Comparative and population genomics landscape of *Phellinus noxius*: a hypervariable fungus causing root rot in trees

Chia-Lin Chung¶1,2, Tracy J. Lee3,4,5, Mitsuteru Akiba6, Hsin-Han Lee1, Tzu-Hao Kuo3, Dang Liu3,7, Huei-Mien Ke3, Toshiro Yokoi6, Marylette B Roa3,8, Meiyeh J Lu3, Ya-Yun Chang1, Pao-Jen Ann9, Jyh-Nong Tsai9, Chien-Yu Chen10, Shean-Shong Tzean1, Yuko Ota6,11, Tsutomu Hattori6, Norio Sahashi6, Ruey-Fen Liou1,2, Taisei Kikuchi12 and Isheng J Tsai¶3,4,5,7

1Department of Plant Pathology and Microbiology, National Taiwan University, Taiwan
2Master Program for Plant Medicine, National Taiwan University, Taiwan
3Biodiversity Research Center, Academia Sinica, Taipei, Taiwan
4Biodiversity Program, Taiwan International Graduate Program, Academia Sinica and National Taiwan Normal University
5Department of Life Science, National Taiwan Normal University
6Department of Forest Microbiology, Forestry and Forest Products Research Institute, Tsukuba, Japan
7Genome and Systems Biology Degree Program, National Taiwan University and Academia Sinica, Taipei, Taiwan
8Philippine Genome Center, University of the Philippines, Diliman, Quezon City, Philippines 1101
9Plant Pathology Division, Taiwan Agricultural Research Institute
10Department of Bio-industrial Mechatronics Engineering, National Taiwan University, Taiwan
11College of Bioresource Sciences, Nihon University, Fujisawa, Japan
12Division of Parasitology, Faculty of Medicine, University of Miyazaki, Miyazaki, Japan

¶ These authors contributed equally to this work

Correspondence: Chia-Lin Chung, clchung@ntu.edu.tw, Taisei Kikuchi (taisei_kikuchi@med.miyazaki-u.ac.jp) and Isheng Jason Tsai (ijtsai@gate.sinica.edu.tw)
Abstract

The order Hymenochaetales of white rot fungi contain some of the most aggressive wood decayers causing tree deaths around the world. Despite their ecological importance and the impact of diseases they cause, little is known about the evolution and transmission patterns of these pathogens. Here, we sequenced and undertook comparative genomics analyses of Hymenochaetales genomes using brown root rot fungus *Phellinus noxius*, wood-decomposing fungus *Phellinus lamaensis*, laminated root rot fungus *Phellinus sulphurascens*, and trunk pathogen *Porodaedalea pini*. Many gene families of lignin-degrading enzymes were identified from these fungi, reflecting their ability as white rot fungi. Comparing against distant fungi highlighted the expansion of 1,3-beta-glucan synthases in *P. noxius*, which may account for its fast-growing attribute. We identified 13 linkage groups conserved within Agaricomycetes, suggesting the evolution of stable karyotypes. We determined that *P. noxius* has a bipolar heterothallic mating system, with unusual highly expanded ~60 kb A locus as a result of accumulating gene transposition. We investigated the population genomics of 60 *P. noxius* isolates across multiple islands of the Asia Pacific region. Whole-genome sequencing showed this multinucleate species contains abundant poly-allelic single-nucleotide-polymorphisms (SNPs) with atypical allele frequencies. Different patterns of intra-isolate polymorphism reflect mono-/heterokaryotic states which are both prevalent in nature. We have shown two genetically separated lineages with one spanning across many islands despite the geographical barriers. Both populations possess extraordinary genetic diversity and
show contrasting evolutionary scenarios. These results provide a framework to further investigate the genetic basis underlying the fitness and virulence of white rot fungi.

Introduction

Under most circumstances, fungi coexist with trees or act as saprotrophs responsible for carbon and nitrogen cycling in forest systems. However, some fungi are also one of the dominant groups of pathogens causing diseases in trees. There has been an emergence of tree disease outbreaks in different parts of the world such as ash dieback (Gross, Holdenrieder, Pautasso, Queloz, & Sieber, 2014), Dutch elm disease (Potter, Harwood, Knight, & Tomlinson, 2011), laminated root rot caused by *Phellinus sulphurascens* (Williams et al., 2014), and brown root rot caused by *Phellinus noxius* (Akiba et al., 2015; P. Ann, Chang, & Ko, 2002; Chung et al., 2015). Factors contributing to this phenomenon include climate change (Goberville et al., 2016) and human activity (Fisher et al., 2012). If interventions are not implemented early and effective, pathogen infections can kill millions of trees and the spread can become very difficult to stop (Cunniffe, Cobb, Meentemeyer, Rizzo, & Gilligan, 2016).
Hymenochaetales is dominated by wood decay fungi and belongs to Agaricomycetes of Basidiomycota. Most species within this order are saprotrophic but some also exhibit pathogenic lifestyles that have been recorded in major forest incidents as early as 1971 in different parts of the world (Hepting, 1971; Norio Sahashi, Akiba, Ishihara, Ota, & Kanzaki, 2012). In particular, *P. noxius* has a very wide host range, spanning more than 200 broadleaved and coniferous tree species (at least 59 families) including many agricultural, ornamental, and forest trees such as longan, litchi, camphor, banyan, and pine (P. Ann et al., 2002; Norio Sahashi et al., 2014). Inoculation assays showed that only seven out of 101 tested tree cultivars (92 species) exhibited high resistance (P. J. Ann, Lee, & Huang 1999). Despite the recognized importance of *P. noxius* as an emerging pathogen, its genome, evolution and global population genetics is poorly understood. Previous reports based on simple sequence repeat (SSR) markers suggest the existence of highly diversified *P. noxius* populations (Akiba et al., 2015; Chung et al., 2015), but the isolates exhibited little to no host specificity (P. J. Ann et al., 1999; Nandris, Nicole, & Geiger, 1987; N. Sahashi, Akiba, Ishihara, Miyazaki, & Kanzaki, 2010). Currently, no gold standard genomes (Chain et al., 2009) are available for this group of wood decay fungi, which is a necessary step toward a better knowledge of their diversity, ecology and evolution.

The life cycle of *P. noxius* (P. Ann et al., 2002) is thought to be similar to other important root-rotting basidiomycetes, such as *Armillaria mellea* and *Heterobasidion annosum*. The new infection can start from previously infected plants/stumps or colonized wood debris, from which the mycelium of *P. noxius* grows to infect the
lateral and tap roots of the host tree (P. Ann et al., 2002). An invasion to the cortex and lignified xylem is usually observed (T.-T. Chang, 1992), sometimes accompanied by gradual expansion of the mycelial mat to the basal stem (Fig. 1A). Diseased trees with decayed roots may then show symptoms of foliar chlorosis, thinning leaves, defoliation, and eventually decline within a few months to several years. The damaged and fragile roots (Fig. 1B) also make the trees easily toppled over by strong winds and heavy rains. Basidiocarps are occasionally formed on trunks of infected trees (Fig. 1C-D). The sexual reproduction system of *P. noxius* has remained unclarified, partly due to the lack of clamp connections for diagnosing compatibility (P. Ann et al., 2002).

Brown root rot disease caused by *P. noxius* is widespread in tropical and subtropical areas in Southeast and East Asia, Oceania, Africa, Central America and the Caribbean. The geographical distribution appears to be related to the growth temperature range of *P. noxius*: 10-12°C to 36 °C, with optimum growth at 30 °C. In the past 20 years, brown root rot disease has become a serious threat to a variety of perennial fruit trees, ornamental and landscape trees, and shade trees in Taiwan (Chung et al., 2015) and in Ryukyu and Ogasawara Islands of Japan (Akiba et al., 2015). In Australia, it occurs in the natural and commercial plantation forests and orchards along the east coast, and has killed many trees in the Greater Metropolitan area of Brisbane Queensland (Schwarze, Jauss, Spencer, Hallam, & Schubert, 2012). In West Africa and China, brown root rot was reported as the most devastating root disease attacking the rubber plantations (Nandris et al., 1987). Trees in urban areas and plantation forests in Hong
Kong and Macao have also been seriously affected (Huang, Sun, Bi, Zhong, & Hu, 2016; Wu et al., 2011). Field observations support the root-to-root spread as a major transmission mode of the epidemic. Recent population genetics studies based on SSR markers revealed highly diverse genotypes within populations and nearly identical genotypes from neighboring infected trees, suggesting that *P. noxius* spreads over short distances via root-to-root contact of the hosts, and the genetically variable basidiospores are likely involved in long-distance dispersal and the establishment of unique clones in new disease foci (Akiba et al., 2015; Chung et al., 2015).

In the present study, we aimed to further understand the evolution, reproductive system, and epidemiology in *P. noxius*, on the whole genomic basis. To achieve this, we first sequenced, assembled and annotated the genome sequences of four species from Hymenochaetales: *Phellinus noxius*, *Phellinus lamaensis* (wood decomposing fungus that causes a white pocket rot only on dead heartwood trees), *Phellinus sulphurascens* (syn. *Coniferiporia sulphurascens* (Zhou, Vlasák, & Dai, 2016); pathogen responsible for laminated root rot in Douglas-fir/true fir), and *Porodaelea pini* (syn. *Phellinus pini*; trunk pathogen of conifers). Focusing on *P. noxius*, its ~31Mb genome was sequenced to a high level of completion containing telomere-to-telomere chromosome sequences. By comparing against representative species of Basidiomycota, we investigated the genomic conservations and specialisations of this group and how these features possibly relate to the lifestyle of a wood-decayer. Second, we collected *P. noxius* isolates from diseased trees across Taiwan and Japanese offshore islands in 2007-2014. We sequenced the
whole-genomes of these 60 isolates and realigned against the *P. noxius* reference genome to identify single nucleotide polymorphisms (SNPs). Based on the genetic variation, the phylogenetic relationship of *P. noxius* populations within this Asia Pacific region was determined. We also quantified the extent of intra-isolate polymorphism to infer the frequencies of morphologically indistinguishable mono- and heterokaryons at infection sites. This allowed the confirmation of heterokaryosis is not necessary for pathogenicity in *P. noxius*.

Methods

Strain preparation and sequencing

Genome sequencing of three *Phellinus* species and *P. pini* (Text S1) was performed using both Pacific Biosciences (*P. noxius*, *P. lamaensis*, *P. sulphurascens* and *P. pini*) and Illumina (*P. noxius*) platforms. DNA was isolated using the CTAB method (Chung et al., 2015). At least 15 μg DNA was used for a 20 kb library prep according to the manufacturer’s instructions. Sequencing was performed on a Pacific Biosciences RS II system using 8 SMRT cell per run, P6C4 chemistry and 360 min movie time. A total of 5-7 SMRT cells were run per species yielding a raw depth of coverage of 173-266X. For samples underwent *de novo* assembly, genomic DNA were sheared in M220 Focused-ultrasonicator™ (Covaris) in microTUBE-50 (Covaris) under 250bp program (duty factor 20%, treatment for 120 sec) with gel size selection after adaptor ligation. For population resequencing samples, genomic DNA were
sheared in microTUBE-130 (Covaris) using 500bp shearing program (duty factor 10%, treatment for 60 sec). Genomic libraries were prepared using TruSeq DNA LT Sample Prep Kit (Illumina). Input of 1 μg sheared DNA was used for end-repair, A-tailing, adaptor ligation, and gel size selection. Size range at 600-700bps range was selected from gel and amplified by 5 cycles of PCR. In addition, Illumina mate-pair libraries were generated using 8 μg of genomic DNA with gel size selection of tagmented DNA at 2-4 kb, 4-7 kb, 7-10 kb, and 10-20 kb, and amplified by 10~15 cycles of PCR. These libraries were normalized by KAPA Library Quantification Kit (KAPA Biosystems), and pooled equally for PE2*300 sequencing on MiSeq V2 sequencer. The descriptions of the raw genomic data are available on Table S1.

To aid annotation for each of the species in this study, 7- to 21-day old mycelia from PDA cultures were used for RNA-seq. RNA-seq libraries were constructed using the Illumina TruSeq Stranded mRNA HT Sample Prep Kit with the dual index barcoded adaptors. Input of 3 μg of total RNA was used for each sample for two rounds of oligo-dT bead enrichment, and the ligated cDNA were amplified by 10 cycles of PCR. The Stranded mRNA libraries were quantified by Qubit and molar concentrations normalized against the KAPA Library Quantification Kit (KAPA Biosystems) for Illumina platform. The transcriptome libraries were pooled at equal molar concentrations, and PE2*151nt multiplexed sequencing was conducted on HiSeq 2500 sequencer. The descriptions of the raw RNA-seq data are available on Table S2.

Nuclear quantification
Nuclei in the growing hyphal tips were stained following the procedure described by Chung et al. (2015). The mycelium on the slide was mounted in 20 μl of a DAPI (4',6-diamidino-2-phenylindole) solution (10 μg/ml in ddH2O) for an hour, destained in ddH2O for 30 min, then observed under a OLYMPUS BX41 microscope (Shinjuku-ku, Tokyo, Japan) equipped with filter cube U-MWU2 (BP 330–385 nm, LP 420 nm). Images were captured by using a Canon (Ohta-ku, Tokyo, Japan) digital camera EOS 700D. One hundred hyphal cells per strain were counted.

**Genome assembly**

Genome assembly of different species was carried out with Falcon (ver 0.5.0; Chin et al., 2016) and were improved using Quiver (Chin et al., 2013) and finisherSC (Lam, LaButti, Khalak, & Tse, 2015). For assembly of individual strains of *P. noxius*, Illumina paired end reads were trimmed with Trimmomatic (ver 0.32; options LEADING:30 TRAILING:30 SLIDINGWINDOW:4:30 MINLEN:50; Bolger, Lohse, & Usadel, 2014) and subsequently assembled using SPAdes (ver 3.7.1; Bankevich et al., 2012). Multiple mate-pair reads were available for three strains of *P. noxius* (KPN91, A42 and 718-S1) and they were assembled using ALLPATH-LG (ver 49688; Butler et al., 2008) assembler and improved using Pilon (Walker et al., 2014). The *P. noxius* assembly was further merged with metassembler (ver 1.5; Wences & Schatz, 2015), misassemblies were identified using REAPR (ver 1.0.18; Hunt et al., 2013) and manually corrected.

**Gene predictions and functional annotation**
For *P. noxius*, the gene predictor Augustus (ver 3.2.1; Stanke, Tzvetkova, & Morgenstern, 2006) was trained on a gene training set of complete core genes from CEGMA (ver 2.5; Parra, Bradnam, & Korf, 2007) and subsequently used for manual curation of ~1000 genes. Annotation was then run by providing introns as evidence from RNA-seq data. For *P. lamaensis*, *P. sulphurascens* and *P. pini*, genes were predicted using Braker1 (ver 1.9; Hoff, Lange, Lomsadze, Borodovsky, & Stanke, 2016) pipeline that automatically use RNA-seq mappings as evidence hints and retraining of GeneMark-ES (Borodovsky & Lomsadze, 2011) and Augustus. Gene product description was assigned using blast2go (ver 4.0.7; Conesa et al., 2005) and GO term assignment were provided by ARGOT2.5 (Lavezzo, Falda, Fontana, Bianco, & Toppo, 2016). The web server dbCAN (HMMs 5.0, last accessed September 5 2016; Yin et al., 2012) was used to predict CAZymes from the protein sequences of all species, while AntiSMASH (ver 3.0; Weber et al., 2015) was used to predict secondary metabolite gene clusters. For dbCAN results, only hits with <= 1 x 10e-5 e-value and >= 30% HMM coverage were considered, while overlapping domains were resolved by choosing hits with the smallest *P*-value. Proteome completeness were assessed with BUSCO (ver 2.0; Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) using the Basidiomycota dataset.

**Comparative genomics analysis**

Gene families were determined by OrthoFinder (ver 1.0.6; Emms & Kelly, 2015). Then, MAFFT (ver 7.271; Katoh & Standley, 2014) was used to align sequences in each of 1,127 single-copy orthogroups. Alignments results with less than 10%
alignment gaps were concatenated, and the outcome was taken to compute a maximum likelihood phylogeny using RAxML (ver 7.7.8; Stamatakis, 2006) with 100 bootstrap replicates. Gene family gain and loss in different positions along the global phylogeny leading to \textit{P. noxius} were inferred using dollop (Felsenstein, 2005). Pfam and GO enrichments were carried out on these gene families using TopGO (ver 2.10.0; Alexa & Rahnenfuhrer, 2016). Sequences to be included in the phylogenetic tree for NACHT domain containing genes were selected on the basis of the presence of pfam domain PF05729.

\textit{MAT} locus

Homologs of the conserved genes in mating locus \textit{A} and \textit{B} were annotated in \textit{P. noxius}, \textit{P. pini}, \textit{P. sulphurascens} and \textit{P. lamaensis} (e-value $< 10^{-5}$) then subjected to InterProScan 5 (ver 5.20-59.0; Jones et al., 2014) and Pfam (ver 30.0; Finn et al., 2016) for protein signature prediction. Syntenic alignment of \textit{A} locus among \textit{Phellinus} spp. and other species (sequences/annotations retrieved from Joint Genome Institute MycoCosm (Grigoriev et al., 2014)) was plotted using genoPlotR package (Guy, Roat Kultima, & Andersson, 2010) in R. The sequences of \textit{A} locus in KPN91, 718-S1 and A42 were further compared by MUMmer (ver 3.20; Stefan Kurtz et al., 2004) and PipMaker (http://pipmaker.bx.psu.edu/cgi-bin/pipmaker?basic; Schwartz et al., 2000). For identification of candidate pheromone genes, all the potential open reading frames were filtered for small peptides with C-terminal CaaX motif, then searched for pheromone homologies against Pfam-A and scanned for farnesylation signal by PrePs (http://mendel.imp.ac.at/PrePS/; Maurer-Stroh & Eisenhaber, 2005).
The sequences of *HD* and *STE3* genes were analyzed for 10 single-basidiospore isolates originating from a basidiocarp by the dideoxy termination method (primers listed in Table S3). Long-range PCR followed by primer walking was performed to sequence the highly diverse regions containing *HD1-HD2* in A locus. The outer primers were designed manually based on the alignment of all the isolates; the inner primers were developed step-by-step according to previous sequencing results. The PCR reaction was performed in 30-μl reaction mixture containing 50 to 100 ng genomic DNA, 0.2 mM dNTP, 1X Ex Taq buffer [proprietary, containing 20mM Mg²⁺] (Takara Bio Inc., Japan), 0.67 μM forward and reverse primers, and TaKaRa Ex Taq® DNA polymerase (Takara Bio Inc., Japan). The thermal cycling parameters were 1 cycle of 95°C for 3 min, 30 cycles of 95°C for 30 s, 54°C for 30 s, and 72°C for 60–270 s (according to different product sizes, ~2 kb/min), and a final extension step of 72°C for 10 min. The PCR products were sequenced by Genomics Biotechnology Inc. (Taipei, Taiwan). DNA trace data were visualized using 4Peaks (http://nucleobases.com/4peaks) and assembled using DNA Sequence Assembler in DNA Baser (http://www.dnabaser.com).

**Population genomics**

Paired end reads of 60 *P. noxius* strains (description listed in Table S4) were aligned to the KPN91 reference using Smalt (ver 5.7; www.sanger.ac.uk/resources/software/smalt/). Removal of PCR duplicates and bam file sorting were implemented with Picard (http://broadinstitute.github.io/picard/) and
samtools (ver 1.3-20-gd49c73b; Li et al., 2009). The first round of variant identification was implemented in Varscan (ver 2.4.0; Koboldt et al., 2012) and degree of heterokaryon was inferred in each strain based on allele frequency and total number of heterozygous SNPs called. The final list of SNPs was ascertained by combining evidences from samtools (ver 1.3-20-gd49c73b; Li et al., 2009) and FreeBayes (ver 1.0.2-16-gd466dde; Garrison & Marth, 2012). A maximum likelihood phylogeny of the SNPs segregating in these 60 strains was produced using FastTree (Price, Dehal, & Arkin, 2009). Plink (ver 1.9; C. C. Chang et al., 2015) was used to subset biallelic SNPs without linkage (filtering options: --maf 0.05 --indep-pairwise 50 5 0.2), which were clustered using fastSTRUCTURE (ver 1.0; Raj, Stephens, & Pritchard, 2014) to determine the optimal number of populations. Strains were phased using samtools (ver 1.3-20-gd49c73b; Li et al., 2009) and one haplotype was chosen at random. Consensus sequences were generated from each strain using bcftools (ver 1.3.1; Danecek et al., 2011) and population genetics analyses were conducted using Variscan (ver 2.0.3; Vilella, Blanco-Garcia, Hutter, & Rozas, 2005).

Results

Genomes and annotations of four Hymenochaetales members

We produced a 31.6 Mb Phellinus noxius genome reference assembly from a Japanese KPN91 isolate combining both Pacific biosciences and Illumina sequencing
platforms (Methods). The nuclear genome of *P. noxius* is assembled into 12 scaffolds with six assembled into chromosomes from telomere to telomere, while the mitochondrial genome is assembled into a single sequence of 163.4 kb. For a comprehensive understanding of genome evolution amongst members of the hymenochaetoid clade, we also sequenced and assembled three additional species: *P. sulphurascens*, *P. lamaensis* and *P. pini*, as well as two more isolates (A42 and 718-S1) of *P. noxius*. The three assemblies of *P. noxius* have N50s of 2.4-2.7 Mb, whilst the other three genome assemblies comprise 30.7-53.3 Mb with N50’s of 570 kb-2.7 Mb.

A total of 9,833-18,103 genes were predicted in each species, which are 82-94% complete (Table S5). To compare these predicted proteins to those of other basidiomycetes to explore chromosome and gene family evolutionary dynamics, we selected the proteomes of fifteen species that are highly finished from the 1000 Fungal Genomes Project (Table S5). The Hymenochaetales species have median intergenic and intron lengths of 507-634 bp and 59-60 bp, respectively, which are comparable with those observed in genomes of other basidiomycetes. The maximum likelihood phylogeny based on 1,127 single-copy orthologs placed these species with *Fomitiporia mediterranea* and *Schizopora paradoxa*, two other species of the Hymenochaetaceae group with strong bootstrap support (Fig. 1E). This phylogenetic relationship is consistent with previous findings (Larsson et al., 2006), and species with similar genome sizes and pathogenic habits are grouped together. For instance, *P. noxius* and *P. lamaensis* with compact genome sizes of ~31 Mb are grouped together,
while the trunk rot pathogens *P. pini* and *F. mediterranea* show an expansion of their genome sizes to 53-63 Mb and are also grouped together.

To explore the genomic architecture underlying the biological attributes of Hymenochaetales, we sought to identify genes and protein domains specific to Hymenochaetales by determining when a new gene family emerged and if the family has expanded or contracted. In total, 7,125-11,659 proteins in the Hymenochaetales order are clustered together in 5,184 families. Acquisition of gene families was mainly found at the tips of the phylogeny (531-8,055 families) suggesting each species has a repertoire of specific genes. The seven Hymenochaetaceae species have a total of 62 enriched protein domains compared to other basidiomycetes (Fig. S1). Gain of domains are highlighted such as fungal specific transcription factors (Fungal_trans; 53.8 vs 41.6 copies) and peroxidase associated protein domains (Peroxidase_ext; 16.7 vs 3.9 copies). These are expected as they are required for efficient degradation of lignin, a tough biopolymer present in woody plants (Dashtban, Schraft, Syed, & Qin, 2010). Consistent with this, CAZymes spanning eight families of lignin-degrading enzymes (AA1-AA8) which include laccases, peroxidases, oxidases, and reductases, and a lytic polysaccharide mono-oxygenase family (LPMO; AA9) were found in all Hymenochaetales species (Table S6).

**Genome conservation and specialisation of Phellinus noxius**

Although *P. noxius* has a compact genome, 74% (7,313) of its 9,833 predicted genes have orthologs from at least one of the other basidiomycetes, suggesting most of the
basidiomycetous core genes are conserved. As a chromosome level assembly is now available in Hymenochaetales, we attempted to characterize chromosome architecture and evolution amongst the Agaricomycotina sub-division. The number of known karyotypes in Basidiomycota ranges from 11 to 14 chromosome pairs (Table S5), which suggests a possible ancestor with similar chromosome numbers. We constructed a linkage network of seven selected species with chromosome sequences based on single-copy orthologs between species pairs. Indeed, we identified 13 distinct linkage groups (LG) providing strong evidence that chromosome macro-synteny is largely conserved since the common ancestor of Agaricomycetes (Fig. 2A). Strikingly, such a relationship even extended to Dacrymycetes where multiple scaffolds can be predominantly assigned to different linkage groups, but it is no longer apparent when compared to Tremellomycetes. Certain scaffolds are found to connect different linkage groups, implying inter-chromosomal rearrangements. For example, *P. noxius* scaffold1 is strongly clustered in LG11 but also shows linkage to LG10, implying a translocation from scaffold5 (Fig. 2). Within each linkage group, gene collinearity is no longer apparent, suggesting high levels of intra-chromosomal rearrangements which have been observed in different fungal groups (Hane et al., 2011) (Fig. 2B).

Consistent with the fact that *P. noxius* is an extremely fast grower (P. Ann et al., 2002), we identified a strikingly 7-fold increase in 1,3-beta-glucan synthase (14 compared to an average of 1.8 copies), which is responsible for the formation of beta-glucan components in fungal cell wall (Douglas, 2001). *P. noxius* contains a
comprehensive repertoire of carbohydrate-active enzymes; a total of 416 proteins of its proteome were identified as CAZymes (Table S6). This number makes up 4.23% of *P. noxius*’s proteome which is more than those in the other Hymenochaetales, suggesting these genes are necessary and have been retained despite a genome size compaction. Taken together, being able to grow fast and the diversity of the CAZymes encoded in the *P. noxius* genome suggest its capability to infect a wide range of hosts. Additionally, counts of WD40 protein domains are highest in *P. noxius* and *P. lamaensis* despite their small genome size (Table S7). This domain is important in protein-protein interactions of cellular networks and is usually associated with additional domains (Leipe, Koonin, & Aravind, 2004). Interrogating this expansion revealed the association of WD40 domains with the AAA and NACHT domains (Table S7), both of which are NTPase domains and such combinations are commonly found in nucleotide-binding-oligomerization-domain like receptors (NLRs). A maximum likelihood phylogeny of the NACHT domain proteins shows different expansions of NLR subfamilies in fungi (Fig. S2). In particular, the C2-NACHT-WD40(n) subfamily has only been found exclusively in a few basidiomycetes (Van der Nest et al., 2014) and is present in the highest copy number in *P. noxius* and *P. lamaensis*. Other expansions include UDP-glucuronosyltransferases, which catalyse conjugation reactions in the biotransformation of xenobiotics released by its host environment (Jancova, Anzenbacher, & Anzenbacherova, 2010).

*P. noxius displays a bipolar heterothallic mating system*
Determination of mating loci and reproductive mode is considered high-priority in fungal genome analyses as they are the primary determinants of how a fungus expands and generates diversity. Mate recognition of sexual reproduction in Basidiomycota is known to be controlled by two unlinked loci, named as \( A \) and \( B \) locus. A conserved head-to-tail orientation of \( HD1-HD2 \) in \( HD \) pair 1 was found in Hymenochaetales (Fig. 3) which is different from in most species in Agaricomycetes (James et al., 2013). Alignment of 100-kb sequences upstream and downstream of \( A \) locus in \( P. noxius \) isolates KPN91, 718-S1 and A42 revealed that \( A \) locus is highly polymorphic (\( HD \) pairs in particular) despite well-conserved flanking regions (Fig. S3). For \( B \) locus, only one \( STE3 \) encoding seven transmembrane helices was identified; four pheromone genes were identified but not physically linked to the highly monomorphic \( STE3 \) (Fig. S4, Table S8, Text S1), which is characteristic of a bipolar mating system.

To further confirm this observation, allele diversity was analyzed by resequencing the \( A \) and \( B \) loci from 10 single-basidiospore isolates originating from a single basidiocarp (Table S9). Sequencing of \( STE3 \) revealed two highly similar alleles (b1, b2), with 99.5% amino acid identity. The only variant (244V/A) is considered a semi-conservative mutation (valine and alanine are nonpolar aliphatic acids) and may have minor or no effect on protein function. Previous studies have shown that although \( STE3 \) is not involved in mating type determination in bipolar fungi, variations can still be observed (James, Lee, & van Diepen, 2011; James, Srivilai, Kues, & Vilgalys, 2006; Niculita-Hirzel et al., 2008). Primer walking of \( A \) locus...
revealed two distinct haplotypes which suggests a heterothallic bipolar reproductive mode in *P. noxius* (Fig. S5): the a1 and a2 alleles of HD pair I shared an overall 78% nucleotide identity; the two alleles of HD pair II were highly divergent and the HD1 domain in a2 allele contains a 1-bp and a 9-bp deletions and has become a pseudogene. The loss of HD1 domain was also found in the HD pair I of 718-S1 (Fig. 3), suggesting that at least one of the HD1 motifs in A locus is dispensable for a functional HD1-HD2 heterodimer in *P. noxius*. The presence or absence of specific HD domains reflects phylogenetic characteristics and has been commonly observed in fungi (Kües, 2015).

*P. noxius contains a highly polymorphic A locus*

Distinct from all the other fungi, *P. noxius* has an exceptionally highly expanded A locus across a ~60-kb region (Fig. 3). There are two pairs of HD1-HD2 gene located ~50-kb away from each other. In the case of *P. noxius* KPN91, 14 genes were annotated at the A locus (Fig. 3). Orthology analysis placed these genes into 9 families that are also present in other fungi (Fig. 4). However, in the genomes of other species, the majority of homologous members of different families are not located in proximity to each other (Fig. 4). This indicates that the large interval in *P. noxius* is a result of accumulating transposition of genes from conserved fungi families not in proximity to A locus (Fig. 4). Interestingly, different *P. noxius* isolates display different interval lengths at A locus, suggesting that transposition events may be frequent (Fig. 3). Whether transposition plays an additional role in the mating system remains to be clarified.
Sequence analysis of 60 P. noxius strains

To further understand the regional dissemination of *P. noxius*, we sequenced the genomes of 60 isolates (Fig 5A) originating from diseased trees across 13 Taiwan and Japanese offshore islands from 2007-2014. This collection was sequenced to a median depth of coverage of 35X. To characterize chromosomal synteny relationship, additional mate pair libraries were sequenced for two isolates (718-S1 and A42). An average of 96% mapping rate was achieved after aligning reads from each strain to the KPN91 reference genome. The descriptions and statistics of the strains are provided in Table S4.

Diversification of P. noxius across Pacific Ocean islands

We hypothesized that there may be three populations segregating in this area of Pacific Islands: Taiwan, Ryukyu and Ogasawara islands. To examine the population structure of *P. noxius* in this area, we performed principal-component analysis (Fig. 5B) on the SNP variants from individual phased haploid genomes. The major principal component divided Ogasawara islands samples from the rest of the samples, which are located geographically 1,210 km apart. The second and third principal components defined a tight cluster of Taiwanese isolates with considerable overlaps between cities despite being 170 km apart. Most isolates from Ryukyu islands can be differentiated from those from Taiwan (Fig. S6). Interestingly, we identified three isolates from Taiwan that are distinctive from both the Taiwan-Ryukyu and
Ogasawara island clusters, suggesting a possibility of more genetically distinct populations within the Taiwan island.

We constructed a maximum likelihood phylogeny based on 1,837,281 high confidence SNPs of *P. noxius* (Fig. 5C). Concordant with the PCA, the nine isolates from Ogasawara islands form a distinctive lineage, while the 51 Taiwanese and Ryukyu islands isolates are grouped together forming another major lineage. We inferred the population structure in 144,426 unlinked bialleic sites using the Bayesian model-based clustering approach implemented by fastSTRUCTURE (Raj et al., 2014). Consistent with the phylogeny, two Taiwan-Ryukyu and Ogasawara lineages were identified and circulating in this region (highest likelihood with \( K = 2 \); Fig. 5C).

Higher \( K \) values show that the isolates from Ogasawara islands remain one cluster, while the rest of the isolates (independently of their geographical origins) are grouped into genetically homogeneous clusters with little admixture (Fig. S7).

Further inspection of the phylogeny indicated the possibility of gene flow in this region despite physical separation by the sea. Within the Ogasawara clade, the isolates can be further grouped by their geographical origin (two main islands: Hahajima and Chichijima island separated by 50km) with the exception of KPN334 strain. Strain KPN260 was collected from Anijima island which is geographically in proximity to Chichijima island but was grouped with the Hahajima island isolates. Strains collected from within Amami and Okinawa islands are not grouped together on the phylogeny, suggesting independent origins of *P. noxius* infection in both
regions. The pattern of gene flow was more apparent in the 42 isolates collected in two cities of Taiwan, where the isolates were not grouped in the phylogenetic tree according to their geographical origins.

Stable structural variation and intra-isolate polymorphism

The detection of frequency of heterokaryons and polyploids of *P. noxius* in nature remain challenging as its arthospores contain 1-5 nuclei (based on quantification of 145 arthospores from A42 in this study) and multiple allelic fragments in SSRs (2-4 alleles per locus) are commonly found in populations (Akiba et al., 2015; Chung et al., 2015). All the strains were isolated from either a single arthospore or fungal mat, and thus allowed for the analysis of variation in genome structure. We found no deviation in coverage across every scaffold in all strains (Fig. S8), suggesting a stable number of chromosome copies in *P. noxius*. To distinguish the frequency of mono- or heterokaryons in nature, heterozygous SNPs and minor allele frequency (MAF) distribution were inferred in each strain (Fig. 6). We identified four groups (A to D) that clearly differ in heterozygosity and MAF. The group A with the lowest heterozygosity of averaging 0.2 % includes strains 718-S1 and A42, each isolated from a single basidiospore implying they are monokaryotic in nature. All strains in this group exhibited a flat MAF distribution, suggesting that there are spontaneous mutations segregating during the growth of monokaryons that originated at different times. The occurrences of monokaryons are 11% and 47% in the Ogasawara and Taiwan/Ryukyu lineages, respectively. The prevalence of monokaryotic isolates
suggests the involvement of basidiospores in disease spread and that heterokaryosis is not required for pathogenicity.

58% of strains display a much higher heterozygosity, with strain Pn103 having a 1.7% heterozygosity displaying a peak around 50% in MAF distribution. These strains can be further grouped into three categories with distinct MAF profile and heterozygosities (Fig. 6). The largest group of the three (n=19; group D of Fig.6) has on average 1.6% heterozygosity displaying a peak around 50% in MAF distribution, suggesting the presence of two genetically distinct nuclei within the population, i.e., dikaryons. The remaining two groups did not follow a typical di-karyon MAF distribution, as they peaked around 27%-33% and exhibited different heterozygosities. The deviations may be associated with the number and composition of multiple nuclei in a cell. However, nuclear quantification of seven randomly selected isolates showed no differences between the groups (Fig. S9). These groups may refer to different compositions of two or more genetically distinct nuclei.

**High nucleotide diversity in P. noxius populations**

The de novo assemblies of three strains of *P. noxius* have on average 97% nucleotide identity and are largely co-linear to each other with apparent genome translocations between Taiwan/Ryukyu (A42 and KPN91) and the Ogasawara (718-S1) isolate (Fig. S10). We quantified sequence diversity using θs and θπ and categorised them into whole genome average, four-fold synonymous sites and replacement sites (Table 1). A large number of segregating SNPs are found in the genome of *P. noxius*, averaging
one SNP identified in every 20-59 bp. Nucleotide diversity at synonymous sites \( \theta_{\text{syn}} \) is 15.8 and 19.2 per kilobase in Ogasawara island and Taiwan-Ryukyu populations, respectively. This is \(~5\) fold higher when compared to the majority of species (Leffler et al., 2012) and is likely an underestimate of true diversity as indels were not considered. Taiwan-Ryukyu lineage has a much higher diversity than Ogasawara which is not due to sample size difference; the same was observed when we reanalyzed with only 9 randomly chosen strains in the Taiwan-Ryukyu populations. Extremely high diversity has been reported in natural populations of \textit{Schizophyllum commune} (Baranova et al., 2015). The Tajima’s D is strongly negative in the Taiwan and Ryukyu islands lineage, suggesting an excess of low frequency alleles present in the population possibly as a result of high mutation rate (Baranova et al., 2015) or population expansion. In the Ogasawara population, nucleotide diversity is reduced compared to the Taiwan/Ryukyu lineage and Tajima’s D is positive implying a reduced effective population size; \textit{P. noxius} may have been introduced recently in these islands. Together our data suggest that the two \textit{P. noxius} lineages may have derived from genetically distinct gene pools and have undergone divergent evolutionary scenarios, possibly as a result of different time of introduction, different environments, and human interference in Taiwan-Ryukyu vs Ogasawara areas.

\textit{Candidates for population differentiation}

The Taiwan/Ryukyu strains were mainly isolated from diseased trees in urban settings, while the Ogasawara strains were from trails in natural parks (Table S4). Genomic
pairwise $F_{ST}$ revealed moderate differentiation (0.12) between the two genetic lineages of *P. noxius*, which is in accordance with both the phylogeny and fastSTRUCTURE analyses. We identified genomic regions that may contain potential candidates for such environmental origins by investigating 99.9% tail for observed $F_{ST}$ 5-kb windows (Fig. 7). This definition revealed 13 outlier regions. Gene ontology analyses of 42 genes within these regions did not reveal any GO enrichment terms, indicating the sites showing evidence of differentiation may be involved in different functions. Interestingly, the largest region spanning 12 kb at scaffold 6 contains four genes. Homologs from three of the genes are implicated in fungal cell wall organization and fruiting body formation (PNOK_0653900; (Szeto, Leung, & Kwan, 2007)), salt tolerance (PNOK_0654000; (Steffens, Brautigam, Jakoby, & Hulskamp, 2015)), and plant cell wall degradation (PNOK_0654100; xylanase). These are all possible drivers for population differentiation (Apse, Aharon, Snedden, & Blumwald, 1999).

**Discussion**

Here we report four high-quality genome sequences of Hymenochaetales species that are global tree pathogens of particular importance. To date chromosome-level assemblies are available only for a few basidiomycetous species (Manuel Alfaro et al., 2016; Foulongne-Oriol et al., 2016; J. E. Stajich et al., 2010) including *P. noxius* (Table S5). Orthologous relationships with other complete genomes of basidiomycetes have confirmed conservation of karyotypes with few fusion or breaks.
in Agaricomycetes. Our study has shown the diversity and abundance of CAZymes, in particularly lignin-degrading enzymes, in the genomes of Hymenochaetaceae species. Such genetic architecture has been demonstrated in other white rot fungi (Riley et al., 2014), and differences in CAZymes have been implicated as the genetic basis of softwood or hardwood utilization (Suzuki et al., 2012). As revealed by comparative genomics analysis, *P. noxius* has highly expanded 1,3-beta-glucan synthase genes and abundant CAZymes, making it an aggressive wood decay pathogen of a wide variety of both broadleaved and coniferous trees. The results can serve as a starting point for understanding the ecological role of *P. noxius*. Our study also identified high copy numbers of C2-NACHT-WD40(n)-containing NLRs and candidate genes associated with population differentiation. The NLR family is highly diverse in fungi and was found to be central to the process of programmed cell death and implicated in fungal vegetative incompatibility and general nonself recognition (Bidard, Clave, & Saupe, 2013). Analysis of global transcription at different pathogenesis stages and detailed functional assays will help resolve their functions.

Population genomics analyses of *P. noxius* suggest that it is a hypervariable species. Our investigations into mating type loci and genome-wide heterozygosity further indicated that the genetic hyperdiversity can be attributed to a bipolar heterothallic reproductive system and heterokaryotic nature, though gene flow and/or high mutation rate may also play some role. The characteristic large interval between *HD* pairs has only been reported in *Schizophyllum commune* (~55 kb) and *Flammulina velutipes* (~70 kb), in which the genomic separation likely emerged through
inversions or transpositions of gene clusters surrounding $HD$ (van Peer et al., 2011). This exceptional large separation between $HD$ pairs would allow higher probability of recombination between the physically distant $HD$ genes (James et al., 2013; van Peer et al., 2011), thus resulting in progeny with more diverse mating types which are ready to mate. Notably, both monokaryotic and heterokaryotic state of $P. noxius$ mycelia are prevalent in the nature, and some isolates likely contain more than two genetically different nuclei (Chung et al., 2015). In addition, some isolates were able to produce basidiocarps by themselves when cultured on sawdust medium (P.-J. Ann, pers. comm.), suggesting dikaryotization or homokaryotic fruiting occurred spontaneously or in response to certain conditions (Wendland, 2016). It would be interesting to further investigate the complex regulatory mechanisms underlying anastomosis, karyogamy, and meiotic division during vegetative growth and basidiocarp formation. Transcriptional diversity among genetically variable individuals is also warrant further exploration.

How $P. noxius$ is spread in regions of Asia has been the focus of a few studies (Akiba et al., 2015; Chung et al., 2015; Hattori, Abe, & Usugi, 1996). Genomic analysis of the strains from across Taiwan and offshore islands of Japan allowed us to re-examine possible mode of disease transmission in the Asia Pacific region. The pattern of gene flow within and between islands suggested that human activity such as planting of infected seedlings may have promoted the movement of $P. noxius$, which provides an artificial environment for population to increase. The within-population hyperdiversity, prevalence of monokaryotic isolates, and sporadic pattern of new
disease foci also support the involvement of basidiospores in *P. noxius* dissemination. Considering its host range and fast-growing ability in warm weather, and the abundant basidiospores that can be produced from perennial fruiting bodies with huge dispersal potential (e.g., the basidiospores of *Heterobasidion annosum* and *Peniophora aurantiaca* were captured 50–500 km and ~1000 km apart from the inoculum source), *P. noxius* may potentially affect more agricultural, ecological, and residential environments. A preliminary model based on 19 bioclimatic variables of known locations of ~100 *P. noxius* isolates from south eastern Asia, Australia, and Pacific islands predicted that extensive global regions are at risk, which includes a big part of the South American continent (Klopfenstein et al., 2016).

The Hymenochaetales is phylogenetically placed between the better-studied Agaricales and Ustilaginales orders, making the reference assembly of *P. noxius* KPN91 an attractive genome to study the evolutionary transition between these orders of Basidiomycota. It should be of continuous interest to confirm this observation when more reference genome assemblies become available. Genetic hyperdiversity of *P. noxius* suggests that the pathogen may have greater adaptability to different environments and stresses. Associating growth phenotypes under a variety of conditions to the variations identified in this study will be a logical next step for a full understanding of the genetic basis underlying the fitness and virulence of white root rot fungi. For disease control and prevention, much more attention needs to be paid to monitor how these fungi will behave in changing or extreme weather conditions.
Figures and Tables

Fig 1. Life stages of *P. noxius* and comparative genomic analysis of Hymenochaetales species. (a) The mycelial mat with young creamy leading front and aged brown section. (b) In advanced stage of decay, the hyphae form a network of brown zone lines permeating the soft and white wood tissue. (Lower left and lower right) (c and d) Basidiocarps are perennial and can be resupinate (c) or grow into a sessile bracket-like conk with a broad basal attachment (d). The distinctive greyish-brown surface is the hymenial layer with irregularly polygonal pores, containing four-spored basidia, ellipsoid and hyaline basidiospores, but no hymenial setae. (e) The phylogeny of four *Phellinus* species with 15 other species of Basidiomycota based on a concatenated alignment of single-copy orthologous genes. All nodes have 100 out of 100 bootstrap replicates. The numbers of gained (“+”) and lost (“-”) gene families along each branch of the phylogeny is annotated.

Fig 2. Linkage group (LG) network of Basidiomycota. (a) The linkage groups were identified by linking single-copy orthologs of scaffolds between species pairs. All scaffolds included in this plot are larger than 500 kb. Each directional edge points toward the reference chromosome. Edges were weighted by the number of single-copy orthologs, and an edge was filtered if it has a weight smaller than 20 or less than 40% of the sum of all weights out of a node. (b) Cross-mapping of single-copy orthologs in LG10 and 11. Scaffold names are shown at the upper-right side of each sequence, with detectable telomeric regions labelled as ‘T’ at the upper-left side.

Fig 3. Synteny around the A mating locus. The analysis included *Phellinus noxius*, *Porodaedalea pini*, *Phellinus sulphurascens*, *Phellinus lamaensis*, two other species
from Hymenochaetales species (*Fomitiporia mediterranea, Schizopora paradoxa*), two species from Agaricales (*Laccaria bicolar, Coprinopsis cinerea*), *Sistotremastrum niveocremeum*, and *Ustilago maydis*.

**Fig 4. Orthology network of genes in the A locus of P. noxius.** Each node in this plot represents a gene. An edge is added if two genes are in proximity (physically separated by less than two genes). Numbers next to the edges and nodes are number of occurrences of different combinations of genes in each species. For example, 77 members of OG0000015 orthogroup in *P. lamaensis* are found adjacent to each other on the genome. Another 37 members of this orthogroup are dispersed throughout the genome and are not located in proximity to any members of the 9 orthogroups.

**Fig 5. The population genomics of 60 P. noxius isolates.** (a) Map of Taiwan and offshore islands of Japan showing origins of the 60 sampled *P. noxius* isolates. Ogasawara islands were conveniently drawn below Ryukyu islands so this does not represent their actual location. (b) PCA plot of 60 *P. noxius* isolates using genome-wide variation data sampled from 13 islands by the first three eigenvectors. (c) Top: Phylogenetic tree with 100 bootstrap using SNPs computed from alignment to KPN91 reference. Point separating the Taiwan-Ryukyu and Ogasawara island isolates was used as root. Nodes with >90% bootstrap were labelled with circles. Bottom: fastSTRUCTURE analysis of the linkage independent pruned set of variation data. A model with two ancestral components (*K*=2) had the highest likelihood to explain the variation of genome-wide structure on the 60 isolates. Also see Fig. S7 for different *K*.

**Fig 6. Heterozygosity and allele frequencies of P. noxius.** (a) Boxplot showing abundances of heterozygous SNPs inferred in each *P. noxius* isolate can be categorised into four groups. (b) Density plot of minor allele frequencies (MAF) of heterozygous SNPs in 60 isolates of *P. noxius*. * Asterisk denote group containing two isolates that were cultured from basidiospores.
Fig. 7 Weighted F_{ST} values for 5kb windows across the P. noxius assembly. Red colour dots indicate windows having F_{ST} value greater than 99.9% tail of 0.55. The 12-kb candidate region is marked in blue star. Annotations and references of genes located in this candidate region are listed.

Table 1. Polymorphism in the two regional lineages of P. noxius.
Availability of data and materials:

Genome assemblies and annotations are available from NCBI under whole genome shotgun (WGS) ID: NBII00000000 (P. noxius), NBBA00000000 (P. sulphurascens), NBAY00000000 (P. pini) and NBAZ00000000 (P. lamaensis). Bioproject and Biosample ID of raw data are available in Table S1 and S2.

Author’s contributions:


Acknowledgements

We thank John Wang (Biodiversity Research Center, Academia Sinica), Yen-Ping Hsueh (IMB, Academia Sinica) and Robert Waterhouse (University of Lausanne) for commenting the manuscript. We thank Chian-mei Wei, Kuan-chun Chen (High Throughput Genomics Core at Biodiversity Research Center, Academia Sinica) for sequencing. We thank Tun-Tschu Chang and Yu-Ching Huang for strain collection and consultation.

Funding:

I.J.T. and J.T.L were funded by Ministry of Science and Technology (105-2628-B-001-002-MY3 and 105-2313-B-001 -003). C.L.C., H.H.L. and Y.Y.C were funded by the Office of General Affairs, National Taiwan University.
References


<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Product Description</th>
<th>Reference of homologs</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNOK_0653900</td>
<td>glycosylphosphatidylinositol-anchored</td>
<td>Szeto et al., 2007</td>
</tr>
<tr>
<td>PNOK_0654000</td>
<td>BEACH-domain-containing protein</td>
<td>Steffens et al., 2015</td>
</tr>
<tr>
<td>PNOK_0654100</td>
<td>endo-1,4-beta xylanase</td>
<td>Bray and Clark (1994)</td>
</tr>
<tr>
<td>PNOK_0654200</td>
<td>hypothetical protein</td>
<td>N/A</td>
</tr>
<tr>
<td>Population</td>
<td>Analyzed Sites</td>
<td>Segregating sites (S)</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>----------------</td>
<td>------------------------</td>
</tr>
<tr>
<td><strong>Whole genome</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ogasawara (n=9)</td>
<td>30,028,158</td>
<td>506,231</td>
</tr>
<tr>
<td>TaiwanRyukyu (n=51)</td>
<td>31,376,691</td>
<td>1,538,598</td>
</tr>
<tr>
<td>TaiwanRyukyu subset (n=9)</td>
<td>30,640,276</td>
<td>749,681</td>
</tr>
<tr>
<td><strong>Synonymous Sites</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ogasawara (n=9)</td>
<td>4,651,751</td>
<td>182,879</td>
</tr>
<tr>
<td>TaiwanRyukyu (n=51)</td>
<td>4,758,211</td>
<td>513,717</td>
</tr>
<tr>
<td>TaiwanRyukyu subset (n=9)</td>
<td>4,704,161</td>
<td>269,060</td>
</tr>
<tr>
<td><strong>Replacement Sites</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ogasawara (n=9)</td>
<td>9,794,069</td>
<td>71,497</td>
</tr>
<tr>
<td>TaiwanRyukyu (n=51)</td>
<td>10,018,114</td>
<td>230,689</td>
</tr>
<tr>
<td>TaiwanRyukyu subset (n=9)</td>
<td>9,904,211</td>
<td>103,907</td>
</tr>
</tbody>
</table>

1 A random of nine isolates were chosen to repeat the analysis.