3). Furthermore, the result of the geographic structure analysis indicated that southwestern CA was similar to OR\_WA and southern CO (Figure A4.3).

Geographic structure analysis also indicated genetic similarities among *G. morbida* haplotypes in northern CO, TN, central CA and northern CA (Figure A4.3), with the latter region representing the natural range of *J. hindsii* (Figure A4.1A). Central CA and northern CA shared 88% and 83% of their haplotypes with other regions (Figure A4.3). It is possible that the TCD epidemic was triggered when infested raw walnut wood products were moved from northern CA to northern CO, or vice versa.

#### *Geosmithia morbida* dispersal

Regardless of its origin, TCD was first noted after its appearance on the highly susceptible *J. nigra* [4,10], probably as a consequence of transportation of infested logs that allowed movement of WTB and *G. morbida* into regions that they would not have reached naturally. Introduction to new sites could also have been facilitated by tourists/campers carrying even small amounts of wood. For example, there were more than 13 million visitors to Arizona's State and National Parks from 2001–2012 ([www.azot.gov](http://www.azot.gov)). Visitor attendance during this period to California's

State Parks alone was even higher, ranging from 85.5 million (2001–2002) to 67.9 million (2011–2012) ([http://www.parks.ca.gov/?page\\_id=23308](http://www.parks.ca.gov/?page_id=23308)). The genetic makeup of the *G. morbida* populations in regions where *J. nigra* has been affected suggests that the outbreaks in most of the USA are related to those in CA and from a fungus population that was already established in at least one species of *Juglans*.

## **Conclusion**

This study helped to explain the genetic structure of *G. morbida* and its multi-factorial and disjunct distribution in the USA. Even though the observation of TCD in *Juglans* and *Pterocarya* species has only been recent, the observed haplotype diversity and the genetic complexity of *G. morbida* indicate it has been in association with at least one *Juglans* spp. and the WTB for a long period of time. Some haplotypes and genetic clusters were found in specific regions and in association only with certain *Juglans* species. The scattered geographic distribution of genetic clusters indicates that *G. morbida* was disseminated several times and from several sources and most likely by transportation in beetle-infested wood.

## Supporting Information

### ITS

Group09 CAGTTGCTTCGGCGGCCCCGCCCGGGGCGCCTCGGGA**CCCCCCCC**-TAGGGGG-CCCCG 58  
Group06 CAGTTGCTTCGGCGGCCCCGCCCGGGGCGCCTCGGGA**TCCCCCCCC**CTAGGGGG-CCCCG 59  
Group01 CAGTTGCTTCGGCGGCCCCGCCCGGGGCGCCTCGGGA**TCCCCCCCC**-TAGGGGG-CCCCG 58  
Group07 CAGTTGCTTCGGCGGCCCCGCCCGGGGCGCCTCGGGA**TCCCCCCCC**TAGGGGG-CCCCG 59  
Group04 CAGTTGCTTCGGCGGCCCCGCCCGGGGCGCCTCGGGA**TC**--CCCC-**TAGGGGG**-CCCCG 56  
Group10 CAGTTGCTTCGGCGGCCCCGCCCGGGGCGCCTCGGGA**TCCCCCCCC**-TAGGGGG**G**CCCCG 59  
Group08 CAGTTGCTTCGGCGGCCCCGCCCGGGGCGCCTCGGGA**TCTCCCCCCCC**-TAGGGGG**G**CCCCG 59  
Group02 CAGTTGCTTCGGCGGCCCCGCCCGGGGCGCCTCGGGA**TCTCCCCCCCC**-TAGGGGG-CCCCG 58  
Group03 CAGTTGCTTCGGCGGCCCCGCCCGGGGCGCCTCGGGA**TC**--CCCC-**TAGGGGG**-CCCCG 56  
Group05 CAGTTGCTTCGGCGGCCCCGCCCGGGGCGCCTCGGGA**TC**-**CCCCCCCC**-TAGGGGG**G**CCCCG 58  
Group11 CAGTTGCTTCGGCGGCCCCGCCCGGGGCGCCTCGGGA**TCTCCCCCCCC**-TAGGGGG-CCCCG 58  
**Loci** \*\*\*\*\*1\*23\*\*\*\*\*4\*\*\*\*\*5\*\*\*\*\*

Group09 CGGGCGACCCGGACCCAGGCGACCCCGGGGGACCCAAACTCTT-GTCTTTTAGTGTATT 117  
Group06 CGGGCGACCCGGACCCAGGCGACCCCGGGGGACCCAAACTCTT-GTCTTTTAGTGTATT 118  
Group01 CGGGCGACCCGGACCCAGGCGACCCCGGGGGACCCAAACTCTT-GTCTTTTAGTGTATT 117  
Group07 CGGGCGACCCGGACCCAGGCGACCCCGGGGGACCCAAACTCTT-GTCTTTTAGTGTATT 118  
Group04 CGGGCGACCCGGACCCAGGCGACCCCGGGGGACCCAAACTCTT-GTCTTTTAGTGTATT 115  
Group10 CGGGCGACCCGGACCCAGGCGACCCCGGGGGACCCAAACTCTT-GTCTTTTAGTGTATT 118  
Group08 CGGGCGACCCGGACCCAGGCGACCCCGGGGGACCCAAACTCTT-GTCTTTTAGTGTATT 118  
Group02 CGGGCGACCCGGACCCAGGCGACCCCGGGGGACCCAAACTCTT-GTCTTTTAGTGTATT 117  
Group03 CGGGCGACCCGGACCCAGGCGACCCCGGGGGACCCAAACTCTT**T**GTCTTTTAGTGTATT 116  
Group05 CGGGCGACCCGGACCCAGGCGACCCCGGGGGACCCAAACTCTT-GTCTTTTAGTGTATT 117  
Group11 CGGGCGACCCGGACCCAGGCGACCCCGGGGGACCCAAACTCTT-GTCTTTTAGTGTATT 117  
**Loci** \*\*\*\*\*6\*\*\*\*\*

Group09 ATCTGAGTGGCATAAGCAATAAAATTAATAACTTTCAGCAACGGATCTCTTGGTTCTGGCA 177  
Group06 ATCTGAGTGGCATAAGCAATAAAATTAATAACTTTCAGCAACGGATCTCTTGGTTCTGGCA 178  
Group01 ATCTGAGTGGCATAAGCAATAAAATTAATAACTTTCAGCAACGGATCTCTTGGTTCTGGCA 177  
Group07 ATCTGAGTGGCATAAGCAATAAAATTAATAACTTTCAGCAACGGATCTCTTGGTTCTGGCA 178  
Group04 ATCTGAGTGGCATAAGCAATAAAATTAATAACTTTCAGCAACGGATCTCTTGGTTCTGGCA 175  
Group10 ATCTGAGTGGCATAAGCAATAAAATTAATAACTTTCAGCAACGGATCTCTTGGTTCTGGCA 178  
Group08 ATCTGAGTGGCATAAGCAATAAAATTAATAACTTTCAGCAACGGATCTCTTGGTTCTGGCA 178  
Group02 ATCTGAGTGGCATAAGCAATAAAATTAATAACTTTCAGCAACGGATCTCTTGGTTCTGGCA 177  
Group03 ATCTGAGTGGCATAAGCAATAAAATTAATAACTTTCAGCAACGGATCTCTTGGTTCTGGCA 176  
Group05 ATCTGAGTGGCATAAGCAATAAAATTAATAACTTTCAGCAACGGATCTCTTGGTTCTGGCA 177  
Group11 ATCTGAGTGGCATAAGCAATAAAATTAATAACTTTCAGCAACGGATCTCTTGGTTCTGGCA 177  
\*\*\*\*\*

Group09 TCGATGAAGAACCGCAGCAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAATCAT 237  
Group06 TCGATGAAGAACCGCAGCAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAATCAT 238  
Group01 TCGATGAAGAACCGCAGCAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAATCAT 237  
Group07 TCGATGAAGAACCGCAGCAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAATCAT 238  
Group04 TCGATGAAGAACCGCAGCAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAATCAT 235  
Group10 TCGATGAAGAACCGCAGCAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAATCAT 238  
Group08 TCGATGAAGAACCGCAGCAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAATCAT 238  
Group02 TCGATGAAGAACCGCAGCAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAATCAT 237  
Group03 TCGATGAAGAACCGCAGCAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAATCAT 236  
Group05 TCGATGAAGAACCGCAGCAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAATCAT 237  
Group11 TCGATGAAGAACCGCAGCAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAATCAT 237  
\*\*\*\*\*

Group09 CGAATCTTTGAACGCACATTGCGCCCGCCAGTACTCTGGCGGGCATGCCTGTCCGAGCGT 297  
Group06 CGAATCTTTGAACGCACATTGCGCCCGCCAGTACTCTGGCGGGCATGCCTGTCCGAGCGT 298  
Group01 CGAATCTTTGAACGCACATTGCGCCCGCCAGTACTCTGGCGGGCATGCCTGTCCGAGCGT 297  
Group07 CGAATCTTTGAACGCACATTGCGCCCGCCAGTACTCTGGCGGGCATGCCTGTCCGAGCGT 298  
Group04 CGAATCTTTGAACGCACATTGCGCCCGCCAGTACTCTGGCGGGCATGCCTGTCCGAGCGT 295  
Group10 CGAATCTTTGAACGCACATTGCGCCCGCCAGTACTCTGGCGGGCATGCCTGTCCGAGCGT 298  
Group08 CGAATCTTTGAACGCACATTGCGCCCGCCAGTACTCTGGCGGGCATGCCTGTCCGAGCGT 298  
Group02 CGAATCTTTGAACGCACATTGCGCCCGCCAGTACTCTGGCGGGCATGCCTGTCCGAGCGT 297  
Group03 CGAATCTTTGAACGCACATTGCGCCCGCCAGTACTCTGGCGGGCATGCCTGTCCGAGCGT 296  
Group05 CGAATCTTTGAACGCACATTGCGCCCGCCAGTACTCTGGCGGGCATGCCTGTCCGAGCGT 297  
Group11 CGAATCTTTGAACGCACATTGCGCCCGCCAGTACTCTGGCGGGCATGCCTGTCCGAGCGT 297  
\*\*\*\*\*

Group09 CATTTC AACCTCAGCACCCCGTTTCGCGCGGGAGCTGGCGTTGGGGATCGGCCGTCT 357  
Group06 CATTTC AACCTCAGCACCCCGTTTCGCGCGGGAGCTGGCGTTGGGGATCGGCCGTCT 358  
Group01 CATTTC AACCTCAGCACCCCGTTTCGCGCGGGAGCTGGCGTTGGGGATCGGCCGTCT 357  
Group07 CATTTC AACCTCAGCACCCCGTTTCGCGCGGGAGCTGGCGTTGGGGATCGGCCGTCT 358  
Group04 CATTTC AACCTCAGCACCCCGTTTCGCGCGGGAGCTGGCGTTGGGGATCGGCCGTCT 355  
Group10 CATTTC AACCTCAGCACCCCGTTTCGCGCGGGAGCTGGCGTTGGGGATCGGCCGTCT 358  
Group08 CATTTC AACCTCAGCACCCCGTTTCGCGCGGGAGCTGGCGTTGGGGATCGGCCGTCT 358  
Group02 CATTTC AACCTCAGCACCCCGTTTCGCGCGGGAGCTGGCGTTGGGGATCGGCCGTCT 357  
Group03 CATTTC AACCTCAGCACCCCGTTTCGCGCGGGAGCTGGCGTTGGGGATCGGCCGTCT 356  
Group05 CATTTC AACCTCAGCACCCCGTTTCGCGCGGGAGCTGGCGTTGGGGATCGGCCGTCT 357  
Group11 CATTTC AACCTCAGCACCCCGTTTCGCGCGGGAGCTGGCGTTGGGGATCGGCCGTCT 357  
\*\*\*\*\*

Group09 CTGCGGCGGCCGCCCCGAAACGAAGTGGCGGTACGTCGCGACCTCCTCTGCGTAGTAG 417  
Group06 CTGCGGCGGCCGCCCCGAAACGAAGTGGCGGTACGTCGCGACCTCCTCTGCGTAGTAG 418  
Group01 CTGCGGCGGCCGCCCCGAAACGAAGTGGCGGTACGTCGCGACCTCCTCTGCGTAGTAG 417  
Group07 CTGCGGCGGCCGCCCCGAAACGAAGTGGCGGTACGTCGCGACCTCCTCTGCGTAGTAG 418  
Group04 CTGCGGCGGCCGCCCCGAAACGAAGTGGCGGTACGTCGCGACCTCCTCTGCGTAGTAG 415  
Group10 CTGCGGCGGCCGCCCCGAAACGAAGTGGCGGTACGTCGCGACCTCCTCTGCGTAGTAG 418  
Group08 CTGCGGCGGCCGCCCCGAAACGAAGTGGCGGTACGTCGCGACCTCCTCTGCGTAGTAG 418  
Group02 CTGCGGCGGCCGCCCCGAAACGAAGTGGCGGTACGTCGCGACCTCCTCTGCGTAGTAG 417  
Group03 CTGCGGCGGCCGCCCCGAAACGAAGTGGCGGTACGTCGCGACCTCCTCTGCGTAGTAG 416  
Group05 CTGCGGCGGCCGCCCCGAAACGAAGTGGCGGTACGTCGCGACCTCCTCTGCGTAGTAG 417  
Group11 CTGCGGCGGCCGCCCCGAAACGAAGTGGCGGTACGTCGCGACCTCCTCTGCGTAGTAG 417  
\*\*\*\*\*

Group09 CAATATCTCGCAGCCTGGATAGCGGCGCGGCCACGCCGTA AAAACCCCAACTATAACCAAG 477  
Group06 CAATATCTCGCAGCCTGGATAGCGGCGCGGCCACGCCGTA AAAACCCCAACTATAACCAAG 478  
Group01 CAATATCTCGCAGCCTGGATAGCGGCGCGGCCACGCCGTA AAAACCCCAACTATAACCAAG 477  
Group07 CAATATCTCGCAGCCTGGATAGCGGCGCGGCCACGCCGTA AAAACCCCAACTATAACCAAG 478  
Group04 CAATATCTCGCAGCCTGGATAGCGGCGCGGCCACGCCGTA AAAACCCCAACTATAACCAAG 475  
Group10 CAATATCTCGCAGCCTGGATAGCGGCGCGGCCACGCCGTA AAAACCCCAACTATAACCAAG 478  
Group08 CAATATCTCGCAGCCTGGATAGCGGCGCGGCCACGCCGTA AAAACCCCAACTATAACCAAG 478  
Group02 CAATATCTCGCAGCCTGGATAGCGGCGCGGCCACGCCGTA AAAACCCCAACTATAACCAAG 477  
Group03 CAATATCTCGCAGCCTGGATAGCGGCGCGGCCACGCCGTA AAAACCCCAACTATAACCAAG 476  
Group05 CAATATCTCGCAGCCTGGATAGCGGCGCGGCCACGCCGTA AAAACCCCAACTATAACCAAG 477  
Group11 CAGTATCTCGCAGCCTGGATAGCGGCGCGGCCACGCCGTA AAAACCCCAACTATAACCAAG 477  
**Loci** \*\*7\*\*\*\*\*

Group09 TTGACCTCGGATCAGGTAGGAATACCCGCTGA ACTTAA 515  
Group06 TTGACCTCGGATCAGGTAGGAATACCCGCTGA ACTTAA 516  
Group01 TTGACCTCGGATCAGGTAGGAATACCCGCTGA ACTTAA 515  
Group07 TTGACCTCGGATCAGGTAGGAATACCCGCTGA ACTTAA 516  
Group04 TTGACCTCGGATCAGGTAGGAATACCCGCTGA ACTTAA 513  
Group10 TTGACCTCGGATCAGGTAGGAATACCCGCTGA ACTTAA 516  
Group08 TTGACCTCGGATCAGGTAGGAATACCCGCTGA ACTTAA 516  
Group02 TTGACCTCGGATCAGGTAGGAATACCCGCTGA ACTTAA 515  
Group03 TTGACCTCGGATCAGGTAGGAATACCCGCTGA ACTTAA 514  
Group05 TTGACCTCGGATCAGGTAGGAATACCCGCTGA ACTTAA 515  
Group11 TTGACCTCGGATCAGGTAGGAATACCCGCTGA ACTTAA 515  
\*\*\*\*\*



K TTACGCACTGGCCGGTCTGGAGGTGAACCTATTTGAGAGGGGACAAACACGTCGTTAGCT 240  
E TTACGCACTGGCCGGTCTGGAGGTGAACCTATTTGAGAGGGGACAAACACGTCGTTAGCT 240  
A TTACGCACTGGCCGGTCTGGAGGTGAACCTATTTGAGAGGGGACAAACACGTCGTTAGCT 240  
I TTACGCACTGGCCGGTCTGGAGGTGAACCTATTTGAGAGGGGACAAACACGTCGTTAGCT 240  
D TTACGCACTGGCCGGTCTGGAGGTGAACCTATTTGAGAGGGGACAAACACGTCGTTAGCT 240  
G TTACGCACTGGCCGGTCTGGAGGTGAACCTATTTGAGAGGGGACAAACACGTCGTTAGCT 240  
P TTACGCACTGGCCGGTCTGGAGGTGAACCTATTTGAGAGGGGACAAACACGTCGTTAGCT 240  
Q TTACGCACTGGCCGGTCTGGAGGTGAACCTATTTGAGAGGGGACAAACACGTCGTTAGCT 240  
J TTACGCACTGGCCGGTCTGGAGGTGAACCTATTTGAGAGGGGACAAACACGTCGTTAGCT 240  
R TTACGCACTGGCCGGTCTGGAGGTGAACCTATTTGAGAGGGGACAAACACGTCGTTAGCT 240  
C TTACGCACTGGCCGGTCTGGAGGTGAACCTATTTGAGAGGGGACAAACACGTCGTTAGCT 240  
B TTACGCACTGGCCGGTCTGGAGGTGAACCTATTTGAGAGGGGACAAACACGTCGTTAGCT 240  
O TTACGCACTGGCCGGTCTGGAGGTGAACCTATTTGAGAGGGGACAAACACGTCGTTAGCT 240  
M TTACGCACTGGCCGGTCTGGAGGTGAACCTATTTGAGAGGGGACAAACACGTCGTTAGCT 240  
H TTACGCACTGGCCGGTCTGGAGGTGAACCTATTTGAGAGGGGACAAACACGTCGTTAGCT 240  
F TTACGCACTGGCCGGTCTGGAGGTGAACCTATTTGAGAGGGGACAAACACGTCGTTAGCT 240  
**Loci** \*\*\*\*\*1\*\*\*\*\*1\*\*\*\*\*1  
3 4 5

K AAAGATTCGACTGACGCAAAGCTGCGAGACCCCTTTCGTTTTCTCCTCGCCTGT--CC 300  
E AAAGATTCGACTGACGCAAAGCTGCGAGACCCCTTTCGTTTTCTCCTCGCCTGT--CC 300  
A AAAGATTCGACTGACGCAAAGCTGCGAGACCCCTTTCGTTTTCTCCTCGCCTGT--CC 300  
I AAAGATTCGACTGACGCAAAGCTGCGAGACCCCTTTCGTTTTCTCCTCGCCTGT--CC 300  
D AAAGATTCGACTGACGCAAAGCTGCGAGACCCCTTTCGTTTTCTCCTCGCCTGT--CC 300  
G AAAGATTCGACTGACGCAAAGCTGCGAGACCCCTTTCGTTTTCTCCTCGCCTGT--CC 300  
P AAAGATTCGACTGACGCAAAGCTGCGAGACCCCTTTCGTTTTCTCCTCGCCTGT--CC 300  
Q AAAGATTCGACTGACGCAAAGCTGCGAGACCCCTTTCGTTTTCTCCTCGCCTGT--CC 299  
J AAAGATTCGACTGACGCAAAGCTGCGAGACCCCTTTCGTTTTCTCCTCGCCTGT--CC 300  
R AAAGATTCGACTGACGCAAAGCTGCGAGACCCCTTTCGTTTTCTCCTCGCCTGT--CC 299  
C AAAGATTCGACTGACGCAAAGCTGCGAGACCCCTTTCGTTTTCTCCTCGCCTGT--CC 298  
B AAAGATTCGACTGACGCAAAGCTGCGAGACCCCTTTCGTTTTCTCCTCGCCTGT--CC 299  
O AAAGATTCGACTGACGCAAAGCTGCGAGACCCCTTTCGTTTTCTCCTCGCCTGT--CC 299  
M AAAGATTCGACTGACGCAAAGCTGCGAGACCCCTTTCGTTTTCTCCTCGCCTGT--CC 300  
H AAAGATTCGACTGACGCAAAGCTGCGAGACCCCTTTCGTTTTCTCCTCGCCTGT--CC 300  
F AAAGATTCGACTGACGCAAAGCTGCGAGACCCCTTTCGTTTTCTCCTCGCCTGT--CC 300  
**Loci** \*\*\*\*\*1111\*\*  
6789

K CCCCCTTTTGGGTTGGAATGTTGACTTGGGTGCACGTTGGCCGGCGGGGTAATAGAAAAG 361  
E CCCCCTTTTGGGTTGGAATGTTGACTTGGGTGCACGTTGGCCGGCGGGGTAATAGAAAAG 360  
A CCCCCTTTTGGGTTGGAATGTTGACTTGGGTGCACGTTGGCCGGCGGGGTAATAGAAAAG 362  
I CCCCCTTTTGGGTTGGAATGTTGACTTGGGTGCACGTTGGCCGGCGGGGTAATAGAAAAG 363  
D CCCCCTTTTGGGTTGGAATGTTGACTTGGGTGCACGTTGGCCGGCGGGGTAATAGAAAAG 362  
G CCCCCTTTTGGGTTGGAATGTTGACTTGGGTGCACGTTGGCCGGCGGGGTAATAGAAAAG 362  
P CCCCCTTTTGGGTTGGAATGTTGACTTGGGTGCACGTTGGCCGGCGGGGTAATAGAAAAG 362  
Q CCCCCTTTTGGGTTGGAATGTTGACTTGGGTGCACGTTGGCCGGCGGGGTAATAGAAAAG 361  
J CCCCCTTTTGGGTTGGAATGTTGACTTGGGTGCACGTTGGCCGGCGGGGTAATAGAAAAG 362  
R CCCCCTTTTGGGTTGGAATGTTGACTTGGGTGCACGTTGGCCGGCGGGGTAATAGAAAAG 361  
C CCCCCTTTTGGGTTGGAATGTTGACTTGGGTGCACGTTGGCCGGCGGGGTAATAGAAAAG 360  
B CCCCCTTTTGGGTTGGAATGTTGACTTGGGTGCACGTTGGCCGGCGGGGTAATAGAAAAG 361  
O CCCCCTTTTGGGTTGGAATGTTGACTTGGGTGCACGTTGGCCGGCGGGGTAATAGAAAAG 361  
M CCCCCTTTTGGGTTGGAATGTTGACTTGGGTGCACGTTGGCCGGCGGGGTAATAGAAAAG 362  
H CCCCCTTTTGGGTTGGAATGTTGACTTGGGTGCACGTTGGCCGGCGGGGTAATAGAAAAG 362  
F CCCCCTTTTGGGTTGGAATGTTGACTTGGGTGCACGTTGGCCGGCGGGGTAATAGAAAAG 361  
**Loci** \*\*\*\*\*2\*\*\*\*\*  
0

**MAP**

```

α      GGGACTGCTATCAATTTGCTGATTGGATATATACCAGGCCTACCCTTCTCCTCTCAACTA 60
δ      GGGACTACTATCAATTTGCTGATTGGATATATACCAGGCCTACCCTTCTCCTCTCAACTA 60
β      GGGACTGCTATCAATTTGCTGATTGGATATATACCAGGCCTACCCTTCTCCTCTCAACTA 60
γ      GGGGCTGCTATCAATTTGCTGATTGGATATATACCAGGCCTACCCTTCTCCTCTCAACTA 60
Loci   **2*2*****
        1 2

α      CAACAATTTCCCAAGTCTTGTGTACATCGATAAACGAAGTCATTTGCCACGGCATTCC 120
δ      CAACAATTTCCCAAGTCTTGTGTACATCGATAAACGAAGTCATTTGCCACGGCATTCC 120
β      CAACAATTTCCCAAGTCTTGTGTACATCGATAAACGAAGTCATTTGCCATGGCATTCC 120
γ      CAACAATTTCCCAAGTCTTGTGTACATCGATAAACGAAGTCATTTGCCACGGCATTCC 120
Loci   *****2*****
        3

α      CGACCAGCGAGTCCTCCTCGACGGAGACATCCTCAATATCGACGTTTCCCTCTACCACGA 180
δ      CGACCAGCGAGTCCTCCTCGACGGAGACATCCTCAATATCGACGTTTCCCTCTACCATGA 180
β      CGACCAGCGAGTCCTCCTCGACGGAGACATCCTCAATATCGACGTTTCCCTCTACCACGA 180
γ      CGACCAGCGAGTCCTCCTCGACGGAGACATCTCTCAATATCGACATTTCCCTCTACCACGA 180
Loci   *****2*****2**
        4 5 6

α      AGGCTACCATGCCGATTTGAATGAGACATACTACATTGGCGACAAGGCCAAAGCAGATCC 240
δ      AGGCTACCATGCCGATTTGAATGAGACATACTACATTGGCGACAAGGCCAAAGCAGATCC 240
β      AGGCTACCATGCCGATTTGAATGAGACATACTACATTGGCGACAAGGCCAAAGCAGATCC 240
γ      AGGCTACCATGCCGATTTGAATGAGACATACTACATTGGCGACAAGGCCAAAGCAGATCC 240
Loci   *****2*****
        7

α      CGATAGTGTGCGCGTCGTCGAGGCGGCCCGTGAGTGCCTGGAAGAATCCATCAAGGCCGT 300
δ      CGATAGTGTGCGCGTCGTCGAGGCGGCCCGTGAGTGCCTGGAAGAATCCATCAAGGCCGT 300
β      CGATAGTGTGCGCGTCGTCGAGGCGGCCCGTGAGTGCCTGGAAGAATCCATCAAGGCCGT 300
γ      CGATAGTGTGCGCGTCGTCGAGGCGGCCCGTGAGTGCCTGGAAGAATCCATCAAGGCCGT 300
*****

α      CAAGCCGGGCACCTTGATCCGTGAATTCGGAACATCATTGAGAAGCATGCCAAGGCCAAA 360
δ      CAAGCCGGGCACCTTGATCCGTGAATTCGGAACATCATTGAGAAGCATGCCAAGGCCAAA 360
β      CAAGCCGGGCACCTTGATCCGTGAATTCGGAACATCATTGAGAAGCATGCCAAGGCCAAA 360
γ      CAAGCCGGGCACCTTGATCCGTGAATTCGGAACATCATTGAGAAGCATGCCAAGGCCAAA 360
*****

α      GAATT 365
δ      GAATT 365
β      GAATT 365
γ      GAATT 365
*****

```

Figure A4.S1 Twenty-seven nucleotide polymorphic loci detected in three genomic regions in *G. morbida*. Sequencing ITS, BT and MAP of *G. morbida* isolates resulted in 7, 13 and 7 polymorphisms, respectively. Those were combined in 57 haplotypes (Table A4.S2). Polymorphisms are depicted in colors and loci are identified.

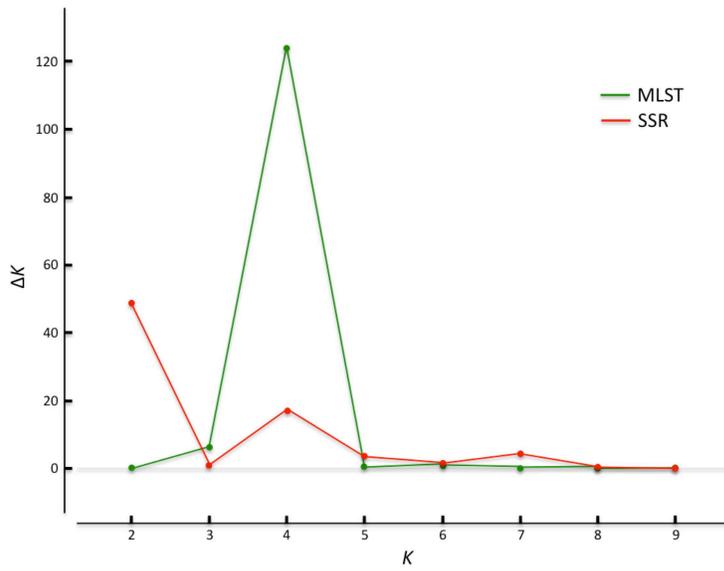


Figure A4.S2 Results of  $\Delta K$  computation for *Geosmithia morbida* isolates. Scenario where  $K=1-10$  (20 iterations) by using STRUCTURE [39]. SSR data supports clusters of 2 and 4, whereas MLST data supports a cluster of 4.

Table A4.S1 Microsatellite (SSR) loci used to characterize *Geosmithia morbida* isolates.

SSR name	GenBank identification	Motif(s)	Number of Alleles	Size (bp)	Primers
Geo4045	KF574259	CT	4	182-195	TCGGTGAATTGATCGGTGTA CTCCCAGCCGTAACGAC
Geo223	KF574250	CT	4	175-185	ACGTCACAGGAGGGAGAGAA CGTGGTGGTTCTCGTAACAT
Geo3896	KF574256	CT	5	200-226	CGGGTAGGGTGACCACGA GTGGTCGAGGACAACAAGGT
Geo3416	KF574255	GT	5	214-244	AGTTCGCGAAAGAGCACAAT GTTACACAGAAGGTTGTTAG TATGG
Geo7713	KF574258	CA	11	293-411	CGTACGATGGTATTTATCTCA TTCA TTACACGTAGGATGGCATGG
Geo194	KF574251	GT/AG	7	253-285	CCCACGTCAAGTCTGTCA TCCTCCTCAGTCGCATACTCT
Geo2849	KF574254	GA/AAT	4	310-341	CGCCTACGGAATACCTACC AGGACGAAACATCACCGAGT
Geo2514	KF574253	AC/TA	4	259-280	TTAACATGGCCAACACAACC TCA TTGA TGTCGGCAAAGAC
Geo1851	KF574252	AGC	6	202-228	GGGCAATATAACGGTTT CCTCGTAGGCGTAAGTCTGG
Geo6823	KF574257	TTG/GA	9	303-331	TTCAGCATCATTACTTGTTGG TCGTCGTCGTTGCTGTAGTC

Table A4.S2 Collection Information of 209 *Geosmithia morbida*-isolates and their MLST and SSR profile.

Iso- lat e	St at e	County	Geo Reg	Host	MLST data four-cluster-MLST- DAPC model				SSR data**								Loc us 771 3		
					Hapl otype #	ITS	BT #	MA P	Locu s 223	Locu s 194	Locu s 1851	Locus 2514 (75% failed)	Locu s 2849	Locus 3416 (63% failed)	Locu s 3896	Locu s 4045		Locu s 6823	
15 82	C A	Merced	04_C eCA	<i>J. hindsii</i>	Clu ster 1	Gro up0 H01	1	R	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
12 17	C O	Boulder	12_N oCO	<i>J. nigra</i>	Clu ster 1	Gro up0 H03	1	B	α	185	270	214	266	316	214	200	184	311	358
12 49	C A	Yolo	05_N oCA	<i>J. californica</i>	Clu ster 1	Gro up0 H03	1	B	α	185	270	205	270	310	229	200	183	311	340
12 50	C O	Boulder	12_N oCO	<i>J. nigra</i>	Clu ster 1	Gro up0 H03	1	B	α	185	270	214	-9	316	-9	200	184	316	340
12 61	C A	San Joaquin	05_N oCA	<i>J. hindsii</i>	Clu ster 1	Gro up0 H03	1	B	α	185	270	214	-9	310	-9	200	184	323	340
12 62	C A	Ventura	03_S wCA	<i>J. californica</i>	Clu ster 1	Gro up0 H03	1	B	α	183	280	205	266	310	-9	-9	184	313	340
13 08	C O	Jeffers on	12_N oCO	<i>J. nigra</i>	Clu ster 1	Gro up0 H03	1	B	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
13 87	C A	Solano	05_N oCA	<i>J. hindsii</i>	Clu ster 1	Gro up0 H03	1	B	α	185	270	214	-9	-9	-9	218	183	320	340

14 O	Clackamas	09_O R_W A	<i>Juglans</i> sp.	1 Cluster 1	H03	1	B	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
14 C	Lake	05_N oCA	<i>J. hindsii</i>	1 Cluster 1	H03	1	B	α	185	270	212	-9	310	-9	200	183	311	358
14 C	Lake	05_N oCA	<i>J. hindsii</i>	1 Cluster 1	H03	1	B	α	185	270	214	-9	-9	-9	204	184	311	358
14 C	Lake	05_N oCA	<i>J. hindsii</i>	1 Cluster 1	H03	1	B	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
14 C	Lake	05_N oCA	<i>J. hindsii</i>	1 Cluster 1	H03	1	B	α	185	270	212	-9	316	-9	208	-9	-9	-9
15 T	Knox	15_T N	<i>J. nigra</i>	1 Cluster 1	H03	1	B	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
15 C	San Benito	04_C eCA	<i>J. regia</i>	1 Cluster 1	H03	1	B	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
15 C	Fresno	04_C eCA	<i>J. regia</i>	1 Cluster 1	H03	1	B	α	183	270	205	266	310	-9	200	184	308	340
15 C	Fresno	04_C eCA	<i>J. regia</i>	1 Cluster 1	H03	1	B	α	-9	270	205	-9	-9	-9	200	184	-9	-9
15 T	Knox	15_T N	<i>J. nigra</i>	1 Cluster 1	H03	1	B	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
15 C	Merced	04_C eCA	<i>J. hindsii</i>	1 Cluster 1	H03	1	B	α	185	270	202	-9	310	-9	208	184	313	348

16 P		17_P		1		1												
59 A	Bucks	A	<i>J. nigra</i>	Cluster 1	H03	Group0 1	B	α	NA									
16 C		04_C		Cluster 1		Group0 1												
99 A	Tulare	eCA	<i>J. hindsii</i>	Cluster 1	H03	Group0 1	B	α	NA									
18 C		05_N	<i>P. stenoptera</i>	Cluster 1		Group0 1												
22 A	Solano	oCA	<i>a</i>	Cluster 1	H03	Group0 1	B	α	NA									
12 W		09_O		Cluster 1		Group0 1												
36 A	Benton	A	<i>J. nigra</i>	Cluster 1	H05	Group0 1	K	α	183	270	205	269	310	229	-9	184	323	369
15 T		15_T		Cluster 1		Group0 1												
05 N	Knox	N	<i>J. nigra</i>	Cluster 1	H08	Group0 1	B	β	185	270	202	-9	316	229	208	184	313	340
12 C	Jeffers	12_N		Cluster 1		Group0 1												
74 O	on	oCO	<i>J. nigra</i>	Cluster 1	H09	Group0 1	C	α	183	270	214	-9	314	-9	208	183	320	296
14 C	Contra	05_N		Cluster 1		Group0 1												
59 A	Costa	oCA	<i>J. hindsii</i>	Cluster 1	H09	Group0 1	C	α	NA									
15 C		12_N		Cluster 1		Group0 1												
99 O	Boulder	oCO	<i>J. nigra</i>	Cluster 1	H09	Group0 1	C	α	NA									
16 V	Dinwidie	16_V		Cluster 1		Group0 1												
62 A		A	<i>J. nigra</i>	Cluster 1	H09	Group0 1	C	α	NA									
16 V	Chester	16_V		Cluster 1		Group0 1												
67 A	field	A	<i>J. nigra</i>	Cluster 1	H09	Group0 1	C	α	NA									
12 C		05_N		Cluster 1		Group0 1												
33 A	Yolo	oCA	<i>J. hindsii</i>	Cluster 1	H11	Group0 1	C	α	185	-9	214	-9	310	-9	208	184	-9	340

12	C		03_S	<i>J.</i>	1		2												
72	A	Ventura	wCA	<i>californica</i>	1	H11	2	C	α	185	270	214	266	316	-9	226	183	313	340
15	N	Bernalil	X_Ce		1		2												
03	M	lo	NM	<i>J. nigra</i>	1	H11	2	C	α	NA									
12	C	Jeffers	12_N		1		3												
39	O	on	oCO	<i>J. regia</i>	1	H16	3	C	α	183	270	208	259	310	-9	200	183	320	296
12	C	Jeffers	12_N		1		3												
71	O	on	oCO	<i>J. nigra</i>	1	H16	3	C	α	183	266	208	-9	316	-9	-9	183	323	296
15	N		01_N		1		4												
73	M	Catron	Z	<i>J. major</i>	1	H17	4	P	α	183	253	208	-9	316	-9	208	182	320	296
13	C	Jeffers	12_N		1		4												
11	O	on	oCO	<i>J. nigra</i>	1	H18	4	B	α	183	253	208	-9	313	-9	200	183	320	293
15	C		12_N		1		4												
47	O	Denver	oCO	<i>J. nigra</i>	1	H18	4	B	α	NA									
12	C	Jeffers	12_N		1		4												
73	O	on	oCO	<i>J. nigra</i>	1	H19	4	C	α	185	270	208	-9	316	229	208	183	316	296
13	C	Jeffers	12_N		1		4												
12	O	on	oCO	<i>J. nigra</i>	1	H19	4	C	α	NA									
13	C	Jeffers	12_N		1		4												
13	O	on	oCO	<i>J. nigra</i>	1	H19	4	C	α	NA									
15	C		12_N		1		4												
96	O	Boulder	oCO	<i>J. nigra</i>	1	H19	up0	C	α	NA									

15	C		12_N		1		4												
97	O	Boulder	oCO	<i>J. nigra</i>	Clu ster	H19	4	C	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
15	C		12_N		Clu ster														
98	O	Boulder	oCO	<i>J. nigra</i>	Clu ster	H19	4	C	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
14	N		01_N		Clu ster														
53	M	Grant	Z	<i>J. major</i>	Clu ster	H21	4	E	α	183	270	208	-9	316	-9	200	184	320	296
15	N		01_N		Clu ster														
40	M	Lincoln	Z	<i>J. major</i>	Clu ster	H21	4	E	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
15	N		01_N		Clu ster														
58	M	Grant	Z	<i>J. major</i>	Clu ster	H21	4	E	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
15	N		01_N		Clu ster														
59	M	Grant	Z	<i>J. major</i>	Clu ster	H21	4	E	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
16	N		01_N		Clu ster														
70	M	Lincoln Santa	Z	<i>J. major</i>	Clu ster	H21	4	E	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
14	C	Barbar	03_S	<i>J.</i>	Clu ster														
78	A	Barbar	wCA	<i>californica</i>	Clu ster	H22	4	H	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
13	U		11_U		Clu ster														
09	T	Cache	T	<i>J. nigra</i>	Clu ster	H23	4	K	α	185	270	205	-9	310	-9	200	183	313	296
14	C	Contra	05_N		Clu ster														
62	A	Costa	oCA	<i>J. hindsii</i>	Clu ster	H25	5	B	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
14	C	Santa	03_S	<i>J.</i>	Clu ster														
76	A	Barbar	wCA	<i>californica</i>	Clu ster	H25	up0	B	α	185	-9	214	-9	316	229	-9	184	311	369

a				1	5															
15	C		04_C	Cluster		5														
18	A	Fresno	eCA	<i>J. regia</i>	1	H25	5	B	$\alpha$	NA	NA	NA	NA	NA	NA	NA	NA	NA		
14	C		04_C	Cluster		5														
07	A	Tulare	eCA	<i>J. regia</i>	1	H26	5	C	$\alpha$	185	270	214	-9	310	-9	208	184	313	358	
14	C	Contra	05_N	Cluster		5														
61	A	Costa	oCA	<i>J. hindsii</i>	1	H26	5	C	$\alpha$	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
14	C		05_N	Cluster		5														
86	A	Lake	oCA	<i>J. hindsii</i>	1	H26	5	C	$\alpha$	183	270	212	-9	316	229	208	184	-9	369	
14	O	Clackamas	09_O R_W A	<i>Juglans</i> sp.	Cluster		5													
40	R				1	H27	5	B	$\beta$	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
15	T		15_T	Cluster		5														
09	N	Knox	N	<i>J. nigra</i>	1	H27	5	B	$\beta$	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
15	T		15_T	Cluster		5														
06	N	Knox	N	<i>J. nigra</i>	1	H29	5	C	$\beta$	185	270	205	-9	316	229	204	184	313	358	
15	C		12_N	Cluster		8														
45	O	Denver	oCO	<i>J. nigra</i>	1	H35	8	C	$\alpha$	183	270	214	-9	316	-9	208	184	325	351	
15	N		01_N M_A	Cluster		0														
72	M	Hidalgo	Z	<i>J. major</i>	1	H43	0	B	$\alpha$	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
13	C	Jeffers	12_N	Cluster		0														
14	O	on	oCO	<i>J. nigra</i>	1	H44	0	E	$\alpha$	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
18	C		05_N	Cluster		0														
21	A	Solano	oCA	<i>P. stenopter</i>	1	H56	0	B	$\alpha$	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	

		a	1	2														
12 C		12_N		Clu														
24 O	Boulder	oCO	<i>J. nigra</i>	ster	H02	1	A	α	183	270	214	-9	310	-9	208	183	311	369
12 C		12_N		Clu														
25 O	Boulder	oCO	<i>J. nigra</i>	ster	H02	1	A	α	183	270	205	-9	316	-9	-9	184	323	358
12 C		05_N	<i>J.</i>	Clu														
27 A	Yolo	oCA	<i>californica</i>	ster	H02	1	A	α	183	270	212	-9	310	-9	208	184	316	369
12		10_I		Clu														
45 ID	Boise San	D	<i>J. nigra</i>	ster	H02	1	A	α	185	270	205	266	316	229	217	184	311	348
12 C		03_S	<i>J.</i>	Clu														
52 A	Obispo	wCA	<i>californica</i>	ster	H02	1	A	α	185	270	205	-9	316	-9	208	184	316	358
12 O		09_O		Clu														
60 R	Hood River	R_W A	<i>Juglans</i> sp.	ster	H02	1	A	α	185	270	214	-9	310	229	200	183	-9	358
12 C		03_S	<i>J.</i>	Clu														
63 A	Ventura	wCA	<i>californica</i>	ster	H02	1	A	α	185	270	205	-9	316	-9	-9	184	323	362
12 C		03_S	<i>J.</i>	Clu														
64 A	Ventura	wCA	<i>californica</i>	ster	H02	1	A	α	185	270	214	-9	310	-9	-9	184	323	340
12 C		05_N		Clu														
67 A	San Joaquin	oCA	<i>J. hindsii</i>	ster	H02	1	A	α	185	270	205	266	310	229	200	183	323	340
12 C		03_S	<i>J.</i>	Clu														
68 A	Ventura	wCA	<i>californica</i>	ster	H02	1	A	α	185	270	212	-9	316	214	200	184	325	358
12 C		05_N		Clu														
70 A	San Joaquin	oCA	<i>J. hindsii</i>	ster	H02	up0	A	α	185	270	212	-9	316	229	200	183	313	340



13	O	Clackamas	09_O R_W A	<i>Juglans</i> <i>sp.</i>	2 Cluster	H02	1	A	α	185	270	205	-9	316	-9	204	184	311	348
13	C	Sutter	05_N oCA	<i>J. nigra X</i> <i>hindsii</i>	2 Cluster	H02	1	A	α	185	270	205	-9	316	-9	218	184	325	340
13	C	Yolo	05_N oCA	<i>J. hindsii</i>	2 Cluster	H02	1	A	α	NA									
14	C	Kings	04_C eCA	<i>J. hindsii</i>	2 Cluster	H02	1	A	α	185	270	214	-9	316	229	-9	183	311	358
14	C	Boulder	12_N oCO	<i>J. nigra</i>	2 Cluster	H02	1	A	α	NA									
14	C	Sutter	05_N oCA	<i>J. nigra X</i> <i>hindsii</i>	2 Cluster	H02	1	A	α	185	270	205	259	316	-9	200	183	311	362
14	C	Yolo Santa	05_N oCA	<i>J. hindsii</i>	2 Cluster	H02	1	A	α	185	270	214	-9	310	-9	200	183	320	369
14	C	Barbara	03_S wCA	<i>J.</i> <i>californica</i>	2 Cluster	H02	1	A	α	NA									
14	C	Barbara	03_S wCA	<i>J.</i> <i>californica</i>	2 Cluster	H02	1	A	α	185	270	212	-9	316	244	200	184	323	358
14	C	Lake	05_N oCA	<i>J. hindsii</i>	2 Cluster	H02	1	A	α	185	270	214	-9	316	229	-9	184	323	369
14	C	Los Angeles	03_S wCA	<i>J.</i> <i>californica</i>	2 Cluster	H02	up0	A	α	NA									

s				2	1													
				Cluster		Group												
15 C		04_C		2		Group0												
14 A	Tulare	eCA	<i>J. regia</i>	2	H02	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
15 T		15_T		2		Group0												
23 N	Knox	N	<i>J. nigra</i>	2	H02	1	A	α	185	270	212	-9	310	-9	208	184	323	358
15 T		15_T		2		Group0												
25 N	Knox	N	<i>J. nigra</i>	2	H02	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
15 W		09_O		2		Group0												
30 A	Klickitat	R_W	<i>J. hindsii</i>	2	H02	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
15 C		14_S		2		Group0												
35 O	Otero	oCO	<i>J. nigra</i>	2	H02	1	A	α	185	270	205	269	316	-9	200	183	323	411
15 C		04_C		2		Group0												
81 A	Merced	eCA	<i>J. hindsii</i>	2	H02	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
15 W	Walla	09_O		2		Group0												
90 A	Walla	R_W	<i>J. nigra</i>	2	H02	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
16 C		12_N		2		Group0												
00 O	Boulder	oCO	<i>J. nigra</i>	2	H02	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
16 C		12_N		2		Group0												
02 O	Boulder	oCO	<i>J. nigra</i>	2	H02	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
16 N	Carson	06_N	<i>Juglans</i>	2		Group0												
31 V	City	V	sp.	2	H02	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
16 N	Carson	06_N	<i>Juglans</i>	2		Group0												
33 V	City	V	sp.	2	H02	up0	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

16	C	Los Angeles	03_S wCA	<i>J. californica</i>	2	H02	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
16	V	Dinwiddie	16_V A	<i>J. nigra</i>	2	H02	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
16	V	Chesterfield	16_V A	<i>J. nigra</i>	2	H02	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
17	N	Washoe	06_N V	<i>J. nigra</i>	2	H02	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
18	C	Solano	05_N oCA	<i>P. fraxinifolia</i>	2	H02	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
20	C	Los Angeles	03_S wCA	<i>P. fraxinifolia</i>	2	H02	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
12	U	Cache	11_U T	<i>J. nigra</i>	2	H04	1	D	α	185	270	214	-9	316	-9	-9	183	-9	358
12	C	Ventura	03_S wCA	<i>J. californica</i>	2	H04	1	D	α	185	270	205	-9	310	-9	-9	184	316	358
13	O	Marion	09_O R_W A	<i>Juglans</i> sp.	2	H04	1	D	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
13	C	Mesa	13_WeC O	<i>J. nigra</i>	2	H04	1	D	α	183	270	205	-9	316	229	200	183	325	369
13	C	Sacramento	05_N oCA	<i>J. hindsii</i>	2	H04	1	D	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

				2		1												
14	O		09_O	Clu		Gro												
32	R	Wasco	R_W	ster		up0												
			A	2	H04	1	D	α	185	270	214	-9	316	229	208	184	320	340
			09_O	Clu		Gro												
14	O		R_W	ster		up0												
39	R	Marion	A	2	H04	1	D	α	NA									
			07_S	Clu		Gro												
15	O	Jackso	wOR	ster		up0												
34	R	n	A	2	H04	1	D	α	NA									
			02_C	Clu		Gro												
13	A	Yavapa	eAZ	ster		up0												
66	Z	i	AZ	2	H06	1	J	δ	183	270	202	-9	316	-9	208	183	325	369
			05_N	Clu		Gro												
14	C	Alamed	oCA	ster		up0												
28	A	a	CA	2	H07	1	M	α	183	270	212	-9	316	229	208	184	313	340
			09_O	Clu		Gro												
12	O		R_W	ster		up0												
58	R	Wasco	A	2	H10	2	A	α	183	270	212	270	341	-9	208	184	323	340
			12_N	Clu		Gro												
12	C	Larimer	oCO	ster		up0												
76	O		CO	2	H10	2	A	α	-9	-9	205	-9	-9	-9	226	184	311	340
			09_O	Clu		Gro												
13	O		R_W	ster		up0												
46	R	Wasco	A	2	H10	2	A	α	183	270	205	-9	310	-9	200	184	313	340
			12_N	Clu		Gro												
16	C	Boulder	oCO	ster		up0												
01	O		CO	2	H10	2	A	α	NA									
			15_T	Clu		Gro												
16	T		N	ster		up0												
72	N	Knox	N	2	H10	2	A	α	NA									
			07_S	Clu		Gro												
15	O	Jackso	wOR	ster		up0												
33	R	n	A	2	H12	up0	D	α	NA									

15	C		04_C		2														
84	A	Merced	eCA	<i>J. hindsii</i>	2	H12	2	D	$\alpha$	NA									
13	C		12_N		2														
93	O	Weld	oCO	<i>J. nigra</i>	2	H13	2	G	$\alpha$	183	270	-9	-9	310	229	-9	183	311	348
12	O	Multno	R_W	<i>Juglans</i>	2														
56	R	mah	A	sp.	2	H14	2	A	$\beta$	183	270	205	269	316	229	208	184	311	340
16	C	Los	03_S	<i>J.</i>	2														
34	A	Angeles	wCA	<i>californica</i>	2	H14	2	A	$\beta$	NA									
16	C		05_N	<i>J. hindsii</i>	2														
10	A	Yolo	oCA	<i>X major</i>	2	H15	2	D	$\beta$	NA									
12	C		13_WeC		2														
85	O	Delta	O	<i>J. nigra</i>	2	H20	4	D	$\alpha$	183	-9	205	-9	316	-9	208	184	320	293
13	C	Jeffers	12_N		2														
05	O	on	oCO	<i>J. nigra</i>	2	H20	4	D	$\alpha$	183	253	208	259	314	-9	200	183	320	296
12	C		12_N		2														
18	O	Boulder	oCO	<i>J. nigra</i>	2	H24	5	A	$\alpha$	185	270	205	269	310	233	208	-9	323	340
12	O		09_O		2														
59	R	Marion	R_W	<i>Juglans</i>	2	H24	5	A	$\alpha$	185	270	205	-9	316	-9	200	184	311	358
13	C		14_S		2														
18	O	Otero	oCO	<i>J. nigra</i>	2	H24	5	A	$\alpha$	183	270	205	-9	-9	229	208	184	311	348
14	C		04_C		2														
04	A	Tulare	eCA	<i>J. regia</i>	2	H24	up0	A	$\alpha$	NA									

14	C	Los Angeles	03_S wCA	<i>J. californica</i>	2	H24	5	A	α	NA									
15	O	Wasco	09_O R_W	<i>J. hindsii</i>	2	H24	5	A	α	185	270	214	-9	310	229	208	183	311	358
14	C	Tulare	04_C eCA	<i>J. regia</i>	2	H28	5	D	α	185	270	214	-9	310	-9	208	184	313	403
12	C	Ventura	03_S wCA	<i>J. californica</i>	2	H30	6	A	α	183	270	214	-9	310	-9	204	184	311	340
13	C	Jefferson	12_N oCO	<i>J. nigra</i>	2	H30	6	A	α	-9	270	205	-9	310	229	208	183	311	340
14	O	Marion	09_O R_W	<i>Juglans</i> sp.	2	H30	6	A	α	NA									
15	C	Fresno	04_C eCA	<i>J. regia</i>	2	H30	6	A	α	NA									
15	T	Knox	15_T N	<i>J. nigra</i>	2	H30	6	A	α	185	270	214	-9	310	-9	200	184	323	358
17	O	Lane	08_C eOR	<i>J. cinerea</i>	2	H30	6	A	α	NA									
17	O	Lane	08_C eOR	<i>J. cinerea</i>	2	H30	6	A	α	NA									
14	C	Lake	05_N oCA	<i>J. hindsii</i>	2	H31	up0	A	β	185	266	214	266	316	-9	208	184	320	340

				2		6													
12	C		14_S	Cluster		Gro													
47	O	El Paso	oCO 01_N	<i>J. nigra</i>	2	H34	8	A	α	183	253	212	266	310	-9	200	184	320	296
16	A	Cochis	M_A		Cluster		Gro												
23	Z	e	Z	<i>J. major</i>	2	H42	0	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
16	A	Cochis	M_A		Cluster		Gro												
24	Z	e	Z	<i>J. major</i>	2	H42	0	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
16	A	Cochis	M_A		Cluster		Gro												
27	Z	e	Z	<i>J. major</i>	2	H42	0	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
16	A	Cochis	M_A		Cluster		Gro												
26	Z	e	Z	<i>J. major</i>	2	H45	0	I	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
16	A	Cochis	M_A		Cluster		Gro												
28	Z	e	Z	<i>J. major</i>	2	H46	0	J	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
14	O		09_O		Cluster		Gro												
02	R	Wasco	R_W A	<i>Juglans</i> sp.	2	H47	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
15	C		04_C		Cluster		Gro												
85	A	Tulare	eCA	<i>J. regia</i>	2	H47	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
15	C		04_C		Cluster		Gro												
86	A	Tulare	eCA	<i>J. regia</i>	2	H47	1	A	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
13	C	Jeffers	12_N		Cluster		Gro												
06	O	on	oCO	<i>J. nigra</i>	2	H48	1	G	α	185	270	212	266	316	-9	204	184	-9	340
14	C	Fremon	14_S		Cluster		Gro												
96	O	t	oCO	<i>J. nigra</i>	ster	H49	up0	A	β	NA	NA	NA	NA	NA	NA	NA	NA	NA	

17	N	Douglas	06_N		2		1												
08	V	s	V	<i>J. nigra</i>	2	H51	6	D	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
16	A	Cochis	01_N		Clu		Gro												
25	Z	e	M_A	<i>J. major</i>	2	H52	8	J	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
15	C		Z		ster		Gro												
13	A	Solano	05_N	<i>J. regia</i>	2	H55	1	H	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
12	C		oCA		Clu		Gro												
46	O	El Paso	14_S	<i>J. nigra</i>	2	H57	4	A	α	183	270	208	-9	314	-9	208	184	320	296
15	N		oCO		ster		Gro												
69	M	Hidalgo	M_A	<i>J. major</i>	3	H36	9	O	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
15	N		Z		Clu		Gro												
76	M	Catron	01_N	<i>J. major</i>	3	H37	9	Q	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
13	C	Jeffers	M_A		ster		Gro												
01	O	on	12_N	<i>J. nigra</i>	3	H38	9	C	α	183	263	208	-9	310	-9	208	184	316	296
13	A	Yavapa	oCO		Clu		Gro												
02	Z	i	02_C	<i>J. major</i>	3	H38	9	C	α	183	270	208	266	316	-9	-9	183	316	296
13	C	Jeffers	eAZ		ster		Gro												
21	O	on	12_N	<i>J. nigra</i>	3	H38	9	C	α	183	253	208	266	316	-9	200	183	320	296
13	N		oCO		Clu		Gro												
40	M	Grant	M_A	<i>J. major</i>	3	H38	9	C	α	NA	NA	NA	NA	NA	NA	NA	NA	NA	
13	A	Yavapa	Z		ster		Gro												
88	Z	i	02_C	<i>J. major</i>	ster	H38	up0	C	α	185	253	208	-9	-9	-9	-9	183	320	296

14	N		01_N			3		9												
01	M	Catron	M_A			Clu		Gro												
			Z	<i>J. major</i>		ster		up0												
						3	H38	9	C	α	183	266	208	259	316	-9	218	183	320	296
15	A	Yavapa	02_C			Clu		Gro												
54	Z	i	eAZ	<i>J. major</i>		ster		up0												
						3	H38	9	C	α	183	269	208	-9	316	-9	200	183	331	293
15	N		01_N			Clu		Gro												
56	M	Grant	M_A			ster		up0												
			Z	<i>J. major</i>		3	H38	9	C	α	NA									
15	N		01_N			Clu		Gro												
74	M	Catron	M_A			ster		up0												
			Z	<i>J. major</i>		3	H38	9	C	α	NA									
13	N		01_N			Clu		Gro												
73	M	Grant	M_A			ster		up0												
			Z	<i>J. major</i>		3	H39	9	E	α	NA									
14	N		01_N			Clu		Gro												
21	M	Grant	M_A			ster		up0												
			Z	<i>J. major</i>		3	H39	9	E	α	NA									
14	N		01_N			Clu		Gro												
22	M	Grant	M_A			ster		up0												
			Z	<i>J. major</i>		3	H39	9	E	α	NA									
15	N		01_N			Clu		Gro												
57	M	Grant	M_A			ster		up0												
			Z	<i>J. major</i>		3	H39	9	E	α	175	261	208	-9	316	-9	200	183	320	296
12	C	Jeffers	12_N			Clu		Gro												
48	O	on	oCO	<i>J. nigra</i>		ster		up0												
						3	H40	9	J	α	183	253	-9	259	-9	222	200	184	320	296
13	A	Yavapa	02_C			Clu		Gro												
22	Z	i	eAZ	<i>J. major</i>		ster		up0												
						3	H40	9	J	α	NA									
13	C	Jeffers	12_N			Clu		Gro												
23	O	on	oCO	<i>J. nigra</i>		ster		up0												
							H40	up0	J	α	NA									



15	A	Yavapa	02_C		4	7													
50	Z	i	eAZ	<i>J. major</i>	4	H33	7	F	γ	NA									
16	A	Yavapa	02_C		4		7												
75	Z	i	eAZ	<i>J. major</i>	4	H33	7	F	γ	NA									
12	U		11_U		no		Gro	no	no										
22	T	Cache	T	<i>J. regia</i>	. info.	4	up0	inf	o.	185	280	214	259	341	-9	208	184	311	362
12	C		05_N	<i>J.</i>	no		Gro	no	inf										
28	A	Yolo	oCA	<i>californica</i>	. info.	2	A	o.		-9	-9	205	-9	-9	-9	208	184	320	358
12	C		05_N	<i>J.</i>	no		Gro	no	inf										
29	A	Yolo	oCA	<i>californica</i>	. info.	5	D	o.		-9	270	205	-9	316	-9	-9	184	311	358
13	O		09_O		no		Gro	no	inf										
47	R	Wasco	R_W	<i>Juglans</i>	info	no	up0	inf											
		Los	A	sp.	. info.	1	o.	α		185	270	214	-9	310	229	217	184	313	358
13	C	Angeles	03_S	<i>J.</i>	no		Gro	no	inf										
48	A	s	wCA	<i>californica</i>	. info.	1	A	o.		185	270	205	-9	310	214	208	183	325	340
13	O		09_O		no		Gro	no	inf										
49	R	Wasco	R_W	<i>Juglans</i>	info	no	up0	inf											
		Los	A	sp.	. info.	1	o.	α		183	270	212	-9	310	229	208	184	316	358
13	C		13_		no		Gro	no	no										
59	O	Mesa	WeC	<i>J. nigra</i>	. info.	1	o.	o.		183	270	205	-9	316	229	200	184	323	293
13	C	Angeles	03_S		no		Gro	no	inf										
83	A	s	wCA	<i>J. hindsii</i>	. info.	1	o.	α		185	270	212	-9	310	-9	204	184	313	340
14	C		04_C		no	no	Gro	no											
05	A	Kings	eCA	<i>J. hindsii</i>	info	info.	up0	R	inf	185	270	212	-9	-9	229	226	183	-9	-9

				.		1		o.											
14	C	Santa Barbar	03_S	<i>J.</i>	no	Gro	no												
73	A	a	wCA	<i>californica</i>	info.	up0	inf	α	185	270	212	-9	316	229	200	183	-9	358	
			09_O		no	Gro	no												
15	W	Walla	R_W		info	no	up0	inf											
28	A	Walla	A	<i>J. nigra</i>	info.	1	o.	α	185	270	205	-9	310	-9	208	183	323	340	
			09_O		no	Gro	no												
15	W	Walla	R_W		info	no	up0	inf											
29	A	Walla	A	<i>J. nigra</i>	info.	1	o.	α	185	270	212	-9	310	-9	200	184	311	340	

# haplotypes are based on the combination of 11 variants of ITS (represented by numbers), 16 variants of BT (represented by letters) and 4 of MAP (represented by Greek letters).

⌘ BT variants "L" and "N" are not listed, because they corresponded to sequence errors.

\* alleles in columns K-T highlighted in blue, red, yellow and green, correspond to private alleles (total of 24) of clusters 1 (3 private alleles), 2 (7 private alleles), 3 (5 private alleles) and 4 (9 private alleles), respectively, according to the four-cluster-MLST-DAPC model (column F).

\* "-9" indicates either null alleles or failed reactions, whereas "NA" indicates isolates not analyzed by SSR approach.

Table A4.S3 Host specificity analysis.  $F_{ST}$  values calculated using both MLST (above diagonal) and SSR (below diagonal) data, grouping isolates by host. Significant ( $P < 0.05$ )  $F_{ST}$  values are denoted by (\*).

Host name					
-----					
	<i>J. californica</i>	<i>J. hindsii</i>	<i>J. major</i>	<i>J. nigra</i>	<i>J. regia</i>
<i>J. californica</i>	-	0.02876	0.15919*	0.03525*	0.05465
<i>J. hindsii</i>	-0.00012	-	0.13289*	0.00158	-0.02118
<i>J. major</i>	0.26018*	0.2149*	-	0.12837*	0.08492*
<i>J. nigra</i>	0.03888*	0.04309	0.11609*	-	-0.0105
<i>J. regia</i>	0.01	0.02239	0.15748*	-0.01454	-

File A4.S1 Development of MLST-primers based on partial sequencing of the *Geosmithia morbida* (CBS 124663) genome.

High-quality total DNA of *G. morbida* isolate 1217 (CBS 124663) was extracted as described in “DNA extraction” (Materials and Methods) and skim-sequenced at the Purdue University Genomics Core Facility. A half-plate run on a 454 FLX standard pyrosequencing [1] of a 3 kb-long paired end library generated 779,553 reads representing 116,376,601 bp. Sequences were assembled in Newbler 2.3 (454 Life Sciences Corp, CT, USA) by using default settings. This resulted in the assembly of 628,023 reads (80.56% of the total), which were organized into 27,933 contigs, representing 15,924,440 bp of the *G. morbida* genome. A total 24,982 sequences (contigs and singletons) larger than 200 bp were automatically annotated with the Blast2GO similarity tool [2], and 92% were homologous to predicted proteins deposited in the NR database (BlastX, e-value cut-off =  $10^{-5}$ ); 81% had a GO term [3] associated with their hit, and 61% were homologous to a predicted protein with known function. Six contigs containing the partial sequences of predicted housekeeping genes (Table A4.2) were selected and primers flanking these sequences were designed. The sequences were: methionine aminopeptidase (KF947520); ribosomal L18ae protein family (KF947521); dolichyl-phosphate-mannose-protein mannosyltransferase (KF947522); amino acid permease (KF947523); 40S ribosomal protein S2 (KF947524); and kinesin (KF947525). Primers were used for the establishment of a MLST scheme for *G. morbida*.

#### **References** for File A4.S1

1. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376-380.
2. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, et al. (2005) Blast2GO: a universal

tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674-3676.

3. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene Ontology: tool for the unification of biology. *Nature Genetics* 25: 25-29.

## **Acknowledgments**

We thank R.M. Bostock, T.W. Coleman, P.L. Dallara, G. Durham, E.J. Fichtner, S. Fraedrich, K.J. Greby, B. Hammon, J.E. Henrich, S.M. Hishinuma, D. Leatherman, C.A. Leslie, A. Liu, J. McKenna, L.M. Ohara, M. Putnam, S. Schlarbaum, C. Utley, and D.L. Wood for assistance in collecting samples. We also thank E. Freeland, E. Luna, N. Thiry, R. Billings, L. Holder, R. Feild, M. Shenk, M. Nelson, R. Sitz and M. Curto for providing technical support. Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the U.S. Dept. of Agriculture and does not imply its approval to the exclusion of other products or vendors that also may be suitable.

## **Author Contributions**

Conceived and designed the experiments: KW WC SJS NT. Performed the experiments: MMZ CH JT. Analyzed the data: MMZ JIC KW KB. Contributed reagents/materials/analysis tools: ADG JP WC SJS NT. Wrote the paper: MMZ JIC KW ADG JP KB WC SJS NT. Performed statistical analyses: JIC.

## References

1. Williams RD (1990) *Juglans nigra* L., black walnut. In: Burns RM, Honkala, BH, tech coords. 2. Hardwoods. Agriculture Handbook 654. U.S. Department of Agriculture, Forest Service, Washington, DC. vol. 2. pp. 391–410.
2. Hardin JW, Leopold DJ, White FM (2000) Harlow & Harrar's Textbook of Dendrology: New York: McGraw-Hill. 544 p.
3. Nicodemus MA, Salifu KF, Jacobs DF (2008) Nitrate reductase activity and nitrogen compounds in xylem exudate of *Juglans nigra* seedlings: relation to nitrogen source and supply. *Trees* 22: 685–695.
4. Tisserat N, Cranshaw W, Putnam ML, Pscheidt J, Leslie CA, et al. (2011) Thousand cankers disease is widespread in black walnut in the western United States. Online. *Plant Health Progress*. doi:10.1094/PHP-2011-0630-01-BR. Accessed 10 July 2014.
5. Walnut (*Juglans* spp.)-Thousand cankers disease (black walnut decline). In: Pscheidt JW, O'camb CM, eds. (2014) Pacific Northwest plant disease management handbook [online]. Corvallis, OR: Oregon State University. <http://pnwhandbooks.org/plantdisease/walnut-Juglans-spp-thousand-cankers-disease-black-walnut-decline>. Accessed 10 July 2014.
6. USDA-FS (2001) Forest Insect and Disease Conditions of the Southwestern Region. USDA-FS Southwestern Region, Forestry and Forest Health Publication R3-02-01. 21 p.
7. Bright DE, Jr. (1981) Taxonomic monograph of the genus *Pityophthorus* Eichhoff in North and Central America (Coleoptera: Scolytidae). *Memoirs of the Entomological Society of Canada* 113: 1–378.
8. Bright DE, Jr. (2014) A catalog of Scolytidae and Platypodidae (Coleoptera), Supplement 3 (2000–2010), with notes on subfamily and tribal reclassifications. *Insecta Mundi* 356: 1–336.
9. Tisserat N, Cranshaw W, Leatherman D, Utlely C, Alexander K (2009) Black walnut mortality in Colorado caused by the walnut twig beetle and thousand cankers disease. Online. *Plant Health Progress*. doi:10.1094/PHP-2009-0811-01-RS. Accessed 10 July 2014.
10. Kolar'ik M, Freeland E, Utlely C, Tisserat N (2011) *Geosmithia morbida* sp.nov., a new phytopathogenic species living in symbiosis with the walnut twig beetle (*Pityophthorus juglandis*) on *Juglans* in USA. *Mycologia* 103: 325–332.
11. Seybold SJ, Haugen D, O'Brien J, Graves AD (2013) Thousand cankers disease. USDA Forest Service, Northeastern Area, State and Private Forestry Pest Alert NA-PR-02-10. <http://www.na.fs.fed.us/pubs/detail.cfm?id=5225>. Accessed 10 July 2014.

12. Grant JF, Windham MT, Haun WG, Wiggins GJ, Lambdin PL (2011) Initial assessment of thousand cankers disease on black walnut, *Juglans nigra*, in eastern Tennessee. *Forests* 2: 741–748. doi:10.3390/f2030741.
13. Haun WG, Powell S, Strohmeier C, Kirksey J (2010) State of Tennessee thousand cankers disease action plan: Tennessee Department of Agriculture, 34 p.
14. Hanson MA, Bush EA, Day E, Griffin G, Dart NL (2011) Walnut Thousand Cankers Disease Alert. Virginia Cooperative Extension, 4 p.
15. Hadziabdic D, Windham MT, Baird R, Vito L, Cheng Q, et al. (2014) First report of *Geosmithia morbida* in North Carolina: The pathogen involved in thousand cankers disease of black walnut. *Plant Disease* 98: 992. <http://dx.doi.org/10.1094/PDIS-06-13-0630-PD>.
16. Services. VDoAaC (2011) VDACS - Plant and pest services - Thousand cankers disease.
17. Force PFPT (2012) Commonwealth of Pennsylvania thousand cankers disease action plan.
18. Fisher JR, McCann DP, Taylor NJ (2013) *Geosmithia morbida*, thousand cankers disease of black walnut pathogen, was found for the first time in southwestern Ohio. Online. *Plant Health Progress*. doi:10.1094/PHP-2013-1201-01-BR. Accessed 10 July 2014.
19. Montecchio L, Faccoli M (2014) First record of thousand cankers disease *Geosmithia morbida* and walnut twig beetle *Pityophthorus juglandis* on *Juglans nigra* in Europe. *Plant Disease*. 98: 696. <http://dx.doi.org/10.1094/PDIS-10-13-1027-PDN>.
20. Cranshaw W (2011) Recently recognized range extensions of the walnut twig beetle, *Pityophthorus juglandis* Blackman (Coleoptera: Curculionidae: Scolytinae), in the western United States. *Coleop Bull* 65: 48–49.
21. Blackman MW (1928) The genus *Pityophthorus* Eichh. in North America: a revisional study of the Pityophthori, with descriptions of two new genera and seventy-one species. *Bull New York State College of Forestry* 1(3–6): 183.
22. Seybold SJ, Coleman TW, Dallara PL, Dart NL, Graves AD, et al. (2012) Recent collecting reveals new state records and geographic extremes in the distribution of the walnut twig beetle, *Pityophthorus juglandis* Blackman (Coleoptera: Scolytidae), in the United States. *Pan-Pac Entomol* 88: 277–280.
23. Graves AD, Coleman TW, Seybold SJ (2011) Monitoring walnut health and decline in response to thousand cankers disease and infestation by the walnut twig beetle, *Pityophthorus juglandis*, in southern California and New Mexico. Year one Progress Report for USDA Forest Service Forest Health Monitoring Project INT-EM-B-11-03. <http://fhm.fs.fed.us/em/funded/12/INT-EM-B-11-03.pdf>. Accessed 10 July 2014.

24. Griffin JR, Critchfield WB (1972) The distribution of forest trees in California. USDA Forest Service Research Paper PSW-82, Pacific Southwest Forest and Range Experiment Station. 114 p.
25. Newton LP, Fowler G, Neeley AD, Schall RA, Takeuchi Y (2009) Pathway Assessment: *Geosmithia* sp. and *Pityophthorus juglandis* Blackman movement from the western into the eastern United States. United States Department of Agriculture Animal and Plant Health Inspection Service Rev. 1: 10.19.2009.  
[http://tn.gov/agriculture/publications/regulatory/tc\\_pathwayanalysis.pdf](http://tn.gov/agriculture/publications/regulatory/tc_pathwayanalysis.pdf) Accessed 10 July 2014.
26. Freeland E, Cranshaw W, Tisserat N (2012) Effect of *Geosmithia morbida* isolate and temperature on canker development in black walnut. Online. Plant Health Progress. doi:10.1094/PHP-2012-0618-01-RS. Accessed 10 July 2014.
27. Hadziabdic D, Vito LM, Windham MT, Pscheidt JW, Trigiano RN, et al. (2013) Genetic differentiation and spatial structure of *Geosmithia morbida*, the causal agent of thousand cankers disease in black walnut (*Juglans nigra*). Curr Genet 60: 75–87.
28. Halkett F, Simon JC, Balloux F (2005) Tackling the population genetics of clonal and partially clonal organisms. Trends Ecol Evol 20: 194–201. DOI:10.1016/j.tree.2005.01.001.
29. Fournier E, Giraud T (2008) Sympatric genetic differentiation of a generalist pathogenic fungus, *Botrytis cinerea*, on two different host plants, grapevine and bramble. J Evol Biol 21: 122–132.
30. Serdani M, Vlach JJ, Wallis KL, Zerillo MM, McCleary T, et al. (2013) First report of *Geosmithia morbida* and *Pityophthorus juglandis* causing thousand cankers disease in butternut. Plant Health Progress. doi:10.1094/PHP-2013-1018-01-BR. Accessed 10 July 2014.
31. Huang W, Marth G (2008) EagleView: a genome assembly viewer for next-generation sequencing technologies. Genome Res 18: 1538–1543.
32. Rozen S, Skaletsky H (1999) Primer3 on the WWW for general users and for biologist programmers. Bioinformatics methods and protocols: Springer. 365–386.
33. Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology 18: 233–234.
34. Debourgogne A, Gueidan C, Hennequin C, Contet-Audonnet N, de Hoog S, et al. (2010) Development of a new MLST scheme for differentiation of *Fusarium solani* species complex (FSSC) isolates. J Microbiol Methods 82: 319–323.
35. Gardes M, Bruns T (1993) ITS primers with enhanced specificity for basidiomycetes—application to the identification of mycorrhizae and rusts. Mol Ecol 2: 113–118.

36. O'Donnell K, Cigelnik E (1997) Two divergent intragenomic rDNA ITS2 Types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. *Mol Phylogenet Evol* 7: 103–116.
37. Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8: 186–194.
38. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95–98.
39. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
40. Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11: 94.
41. Jombart T, Ahmed I (2011) ADEGENET 1.3–1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27: 3070–3071.
42. Earl DA, vonHoldt BM (2011) Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conserv Gen Res* 4: 359–361.
43. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software Structure: a simulation study. *Mol Ecol* 14: 2611–2620.
44. Excoffier L (1995) AMOVA: Analysis of molecular variance (version 1.55). Geneva, Switzerland: University of Geneva. 47 p.
45. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1: 47.
46. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
47. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731–2739.
48. Green PJ (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82: 711–732.
49. Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.
50. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, et al. (2010) New

algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *System Biol* 59: 307–321.

51. Arnaud-Haond S, Belkhir K (2007) GenClone: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Mol Ecol Notes* 7: 15–17.
52. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496–2497.
53. Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. *Mol Ecol Notes* 1: 101–102.
54. Hudson RR, Boos DD, Kaplan NL (1992) A statistical test for detecting geographic subdivision. *Mol Biol Evol* 9: 138–151.
55. Freeland E (2012) Intraspecific variability of *Geosmithia morbida* the causal agent of thousand cankers disease, and effects of temperature, isolate and host family (*Juglans nigra*) on canker development. Fort Collins, CO: Colorado State University. 91 p.
56. Kolarčík M, Jankowiak R (2013) Vector affinity and diversity of *Geosmithia* fungi living on subcortical insects inhabiting Pinaceae species in central and northeastern Europe. *Microb Ecol* 66: 682–700.
57. Broders KD, Boraks A, Sanchez AM, Boland GJ (2012) Population structure of the butternut canker fungus, *Ophiognomonia clavignenti-juglandacearum*, in North American forests. *Ecol Evol* 2: 2114–2127.
58. Goss EM, Larsen M, Vercauteren A, Werres S, Heungens K, et al. (2011) *Phytophthora ramorum* in Canada: Evidence for migration within North America and from Europe. *Phytopathology* 101: 166–171.
59. Keller SR, Gilbert KJ, Fields PD, Taylor DR (2012) Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, *Silene latifolia*. *Molecular Ecology* 21: 4721–4734.
60. Ciosi M, Miller NJ, Kim KS, Giordano R, Estoup A, et al. (2008) Invasion of Europe by the western corn rootworm, *Diabrotica virgifera*: multiple transatlantic introductions with various reductions of genetic diversity. *Molecular Ecology* 17: 3614–3627.
61. Delmotte F, Giresse X, Richard-Cervera S, M'Baya J, Vear F, et al. (2008) Single nucleotide polymorphisms reveal multiple introductions into France of *Plasmopara halstedii*, the plant pathogen causing sunflower downy mildew. *Infection Genetics and Evolution* 8: 534–540.
62. Utley C, Nguyen T, Roubtsova T, Coggeshall M, Ford TM, et al. (2013) Susceptibility of walnut and hickory species to *Geosmithia morbida*. *Plant Disease* 97: 601–607.

63. Sitz R (2013) Management options for the walnut twig beetle, *Pityophthorus juglandis* Blackman, vector of the fungal canker pathogen *Geosmithia morbida*: Colorado State University. 68 p.
64. Little EL, Jr. (1979) Checklist of United States trees (native and naturalized). US Department of Agriculture Forest Service, Agriculture Handbook No. 541. 375 p.