Genomics Data 14 (2017) 106–113

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Comparative analysis of genome sequences of the conifer tree pathogen, *Heterobasidion annosum s.s.*

Jaeyoung Choi[a,1], Gir-Won Lee[b,2], Ki-Tae Kim[c], Jongbum Jeon[c], Nicolas Détry[a], Hsiao-Che Kuo[a], Hui Sun[a], Fred O. Asiegbu[a], Yong-Hwan Lee[a,c,d,*]

[a] *Department of Forest Sciences, University of Helsinki, 00014 Helsinki, Finland*
[b] *National Instrumentation Center for Environmental Management, Seoul National University, Seoul 08826, Republic of Korea*
[c] *Department of Agricultural Biotechnology, Seoul National University, Seoul 08826, Republic of Korea*
[d] *Center for Fungal Genetic Resources, Plant Genomics and Breeding Institute, and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea*

ARTICLE INFO

Keywords:
Heterobasidion annosum
Annosum root rot
Conifer tree pathogen
Whole genome sequencing
Comparative genomics

ABSTRACT

The causal agent of root and butt rot of conifer trees, *Heterobasidion annosum*, is widespread in boreal forests and economically responsible for annual loss of approximately 50 million euros to forest industries in Finland alone and much more at European level. In order to further understand the pathobiology of this fungus at the genome level, a Finnish isolate of *H. annosum sensu stricto* (isolate 03012) was sequenced and analyzed with the genome sequences of 23 white-rot and 13 brown-rot fungi. The draft genome assembly of *H. annosum* has a size of 31.01 Mb, containing 11,453 predicted genes. Whole genome alignment showed that 84.38% of *H. annosum* genome sequences were aligned with those of previously sequenced *H. irregulare* TC 32-1 counterparts. The result is further supported by the protein sequence clustering analysis which revealed that the two genomes share 6719 out of 8647 clusters. When sequencing reads of *H. annosum* were aligned against the genome sequences of *H. irregulare*, six single nucleotide polymorphisms were found in every 1 kb, on average. In addition, 98.68% of SNPs were found to be homo-variants, suggesting that the two species have long evolved from different niches. Gene family analysis revealed that most of the white-rot fungi investigated had more gene families involved in lignin degradation or modification, including laccases and peroxidase. Comparative analysis of the two *Heterobasidion* spp. as well as white-/brown-rot fungi would provide new insights for understanding the pathobiology of the conifer tree pathogen.

1. Introduction

Heterobasidion annosum is the causal agent of root and butt rot of conifer trees in the forests of northern Europe and North America, causing annual economic loss of 800 million euros in Europe alone [1,2]. *Heterobasidion* species complex comprises the European *H. annosum*, *H. abietinum*, *H. parviporum*, and the North American *H. irregulare* and *H. occidentale*. They exhibit differences in geographic distribution, lifestyles, interfertility, and host ranges [3]. *Heterobasidion* spp. have been studied with regard to ecology, control methods, physiology, and evolution [4–6]. Due to their socioeconomic and biological importance, *Heterobasidion* species complex has been one of the most studied forest fungi [3].

Recently *H. irregulare* TC 32-1 was fully sequenced, which was the first sequenced basidiomycete forest phytopathogen [7]. With the aid of this, genome-wide association study was able to identity single nucleotide polymorphisms (SNPs) associated with fungal virulence [8]. Transcriptomic profiling became more informative by using the gene models from the taxonomically close species, vitalizing gene expression-associated studies [9–11]. However, the sequenced North American isolate *H. irregulare* is genetically and physiologically different from the European *H. annosum* [12,13]. The more genome sequences from various geographic locations would enable more accurate analysis and capture species-specific characteristics as well.

Here, we report the genome sequences of a Finnish isolate of *H. annosum sensu stricto* (isolate 03012) to further understand the pathobiology of this fungus at genome level. In order to assess the macroscopic similarity between *H. annosum* and *H. irregulare*, genome-wide

* Corresponding author at: Department of Agricultural Biotechnology, Seoul National University, Seoul 08826, Republic of Korea.
E-mail address: yonglee@snu.ac.kr (Y.-H. Lee).
[1] Present address: Noble Research Institute, 2510 Sam Noble Parkway, Ardmore, OK 73401, USA.
[2] Present address: Korea DNA Bank, 117, Unjung-ro, Bundang-gu, Seongnam-si, Gyeonggi-do 13461, Republic of Korea.

http://dx.doi.org/10.1016/j.gdata.2017.10.003
Received 21 June 2017; Received in revised form 21 September 2017; Accepted 15 October 2017
2213-5960/ © 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).

alignment and proteins clustering were conducted. Gene synteny analysis showed species-specific genes, implying genomic rearrangements after speciation. In addition, SNPs and short insertions/deletions (InDels) were identified to elucidate microscopic differences between the two species. With the genome sequences of 35 wood-rotting fungi obtained from JGI MycoCosm [14], comparative gene family analysis was conducted to review the white- and brown-rot classification, focusing on their capability to degrade plant materials. As a result, white-rot fungi had more gene families involved in lignin degradation or modification, including laccases and peroxidase, than brown-rot fungi. Collectively, the results presented in this study may provide new insights for understanding the pathobiology of this conifer tree pathogen.

2. Results and discussion

2.1. Comparisons between the genome and proteome sequences of H. annosum and H. irregulare

The genome of *H. annosum sensu stricto* (isolate 03012) was assembled into 31,007,627 bp with GC content of 51.55%, representing 66.5-fold coverage. Subsequent gene prediction showed that *H. annosum* was predicted to have 11,453 protein-coding genes. A series of genomic analyses were carried out so as to identify similarities and differences between the genome sequences of *H. annosum* and *H. irregulare* (Fig. 1). As a result, the genome sequences of *H. annosum* showed similar genomic characteristics with those of *H. irregulare* (Table 1). Protein clustering results showed that 6719 out of 8647 clusters were shared between *H. annosum* and *H. irregulare*. The 6719 clusters encompassed 10,701 and 12,193 proteins from *H. annosum* and *H. irregulare*, respectively, which account for 93.43% and 90.96% of the whole protein sequences. Among the shared clusters, 4804 clusters contained a single protein member from each species. A large number of shared clusters clearly showed their close relatedness with each other. Their close connection was also shown in the whole genome alignment. It showed that 84.38% of *H. annosum* genome sequences were aligned with *H. irregulare* counterparts. In addition, the aligned regions showed 96.06% of sequence identity across the whole genome, on average (Fig. S1). Meanwhile, structural variations, including SNPs and short InDels, further confirmed the taxonomic differences between the two species. A total of 188,190 SNPs were found when the sequencing reads of *H. annosum* were mapped against the genome sequences of *H. irregulare*. The majority of SNPs were found in genic regions (145,640), including exons and introns and untranscribed regions. In spite of macro level similarity between the two species, approximately six SNPs were found in every 1 kb, implying that there were great extent of genome modifications in micro level. In addition, 98.68% of the whole variations were found to be homo-variant, suggesting that there was long enough time for those variations to be fixed. Effects of SNPs and InDels were also pronounced in genic regions, followed by upstream, downstream and intergenic regions (Fig. 2a). Among the SNPs, transitions were far more frequently found than transversions, showing the ratio of transitions to transversions to be 3.57. When investigated the dinucleotide changes, the ratio of two transitions, AT to GC and GC to AT, accounted for 73.90%, on average. Each of the four transversions only accounted for 5.00 to 10.10% of the corresponding types (Fig. 2b). Though 92.11% of InDels showed their length within 10 bp, longer ones, up to 42 bp, were also detected (Fig. 2c). It might also suggest that the two species have gone through long-standing active mutations. This is in the line with the earlier report that the divergence of *H. annosum* and *H. irregulare* occurred around 30 mya [6]. In addition, a genome-wide association study by sequencing 23 haploid *H. annosum s.s.* in low-depth sequencing showed that there were approximately 33,000 SNPs for all individuals [8]. Considering that it was in low coverage, it is in agreement with our results in a way that there are many SNPs to convince they have diverged a long time ago.

2.2. H. annosum-specific genes and unique genomic regions in H. irregulare

From the protein clustering result of the two species, we found 735 and 1624 species-specific genes in 616 and 1312 clusters of *H. annosum* and *H. irregulare*, respectively. Only 40 of 735 (5.44%) protein sequences, belonging to *H. annosum*-specific clusters, had protein domains predicted by InterPro scan. It showed huge difference to the fact that 58.38% of whole protein sequences in *H. annosum* had at least one domain. We hypothesized that species-specific sequences may be more unstructured than the rest of the proteome, since they presumably have emerged in comparatively recent date. In fact, 25.79% and 30.52% of amino acids in the *H. annosum* and *H. irregulare*-specific protein sequences were predicted to be global and local disorder, respectively. The rest of protein sequences in *H. annosum*, however, showed that 15.33% and 18.33% of amino acid residues were predicted to be locally and globally unstructured, respectively. It was also true for *H. irregulare*-specific proteins. Among the *H. irregulare*-specific proteins, 32.58% and 39.38% of amino acid residues were predicted as disordered, while 15.81% and 19.08% were marked as disordered in the others. It supports the fact that the proteins coded by species-specific genes tend to be unstructured than those conserved throughout other species [15]. It may address the observation that *H. annosum*-specific proteins had much less predicted domains compared to the whole proteome.

In order to see distribution of the *H. annosum* and *H. irregulare*-specific genes in other species, BLAST search was performed against the 37 rot fungi including the two *Heterobasidion* spp. A total of 468 and 806 of *H. annosum* and *H. irregulare*-specific protein sequences, respectively, did not show any hits in the other rot fungi genomes. In addition, 464 out of the 468 *H. annosum* sequences did not have any significant hits when searched against the NCBI non-redundant (NR) database, suggesting that orphan genes may encode them. Interestingly, four of the 468 protein sequences only had hits from bacterial species in the BLAST search, of which the hits were found in mainly sphingomonads belonging to the genera *Sphingobium* and *Novosphingobium* (Table S1). These bacterial species have been reported to have similar natural wood habitat as *H. annosum* [16]. Moreover, some isolates of these species are capable of degrading lignin or aromatic compounds like lignin-derived monoaryls and biaryls [17,18]. Since these genes were not found in *H. irregulare*, it is speculated that genes encoding these proteins might be acquired from cohabiting sphingomonad bacteria. On the other hand, there were *H. irregulare*-specific regions retrieved from the dot plot (Fig. S1), which are 10 kb or longer. A total 1.93 Mb covering 35 regions showed that the repetitive sequences were enriched in comparison with the rest of *H. irregulare* genome sequences (Table S2). These regions encompassed 633 genes, of which 278 genes did not show any significant hits in any species other than *H. irregulare* from BLASTP searches against the NR database. Furthermore, 190 out of 278 genes were not found in *H. annosum* (Fig. S2). The results suggest that species-specific evolution is differently occurring in the two species for adapting to their own niches.

2.3. Evolutionary implication of gene synteny between the two Heterobasidion spp.

In order to speculate potential driving force underpinning the speciation of the two *Heterobasidion* spp., macroscopic and microscopic comparisons were conducted as shown above. The macroscopic comparison showed that the majority of contigs of *H. annosum* genome were mapped against the scaffolds in *H. irregulare*. According to SNP/InDel identification result, which is a microscopic level of genome comparison, six SNPs were identified in every 1 kb, on average. These comparisons showed that there might have been prolonged genomic rearrangements for adapting to their own niches. To further compare the two genomes at an intermediate scale, genome-wide gene synteny was investigated as well as non-syntenic regions in between. As a result, a total of 528 sets of syntenic genes was identified, encompassing 6894

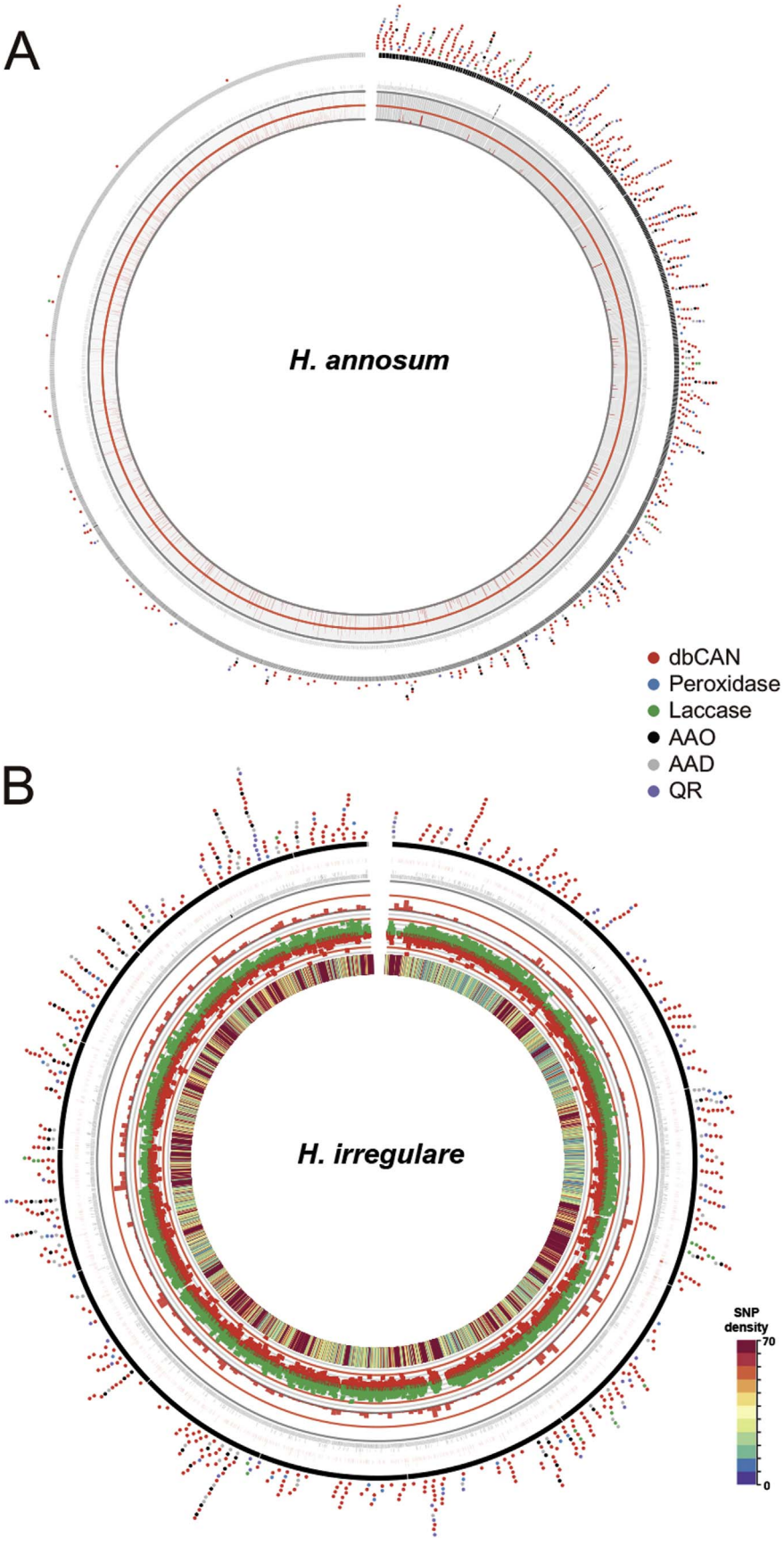

Fig. 1. Genome comparisons of the two *Heterobasidion* spp. (a) Genomic characteristics of *H. annosum* are displayed, including density of repetitive elements, predicted genes, orthologous genes found in *H. irregulare*, scaffold representation, and genes involved in degradation of plant materials (aryl-alcohol dehydrogenase in gray, aryl-alcohol oxidase in black, CAZyme predicted by dbCAN in red, laccase in green, peroxidase in blue and quinone reductase in purple). (b) Genomic characteristics of *H. irregulare* genome and comparisons to *H. annosum* genome are presented, including SNP density (purple as the lowest and red as the highest), frequency of insertions/deletions (shown in green and red, respectively), density of repetitive elements, predicted genes, orthologous genes found in *H. annosum*, scaffold representation, and genes involved in degradation of plant materials with the same color codes used in Fig. 1a. The tracks were described from the innermost circle.

Table 1
Summary of *H. annosum* 03012 genome in comparison with that of *H. irregulare* TC 32-1.

	H. annosum 03012	*H. irregulare* TC 32-1
Assembled genome size	31.01 Mb	33.65 Mb
Number of contigs/scaffolds	2415	15
Number of ambiguous bases (Ns)	1574 bp	323 bp
GC contents	51.55%	52.23%
Number of predicted genes	11,453	13,405
Mean gene length (in amino acids)	405.8 aa	377.8 aa
Average number of exons per gene	5.64	5.39

genes (Table S3). Many sets of genes had small non-syntenic regions, which could be caused by different gene models predicted in two genome sequences. As a result, insertions/deletions of a few kb could be frequently found. For example, around 6 kb of insertion in *H. annosum* contained five predicted genes (Fig. 3). Interestingly, three of them had the best hit against the same gene in *H. irregulare*, a cytochrome P450 protein-encoding gene, implying that the inserted region might be caused by duplication. Another example showed that an insertion harboring four genes was found in *H. irregulare* (Fig. 4). The synteny analysis showed that there are many subtle differences at small and intermediate scale between the two genomes of *Heterobasidion* spp. A series of insertions/deletions of a few kb might have contributed to the adaptation for each species at different geographic locations.

2.4. Comparative gene family analysis of 37 rot fungi including Heterobasidion spp.

Many wood-decaying fungi usually have the ability to deconstruct plant materials, including cellulose, hemicellulose and lignin. In particular, white-rot fungi are well known for their ability to delignify wood materials [19]. Enzymes involved in lignin degradation, such as peroxidases and laccases, have been highlighted for their ecological significance and potential applications in industry [20,21]. Accordingly, the identification of gene families involved in plant material degradation was conducted with 37 genome sequences of rot fungi (Table S4). For the two *Heterobasidion* spp., the results of identified gene families closely resembled with each other, since their genomes showed high similarity and they cause the same disease. Interestingly, CAZymes, aryl-alcohol oxidases (AAOs), peroxidases and laccases were more frequently found in white-rot fungi than in brown-rot with statistical significance (*t*-test; $P < 0.0078$) (Fig. S3). Considering the fact that AAOs, peroxidases and laccases are known to be major players in lignin degradation [22], it might suggest that white-rot fungi would have better performance in lignin degradation.

On the other hand, secretory proteins play important roles in plant-microbe interactions, including colonization of host cells, defense against oxidative stress from plant cells, environmental sensing and acquisition of nutrients [23]. When secretory potential was assessed for the whole protein sequences for each species, white-rot fungi showed higher number of putative secretory proteins than brown-rot (*t*-test; $P = 0.0001$). This result may possibly be interpreted that fungi

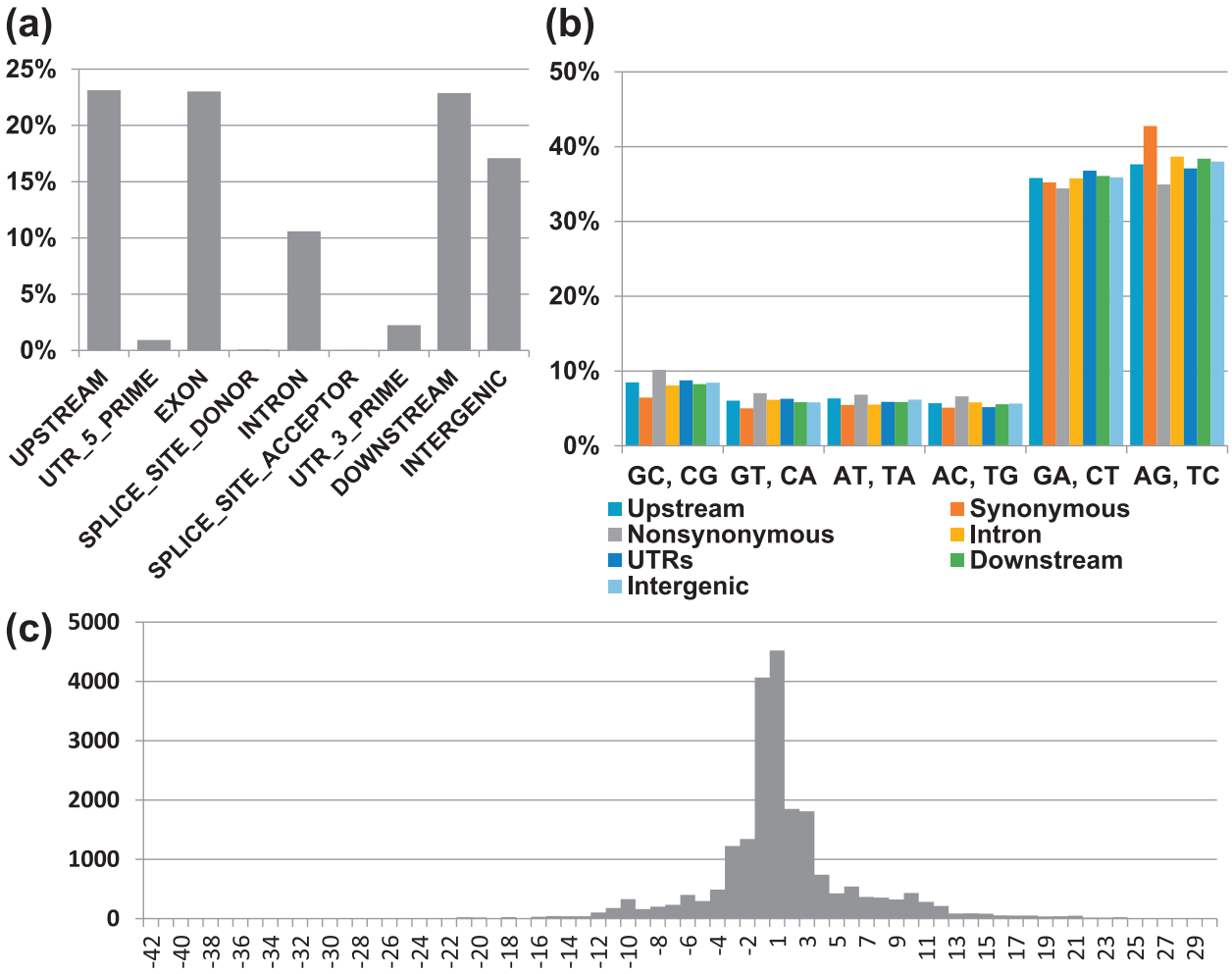

Fig. 2. Summary of single nucleotide changes. (a) Distribution of SNPs by genetic positions, (b) ratio of mutational changes for across genetic positions, and (c) length distribution of insertions and deletions.

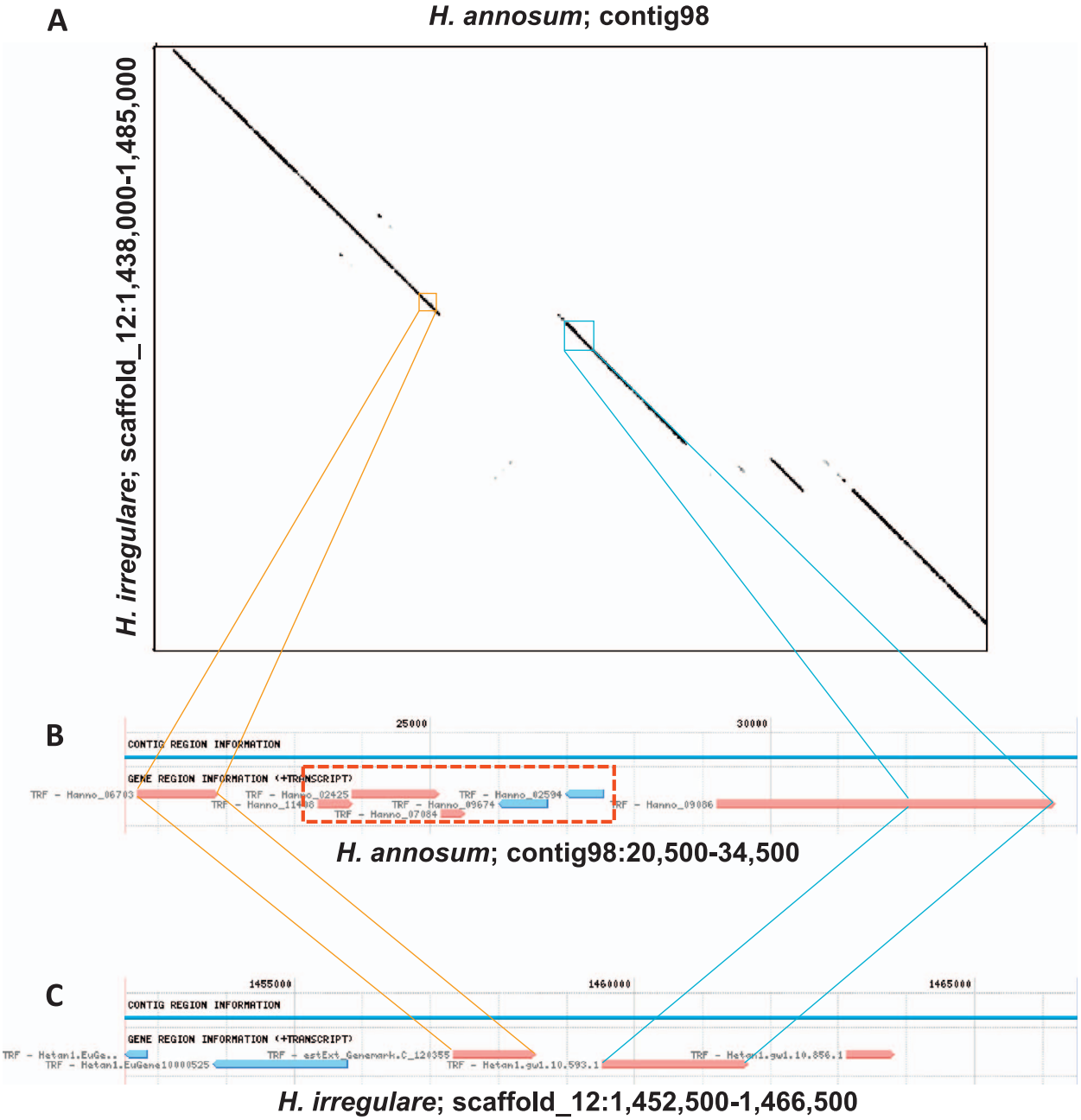

Fig. 3. An example of non-syntenic regions found in *H. annosum*. (a) A dot plot presenting the alignment between *H. annosum* contig98 (64,767 bp) on X-axis and the corresponding region of *H. irregulare* on Y-axis. (b) Closer outlook of genomic context in *H. annosum* where an insertion containing the five predicted genes might occurred. (c) Genomic context showing the corresponding region in *H. irregulare*.

encoding more proteins involved in plant material degradation need higher secretory potential to convey them to the appropriate subcellular locations. In *H. annosum*, for example, considerable number of laccases, peroxidases, AAOs, and CAZymes were predicted to be secretory (Table S4). However, it has been reported that some basidiomycetes do not show the characteristics of either typical white- or brown-rot [24,25]. It may underlie the plasticity in fungal life styles and imply that classical white and brown-rot separation has a room for improvement. Availability of more sequenced genomes of wood-decaying fungi may help to elucidate the complexity behind this classification.

3. Materials and methods

3.1. Fungal source and preparation of genomic DNA

The strain used for genome sequencing was *Heterobasidion annosum* P-type (isolate 03012 provided by Kari Korhonen, METLA, Finland). The strain was grown on malt extract medium. Genomic DNA from *H. annosum* was prepared by following a published method [26]. Mycelium was grinded in liquid nitrogen, and pulverized and incubated in equal volumes of lysis buffer (0.2 M sodium borate, 30 mM EDTA, 1% SDS, pH 9.0 and phenol) at 60 °C for 5 min. After centrifugation, the supernatant was treated with RNase, and afterwards with an equal volume phenol/chloroform (1:1). Genomic DNA was precipitated with 0.7 volume 2-propanol, washed with 70% ethanol, dried, and re-suspended in 100–200 μL of double distilled H_2O.

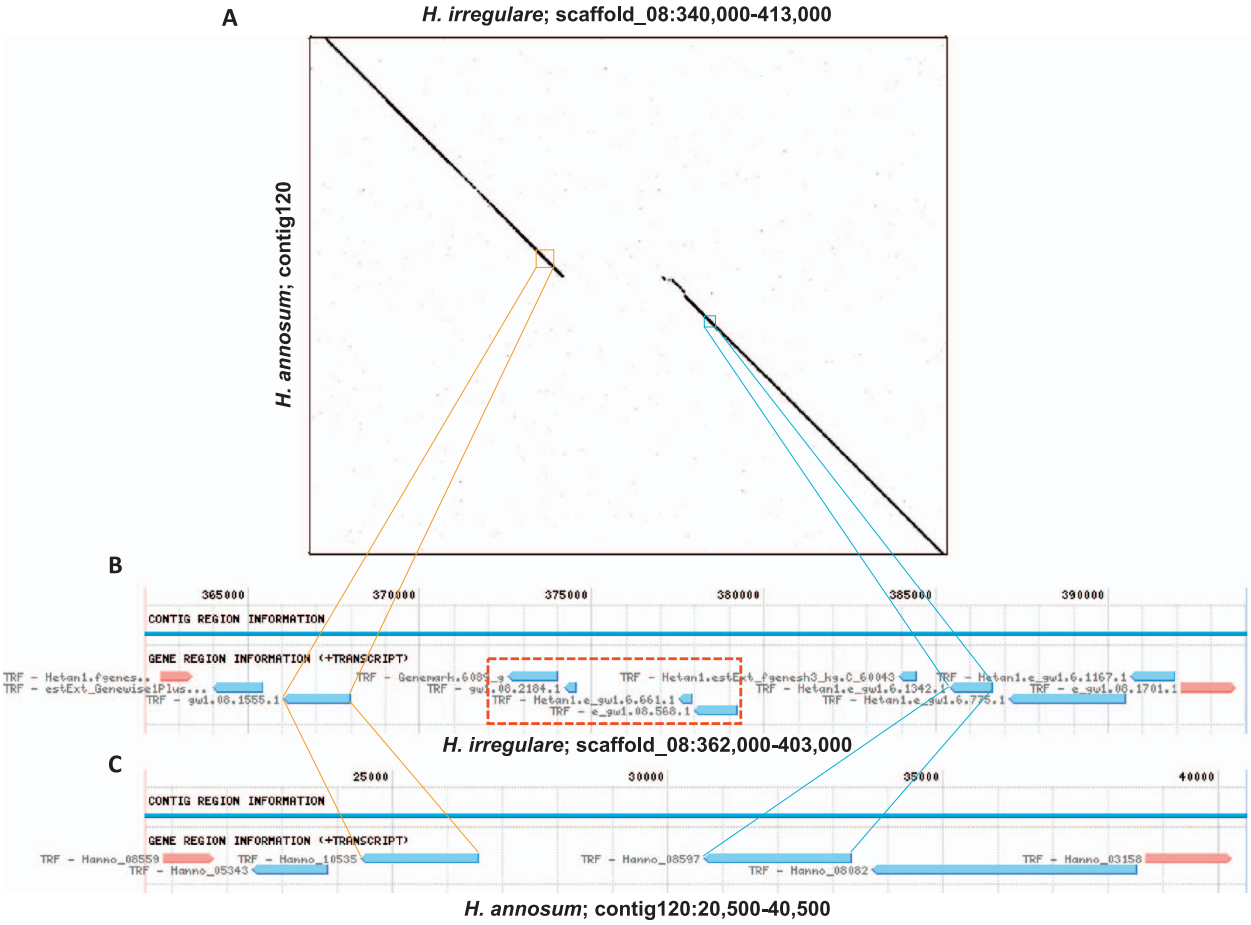

Fig. 4. An example of non-syntenic regions found in *H. irregulare*. (a) A dot plot presenting the alignment between *H. irregulare* scaffold_08:340,000–413,000 on X-axis and *H. annosum* contig120 on Y-axis. (b) Overview of genomic context in *H. irregulare* where an insertion containing the four predicted genes might occurred. (c) Genomic context showing the corresponding region in *H. annosum*.

3.2. Genome sequencing, assembly and gene prediction

Whole genome shotgun strategy was used in generation of short reads for *H. annosum* 03012 genome by using Illumina Genome Analyser IIx platform with the insert size of 500 bp. The short reads were assembled by SOAP package including SOAPdenovo (version 2.04) [27] and GapCloser (version 1.12). The generated contigs were further used in generation of scaffolds by SSPACE (version 2.0) [28]. Gene prediction was performed by using MAKER (version 2.28) [29] pipeline, integrating the results from SNAP (version 2006-07-28) [30], AUGUSTUS (version 2.7) [31] and GeneMark-ES (version 2.3e) [32].

3.3. Preparation of genome sequences of selected rot fungi

Genome sequences of 36 rot fungi were downloaded from MycoCosm at U.S. Department of Energy Joint Genome Institute website [14]. The data set includes 13 brown-rot and 23 white-rot fungi (Table S5). All the genome data was also archived in the standardized data warehouse of Comparative Fungal Genomics Platform 2.0 (CFGP 2.0; http://cfgp.snu.ac.kr/) for further analysis [33].

3.4. Whole genome alignment and detection of structural variations

In order to see genome-wide conservation, contig sequences of *H. annosum* were aligned against each of the scaffold sequences of *H. irregulare* by using BLAT [34]. By taking the hits showing ≥ 1 kb matches into consideration, dot plots for each *H. irregulare* scaffolds were created by using Gepard (version 1.30) [35]. The best hit for each *H. annosum* contig against the *H. irregulare* genome was collected for calculation of the net aligned regions between the two genomes. Single nucleotide polymorphisms (SNPs), short insertions and deletions were identified and analyzed by using BWA (version 0.5.9-r16) [36], SAMtools (version 0.1.18) [37] and snpEff (version 3.4i) [38]. The reads were trimmed from 3′ termini if base quality is < 20 in the Phred scale. The SNPs supported by < 4 reads or > 100 reads were discarded for under- and over-representation. The size of upstream and downstream for each transcript were set to be 1 kb each.

3.5. Identification of gene synteny between the Heterobasidion spp.

In order to identify syntenic regions between *H. annosum* and *H. irregulare* genomes, candidate orthologous gene pairs were defined by using BLASTP [39]. Protein sequences in *H. annosum* were used as query and searched against those in *H. irregulare*, followed by filtering the best hits. Syntenic regions were identified by using DAGchainer [40] with default settings, and the micro-syntenic differences were shown with the selected regions.

3.6. Gene family annotation

Previously published pipelines were used in identification of putative genes encoding plant cell wall-degrading enzymes (PCWDEs), secretory proteins and peroxidases [23,33,41,42]. To capture the maximum extent of PCWDEs, carbohydrate-active enzymes (CAZymes) were identified by using dbCAN [43] with the E-value cutoff of 1e − 5. And an in-house pipeline was used in laccase and laccase-like

multi-copper oxidase prediction (unpublished). Additional genes involved in lignin degradation, including aryl-alcohol dehydrogenase (AAD), aryl-alcohol oxidase (AAO) and quinone reductase (QR), were identified based on protein domain profiles predicted by InterPro scan [44] (Table S6) The domain profiles were retrieved by using the protein sequences annotated as AAD, AAO and QR in UniProt database [45].

3.7. Protein clustering analysis

Whole proteome sequences were subjected to Markov clustering using tribeMCL [46]. All-by-all BLASTP was performed to prepare input data for clustering analysis with the E-value cutoff of 1e − 5 and the inflation factor of 2.0.

3.8. Prediction of unstructured protein regions

Whole proteome sequences of the two *Heterobasidion* spp. were scanned by using IUPred, a software predicting global and short disorder in proteins sequences [47]. Subsequently amino acid residues showing disorder tendency above 0.5 were considered as disordered. In addition, residues predicted as structured were removed from the list of local and global disorder.

4. Data deposition

The draft genome sequence of *Heterobasidion annosum s.s.* isolate 03012 was deposited at DDBJ/EMBL/GenBank under accession no. AOSL00000000.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gdata.2017.10.003.

Transparency document

The Transparency document associated with this article can be found, in the online version.

Acknowledgements

This work was supported by the National Research Foundation of Korea grants funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A2A1A10051434, NRF-2015M3A9B8028679) and the Cooperative Research Program for Agricultural Science & Technology Development (PJ01115401), Rural Development Administration, Republic of Korea. KTK and JJ are grateful for a graduate fellowship through the Brain Korea 21 Plus Program. This work was also supported by the Finland Distinguished Professor Programme (FiDiPro #138116) from Academy of Finland to YHL.

References

[1] F.O. Asiegbu, A. Adomas, J. Stenlid, Conifer root and butt rot caused by *Heterobasidion annosum* (Fr.) Bref. *s.l*, Mol. Plant Pathol. 6 (2005) 395–409.
[2] C.S. Hodges, *Heterobasidion annosum*. Biology, ecology, impact and control, Plant Pathol. 48 (1999) 564–565.
[3] M. Garbelotto, P. Gonthier, Biology, epidemiology, and control of *Heterobasidion* species worldwide, Annu. Rev. Phytopathol. 51 (2013) 39–59.
[4] M. Lind, J. Stenlid, Å. Olson, *Heterobasidion annosum* s.l. genomics, Adv. Bot. Res. 70 (2014) 371–396.
[5] F. Sillo, M. Garbelotto, M. Friedman, P. Gonthier, Comparative genomics of sibling fungal pathogenic taxa identifies adaptive evolution without divergence in pathogenicity genes or genomic structure, Genome Biol. Evol. 7 (2015) 3190–3206.
[6] K. Dalman, A. Olson, J. Stenlid, Evolutionary history of the conifer root rot fungus *Heterobasidion annosum sensu lato*, Mol. Ecol. 19 (2010) 4979–4993.
[7] A. Olson, A. Aerts, F. Asiegbu, L. Belbahri, O. Bouzid, A. Broberg, B. Canback, P.M. Coutinho, D. Cullen, K. Dalman, G. Deflorio, L.T.A. van Diepen, C. Dunand, S. Duplessis, M. Durling, P. Gonthier, J. Grimwood, C.G. Fossdal, D. Hansson, B. Henrissat, A. Hietala, K. Himmelstrand, D. Hoffmeister, N. Hogberg, T.Y. James, M. Karlsson, A. Kohler, U. Kues, Y.H. Lee, Y.C. Lin, M. Lind, E. Lindquist, V. Lombard, S. Lucas, K. Lunden, E. Morin, C. Murat, J. Park, T. Raffaello, P. Rouze, A. Salamov, J. Schmutz, H. Solheim, J. Stahlberg, H. Velez, R.P. de Vries, A. Wiebenga, S. Woodward, I. Yakovlev, M. Garbelotto, F. Martin, I.V. Grigoriev, J. Stenlid, Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen, New Phytol. 194 (2012) 1001–1013.
[8] K. Dalman, K. Himmelstrand, A. Olson, M. Lind, M. Brandstrom-Durling, J. Stenlid, A genome-wide association study identifies genomic regions for virulence in the non-model organism *Heterobasidion annosum* s.s, PLoS One 8 (2013) e53525.
[9] M.A. Van der Nest, A. Olson, M. Karlsson, M. Lind, K. Dalman, M. Brandstrom-Durling, M. Elfstrand, B.D. Wingfield, J. Stenlid, Gene expression associated with intersterility in *Heterobasidion*, Fungal Genet. Biol. 73 (2014) 104–119.
[10] K. Lunden, M. Danielsson, M.B. Durling, K. Ihrmark, M. Nemesio Gorriz, J. Stenlid, F.O. Asiegbu, M. Elfstrand, Transcriptional responses associated with virulence and defence in the interaction between *Heterobasidion annosum s.s.* and Norway spruce, PLoS One 10 (2015) e0131182.
[11] T. Raffaello, H. Chen, A. Kohler, F.O. Asiegbu, Transcriptomic profiles of *Heterobasidion annosum* under abiotic stresses and during saprotrophic growth in bark, sapwood and heartwood, Environ. Microbiol. 16 (2014) 1654–1667.
[12] T. Niemelä, K. Korhonen, Taxonomy of the genus *Heterobasidion*, *Heterobasidion annosum*. Biology, Ecology, Impact and Control. Ed. xWoodward, S., Stenlid, J., Karjalainen, R. & Hüttermann, A., (1998).
[13] W.J. Otrosina, M. Garbelotto, *Heterobasidion occidentale* sp. nov. and *Heterobasidion irregulare* nom. nov.: a disposition of North American *Heterobasidion* biological species, Fungal Biol. 114 (2010) 16–25.
[14] I.V. Grigoriev, R. Nikitin, S. Haridas, A. Kuo, R. Ohm, R. Otillar, R. Riley, A. Salamov, X. Zhao, F. Korzeniewski, T. Smirnova, H. Nordberg, I. Dubchak, I. Shabalov, MycoCosm portal: gearing up for 1000 fungal genomes, Nucleic Acids Res. 42 (2014) D699–D704.
[15] Z.W. Arendsee, L. Li, E.S. Wurtele, Coming of age: orphan genes in plants, Trends Plant Sci. 19 (2014) 698–708.
[16] H. Sun, E. Terhonen, K. Koskinen, L. Paulin, R. Kasanen, F.O. Asiegbu, The impacts of treatment with biocontrol fungus (*Phlebiopsis gigantea*) on bacterial diversity in Norway spruce stumps, Biol. Control 64 (2013) 238–246.
[17] Y. Chen, L. Chai, C. Tang, Z. Yang, Y. Zheng, Y. Shi, H. Zhang, Kraft lignin biodegradation by *Novosphingobium* sp. B-7 and analysis of the degradation process, Bioresour. Technol. 123 (2012) 682–685.
[18] E. Masai, N. Kamimura, D. Kasai, A. Oguchi, A. Ankai, S. Fukui, M. Takahashi, I. Yashiro, H. Sasaki, T. Harada, S. Nakamura, Y. Katano, S. Narita-Yamada, H. Nakazawa, H. Hara, Y. Katayama, M. Fukuda, S. Yamazaki, N. Fujita, Complete genome sequence of *Sphingobium* sp. strain SYK-6, a degrader of lignin-derived biaryls and monoaryls, J. Bacteriol. 194 (2012) 534–535.
[19] D. Gao, L. Du, J. Yang, W.M. Wu, H. Liang, A critical review of the application of white rot fungus to environmental pollution control, Crit. Rev. Biotechnol. 30 (2010) 70–77.
[20] M. Puhse, R.T. Szweda, Y. Ma, C. Jeworrek, R. Winter, H. Zorn, *Marasmius scorodonius* extracellular dimeric peroxidase — exploring its temperature and pressure stability, Biochim. Biophys. Acta 1794 (2009) 1091–1098.
[21] C. Torres-Duarte, R. Vazquez-Duhalt, Applications and prospective of peroxidase biocatalysis in the environmental field, in: E. Torres, M. Ayala (Eds.), Biocatalysis Based on Heme Peroxidases: Peroxidases as Potential Industrial Biocatalysts, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 179–206.
[22] A. Hernandez-Ortega, P. Ferreira, A.T. Martinez, Fungal aryl-alcohol oxidase: a peroxide-producing flavoenzyme involved in lignin degradation, Appl. Microbiol. Biotechnol. 93 (2012) 1395–1410.
[23] J. Choi, J. Park, D. Kim, K. Jung, S. Kang, Y.H. Lee, Fungal secretome database: integrated platform for annotation of fungal secretomes, BMC Genomics 11 (2010) 105.
[24] D. Floudas, B.W. Held, R. Riley, L.G. Nagy, G. Koehler, A.S. Ransdell, H. Younus, J. Chow, J. Chiniquy, A. Lipzen, A. Tritt, H. Sun, S. Haridas, K. LaButti, R.A. Ohm, U. Kues, R.A. Blanchette, I.V. Grigoriev, R.E. Minto, D.S. Hibbett, Evolution of novel wood decay mechanisms in Agaricales revealed by the genome sequences of *Fistulina hepatica* and *Cylindrobasidium torrendii*, Fungal Genet. Biol. 76 (2015) 78–92.
[25] R. Riley, A.A. Salamov, D.W. Brown, L.G. Nagy, D. Floudas, B.W. Held, A. Levasseur, V. Lombard, E. Morin, R. Otillar, E.A. Lindquist, H. Sun, K.M. LaButti, J. Schmutz, D. Jabbour, H. Luo, S.E. Baker, A.G. Pisabarro, J.D. Walton, R.A. Blanchette, B. Henrissat, F. Martin, D. Cullen, D.S. Hibbett, I.V. Grigoriev, Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 9923–9928.
[26] H.C. Kuo, T.Y. Wang, P.P. Chen, R.S. Chen, T.Y. Chen, Genome sequence of *Trichoderma virens* FT-333 from tropical marine climate, FEMS Microbiol. Lett. 362 (2015) fnv036.
[27] R. Luo, B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, Y. Liu, J. Tang, G. Wu, H. Zhang, Y. Shi, Y. Liu, C. Yu, B. Wang, Y. Lu, C. Han, D.W. Cheung, S.M. Yiu, S. Peng, Z. Xiaoqian, G. Liu, X. Liao, Y. Li, H. Yang, J. Wang, T.W. Lam, J. Wang, SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler, Gigascience 1 (2012) 18.
[28] M. Boetzer, C.V. Henkel, H.J. Jansen, D. Butler, W. Pirovano, Scaffolding pre-assembled contigs using SSPACE, Bioinformatics 27 (2011) 578–579.
[29] B.L. Cantarel, I. Korf, S.M.C. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A.S. Alvarado, M. Yandell, MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes, Genome Res. 18 (2008) 188–196.
[30] I. Korf, Gene finding in novel genomes, BMC Bioinf. 5 (2004) 59.
[31] M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel, Bioinformatics 19 (2003) ii215–ii225.
[32] V. Ter-Hovhannisyan, A. Lomsadze, Y.O. Chernoff, M. Borodovsky, Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training,

Genome Res. 18 (2008) 1979–1990.
[33] J. Choi, K. Cheong, K. Jung, J. Jeon, G.W. Lee, S. Kang, S. Kim, Y.W. Lee, Y.H. Lee, CFGP 2.0: a versatile web-based platform for supporting comparative and evolutionary genomics of fungi and oomycetes, Nucleic Acids Res. 41 (2013) D714–D719.
[34] W.J. Kent, BLAT—the BLAST-like alignment tool, Genome Res. 12 (2002) 656–664.
[35] J. Krumsiek, R. Arnold, T. Rattei, Gepard: a rapid and sensitive tool for creating dotplots on genome scale, Bioinformatics 23 (2007) 1026–1028.
[36] H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform, Bioinformatics 26 (2010) 589–595.
[37] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, G.P.D. Proc, The sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079.
[38] P. Cingolani, A. Platts, L.L. Wang, M. Coon, T. Nguyen, L. Wang, S.J. Land, X.Y. Lu, D.M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain *w*(1118); *iso*-2; *iso*-3, Fly 6 (2012) 80–92.
[39] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST +: architecture and applications, BMC Bioinf. 10 (2009) 421.
[40] B.J. Haas, A.L. Delcher, J.R. Wortman, S.L. Salzberg, DAGchainer: a tool for mining segmental genome duplications and synteny, Bioinformatics 20 (2004) 3643–3646.
[41] J. Choi, N. Detry, K.T. Kim, F.O. Asiegbu, J.P.T. Valkonen, Y.H. Lee, fPoxDB: fungal peroxidase database for comparative genomics, BMC Microbiol. 14 (2014) 117.
[42] K.T. Kim, J. Jeon, J. Choi, K. Cheong, H. Song, G. Choi, S. Kang, Y.H. Lee, Kingdom-wide analysis of fungal small secreted proteins (SSPs) reveals their potential role in host association, Front. Plant Sci. 7 (2016) 186.
[43] Y.B. Yin, X.Z. Mao, J.C. Yang, X. Chen, F.L. Mao, Y. Xu, dbCAN: a web resource for automated carbohydrate-active enzyme annotation, Nucleic Acids Res. 40 (2012) W445–W451.
[44] S. Hunter, P. Jones, A. Mitchell, R. Apweiler, T.K. Attwood, A. Bateman, T. Bernard, D. Binns, P. Bork, S. Burge, E. de Castro, P. Coggill, M. Corbett, U. Das, L. Daugherty, L. Duquenne, R.D. Finn, M. Fraser, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, C. McMenamin, H. Mi, P. Mutowo-Muellenet, N. Mulder, D. Natale, C. Orengo, S. Pesseat, M. Punta, A.F. Quinn, C. Rivoire, A. Sangrador-Vegas, J.D. Selengut, C.J. Sigrist, M. Scheremetjew, J. Tate, M. Thimmajanarthanan, P.D. Thomas, C.H. Wu, C. Yeats, S.Y. Yong, InterPro in 2011: new developments in the family and domain prediction database, Nucleic Acids Res. 40 (2012) D306–D312.
[45] UniProt Consortium, UniProt: a hub for protein information, Nucleic Acids Res. 43 (2015) D204–D212.
[46] A.J. Enright, S. Van Dongen, C.A. Ouzounis, An efficient algorithm for large-scale detection of protein families, Nucleic Acids Res. 30 (2002) 1575–1584.
[47] Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon, IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, Bioinformatics 21 (2005) 3433–3434.