Title

Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen

Journal

New Phytologist, 194(4)

Authors

Olson, Ake
Aerts, Andrea
Asiegbu, Fred
et al.

Publication Date

2012-03-28

Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen

Åke Olson[1], Andrea Aerts[2], Fred Asiegbu[3], Lassaad Belbahri[4], Ourdia Bouzid[5], Anders Broberg[6], Björn Canbäck[1], Pedro M. Coutinho[7], Dan Cullen[8], Kerstin Dalman[1], Giuliana Deflorio[9], Linda T.A. van Diepen[10], Christophe Dunand[11], Sébastien Duplessis[12], Mikael Durling[1], Paolo Gonthier[13], Jane Grimwood[14], Carl Gunnar Fossdal[15], David Hansson[6], Bernard Henrissat[7], Ari Hietala[15], Kajsa Himmelstrand[1], Dirk Hoffmeister[16], Nils Högberg[1], Timothy Y. James[10], Magnus Karlsson[1], Annegret Kohler[12], Ursula Kües[17], Yong-Hwan Lee[18], Yao-Cheng Lin[19], Mårten Lind[1], Erika Lindquist[2], Vincent Lombard[7], Susan Lucas[2], Karl Lundén[1], Emmanuelle Morin[12], Claude Murat[12], Jongsun Park[18], Tommaso Raffaello[3], Pierre Rouzé[19], Asaf Salamov[2], Jeremy Schmutz[14], Halvor Solheim[15], Jerry Ståhlberg[20], Heriberto Vélëz[1], Ronald P. de Vries[5,21], Ad Wiebenga[21], Steve Woodward[9], Igor Yakovlev[15], Matteo Garbelotto[22,], Francis Martin[12,], Igor V. Grigoriev[2,], Jan Stenlid[1,]

1. Department of Forest Mycology and Pathology, Swedish University of Agricultural Sciences, Box 7026, Ullsväg 26, 750 05 Uppsala, Sweden
2. US DOE Joint Genome Institute, Walnut Creek, CA 94598, USA
3. Department of Forest Ecology, PO Box 27 Latokartanonkaari 7, 00014 University of Helsinki, Helsinki, Finland
4. Laboratory of Soil Biology, University of Neuchâtel, Rue Emile Argand 11, CH-2000 Neuchâtel, Switzerland
5. Microbiology, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands
6. Department of Chemistry, Swedish University of Agricultural Sciences, Box 7015, 750 05 Uppsala, Sweden
7. AFMB UMR 6098 CNRS/UI/UII, Case 932, 163 Avenue de Luminy 13288 Marseille cedex 9, France
8. Forest Products Laboratory, Madison, WI 53726, USA
9. Department of Plant and Soil Science, Institute of Biological and Environmental Sciences, University of Aberdeen, Cruickshank Building, St. Machar Drive, Aberdeen, AB24 3UU, Scotland UK
10. Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA
11. Laboratory of Cell Surfaces and Plant Signalisation 24, University Paul Sabatier (Toulouse III), UMR5546- CNRS, Chemin de Borde-Rouge, BP 42617, Auzeville 31326 Castanet-Tolosan, France
12. UMR INRA-UHP 'Interactions Arbres/Micro-Organismes' IFR 110 'Genomique, Ecophysiologie et Ecologie Fonctionnelles' INRA-Nancy 54280 Champenoux, France

13. Department of Exploitation and Protection of Agricultural and Forest Resources (Di. Va. P. R. A.) – Plant Pathology, University of Torino, Via L. da Vinci 44, I-10095 Grugliasco, Italy
14. HudsonAlpha Institute for Biotechnology, 601 Genome Way Huntsville, AL 35806, USA
15. Norwegian Forest and Landscape Institute, Høgskoleveien 8, NO-1432 Ås, Norway
16. Pharmaceutical Biology, Friedrich-Schiller-Universität Jena, Winzerlaer Str. 2, 07745 Jena, Germany
17. Büsgen-Institute, Section Molecular Wood Biotechnology and Technical Mycology, University of Göttingen, Büsgenweg 2, D-37077 Göttingen, Germany
18. Department of Agricultural Biotechnology, Seoul National University, Seoul 151-921, Korea
19. VIB Department of Plant Systems Biology, Ghent University, Bioinformatics and Evolutionary Genomics, Technologiepark 927, B-9052 Gent, Belgium
20. Department of Molecular Biology, Swedish University of Agricultural Sciences, Box 590, Husargatan 3, 751 24 Uppsala, Sweden
21. CBS-KNAW Fungal Biodiversity Centre, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands
22. University of California, 338 Hilgard Hall Berkeley CA 94720 USA

JUNE 2012

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

DISCLAIMER

Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen

Åke Olson[1#], Andrea Aerts[8], Fred Asiegbu[2], Lassaad Belbahri[18], Ourdia Bouzid[20], Anders Broberg[21], Björn Canbäck[1], Pedro M. Coutinho[4], Dan Cullen[17], Kerstin Dalman[1], Giuliana Deflorio[14], Linda T.A. van Diepen[15], Christophe Dunand[12], Sébastien Duplessis[3], Mikael Durling[1], Paolo Gonthier[22], Jane Grimwood[9], Carl Gunnar Fossdal[5], David Hansson[21], Bernard Henrissat[4], Ari Hietala[5], Kajsa Himmelstrand[1], Dirk Hoffmeister[11], Nils Högberg[1], Timothy Y. James[15], Magnus Karlsson[1], Annegret Kohler[3], Ursula Kües[7], Yong-Hwan Lee[13], Yao-Cheng Lin[6], Mårten Lind[1], Erika Lindquist[8], Vincent Lombard[4], Susan Lucas[8], Karl Lundén[1], Emmanuelle Morin[3], Claude Murat[3], Jongsun Park[13], Tommaso Raffaello[2], Pierre Rouzé[6], Asaf Salamov[8], Jeremy Schmutz[9], Halvor Solheim[5], Jerry Ståhlberg[16], Heriberto Vélëz[1], Ronald P. de Vries[19,20], Ad Wiebenga[19], Steve Woodward[14], Igor Yakovlev[5], Matteo Garbelotto[10*], Francis Martin[3*], Igor V. Grigoriev[8*], Jan Stenlid[1*]

To whom correspondence should be addressed
* These authors contributed to this work as senior authors

1 Department of Forest Mycology and Pathology, Swedish University of Agricultural Sciences, Box 7026, Ullsväg 26, 750 05 Uppsala, Sweden
2 Department of Forest Ecology, PO Box 27 Latokartanonkaari 7, 00014 University of Helsinki, Finland
3 UMR INRA-UHP "Interactions Arbres/Micro-Organismes" IFR 110 "Genomique, Ecophysiologie et Ecologie Fonctionnelles" INRA-Nancy 54280 Champenoux, France
4 AFMB UMR 6098 CNRS/UI/UII, Case 932, 163 Avenue de Luminy 13288 Marseille cedex 9, France
5 Norwegian Forest and Landscape Institute, Høgskoleveien 8, NO-1432 Ås, Norway
6 VIB Department of Plant Systems Biology, Ghent University, Bioinformatics and Evolutionary Genomics, Technologiepark 927, B-9052 Gent, Belgium,
7 Institute for Forest Botany, University of Göttingen, Büsgenweg 2, 37077 Göttingen, Germany
8 US DOE Joint Genome Institute, Walnut Creek, CA 94598, USA
9 HudsonAlpha Institute for Biotechnology, 601 Genome Way Huntsville, AL 35806, USA
10 University of California, 338 Hilgard Hall Berkeley CA 94720 USA,
11 Pharmaceutical Biology, Friedrich-Schiller-Universität Jena, Winzerlaer str. 2, 07745 Jena, Germany
12 University Paul Sabatier (Toulouse III), UMR5546- CNRS, Laboratory of Cell Surfaces and Plant Signalisation 24, Chemin de Borde-Rouge, BP 42617, Auzeville 31326 Castanet-Tolosan, France
13 Department of Agricultural Biotechnology, Seoul National University, Seoul 151-921, Korea

14 University of Aberdeen, Institute of Biological and Environmental Sciences, Department of Plant and Soil Science, Cruickshank Building, St. Machar Drive, Aberdeen, AB24 3UU, Scotland UK
15 Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA
16 Department of Molecular Biology, Swedish University of Agricultural Sciences, Box 590, Husargatan 3, 751 24 Uppsala, Sweden
17 Forest Products Laboratory, Madison, WI 53726, USA
18 Laboratory of Soil Biology, University of Neuchâtel, Rue Emile Argand 11, CH-2000 Neuchâtel, Switzerland
19 CBS-KNAW Fungal Biodiversity Centre, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands
20 Microbiology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands,
21 Department of Chemistry, Swedish University of Agricultural Sciences, Box 7015, 750 05 Uppsala, Sweden
22 Department of Exploitation and Protection of Agricultural and Forest Resources (Di. Va. P. R. A.) – Plant Pathology, University of Torino, Via L. da Vinci 44, I-10095 Grugliasco, Italy

Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen

Abstract

- Parasitism and saprotrophic wood decay are two fungal strategies fundamental for succession and nutrient cycling in forest ecosystems. An opportunity to assess trade-off between these strategies is provided by the major forest pathogen and wood decayer *Heterobasidion annosum sensu lato*.
- We report on the annotated genome sequence and transcript profiling as well as quantitative trait loci mapping of one member of the species complex; *H. irregulare.* Quantitative trait loci critical for pathogenicity and rich in transposable elements, orphan and secreted genes, were identified.
- A wide range of cellulose degrading enzymes is expressed during wood decay. In contrast, pathogenic interaction between *H. irregulare* and pine engages fewer carbohydrate active enzymes, but involves an increase in pectinolytic enzymes, transcription modules for oxidative stress, and secondary metabolite production.
- Our results show a trade-off in terms of constrained carbohydrate decomposition and membrane transport capacity during interaction with living host. The findings establish that saprotrophic wood decay and necrotrophic parasitism involve two distinct yet overlapping processes.

Introduction

Fungi are heterotrophs that play several distinctive roles in ecosystems as saprotrophs, parasites of plants and animals, and mutualistic symbionts of photosynthetic organisms. Generally, each species is specialized in one of these strategies although recent findings indicate that there might be partial physiological capacity overlap among fungi with different primary trophic strategies (Vasiliauskas *et al.*, 2007; Newton *et al*., 2010). The use of more than one strategy by a single species might convey flexibility towards local changes in environment and competition with other organisms, as well as access to a wider ecological niche but is likely to result in a trade-off in terms of constrained use of its full genomic capacity under specific environmental conditions.

The *Heterobasidion annosum sensu lato* (s.l.) is a cosmopolitan fungal pathogen in conifer forests. In 1995 the economic losses were in the order of € 600 million annually to forest owners in Europe through tree mortality and wood decay (Woodward *et al.,* 1998). Although the economic consequences for North America forestry are less well documented, they are expected to be of similar magnitude. The frequency of root rot is increasing with about 23% per decade in managed forests in northern Europe (Thor *et al.,* 2005). In addition to threaten forest health this white rot fungus also causes massive release of CO_2 by decaying wood, thus representing a major threat to coniferous forests´ ability to serve as a natural carbon sink. The species complex is comprised of three Eurasian (*Heterobasidion annosum sensu stricto* (s.s.), *Heterobasidion parviporum* and *Heterobasidion abietinum*) and two North American (*Heterobasidion occidentale* and *Heterobasidion irregulare*) species, each with a different but overlapping host range (Niemelä & Korhonen, 1998; Otrosina & Garbelotto, 2010). Infections by *Heterobasidion* spp. are initiated in fresh wounds or newly created tree stump surfaces followed by spread via root to root infection through living bark and subsequent decay of and survival in the root and trunk of standing trees (Woodward *et al.,* 1998). This infection cycle relies on mechanisms for both saprotrophic wood decay and pathogenic interactions with a living host, allowing us to study potential trade-off between the two trophic strategies. The switch between saprotrophy and parasitism could be associated with an activation of distinct gene sets for the two growth modes, or governed by

differential regulation of common metabolic processes. A major goal of this genomic study was to identify key elements in the molecular repertoire required for balancing the two trophic strategies.

Materials and Methods

Selection of *H. irregulare* strain and isolation of genomic DNA and RNA

The sequenced *H. irregulare* strain TC 32-1 (Chase, 1985) is well characterized and have been used in many studies, e.g. the strain constitute one of the monokaryotic parental isolates in a genetic hybrid AO8, used for generating the mapping population for the *H. annosum* (s.l.) genetic linkage map (Lind *et al.*, 2005). Already published ESTs (expressed sequence tags) from the TC32-1 strain were available (Karlsson *et al.*, 2003). In addition, a BAC library of 2,688 clones was made by BIO S&T Inc.according to their standard protocols. DNA was extracted using established methods (Sambrook & Russell, 2001) and extractions of total RNA were performed according to CTAB and phenol/chloroform methods (Karlsson *et al.*, 2003).

Genome sequencing, assembly, and annotation

All sequencing reads for the whole genome shotgun sequencing were collected with standard Sanger sequencing protocols on ABI 3730XL capillary sequencing machines at the Department of Energy Joint Genome Institute in Walnut Creek, California. Three different sized libraries were used as templates for the plasmid subclone sequencing process and both ends were sequenced. 214,143 reads from the 2.7 kb sized library, 192,768 reads from the 6.0 kb sized library, and 63,168 reads from a 39.1 kb fosmid library were sequenced (Table S1).

A total of 406,752 reads were assembled using a modified version of Arachne v.20071016 (Jaffe *et al.*, 2003) with parameters maxcliq1=100, correct1_passes=0 and BINGE_AND_PURGE=True. This produced 53 scaffold sequences, with L50 of 2.2 Mb (the length of the scaffold that separates the top half (N50) of the assembled genome

from the rest), 19 scaffolds larger than 100 kb, and total scaffold size of 33.9 Mb. Each scaffold was screened against bacterial proteins, organelle sequences and GenBank and removed if found to be a contaminant. Additional scaffolds were removed if the scaffold contained only unanchored rDNA sequences. The final draft whole genome shotgun assembly contained scaffolds that cover 33.1 Mb of the genome with a contig L50 of 127.0 kb and a scaffold L50 of 2.2 Mb (Table S2).

Genome improvement

All genome improvement reactions were performed at the HudsonAlpha Genome Sequencing Center in Huntsville, Alabama. In order to improve and finish the genome of *H. irregulare*, the whole genome shotgun assembly was broken into scaffold size pieces and each scaffold piece was reassembled with phrap (Green, 1999). The scaffold pieces were then finished using a Phred/Phrap/Consed pipeline (Gordon *et al.*, 1998). Initially, all low quality regions and gaps were targeted with computationally selected sequencing reactions completed with 4:1 BigDye terminator: dGTP chemistry (Applied Biosystems). These automated rounds included directed primer walking on plasmid subclones using custom primers.

After the completion of the automated rounds each assembly were manually inspected. Further reactions were manually selected to complete the genome. These reactions included additional custom primer walks on plasmid subclones or fosmids using 4:1 BigDye terminator:dGTP chemistry. Smaller repeats in the sequence were resolved by transposon-hopping 8kb plasmid clones. To fill large gaps, resolve larger repeats or to resolve chromosome duplications and extend into chromosome telomere regions shotgun sequencing and finishing of BAC fosmid clones were used. During the course of the improvement project, 5,376 BAC ends were sequenced to add additional contiguity. Finally, the sequences were compared to markers on the available genetic map (Lind *et al.*, 2005) and two map joins were made based on map evidence.

Each assembly were validated by independent quality assessment that included a visual examination of subclone paired ends and visual inspection of discrepancies containing high quality sequence and all remaining low quality areas. All available EST resources were also placed on the assembly to ensure completeness.

cDNA library construction and sequencing

H. irregulare TC 32-1 poly A+ RNA was isolated from total RNA for two RNA samples; RNA1 - cells grown in Liquid Hagem-medium (Stenlid, 1985) and RNA2 - cells grown in Liquid high nitrogen MMN-medium (Marx, 1969) using the Absolutely mRNA Purification kit and manufacturers instructions (Stratagene, La Jolla, CA). cDNA synthesis and cloning was a modified procedure based on the "SuperScript plasmid system with Gateway technology for cDNA synthesis and cloning" (Invitrogen, Carsbad, CA). 1-2 µg of poly A+ RNA, reverse transcriptase SuperScript II (Invitrogen) and oligo dT-NotI primer (5' GACTAGTTCTAGATCGCGAGCGGCCGCCCT15VN 3') were used to synthesize first strand cDNA. Second strand synthesis was performed with *E. coli* DNA ligase, polymerase I, and RNaseH followed by end repair using T4 DNA polymerase. The SalI adaptor (5' TCGACCCACGCGTCCG and 5' CGGACGCGTGGG) was ligated to the cDNA, digested with NotI (New England Biolabs, Ipswich, MA), and subsequently size selected by gel electrophoresis (1.1% agarose). Size ranges of cDNA were cut out of the gel for the RNA1 sample yielding two cDNA libraries (JGI library codes CCPA for range 0.6k-2kb and CCOZ for the range >2kb), and JGI library codes CCPC and CCPB (same respective sizes) for the RNA2 sample. The cDNA inserts were directionally ligated into the SalI and NotI digested vector pCMVsport6 (Invitrogen). The ligation was transformed into ElectroMAX T1 DH10B cells (Invitrogen).

Library quality was first assessed by randomly selecting 24 clones and PCR amplifying the cDNA inserts with the primers M13-F (5' GTAAAACGACGGCCAGT) and M13-R (5' AGGAAACAGCTATGACCAT) to determine the fraction of insertless clones. Colonies from each library were plated onto agarose plates (254mm plates from Teknova, Hollister, CA) at a density of approximately 1000 colonies per plate. Plates were grown

at 37 °C for 18 hours after which individual colonies were picked and each used to inoculate a well containing LB media with appropriate antibiotic in a 384 well plate (Nunc, Rochester, NY). Clones in 384 well plates were grown at 37 °C for 18 hours. Contained plasmid DNA for sequencing was produced by rolling circle amplification (Templiphi, GE Healthcare, Piscataway, NJ). Subclone inserts were sequenced from both ends using primers complementary to the flanking vector sequence (Fw: 5' ATTTAGGTGACACTATAGAA, Rv: 5' TAATACGACTCACTATAGGG) and Big Dye terminator chemistry and then run on ABI 3730 instruments (Applied Biosystems, Foster City, CA).

EST sequence processing and assembly

A total of 40,807 ESTs including; 8,840 from CCPA, 8,759 from CCOZ, 13,280 from CCPC, 8,263 from CCPB and 1,665 from external sources were processed through the JGI EST pipeline (ESTs were generated in pairs, a 5' and 3' end read from each cDNA clone). To trim vector and adaptor sequences, common sequence patterns at the ends of ESTs were identified and removed using an internally developed tool. Insertless clones were identified if either of the following criteria were met: >200 bases of vector sequence at the 5' end or less than 100 bases of non-vector sequence remained. ESTs were then trimmed for quality using a sliding window trimmer (window = 11 bases). Once the average quality score in the window was below the threshold (Q15) the EST was split and the longest remaining sequence segment was retained as the trimmed EST. EST sequences with less than 100 bases of high quality sequence were removed. ESTs were evaluated for the presence of polyA or polyT tails (which if present were removed) and the EST reevaluated for length, removing ESTs with less than 100 bases remaining. ESTs consisting of more than 50% low complexity sequence were also removed from the final set of "good ESTs". In the case of resequencing the same EST, the longest high quality EST was retained. Sister ESTs (end pair reads) were categorized as follows: if one EST was insertless or a contaminant then by default the second sister was categorized as the same. However, each sister EST was treated separately for complexity and quality scores. Finally, EST sequences were compared to the Genbank nucleotide database in order to

identify contaminants; non-desirable ESTs such as those matching rRNA sequences were removed.

For clustering, ESTs were evaluated with malign, a kmer based alignment tool, which clusters ESTs based on sequence overlap (kmer = 16, seed length requirement = 32, alignment ID >= 98%). Clusters of ESTs were further merged based on sister ESTs using double linkage. Double linkage requires that 2 or more matching sister ESTs exist in both clusters to be merged. EST clusters were then each assembled using CAP3 (Huang & Madan 1999) to form consensus sequences. For cluster consensus sequence annotation, the consensus sequences were compared to Swissprot using blastx and the hits were reported. Clustering and assembly of all 33,539 ESTs resulted in 10,126 consensus sequences and 1,503 singlets.

Whole-genome exon oligoarray

The *Heterobasidion irregulare* custom-exon expression array (4 x 72K) manufactured by Roche NimbleGen Systems Limited (Madison, WI) (http://www.nimblegen.com/products/exp/index.html) contained five independent, nonidentical, 60-*mer* probes per gene model coding sequence. For 12,199 of the 12,299 annotated protein-coding gene models probes could be designed. For 19 gene models no probes could be generated and 81 gene models shared all five probes with other gene models. Included in the array were 916 random 60-*mer* control probes and labelling controls. For 2032 probes, technical duplicates were included on the array.

Total RNA was extracted using CTAB/phenol/chloroform and LiCl precipitation. The RNA was DNase I treated and cleaned with Qiagen RNA cleanup Kit. Arrays were performed from *H. irregulare* mycelium grown in liquid MMN medium (three biological replicates), from cambial zone of necrotic bark of pines inoculated with *H. irregulare* (21dpi; three biological replicates), from fruiting bodies collected in California (four biological replicates) as well as from *H. irregulare* grown on wood shavings from pine (four biological replicates), grown in liquid medium amended with lignin (Kraft Pine

lignin B 471003-500G; SIGMA-Aldrich) (2 biological replicates) and growth in liquid medium amended with cellulose from Spruce (22182-KG Fluka; SIGMA-Aldrich) (2 biological replicates). Cultures were harvested after 3 weeks of incubation 22 °C in darkness.

Total RNA preparations were amplified using the SMART PCR cDNA Synthesis Kit (Clontech) according to the manufacturer's instructions. Single dye labeling of samples, hybridization procedures, data acquisition, background correction and normalization were performed at the NimbleGen facilities (NimbleGen Systems, Reykjavik, Iceland) following their standard protocol. Microarray probe intensities were quantile normalized across all chips. Average expression levels were calculated for each gene from the independent probes on the array and were used for further analysis. Raw array data were filtered for non-specific probes (a probe was considered as non-specific if it shared more than 90% homology with a gene model other than the gene model it was made for) and renormalized using the ARRAYSTAR software (DNASTAR, Inc. Madison, WI, USA). For 621 gene models no reliable probe was left. A transcript was deemed expressed when its signal intensity was three-fold higher than the mean signal-to-noise threshold (cut-off value) of the random oligonucleotide probes present on the array (50 to 100 arbitrary units). Gene models with an expression value higher than three-fold the cut-off level were considered as transcribed. The maximum signal intensity values were ~65000 arbitrary units. A Student *t*-test with false discovery rate (FDR) (Benjamini-Hochberg) multiple testing correction was applied to the data using the ARRAYSTAR software (DNASTAR). Transcripts with a significant *p*-value (<0.05) were considered as differentially expressed. The complete expression dataset is available as series (accession number GSE30230) at the Gene Expression Omnibus at NCBI (http://www.ncbi.nlm.nih.gov/geo/).

Genome annotation

The *H. irregulare* genome was annotated using the JGI annotation pipeline, which takes multiple inputs (scaffolds, ESTs, and known genes) and runs several analytical tools for

gene prediction and annotation, and deposits the results in the JGI Genome Portal (http://www.jgi.doe.gov/Heterobasidion) for further analysis and manual curation. Genomic assembly scaffolds were masked using RepeatMasker (Smit *et al.*, 1996-2010) and the RepBase library of 234 fungal repeats (Jurka *et al.*, 2005). Using the repeat-masked assembly, several gene prediction programs falling into three general categories were used: 1) *ab initio* - FGENESH (Salamov & Solovyev, 2000); GeneMark (Isono *et al.*, 1994), 2) *homology-based* - FGENESH+; Genewise (Birney & Durbin, 2000) seeded by BLASTx (Altschul *et al.*, 1990) alignments against GenBank's database of non-redundant proteins (NR: http://www.ncbi.nlm.nih.gov/BLAST/), and 3) *EST-based* - EST_map (http://www.softberry.com/) seeded by EST contigs. Genewise models were extended where possible using scaffold data to find start and stop codons. EST BLAT alignments (Kent, 2002) were used to extend, verify, and complete the predicted gene models. The resulting set of models was then filtered for the best models, based on EST and homology support, to produce a non-redundant representative set. This representative set of 11,464 gene models was subject to further analysis and manual curation. Measures of model quality include proportions of the models complete with start and stop codons (86% of models), consistent with ESTs (48% of models) and supported by similarity with proteins from the NCBI NR database (70% of models) (Table S3). About 90% of the models showed expression in at least one of the conditions (growth in liquid MMN medium, cambial zone of necrotic bark of pines inoculated with *H. irregulare*, fruiting bodies, *H. irregulare* growth on wood shavings from pine, growth in liquid medium amended with lignin or cellulose) analyzed in the NimbleGen array. Characteristics of predicted genes are listed in Table S4. Multigene families were predicted with the Markov clustering algorithm (Enright *et al.*, 2002) to cluster the proteins, using BLASTp alignment scores between proteins as a similarity metric. Orthologs with other sequenced basidiomycete genomes were determined based on best bi-directional blast hits (Table S5).

All predicted gene models were functionally annotated using SignalP (Nielsen *et al.*, 1997), TMHMM (Melen *et al.*, 2003), InterProScan (Zdobnov & Apweiler, 2001), BLASTp (Altschul *et al.*, 1990) against nr, and hardware-accelerated double-affine

Smith-Waterman alignments (deCypherSW; http://www.timelogic.com/decypher_sw.html) against SwissProt (http://www.expasy.org/sprot/), KEGG (Kanehisa *et al.*, 2008), and KOG (Koonin *et al.*, 2004). KEGG hits were used to assign EC numbers (http://www.expasy.org/enzyme/), and Interpro and SwissProt hits were used to map GO terms (http://www.geneontology.org/). Functional annotations are summarized in Table S6. The top 30 PFAM domains are listed in Table S7. Community-wide manual curation of the automated annotations was performed by using the web-based interactive editing tools of the JGI Genome Portal (http://www.jgi.doe.gov/Heterobasidion) to assess predicted gene structures, assign gene functions, and report supporting evidence.

Results

Genome structure

Using a whole genome shotgun approach, the 33.6 MB genome of *H. irregulare* (formerly *H. annosum* North American P-type) was sequenced to 8.5 X coverage. Genome improvement, finishing and gap closure resulted in 33,649,967 bp with an estimated error rate of less than 1 error in 100,000 base pairs. The genome is represented in 15 scaffolds ranging in size from 3,591,957 to 8087 bp. Six of the scaffolds represent complete chromosomes with sequence spanning from telomere to telomere. Seven other scaffolds have an identified telomere only at one end (Fig. 1, Supporting Information). The final assembly statistics are shown in Table S8.

The published linkage map (Lind *et al.*, 2005) was anchored to the sequenced genome using Simple Sequence Repeats (SSR) markers designed from the genome sequence and evenly distributed across the scaffolds (Fig. 1). Segregation analysis of 179 sequence and SSR markers supported a genome organized into 14 chromosomes which is consistent with pulsed-gel electrophoresis data (Anderson *et al.*, 1993). The linkage map was used to locate quantitative trait loci (QTL) for pathogenicity, growth rate and fungal interactions (Olson, 2006; Lind *et al.*, 2007a; Lind *et al.*, 2007b) onto the genome sequence, allowing identification of the genes in the targeted regions.

Transposable elements (TEs) comprised 16.2% of the *H. irregulare* genome which were not uniformly distributed across the scaffolds ($P<0.05$) (Fig. S1 and S2). The Gypsy-like elements were the most frequent TE, corresponding to 9.3% of the assembly. Class II TIR represented the second most frequently categorized elements (1.1%) while 3.7% of the genome comprised TEs belonging to unknown families. The insertion age of full length LTRs shows that *H. irregulare* underwent a recent transposon activity which peaked at an estimated 0.2 Mya and an old activity that occurred at 4-8 Mya (Fig. S3, Supporting Information). The genome of *H. irregulare* contains 100,467 SSRs corresponding to on average 2895 SSR/MB with a distribution about twice as dense in the intergenic as in the intragenic regions (Fig. S2). High frequencies of tri- and hexarepeats, which are less likely to cause frameshift mutations than SSR's with other repeat unites, were found in exonic regions as well as in the 5' UTRs and regions immediately upstream from genes, indicating a possible role of these repeats in overall gene expression (Table S9).

The mitochondrial genome (mt-genome), shown to influence *H. irregulare* virulence (Olson & Stenlid, 2001), spans 114,193 bp and is one of the largest sequenced in fungi. In addition to genes coding for proteins of the oxidative phosphorylation system, we found 14 intron-containing genes, two genes and two pseudogenes probably derived from a mitochondrial plasmid and six non-conserved hypothetical genes (Table S10).

A total of 11,464 gene models have been predicted in *H. irregulare* with half of them shared across Basidiomycotina (Table S5 and S11-S15). The transcription factor distribution comparable with other fungal taxa and the signal transduction pathway are conserved (Fig. S8 – S10). The largest gene families include transporters and signaling domains (MFS, p450, WD40, protein kinases) (Table S7). Sequenced ESTs and microarray analysis supported 90% of the predicted genes. In the microarray experiment, 1615 gene models showed differential expression among probes representing a particular gene model in a given growth condition, indicating alternative gene model structure to those predicted or alternative splice variants present (Supporting Information). In

genomic regions rich in TE, such models were more abundant (Fig. 1). A relatively small fraction (59%) of the gene models in the TE-rich regions compared to the total genome showed similarity to other genes from NCBI. However, the existence of most gene models without homology was verified by ESTs or microarray expression data.

The sexual cycle in *H. irregulare* is controlled by *Mat-A* genes similar to that of other bipolar Agaricomycetes, but the mating type locus may encode either a novel class of *MAT* protein or a highly derived homeodomain transcription factor (Supporting Information). The *MAT–A* locus was found on the largest chromosome and displayed extensive locally conserved gene order with other Agaricomycetes (Fig. S5). A second locus was also identified, encoding a cluster of five pheromone receptor genes and three putative pheromone genes homologous to the genes (*MAT-B*) controlling nuclear behavior and communication in other basidiomycetes. As predicted under the intense balancing selection observed at *MAT* loci, we found that *MAT-A* was highly polymorphic, whereas *MAT-B* genes were not. In a segregation analysis of a progeny array of *H. irregulare*, the putative *MAT-A* genes, but not the putative *MAT-B* genes, co-segregated with mating type demonstrating that the mating is controlled by the MAT-A and not MAT-B locus in this bipolar fungus (Fig. S6).

Wood degradation machinery

Heterobasidion irregulare showed a broad spectrum of carbohydrate active enzymes. The number of glycoside hydrolase (GH), polysaccharide lyase (PL) and carbohydrate esterase (CE) genes is comparable with the white rotter *Phanerochaete chrysosporium* and the mycorrhizal symbiont *Laccaria bicolor* but smaller than in the saprotrophs *Coprinopsis cinerea* and *Schizophyllum commune*, and in the pathogen *Magnaporthe oryzae* (Table S16). However, *H. irregulare* is almost as well equipped regarding gene families involved specifically in plant cell wall degradation as *S. commune* and *C. cinerea*, encoding the enzymes required to digest cellulose, hemicellulose (xyloglucan and its side chains), and pectin and its side chains (Table S17). In contrast, *H. irregulare* has a limited potential for degradation of a second hemicellulose structure, xylan, with only two xylanases from family GH10 and none from family GH11 (Table S17). *H.*

irregulare possesses two GH29 fucosidase genes that might act on living/fresh material and two GH5 (β-mannanases) that may have a role in softwood hemicellulose degradation since softwood is known to contain large proportion of glucomannan (Wiedenhoeft & Miller, 2005).

Growth of *H. irregulare* on various carbohydrate substrates correlated well with its enzymatic repertoire (Supporting Information). For example, as suggested by the presence of an invertase (GH32) gene, a feature shared with many phytopathogens (Parrent *et al.*, 2009), *H. irregulare* thrives on sucrose. Enzymes active on cell wall polymers were prominent during transcriptome analysis; of the 282 carbohydrate-active enzymes present in the *H. irregulare* genome, 36 are more than two fold up-regulated ($P<0.05$) during early wood degradation. This subset is dominated by putative cellulose degrading enzymes in the groups GH1, GH5, GH6, GH7, GH10, GH12, GH45 and GH61 (Supporting Information, Table S18). The pectate lyase and pectin hydrolase (GH28), both up-regulated upon early wood decay of pine, are likely to act on the middle lamellae in a softwood specific manner. Hemicellulose of wood fibres can be degraded by GH5 and GH10, of which five and two members, respectively, were up-regulated. In addition, 14 sugar transporters showed elevated transcript levels during wood degradation as compared to liquid culture growth (Table S18).

Oxidative enzymes implicated in ligninolysis by white rot fungi include lignin peroxidase, manganese peroxidase, glyoxal oxidase and laccase (Hatakka, 1994). With its eight Mn-peroxidases and lack of lignin peroxidases *H. irregulare* has a lower peroxidase potential than *P. chrysosporum* (Fig. S11, Table S19) but a higher number of phenol oxidases and laccases, 18 and 5, respectively. To generate H_2O_2, *H. irregulare* possesses 17 quinone oxidoreductases, 5 glyoxal oxidases, 34 glucose-methanol-choline oxidoreductases and four Mn-superoxide dismutases (Supporting Information, Fig. S12, Table S19).

Generally, genes involved in oxidative lignocellulose degradation showed a lower expression level than carbohydrate hydrolysing enzymes during wood degradation (Table

S18). However, when compared to growth in liquid medium, some of these genes were significantly up-regulated: one of the eight *H. irregulare* Mn-peroxidase genes was expressed three fold higher during wood degradation than in liquid medium. Two of the five glyoxal oxidase genes were significantly up-regulated and one of the GMC-oxidoreductase genes showed 15 fold higher expression levels, whereas only one of the laccase genes was moderately up-regulated. The gene for cellobiose dehydrogenase responsible for the oxidation of cellobiose was up-regulated 14 fold (Table S18). Lipids and proteins are minor constituents of softwood but, being easily digestible substrates, they may be of importance in the early stages of wood colonization. During *H. irregulare* degradation of pine sapwood, lipase and protease genes showed significantly higher expression compared to the liquid culture conditions.

Pathogenicity

Heterobasidion spp. are recognized as producers of at least 10 different secondary metabolites which are produced both in axenic cultures and during interaction with plants and other fungi (Sonnenbichler *et al.*, 1989). Genome analysis using a secondary metabolite unique regions finder (SMURF) web-tool (http://www.jcvi.org/smurf/index.php) and manual curation identified solitary and clustered putative natural product genes (Table S20, S21). The genome of *H. irregulare* contained genes for three polyketide synthases (PKS), 13 nonribosomal peptide synthetase like (NRP-like) enzymes, 3 terpene cyclases and several tailoring enzymes, including one dimethylallyltryptophan synthase (DMATS) predicted to be involved in secondary metabolite production. The phytotoxins fomannosin and fomannoxin were accumulated in culture filtrates of *H. irregulare* (Fig. 2) and terpene cyclase (fomannosin) and DMATS (fomannoxin) genes were identified in the genome (Supporting Information, Table S20).

Three genes encode members of the small (~150 aa) secreted protein family Cerato-platanin (CP) (Supporting Information) originally identified in *Ceratocystis platani* (Pazzagli *et al.*, 1999) (Fig. S7). During interaction with host tissue, necrotrophic plant pathogens produce reactive oxygen species (ROS) that contribute to the host mediated

oxidative stress and facilitate infection. For the production of ROS, fungi utilize NADPH oxidase homologues (NOx) and ferric reductase (FRe) (Gessler *et al.,* 2007). NOxs are necessary for superoxide generation during developmental processes, whereas FRes are required to acquire iron from the infected host, and are potentially related to pathogenicity since this gene family of seven members in *H. irregulare* is large compared to in non-pathogenic basidiomycetes (Supporting Information).

Transcriptome analysis showed that 55 of the 250 most highly expressed genes during pathogenic interaction had secretion signal sequences, and 18 of these encode enzymes putatively active in carbohydrate degradation (Supporting Information, Table S22). Sixty-two genes were differentially expressed ($P<0.05$) in the pathogenic interaction compared to growth on defined liquid medium with 47 up- and 15 down-regulated (Supporting Information, Table S23). Sixteen differentially expressed genes encode proteins which were predicted to be secreted. among which there are five likely to act on carbohydrate substrates (PL1, GH5, GH28, CE16, CBM1), one lipase, two oxidases active on saccharide molecules. The remaining eight with secretion signal showed no similarity with any protein of known function. Two of the differentially expressed genes with secretion signal had conserved four transmembrane domains which suggest a membrane localisation.

By re-mapping virulence data from Lind and colleagues (Lind *et al.,* 2007a) we located three major QTL regions important for pathogenic interactions with Norway spruce and Scots pine, one on scaffold 1 and two on scaffold 12 (Fig. 1). These QTL regions include 178, 142 and 299 predicted gene models, in respectively scaffold, out of which one third had an expression distinct from background and showed no cross-hybridization with plant transcripts (Fig. 3). Transcriptional data from mycelia grown in cambium limited the virulence candidates within the QTLs to a handful of significantly ($P<0.05$) differentially expressed genes The most highly up-regulated ones were a high affinity sugar transporter (70 fold) present in QTL 2 on scaffold 12. In this QTL, a gene model with no sequence homology was found with 4 fold higher expressed in mycelia grown in cambium compared to liquid culture (Fig. 3). In QTL 1 on scaffold 1their was one gene

model with no sequence homology which was significantly lower expressed in mycelia grown on pine compared to in liquid media and a putative flavin containing Baeyer-Villiger monooxygenase, with 9.4 fold higher expressed during infection (Fig. 3). This type of monooxygenase is needed for one of the biosynthetic steps required for the synthesis of phytotoxin fomannosin making it a very strong pathogenicity candidate. Two overlapping secondary metabolite clusters harboring altogether 43 gene models were located in the QTL region on scaffold 12. The clusters include three NRPS (3, 4 and 11), several oxidative enzymes and transport proteins (Table S21). Furthermore, the sequence similarities of the QTL regions to other Basidiomycota genomes were low (Supporting Information) and the frequency of orphan genes was higher than in other parts of the genome (53% and 34%, respectively). Also transposable element density was higher within the pathogenicity QTLs than in other parts of the genome (Fig. 1).

Two way hierarchical cluster analysis of mean gene expression showed that the biological samples were separated according to if mycelia had been grown together with wood components or not (Fig. 4). Mycelia grown in the cambial zone show a distinct pattern in the heat map indicating that interaction with living tissue is very different from the other growth conditions analyzed. Cluster 3, represent a fruit body specifically expressed gene models (Fig. 4). The gene models higher expressed during interaction with living tissue are represented in cluster 1 and 5 (Fig. 4). In cluster 13 gene models are grouped together that are higher expressed in contact with wood components compared with the expression in fruit body and mycelia grown in liquid media (Fig. 4).

Global transcript profiling demonstrated that genes induced during saprotrophic wood degradation but not upon interaction with living host tissue represent a trade-off between the two trophic strategies (Table S24, S25). Gene expression during saprotrophic growth on wood showed highest correlations with gene expression during growth on cellulose and lignin, but much lower with that in the fruiting body. Gene expression during growth in cambial zone of pine showed an intermediate correlation to the levels in wood (Table 1). The number of significantly ($P<0.05$) higher expressed genes was highest during saprotrophic growth on wood followed by fruit body and lowest during growth in

cambial zone of pine (Fig. 5). The majority of genes higher expressed during growth in cambial zone of pine were also significantly ($P<0.05$) higher expressed during growth on wood and include pectinolytic enzymes and part of the cellulolytic capacity (Fig. 5). The significantly ($P<0.05$) higher expressed genes during growth in cambial zone of pine and in fruit body are distinct while some overlap can be found between fruit body and saprotrophic growth on wood.

Discussion

The forest pathogen and wood decayer *H. annosum* (s.l.) uses two ecological strategies; parasitism and saprotrophic wood decay. During the life cycle it infects and lives within standing conifer trees but it also continues colonizing and degrading the dead tissue. Our analyses of gene expression during these two trophic stages reveal a trade-off in terms of restricted energy acquisition. Fewer genes encoding carbohydrate active enzymes and transporters are expressed during pathogenic growth than during saprotrophic wood decay, indicating that the fungus is not using its full capacity for energy acquisition during necrotrophic growth, i.e. its full arsenal of wood degrading enzymes. Instead, the living tissue triggered an expanded metabolic repertoire involving genes associated with e.g., toxin production, protection against plant defenses, handling low oxygen pressure, and other abiotic stresses. We conclude that there is a trade-off present between maximal nutritional gain and access to a different ecological niche that the fungus has to balance.

Gene expression during saprotrophic growth on wood correlated the most with expression during growth on cellulose and lignin, but just intermediately so with expression during growth in cambial zone of pine. Presumably, *H. irregulare* detects wood as a source of cellulose-derived energy, whereas living tissue only partly function in this manner. The genes induced specifically during infectious growth enable *H. irregulare* to access energy sources, such as carbon bound in macro molecules of living organisms, unavailable to other organisms with which it would otherwise compete.

Re-mapping of two different virulence measurements on two hosts, pine and spruce, revealed three major regions on two chromosomes to be involved, harboring 178, 142

and 299 gene models, respectively. These regions are characterized by a high number of transposable elements and orphan genes with no homologues genes reported from other species. These orphan models constitute a resource for the exploration of novel enzymatic functions and biological mechanisms. Together the enrichment in orphan genes and repetitive elements indicates that these are highly dynamic regions with high evolutionary rate. The characteristics, with high number of orphan genes and many repetitive elements, in these regions are comparable with the effector regions identified in *Phytophthora infestans* (Haas *et al.,* 2009).

Transposable elements are not equally distributed within and among the chromosomes. The major TE´s observed are younger than 12 Mya and the decrease detected probably reflects element deterioration leading to loss of ability to detect older elements. Since TE proliferation within the pathogenicity QTLs is clearly younger than speciation (Fig. S3) (Dalman *et al.,* 2010), we hypothesize that transposon activity may have contributed to shaping the species specific characteristics of *H. irregulare*. Transposable elements have been implicated to co-locate with important factors for pathogenicity also in other pathosystems, eg. *Phytophthora infestans* (Cuomo *et al.,* 2007; Haas *et al.,* 2009).

Transcriptome analyses combined with QTL approach proved to be a powerful approach to reduce the number of candidate virulence genes of the QTL regions. Gene models who are present in QTL regions for virulence and significantly up-regulated during pathogenic interaction with pine are strong candidates. Three candidate genes fulfill the criteria; a sugar transporter, a monooxygenase and a gene model without homology to known genes. Since the QTLs are based on a mapping population derived from a cross between *H. irregulare* and *H. occidentale*, these genes are the main candidates to explain the difference in virulence between the species. As host specificity probably plays an important role in speciation, these genes could constitute a crucial step towards understanding the separation of *H. annosum* (s.l.) into separate species.

Hybrids between *H. irregulare* and *H. occidentale* showed that the mitochondria had an influence on pathogenicity in these species (Olson & Stenlid, 2001). Analysis of the

mitochondrial genome also revealed the largest mt-genome among fungal sequences so far, containing 24 gene models in addition to those involved in oxidative phosphorylation. These genes constitute interesting candidates for the mitochondrial connection to virulence.

The *H. irregulare* genome shows a great potential for both saprotrophic and biotrophic lifestyles. It encodes a wide arsenal of enzymes required to digest cellulose, hemicellulose and pectin, making it almost as well equipped regarding plant cell wall degradation as obligate saprotrophs such as *S. commune* and *C. cinerea.* In contrast to *P. chrysosporum*, *H. irregulare* possesses two GH29 fucosidase genes that might act on living/fresh material and two GH5 (β-mannanases) that may have a role in softwood-specific glucomannan-degradation. Furthermore, *H. irregulare* growth on sucrose correlated well with the presence of an invertase (GH32) gene, a feature shared with many phytopathogens. Sucrose is one of the major sugars found in fresh pine stump surface (Asiegbu, 2000) and the capacity to utilize it during the initial phase of colonization might provide *H. irregulare* with a selective advantage compared to saprotrophs lacking invertase activity, which could help explain why *H. annosum* spp. are so competitive in industrially managed forests.

Analysis of culture filtrates revealed presence of fomannosin and fomannoxin while genome analysis identified terpene cyclase and DMATS, possible involved in the respective synthesis of these known phytotoxins. In addition genes predicted to be involved in production of other secondary metabolites were identified. The presence of genes for putative PKS, NRPS-like enzymes and halogenase, however, implies that the biosynthetic capacity of *H. irregulare* has not been fully explored, as to date, no polyketides, non-ribosomal peptides or halogenated compounds have been identified.

Conclusion

Heterobasidion irregulare is an economically vastly important non-model organisms with a uniquely strong potential for versatility between pathogenic interaction and wood decay. Key enzymes and pathways of these central processes come within reach by our

genomic approach. Comparing growth on dead and living tissue, we reveal a switch in gene expression during living host interaction towards toxin production, protection against plant defenses and handling abiotic stress, at the expense of carbohydrate decomposition and membrane transport capacity. Combining QTL and transcriptome analyses proved a powerful approach to elucidate gene sets involved in important phenotypic traits of a species. The virulence QTL regions are characterized by over-representation of transposable elements, orphan and secreted genes. We demonstrated that a limited number of genes fulfill the criteria of both being located within a QTL and significantly differentially expressed during host interaction compared to in liquid media. Some of these genes encode proteins that are linked to secondary metabolite production. This approach enabled the identification of new candidate pathogenicity factors, whilst at the same time limited the number.

Acknowledgements

The work conducted by the U.S. Department of Energy Joint Genome Institute and was supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Financial support from the Swedish Foundation for Strategic Research is gratefully acknowledged.

Authors Information

Assembly and annotations of the *H. irregulare* genome are available from JGI Genome Portal at http://www.jgi.doe.gov/Heterobasidion and deposited at DDBJ/EMBL/GenBank under the accession [TO BE PROVIDED UPON PUBLICATION]. The complete expression dataset is available as series (accession number GSE30230) at the Gene Expression Omnibus at NCBI (http://www.ncbi.nlm.nih.gov/geo/).

References

1. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990.** Basic local alignment search tool. *Journal Molecular Biology* **215**: 403–410.
2. **Anderson M, Kasuga T, Mitchelson K. 1993.** A partial physical karyotype of *Heterobasidion annosum*. In: Johansson M, Stenlid J, eds. Proceedings of the eighth International Conference on Root and Butt Rots, Wik, Sweden and Haikko, Finland, 303-313.
3. **Asiegbu FO. 2000.** Adhesion and development of the root rot fungus (*Heterobsidion annosum*) on conifer tissue: effects of spore and host surface constituents. *FEMS Microbiology Ecology* **33**: 101-110.
4. **Birney E, Durbin R. 2000.** Using GeneWise in the Drosophila annotation experiment. *Genome Research* **10**: 547-548.
5. **Chase TE. 1985.** *Genetics of sexuality and speciation in the fungal forest pathogen* Heterobasidion annosum. PhD thesis, University of Vermont, Burlington, VT, USA.
6. **Cuomo CA, Güldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, Walton J D, Ma LJ, Baker SE, Rep M. *et al.* 2007.** The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* **317**: 1400-1402.
7. **Dalman K, Olson Å, Stenlid J. 2010.** Evolutionary history of the conifer root rot fungus *Heterobasididion annsosum sensu lato. Molecular Ecology* **19**: 4979–4993.
8. **Enright AJ, Van Dongen S, Ouzounis CA. 2002.** An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* **30**: 1575-1584.
9. **Gessler NN, Aver'yanov AA, Belozerskaya TA. 2007.** Reactive oxygen species in regulation of fungal development. *Biochemistry (Mosc)* **72**: 1091-1109.
10. **Gordon D, Abajian C, Green P. 1998.** Consed: A Graphical Tool for Sequence Finishing. *Genome Research* **8**: 195-202.
11. **Green P. 1999.** Phrap, version 0.990329. http://phrap.org.

12. **Haas B, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T. *et al.* 2009.** Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* **461**: 393-398.

13. **Hatakka A. 1994.** Lignin-modifying enzymes from selected white-rot fungi: production and role from in lignin degradation. *FEMS Microbiology Review*. **13:** 125-135.

14. **Huang X, Madan A. 1999.** CAP3: A DNA Sequence Assembling Program. *Genome Research* **9**: 868-877.

15. **Isono K, McIninch JD, Borodovsky M. 1994.** Characteristic features of the nucleotide sequences of yeast mitochondrial ribosomal protein genes as analyzed by computer program GeneMark. *DNA Research* **1**: 263-269.

16. **Jaffe DB, Butler J, Gnerre S, Mauceli E, Lindblad-Toh K, Mesirov JP, Zody MC, Lander ES. 2003.** Whole-genome sequence assembly for mammalian genomes: Arachne 2. *Genome Research* **13**: 91-6.

17. **Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005.** Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. *Genome Research* **110**: 462–467.

18. **Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T. 2008.** KEGG for linking genomes to life and the environment. *Nucleic Acids Research* **36**: 480–484.

19. **Karlsson M, Olson Å, Stenlid J. 2003.** Expressed sequences from the basidiomycetous tree pathogen *Heterobasidion annosum* during early infection of Scots pine. *Fungal Genetics and Biology* **39**: 51–59.

20. **Kent WJ. 2002.** BLAT--the BLAST-like alignment tool. *Genome Research* **12**: 656-664.

21. **Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS. 2004.** A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biology* **5**: R7.

22. **Lind M, Dalman K, Stenlid J, Karlsson B, Olson Å. 2007.** Identification of quantitative trait loci affecting virulence in the basidiomycete *Heterobasidion annosum* s.l. *Current Geneicst* **52**: 35-44.

23. **Lind M, Olson Å, Stenlid J. 2005.** An AFLP-marker based genetic linkage map of *Heterobasidion annosum* locating intersterility genes. *Fungal Genetics and Biology* **42**:519-527.

24. **Lind M, Stenlid J, Olson Å. 2007.** Genetics and QTL mapping of somatic incompatibility and intraspecific interactions in the basidiomycete *Heterobasidion annosum* s.l. *Fungal Genetics and Biology* **44**: 1242-1251.

25. **Marx DH. 1969.** The influence of ectomycorrhizal fungi on the resistance of pine roots to pathogenic infections. *Phytopathology* **59:** 153-163.

26. **Melen K, Krogh A, von Heijne G. 2003.** Reliability measures for membrane protein topology prediction algorithms. *Journal Molecular Biology* **327**: 735-744.

27. **Newton AC, Fitt BDL, Atkins SD, Walters, DR, Daniell TJ. 2010.** Pathogenesis, parasitism and mutualism in the trophic space of microbe-plant interactions. *Trends in Microbiol.* **18:** 365-373..

28. **Nielsen H, Engelbrecht J, Brunak S, von Heijne G. 1997.** Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10**: 1-6.

29. **Niemelä T, Korhonen K. 1998.** *Taxonomy of the Genus* Heterobasidion. In: Woodward S, Stenlid J, Karjalainen R, Hüttermann A, eds. *Heterobasidion annosum*. Biology, Ecology, Impact and Control. Wallingford, UK: CAB International, 1–25.

30. **Olson Å. 2006.** Genetic linkage between growth rate and the intersterility genes S and P in the basidiomycete *Heterobasidion annosum* sensu lato. *Mycoogical Research* **110**: 979-984.

31. **Olson Å, Stenlid J. 2001.** Mitochondrial control of fungal hybrid virulence. *Nature* **411**: 438.

32. **Otrosina WJ, Garbelotto M. 2010.** *Heterobasidion occidentale* sp. nov. and *Heterobasidion irregulare* nom. nov.: A disposition of North American Heterobasidion biological species. *Fungal Bioogyl* **114**:16-25.

33. **Parrent JL, James TY, Vasaitis R, Taylor AFS. 2009.** Friend or foe? Evolutionary history of glycoside hydrolase family 32 genes encoding for sucrolytic activity in fungi and its implications for plant-fungal symbioses. *BMC Evolutionary Biology* **9**: 1-16.

34. **Pazzagli L, Cappugi G, Manao G, Camici G, Santini A, Scala A. 1999.** Purification of cerato-platanin, a new phytotoxic protein from *Ceratocystis fimbriata f. sp. platani*. *Journal of Biological Chemistry* **274**: 24959-24964.

35. **Salamov AA, Solovyev VV. 2000**. Ab initio gene finding in Drosophila genomic DNA. *Genome Research* **10**: 516-522.

36. **Sambrook J, Russell DW. 2001.** *Molecular Cloning: A Laboratory Manual*. New York, NY, USA: Cold Spring Harbor Laboratory Press.

37. **Smit AFA, Hubley R, Green P. 1996-2010.** RepeatMasker Open-3.0.

38. **Sonnenbichler J, Bliestle IM, Peipp H, Holdenrieder O. 1989.** Secondary fungal metabolites and their biological activity, I. isolation of antibiotic compounds from cultures of *Heterobasidion annosum* synthesized in the presence of antagonistic fungi or host plant cells. *Biological Chemistry* **370**: 1295-1303.

39. **Stenlid J. 1985.** Population structure of *Heterobasidion annosum* as determined by somatic incompatibility, sexual incompatibility, and isoenzyme patterns. *Canadian Journal of Botany* **63:** 2268-2273.

40. **Thor M, Ståhl G, Stenlid J. 2005.** Modeling root rot incidences in Sweden using tree, site and stand variables. *Scandinavian Journal of Forest Research* **20**: 165-176.

41. **Vasiliauskas R, Menkis A, Finlay R, D, Stenlid J. 2007.** Wood-decay fungi in fine living roots of conifer seedlings. *New Phytologist* **174**: 441-446.

42. **Woodward S, Stenlid J, Karjalainen R, Huttermann A. 1998.** *Heterobasidion annosum. Biology, Ecology, Impact and Control.* Wallingford, UK: CAB International.

43. **Zdobnov EM, Apweiler R. 2001.** InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**: 847-848.

Table 1. Correlation of global gene expression under different growth conditions.

Condition	Lignin	Cellulose	Wood	Fruit body	Cambial zone
Lignin	1	0.6934[a]	0.7256	0.2965	0.5328
Cellulose		1	0.7365	0.2244	0.4880
Wood			1	0.3699	0.6158
Fruit body				1	0.2489
Cambial zone					1

a r^2

N=3590

Figure legends

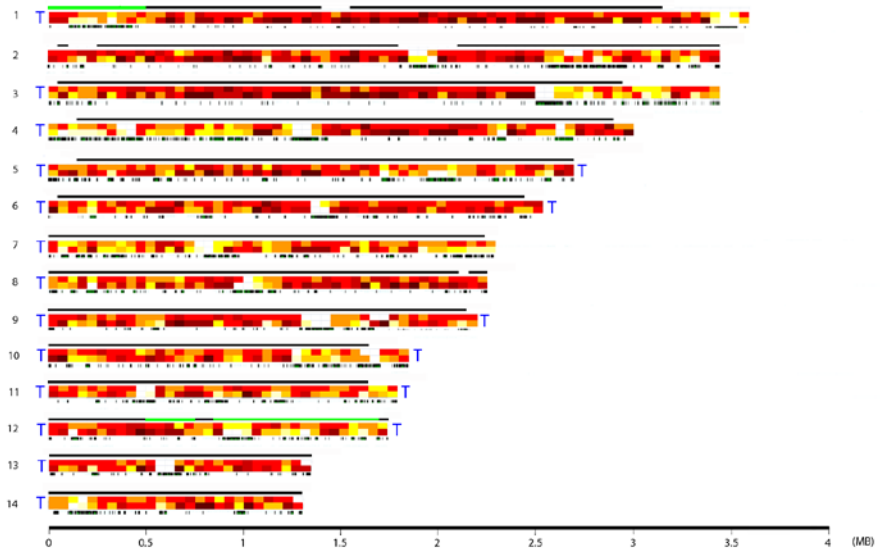


Figure 1. The 14 postulated chromosomes of *Heterobasidion irregulare*. The upper black bar of each chromosome denotes linkage map coverage, with pathogeniticy QTLs marked in green. The wide yellow-to-brown bar describes gene density (upper half) and gene model quality (lower half) for every 50 kB segment of the sequence. Gene density is calculated in number of gene models, ranging from over 27 (brown) to under 10 (white). Gene model quality is calculated based on microarray experiments using five probes for each gene model. The color indicates the percentage of models where all five probes hybridized, ranging from 100% (brown) to below 10% (white). The lowest bar of each chromosome indicates transposon regions, as per masked by RepeatMasker. Blue Ts stands for an identified telomere region in the corresponding chromosome end.

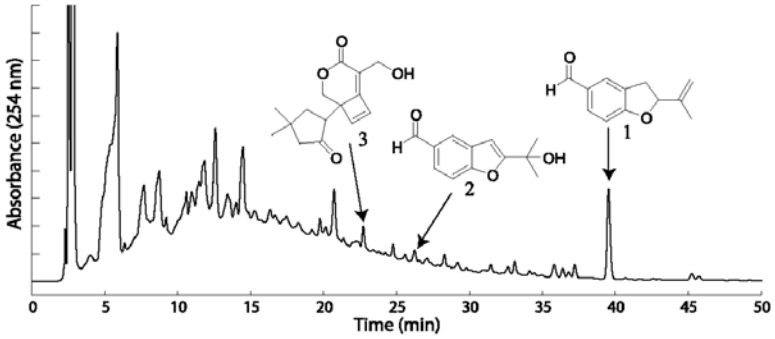


Figure 2. Preparative gradient reversed-phase HPLC chromatogram of *H. irregulare* mono

culture filtrate after solid phase extraction. Selected identified compounds are shown; fomannoxin (**1**) (t_R 39.5), 5-formyl-2-(isopropyl-1´-ol)benzofuran (**2**) (t_R 26.2) and fomannosin (**3**) (t_R 22.7).

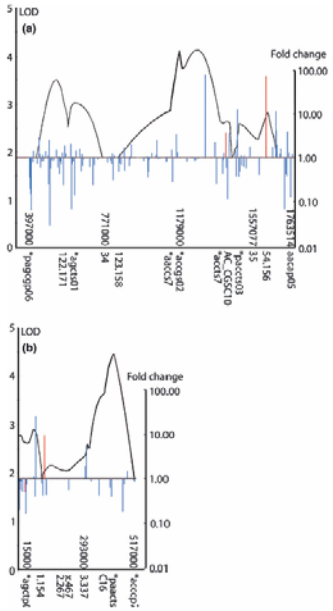


Figure 3. The QTL regions on scaffolds 12 (A) and 1 (B), and within them the up- or down-regulated gene models during mycelial growth in cambium compared to growth in liquid culture. Left y-axis denotes LOD value for QTL effect, with the horizontal bar (LOD 1.9) indicating 5% level of significance. Right y-axis has a logarithmic scale of fold change for gene models, where 1 indicates no change in expression level above background. Blue vertical bars indicate gene models with fold changes not significantly different from background level; red bars indicate models with fold changes significantly ($P<0.05$) different from background. Bottom x-axis shows the markers of the QTL regions and their scaffold positions (bp). The QTL-curve is adjusted to fit physical distance between the markers rather than genetic distance.

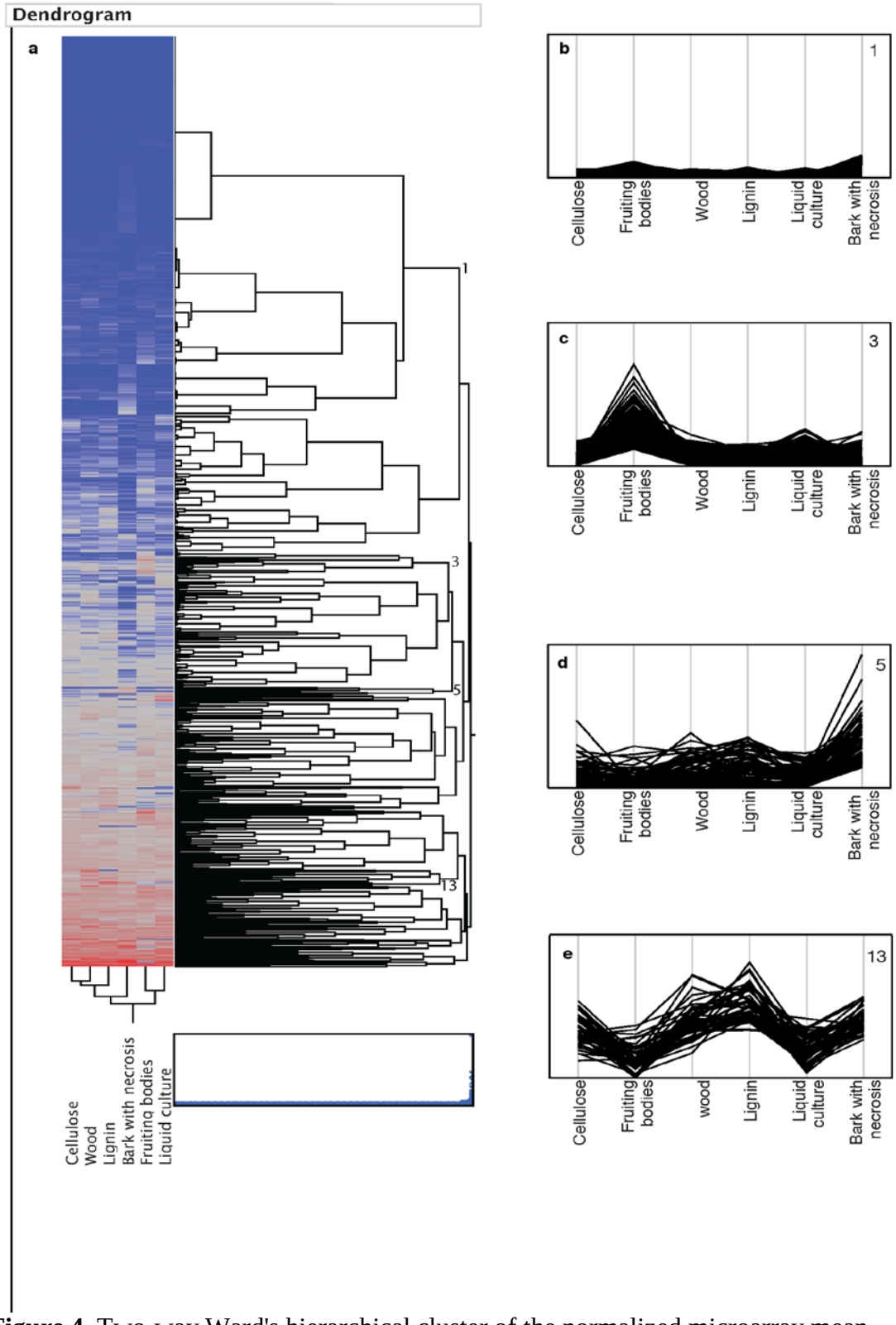


Figure 4. Two-way Ward's hierarchical cluster of the normalized microarray mean

expression, cluster 1, 3, 5 and 13 marked. Parallel plots of B, cluster 1, C, cluster 3, D, cluster 5 and E cluster 10.

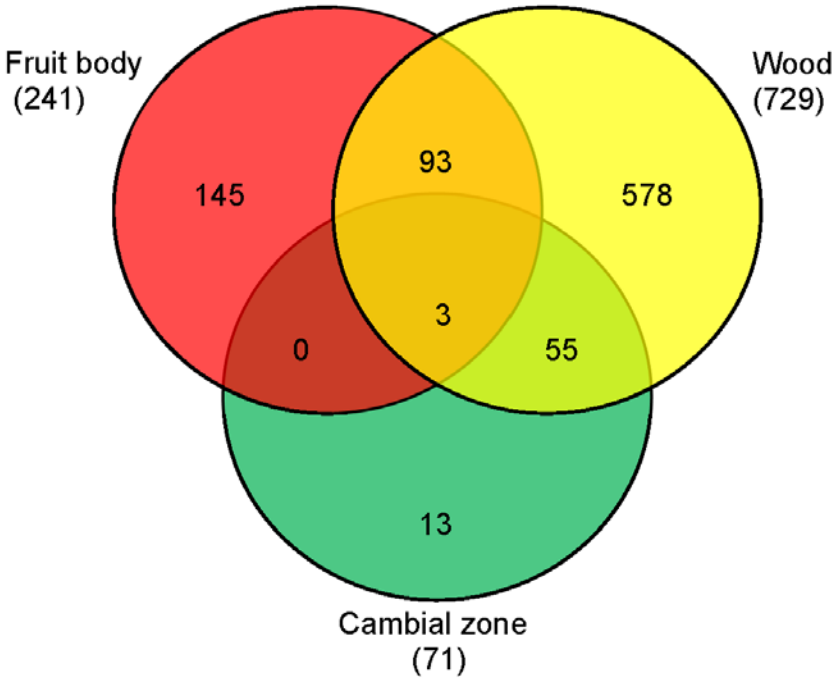
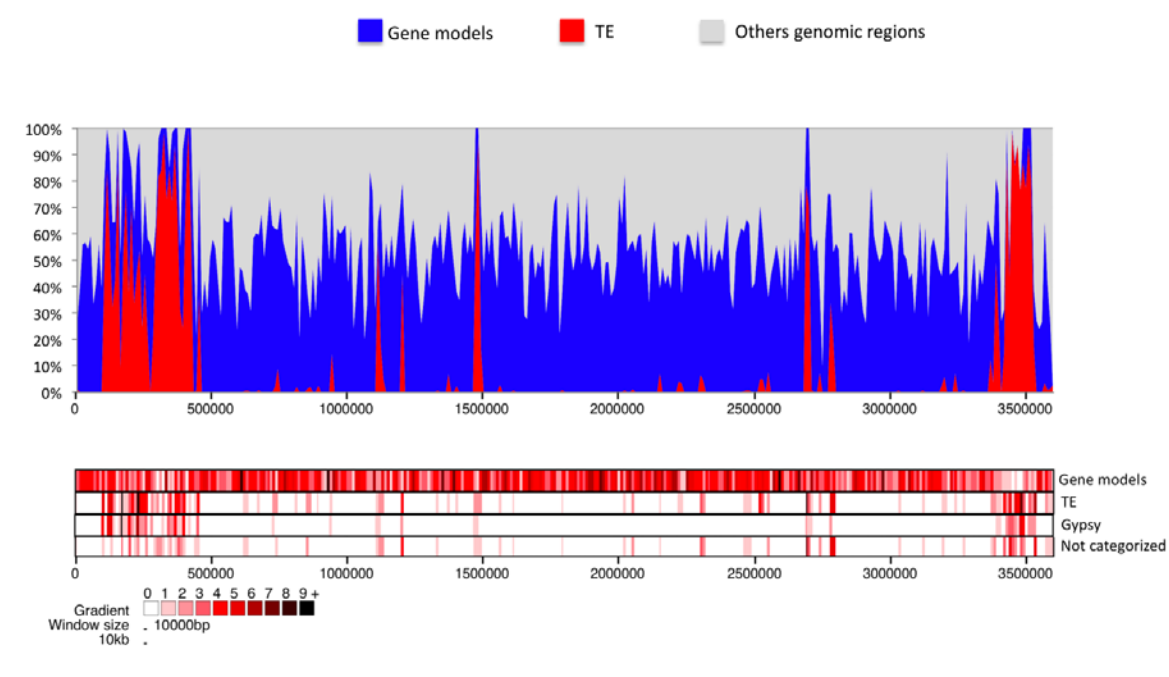


Figure 5. Venn diagram showing the number of significantly up-regulated, unique and common, genes in *H. irregulare* during mycelial growth in wood, growth in cambial zone of pine or in fruit body, compared to gene expression from mycelia grown in liquid culture.

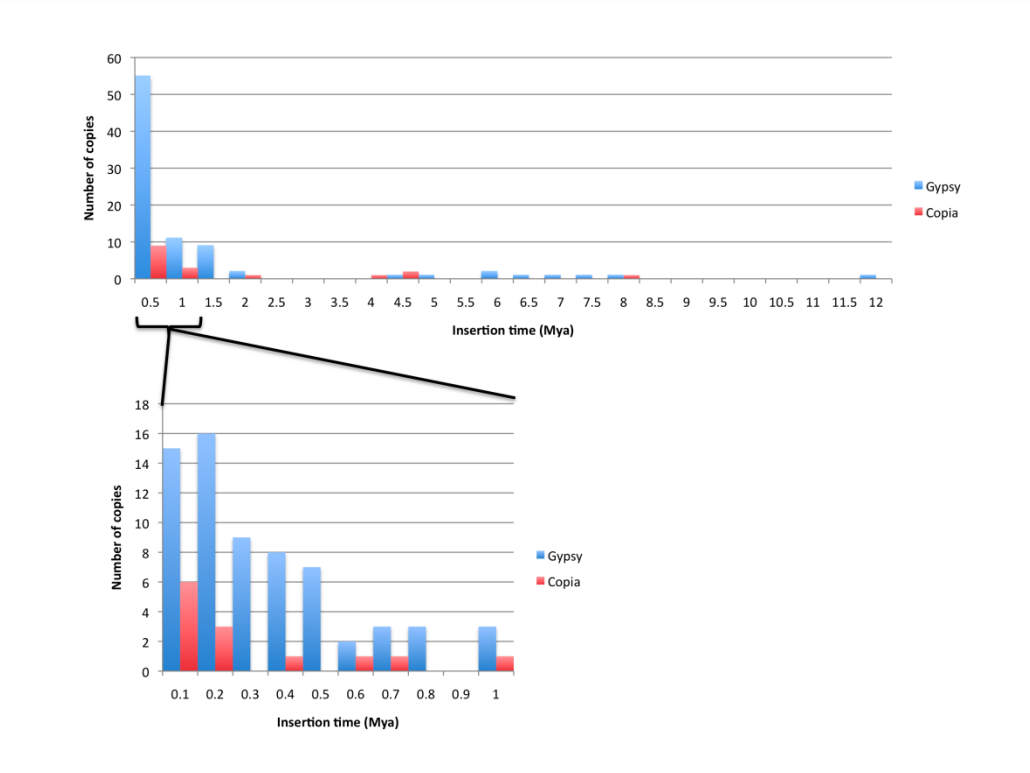
Supporting figures:

Classification		Structure	Number of Consensus Sequences	Number of Occurrences	% Assembly Coverage
Order	**Superfamily**				
Class I (retrotransposons)					
LTR	*Copia*		3 (17)*	184	0.84
	Gypsy		16 (90)*	1,639	9.28
	Others LTR		4	23	0.03
Non LTR			5	152	0.24
LINE			2	30	0.07
***Class II (DNA-transposons)* – Subclass 1**					
TIR	*PIF-Harbinger*		1	49	0.10
	Others TIR		7	341	0.95
***Class II (DNA-transposons)* – Subclass 2**					
Helitron	*Helitron*		6	196	0.64
Elements not categorized			225	2,654	3.67
Total			379**	5,328	16.21

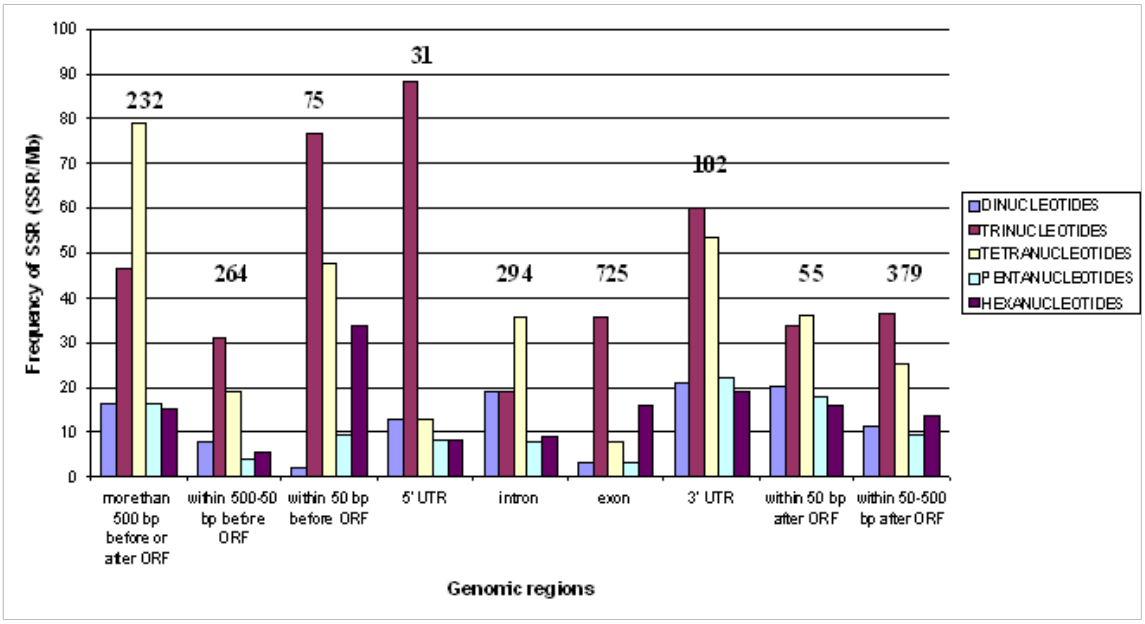
Supplementary Figure 1. The diversity and distribution of class I and class II transposable elements in *H. irregulare*. The putative elements TE structures are depicted according to (Wicker et al., 2007). The number of TE occurrences and the % genome coverage were identified with RepeatMasker (www.repeatmasker.org) using the 379 consensus sequences (**) corresponding to the 272 consensus sequences coming from the RepeatScout/REPCLASS pipeline and the 90 *Gypsy*/*Ty3*-like Full length elements (*) and 17 to *Copia*/*Ty1*-like (*) Full length elements identified by LTR_STRUC (Mc Carthy & Mc Donald, 2003).



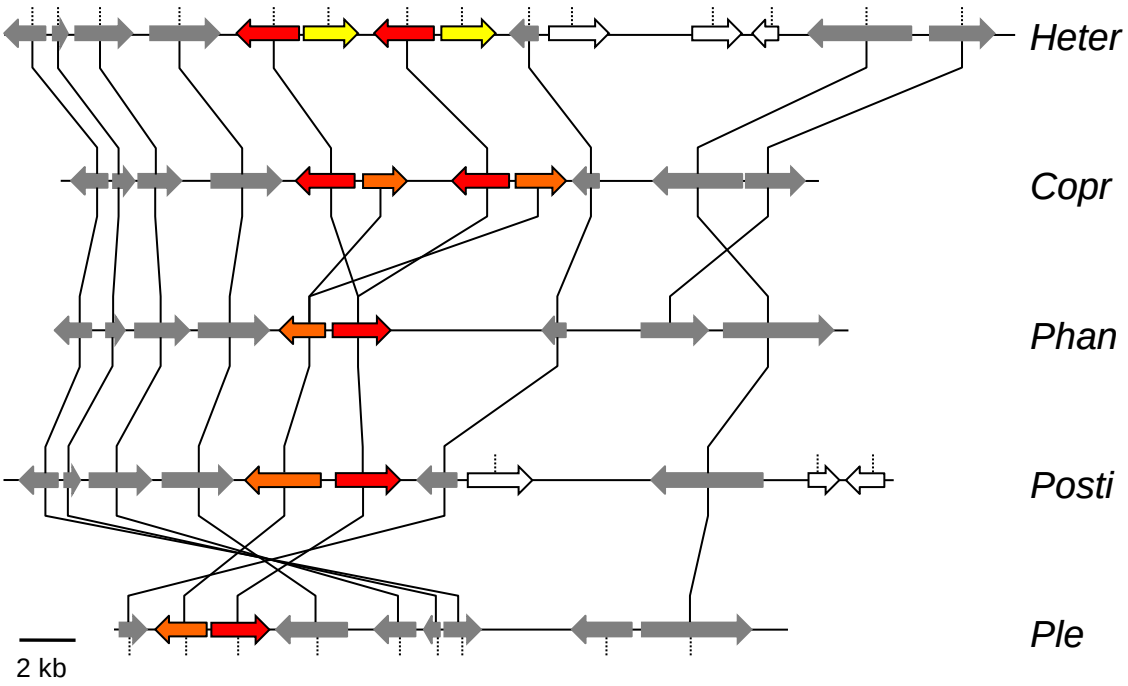
Supplementary Figure 2. Genomic landscape of *H. irregulare*. Area charts quantify transposable elements, genes models and others genomic regions in 10,000 bp windows on the scaffold 1. The y axis represents the percentage of base pairs corresponding to TE (red), genes (blue), and other regions and gaps (grey) in 10,000-bp sliding windows. Heat maps tracks detail the distribution of selected elements (Gene models; TE, transposable elements; Gypsy, LTR-Retrotransposons gypsy and not categorized elements). The maps are realized counting the number of elements in 10,000 bp windows on the scaffold 1. The density has been calculated with Perl and Python scripts available at INRA TuberDB (http://mycor.nancy.inra.fr/IMGC/TuberGenome/).



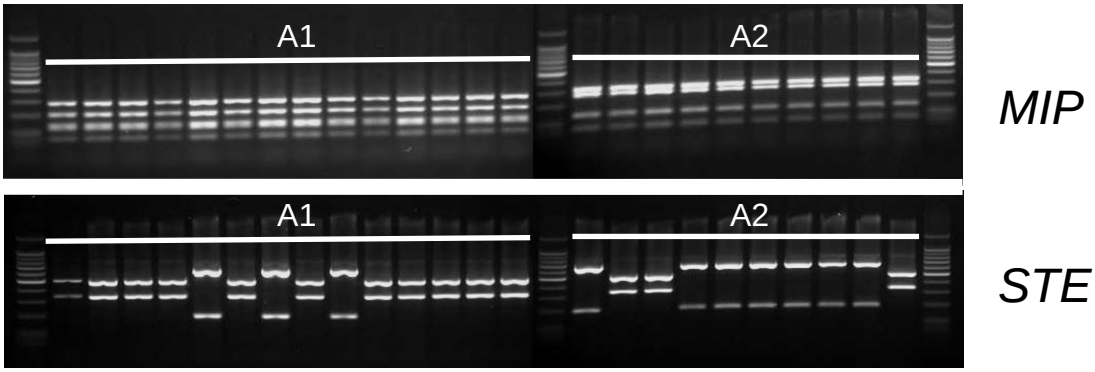
Supplementary Figure 3. Estimated time since the major LTR retrotransposon activity in the *H. irregulare* genome. Consensus full-length copies of the Gypsy and Copia elements are shown. A substitution mutation rate of 1.3 x 10^{-8} was used (Ma & Bennetzen 2004).



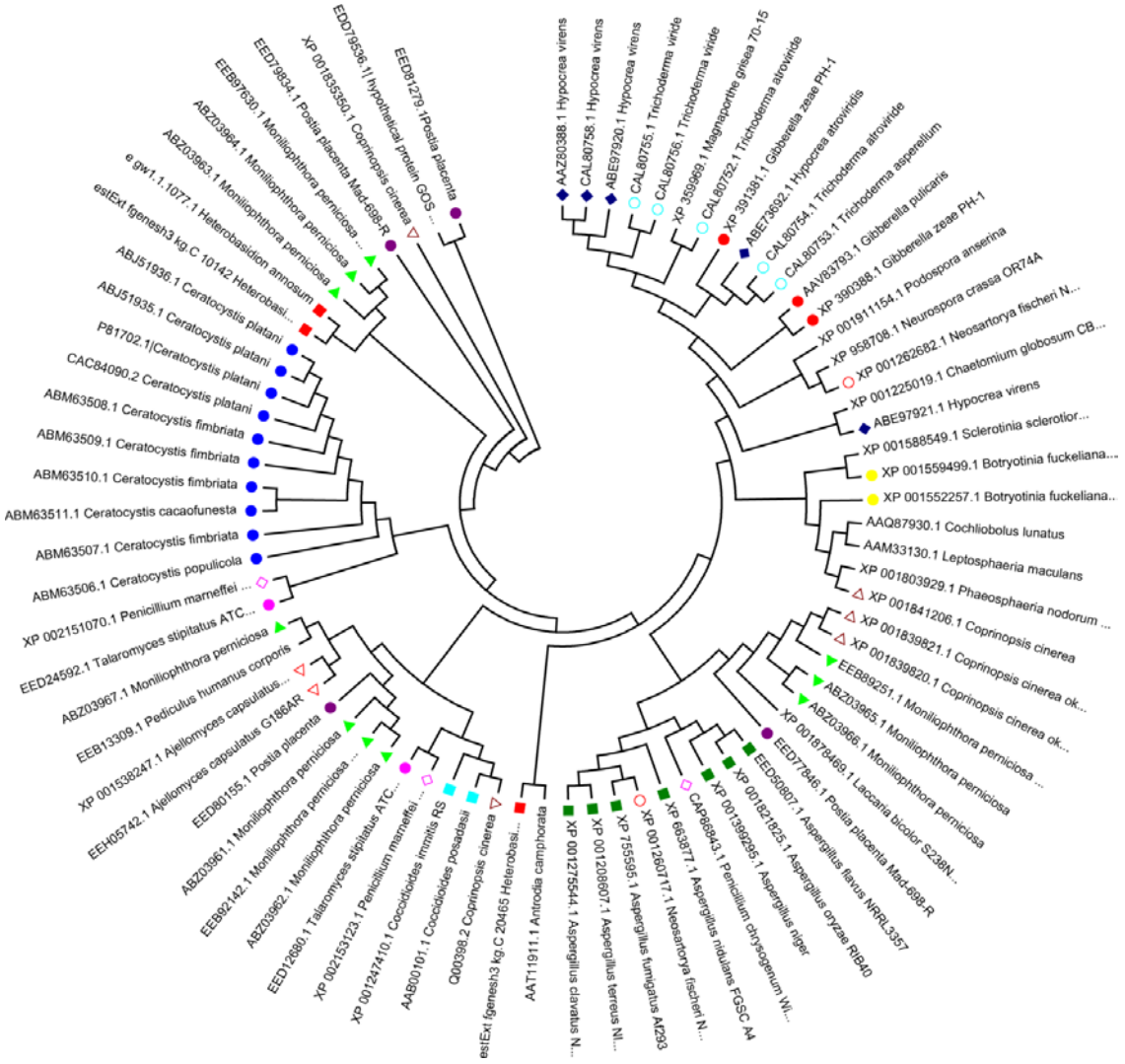
Supplementary Figure 4. Frequency of SSRs in selected genome fractions of *Heterobasidion irregulare*. Numbers represent the absolute frequency of SSRs in each fraction of the 10 largest scaffolds.



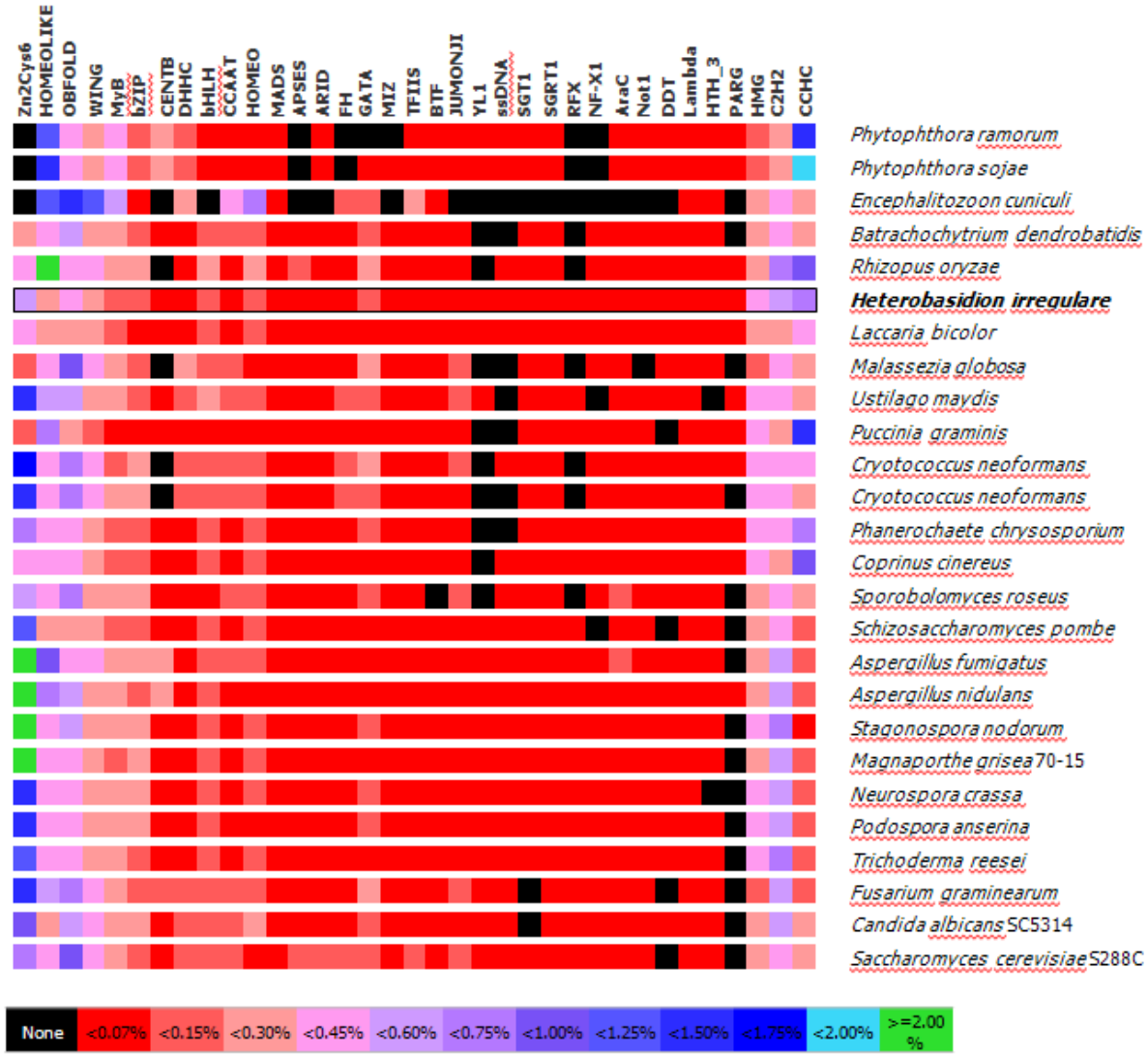
Supplementary Figure 5. Comparative map of gene order surrounding the *MAT-A* locus in representative Agaricomycetes. Lines connect homologous genes. Arrows indicate order of transcription. Grey genes represent those conserved (found near *MAT-A*) in all species, white genes represent those that are not completely conserved, red genes represent the HD1 homeodomain encoding genes, and the orange genes represent the HD2 homeodomain encoding genes. The genes of the *H. irregulare MAT* locus that are located where the HD2 genes are found in other Agaricomycetes are shown in yellow as these *H. irregulare* genes appear to lack the characteristic HD2 DNA-binding domain.



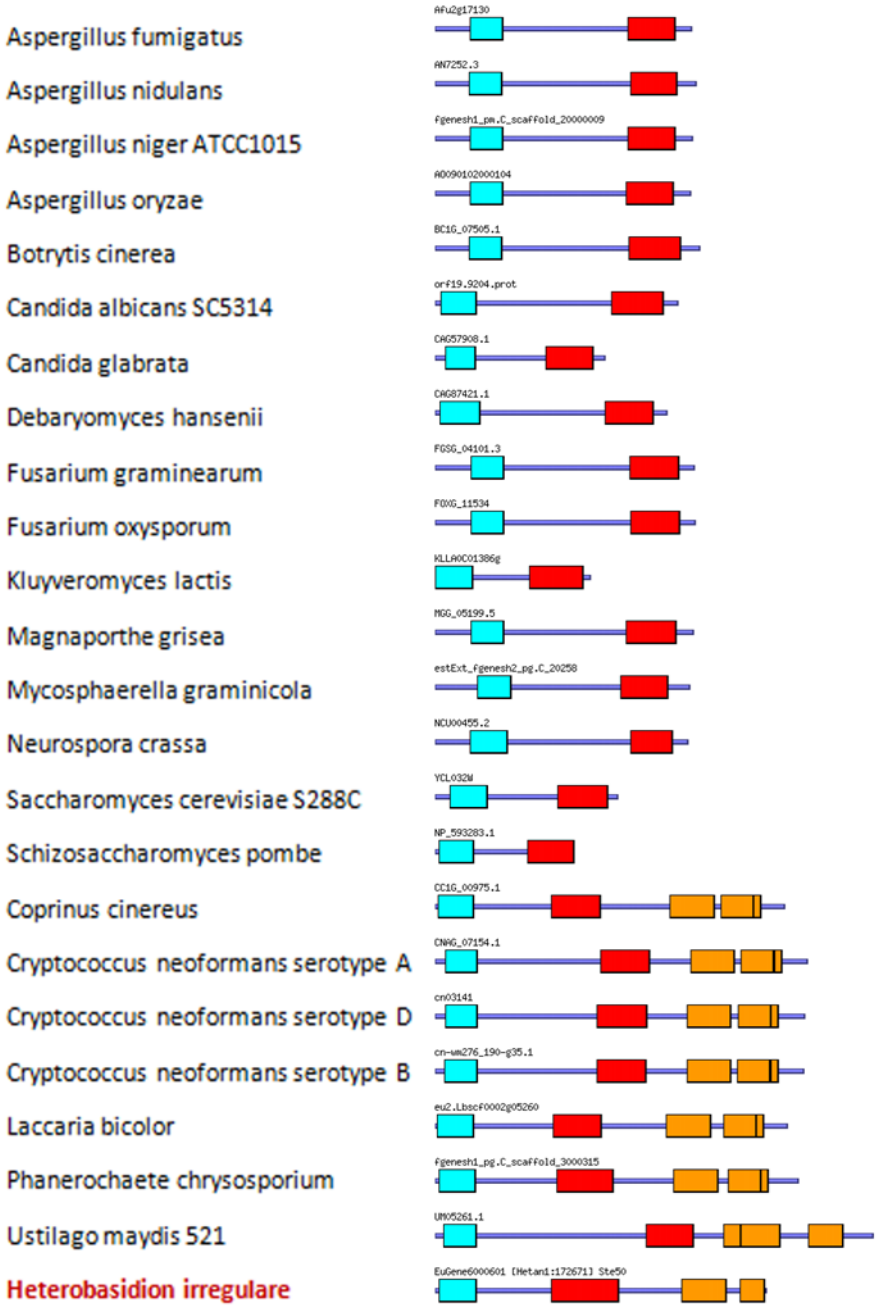
Supplementary Figure 6. Cosegregation of the *MAT-A* region (*MIP*) and mating type among a progeny array of *H. irregulare*. In the top panel the segregation of the 24 isolates at the *MIP* locus is shown, and the bottom panel shows the segregation at the *STE3.3* locus. Homokaryons were assigned to two mating types (A1 or A2) based on heterokaryon formation by pairings on malt extract agar. Genotypes at the putative *MAT-A* region were determined by PCR amplification of the *MIP* gene followed by digestion with *Rsa*I, and genotypes at the putative *MAT-B* region were determined by amplification of a pheromone receptor (*STE3.3*) and digestion with *Aci*I.



Supplementary Figure 7. Unrooted phylogram of Cp protein family including the three *Heterobasidion irregulare* proteins. A total of 77 protein sequences were retrieved from GenBank, edited and aligned using ClustalW, and the unrooted phylogram was prepared using the neighbor-joining method (MEGA 3.1). Protein sequences accession number and species of origin are indicated.

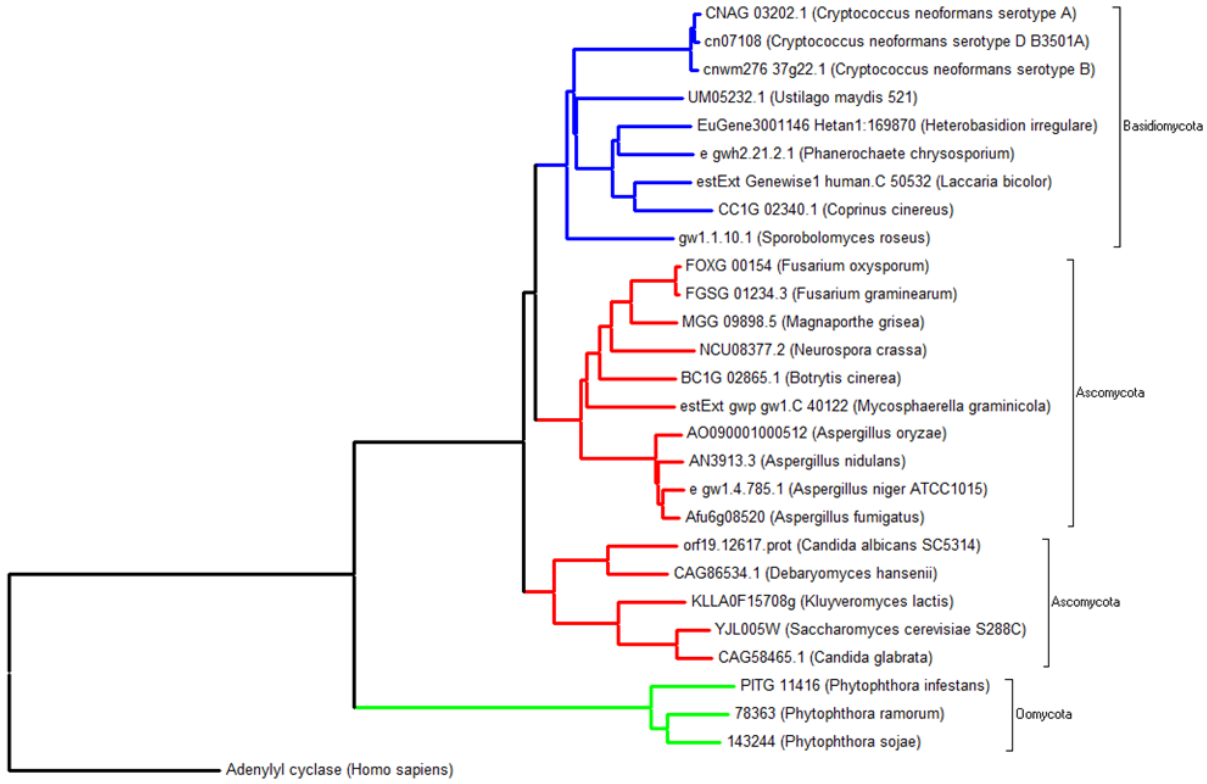


Supplementary Figure 8. TF family distribution across fungal taxa. The matrix of transcription factor configuration from 26 fungal and oomycete genomes. The Legend reflects the proportion of each TF family members against total proteome. *Heterobasidon irregulare* genome is marked as bold and black box

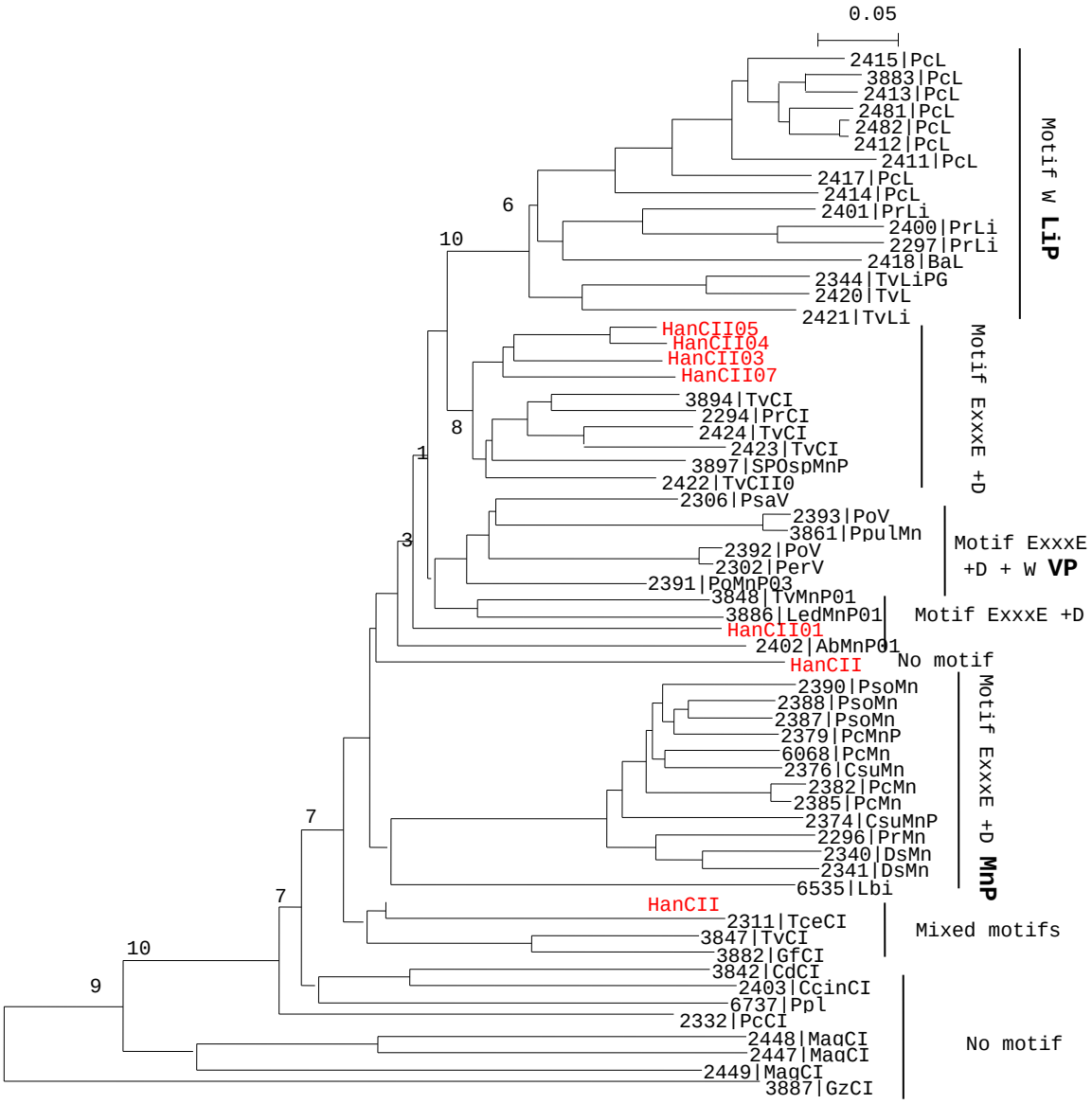


Supplementary Figure 9. Ste50 proteins from basidiomycota and ascomycota.

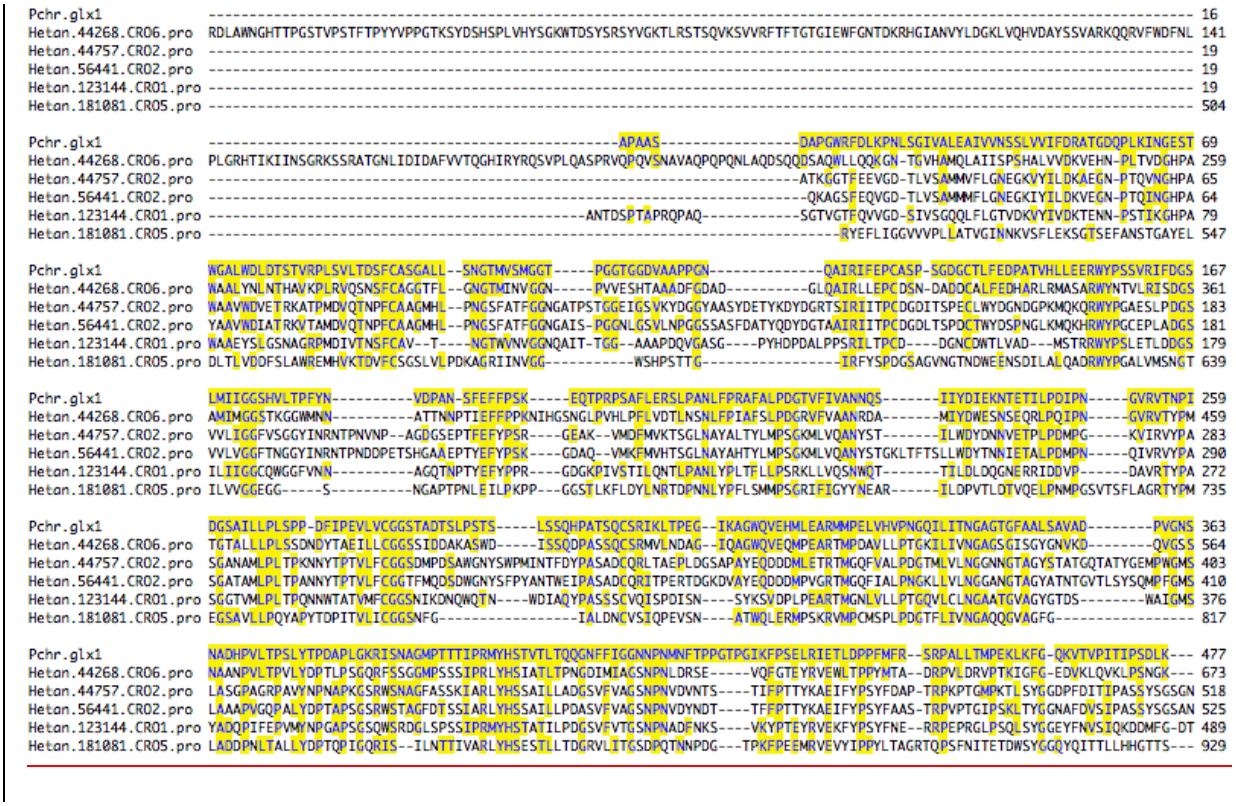
Light-Bluebox = SAMdomain, redbox = kinasedomain, orangebox = SH3domain



Supplementary Figure 10. Phylogenetic analysis of the adenylatecyclase proteins from basidiomycota, ascomycota and oomycota.



Supplementary Figure 11. Phylogenetic analysis of class II peroxidases from various fungal taxa. Protein sequences corresponding to accession numbers and protein names can be found in the PeroxiBase (http://peroxibase.toulouse.inra.fr/). Class II-like protein sequences from non-lignolytic fungi (*Magnaporthe grisea* and *Gibberella zeae*) were used to root the tree.



Supplementary Figure 12. Alignment of five *H. irregulare* protein models and the *P. chrysosporium glx1*. Secretion signals have been removed, as has 4 N-terminal repeats of the WSC domain from *cro5*. Yellow highlights agreement with the *P. chrysosporium* sequence.

Supporting Tables:

Table S1. Genomic libraries included in the *H. irregulare* genome assembly and their respective assembled sequence coverage levels in the whole genome shotgun assembly.

Table S2. Summary statistics of the draft whole genome shotgun assembly of *H. irregulare* before screening and removal of organelles and contaminating scaffolds.

Table S3. Predicted gene models in *H. irregulare* and supporting lines of evidence.

Table S4. Characteristics of predicted gene models in *H. irregulare.*

Table S5. Number of orthologs between *H. irregulare* and 6 other basidiomycetes.

Table S6. Functional annotation of *H. irregulare* proteins.

Table S7. Top 30 Pfam domains in *H. irregulare* proteome.

Table S8. Statistics of the *H. irregulare* assembly.

Table S9. Total length (bp per megabase of DNA) of fully standardized di- and tri- nucleotide repeats in different genomic regions of the *H. irregulare* genome and of other fungal genomes.

Table S10. Protein coding genes, rRNA genes and introns in the mt-genome of *H. irregulare.*

Table S11. Distribution of ROS related proteins among various fungal species.

Table S12. Comparison of numbers of genes putatively involved in lignin degradation in the genomes of several wood degrading basidiomycete fungi.

Table S13. Distribution of genes coding for membrane transporter families in *H. irregulare,* and comparison with other sequenced basidiomycetes.

Table S14. Distribution of genes coding for proteinase families in *H. irregulare,* and comparison with other sequenced basidiomycetes

Table S15. Distribution of proteinase family members in *H. irregulare.*

Table S16. Summary of glycoside hydrolases (GH), polysaccharide lyases (PL) and carbohydrate esterases (CE) genes in *H. irregulare.*

Table S17. Carbohydrate active enzymes of *H. irregulare* active on plant cell walls.

Table S18. Gene models up-regulated >10 fold during *H. irregulare* growth on wood.

Table S19. Location and characteristics of gene models putatively involved in lignin degradation in the *H. irregulare* genome.

Table S20. Annotated putative natural product genes in the *H. irregulare* genome.

Table S21. Putative natural product gene clusters in the *H. irregulare* genome.

Table S22. The 250 highest expressed gene models during *H. irregular* growth in cambial zone of necrotic bark tissue.

Table S23. Gene models up-regulated during *H. irregulare* growth in cambial zone of necrotic bark tissue.

Table S24. Number of carbohydrate active enzymes significantly up-regulated during *H. irregulare* growth in wood and in cambial zone of necrotic bark tissue.

Table S25. Number of transporters significantly up-regulated during *H. irregulare* growth in wood and in cambial zone of necrotic bark tissue.

Supplementary Table 1. Genomic libraries included in the *H. irregulare* genome assembly and their respective assembled sequence coverage levels in the whole genome shotgun assembly.

Library Type	Average Insert Size	Read Number	Assembled Sequence Coverage (X)
3kb	2,716	214,143	3.71
8kb	6,012	192,768	3.56
Fosmid	39,125	63,168	0.96
Total		470,079	8.23

Supplementary Table 2. Summary statistics of the draft whole genome shotgun assembly of *H. irregulare* before screening and removal of organelles and contaminating scaffolds.

Size	Scaffolds (#)	Contigs (#)	Scaffold Size (bp)	Base pairs (bp)	Non-gap (%)
5,000,000	0	0	0	0	0.00
2,500,000	4	193	13,417,098	13,220,956	98.54
1,000,000	14	540	31,427,294	30,865,195	98.21
500,000	15	566	32,363,004	31,754,554	98.12
250,000	18	601	33,597,426	32,962,627	98.11
100,000	19	602	33,715,914	33,081,115	98.12
50,000	19	602	33,715,914	33,081,115	98.12
25,000	20	603	33,751,153	33,116,354	98.12
10,000	24	608	33,803,041	33,168,110	98.12
5,000	30	615	33,845,098	33,209,682	98.12
2,500	45	640	33,896,119	33,255,494	98.11
1,000	51	646	33,904,916	33,264,291	98.11
0	53	648	33,905,970	33,265,345	98.11

Supplementary Table 3. Predicted gene models in *H. irregulare* and supporting lines of evidence.

Gene models (#)	11,464
Complete* gene models (%)	86
With homology support (%)	70
With Pfam domains (%)	44
With EST support (%)	48

* With start and stop codons.

Supplementary Table 4. Characteristics of predicted gene models in *H. irregulare.*

	Average	Median
Gene length (bp)	1601	1283
Protein length (aa)	379	303
Exon frequency per gene (#)	5.39	4
Exon length (bp)	232	145
Intron length (bp)	82	60

Supplementary Table 5. Number of orthologs between *H. irregulare* and 6 other basidiomycetes.

	Orthologs (#)	Identity* (%)
S. commune	6109	57
C. cinerea	6076	56
L. bicolor	5975	59
P. chrysosporium	5576	60
C. neoformans	4533	46
U. maydis	4154	45

* Amino acid identity.

Supplementary Table 6. Functional annotation of *H. irregulare* proteins.

Classification	Proteins # (%)	Categories #
KOG	6698 (58%)	3204
GO	5718 (50%)	2185
EC	2238 (20%)	664
Pfam	5942 (52%)	2306

Supplementary Table 7. Top 30 Pfam domains in *H. irregulare* proteome and 6 other basidiomycetes.

PfamID	Hetan*	Copci	Cneof	Lacbi	Pchr	Umay	Scom	PfamName
PF07690	143	111	161	101	141	90	181	MFS_1
PF00067	129	122	5	69	113	19	106	p450
PF00400	115	109	90	122	109	90	113	WD40
PF00069	102	111	81	120	106	79	116	Pkinase
PF00096	94	60	24	57	35	28	95	zf-C2H2
PF00271	77	91	71	99	59	65	91	Helicase_C
PF00106	62	50	38	48	84	45	114	adh_short
PF00076	61	59	56	65	50	51	65	RRM_1
PF00646	59	178	17	122	70	17	199	F-box
PF00172	58	42	75	69	35	86	79	Zn_clus
PF00023	50	92	17	25	20	15	29	Ank
PF00097	50	38	30	55	28	34	53	zf-C3HC4
PF01266	45	25	28	20	35	18	36	DAO
PF00270	44	52	42	56	33	42	64	DEAD
PF01370	43	26	21	24	50	18	42	Epimerase
PF04082	42	38	60	45	36	31	57	Fungal_trans
PF00083	41	23	62	37	39	28	53	Sugar_tr
PF00248	40	22	20	21	53	14	56	Aldo_ket_red
PF02985	40	38	37	41	42	37	39	HEAT
PF01494	39	22	7	14	31	9	40	FAD_binding_3
PF00004	39	38	31	34	39	28	39	AAA
PF08240	37	29	27	28	47	24	40	ADH_N
PF05199	36	35	2	11	32	10	21	GMC_oxred_C
PF00732	35	36	2	9	31	10	21	GMC_oxred_N
PF00107	35	28	24	21	52	22	44	ADH_zinc_N
PF07993	34	18	16	17	29	16	32	NAD_binding_4
PF00226	33	32	23	36	22	26	27	DnaJ
PF00153	33	35	33	35	34	36	33	Mito_carr
PF08477	32	34	28	54	32	33	33	Miro
PF00515	32	31	22	36	23	25	25	TPR_1
PF07719	32	39	23	33	27	27	29	TPR_2
PF00005	32	43	34	43	50	38	45	ABC_tran

* Hetan = *Heterobasidion irregular*, Copci = *Coprinopsis cinereus*, Cneof = *Cryptococcus neoformans*, Lacbi = *Laccaria bicolor*, Pchr = *Phanerochaete chrysosporium*, Umay = *Ustilago maydis*, Scom = *Schizophyllum commune*

Supplementary Table 8. **Statistics of the *H. irregulare* assembly.**

Main genome scaffold (#)	15
Main genome contig (#)	18
Main genome scaffold sequence (Mb)	33.6
Main genome contig sequence (Mb)	33.6
Main genome scaffold N/L50 (Mb)	6/2.6
Main genome contig N/L50 (Mb)	6/2.3
Number of scaffolds > 50 KB	14
Main genome in scaffolds > 50 KB (%)	100

Supplementary Table 9. Total length (bp per megabase of DNA) of fully standardized di- and tri- nucleotide repeats in different genomic regions of the *H. irregulare* genome and of other fungal genomes.

Motif	*H. irregulare*						Fungi[1]			
	All	Inter[2]	Introns	Exons	5'UTR	3'UTR	All	Inter[2]	Introns	Exons
AC	41	53	102	10	76	79	81	140	511	2
AG	42	46	123	9	89	110	58	136	140	–
AT	16	22	48	1	-	26	133	279	362	2
CG	27	21	57	24	-	26	–	–	–	–
AAC	26	17	10	26	63	263	108	104	145	107
AAG	20	27	5	16	72	57	59	84	34	55
AAT	11	26	5	-	-	-	119	162	537	34
ACC	53	42	17	63	299	148	36	60	14	26
ACG	128	88	79	174	198	218	12	6	–	12
ACT	17	16	53	5	105	33	15	30	–	5
AGC	108	82	52	144	152	175	55	19	103	67
AGG	65	50	31	88	67	108	31	34	45	27
ATC	14	26	-	4	190	-	32	43	53	22
CCG	125	73	105	170	274	222	18	8	20	26

[1] data from Toth et al. (2000).

[2] Intergenic region.

Supplementary Table 10. Protein coding genes, rRNA genes and introns in the mt-genome of *H. irregulare*.

Gene	Feature	Strand	Start	Stop	Note
rnl	rRNA	+	145	5845	Large subunit 23S ribosomal RNA
	intron	+	2245	3852	Intron in rnl
io-rnl	CDS	+	2963	3823	Intronic putative GIY-YIG endonuclease
cox1	CDS	+	7850	23343	cytochrome c oxidase subunit 1
	intron	+	8123	9438	Intron 1 of cox1
	intron	+	9549	10785	Intron 2 of cox1
io-cox1I2	CDS	+	9574	10683	Intronic putative LAGLIDADG endonuclease
	intron	+	11015	12443	Intron 3 of cox1
io-cox1I3	CDS	+	11037	12047	Intronic putative LAGLIDADG endonuclease
	intron	+	12537	16413	Intron 4 of cox1
io-cox1I4a	CDS	+	12546	13439	Intronic putative 2xLAGLIDADG endonuclease
io-cox1I4b	CDS	+	14048	15037	Intronic putative 2xLAGLIDADG endonuclease
io-cox1I4c	CDS	+	15093	15592	Intronic putative endonuclease
ps-io-cox1I4d	CDS	+	15756	16302	Intronic putative pseudo LAGLIDADG endonuclease
	intron	+	16489	17766	Intron 5 of cox1
io-cox1I5	CDS	+	16504	17703	Intronic putative LAGLIDADG endonuclease
	intron	+	17859	18868	Intron 6 of cox1
io-cox1I6	CDS	+	17965	18819	Intronic putative 2xLAGLIDADG endonuclease
	intron	+	19025	20327	Intron 7 of cox1
io-cox1I7	CDS	+	19556	20320	Intronic putative GIY-YIG endonuclease
	intron	+	20379	21605	intron 8 of cox1
io-cox1I8	CDS	+	20484	21173	Intronic putative 2xLAGLIDADG endonuclease
	intron	+	21804	23058	Intron 9 of cox1
io-cox1I9	CDS	+	22023	22772	Intronic putative 2xGIY-YIG endonuclease
rns	rRNA	+	23941	25862	16S ribosomal RNA
cox2	CDS	+	28421	31918	Cytochrome c oxidase subunit 2
	intron	+	28655	30115	Intron 1 of cox2
	intron	+	30572	31852	Intron 2 of cox2
atp8	CDS	+	32525	32683	ATP synthase protein 8
nc-ORF6	CDS	+	33776	35113	Three transmembrane regions found by InterProScan.
ppl2	CDS	+	35371	36093	Putative plasmid protein 2
nad4L	CDS	+	37665	37940	NADH-ubiquinone oxidoreductase chain 4L
nad5	CDS	+	39025	42400	NADH-ubiquinone oxidoreductase chain 5
	intron	+	39742	41143	Intron of nad5
nad6	CDS	+	44999	45616	NADH-ubiquinone oxidoreductase chain 6
nc-ORF5	CDS	+	46939	47874	Three transmembrane regions found by InterProScan.
nc-ORF4	CDS	+	51526	51948	No transmembrane regions found by InterProScan.
nc-ORF3	CDS	-	52918	53775	Two transmembrane regions found by InterProScan.
nc-ORF2	CDS	-	54527	55276	Three transmembrane regions found by InterProScan.
prt_nad2	CDS	-	55334	55552	N-terminal part of NADH-ubiquinone oxidoreductase chain 2
rps3	CDS	-	57959	59218	Putative ribosomal protein subunit 3
nad2	CDS	+	60691	62421	NADH dehydrogenase subunit 2
nad3	CDS	+	62421	62780	NADH dehydrogenase subunit 3
nad1	CDS	+	65252	69492	NADH-ubiquinone oxidoreductase chain 1
	intron	+	65645	67411	Intron 1 of nad1
	intron	+	67676	69132	Intron 2 of nad1

cob	CDS	+	71194	84090	Cytochrome b
	intron	+	71395	73309	Intron 1 of cob
	intron	+	73394	74192	Intron 2 of cob
	intron	+	74287	76437	Intron 3 of cob
	intron	+	76474	78607	Intron 4 of cob
io-cobI4a	CDS	+	76540	77298	Intronic putative endonuclease
io-cobI4b	CDS	+	77447	78133	Intronic putative LAGLIDADG endonuclease
	intron	+	78779	80274	Intron 5 of cob
	intron	+	80494	82023	Intron 6 of cob
	intron	+	82069	83803	Intron 7 of cob
atp9	CDS	+	86234	86455	ATP synthase A chain subunit 9
nad4	CDS	+	87065	88522	NADH-ubiquinone oxidoreductase chain 4
atp6	CDS	+	94023	94808	ATP synthase A chain subunit 6
cox3	CDS	+	99317	102611	cytochrome oxidase subunit 3
	intron	+	99770	101340	Intron 1 of cox3
io-cox3I1	CDS	+	99923	100458	Intronic putative LAGLIDADG endonuclease
	intron	+	101647	102552	Intron 2 of cox3
nc-ORF1	CDS	+	102632	104398	Five transmembrane regions found by InterProScan.
ppl1	CDS	+	106576	107316	Putative plasmid protein 1
ps-dpo1	CDS	-	108422	110487	Pseudo DNA-directed DNA polymerase 1
ps-dpo2	CDS	-	110675	112915	Pseudo DNA-directed DNA polymerase 2

Supplementary Table 11. Distribution of ROS related proteins among various fungal species.

	Trophic strategy	Animal Prx (LDS)	Catalase	DyP-type Prx	Haloperox (haem-non haem)	Non-animal Prx (class I) CP-CcP-hybrid	Class II (LiP-MnP-VP-CII)	Alkylhydroperoxidase D like	Glutathion	Peroxiredoxin	NOx-Fre	Total seq
Ascomycota												
Gibze	NP	3-1	4	ns	5	3-2-1	- - -1	2	1	4	3-1	(31)
Basidiomycota												
Copci	S	2	4	2	3	ns-1-2	- - -1	1	1	5	2-2	(26)
Hetan	NP	**2**	**3**	**1**	**5**	**? -1-4**	**- - -7**	**?**	**?**	**5**	**2-7**	**(32)**
Lacbi	Sy	2	1	2	4	ns-1-1	- - -1	1	1	5	2-ns	(20)
Pchr	S	3	5	ns	3+2	ns-1-ns	9-5- -1	Ns	1	6	2-1	(39)
Pospl	S	2	4	2	5	ns-1-ns	- - -1	Ns	2	5	4-ns	(26)
Umay	BP	1	ns	ns	3	1-2-ns	- - -	2	1	5	ns-2	(17)

Gibze = *Gibberella zeae*, Copci = *Coprinopsis cinereus*, Hetan = *Heterobasidion irregulare*, Lacbi = *Laccaria bicolor*, Pchr = *Phanerochaete chrysosporium*, Umay = *Ustilago maydis*, *Pospl = Postia placenta;* S = saprotroph; P = pathogen; NP = necrotrophic pathogen; BP = biotrophic pathogen; HP = hemotrophic pathogen; Sy: symbiont. ns = no sequence found; ? = lack of sequence probably due to incomplete genome or real absence of sequence.

Supplementary Table 12. Comparison of numbers of genes putatively involved in lignin degradation in the genomes of several wood degrading basidiomycete fungi.

	Hetan	Pchr	Lacbi	*Pospl*	Copci
mcos	18	5	14	7	17
mnp	8	16	1	0	1
cros	5	6	9	3	6
mnsod	4	3	3	1	2
qor	17	23	17	13	4
nor	3	9	1	8	2
sqor	1	1	1	0	1
nuor	1	1	1	0	8
****aao, gor, chd***	34 (16+11+7)	34	14	21	13 (11+1+1)
pdh	–	–	–	–	6
akr	22	24	12	21	18

Hetan = *Heterobasidion irregulare*, Pchr = *Phanerochaete chrysosporium*, Lacbi = *Laccaria bicolor*, *Pospl = Postia placenta,* Copci = *Coprinopsis cinerea;*

* - classification differing among different sources

mcos - Multi copper oxidases (includes laccases)
mnp – Manganese peroxidase (Fungal peroxidase)
cros - Copper radical oxidases (includes glyoxal oxidase, *glx*)
mnsod - Manganese superoxide dismutases
qor - quinone oxidoreductase
nor - NADP-dependent oxidoreductase
sqor - sulfide:quinone oxidoreductase
nuor - NADH:ubiquinone oxidoreductase
aao - aryl-alcohol oxidase
gor - glucose-methanol-choline **(**GMC) oxidoreductase
chd - choline dehydrogenase
akr - Aldo-keto reductase
pdh - pyranose dehydrogenase

Supplementary Table 13. Distribution of genes coding for membrane transporter families in *H. irregulare*, and comparison with other sequenced basidiomycetes.

Transporter Type Family	Hann[1]	Pplac[2]	Lbic[3]	Ccin[3]	Pchrys[3]	Cneo[3]	Umay[3]
ATP-Dependent							
ABC (ATP-binding Cassette)	**35**	67	53	47	45	32	38
ArsAB (Arsenite-Antimonite Efflux Family)	**2**	2	1	1	1	1	1
F-ATPase (F-type, V-type and A-type ATPase)	**55**	57	23	22	21	25	19
MPT (Mitochondrial Protein Translocase)	**20**	11	18	16	13	18	14
P-ATPase (P-type ATPase)	**23**	31	24	16	25	15	18
Sec (General Secretory Pathway)	**17**	15	8	10	8	10	8
TOTAL ATP-Dependent Proteins	**152**	**183**	**127**	**112**	**113**	**101**	**98**
Ion Channels							
Amt (Ammonium Transporter)	**2**	4	8	4	2	2	2
Annexin	**0**	0	1	2	0	1	1
ClC (Chloride Channel)	**3**	4	3	3	3	3	2
Mid1 (Yeast Stretch Activated, Cation-Selective, Ca^{2+} Channel)	**1**	1	1	1	1	1	1
MIP (Major Intrinsic Protein)	**3**	3	7	2	8	1	5
MIT (CorA Metal Ion Transporter)	**4**	4	5	4	5	1	3
MPP (Mitochondrial and Plastid Porin)	**1**	1	1	1	1	1	1
MscS (Small Conductance Mechanosensitive Ion Channel)	**3**	2	3	3	3	1	3
NSCC2 (Non-selective Cation Channel-2)	**1**	1	1	1	1	2	1
TRP-CC (Transient Receptor Potential Ca^{2+} Channel)	**3**	2	2	3	3	4	2
VIC (Voltage Gated Ion Channel)	**4**	8	4	3	2	3	2
TOTAL Ion Channels	**25**	**30**	**36**	**27**	**29**	**20**	**23**
Secondary Transporters							
AAAP (Amino Acid/Auxin Permease)	**4**	3	5	4	6	8	9
ACR3 (Arsenical resistance-3)	**1**	1	3	1	1	1	1
AE (Anion Exchanger)	**1**	2	2	1	1	1	1
APC (Amino Acid-Polyamine Organication)	**20**	34	28	14	19	15	15
ArAE (Aromatic Acid Exporter)	**1**	2	1	1	1	2	2
CaCA (Ca^{2+} Cation Antiporter)	**7**	5	7	7	7	5	4
CCC (Cation-Chloride Cotransporter)	**0**	0	0	0	0	0	1
CDF (Cation Diffusion Facilitator)	**4**	8	8	5	4	5	5
CHR (Chromate Ion Transporter)	**0**	0	4	3	0	0	1
CNT (Concentrative Nucleoside Transporter)	**1**	1	1	1	1	0	1
CPA (Monovalent Cation:Proton Antiporter-1 and -2)	**6**	10	7	7	5	3	6
DASS (Divalent Anion:Na^{+} Symporter)	**1**	1	1	1	1	1	1
DMT (Drug/Metabolite Transporter)	**9**	7	10	11	10	11	10
ENT (Equilibrative Nucleoside Transporter)	**4**	2	1	1	0	1	1
FNT (Formate-Nitrite Transporter)	**1**	0	0	0	1	0	1
GPH (Glycoside-Pentoside-Hexuronide: Cation Symporter)	**2**	2	2	2	2	2	2
GUP (Glycerol Uptake)	**1**	1	1	1	1	0	1
KUP (K^{+} Uptake Permease)	**1**	2	1	0	2	0	0
LCT (Lysosomal Cystine Transporter)	**1**	0	0	0	0	1	1
MC (Mitocondrial Carrier)	**31**	36	36	34	34	34	36
MFS (Major Facilitator Superfamily)	**140**	145	96	97	130	159	89
MOP (Multidrug/Oligosaccharidyl-lipid/Polisaccharide Flippase)	**3**	2	6	2	3	2	2
MTC (Mitochondrial Tricarboxylate Carrier)	**2**	3	0	0	0	1	1
NCS (Nucleobase:Cation Symporter-1 and -2)	**8**	15	14	5	9	8	5
NiCoT (Ni^{2+} -Co^{2+} Transporter)	**1**	0	1	1	1	1	1
Nramp (Metal ion (Mn^{2+}-Iron) Transporter)	**2**	2	1	1	2	1	1
OPT (Oligopeptide Transporter)	**11**	20	10	10	17	6	7
Oxa1 (Cytochrome Oxidase Biogenesis)	**1**	1	1	1	1	1	1
PiT (Inorganic Phosphate Transporter)	**4**	10	0	3	0	1	1
POT (Proton-dependent Oligopeptide Transporter)	**1**	1	2	2	1	1	1
RND (Resistance-Nodulation-Cell Division)	**0**	0	1	1	1	1	1
SSS (Solute:Sodium Symporter)	**3**	8	2	1	2	2	3
SulP (Sulfate Permease)	**3**	4	5	3	4	3	3
TDT (Telurite-resistance/Dicarboxylate Transporter)	**7**	3	1	2	2	2	1
Trk (K^{+} Transporter)	**4**	2	2	2	2	2	1
ZIP (Zinc-Iron Permease)	**4**	3	5	4	6	4	2
TOTAL Secondary Transporters	**290**	**336**	**300**	**245**	**299**	**302**	**238**

Incompletely Characterized Transport Systems							
ATP-E (ATP Exporter)	**1**	1	1	1	1	0	1
Ctr (Copper Transporter)	**3**	2	4	3	3	2	3
Ductin (Putative Ductin Channel)	**2**	2	3	3	3	2	1
FP (Ferroportin)	**1**	1	2	0	1	1	1
ILT (Iron/Lead Transporter)	**1**	1	2	0	1	3	1
LTE (Lipid-Translocating Exporter-I and -II)	**4**	4	7	3	9	8	4
MgtE (Mg^{2+} Transporter-E)	**1**	1	2	1	0	1	1
MHP (Metal Homeostasis Protein)	**2**	1	1	2	2	1	1
MnHP (Mn^{2+} Homeostasis Protein)	**1**	1	0	0	1	1	0
PF27	**1**	1	4	1	1	1	1
PLI (Phospholipid Importer)	**1**	1	1	1	1	1	1
PPI (Protein Importer)	**9**	7	8	7	4	6	6
SHP (Stress-Induced Hydrophobic Peptide)	**4**	2	5	3	2	4	3
YaaH (ATO)	**1**	1	2	3	1	3	3
TOTAL Incompletely Characterized Transport Systems	**32**	**26**	**42**	**28**	**30**	**34**	**27**
TOTAL TRANSPORTER PROTEINS	**499**	**575**	**505**	**412**	**471**	**457**	**386**

Family names are based on the Transport Classification Database (http://www.tcdb.org).
Abbreviations: Hann = *Heterobasidion irregulare*; Pplac = *Postia placenta*; Lbic = *Laccaria bicolor*; Ccin = *Coprinopsis cinerea*; Pchrys = *Phanerochaete chrysosporium*; Cneo = *Cryptococcus neoformans*; Umay = *Ustilago maydis*. [1] = http://genome.jgi-psf.org/Hetan1/Hetan1.home.html
[2] = http://genome.jgi-psf.org/Pospl1/Pospl1.home.html, [3] = Martin et al. 2008 doi:10.1038/nature06556

Supplementary Table 14. Distribution of genes coding for proteinase families in *H. irregulare,* and comparison with other sequenced basidiomycetes

Peptidase Type	Hann[1]	Lbic[2]	Pplac[3]	Pchrys[4]	Ccin[5]
Aspartic peptidases	12	18	22	6	27
Cysteine peptidases	18	22	22	18	24
Glutamic peptidases	0	0	2	11	0
Metallo peptidases	39	51	35	28	54
Serine peptidases	25	27	24	18	20
Threonine peptidases	12	18	8	12	7
Total	**106**	**136**	**113**	**93**	**132**

Classification based on the MEROPS Database (http://merops.sanger.ac.uk/)
1 Hann = *Heterobasidion irregulare*, 2 Lbic = *Laccaria bicolor*, 3 Pplac = *Postia placenta*, 4 Pchrys = *Phanerochaete chrysosporium*, 5 Ccin = *Coprinopsis cinerea*.

Supplementary Table 15. Distribution of proteinase family members in *H. irregulare.*

Family Aspartic	**Hann[1]**	**Lbic[2]**	**Pplac[3]**	**Pchrys[4]**	**Ccin[5]**
A1	10	15	20	1	5
A2	0	1	0	0	0
A11	1	1	0	4	22
A22B	1	1	2	1	0
Total	**12**	**18**	**22**	**6**	**27**
Family Cysteine	**Hann[1]**	**Lbic[2]**	**Pplac[3]**	**Pchrys[4]**	**Ccin[5]**
C1B	1	2	3	1	1
C2	1	1	2	0	3
C12	1	3	2	3	3
C13	1	1	1	1	1
C15	0	0	2	1	0
C19	4	3	1	4	2
C44	3	2	4	1	2
C45	0	0	1	1	0
C48	2	3	1	2	3
C50	1	2	1	1	1
C54	1	1	1	1	1
C56	1	2	1	1	5
C86	1	1	1	0	1
C64-C85-C88	1	1	1	1	1
Total	**18**	**22**	**22**	**18**	**24**
Family Glutamic	**Hann[1]**	**Lbic[2]**	**Pplac[3]**	**Pchrys[4]**	**Ccin[5]**
G1	**0**	**0**	**2**	**11**	**0**
Family Metallo	**Hann[1]**	**Lbic[2]**	**Pplac[3]**	**Pchrys[4]**	**Ccin[5]**
M1	2	1	4	2	1
M3A	2	2	2	2	2
M12A	0	1	0	0	0
M12B	1	1	0	1	1
M13	1	1	1	1	1
M14	1	1	1	1	1
M16	3	7	3	3	6
M17	1	1	1	1	1
M18	1	1	1	1	1
M19	1	0	1	0	1
M20	1	2	1	1	1
M22	2	2	3	2	2
M24	5	5	3	2	3
M28	3	4	1	1	13
M35	1	2	0	0	1
M36	2	6	2	1	8
M38	1	2	1	1	0
M41	2	2	3	3	2
M43B	1	1	0	0	0

M48	1	1	1	1	1
M49	1	1	1	1	1
M50A	1	1	0	0	0
M56	1	0	1	1	1
M67	3	3	1	1	3
M76	1	1	1	1	1
M77	0	2	2	0	2
Total	**39**	**51**	**35**	**28**	**54**
Family Serine	**Hann[1]**	**Lbic[2]**	**Pplac[3]**	**Pchrys[4]**	**Ccin[5]**
S8-S53	10	7	9	2	2
S9	2	6	4	2	4
S10	5	4	2	5	4
S14	1	1	1	1	1
S16	2	2	1	2	2
S26B	1	2	1	1	0
S28	1	2	2	3	3
S33	1	1	2	1	2
S54	1	1	1	0	1
S59	1	1	1	1	1
Total	**25**	**27**	**24**	**18**	**20**
Family Threonine	**Hann[1]**	**Lbic[2]**	**Pplac[3]**	**Pchrys[4]**	**Ccin[5]**
T1	8	11	3	8	4
T2	1	3	1	1	0
T3	2	3	3	2	2
T5	1	1	1	1	1
Total	**12**	**18**	**8**	**12**	**7**

Classification based on the MEROPS Database (http://merops.sanger.ac.uk/)
1 Hann = *Heterobasidion irregulare*, 2 Lbic = *Laccaria bicolor*, 3 Pplac = *Postia placenta*, 4 Pchrys = *Phanerochaete chrysosporium*, 5 Ccin = *Coprinopsis cinerea*.

Supplementary Table 16. Summary of glycoside hydrolases (GH), polysaccharide lyases (PL) and carbohydrate esterases (CE) genes in *Heterobasidion irregulare.*

	GH	PL	CE
H_irre	179	7	18
P_chry	181	4	17
L_bico	163	7	17
C_cine	211	13	51
S_comm	240	16	30
M_gris	232	5	49

H_irre = *Heterobasidion irregulare*
P_chry = *Phanerochaete chrysosporium*
L_bico = *Laccaria bicolor*
C_cine = *Coprinopsis cinerea*
S_comm = *Schizophyllum commune*
M_gris = *Magnaporthe grisea*

Supplementary Table 17. Carbohydrate active enzymes of *H. irregulare* active on plant cell walls.

Cazy family	**Substrate**	**Hetan**	**Lac**	**Cop**	**Phan**	**Cryp**	**Ust**	**Mag**	**Gib**
GH9	CW	1	1	1	1	1	1	0	0
GH1	CW (b-glyc)	2	0	2	2	0	0		
GH2	CW (b-glyc)	3	2	2	2	0	1		
GH5 with CBM1	CW (b-glyc)	3	1	1	4	1	0	1	1
GH6	PCW (cell)	1	0	5	1	0	0	3	1
GH7	PCW (cell)	1	0	7	9	0	0	6	2
GH12	PCW (cell)	4	3	1	2	0	0		
GH45	PCW (cell)	2	0	0	0	0	3	1	1
GH61	PCW (cell)	10	8	33	14	1	0	17	15
GH67	PCW (cell)	0	0	0	0	0	0	1	1
GH74	PCW (cell)	1	0	1	4	0	0	1	1
GH10	PCW (hemi)	2	0	5	6	0	2	5	5
GH11	PCW (hemi)	0	0	6	1	0	1	5	3
GH27	PCW (hemi)	4	1	0	3	0	1		
GH29	PCW (hemi)	2	0	0	0	0	0		
GH35	PCW (hemi)	4	1	0	3	0	1		
GH43	PCW (pect + hemi)	4	0	4	4	0	4	19	16
GH51	PCW (hemi)	1	0	1	2	1	2	3	2
GH53	PCW (hemi)	1	0	1	1	0	0		
Total		**45**	**16**	**69**	**58**	**3**	**15**	**62**	**48**
PL1	PCW (pect)	2	0	1	0	0	1	2	9
PL3	PCW (pect)	0	0	2	0	0	0	1	7
PL4	PCW (pect)	1	0	2	0	1	0	1	3
PL9	PCW (pect)	0	0	0	0	0	0	0	1
GH28	PCW (pect)	8	6	3	4	1	1	3	6

GH78	PCW (pect)	2	0	0	1	2	0		
GH88	PCW (pect)	1	2	1	1	0	0	1	1
GH105	PCW (pect)	2	0	0	0	1	1		
CE8	PCW (pect)	3	3	0	2	0	1	1	5
CE12	PCW (pect)	2							
Total		**21**	**11**	**9**	**8**	**5**	**4**	**9**	**32**

GH = Glycoside hydrolases, PL = Polysaccharide lyases, CE = Carbohydrate esterases, CW = Cell Wall, b-glyc = β-glycans, PCW = Plant Cell Wall, cell = cellulose, hemi = hemicellulose, pect = pectin; Hetan = *H. irregulare*, Lac = *L. bicolor*, Cop = *C. cinerea*, Phan = *P. chrysosporium*, Cryp = *C. neoformans*, Ust = *U. maydis*, Mag = *M. oryzae*, Gib = G. zeae.

Supplementary Table 18. Gene models up-regulated >10 fold during *H. irregulare* growth on wood.

Protein Id.	Fold change*	Description
105463	4785	Glycosyl hydrolase family 61
56987	129	Glycosyl hydrolase family 12
105739	117	No homology
103225	109	No homology
63436	104	Major Facilitator Superfamily
150359	88	FMN-dependent dehydrogenase
63659	85	Glycosyl hydrolase family 61
162562	78	Glycosyl hydrolase family 5
170185	64	No homology
166613	62	Glycosyl hydrolase family 61
16710	59	No homology
152014	55	Glycosyl hydrolases family 28
56288	53	Glycosyl hydrolase family 10
66839	53	Glycosyl hydrolase family 5
151850	52	CBM_1, Fungal cellulose binding domain
119225	47	No homology
68200	40	No homology
104663	33	Pectate lyase
60114	29	Glycosyl hydrolases family 6
163101	27	No homology
42739	26	No homology
104355	26	No homology
64914	25	No homology
14009	23	No homology
145812	23	D-isomer specific 2-hydroxyacid dehydrogenase
11264	23	Aldehyde dehydrogenase
126699	23	Sugar transporter
62185	22	Sugar transporter

100630	21	No homology
125540	21	CBM_1, Fungal cellulose binding domain
153627	21	No homology
55110	19	Sugar transporter
100706	19	No homology
166634	18	Ribonuclease
62980	18	No homology
46058	17	No homology
66540	17	Ammonium Transporter
120243	15	GMC oxidoreductase
168850	15	No homology
107549	15	No homology
108555	15	CBM_1, Fungal cellulose binding domain
157537	14	GMC oxidoreductase
115868	14	No homology
156676	14	Oligopeptide transporter protein
64810	14	Major Facilitator Superfamily
103640	13	No homology
163743	13	No homology
64799	12	No homology
157104	12	Homogentisate 1,2-dioxygenase
65311	12	No homology
101612	12	No homology
44734	12	No homology
106301	12	No homology
120363	12	Glycosyl Hydrolase Family 88
43914	12	RNA 3'-terminal phosphate cyclase
66331	12	G-protein alpha subunit
11198	11	Amino acid permease
52318	11	Glycosyl hydrolase family 1
152576	11	No homology
157157	10	Major Facilitator Superfamily

156533	10	ATPase family associated with various cellular activities

* Fold change between *H. irregulare* samples grown on wood vs. liquid culture (P>0.05).

Supplementary Table 19. Location and characteristics of gene models putatively involved in lignin degradation in the *H. irregulare* genome.

Gene	Gene model	Protein ID	Number of exons	Length (bp)	Length (aa)
Multi copper oxidases					
HaLcc1	EuGene17000143	165789	17	2485	534
HaLcc2	estExt_Genewise 1.C_41971	35202	14	2361	530
HaLcc3	EuGene17000142	165788	17	2556	542
HaLcc4	Genemark.4263_ g	103190	19	2608	534
HaLcc5	Genemark.4967_ g	103894	18	2500	514
HaLcc6	EuGene11000363	163392	21	2708	543
HaLcc7	YAI_1_e_gw1.4. 259.1	181088	18	2567	535
HaLcc8	Genewise1Plus.C _130048	67601	13	2213	518
HaLcc9/ Fet3	estExt_Genewise 1Plus.C_100189	66247	10	2483	629
HaLcc10	EuGene5000736	172039	17	2487	539
HaLcc11	YAI_e_gw1.4.27 2.1	181060	18	2638	525
HaLcc12	YAI_EuGene160 00107	181063	17	2506	535
HaLcc13	dfl_EuGene16000 086	181233/ 127340	21	2688	533
HaLcc14	fgenesh2_pg.C_sc affold_11000009	119423	21	3147	675
HaLcc15	estExt_fgenesh2_ pm.C_110017	157048	22	3094	649
HaLcc16	YAI_estExt_Gene wise1Plus.C_800 82	181064	10	2715	617
HaLcc17	EuGene5000719	172022	5	827	174
HaLcc18	dfl_dfl_EuGene4 001229	181231	17	2536	534
Manganese peroxidases					

(Fungal peroxidases)					
HaMnP1	Genemark.7162_g	106089	11	1659	362
HaMnP2	YAI_EuGene9000355	181068	11	1641	362
HaMnP3	Genemark.2444_g	101371	11	1663	365
HaMnP4	YAI_estExt_fgenesh3_kg.C_30117	181069	10	1668	349
HaMnP5	Genemark.2653_g	101580	11	1617	357
HaMnP6	Genemark.9449_g	108376	18	2018	362
HaMnP7	fgenesh2_pm.C_scaffold_14000147	127157	11	1644	362
HaMnP8	YAI_e_gw1.7.441.1	181106/51741	7	934	183
Quinone oxidoreductases					
HaQOR1	fgenesh2_pm.C_scaffold_1000182	121361 (99223)	6	1333	343
HaQOR2	YAI_gw1.12.856.1	181092	5	1336	371
HaQOR3	estExt_fgenesh3_kg.C_20245	145876	7	2008	346
HaQOR4	estExt_Genewise1Plus.C_50858	63899	6	1374	351
HaQOR5	estExt_Genewise1Plus.C_30278	61341	9	1995	338
HaQOR6	e_gw1.2.1473.1	43155	6	1319	354
HaQOR7	fgenesh2_pm.C_scaffold_4000019	123437	6	1387	354
HaQOR8	EuGene4000635	170620	6	1010	242
HaQOR9	estExt_Genewise1.C_20172	31829	8	1819	366
HaQOR10	estExt_Genewise1.C_21879	32896	6	1432	343
HaQOR11	estExt_fgenesh2_pm.C_20683	154626	6	2770	344
HaQOR12	estExt_fgenesh2_pm.C_40576	155493	9	1527	338
HaQOR13	Genemark.9115_g	108042	6	1293	332

HaQOR14	Genemark.9116_g	108043	4	1440	394
HaQOR15	e_gw1.13.610.1	56575 (126886)	5	1312	352
HaQOR16	estExt_fgenesh3_kg.C_30410	146509	5	1748	356
HaQOR17	estExt_fgenesh2_pg.C_20719	150317	6	1922	367
HaQOR18	estExt_fgenesh2_pm.C_130083	157553	5	1483	350
HaQOR19	estExt_fgenesh2_pm.C_170054	157950	7	1449	261
HaQUI	estExt_Genewise1Plus.C_170060	68456	6	2210	546
HaNOR1	estExt_fgenesh3_kg.C_110174	148544	10	1692	343
HaNOR2	Genemark.9103_g	108030	10	1527	348
HaNOR3	estExt_fgenesh2_pm.C_130077	157548	9	1663	344
HaNuOR	estExt_Genewise1Plus.C_30173	61275	7	2081	496
HaSQOR	estExt_Genewise1.C_20532	32066	9	1961	448
Copper radical oxidases					
HaCRO1	e_gw1.2.104.1	44268 (42959)	3	2464	783
HaCRO2	estExt_fgenesh3_kg.C_30237	146345 (44757)	6	2268	662
HaCRO3	fgenesh2_pm.C_scaffold_3000347	123144	12	2740	638
HaCRO4	e_gw1.13.461.1	56441 (67755)	4	2558	788
HaCRO5	estExt_fgenesh2_pm.C_90119	156743 (181081)	15	4031	997
Glucose-methanol-choline (*GMC*) oxidoreductases					
HaAAO1	estExt_Genewise1Plus.C_41984	63456	16	2641	597
HaAAO2	estExt_Genewise1Plus.C_50835	63889	11	2542	637
HaAAO3	estExt_Genewise1.C_80752	37439	11	2619	613

HaAAO4	estExt_Genewise1Plus.C_90792	66113	10	2345	601
HaAAO5	Genemark.4765_g	103692	11	2508	655
HaAAO6	estExt_Genewise1.C_130445	39531	9	2299	608
HaAAO7	EuGene12000294	163945 (55943)	11	2616	646
HaAAO8	EuGene12000295	163946	10	2355	621
HaAAO9	estExt_Genewise1Plus.C_130459	67820	9	2295	601
HaAAO10	fgenesh2_pg.C_scaffold_13000100	120243	11	2512	592
HaAAO11	estExt_Genewise1Plus.C_130206	67689	12	2683	605
HaChD1	EuGene8000607	174306	1	2584	668
HaChD2	estExt_fgenesh2_pg.C_50196	151291	13	2885	707
HaChD3	estExt_fgenesh2_pg.C_100028	152387	23	3663	556
HaChD4	estExt_fgenesh2_pm.C_60042	155866	8	3261	693
HaChD5	estExt_fgenesh2_pm.C_70087	156219	10	2634	608
HaChD6	estExt_Genewise1Plus.C_41986	63458	17	3169	608
HaChD7	Genemark.4807_g	103734	20	3019	627
HaGOr1	estExt_fgenesh3_kg.C_100019	148218	25	3391	602
HaGOr2	estExt_fgenesh3_kg.C_130049	148859	6	1185	258
HaGOr3	dfl_EuGene9000481	181240	26	3340	656
HaGOr4	Genemark.8837_g	107764	11	2571	643
HaGOr5	EuGene10000119	162514	14	2582	621
HaGOr6	Genemark.3400_g	102327	17	2925	613
HaGOr7	e_gw1.5.100.1	49580	5	2097	593
HaGOr8	estExt_Genewise1Plus.C_160207	68439	10	2338	612

HaGOr9	e_gw1.11.644.1	54732	15	3212	666
HaGOr10	Genemark.7344_g	106271	23	3234	518
HaGOr11	EuGene9000612	174960	25	3267	629
HaGOr12	estExt_fgenesh3_kg.C_130107	148913	22	4326	609
HaGOr13	EuGene7000149	173088	19	3151	628
HaGOr14	Genemark.4871_g	103798	12	2609	661
HaGOr15	e_gw1.4.1199.1	48370	16	3064	621
HaGOr16	Genemark.4868_g	103795	12	2601	655
Manganese superoxide dismutases					
HaMnSOD1	EuGene8000020	173719 (65281)	4	874	206
HaMnSOD2	fgenesh2_pg.C_scaffold_5000056	116985	11	1481	268
HaMnSOD3	EuGene5000065	171368	8	1186	221
HaMnSOD4	EuGene4001014 (estExt_fgenesh3_kg.C_40430)	170999 (146931)	3	1210	368

Supplementary Table 20. Annotated putative natural product genes in the *H. irregulare* genome.

Annotated function	Protein ID	Scaffold	CDS start	CDS end	Protein length (aa)	Domain organization (PKS and NRPS)	Preliminary designation
Polyketide synthase							
	50938	7	280777	287533	2026	KS-AT-ACP-TE/CYC (wA-type)	pks1
	174227	8	1445471	1450470	1537	A-PCP-KS	pks2
	174228	8	1451692	1456389	1452	KS-Red-AT-ACP	pks3
Prenyl transferase (DMATS-type)							
	108351	14	285796	287319	455		ppt4
Prenyl transferase (UbiA-type)							
	47794	4	2361132	2362408	360		ppt1
	142593	5	667698	669322	328		ppt2
	37087	8	293985	295311	327		ppt3
Terpene cyclase							
	181194	6	502645	503864	314		cyc1
	115814	3	1514452	1515667	308		cyc2
	169607	3	2401190	2402411	310		cyc3
Halogenase							
	181189	10	1487768	1489943	525		hal1
	181184	10	1425026	1427204	530		hal2
	181191	5	1919369	1921555	536		hal3
	181192	10	504204	507385	568		hal4
	174229	8	1456740	1458853	536		hal5
Nonribosomal peptide synthetase-like genes							
	66316	10	567244	571769	1420	A-PCP-Red	lys2
	65036	7	1253120	1257095	988	A-PCP-Red	nps1
	173809	8	288211	292432	1080	A-PCP-Red	nps2
	106754	10	1493229	1497466	1099	A-PCP-Red	nps3
	162913	10	1430512	1434727	1092	A-PCP-Red	nps4
	171239	4	2888033	2891777	1005	A-PCP-Red	nps5
	171317	5	22961	26757	1099	A-PCP-Red	nps6
	171792	5	1518523	1523053	1297	A-PCP-Red	nps7

	45017	3	3168510	3172948	1070	A-PCP-Red	nps8
	53480	9	17660	21951	1065	A-PCP-Red	nps9
	57251	14	479146	483397	1091	A-PCP-Red	nps10
	66488	10	1391587	1395814	1089	A-PCP-Red	nps11
	153301	15	487439	491529	1043	A-PCP-Red	nps12
Transcriptional regulator							
	51550	7	291575	294289	711		zfp1

Abbreviations for domains are: KS = keto synthase, AT = acyl transferase, ACP = acyl carrier protein, TE = thioesterase, CYC = cyclase, A = adenylation domain, PCP = peptidyl carrier protein, Red = reductase.

Supplementary Table 21. Putative natural product gene clusters in the *H. irregulare* genome.

Protein ID	Gene position[1]	Scaffold	Putative function	Preliminary designation
Cluster 1				
118599	-5	8	C2HC type Zn finger Transcription factor	
174231	-4	8	No homology	
52305	-3	8	Major facilitator superfamily transporter	
118597	-2	8	No homology	
174229	-1	8	Halogenase	hal5
174228	0	8	Polyketide synthase	pks3
174227	1	8	Polyketide synthase	pks2
Cluster 2				
53871	-2	10	Major facilitator superfamily transporter	
66536	-1	10	Phenylalanine and histidine ammonia-lyase	
106754	0	10	Nonribosomal peptide synthetase	nps3
66532	1	10	Hydroxyindole-O-methyltransferase	
181189	2		Halogenase	hal1
38264	3	10	Isoprenylcysteine carboxyl methyltransferase	
54099	4	10	Cytochrome b/b6	
53908	5	10	Translation initiation factor 3, subunit g	
162935	6	10	No homology	
38257	7	10	Cystathionine beta-lyases/gamma-synthases	
126108	8	10	Major facilitator superfamily transporter	
66519	9	10	No homology	
106746	10	10	No homology	
148353	11	10	Major facilitator superfamily transporter	
Cluster 3				
120490	-14	14	Cytochrome c heme-binding	
67962	-13	14	Glucose/ribitol dehydrogenase	
120488	-12	14	ATP-NAD/AcoX kinase	
164833	-11	14	No homology	
164832	-10	14	No homology	
164831	-9	14	No homology	
157674	-8	14	No homology	
148995	-7	14	Glycosyltransferase Family 1 protein	

57513	-6	14	Flavoprotein monooxygenase	
67959	-5	14	Major facilitator superfamily transporter	
127042	-4	14	Probably part of the transporter model 67959	
12246	-3	14	Flavoprotein monooxygenase	
57258	-2	14	HpcH/HpaI aldolase	
127040	-1	14	Branched chain aminotransferase	
108351	0	14	Prenyl transferase (DMATS-type)	ppt4
Cluster 4				
153301	0	15	Nonribosomal peptide synthetase	nps12
165327	1	15	No homology other model	
165326	2	15	Major facilitator superfamily transporter	
165325	3	15	Rac1 GTPase	
149146	4	15	Uricase (urate oxidase)	
165322	5	15	Mitochondrial ATP synthase g subunit	
127261	6	15	Inorganic phosphate transporter	
165320	7	15	Phosphatidylinositol 4,5-bisphosphate-binding protein	
Cluster 5				
126108	-16	10	Major facilitator superfamily transporter	
66519	-15	10	No homology	
106746	-14	10	No homology	
148353	-13	10	Major facilitator superfamily transporter	
119355	-12	10	No homology	
66514	-11	10	HpcH/HpaI aldolase	
162925	-10	10	No homology	
27192	-9	10	No homology	
21632	-8	10	Cytochrome P450 monooxygenase	cpm34
157009	-7	10	No homology	
162920	-6	10	Amino acid transporter	
162919	-5	10	Amino acid transporter	
162918	-4	10	Cytochrome P450 monooxygenase	cpm20
119348	-3	10	No homology	
54234	-2	10	Major facilitator superfamily transporter	
66504	-1	10	Phenylalanine and histidine ammonia-lyase	
162913	0	10	Nonribosomal peptide	nps4

			synthetase	
157005	1	10	O-methyltransferase, family 2	
181184	2	10	Halogenase	hal2
66497	3	10	Amino acid transporter	
66494	4	10	No homology	
106730	5	10	No homology	
152523		10	No homology	
162905		10	No homology	
162904		10	No homology	
106727	6	10	No homology	
119338	7	10	No homology	
106724	8	10	No homology	
119335	9	10	No homology	
162898	10	10	No homology	
66488	11	10	Nonribosomal peptide synthetase	nps11
Cluster 6				
47230	-3	4	Farnesyl cysteine-carboxyl methyltransferase	
103253	-2	4	Major facilitator superfamily transporter	
63463	-1	4	Major facilitator superfamily transporter	
171239	0	4	Nonribosomal peptide synthetase	nps5
171238	1	4	Aryl-alcohol oxidase 1	aao1
171237	2	4	Aryl-alcohol dehydrogenase 1	aad1
155628	3	4	No homology	
35213	4	4	Major facilitator superfamily transporter	
Cluster 7				
37102	-9	8	ABC transporter	abc2
37098	-8	8	Purine-cytosine permease	
125392	-7	8	Trehalase	
65351	-6	8	Major facilitator superfamily transporter	
105355	-5	8	Metallopeptidase	
37090	-4	8	No homology	
125388	-3	8	Silent information regulator protein Sir2	
51948	-2	8	Glycosyltransferase Family 90 protein	
37087	-1	8	Prenyl transferase (UbiA-type)	ppt3
173809	0	8	Nonribosomal peptide synthetase	nps2
Cluster 8				
147551	-4	7	Major facilitator superfamily transporter	
36468	-3	7	Cell cycle-associated protein	

			Mob1-1	
51550	-2	7	Zinc finger/binuclear cluster transcriptional regulator	zfp1
36465	-1	7	Halotolerance protein HAL3	
50938	0	7	Polyketide synthase	pks1
51087	1	7	No homology	
173037	2	7	Glucose/ribitol dehydrogenase	
Cluster 9				
65741	-12	9	Polyprenyl synthetase	
152162	-11	9	Transcription factor	
152161	-10	9	No homology	
143606	-9	9	Probably part of 65736	
65736	-8	9	Glycoside Hydrolase Family 16 protein	
105798	-7	9	No homology	
105797	-6	9	No homology	
105796	-5	9	Methyltransferase	
148004	-4	9	No homology	
148003	-3	9	Glutathione S-transferase	gst1
105793	-2	9	No homology	
105792	-1	9	No homology	
53480	0	9	Nonribosomal peptide synthetase	nps9
105790	1	9	No homology	
65732	2	9	Mitochondrial inner membrane protein	
156656	3	9	Phosphopantetheine-binding protein	
156655	4	9	Flavoprotein monooxygenase	
Cluster 10				
148408	-11	11	Major facilitator superfamily transporter	
148407	-10	11	Galactokinase	
148406	-9	11	Probably part of model 148407	
148402	-8	11	Major facilitator superfamily transporter	
119449	-7	11	No homology	
144043	-6	11	No homology	
66612	-5	11	Plasma membrane fusion protein PRM1	
106859	-4	11	Mevalonate pyrophosphate decarboxylase	
157056	-3	11	No homology	
157055	-2	11	Vacuolar transporter chaperone 4	
54348	-1	11	GTPase	
66606	0	11	Fatty acid synthase, Acyl-carrier protein synthase	

[1] Position in relation to the putative natural product backbone gene, by definition assigned position 0.

Supplementary Table 22. The 250 highest expressed gene models during *H. irregular* growth in cambial zone of necrotic bark tissue.

Protein Id.	Relativ expression*	Signal peptide	Description of model
37102	51569	NO	ABC-2 type transporter
27192	51053	YES	No homology
125826	50943	NO	GMC oxidoreductase
149112	47321	NO	Aldo/keto reductase family
67711	43642	NO	No homology
171070	43007	NO	No homology
145367	41854	YES	Protein priA
148906	41772	YES	No homology
149208	41461	YES	Pectate lyase
124093	41026	NO	No homology
165326	40899	NO	Major Facilitator Superfamily
60386	40812	NO	Major intrinsic protein
107421	40677	NO	No homology
148898	39637	YES	No homology
44857	38616	NO	Alcohol dehydrogenase
157374	38299	NO	Cytochrome P450
156457	37960	NO	Peroxisomal multifunctional beta-oxidation protein
172978	37442	YES	Glycosyl hydrolases family 43
147528	37366	NO	No homology
23076	37100	YES	No homology
60674	36974	YES	No homology
150620	36751	YES	No homology
51382	36343	YES	No homology
163830	36006	YES	Glycosyl hydrolases family 16
163550	35827	YES	Lipase
59424	35395	NO	Major Facilitator Superfamily
145525	34652	NO	No homology
145128	34359	NO	Metallopeptidase family
106809	34093	NO	Sugar transporter
149091	33817	NO	No homology
121815	33600	YES	SCP-like extracellular protein
155598	33550	NO	Cytochrome P450
147966	33522	YES	Cytochrome P450
156364	33479	YES	Rhamnogalacturonase B
38882	33322	NO	No homology
168124	33141	NO	No homology
43846	33109	NO	No homology
154644	32935	NO	Alcohol dehydrogenase
148394	32877	NO	Aldo/keto reductase family
121927	32753	NO	Oligopeptide transporter protein

156898	32593	NO	Short chain dehydrogenase
148033	32434	NO	No homology
63331	32432	NO	Major Facilitator Superfamily
140638	32279	NO	No homology
172384	32167	YES	Lipase
33761	31819	NO	No homology
155703	31809	NO	Lactate/malate dehydrogenase
33779	31698	NO	ABC transporter
145132	31648	NO	No homology
33584	31548	NO	Mitochondrial carrier protein
153952	31260	NO	Amino acid permease
52607	31206	NO	Translation initiation factor 2
164295	31166	YES	No homology
59167	30962	NO	Short chain dehydrogenase
45732	30909	YES	Pectinesterase
35992	30879	YES	2-nitropropane dioxygenase
171688	30830	NO	No homology
157588	30786	NO	Profilin
36572	30568	YES	Glycosyl hydrolases family 15
146869	30461	NO	Thiolase
147851	30412	NO	ATP synthase
46342	30394	NO	ABC transporter
51894	30170	NO	Major intrinsic protein
116432	30075	NO	NARE associated Golgi protein
64350	29827	NO	Serine/threonine protein phosphatase 2A,
50287	29709	NO	Lipase
66138	29292	NO	No homology
60313	29195	NO	Transcription factor
170624	29168	NO	No homology
34308	29168	NO	Ribosomal protein L6
115608	29016	YES	No homology
154023	28920	NO	No homology
40228	28753	YES	No homology
145190	28738	NO	No homology
116097	28706	NO	ATP synthase subunit 5
149037	28675	NO	Heat shock 70 kDa protein
168597	28443	YES	No homology
117225	28334	YES	Glycosyl hydrolases family 35
68439	28298	YES	GMC oxidoreductase
146825	28271	NO	Alcohol dehydrogenase
43178	28169	NO	No homology
60207	28061	NO	Dihydrodipicolinate synthetase family
39116	27963	NO	Flavin-binding monooxygenase-like
155393	27957	NO	Ubiquitin-like protein
157171	27860	YES	Glycosyl hydrolases family 43
104663	27755	YES	Pectate lyase

36368	27742	NO	Small heat-shock protein
145747	27687	YES	No homology
37087	27665	NO	UbiA prenyltransferase
156257	27524	NO	Ribosomal protein S12
124845	27297	NO	RNA polymerase
35598	27226	NO	Mitochondrial carrier protein
174003	27211	YES	Major Facilitator Superfamily
65379	27158	NO	2OG-Fe(II) oxygenase superfamily
46789	27066	NO	Snf7
156198	27043	NO	No homology
163187	27012	NO	Terpene synthase family
127058	27004	NO	N-Acetylglucosamine kinase
49485	27001	NO	No homology
145402	26944	NO	Aminotransferase
156868	26837	NO	Histone promoter control 2
47144	26725	NO	Arylacetamide deacetylase
146651	26476	NO	Alcohol dehydrogenase
34945	26455	NO	Ribosomal protein S7
122349	26452	NO	Porin
37838	26446	NO	Sugar transporter
39439	26414	NO	Ribosomal L38
143978	26354	YES	No homology
62388	26265	NO	Tubulin
146308	26203	NO	No homology
127052	26110	NO	Acyl-CoA synthetase
102951	26102	NO	No homology
67107	26098	YES	Sugar transporter
62354	26098	NO	ATP-dependent RNA helicase
103048	26093	NO	Myosin tail
154319	26071	NO	1,3-beta-glucan synthase
63020	26060	YES	No homology
155912	26020	YES	No homology
165638	26010	NO	No homology
145565	25951	NO	No homology
67106	25946	NO	No homology
147426	25829	NO	No homology
144624	25767	NO	No homology
146577	25641	NO	Tubulin
147705	25561	YES	No homology
59548	25507	NO	NADH-cytochrome b-5 reductase
64392	25480	NO	Mitochondrial import inner membrane translocase
32273	25475	YES	Glycosyl hydrolases family 13
66401	25452	NO	SURF4
156970	25362	NO	Major Facilitator Superfamily
16804	25317	NO	Ribosomal L28

148542	25276	YES	No homology
156245	25238	NO	No homology
145946	25199	NO	Phosphoglycerate kinase
146268	25179	NO	Phosphatase
108009	25139	NO	Ribosomal_L14
166682	25134	NO	Alkaline phytoceramidase
148178	25071	NO	No homology
63221	25011	NO	No homology
33983	24915	NO	G-protein alpha subunit
152014	24746	YES	Glycosyl hydrolases family 28
165049	24741	NO	No homology
36254	24704	NO	Vesicle coat complex COPI
28158	24684	NO	Heat shock factor binding protein
100407	24604	NO	Transport protein particle
148443	24579	NO	Tubulin
28098	24549	NO	NADH-ubiquinone oxidoreductase
63493	24537	YES	Aspartyl protease
153265	24491	YES	No homology
35032	24471	YES	No homology
66839	24464	YES	Glycosyl hydrolase family 5
148291	24424	NO	S-adenosyl-L-homocysteine hydrolase
173445	24332	NO	No homology
147131	24282	NO	Farnesyl cysteine-carboxyl methyltransferase
166472	24191	NO	No homology
174738	24144	NO	No homology
39599	24119	NO	Aldehyde dehydrogenase
62758	24114	NO	Ribosomal L29
169660	24112	NO	No homology
62256	24018	NO	Proteasome
154292	24015	YES	No homology
172869	24009	NO	No homology
154939	23960	NO	Transcriptional regulatory protein
167124	23872	NO	No homology
147192	23796	YES	No homology
148179	23681	NO	No homology
145578	23587	YES	Glycosyl hydrolases family 72
53076	23526	NO	Glycosyl hydrolase family 5
36688	23506	NO	No homology
140881	23485	NO	Ribosomal protein L13
60007	23463	NO	S-adenosylmethionine synthetase
31216	23459	NO	Short chain dehydrogenase
64341	23438	NO	Sterol O-acyltransferase
146709	23437	NO	Chitin synthase
99958	23405	NO	No homology
145857	23345	NO	Succinyl-CoA:alpha-ketoacid-CoA transferase
156180	23326	NO	No homology

124018	23314	NO	Ribosomal protein L12
156489	23259	NO	NADH-ubiquinone oxidoreductase
152200	23258	NO	20S proteasome
155263	23238	YES	Glycosyl hydrolases family 13
62621	23236	NO	Exosomal 3'-5' exoribonuclease
59062	23151	YES	No homology
35295	23141	NO	Aconitate hydratase
116059	23011	NO	No homology
36248	23000	NO	Ubiquitin-protein ligase
149420	22986	NO	NADH-ubiquinone oxidoreductase
166956	22962	NO	No homology
31921	22951	NO	Arginase
39297	22803	YES	Aspartyl protease
145167	22735	NO	Putative voltage-gated potassium channel subunit beta
39125	22706	NO	Triose-phosphate Transporter
152726	22538	NO	No homology
146873	22447	NO	Aldehyde dehydrogenase
145534	22411	YES	No homology
155410	22410	NO	No homology
147551	22314	NO	Major Facilitator Superfamily
151860	22283	NO	Cyclophilin type peptidyl-prolyl cis-trans isomerase
31144	22177	YES	No homology
118397	22163	YES	Glycosyl hydrolases family 28
146647	22142	NO	Ribosomal protein L6
145615	22064	NO	No homology
67127	22055	NO	Ribosomal S13
38002	22030	NO	Dienelactone hydrolase
58715	22015	NO	Aldo/keto reductase
147624	22006	NO	No homology
155557	21906	NO	Major Facilitator Superfamily
59095	21860	NO	No homology
156524	21805	NO	Peroxidase
154562	21801	NO	Ras family
146969	21621	NO	No homology
105902	21570	NO	No homology
63766	21544	NO	Alpha-ketoglutarate-dependent sulfonate dioxygenase
57035	21488	NO	No homology
157155	21401	NO	GTP-binding protein
153948	21398	NO	No homology
151475	21385	NO	Ergosterol biosynthesis
33531	21028	NO	Ribosomal L39
146879	20993	NO	No homology
31323	20921	NO	Amino acid transporters
154587	20848	NO	Nuclear transport factor 2

155889	20705	YES	Cytochrome P450
145422	20647	NO	Ribosomal L22
147722	20547	NO	F-actin capping protein
157976	20454	NO	Oligopeptide transporter protein
109183	20421	NO	2OG-Fe(II) oxygenase superfamily
154589	20217	NO	ABC transporter
54368	20193	NO	No homology
65589	20163	NO	Cytochrome P450
148385	20145	NO	Hydrolase/acyltransferase
62415	20125	NO	ABC transporter
64200	20074	NO	Fatty acid hydroxylase superfamily
125540	20008	YES	CBM_1, Fungal cellulose binding domain
61998	19749	NO	Sugar transporter
38469	19724	NO	Terpene synthase
100446	19362	NO	No homology
145932	19216	NO	Cell cycle control protein
62632	19143	NO	Ribosomal L18
35255	19045	NO	Glutathione S-transferase
62063	18941	YES	Glycosyl hydrolase family 5
45491	18843	NO	2-methylcitrate dehydratase
43930	18741	NO	Adenylate kinase
147116	18493	NO	Arylacetamide deacetylase
66891	18297	NO	Alpha-methylacyl-CoA racemase
65927	18241	NO	Protein kinase
105516	18182	NO	No homology
126495	17491	NO	Arginase
151850	17418	YES	CBM_1, Fungal cellulose binding domain
173293	17327	NO	Transmembrane amino acid transporter protein
65276	17080	YES	No homology

* Relative expression of gene model during *H. irregulare* growth on bark in units.

Supplementary Table 23. Gene models up-regulated during *H. irregulare* growth in cambial zone of necrotic bark tissue.

Protein ID	Signal peptide	Fold change*	Description
104663	YES	44	candidate pectin lyase, Polysaccharide Lyase Family 1 protein
152014	YES	39	Glycoside Hydrolase Family 28 protein, polygalacturonase
66839	YES	38	Glycoside Hydrolase Family 5 protein, endo-b-1,4-glucanase; N-terminal CBM1 module
115608	YES	12	No homology
145796	YES	12	No homology
125540	YES	11	Carbohydrate-Binding Module Family 1 protein
108555	YES	8	Carbohydrate Esterase Family 16 protein
35715	YES	4	Lipase, class 3
153265	YES	4	No homology
118836	YES	4	No homology
61332	YES	4	No homology
148906	YES	3	No homology
150620	YES	3	No homology
37721	YES	3	Proteins containing the FAD binding domain, FAD linked oxidase
51382	YES	3	No homology
68439	YES	2	Glucose dehydrogenase/choline dehydrogenase/mandelonitrile lyase (GMC oxidoreductase family)
102296	NO	31	No homology
117358	NO	18	No homology
14009	NO	14	No homology
153301	NO	12	L-aminoadipate-semialdehyde dehydrogenase
154438	NO	10	No homology
58532	NO	9	Flavin-containing monooxygenase
64810	NO	9	Major facilitator superfamily
31733	NO	6	Peroxisomal 3-ketoacyl-CoA-thiolase P-44/SCP2
124446	NO	6	Aldehyde dehydrogenase
59167	NO	5	Mitochondrial/plastidial beta-ketoacyl-ACP reductase
154885	NO	5	Flavoprotein monooxygenase, Aromatic-ring hydroxylase
60207	NO	5	Dihydrodipicolinate synthase-like protein
67622	NO	5	Major facilitator superfamily
157466	NO	4	Mandelate racemase/muconate lactonizing enzyme
105333	NO	4	No homology
169569	NO	4	FOG: Zn-finger, C2H2 type
37838	NO	3	MFS sugar transporter
149017	NO	3	Glyoxylate/hydroxypyruvate reductase (D-isomer-specific 2-hydroxy acid dehydrogenase superfamily)
146825	NO	3	Alcohol dehydrogenase, class III;
63331	NO	2	Major facilitator superfamily
169461	NO	2	Isoflavone reductase
149112	NO	2	Aldo/keto reductase family proteins
58654	NO	2	andidate lipase/esterase from carbohydrate esterase family CE10
151860	NO	2	Peptidyl-prolyl cis-trans isomerase
145565	NO	2	No homology
45919	NO	2	No homology

* Fold change between *H. irregulare* samples grown on bark vs. liquid culture (P>0.05).

Supplementary Table 24. Number of carbohydrate active enzymes significantly up-regulated during *H. irregulare* growth in wood and in cambial zone of necrotic bark tissue.

Cazy family	Substrate	On wood	On bark	Trade off*
CBM1		2	1	1
GH3	b-glyc	1	0	1
GH5	b-glyc	2	1	1
GH6	Cell	1	0	1
GH10	Hemi	2	1	1
GH12	Cell	1	1	0
GH61	Cell	3	0	3
GH74	Cell	1	0	1
GH28	Pect	1	1	0
GH43	Pect/Hemi	1	1	0
PL1	Pect	1	1	0
CE16		1	1	0

GH = Glycoside hydrolases, PL = Polysaccharide lyases, CE = Carbohydrate esterases, b-glyc = β-glycans, cell = cellulose, hemi = hemicellulose, pect = pectin;
* Trade off defined as the number of genes expressed on wood but not in bark.

Supplementary Table 25. Number of transporters significantly up-regulated during *H. irregulare* growth in wood and in cambial zone of necrotic bark tissue.

	On wood	On bark	Trade off*
MFS1	12	4	8
Sugar transport	6	1	5
ABC	2	0	2
Oligopeptide	3	0	3
AA permease	1	0	1
Nuclioside permease	3	0	3

* Trade off defined as the number of genes expressed on wood but not in bark.